Generating full-length Killer-cell Immunoglobulin-like Receptor (KIR) gene sequences using Third Generation long-amplicon sequencing to assess the impact of KIR polymorphism on Haematopoietic Cell Transplantation outcomes

Will P Bultitude

Anthony Nolan Research Institute, London, UK

University College London, UK

For the award of PhD

I, William Paul Bultitude, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

Throughout my time at Anthony Nolan, I have received help from far more people than I can acknowledge in this brief statement. However, thanks should first go to Prof Steven GE Marsh, who has managed my research over the last eight years and without whom I would not have had this opportunity. You have always offered entertaining advice on topics from all aspects of life and believed in my abilities from the start.

Enormous gratitude also goes to Dr Neema Mayor, without whom I would almost certainly have got lost down a rabbit hole from which never to return. You have guided me and my research with incredible calm, a skill that I hope to have adopted at least in part. But thanks most for ensuring I never felt lost or in the dark in this project.

There were several people who suggested that I avoid a KIR sequencing PhD project at all costs. Although I hope this thesis goes some way to prove that it can be done, the same group of people also checked in to make sure my research was on track and moving forwards. Your humour and support were always hugely appreciated; I wouldn't have made it to the end without it. To my all my other colleagues, past and present, thank you. Because of you, work has never felt like a chore to me. I wish you all, and everyone at Anthony Nolan, good luck in the future.

My final thanks go to my family. Your support ranged from sustained apparent interest in the minutiae of KIR alleles to putting a roof over my head for several of the years I was a PhD student. To my wife, Ana, despite all the late nights, missed dinners and delays to our plans, I could always rely on your unwavering support of this PhD, even when it felt like it would never end. From the bottom of my heart, thank you.

Abstract

Killer-cell Immunoglobulin-like Receptor (KIR) polymorphism is extensive in both allelic and copy number variation. Although multiple assays have been designed to assess the latter form of polymorphism, KIR allelic diversity is less well understood owing to the homologous nature of different KIR genes and, until recently, limitations in sequencing technology. To better understand KIR allelic diversity in the UK, and its impact on haematopoietic cell transplant (HCT) outcomes, I have designed and validated a whole-gene, fully-phased allele sequencing strategy that encompasses third generation sequencing technology to deliver unambiguous genotypes for several different KIR genes. Subsequently, this strategy was applied to a novel, largely T cell deplete UK cohort of patients receiving HCT to treat acute myeloid leukaemia, and their respective donors. This assay utilises a semi-generic targeted polymerase chain-reaction amplification prior to multiplexed library preparation and sequencing, providing a relatively high-throughput methodology amenable to clinical laboratories. Initially, to assess the relevance of presence/absence KIR polymorphism on HCT outcomes, I utilised existing genotyping methods to establish baseline characteristics. In contrast to previous publications, relapse was largely unaffected by KIR polymorphism. However, striking differences related to preparative conditioning regimen were observed in KIR-mediated effects in other HCT outcomes. Presence of donor-encoded centromeric (Cen)-B motifs relayed increased risk of detrimental non-relapse mortality following myeloablative conditioning, whereas the opposite appeared to be true of reduced-intensity conditioning transplants. When the impact of allelic diversity at the KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 loci on HCT outcomes was assessed in my cohort, allelic differences within the Cen-A, Cen-B and telomeric A haplotype motifs provided additional insight into the possible mechanisms of KIR-mediated influence on HCT

outcomes. By estimating frequencies of different KIR alleles within a UK population, I have demonstrated that donor selection algorithms could feasibly incorporate allelic KIR genotypes that may improve quality of life after HCT.

Summary of key findings

Single Molecular Real-Time long amplicon sequencing is an accurate method to determine Killer-cell Immunoglobulin-like Receptor (KIR) allele types (Chapter 4, Section 4.05, p.250-267)

Having developed a long-range PCR approach to specifically amplify KIR genes, I validated the technique using a variety of previously characterised cell lines. This demonstrated high concordance to existing allele types. Furthermore, it was possible to identify and correct several errors within previous genotyping, demonstrating the accuracy and suitability of SMRT sequencing for KIR allele typing.

Haematopoietic cell transplant recipients' conditioning regimen influences the effect KIR polymorphism has on transplant outcomes (Chapters 3 and 6)

The influence of KIR polymorphism was assessed in cohorts stratified by conditioning regimen (myeloablative vs reduced intensity). The effect of polymorphism was not consistent between these groups, and often had opposing consequences (detrimental in one whilst beneficial/neutral in the other). This highlights the need to encompass all available recipient information when selecting donors based on KIR genotype information.

Donor Centromeric (Cen) B content associates with detrimental outcomes in recipients of haematopoietic cell transplant following myeloablative conditioning (Chapter 3, Section 3.10.5, p.175-181)

Despite previous analysis implicating relapse protection with donor Cen-B content in a cohort from the USA, an equivalent analysis using this UK cohort links donor Cen-B

content with increased infection-related mortality at one year post-myeloablative transplant (Hazard Ratio [HR]=5.5, 95% Confidence Interval [CI]=1.5-20.3, p=0.01). This also significantly reduces overall and disease-free survival at five years post-transplant (OS: HR=1.89, CI=1.16-3.13, p=0.01; DFS: HR=1.75, CI=1.09-2.78, p=0.02).

Allelic differences within donors' centromeric motifs distinguish beneficial, neutral and detrimental genotypes for myeloablative conditioning haematopoietic cell transplant recipients (Chapter 6, Section 6.06.01, p.418-420)

Given the importance of the centromeric motif evidenced above, the common KIR2DL2/3~KIR2DL1 allelic haplotypes were assessed to determine their effect on HCT outcomes. This revealed that recipients of myeloablative conditioning transplants utilising donors encoding at least one copy of the KIR2DL3*00101~KIR2DL1~00302 motif were significantly less likely to suffer non-relapse mortality at one year post-transplant (16% *vs* 36%, p=0.003). Furthermore, presence of at least one copy of the KIR2DL2*00101~KIR2DL1*00401 motif within equivalent donor genotypes correlated with increased non-relapse mortality at one year post-transplant (43% *vs* 21%, p=0.01). This, in turn, also conferred a detrimental risk to overall survival probability at five years post-transplant (22% *vs* 40%, p=0.02). Stratification of the myeloablative conditioning cohort according to the presence of other Cen-A or Cen-B allele haplotypes showed no significant differences in any of the outcomes assessed.

Donors encoding R238 KIR3DL1 alleles relay greater protection against relapse but increased risk of non-relapse mortality in recipients of myeloablative conditioning haematopoietic cell transplant (Chapter 6, Section 6.05.03.03, p.408-410)

In addition to centromeric allelic variation, polymorphism at the KIR3DL1/S1 locus of donors also influenced the risk of non-relapse mortality. The high functionality variant, R^{238}, was associated with increased non-relapse mortality at one year post-transplant (HR=2.9, CI=1.5-5.5, p=0.002). This detrimental effect was, in part, balanced by a borderline significant reduction in five year relapse probability (HR=0.5, CI=0.3-1.0, p=0.05). As such, selection of donors encoding this polymorphism for patients with high relapse risk may still confer some benefit.

Donor-recipient KIR genotype and allele mismatching in graft-versus-host direction confers a detrimental impact on overall survival, non-relapse mortality and acute graft-versus-host disease in myeloablative conditioning haematopoietic cell transplant recipients (Chapter 3, Section 3.10.04, p.167-175; Chapter 6, Section 6.07.03, p.429-432)

When considering KIR genotype matching status, donors and recipients who were mismatched in the graft-versus-host direction at all genes excluding KIR2DL2/S2 were significantly less likely to survive (HR=1.9, CI=1.1-3.2, p=0.02) and more likely to suffer non-relapse mortality (HR=5.9, CI=1.7-20.5, p=0.005). When KIR genotypes were compared at allele resolution, donors and recipients mismatched in the graft-versus-host direction for KIR2DL1 were associated with more severe acute graft-versus-host disease (HR=8.4, CI=2.3-30.8, p=0.001), although graft-versus-host

direction mismatching at KIR2DL2/3 or KIR3DL1/S1 did not confer significant differences in any transplant outcome assessed.

Donor-recipient KIR allele mismatching in graft-versus-host direction improves overall and disease-free survival in recipients of reduced intensity conditioning haematopoietic cell transplant (Chapter 6, Section 6.07.02, p.427-429)

Despite evidence suggesting KIR2DL1 allele mismatching following myeloablative conditioning haematopoietic cell transplantation conferred detrimental acute graft-versus-host disease probability, this was not observed in the reduced intensity conditioning cohort. Instead, high frequency (more than 3) KIR allele graft-versus-host mismatching across the KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 loci conferred a beneficial increase in five year overall survival (HR=0.4, CI=0.2-0.8, p=0.01) and disease-free survival (HR=0.5, CI=0.3-0.9, p=0.02).

Selection of haematopoietic cell transplantation donors to incorporate beneficial and avoid detrimental KIR polymorphisms is eminently feasible (Chapter 5, Sections 5.03.02 and 5.05.02, p.282-286 and p.304-307)

As discussed above, polymorphisms at both the KIR2DL1 and KIR3DL1/S1 loci conferred significant differences on haematopoietic cell transplantation outcomes. Several such polymorphisms associated with beneficial outcomes exhibit high frequency. For instance, the L^{114}~R^{245} motif defines the KIR2DL1*003-like allele group. KIR2DL1*00302 was the most common allele in this cohort and has allele frequency of approximately 40%. As such, selection of donors encoding this allele is a viable approach to reducing non-relapse mortality risk in myeloablative conditioning haematopoietic cell transplant recipients. Furthermore, the KIR3DL1 R^{238} polymorphism, found only in KIR3DL1*002, was also identified at high frequency;

exhibiting allele frequency of 12%. Deliberate selection to avoid this polymorphism, and the associated increase in non-relapse mortality probability, would also be straightforward. To improve survival in reduced intensity conditioning haematopoietic cell transplant recipients, selection of donors encoding multiple graft-versus-host direction allele mismatches may be desirable. As a direct result of KIR loci exhibiting high levels of allelic diversity, this is also entirely possible.

More detailed suggestions for adaptations to incorporate KIR genotype information into donor selection algorithms are given in Chapter 7.

Impact statement

The study of KIR polymorphism to date has predominantly focussed on the copy number variation of these genes into haplotype motifs and extended haplotypes. Despite efforts to resolve functional differences between different KIR alleles at the molecular biology level, much is yet to be learned regarding KIR biology. Furthermore, allelic resolution KIR genotyping has rarely been translated into clinical science, perhaps contributing to the conflicting reports within the current literature. As such, my development of a strategy to determine unambiguous allelic KIR genotypes has potential to impact several sectors. The significant differences in outcomes of haematopoietic cell transplantation (HCT) observed when comparing different alleles and allele matching scenarios, even within the modest cohort described herein, demonstrate the potential impact this technology may have on patients' quality (and quantity) of life post-transplant. Further clinical benefits will undoubtedly arise when more is understood of KIR biology, a factor that requires a more in-depth knowledge of the complex genetics of this hyperpolymorphic locus and interacting gene systems.

I anticipate that adoption of this genotyping will also improve the understanding of the extent and function of KIR polymorphism. In addition to transplantation, which is discussed in depth in the main text, potential benefits may also be observed within natural killer (NK) cell research topics, including pregnancy and immunotherapy. Additionally, by sequencing many thousands of samples from individuals of different ethnicities, I foresee an even greater range of KIR alleles will be discovered, helping to determine key functional residues within the proteins and offering insight into evolution across this intriguing locus. I demonstrate interesting distinctions of allelic polymorphism between the three KIR loci studied in-depth but anticipate that far more

is yet to be uncovered in broader populations and the remaining KIR loci. This may reveal markers of human migration and evolution, as well as other potentially clinically relevant factors such as disease association indicators.

Finally, by performing genotyping and clinical analysis in a novel UK cohort, in which HCT protocol differs to much of the rest of the world, I expect this research to have both national and international importance. National benefit may arise through observation that existing donor selection strategies, modelled and appropriate in the USA, are not necessarily applicable within the UK HCT setting. International benefit may develop from my theoretical analysis relating to how the impact of KIR polymorphism on HCT outcomes differs by transplantation protocol. This may guide future prospective studies anywhere in the world to encompass KIR genotyping within transplant protocol-dependent donor selection.

To disseminate this impact, I have presented my findings at national (British Society of Histocompatibility and Immunogenetics) and international conferences (European Federation for Immunogenetics, International HLA and Immunogenetics Workshop and the KIR Workshop). Preparation of manuscripts detailing the results for publication within peer-reviewed journals is underway.

Publications

During my time at Anthony Nolan, I was fortunate enough to publish several manuscripts and abstracts, as well as contributing as co-author on several others. In addition, the research output of this PhD is currently being prepared into manuscripts that we hope to publish in the near future. A full list of publications and planned manuscripts is provided below:

Manuscripts:

Bultitude WP, Schellekens J, Szydlo RM, Anthias C, Cooley SA, Miller JS, Weisdorf DJ, Shaw BE, Roberts Ch, Garcia-Sepulveda CA, Perry J, Pearce RM, Wilson MC, Potter MN, Byrne JL, Russell NH, MacKinnon S, Bloor AJ, Patel A, McQuaker G, Malladi R, Tholouli E, Orchard K, Potter VT, Madrigal JA, Mayor N, Marsh SGE "Presence of donor-encoded centromeric KIR B content increases the risk of infectious mortality in adult recipients of myeloablative, HLA-matched HCT to treat AML". *Manuscript submitted*

Bultitude WP, Gymer AW, Robinson J, Szydlo RM, Anthias C, Cooley SA, Miller JS, Weisdorf DJ, Shaw BE, Roberts Ch, Garcia-Sepulveda CA, Perry J, Pearce RM, Wilson MC, Potter MN, Byrne JL, Russell NH, MacKinnon S, Bloor AJ, Patel A, McQuaker G, Malladi R, Tholouli E, Orchard K, Potter VT, Madrigal JA, Mayor N, Marsh SGE "The impact of allelic polymorphism at the KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 loci on the outcomes of haematopoietic cell transplantation to treat acute myeloid leukaemia". *Manuscript in preparation*

Mayor NP, Hayhurst JD, Turner TR, Szydlo RM, Shaw BE, Bultitude WP, Sayno JR, Tavarozzi F, Latham K, Anthias C, Robinson J, Braund H, Danby R, Perry J, Wilson MC, Bloor AJ, McQuaker IG, MacKinnon S, Marks DI, Pagliuca A, Potter MN, Potter VT, Russell NH, Thomson KJ, Madrigal JA, Marsh SGE "Recipients Receiving Better HLA-Matched Hematopoietic Cell Transplantation Grafts, Uncovered by a Novel HLA Typing Method, Have Superior Survival: A Retrospective Study". *Biol Blood Marrow Transplant.* (2019) **25(3)** p443-450

Creary LE, Guerra SG, Chong W, Brown CJ, Turner TR, Robinson J, Bultitude WP, Mayor NP, Marsh SGE, Saito K, Lam K, Duke JL, Mosbruger TL, Ferriola D, Monos D, Willis A, Askar M, Fischer G, Loong Saw C, Ragoussis I, Petrek M, Serra-Pages C, Juan Otero M, Stavropoulos-Giokas C, Dinou A, Ameen R, Al Shemmari S, Spierings E, Gendzekhadze K, Morris GP, Zhang Q, Kashi Z, Hsu S, Gangavarapu S, Mallempati KC, Yamamoto F, Osoegawa K, Vayntrub T, Chang CJ, Hansen JA, Fernandez-Vina MA "Next-generation HLA typing of 382 International Histocompatibility Working Group reference B-lymphoblastoid cell lines: Report from the 17th International HLA and Immunogenetics Workshop". *Human Immunology* (2019) **80(7)** p449-460

Osoegawa K, Vayntrub TA, Wenda S, De Santis D, Barsakis K, Ivanova M, Hsu S, Barone J, Holdsworth R, Diviney M, Askar M, Willis A, Railton D, Laflin S, Gendzekhadze K, Oki A, Sacchi N, Mazzocco M, Andreani M, Ameen R, Stavropoulos-Giokas C, Dinou A, Torres M, Dos Santos Francisco R, Serra-Pages C, Goodridge D, Balladares S, Bettinotti MP, Iglehart B, Kashi Z, Martin R, Saw CL, Ragoussis J, Downing J, Navarrete C, Chong W, Saito K, Petrek M, Tokic S, Padros K, Beatriz Rodriguez M, Zakharova V, Shragina O, Marino SR, Brown NK, Shiina T, Suzuki S, Spierings E, Zhang Q, Yin Y, Morris GP, Hernandez A, Ruiz P, Khor SS, Tokunaga K, Geretz A, Thomas R, Yamamoto F, Mallempati KC, Gangavarapu S, Kanga U, Tyagi S, Marsh SGE, Bultitude WP, Liu X, Cao D, Penning M, Hurley CK, Cesbron A, Mueller C, Mytilineos J, Weimer ET, Bengtsson M, Fischer G, Hansen JA, Chang CJ, Mack SJ, Creary LE, Fernandez-Vina MA "Quality control project of NGS HLA genotyping for the 17th International HLA and Immunogenetics Workshop". *Human Immunology* (2019) **80(4)** p228-236

Misra MK, Augusto DG, Martin GM, Nemat-Gorgani N, Sauter J, Hofmann JA, Traherne JA, Gonzalez-Quezada B, Gorodezky C, Bultitude WP, Marin W, Vierra-Green C, Anderson KM, Balas A, Caro-Oleas JL, Cisneros E, Colucci F, Dandekar R, Elfishawi SM, Fernandez-Vina MA, Fouda M, Gonzalez-Fernandez R, Grosse A, Herrero-Mata MJ, Hollenbach SQ, Marsh SGE, Mentzer A, Middleton D, Moffett A, Moreno-Hidalgo MA, Mossallam GI, Nakimuli A, Oksenberg JR, Oppenheimer SJ, Parham P, Petzl-Erler ML, Planelles D, Sanchez-Garcia F, Sanchez-Gordo F, Schmidt AH, Trowsdale J, Vargas LB, Vicario JL, Vilches C, Norman PJ, Hollenbach JA "Report from the Killer-cell Immunoglobulin-like Receptors (KIR) component of the 17th International HLA and Immunogenetics Workshop". *Human Immunology* (2018) **79(12)** p825-833

Bultitude WP, Gymer AW, Robinson J, Mayor NP, Marsh SGE. "The novel KIR2DL1 allele, KIR2DL1*037, defined in the cell line SPO010 (IHW9036)". *HLA* (2018) **91(6)** p547-548

Bultitude WP, Gymer AW, Robinson J, Mayor NP, Marsh SGE. "KIR2DL1 allele sequence extensions and discovery of 2DL1*0010102 and 2DL1*0010103 alleles by DNA sequencing". *HLA* (2018) **91(6)** p546-547

Turner TR, Hayhurst JD, Hayward DR, Bultitude WP, Barker DJ, Robinson J, Madrigal JA, Mayor NP, Marsh SGE, "Single molecule real-time DNA sequencing of HLA genes at ultra-high resolution from 126 International HLA and Immunogenetics Workshop cell lines" (2018) HLA 91 (2), p.88-101

Shaw BE, Mayor NP, Szydlo RM, Bultitude WP, Anthias C, Kirkland K, Perry J, Clark A, Mackinnon S, Marks DI, Pagliuca A, Potter MN, Russell NH, Thomson K, Madrigal JA, Marsh SGE "Recipient/donor HLA and CMV matching in recipients of T-cell-depleted unrelated donor haematopoietic cell transplants". *Bone Marrow Transplant.* (2017) **52(5)** p717-725

Hayward DR, Bultitude WP, Mayor NP, Madrigal JA, Marsh SGE. "The novel HLA-B*44 allele, HLA-B*44:220, identified by Single Molecule Real-Time DNA sequencing in a British Caucasoid male" *Tissue Antigens* (2015) **86(1)** p61-63

Mayor NP, Robinson J, McWhinnie AJ, Ranade S, Eng K, Midwinter W, Bultitude WP, Chin CS, Bowman B, Marks P, Braund H, Madrigal JA, Latham K and Marsh SGE "HLA Typing for the Next Generation". *PLOS ONE* (2015) **10(5)** e0127153

Bultitude WP, Mayor NP, McWhinnie AJ, Madrigal JA, Marsh SGE. "Genomic sequence of the rare HLA-A*02:95 allele identified by sequence-based typing and cloning". *Tissue Antigens* (2014) **84(3)** p324-326

Published abstracts

Bultitude WP, Gymer AW, Robinson J, Anthias C, Potter MN, Russell NH, Pagliuca A, Potter VT, Bloor AJ, Marks DI, Patel A, MacKinnon S, Carpenter B, McQuaker IG, Szydlo RM, Perry J, Wilson MC, Mayor NP, Marsh SGE "The effect of donor KIR2DL1 allelic diversity on the outcomes of HSCT is influenced by conditioning regimen". *HLA* (2019) **94(2)** p122-123

Mayor NP, Hayhurst JD, Turner TR, Szydlo RM, Shaw BE, Bultitude WP, Sayno J-R, Tavarozzi F, Latham K, Anthias C, Braund H, Danby R, Perry J, Wilson MC, Bloor AJ, Clark A, MacKinnon S, Marks DI, Pagliuca A, Potter MN, Russell NH, Thomson KJ, Madrigal JA, Marsh SGE "Better HLA matching as revealed only by Next Generation Sequencing technology results in superior overall survival post-allogeneic haematopoietic cell transplantation with unrelated donors". *Biol Blood Marrow Transplant.* (2018) **24(3)** S63-64

Bultitude WP, Guijarro C, Mayor NP, Madrigal JA, Marsh SGE. "O14: Generation of full- length KIR2DL1/2/3 gene sequences using Single Molecule Real-Time (SMRT) DNA sequencing". *International Journal of Immunogenetics* (2016) **43(5)** p336

Merlo D, Latham K, Mayor NP, Robinson J, Tavarozzi F, Wallis-Jones S, Grewal R, Boix-Giner F, Lepore D, Brosnan N, Bultitude WP, Turner TR, Marsh SGE "Introducing Single Molecule Real-Time (SMRT®) sequencing in a clinical routine setting" *HLA* (2015) **87(4)** p277

Bultitude WP, Mayor NP, Robinson J, Madrigal JA, Latham K, Marsh SGE. "Further optimisation of SMRT DNA sequencing for HLA typing in a high throughput clinical laboratory". *International Journal of Immunogenetics* (2015) **42(5)** p389

Shaw BE, Mayor NP, Szydlo R, Bultitude WP, Kirkland K, Perry J, Clark A, *et al*. "HLA-DQB1 and -DPB1 mismatches are significantly associated with worse survival in unrelated donor haematopoietic stem cell transplantation". *Bone Marrow Transplant.* (2015) **50** S152

Shaw BE, Mayor NP, Szydlo R, Bultitude WP, Kirkland K, Perry J, Wilson M, Clark A, MacKinnon S, Marks DI, Pagliuca A, Potter M, Russell N, Thomson K, Madrigal JA, Marsh SGE "Patient/donor CMV matching is a critical determinant of survival in unrelated donor haematopoietic stem cell transplantation". *Blood* (2014) **124(21)** p1207

Mayor NP, Robinson J, McWhinnie AJ, Ranade S, Eng K, Bultitude WP, Midwinter W, Bowman B, Marks P, Braund H, Madrigal JA, Latham K, Marsh SGE. "Genomic DNA sequences of HLA class I alleles generated using multiplexed barcodes and SMRT® DNA sequencing technology" *Tissue Antigens* (2014) **84(1)** p98

Mayor NP, Robinson J, McWhinnie AJ, Ranade S, Eng K, Midwinter W, Bultitude WP, Chin J, Bowman B, Marks P, Hepler L, Braund H, Madrigal JA, Latham K, Marsh SGE "Generation of multiple full length HLA class I gene sequences in a single sequencing reaction using SMRT DNA sequencing technology" *Tissue Antigens* (2014) **84(1)** p11

Mayor NP, Robinson J, Ranade S, Eng K, Wallis-Jones S, McWhinnie AJ, Bultitude WP, Midwinter W, Bowman B, Hepler L, Braund H, Madrigal JA, Latham K, Marsh SGE "Generation of 252 HLA class I genomic sequences in a single sequencing reaction using DNA barcodes and Single Molecule Real-Time (SMRT) DNA sequencing technology" *Human Immunology* (2014) **75(Supplement)** p49

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Abbreviations

µL	Microlitre
µM	Micromolar
ADCC	Antibody-Dependent Cellular Cytotoxicity
aGVHD	Acute Graft *Versus* Host Disease
AML	Acute Myeloid Leukaemia
ALL	Acute Lymphoblastic Leukaemia
APC	Antigen-Presenting Cells
AT	AlleleTeapot
ATG	Antithymocyte Globulin
ATP	Adenosine Triphosphate
BCR	B Cell Receptor
BCR-ABL	Breakpoint Cluster Region-Abelson Murine Leukaemia viral oncogene homologue 1
BDT	BigDye Terminator™
BM	Bone Marrow
BLAST-n	Basic Local Alignment Search Tool for Nucleotides
BLCL	B Lymphoblastoid Cell Line
bp	Base Pairs
BSBMT	British Society for Blood and Marrow Transplantation
Bu	Busulfan
CAD	Caspase-Activated DNase
cAMP	Cyclic Adenosine Monophosphate
CBT	Cord Blood Transplantation
CBU	Cord Blood Unit
CCD	Charge-Coupled Device
CCS	Circular Consensus Sequencing
CD	Cluster of Differentiation
CDS	Coding Domain Sequence
C/EBP	CCAAT-Enhancer-Binding Protein
Cen	Centromeric
cGVHD	Chronic Graft *Versus* Host Disease
cGRFS	cGVHD and Relapse-Free Survival
CLR	Continuous Long Read
CML	Chronic Myeloid Leukaemia
CMV	Cytomegalovirus
CNV	Copy Number Variation
Conc.	Concentration
CRE	cAMP Response Element
CREB	CRE Binding Protein

Cy	Cyclophosphamide
DAP	DNAX-Activating Protein
DFS	Disease-Free Survival
DLI	Donor Leukocyte Infusion
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic Acid
dNTP	Dinucleotide Triphosphate
DTT	Dithiothreitol
EBMT	European Society for Blood and Marrow Transplantation
EDTA	Ethylenediaminetetraacetic
Exo	Exonuclease
FADD	Fas-Associated Death Domain
Fc	Fragment crystallizable
FBS	Foetal Bovine Serum
Flu	Fludarabine
FLT	FMS-Like Tyrosine Kinase
gDNA	Genomic DNA
GVT	Graft *Versus* Tumour
GVL	Graft *Versus* Leukaemia
GVHD	Graft *Versus* Host Disease
HIV	Human immunodeficiency virus
HLA	Human Leukocyte Antigen
HR	Hazard Ratio
HSC	Haematopoietic Stem Cell
HCT	Haematopoietic Cell Transplantation
HSV	Herpes simplex virus
IHIW	International HLA and Immunogenetics Workshop
IHWG	International Histocompatibility Working Group
ICAD	Inhibitor of Caspase-Activated DNase
IDH	Isocitrate Dehydrogenase
IFN-γ	Interferon gamma
Ig	Immunoglobulin
IL	Interleukin
ITAM	Immunoreceptor Tyrosine-based Activation Motif
ITD	Internal Tandem Duplication
ITIM	Immunoreceptor Tyrosine-based Inhibitory Motif
IUPAC	International Union of Pure and Applied Chemistry
kbp	Kilobase pairs
KIR	Killer-cell Immunoglobulin-like Receptor
LAIR	Leukocyte-Associated Immunoglobulin-like Receptor
LED	Light-Emitting Diode

LILR	Leukocyte Immunoglobulin-Like Receptor
LM	Lower Marker
LRC	Leukocyte Receptor Complex
M	Molar
MAC	Myeloablative Conditioning
MAFFT	Multiple Alignment using Fast Fourier Transform
Mb	Megabase
mg	Milligram
MHC	Major Histocompatibility Complex
MIC	MHC Class I polypeptide-related sequence
minLength	Minimum Length
minSNR	Minimum Signal-to-Noise Ratio
mL	Millilitre
mm	Millimetre
mM	Millimolar
MUSCLE	Multiple Sequence Comparison by Log-Expectation
MW	Molecular Weight
MZF-1	Myeloid Zinc Finger 1
ng	Nanogram
NGS	Next Generation Sequencing
NK	Natural Killer
NKC	Natural Killer Complex
NPM	Nucleophosmin
NRM	Non-Relapse Mortality
OCPW	One-Cell-Per-Well
OS	Overall Survival
PBMC	Peripheral Blood Mononuclear Cell
PBSC	Peripheral Blood Stem Cell
PCR	Polymerase Chain Reaction
piRNA	PIWI-interacting RNA
PIWI	P-element Induced Wimpy testis
pM	Picomolar
QC	Quality Control
QV	Quality Value
RCLB	Red Cell Lysis Buffer
RFU	Relative Fluorescence Units
RIC	Reduced Intensity Conditioning
rpm	Revolutions Per Minute
RPMI	Roswell Park Memorial Institute
rSSOP	Reverse Sequence-Specific Oligonucleotide Probe
SBT	Sequencing-Based Typing

SDS	Sodium Dodecyl Sulphate
SHP	Src Homology region 2 domain-containing protein tyrosine Phosphatase
SIV	Simian Immunodeficiency Virus
SMRT	Single Molecule Real-Time
SNARE	Soluble N-ethylmaleimide-sensitive factor Attachment protein Receptor
SNP	Single Nucleotide Polymorphism
SSP	Sequence-Specific Primer
TAP	Transporter associated with Antigen Processing
TBE	Tris-Borate-EDTA
TBI	Total Body Irradiation
TCD	T Cell Depletion
TCE	T Cell Epitope
TCR	T Cell Receptor
Tel	Telomeric
TF	Transcription Factor
TGS	Third Generation Sequencing
TM	Transmembrane
TNF	Tumour Necrosis Factor
TRAIL	TNF-Related Apoptosis Inducing Ligand
ULBP	Unique Long 16 Binding Protein
UM	Upper Marker
UTR	Untranslated Region
V	Volt
VUD	Volunteer Unrelated Donor
WMDA	World Marrow Donor Association
Zeb1	Zinc-Finger E-Box-Binding Homeobox 1
ZMW	Zero-Mode Waveguide

Chapter 1 Introduction

1.01 The Immune System

The immune system is a series of connected elements that perform two key functions: protection from harm caused by pathogens and removal of dead or dying cells to maintain cellular homeostasis. Broadly, the immune system can be divided into two main subsystems: adaptive and innate immunity.

1.01.01 Adaptive Immunity

Adaptive immunity is a highly targeted and specific method of pathogen removal by two main cell groups, B and T cells. Both of these cell types undergo gene rearrangement to create receptors with exquisite specificity for unique epitopes [1,2].

1.01.01.01 T cells and the T cell receptor (TCR)

T cells are lymphocytes that originate from lymphoid progenitor cells that migrate from the bone marrow (BM) to the thymus to undergo lineage determination. During their maturation in the thymus, the genes encoding the immunoglobulin (Ig) proteins that make up the TCR undergo a rearrangement process [3]. Although this results in each cell having an individual specificity, the number of cells in which this occurs produces a vast pool of specificity potential [1]. To ensure that cells with autoreactive ability are removed, thymocytes (T cell precursors) undergo a stringent thymic selection process. Initially, only thymocytes that express TCR with sufficient binding affinity to HLA molecules are maintained; a step known as positive selection. Subsequent selection is based on the TCR not binding to any peptides derived from the host's own tissue with too strong an avidity. This step is the negative selection [4].

Those T cells surviving thymic selection may be divided on the presence of different markers on their cell surface. Cells expressing cluster of differentiation (CD) 4 are activated upon TCR binding to HLA class II molecules expressing cognate antigen, after which they undergo clonal expansion. Importantly, specialised antigen-presenting cells (APC) utilise HLA class II molecules to present extrinsically derived peptides to CD4+ T cells. These T cells, also known as helper T cells, recruit and promote other cells of both the innate and adaptive immune systems to perform their immunological function [5]. Contrarily, CD8+ T cells are activated upon interaction with antigen-associated HLA class I molecules and exhibit a cytotoxic response; directly killing virally infected or cancerous cells. This cytotoxicity may be elicited by inducing apoptosis in the target cell through the release of lytic granules [6] or expression of the Fas-ligand on the surface of CD8+ T cells, which is subsequently bound by Fas receptor on the target cell (discussed further in Section 1.01.03.01) [7]. The peptide presented by HLA class I molecules also relates to their function: expression of intrinsic peptides, giving a sample of the interior of the cell, allows T cells to specifically recognize cancerous or virally infected cells, thus allowing a targeted cytotoxic response.

Importantly, following their activation, a small subset of the clonally expanded cells persists as memory T cells. These cells are able to rapidly respond to future infections displaying an antigen expressing the epitope recognized by that cell's TCR [8].

1.01.01.02 B cells, the B cell receptors (BCR) and antibodies

B cells are the other main group of adaptive immune cells. These cells function by creating a humoral response to pathogenic stimuli, following recognition of antigen via the BCR, another protein whose genes undergo gene rearrangement to produce a wide

variety of specificities [9]. The BCR utilises a membrane-bound Ig structure to recognize their targets that, in the presence of T cell help, instigates activation and proliferation of the B cell, maturation into plasma cells and secretion of soluble Ig (antibodies) specific to the antigen. The affinity of antibodies towards a particular antigen increases over time as B cells undergo somatic hypermutation. The downstream effects of antibody-mediated immunity are to induce and recruit elements of the innate immune system such as complement and natural killer (NK) cells that are able to assist in the clearance of the pathogen [10]. A memory response, with the ability to rapidly produce antibodies against the same stimulus, may also be observed following B cell clonal expansion [11].

The equivalent tolerising selection of B cells occurs in BM, whereby B cells expressing BCRs with affinity to self are removed in a clonal deletion process similar to thymic selection (via apoptosis) or made anergic (via downregulation of the specific BCR). This, alongside peripheral tolerance exerted by helper T cells, prevents the production of soluble antibodies in response to BCR binding self-antigens in healthy individuals [12].

1.01.02 Innate Immunity

Contrary to adaptive immunity, innate immunity is found even within simple organisms. Physical barriers such as skin, hair and mucosal membranes provide an initial obstacle against pathogens accessing internal tissues. However, some organisms have evolved more elaborate innate immunity systems to help protect themselves from the wide range of pathogens. To encompass this broad range of physical, humoral and cellular systems, innate immunity may be described as those factors that do not undergo

gene rearrangement and act as a rapid but relatively non-specific means of dealing with infectious agents.

Some animal species, including humans, have evolved a more extensive range of cell types to elicit innate immune responses. Some cells, such as macrophages and neutrophils, actively search out and destroy their targets via pattern recognition receptors specific to components found only on the surface of pathogens: pathogen-associated molecular patterns. Upon these cells binding to the pathogen, macrophages and other phagocytes engulf the pathogen or affected cell, before destroying it internally [13]. Neutrophils are capable of phagocytosis, but also destroy target cells by release of lytic granules in a similar way to cytotoxic T cells [14]. Other cell types, such as mast cells, release a series of cytokines that initiate an inflammatory response [15]. A further subset, including dendritic cells, act as a link between the innate and adaptive immune systems; able to present antigenic peptides to the cells of the adaptive immunity, allowing them to generate a specific, targeted response [16]. However, as the cell that most commonly expresses the molecules of focus in this study, Killer-cell Immunoglobulin-like Receptors (KIR), the innate cell type with most relevance to this project is the NK cell, discussed below.

1.01.03 NK cells

NK cells, originating from lymphoid progenitors, are a group of circulating innate immune cells involved in reactivity against virally infected and cancerous cells. However, despite their functional similarity to cytotoxic CD8+ T cells, NK cells do not encode receptors that undergo gene rearrangement or somatic mutation to generate vast

repertoires of ultra-specific receptors. Instead, more generic signals of pathogenesis are detected, leading to NK cell killing of infected or transformed cells.

1.01.03.01 NK cell immunity

The means by which NK cells carry out their cytolytic function are expansive. Directly, NK cells are able to destroy target cells through a targeted, soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE)-mediated release of lysosomes containing cytolytic factors such as perforin and granzyme. Perforin targets the cell membrane of the susceptible cell, polymerising to create structures which perforate the membrane. This allows uninhibited movement of ions and low molecular weight molecules across the cell membrane, ultimately leading to necrosis [17]. Granzymes, a collective term for a group of proteases also delivered to the target cell by lysosomes, function by degrading target cell cytoplasmic proteins and inducing apoptosis. Just one of many methods by which this may be achieved is granzyme activity against proteins involved in preventing DNA damage, such as inhibitor of caspase-activated DNase (ICAD). Upon ICAD degradation, CAD becomes uninhibited, targeting DNA and furthering the apoptotic process [18,19].

In addition to the release of cytolytic granules, NK cell cytotoxicity may be mediated by activation of death receptors on the target cell surface. TNF-related apoptosis-inducing ligand receptors (TRAIL-R), one example of these death receptors, bind to TRAIL expressed on the surface of NK cells. Upon binding to its ligand, Fas-associated death domain (FADD) proteins are recruited to the cytoplasmic domains of TRAIL-R, initiating a caspase cascade and, ultimately, apoptosis [20]. Fas, an alternative death receptor (and basis for the naming of FADD), unsurprisingly functions in a similar manner upon binding its ligand, FasL [21,22]

In addition to its role in direct, cell-mediated cytotoxicity, FasL is also capable of contributing to the alternative method by which NK cells function: the recruitment of alternative immune cell types to instigate a broader immune response [23]. Soluble FasL, alongside cytokines such as interferon gamma (IFN-γ) and tumour necrosis factor alpha (TNF-α), is responsible for stimulating cells of both the innate and adaptive arms of an immune response, as well as initiating unmediated apoptosis. IFN-γ is able to instigate a T cell response by promoting antigen-presentation in dendritic cells, or by binding directly to IFN-γ receptors on the T cell itself. TNF-α functions predominantly as an inflammatory agent, promoting migration of other immune cells [24-26].

Another example of interactions between innate and adaptive immunity may be seen in the opposite direction, whereby antibodies generated by the adaptive branch of immunity are detected by NK cells (and other innate immune cells) to initiate an innate response. This immunity, designated antibody-dependent cellular cytotoxicity (ADCC), is mediated via fragment crystallizable (Fc) receptors, which recognize the constant component of antibodies. NK cell killing of the target cell is then facilitated by cytolytic granules [27].

In addition to the direct detection of pathogenesis via the Fc receptors, more subtle signals, both inhibitory and activating, are also detected by a range of receptors expressed by NK cells to constantly monitor the health of the cells in the immediate proximity. The balance of these signals determines the reactivity of the NK cell.

1.01.03.02 NK cell detection of "missing self"

NK cells' lytic potential can be observed when the inhibitory signals created by binding between inhibitory NK cell receptors and their ligands are overcome by the activation signals via the activating NK cell receptors. Alternatively, NK cells can become activated in the absence of an inhibitory signal, the so-called "missing self" hypothesis [28]. By this method, NK cells are maintained in an unreactive state by the constant presentation of HLA class I to NK cell receptors by healthy cells. However, as a means of evading CD8+ T cell mediated cytotoxicity, virally-infected and cancerous cells may downregulate HLA class I cell surface expression, resulting in NK cells that were previously inhibited by the expression of HLA class I now detecting the lack of ligand [29,30]. These NK cells become reactive to specifically target the unhealthy cell. However, due to the highly polymorphic nature of HLA class I and some of their associated receptors, these situations pose a conundrum in individuals who encode inhibitory NK cell receptors for which the cognate ligand is not also encoded, as can be the case with certain KIR receptor-ligand combinations. As an example, KIR2DL1 is an inhibitory KIR encoded by more than 90% of the European population. Its ligand, HLA-C2, however, is absent in approximately 30% of those individuals who encode KIR2DL1 (additional receptor-ligand specificities are given in Section 1.03.05) [31]. This example of missing self could, unchecked, cause autoreactivity.

1.01.03.03 NK cell licensing and memory

To overcome this potential autoreactivity, NK cell immunity is thought to be controlled through a licensing process whereby only NK cell clones whose natural state is inhibition are able to function in a cytotoxic fashion. This is achieved via a maturation process that requires an individual NK cell to express at least one inhibitory NK cell receptor for which the cognate HLA ligand is present [32,33]. However, unlike thymic

education, NK cell licensing occurs in the periphery and is adjustable; cytotoxicity is dependent on the environment in which the NK cells are placed [34,35]. As such, NK cells that do not meet this criterion are not destroyed, but rather reside within a quiescent pool of anergic cells. More recent findings imply that this licensing may be affected not only by *trans* interactions (those between an NK cell and a different, HLA-presenting cell), but also interactions between NK cell receptor and ligand sharing the same cell surface (*cis* interactions) [36,37].

Due to their potential as a cellular therapy, research is continuing into how best to bypass this licensing to allow allogeneic NK cells to be infused as a medicine in an active state [38]. As such, it has been observed that NK cells can be induced to undergo a two-stage activation involving "priming" or "triggering" using specific tumour cell lines or interleukin (IL)-2, respectively [39,40].

In addition, evidence is emerging of NK cell subsets exhibiting a memory-like response, whereby lasting NK cell alloreactivity is targeted against a specific pathogen. This has been demonstrated following *Cytomegalovirus* (CMV) reactivation in immunocompromised individuals, whereby a population of mature NK cells persist and respond more quickly and effectively upon re-exposure to the pathogen [41,42]. A similar response has also been observed in the rhesus macaque *(Macaca mulatta)* model, whereby memory NK cells against simian immunodeficiency virus (SIV) have been isolated [43].

Examples of adaptive immunity characteristics, such as immune memory, and factors such as the crosstalk that exists both between and within the innate and adaptive arms of

immunity, demonstrates the complexity of the immune system. This complexity is both controlled and compounded at a genetic level, a factor that has driven immunogenetics to become a focus for many immunologists.

1.02 Immunogenetic structures

There are hundreds of genes involved in control of the immune response arranged throughout the human genome. However, there are several regions of the genome that are enriched for genes involved in modulating immunity. The section below discusses several of these immunogenetic structures.

1.02.01 The Major Histocompatibility Complex

The Human extended Major Histocompatibility Complex (MHC) is an exceptionally gene-dense, 7.6 megabase (Mb) region of the human genome that contains a large proportion of the genes involved in immunity [44]. Included amongst these are genes encoding the HLA and MHC class I polypeptide-related sequence A and B (MICA and MICB) molecules which act as ligands for a range of different NK cell receptors. HLA molecules are also responsible for presenting peptides to T cells, instigating an adaptive immune response via the TCR. The genes of the MHC can be divided into three major regions according to their function and position within the MHC on the short (p) arm of chromosome 6: MHC class I (encoding HLA class I genes), MHC class II (encoding HLA class II genes) and MHC class III [45].

1.02.01.01 HLA class I molecules

The HLA class I genes are located on the telomeric region of the human MHC and encode proteins expressed on the surface of most nucleated cells. These proteins express short (predominantly nonamer), intrinsically-derived peptides, allowing cytotoxic CD8+

T cells to constantly monitor host cells for potential infection or cancerous transformation [46]. In addition, the expression level of HLA class I molecules are monitored by NK cells, allowing the detection and selective targeting of virally-infected or cancerous cells which have downregulated the expression of these molecules [30,47].

HLA class I molecules are membrane-anchored glycoproteins consisting of a short cytoplasmic tail, transmembrane domain and three extracellular immunoglobulin domains termed α1, α2 and α3. In addition to this MHC-encoded protein chain, each complete HLA class I heterotrimeric molecule includes a single β_2-microglobulin molecule (encoded outside of the MHC on chromosome 15) and an endogenously-derived peptide which resides in the groove created between α1 and α2 domains [48,49].

The HLA class I subset can be further divided into classical and non-classical genes. The classical HLA class I loci, HLA-A, HLA-B and HLA-C, are key molecules in transplantation. As well as acting as ligands for some NK and T cell receptors, allele matching donors and patients for these hyperpolymorphic genes limits recipients' alloreactivity [50]. In addition, non-classical HLA loci also act as ligands for other NK cell receptors: HLA-E interacts with NKG2A/CD94 [51], HLA-F open conformers are ligands for KIR3DS1 [52-54] and HLA-G, expressed on decidual cells, binds to KIR2DL4 and may have a key influence on the outcomes of pregnancy [55,56]. The HLA class I region also accommodates several MHC-related genes, such as MICA and MICB, which are ligands for the activating NKG2D NK cell receptor [57].

1.02.01.02 HLA class II molecules

HLA class II genes are found at the opposing, centromeric region of the MHC structure. Each HLA class II molecule is comprised of an α and β chain encoded by separate, but related, genes and are primarily expressed on the surface of specialised APCs. Three classical HLA class II molecules exist: HLA-DR, HLA-DQ and HLA-DP. Each of these is characterised by one gene exhibiting low allelic polymorphism (HLA-DRA, HLA-DQA1 and HLA-DPA1) and the alternative gene encoding considerably more allelic polymorphism in the human population (HLA-DRB1, HLA-DQB1 and HLA-DPB1) [58]. When considering HLA-DR, three further genes; HLA-DRB3, HLA-DRB4 and HLA-DRB5, encode β chains capable of surface expression, albeit at a lower level. These genes are variably present in the human population, such that an individual may encode none, one or two of these additional genes [59].

In addition to the α and β chains, an exogenously-derived peptide is bound at the interface of the most membrane-distal domains of the molecule, completing these HLA class II heterotrimeric structures [60]. Owing to the more open structure of the antigen-binding domains of HLA class II molecules, longer peptides can be accommodated, up to more than 20 residues in length [61]. The longer peptides presented by HLA class II molecules are recognized by CD4+ T cells and, if non-self peptides are presented to these T helper cells, an immune response may be instigated.

1.02.01.03 MHC class III molecules

Residing between the HLA class I and II regions is the MHC class III locus. This region, although encoding no classical HLA molecules, does contain genes which encode several key immune system components. These include several members of the

complement family of proteins which are responsible for the cascade that occurs upon antibody binding that ultimately results in cell lysis. In addition, TNF, a molecule involved in the inflammatory response, is also encoded in this region [44]. The high density of immune system related genes within the MHC highlights this genetic structure's importance in the control of immunity.

1.02.01.04 HLA polymorphism

Extremely high levels of polymorphism have been discovered across the classical HLA genes. At present, there are over 23,000 unique HLA alleles known at these six loci, encoding over 15,000 unique proteins, making these some of the most polymorphic genes in the human genome (IPD-IMGT/HLA Database v3.37.0, released July 2019) [58,62]. This polymorphism has arisen predominantly as a mechanism to ensure that, as a whole, the human population is able to present peptides to the immune system from ever-evolving pathogens [63,64]. As well as the repertoire of presentable peptides, the degree of polymorphism is responsible for differences in cell surface expression, perhaps best categorised by the HLA-C locus [65]. Allelic variation also affects the specific structure of the HLA protein which, in the context of allogeneic transplantation, may act as an antigen to allogeneic immune receptor molecules [66].

1.02.02 The Natural Killer Complex

Another genetic structure, the natural killer complex (NKC), encodes receptor proteins expressed by NK cells that function as either activating or inhibitory molecules. The NKC is located on the short (p) arm of chromosome 12 (12p13.1) and the encoded glycoproteins are not members of the immunoglobulin superfamily. Instead, a type II transmembrane (TM) C-type lectin conformation is adopted. Genes encoding the NKG2 proteins and their dimerization partner, CD94, are located within the NKC. Although

the CD94 gene is monomorphic, the NKG2 genes do exhibit limited allelic polymorphism, although not to the same extent as classical HLA genes. In addition to the NKG2 genes, the human NKC also accommodates the Ly49L pseudogene, an ortholog of the murine functional equivalent of human KIR genes [67,68].

The majority of the NKG2 receptor complexes act to recognize HLA-E molecules that presents peptides of the classical HLA class I proteins, a function that assists in identifying "stressed" target cells [69,70]. This interaction has been proposed as a key mediator in the education of NK cells [51]. Another member of the NKG2 family of receptors, NKG2D, does not utilise HLA-E as a ligand but instead binds alternative stress-markers: MICA, MICB and unique long 16 binding proteins (ULBP) [57].

1.02.03 The Leukocyte Receptor Complex

The Leukocyte Receptor Complex (LRC) is an additional gene-dense region of the human genome involved in encoding receptors involved in immunity. It is located on the long (q) arm of chromosome 19 (19q13.4) [71]. Unlike the NKG2 glycoproteins described above, the receptors encoded in the LRC belong to the Ig superfamily. However, their expression is also limited to the surface of immune cells, NK cells particularly. Three main groups of receptors are encoded: KIR, leukocyte immunoglobulin-like receptors (LILR) and leukocyte-associated immunoglobulin-like receptors (LAIR). There are several other genes encoding proteins involved in immunity contained within the LRC, such as FCAR: a receptor specific for the Fc fragment of IgA [72,73].

Briefly, LILRs exist as either membrane-bound or soluble, activating or inhibitory receptors. Initial studies revealed the ligands for these receptors were dominated by HLA class I molecules. However, more recent research has implicated several other potential ligands, suggesting that these genes are involved in several different aspects of biology. Additionally, two LILR pseudogenes have been described [74,75]. The loci of the LILR genes are separated into two clusters, either side of the LAIR cluster. The two LAIR genes each encode collagen receptors. However, LAIR-1 is a membrane-bound protein that, upon binding its ligand, inhibits an immune response, whilst LAIR-2 is a soluble form of the same receptor that is hypothesized to control the degree of LAIR-1-mediated inhibition [73,76]. The KIR genes comprise the majority of the remainder of the LRC region. These are discussed in greater depth below.

1.03 Killer-cell Immunoglobulin-like Receptors (KIR)

The KIR genes form the third member of the LRC family. The members of this family, expressed on the surface of NK cells and some T cell subsets, have been repeatedly implicated in affecting the susceptibility to cancer [77,78], viral infection [79,80], pregnancy complications [81,82] and certain stem cell transplantation outcomes [83,84]. However, our current understanding of the influence of the vast degree of polymorphism is still lacking, prompting this study. In the section below, I briefly review what is already known about this family of fascinating genes.

1.03.01 KIR nomenclature

The nomenclature of KIR is based on the structure of the molecule and gene (Figure 1.1). The number of extracellular domains (either 2D or 3D) is followed by the length of the cytoplasmic tail: long (L) or short (S). This alphabetical position in the gene name may also be used to denote the lack of expression exhibited by pseudogenes (P). The

subsequent digit (1-5) denotes the narrower sub-grouping within this faction, defining individual genes in the order by which they were discovered. In the case of KIR2DL5 genes, a further character, A or B, is added to distinguish unique genes with similar structure and sequence. An asterisk (*) is used to separate gene name from allele name. Allelic information is conferred by a series of numbers: the first three digits distinguish alleles of a gene with differing amino acid sequence, the next two differentiate alleles containing synonymous DNA mutations within the coding domain sequence (CDS) and the final two digits refer to alleles differing by polymorphisms only within introns or untranslated regions (UTRs) [85].

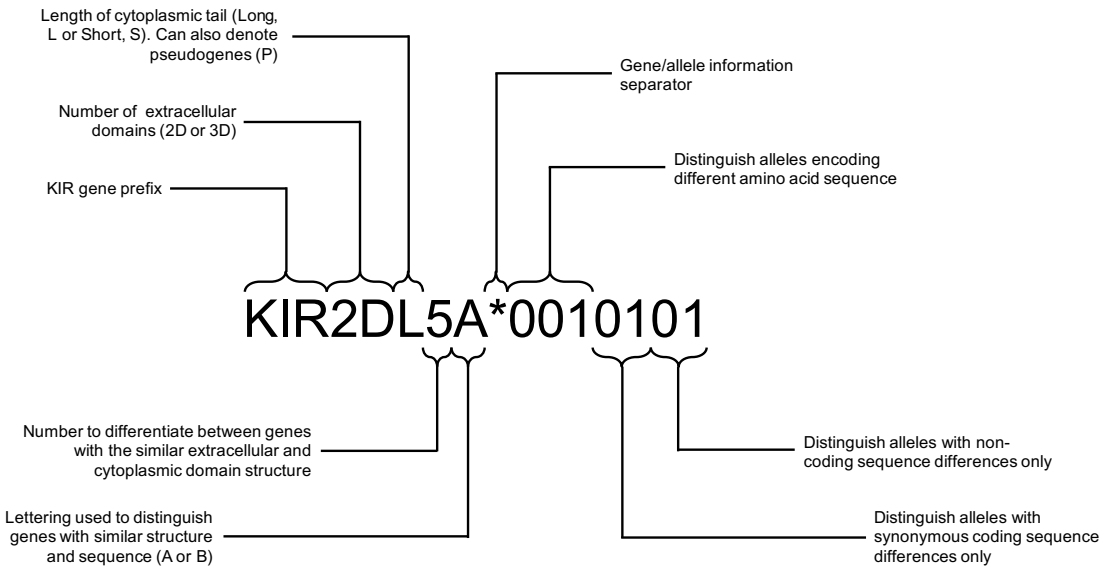


Figure 1.1 KIR gene and allele nomenclature

The nomenclature of KIR genes is based around the domain structure of the protein, whilst the allele name, following the asterisk separator, is based on the specific amino acid and DNA sequence of the gene.

1.03.02 KIR haplotype structure

The KIR locus is located on the long arm of chromosome 19 (19q13.4), occupying a region approximately 100-350 kbp in length (depending on the KIR gene content),

within the LRC genetic structure. Fifteen distinct KIR genes (KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR2DL5A, KIR2DL5B, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DS5, KIR3DL1, KIR3DL2, KIR3DL3 and KIR3DS1) and two KIR pseudogenes (KIR2DP1 and KIR3DP1) have been described [71].

KIR3DL3, KIR3DP1, KIR2DL4 and KIR3DL2 are found on almost all haplotypes and are frequently referred to as KIR framework genes. The remainder of the haplotype gene content, however, can be considerably more variable. Broadly, there are two main haplotype groups. KIR A haplotypes are relatively conserved in gene content and contain, in addition to the framework genes, KIR2DL1, KIR2DL3, KIR2DS4, KIR3DL1 and KIR2DP1, or a lesser combination of these genes. KIR B haplotypes have a more variable gene content and may include any of the KIR genes (Figure 1.2) [86]. If one were to consider all the KIR genes on a single haplotype, there are several loci that accommodate different KIR genes that we now consider to be alleles of the same locus. KIR2DL2/KIR2DL3, KIR2DS3/KIR2DS5 and KIR3DL1/KIR3DS1 are examples.

In addition to differentiating KIR haplotypes based on the presence of KIR B-specific genes, KIR haplotypes may be segregated based on the specific organisation of genes along the chromosome. Due to the short sequence between most KIR genes (~2 kbp), there is only a small chance of intergenic recombination. However, a larger intergenic sequence exists between KIR3DP1 and KIR2DL4 in the centre of the haplotype and forms a recombination hotspot [87-89]. As such, the KIR haplotype may be defined according to the gene motifs KIR3DL3~KIR3DP1 (i.e. centromeric, Cen) and KIR3DP1~KIR3DL2 (i.e. telomeric, Tel).

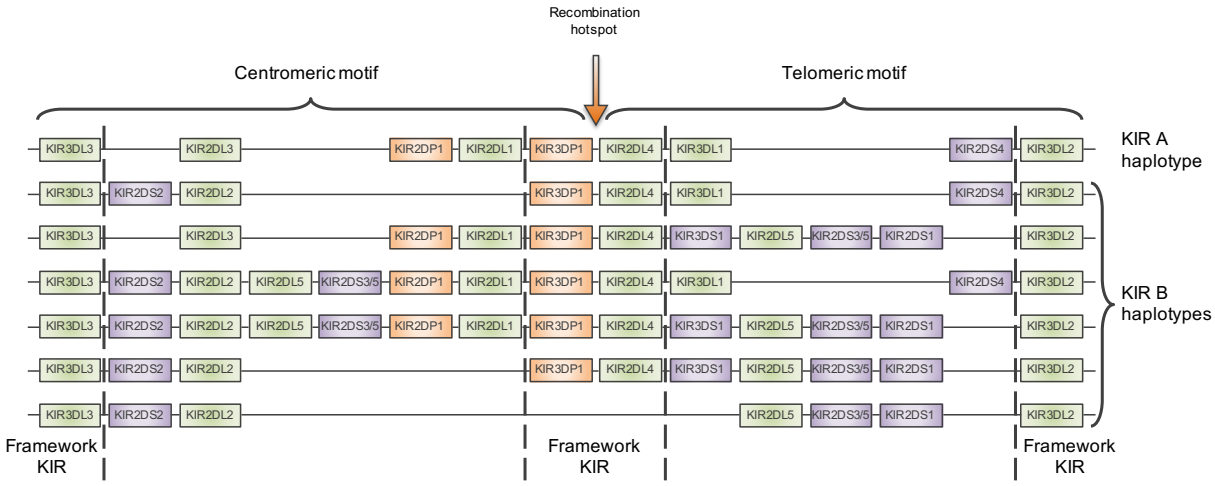


Figure 1.2 Common KIR haplotypes

Several of the more common KIR haplotypes are displayed. Genes with different function are coloured coded: inhibitory KIR (green), activating KIR (purple) and pseudogene KIR (orange). The gene content of the KIR A haplotype (top) is generally conserved, although lesser combinations of these genes have been observed. KIR B haplotypes, however, exhibit more variable gene content, and can incorporate either the centromeric or telomeric motifs from the KIR A haplotype due to the recombination hotspot between KIR3DP1 and KIR2DL4. The loci for KIR2DL5 and KIR2DS3/5 have become duplicated and appear on both the centromeric and telomeric motifs. In the bottom haplotype, the central genes encoding KIR3DP1~KIR2DL4~KIR3DS1 have been deleted. Dashed lines demarcate the framework KIR genes common to the remaining KIR haplotypes.

As more is discovered about KIR haplotypes, it is becoming apparent that gene copy number variation (CNV), arising from unequal gene transfer recombination events, is also a key feature, particularly in the KIR B haplotype group [90-92]. Perhaps the best example of this is the existence of two separate genes encoding the KIR2DL5 protein: KIR2DL5A and KIR2DL5B. These two genes may both be absent from, or present on, the same haplotype, indicative of multiple loci. It is also possible for only one of these loci to appear on an individual haplotype [93]. Interestingly, the KIR2DL5A gene appears on the Tel motif, whilst the KIR2DL5B gene is located most frequently on the Cen motif, although exceptions to this rule have been observed [94]. Other examples of CNV show how larger regions of DNA, such as the KIR3DP1~KIR2DL4~KIR3DL1/S1

motif, have undergone unequal chromosome recombination, thus creating haplotypes with multiple copies of several consecutive genes within a single haplotype. Alternatively, large multigene deletions may also be observed [95].

1.03.03 KIR gene structure

The evolution of the KIR gene family in various species has been characterised by expansion and contraction of different lineages of KIR genes, partly driven by the simultaneous evolution of their HLA class I ligands [96,97]. This duplication-based history has resulted in the human KIR genes each sharing the same basic nine exon structure (Figure 1.3, adapted from Robinson *et al.,* 2013 [98]). Exons 1 and 2 encode the leader peptide responsible for directing the mature KIR protein to the cell surface. Exons 3, 4 and 5 are each capable of encoding one extracellular domain, denoted D0, D1 and D2, respectively. However, in the majority of KIR2D molecules, exon 3 has become silenced via either a mutation at the intron 2-exon 3 splice site or the presence of a premature stop codon within the pseudoexon sequence. This results in a D1-D2 extracellular domain structure. By contrast, the other two-domain KIR in humans, KIR2DL4 and KIR2DL5, translate exon 3 but feature a deleted exon 4 and thus encode the different, D0-D2 extracellular domain structure [99]. As all three exons are present and translated in the KIR3D genes, the mature protein exhibits the D0-D1-D2 structure.

Exon deletion is also observed in the sequences of the KIR3DL3 alleles, whereby the sequence of exon 6 has been removed. Although this exon encodes the stem region of the molecule, low levels of KIR3DL3 expression has been observed on decidual cells and subsets of adult NK cells [100]. Exon 6 has also been deleted from the KIR3DP1 pseudogene sequence along with exons 7, 8 and 9, which encode the transmembrane

and cytoplasmic tails of KIR molecules. The KIRs with short cytoplasmic tails have a shortened exon 9 sequence.

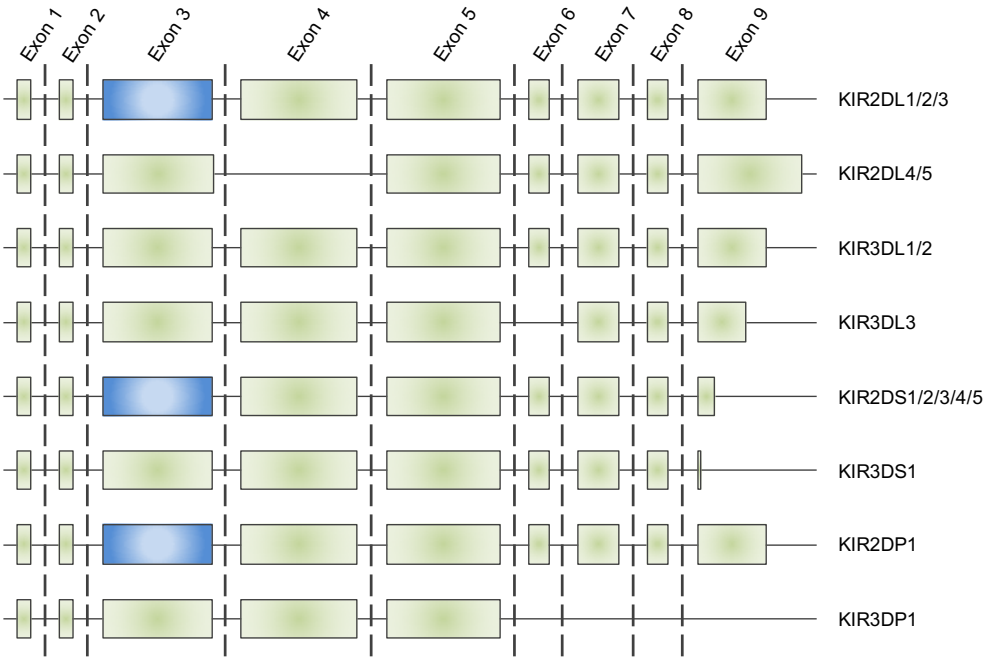


Figure 1.3 KIR exon structure

Due to the duplication-based emergence of the individual KIR genes, a common nine exon structure can be observed, separated by dashed lines. Exons in green are expressed (other than in pseudogenes), whilst exons in blue are pseudoexons whose template is not converted to amino acid sequence. Exon 4 in the KIR2DL4/5 genes, exon 6 in the KIR3DL3 gene and exon 6 onwards in the KIR3DP1 pseudogene have been deleted. The lengths of exons/introns in this figure are not to scale. Adapted from Robinson *et al.* (2013) [98].

1.03.04 KIR molecule structure

As eluded to in previous sections, all KIR molecules are cell surface receptors comprising of multiple extracellular domains, a stem, transmembrane domain and cytoplasmic tail. However, the products of different KIR genes are distinguishable from one another owing to specific arrangement of both extra- and intracellular domains (Figure 1.4, adapted from Robinson *et al.*, 2013 [98]). Despite the extracellular domains of the KIR3D molecules exhibiting similar structure, these gene products can be

differentiated by the frequency or absence of immunoreceptor tyrosine-based inhibitory motifs (ITIMs) encoded within the cytoplasmic tail. Both KIR3DL1 and KIR3DL2 encode two ITIMs, whilst KIR3DL3 encodes only one and KIR3DS1 does not encode any. This ITIM protein structure, [I/L/V/S]-X-Y-X-X-[L/V], is responsible for relaying an inhibitory signal, as discussed further in Section 1.03.07. Those KIR molecules characterised by ITIM-bearing cytoplasmic tails are known as inhibitory KIR. The remaining KIR molecules, each encoding truncated cytoplasmic tails with no ITIM, are known as activating KIR. Although unable to relay an activating signal independently, activating KIR associate with immunoreceptor tyrosine-based activation motif (ITAM)-encoding molecules to elicit their function. The KIR2D proteins also exhibit differences between their cytoplasmic domains. Similarly to KIR3DS1, two-domain KIR with short cytoplasmic tails, i.e. activating KIR2D, also fail to encode any ITIMs owing to their truncated exon 9 sequence. KIR2DL5 molecules encode one ITIM. The remaining KIR2D genes each encode two ITIMs within their cytoplasmic domain. However, both KIR2DL4 and KIR2DL5 may be distinguished from the other KIR2D by the presence of the D0, and absence of the D1, extracellular KIR domains. As it is the extracellular domains of KIR molecules that are responsible for ligand specificity, differences between the ligands bound by KIR3D and KIR2D have been observed, as discussed in the following section.

1.03.05 KIR ligands

HLA molecules were confirmed as ligands for NK cell receptors when their absence or downregulation in BM transplantation models led to NK-mediated rejection of grafts [28]. Soon after, the specificities of individual KIR genes were elucidated; work which

demonstrated the relevance of extracellular domain structure in altering binding specificity.

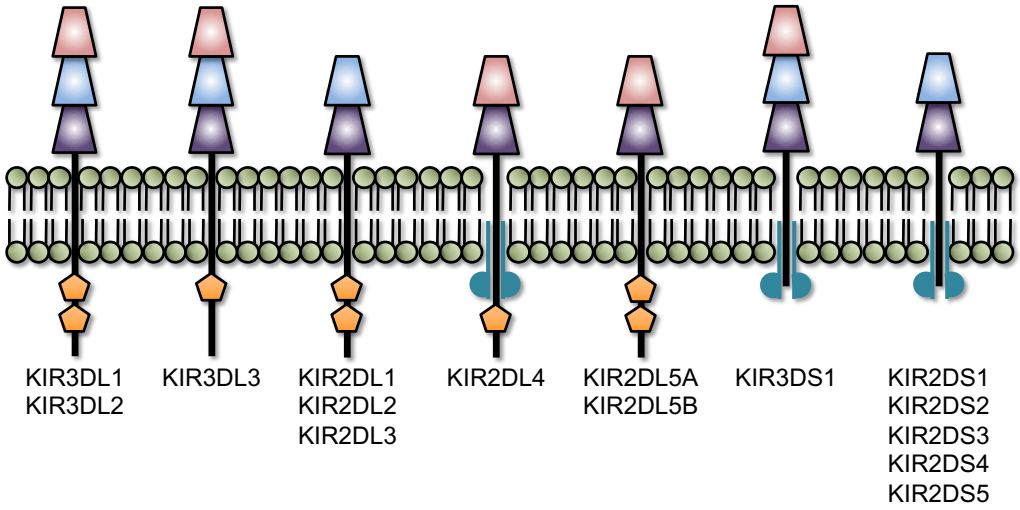


Figure 1.4 KIR molecule structures

The structural differences in both the extra- and intracellular domains can distinguish different groups of KIR molecules. The extracellular KIR domains, D0 (red), D1 (blue) and D2 (purple) control ligand specificity, whereas the transmembrane and cytoplasmic domains (black), encoding either no, one or two ITIMs (orange), are involved in relaying activating or inhibitory signals. Those KIR encoding no ITIM rely on association with adapter molecules (teal) to relay their signal via ITAMs. KIR2DL4, although encoding one ITIM, is also capable of associating with adapter molecules to relay an activating signal. Adapted from Robinson *et al.* (2013) [98].

The inhibitory KIR2D molecules exhibiting the D1-D2 extracellular domain structure were shown to bind to HLA-C allotypes. Furthermore, KIR2DL1 and KIR2DS1 were shown to bind to HLA-C alleles encoding asparagine at residue 80 (HLA-C2), albeit with different avidities [101,102], whilst KIR2DL2 and KIR2DL3 were demonstrated binding to HLA-C allotypes encoding lysine at residue 80 (HLA-C1) [103,104]. In addition to binding to HLA-C1 alleles, certain alleles of KIR2DL2 and KIR2DL3 have also been shown to bind to HLA-C2, although binding is at a weaker avidity than that of KIR2DL1 [105,106].

An investigation into NK cells expressing three-domain KIR demonstrated that they were selectively inhibited by HLA-Bw4, a motif encoded by certain alleles of HLA-A and -B [107]. The individual KIR responsible for this binding was subsequently shown to be KIR3DL1, and studies have proposed a differential functional strength dependent on the isoleucine/threonine dimorphism at position 80 of the Bw4 motif and the peptide bound within the HLA binding groove [108-110]. HLA-Bw4 was also proposed to bind KIR3DS1, due to the high homology with the extracellular domains of KIR3DL1. However, subsequent assays suggested that KIR3DS1-HLA-Bw4 interaction may be peptide dependent [111], whilst more recent evidence suggests that open-conformers of HLA-F are key ligands for this activating KIR molecule [52-54]. Only one allele, KIR3DS1*014, has been implicated in HLA-Bw4 interaction [112].

Extracellular homology between KIR2DL molecules and their KIR2DS equivalents was also proposed to result in similar ligand specificity. However, with the exception of KIR2DS1 and HLA-C2, binding assays revealed that, in most cases, binding to HLA-C was non-existent or significantly weaker [102,113-115]. As such, the ligands for KIR2DS2, KIR2DS3, KIR2DS4 and KIR2DS5 are less well defined. However, HLA-A*11:02 has been proposed as the ligand for KIR2DS4, whilst binding between KIR2DS2 and the HLA-A*11:01 was confirmed when the crystal structure was determined [116,117]. Certain alleles of KIR2DS5 have been demonstrated to bind HLA-C2 alleles, although a ligand shared amongst all allotypes has yet to be discovered [118]. Presently, there are no known ligands for KIR2DS3.

Several ligands for the KIR3DL2 molecule have been proposed. NK cells expressing KIR3DL2 have been demonstrated to be inhibited by target cells expressing HLA-A*03 and -A*11 molecules bound to certain peptides and, more recently, multimers of the HLA-B*27 molecule in both open-conformer and β_2-microglobulin-associated forms [119-121]. To date, no ligands have been found for the highly downregulated KIR3DL3 molecule. It has instead been proposed that this locus may act as a reservoir of polymorphism for the other KIR genes, or function to assist chromosome alignment during cell division [100].

Of the KIR2D molecules exhibiting the D0-D2 extracellular domain structure, only KIR2DL4 has a demonstrated ligand: HLA-G [55,122,123]. Expression of this non-classical HLA class I molecule is exclusively limited to the surface of trophoblast cells, perhaps indicating a key role for KIR2DL4 in the regulation of pregnancy, further discussed in Section 1.03.06.03. At the time of writing, no ligands have been proposed for KIR2DL5.

1.03.06 The influence of KIR on NK cell function

As one of the main groups of receptors on the surface of NK cells, KIR are involved in three vital components of NK cell function: licensing, alloreactive immunity and regulating arterial remodelling during pregnancy.

1.03.06.01 Licensing by KIR

As outlined in Section 1.01.03, the body's NK cells must undergo a licensing procedure in order to become fully matured. This process requires the expression of an HLA-specific inhibitory NK cell receptor on the surface of each mature NK cell for which the

corresponding HLA ligand is also present. As a key family of HLA-specific receptors expressed on NK cells, KIR play an important role in fulfilling this condition.

However, as outlined above, the class HLA class I genes are extremely polymorphic, and only certain groups of HLA alleles are bound by certain KIR. To further compound the issue, KIR itself is also highly polymorphic in gene content, whereby only a very few loci may be found on all haplotypes. For instance, even though an HLA-C1-specific KIR is encoded on every common haplotype (KIR2DL2/KIR2DL3), more than 10% of individuals encode the HLA-C2/C2 genotype, thus lacking the HLA-C1 ligand [31]. A similar scenario exists for KIR2DL1 and KIR3DL1 (specific for HLA-C2 and HLA-Bw4, respectively) found in a majority of individuals, but for which the ligands are not guaranteed. Regardless, the diversity of KIR specificities allows many NK cells to become licensed via KIR [28,32].

In those individuals without the necessary inhibitory KIR-HLA combination, the older, more conserved NKG2A-HLA-E interaction also exists to license NK cells. This system allows the constant monitoring of classical HLA class I expression through the necessity of the leader peptide binding to HLA-E to maintain its own expression. An elegant study investigating polymorphisms both within the leader peptide and remainder of the allele implicate two "schools" of NK cell licensing: one via KIR, the other via NKG2A, that have led to the development of specialised HLA haplotypes [51]. To complement this, the 'rheostat' model has been proposed as a mechanism to fine tune NK cell reactivity. This model accounts for the number and strength of each inhibitory receptor-ligand interaction, including both KIR and NKG2A interactions, in an additive sense to determine an overall licensing strength [124,125].

In addition to interactions between inhibitory NK cell receptors and their ligands, several other studies have associated an involvement of activating KIR genes in the licensing of NK cells [126,127]. Signals mediating from activating KIR may be responsible for increasing the likelihood of expression of alternative KIR genes at the cell surface, consequently boosting the probability of a successful inhibitory KIR-HLA combination [128].

1.03.06.02 KIR-mediated alloreactivity

The inhibitory interactions between NK cells and their potential targets are responsible for their licensing. However, upon removal of these interactions, the inhibition is lifted and NK cells rapidly become activated. HLA downregulation is characteristic of certain tumours and viral infections as a means of evading the adaptive immune system [129]. However, this decreased expression is sufficient for NK cells to detect and target transformed or infected cells. Unsurprisingly, viruses such as *human immunodeficiency virus* (HIV)-1 have evolved mechanisms to selectively remove HLA-A and HLA-B from the cell surface via endocytosis, whilst leaving expression of HLA-C, the predominant KIR ligand, unaffected. By doing so, T-cell reactivity via HLA-A and HLA-B is diminished, whilst KIR-mediated inhibition of NK cells is maintained [29,130]. The makeup of CMV also includes a variety of different NK cell evasion strategies including the production of HLA decoy molecules, such as UL18, to maintain NK cell inhibition [131].

Although much of the early research focused on NK cell stimulation via inhibitory KIR detecting missing self, new evidence suggests that expression of stress-inducible ligands specific to KIR is also capable of shifting the balance of NK cell inhibition to instigate

an activating KIR-mediated response. For example, HLA-F, expressed on the surface of stressed cells, is able to induce KIR3DS1-mediated NK cell lytic activity [54]. This, in combination with studies demonstrating that NK cells encoding more activating KIR have more immune potential, relays the importance of activating KIR in NK cell immune responses [132,133].

1.03.06.03 KIR in pregnancy

As well as their function in reactivity against pathogens, NK cells are modulators of spiral artery remodelling during pregnancy. This process, controlled by KIR as well as other NK cell receptors, is responsible for adjusting the rate of placental blood flow to ensure sufficient foetal growth [134]. Although an in-depth analysis is beyond the scope of this thesis, there are several key points that can be learned about KIR function from looking at its role in pregnancy. Firstly, it was possible to resolve the ligand for KIR2DL4, highly expressed on the surface of decidual NK cells, as the non-classical HLA-G molecule [56,122]. Secondly, the correct balance of NK cell activation is required for successful pregnancy: too much inhibition results in low birth weights and complications such as pre-eclampsia and recurrent abortions, whilst overly active NK cells may cause excessive foetal growth and result in complications during labour [135]. By correlating KIR and HLA genotypes from the mother, father and foetus with clinical observations, it was possible to determine that the inhibitory interaction between maternal KIR2DL1 and paternally-derived HLA-C2 is responsible for a lower-than-average birth weight in those foetuses. However, the presence of KIR2DS1, which also interacts with HLA-C2, is sufficient to overcome this inhibition, restoring average birth weight and reducing the risk of pregnancy complications [135,136]. This balance of NK cell activity is believed to contribute to the evolutionary balance between the inhibitory KIR A and activating KIR B haplotypes [137,138].

1.03.07 KIR signalling

Upon KIR binding to its ligand on target cells, immunological synapses are formed. The structure of these synapses is controlled by the specific KIR involved and determines whether an inhibitory or stimulatory response is initiated.

1.03.07.01 Inhibitory signalling via KIR

Upon binding cognate HLA class I ligand on the potential target cell, inhibitory KIR molecules cluster at the synapse between the cells. This process has been shown to be dependent both on the KIR transmembrane sequence and the presence of zinc [139,140]. In addition to the transmembrane domain, the cytoplasmic region of KIR molecules is also a key determinant of immunological function. The cytoplasmic ITIMs, characteristic of inhibitory KIRs, become phosphorylated at the tyrosine residues upon KIR binding its ligand [141]. This phosphorylation permits interaction between the ITIMs and Src homology region 2 domain-containing protein tyrosine phosphatases (SHP), signalling molecules responsible for relaying the inhibitory stimulus within the cell [142]. SHP-1 can only interact when both the tyrosine residues in both ITIMs are phosphorylated. SHP-2, in contrast, is able to interact when only one ITIM is phosphorylated, or even when the tyrosine is substituted with phenylalanine, although the inhibitory power is reduced [143,144]. Upon binding inhibitory KIR molecules, SHPs dephosphorylate the ITIMs and relinquish their autoinhibitive state. The downstream effect of activated SHPs is believed to be exhibited by dephosphorylation of activation signalling factors such as Vav-1; thereby preventing the stimulatory response [145].

1.03.07.02 Activating signalling via KIR

The resting, inhibitive state of NK cells; maintained by the signalling described above, prevents their autoreactivity. However, upon acquisition of sufficient activating stimulus, or lack of necessary inhibitory stimulus, the activation balance shifts to allow a cytolytic response. This response can be initiated when clusters of activating KIR molecules are simultaneously bound to their respective ligand and their ITAM-encoding accessory protein, DNAX-activating protein (DAP)-12. This complex is formed at the activation synapse, where the lipid raft arrangement brings Src family kinases into proximity with the ITAMs, allowing their phosphorylation in large clusters [139]. Phosphorylated ITAMs then signal via a pathway similar to the T cell receptor pathway, whereby Syk/ZAP-70 is utilised to continue a kinase cascade resulting in cytolysis of the target cell [146].

Although traditionally thought to be restricted to only those KIR unable to encode an ITIM, it is known that KIR2DL4 molecules are also capable of initiating an activating response. Whereas most inhibitory KIR do not encode a charged residue capable of interacting with ITAM-bearing accessory molecules, KIR2DL4 encodes a transmembrane lysine which interacts with DAP-12 and an arginine residue near the extracellular edge of the TM domain that transmits an activating signal via an Fc receptor [147,148]. Signalling via this method results in similar downstream affects to the traditional activating KIR response.

1.03.08 KIR allele polymorphism

Although not as polymorphic as the HLA genes located within the human MHC, KIR is still recognized as a highly polymorphic family of genes, comprising nearly 1,000

alleles defined to date (Table 1.1, IPD-KIR Database, v.2.8.0, released November 2018) [58,98]. Although their gene-duplication-based evolution has resulted in each gene showing remarkable homology in sequence to the other KIR, each KIR gene exhibits multiple allelic forms generated through substitution, insertion, deletion and recombination-based mutations. These polymorphisms may affect several different aspects of KIR function.

Table 1.1 KIR allele polymorphism table

KIR gene name	Number of proteins	Number of alleles
KIR2DL1	36	64
KIR2DL2	15	33
KIR2DL3	34	59
KIR2DL4	41	70
KIR2DL5A	9	19
KIR2DL5B	15	35
KIR3DL1	91	147
KIR3DL2	111	161
KIR3DL3	92	164
KIR2DS1	8	16
KIR2DS2	9	24
KIR2DS3	7	16
KIR2DS4	16	37
KIR2DS5	17	24
KIR3DS1	22	39
KIR2DP1	-	40
KIR3DP1	-	29

Allele numbers obtained from IPD-KIR Database, v2.8.0, November 2018

1.03.08.01 Alleles of KIR are expressed differently

It has been demonstrated that every human carries between 6,000 and 30,000 unique NK cell receptor phenotypes within their NK cell population [149]. Accordingly, the expression of KIR molecules on the surfaces of NK cells does not necessarily reflect the total KIR gene content of individual cells, but rather subsets of different receptor combinations being stochastically expressed [150]. Importantly, the overall phenotype is not influenced by the HLA type encoded by an individual and, in adults, the observed

skew towards expression of KIR for which the ligands are present may simply reflect the immunological history of the individual [151].

To add to this complexity, allelic polymorphism within each locus also affects both the level of cell surface expression on a single NK cell and the percentage of NK cells expressing the allele. The most well-studied example is KIR3DL1, whose different alleles form a spectrum of expression and, as a result, functionality [152]. Although a functional assay probing four positions allows the distinction between high expression, low expression and null allele categories, a single nucleotide polymorphism (SNP) is responsible for the S86L amino acid substitution that has been shown to disrupt a folding motif, resulting in the intracellular retention responsible for majority of the range of cell surface expression [153-155]. Intracellular retention has also been associated with the low expression of KIR2DL2*004 [156], KIR2DL5A*005 [157] and KIR2DS5*001 [158] at the cell surface, in each case also associated by destabilising amino acids in the extracellular domains of the mature protein.

As with HLA, examples of null alleles also occur in several of the KIR genes. For instance, certain alleles of KIR2DL4 feature a truncated homopolymer sequence within exon 6 that prevents cell surface expression [159,160]. KIR2DS4, the only classical activating KIR on the A haplotype, also commonly exists as alleles featuring a 22 base pair deletion that renders the protein null [161]. In an interesting scenario, should the deletion variants of KIR2DL4 and KIR2DS4 both be present on an A haplotype, this haplotype may be considered as having no activating potential [162-164]. In a study of 506 European-Americans, the most common allelic KIR A haplotype was found to be an example of this, with a frequency of 2.77% of the total observed haplotypes [165].

Other common KIR A haplotypes from this study were also found to lack any functional activating KIRs.

The percentage of NK cells expressing specific KIR are tightly regulated by methylation of the promoter region of KIR genes. In a broad sense, polymorphism in promoter sequences between different KIR genes may direct the expression of specific KIR subsets in different NK cell subsets [166]. Within individual genes, however, polymorphism between promoters can affect the percentage of NK cells expressing each KIR allele, perhaps best characterised by certain alleles of KIR2DL5B and KIR3DL1, whose hypermethylated promoters prevent transcription and renders these alleles null [167,168].

1.03.08.02 Allelic polymorphism affects KIR function

Polymorphism within genes can also affect their functional efficacy. For example, the amino acid encoded at residue 245 of KIR2DL1 (arginine or cysteine) effects the recruitment of SHP-2 to ITIMs and subsequent downstream inhibitory signalling [169]. Interestingly, those alleles encoding cysteine at residue 245, relating to lower recruitment of SHP-2 and reduced inhibition, tend to reside on the KIR B centromeric haplotype motif [165]. This may be another mechanism by which this KIR B motif has increased activating potential.

As well as SNPs affecting KIR function, as with KIR2DL1 above, larger genetic changes, involving long segments of genes shuffling on a haplotype, can also alter a KIR gene's function. The similarity of the KIR genes contributes to the high recombination rate between alleles of the same KIR gene, but also between alleles of different KIR genes. For instance, a recombination event between KIR2DL1*001 and

KIR2DS1*00201 created KIR2DS1*005, converting the activating potential of KIR2DL1 whilst maintaining the extracellular domains required for ligand specificity [170]. Perhaps more importantly, this hybrid allele and the largely deleted haplotype it resides on may be completely overlooked by commonly employed gene presence/absence assays as the remainder of the genotype appears relatively typical. However, this deletion could potentially have dramatic effects on the function of NK cells encoding this haplotype [171].

Additionally, the two human KIR pseudogenes, KIR2DP1 and KIR3DP1 also exist in multiple allelic forms. Although not expressed, their sequence remains of interest to the immunogenetic community as both KIR2DP1 and KIR3DP1 have been shown to recombine with other KIR genes to form hybrid molecules capable of expression at the cell surface, thereby acting as a reservoir of diversity [170,172].

1.03.08.03 KIR-ligand avidity is influenced by allelic polymorphism

Although interactions between the KIR receptor and peptide bound by the target HLA molecule is known to affect KIR specificity, its contribution to the total binding avidity is often minimal [173]. However, polymorphism in the KIR molecules is capable of dramatically altering this binding strength. Targeted mutagenesis analysis has demonstrated key extracellular domain residues in KIR2DL1 [106], KIR2DL2/KIR2DL3 [174], KIR3DL1 [175] and even KIR2DP1, although this was assuming that the defining deletion in this pseudogene somehow became ‘repaired’ [176]. Recent research on KIR3DL1 variation has demonstrated that it can significantly affect the control of HIV in some patients [177].

1.03.09 KIR typing

The high degree of homology between different KIR genes has, until recently, made sequencing the KIR genes an onerous task. However, a variety of different genotyping methods, by which it was possible to determine the presence or absence of individual KIR genes, have been designed. These include polymerase chain reaction (PCR-SSP) assays, whereby primer sequences specific to individual KIR genes are used to amplify target genes [178]; reverse sequence-specific oligonucleotide probe (rSSOP) assays, by which fluorescently-tagged oligonucleotides specific to particular KIR sequence motifs are used in combination to determine presence of individual genes [179]; and next generation sequencing (NGS) assays, whereby short, exonic KIR sequences are generated and probabilistically assigned to KIR genes to determine presence [180]. Presence/absence assessment has been employed as the KIR genotyping strategy in the majority of clinical studies conducted to date. To complement presence/absence assays, quantitative PCR (qPCR) and droplet digital PCR methodologies were designed to determine gene copy number for each of the KIR loci [91,92].

While useful, these genotyping methods focussed on small regions of the KIR genes, and as such were not able to assess the vast degree of allelic polymorphism present at the KIR locus. Although sequencing-based typing (SBT) was designed to sequence the coding domains of KIR genes in their entirety [181], the recent developments in NGS technologies have allowed the development of several different approaches to tackle this issue. The study of Norman *et al.* (2016) [94] designed a panel of KIR-specific probes capable of capturing DNA fragments containing KIR sequences prior to massively parallel sequencing using an Illumina NGS platform. The Illumina platform was also used by Maniangou *et al.* (2017) [182] who, instead, utilised the high

homology within KIR intergenic sequences to develop PCR primers to coamplify the KIR genes within one reaction. Both elegant designs allow the definition of the whole gene sequences for each of the KIR genes but, by using short read technology, differ crucially to the method I have utilised and will discuss below.

1.04 Single Molecule Real-Time DNA sequencing

Single molecule real-time (SMRT) DNA sequencing, a Third Generation Sequencing (TGS) technology, was developed by Pacific Biosciences as a means to sequence a circularised DNA molecule. By circularising the DNA molecules, sequencing by synthesis is continuous for the duration of the sequencing reaction, thus allowing very long DNA sequences to be determined in phase. For long, highly polymorphic genes such as KIR, the ability to reliably define phase and accurately determine polymorphism is highly desirable.

1.04.01 Library preparation and MagBead loading

SMRT DNA sequencing is used as a technique to sequence double-stranded DNA, either as sheared genomic DNA or intact PCR amplicons, depending on the user's need. As a high throughput sequencing methodology, large volumes of data may be obtained. As such, it is often beneficial to uniquely label several different products (PCR amplicon or sheared DNA) with a DNA barcode, thus allowing the simultaneous sequencing of each product within a single library, a technique denoted: multiplexing. This allows sequences from multiple samples to be generated in one sequencing reaction but assigned to separate samples during bioinformatic analysis (Figure 1.5) [183].

DNA fragments then undergo enzymatic repair to fix nicks and other errors within the strands, and blunt-ending to allow efficient annealing to SMRTbell adapters. These adapters are ligated to circularise the double-stranded DNA fragment, which will ultimately allow the same DNA polymerase molecule to sequence the same section of DNA several times. The SMRTbell adapter sequence also includes a sequencing primer binding site from which the sequencing reaction commences as the annealed primer acts as a binding site for a DNA polymerase molecule [184]. Single-stranded polythymidine (polyT)-coated MagBeads are then added to the library-primer-polymerase complex, which positively interacts via a polyadenosine section encoded within the sequencing primer [185]. This MagBead-loaded complex is now ready for sequencing on the RS II platform (Figure 1.5).

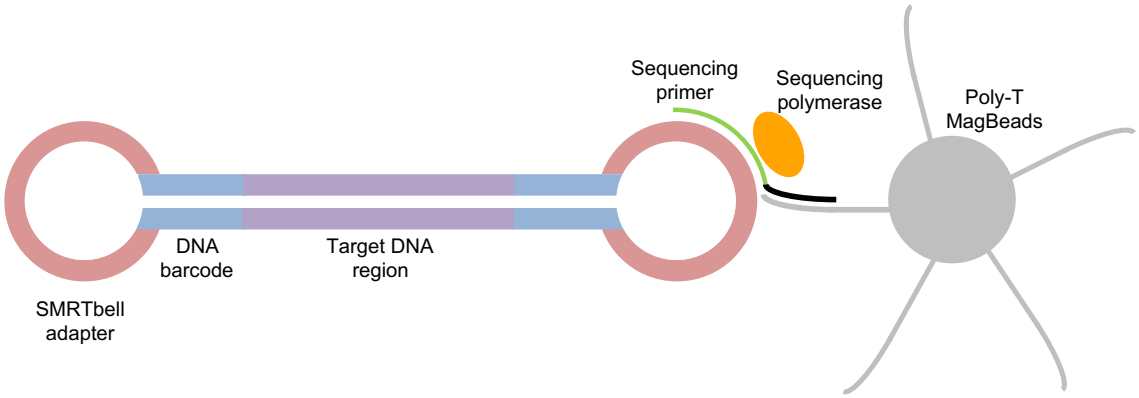


Figure 1.5 Annealed template-MagBead complex

A schematic of a complete SMRT library complex bound to a MagBead. During amplification of the target region (purple), DNA barcodes (blue) are incorporated via unique primer sequence tags. Blunt-ended SMRTbell adapters (red) are then annealed to circularise the double-stranded, linear amplicon. A sequencing primer (green) with complementarity to the SMRTbell adapter acts not only as a docking site for the sequencing polymerase (orange), but also contains a polyadenosine sequence (black) responsible for interacting with poly-T-coated MagBeads (grey).

1.04.02 Sequencing

The sequencing reaction is all controlled within the Pacific Biosciences' RS II instrument. The MagBead complex is transferred to and rolled over the surface of a

SMRT Cell, Pacific Biosciences' sequencing 'chip'. This deposits the complex over 150,000 zero-mode waveguides (ZMWs), which function by trapping light within the lowest compartments of a well, thereby limiting the volume within which excitation of fluorescently-tagged phosphonucleotides can take place. During incorporation of nucleotide into the DNA strand being synthesised, the fluorescent tag is held in an excited state for a number of milliseconds as the nucleotide is linked into the DNA sequence; discerning this signal from the remainder of the "background" signal corresponding to free nucleotides rapidly (within microseconds) diffusing in and out of the excitation zone at the bottom of the ZMW. Fluorescence is detected and recorded by an exquisitely sensitive charge coupled device (CCD) camera in the form of a movie. Each different nucleotide is associated with a different colour fluorescent tag, thus allowing the exact DNA sequence to be decoded [186].

Sequencing by synthesis continues until the polymerase fails or the reaction is halted. In this time, a polymerase molecule may generate a complimentary strand to the same template molecule several times, a process denoted circular consensus sequencing (CCS). This is achieved through a rolling circle, displacement model, whereby the polymerase enzyme is able to displace the nucleotides immediately ahead of it from their associated strand. As the template has been circularised during library preparation, there is no fixed end point other than that set by the experiment design. A single, optimally loaded, 6 hour Pacific Biosciences' RS II sequencing run is able to generate in excess of 2 Gb of sequence (2,000,000,000 bases) [186].

The polymerase and reagents that perform the sequencing have been engineered to perform long-range synthesis with high fidelity, allowing polymorphisms that are long

distances from one another to be accurately determined; a factor particularly important in the sequencing of polymorphic gene systems such as KIR and HLA. Additionally, the polymerase has been adapted to reduce photosensitivity, a factor that is further improved by use of an oxygen-scavenging enzyme to prevent excess damage to the polymerase by reactive oxygen species [187].

1.04.03 Analysis

The fluorescence sequence movie that is generated from each ZMW is decoded into a continuous long read (CLR): a DNA sequence starting with the primer, extending through one direction of the amplicon, through the SMRTadapter at the distal end of the library, through the reverse direction of the amplicon, back through the proximal SMRTadapter, and so forth (Figure 1.6). As such, several passes of the same template molecule can be obtained. Each pass is denoted a subread. Subreads are then subjected to rigorous quality scoring to ensure that only the high quality sequences are utilised in the initial analysis. Reasons to demote the quality of a subread include incomplete sequence or noisy signals resulting from multiple complexes loaded within a single ZMW.

The high quality sequences are first separated by barcode (if multiplexing has been applied), and then iteratively clustered to form sequences from individual alleles. This clustering aligns the high quality sequences from each barcode, before assigning them to a cluster and a phase. For example, in a multiplexed sequencing reaction housing KIR3DL1 and KIR3DS1 template molecules from several samples, the initial separation of subreads is based on the incorporated barcode sequence. Then, the subreads from each different gene are split into separate clusters. Different alleles of the same gene are

housed within the same cluster, but, by definition are distinguishable from their counterparts and are consequently designated different phases (Figure 1.6) [188].

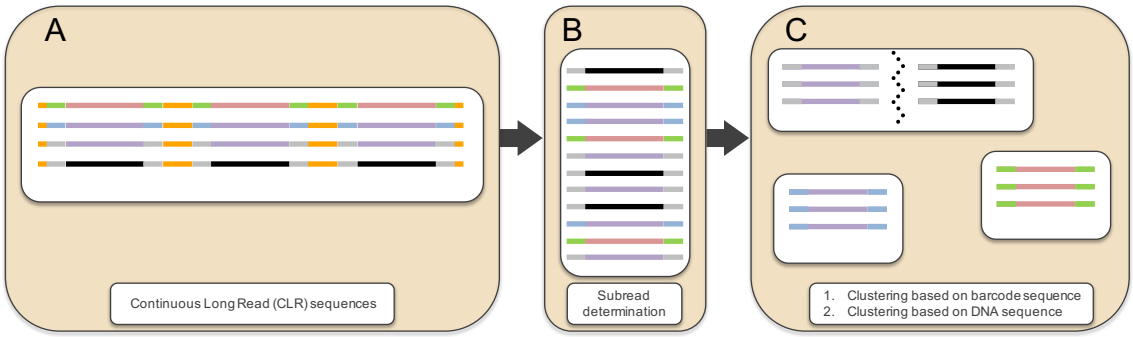


Figure 1.6 Conversion of continuous long reads to individual allele clusters

A) Following SMRT DNA sequencing, the CLR sequences from each ZMW consist of allele sequences (red, purple or black) interspersed with SMRTbell adapter (orange) and DNA barcode sequence (green, blue or grey). B) By trimming away the SMRTbell adapter sequence, individual subreads are obtained. C) Iterative clustering of subreads, initially on barcode sequence only, separates unique samples from one another. This allows identical allele sequences from different samples to be correctly assigned, as shown by the purple allele sequences. Consequent clustering of allele sequences separates alleles of the same locus, should more than one be present, as shown by the grey DNA barcode sequence.

Since many copies of the template are being sequenced concomitantly across the SMRT Cell, each allele is sequenced as many subreads, developing a considerable depth of coverage. Although efforts have been made to minimize misincorporation of nucleotides, the base error rate of the current sequencing chemistry is approximately 10-15%. However, as the errors in SMRT DNA sequencing are non-biased and, as such, occur randomly throughout the generated sequence, sufficient depth of coverage allows the sequences of each individual phase to be aligned and a consensus sequence formed, thereby diluting out any sequencing errors. As a result, the overall error rate of a consensus sequence constructed using in excess of 100 subreads approaches 0% [189].

The generated consensus sequences may then be subjected to further analysis, for example by basic local alignment search tools for nucleotides (BLAST-n) to determine the individual allele identification.

1.05 Haematopoietic cell transplantation (HCT)

Haematopoietic cell transplantation (HCT) is a potentially curative therapy for a number of blood cancers and other haematological disorders. Its basis centres on destruction of the defective BM and peripheral blood cells whilst replacing those cells with healthy haematopoietic stem cells (HSCs) whose pluripotency allows them to repopulate the entire blood system. By replacing HSCs in the patient, the use of extremely high dose chemotherapy is possible. Although the donor source of replacement cells can include the patient themselves (autograft), this thesis focuses on the use of allogeneic donor cells, as it is these transplants that continue to contribute to disease clearance after the initial removal of the faulty cells, termed 'conditioning', discussed below.

1.05.01 Conditioning

Conditioning is the term used to define the preparative treatment a patient undergoes prior to receiving an infusion of HSCs. There are three main purposes of conditioning: to reduce the disease burden; to create a niche for infused cells to populate; and, in the case of allogeneic transplantation, to suppress the recipient's immune system in order to prevent rejection of the infused cells. Although regimens vary widely, three main categories exist: myeloablative, reduced intensity and non-myeloablative.

1.05.01.01 Myeloablative conditioning (MAC)

HCT was first investigated as a potential treatment for people exposed to radiation in the event of nuclear war [190]. The observation that radiation could cause the

irreversible ablation of BM led scientists to attempt to use it as a treatment for BM disorders in mice and other animals. Today, myeloablative conditioning (MAC), from the Latin *muelos* (meaning 'marrow') and *ablatum* (meaning 'remove'), is performed both by total body irradiation (TBI) and chemotherapy, although most commonly-used regimens combine both elements (Figure 1.7). This therapy is much harsher than standard induction therapies in an attempt to completely remove the burden of the malignancy. However, in doing so, the patient's healthy cells are targeted so extensively that the ability to perform normal haematopoiesis is lost. Nevertheless, MAC establishes a niche within bones that healthy, infused cells of the transplant may reside and restore haematopoiesis. In addition, MAC also provides immunoablation, whereby a patient's immune cells, capable of recognizing and rejecting the healthy engrafting cells, are also removed [191].

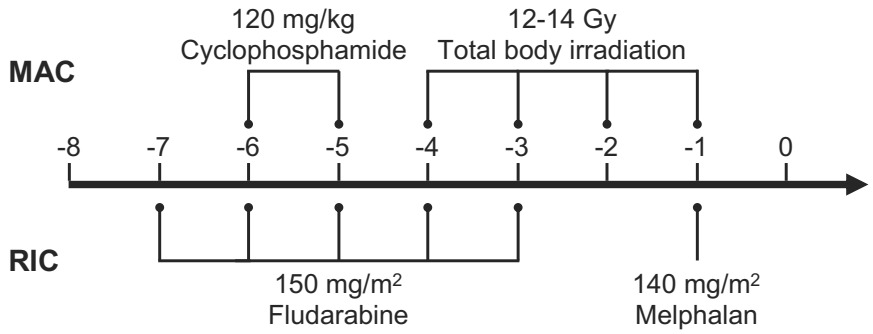


Figure 1.7 Conditioning regimen timeline

A schematic timeline of two common conditioning regimens. On top, a myeloablative "Cy/TBI" protocol, consisting of 120 mg/kg cyclophosphamide on days -6 and -5. This is followed by high dose TBI on each of the four days preceding HSC infusion (day 0). Below the timeline, a common reduced-intensity conditioning (RIC) is shown. This consists of five consecutive days of 150 mg/m^2 fludarabine from day -7, followed by 140 mg/m^2 on the day prior to HSC infusion. Many variations of drug, timing and dosage are used according to transplant centre preference [192].

1.05.01.02 Reduced intensity conditioning (RIC)

Although effectively reducing disease burden, the associated toxicity of MAC is not tolerated by all patients. In particular, one of the largest growing population of HCT recipients, the elderly, are often not fit enough to cope with the severity of MAC preparative regimens (Figure 1.8). Following the discovery that engraftment is possible without myeloablation providing that sufficient immunosuppressive therapy is administered, less toxic conditioning regimens were developed. These use lower doses of chemotherapy alone or in combination with minimal doses of TBI. As this more tolerable preparation is less effective at reducing disease burden, reduced-intensity conditioning (RIC) patients rely more heavily on the engrafting immune cells to target and attack residual diseased cells via a graft *versus* tumour (GVT) effect, discussed in more detail in Section 1.05.03.02.

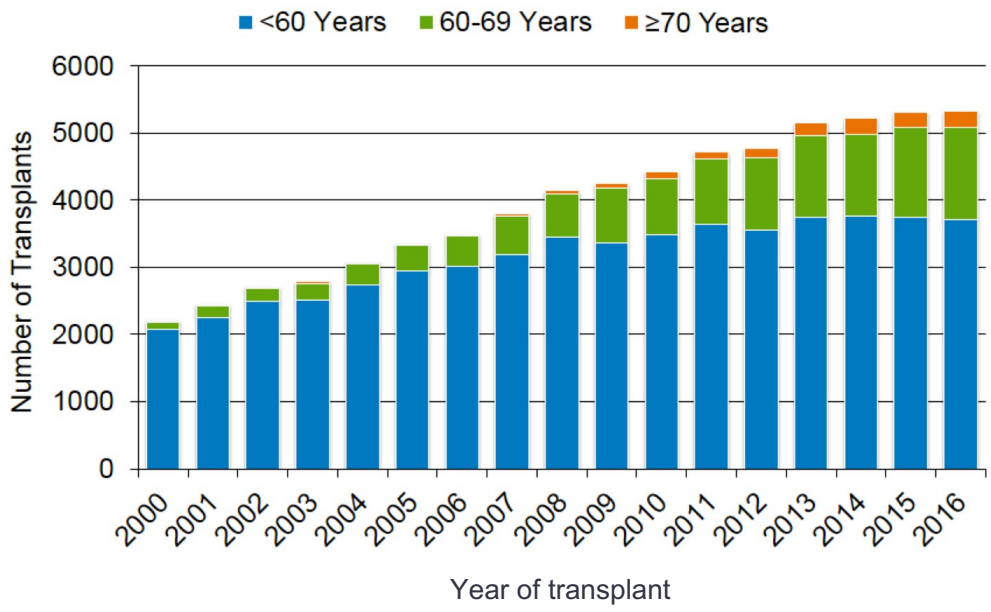


Figure 1.8 CIBMTR summary slide of allogeneic HCT patient age over time

This chart of HCT frequency between different age groups demonstrates that elderly patients (≥60 Years) are the current fastest growing group of HCT recipients, a factor largely related to improvements in RIC transplantation. HCT indications included for this chart: acute myeloid leukaemia (AML), acute lymphoblastic leukaemia (ALL), non-Hodgkin's lymphoma, Hodgkin's disease and multiple myeloma. Reproduced with permission [193].

1.05.01.03 Non-myeloablative conditioning

Non-myeloablative conditioning reflects an even less potent conditioning, from which a patient would be expected to regain haematopoiesis even without receiving a donor HSC infusion. However, sufficient immunodepletion is achieved to allow donor cell engraftment, although recipients often endure a prolonged period of mixed chimærism [194].

1.05.01.04 T cell depletion (TCD)

T cells are one of the main subsets of cells involved in targeting residual diseased cells. However, T cells educated outside of the recipient's body will also recognize their new host as "foreign" and stimulate an immune attack known as graft *versus* host disease (GVHD, discussed further in Section 1.05.03). Accordingly, standard T cell replete transplants rely on administration of immunosuppressive drugs after cell infusion to prevent GVHD. Unfortunately, the rate of GVHD is still high following T cell replete HCT. As such, antibody preparations have been designed to deplete the number of T cells from the infused cells either as *ex vivo* T cell depletion (TCD), whereby agents are applied to the donor HSCs prior to infusion; or, more commonly, *in vivo* TCD, by which the TCD agents are given to the patient as part of their conditioning regimen [195]. There are two main TCD agents: antithymocyte globulin (ATG) and alemtuzumab. ATG is an antibody treatment that is prepared by infusing human thymocytes into a non-human animal (rabbits and horses are commonly used) before isolating the antibodies that the animal's immune system raises against the human thymocytes. When applied as a medical treatment, this polyclonal antibody mixture binds to the recipient's T cells and stimulates their depletion, thus reducing GVHD [196,197]. Alemtuzumab, a humanised monoclonal antibody, works in a similar fashion, but is specific to a single antigen which marks mature lymphocytes, CD52

[198]. Importantly, neither of these immunosuppressive treatments binds to HSCs, therefore allowing their engraftment in the recipient [199]. With relevance to this project, TCD also favours recovery of NK cells relative to T cells, perhaps giving opportunity for KIR-mediated effects to be observed more clearly [200].

1.05.02 Donor selection

The topic of the "best" donor is one of intense controversy, with many factors being involved and it being highly unlikely that the most optimal donor for a patient with an aggressive malignancy be identical to that for a patient with an immunodeficiency. Although the section below is focussed towards HLA matching, each of the following factors have also been implicated in modulating the outcomes of HCT: donor age, patient-donor CMV status matching, cell dose, donor gender, blood group matching, and, of course, patient disease [201-207]. However, broadly, the right balance of genetic parity is vital to the success of HCT.

1.05.02.01 HLA matching

In the initial experiments investigating HCT in humans, HSCs were infused from a wide variety of unusual sources [208]. However, it wasn't until the HSCs from identical twins of the recipient were transplanted that success was observed. The theory followed observations in other forms of transplantation that suggested that related, and thus genetically similar, individuals would share the factors that prevented rejection of the graft [209]. However, not all relatives are equally compatible, leading to the discovery of the genes responsible for this transplant compatibility: HLA. As the understanding of immunogenetics and testing methods improved, HLA-matched siblings were identified as potential donors, and are still currently considered the preferred donor option if available [210,211]. These advances also led to efforts to identify suitable unrelated

donors, for patients without a suitable related donor, leading to the first successful volunteer unrelated donor (VUD) transplantation in 1973 [212]. Since then, a vast amount of research has been performed into how different elements of the immune system, both cellular and genetic, interact with one another to affect the outcome of transplantation, resulting in the outcomes from VUD HCT approaching a similar success rate to those of HLA-matched siblings [213].

The current gold standard for genetic matching in VUD HCT is based on allelic-resolution matches for the HLA-A, HLA-B, HLA-C, HLA-DRB1 and HLA-DQB1 alleles (referred to as 10/10) [205,214,215]. There is also a growing body of evidence demonstrating the importance of T cell epitope (TCE) matching for HLA-DPB1 alleles, whereby an allelic match is not necessarily required (although preferred [216]), but rather matching for a broader range of alleles that share particular motifs confers better outcomes [217-221]. Recent research has explored the permissibility of other HLA mismatches as a means to find additional acceptable donors for those patients for whom a 10/10 matched donor is not available. For example, the HLA-C*03:03/03:04 mismatch has been demonstrated to be more permissible than other HLA-C mismatches, presumably due to the lower expression of these alleles compared to other HLA-C [222,223]. HLA mismatching has been found to be particularly permissible in HCT utilising umbilical cord blood units (CBU), proposed to be due to the naïvety of the infused immune cells [224]. As such, HSC source forms another interesting area of discussion around donor selection.

1.05.02.02 Donor and HSC sources

HLA-matched, related individuals continue to provide the current first choice donor in most transplantation scenarios due to the favourable degree of genetic parity, reduced

cost and rapid availability of alloreactive cells. However, due to average family sizes, and some relatives being unsuitable or unavailable as donors, it has been estimated that only approximately 30% of patients have such a donor available. In addition, patient age and ethnicity have been suggested to cause considerable variation to this figure meaning that, for many patients, transplantation using cells from an HLA-matched, related donor is simply not an option [225].

Initial studies utilising related individuals who share only one HLA haplotype with the patient, so called "haploidentical" donors, had very poor results as a direct result of HLA disparity. This presented as high rates of graft rejection and, in the patient subgroup who did engraft, unacceptable mortality from GVHD [226,227]. However, owing to improvements to transplant practice such as TCD and use of post-transplant cyclophosphamide, haploidentical donors are now also considered an acceptable donor source, provided that sufficient immunosuppressive measures are in place to reduce GVHD that may result from the strong genetic disparity across the mismatched HLA haplotype [228]. In this scenario, each parent and child of a patient becomes a potential donor, as do 50% of siblings. As such, haploidentical transplantation greatly increases the probability of finding a donor, but at risk of developing GVHD and/or prolonged immune reconstitution [229].

An alternative to using related donors is to search the worldwide network of VUD registries. At the time of writing, 35.9 million VUDs and cords were listed on the World Marrow Donor Association (WMDA) database [230]. This large database of donors increases the probability of finding a 10/10 HLA-matched donor, so much so that 69% of UK HCT patients are able to find an HLA compatible donor. However, this

probability is greatly reduced for patients of black, Asian or other ethnic minorities, to approximately 20% [231]. Furthermore, despite 10/10 HLA matching, the degree of genetic disparity across the remainder of the genome can still result in alloreactivity complications post-transplant. In addition, attrition and unavailability of VUDs can prolong pre-transplant waiting periods, increasing the mortality and relapse rates at this stage [232,233]. However, the wider range of donors, and the associated increased probability of finding HLA-compatible donors with other favourable, non-HLA factors such as young donor age and CMV-matched status, make this an attractive option when considering potential donors.

HSCs from each of the above donor options can be obtained from two sources: BM and mobilised peripheral blood stem cells (PBSCs). Historically, BM was used as the only available cell source [234,235]. However, the desire for less intrusive donation, and owing to improvements in HSC mobilisation pharmacology, PBSC donation now forms the majority of both related and unrelated donations [236]. In addition, it has been demonstrated that haematopoietic recovery is faster following PBSC transplantation, although the presence of T cells in the graft has been associated with an increased risk of acute GVHD (aGVHD) in several HCT scenarios [237,238].

Due to difficulties in finding suitable donors for patients from ethnic minorities or with unusual HLA genotypes, HCT with CBU was proposed as a method to circumvent stringent HLA matching. This is possible as a result of cord blood HSCs being naïve and thus less likely to cause chronic GVHD (cGVHD), allowing HLA mismatching that would not be permissible from adult donor sources [239]. As such, cryogenic banks of frozen CBU with relevant testing performed in advance have been established across

the world. Certain drawbacks do exist, however. Due to the smaller doses of HSCs from CBU, pre-engraftment periods are extended, increasing the risk of infection [240,241]. Additionally, cell doses per CBU vary and can often be insufficient for adult patients, resulting in costly double CBU transplants [242].

1.05.03 Complications following HCT

Although HCT has dramatically improved since its conception, to the point where the transplant outcomes from VUDs are comparable to that of an HLA-identical sibling, only a minority of recipients undergo a transplant without complication [243].

1.05.03.01 Graft versus host disease

Our understanding of, and ability to perform, genetic matching has vastly improved over the past 40 years. However, complications as a direct result of genetic disparity between patient and donor still arise. GVHD, an alloreactive response which occurs when the transplanted donor cells recognise the recipient's own cells as foreign and mount an immune response, can still occur despite HLA 10/10 matching. Although studies have shown that frequency and severity of GVHD may be affected by many different factors, it is estimated that 40-60% of allogeneic HCT recipients will develop GVHD post-transplant, although in many cases the severity may be relatively minimal [244].

The classical categorisation of GVHD is dependent on the time point of onset of GVHD symptoms. Acute GVHD typically begins within the first 100 days after HCT and can be triggered by tissue damage caused by the chemo/radiotherapy agents given as part of the conditioning regimen. It typically affects the skin, gut and/or liver tissues causing rash, diarrhoea and/or liver damage [245]. Variation in the severity of symptoms and

clinical presentation led to the development of a severity grading system, the basis of which is still widely employed today [246].

Chronic GVHD is a result of alloreactive T cells from the donor recognising the recipients' cells as foreign and, according to the classical definition, symptoms do not present until after 100 days following the HCT. Symptoms of cGVHD are also more variable than aGVHD and can affect most organ systems. Although treatable with immunosuppressive therapy, GVHD symptoms may persist for years after transplant and may cause severe morbidity [247].

In addition to the classical classification above, several revisions have been made to GVHD categorisation to account for clinical symptoms, as opposed to time of onset. Included in these alterations are the novel categories: late onset aGVHD, which describes symptoms normally associated with aGVHD presenting at time points past 100 days post-transplant; and overlap syndrome, a condition whereby symptoms of both acute and chronic GVHD present concomitantly [248,249]. Novel scoring systems to assess risk of GVHD severity have also been proposed [250].

1.05.03.02 Relapse

Disease relapse, the return of initial disease and another leading cause of mortality following HCT for malignant disease, is also linked to histocompatibility between donor and recipient. This is because the same alloreactive T cells that cause GVHD are also responsible for a phenomenon known as the GVT effect, also known as graft-*versus*-leukaemia (GVL), in which the transfused immune cells recognize diseased cells. As such, a certain degree of disparity between the grafted cells and the host's own may help to eliminate residual cancerous cells capable of causing relapse. The same

principle also supports the use of donor lymphocyte infusion (DLI) following TCD HCT to aid in elimination of persisting disease, reducing the risk of relapse [251]. If, however, the combination of conditioning and immune-mediated tumour reduction is not sufficient to maintain remission in the patient, return of the tumour to detectable levels is known as relapse.

1.05.03.03 Delayed engraftment and infection

Following the immunoablative conditioning prior to transplant, recipients' immune systems are severely compromised. The reconstitution of the immune system may begin to occur following donor cell engraftment. However, the immune systems of HCT recipients remain profoundly supressed as a result of conditioning treatment and GVHD prophylaxis for up to one year post-transplant. Moreover, further delays in immune reconstitution can occur in patients who experience cGVHD, owing to additional immunosuppressive treatments. As such, during the initial post-transplant period, patients are at greater risk of infection from common bacterial and fungal infections, as well as reactivation of viruses including CMV and *Herpes simplex virus* (HSV). Although often trivial in immunocompetent individuals, opportunistic infections such as those above can be fatal in immunocompromised HCT recipients [252].

1.05.04 HCT for acute myeloid leukaemia

There are many indications for HCT. Amongst these are: immunodeficiencies, such as Wiskott-Aldrich syndrome; bone marrow failure syndromes; and blood cancers, including acute myeloid leukaemia (AML) [253].

AML is a heterogeneous disease and, unlike chronic myeloid leukaemia (CML), that is commonly characterised by the breakpoint cluster region-Abelson murine leukaemia

viral oncogene homologue 1 (BCR-ABL) translocation mutation (also known as the Philadelphia chromosome) [254], AML cannot be defined by a single component mutation, but occurs when a myeloid progenitor cell loses the ability to control its division and the unchecked replication results in the BM niche becoming overly occupied – displacing the healthy blood cells. Symptoms include fatigue, anaemia and contusion [255].

Several common disease-causing genetic mutations, such as internal tandem duplication of FMS-like tyrosine kinase 3 (FLT3-ITD) and single point mutations in nucleophosmin 1 (NPM1) and isocitrate dehydrogenase 1 and 2 (IDH1 and 2), are able to generate response estimates for the outcome of induction therapy [256]. This is usually a course of continuous infusion of a combination of cytarabine and anthracycline and is usually followed by consolidation therapy consisting of either further chemotherapy or, for higher risk patients, HCT. Cytogenetic classification of AML disease is also used to identify patients who would benefit from HCT [255].

AML is of particular relevance to this study as, of all the malignancies tested thus far, AML has consistently been shown to be most susceptible to KIR-mediated NK cell immunity in the HCT setting. Interestingly, unlike certain viral infections and other cancers [29,257], HLA class I downregulation does not appear to be an immune evasion strategy inherent to AML [258]. Instead, in addition to the variety of other NK cell receptor ligands expressed on the surface of AML blasts, the mismatch of HLA ligands between patient and donor may provide the ‘missing self’ stimulus required to activate educated donor NK cells [259]. This is discussed in more depth below.

1.05.05 KIR in HCT

Several different models have evolved to consider the effects of KIR-mediated immunity following HCT.

1.05.05.01 The KIR ligand-ligand model

The influence of KIR-mediated immunity on the outcomes of HCT initially focussed on the missing self hypothesis. In healthy individuals, the lack of cognate HLA ligand for an inhibitory KIR does not usually result in NK cell-mediated lysis: licensing prevents these NK cells from being autoreactive. However, NK cells licensed in an allogeneic donor (who expresses both KIR and cognate ligand) that are then transplanted into a patient (who does not express this ligand) are capable of targeting their new host's cells, particularly the leukaemic cells, via this 'missing ligand'.

This theory was first considered in a TCD haploidentical HCT setting. The result was a landmark paper by Ruggeri *et al.* (2002) [259] which demonstrated that immunity resulting from the mismatching of KIR ligands was able to dramatically reduce the incidence of relapse, without an associated increase in aGVHD. Specifically, the probability of relapse at five years was completely abolished in the KIR ligand mismatched group, compared to 75% in the KIR ligand matched recipients ($p<0.0008$). Interestingly, this result was observed only in AML transplants: relapse probability was comparable for KIR ligand matched and mismatched acute lymphoblastic leukaemia (ALL) recipients (90% *vs* 85%, respectively, p=non-significant) [259]. Several other studies have repeated this experiment in haploidentical cohorts, both with and without TCD, and in partially matched VUD and CBU transplants. On many occasions, the results have confirmed the original findings. However, contradictory findings have also been published, including examples where no benefit was observed and others where

detrimental outcomes, such as increase in risk of GVHD, have been observed [260-265].

1.05.05.02 The KIR receptor-ligand model

Although findings in the KIR ligand-ligand model suggested that the ligands alone could elicit this effect, it was always known that specific KIR were responsible for detecting these ligands (or their absence) and that, as these KIR were not inherited on the same chromosome as their HLA ligands, their presence or absence on donor NK cells could vary compared to the recipient. As such, the next iteration of a KIR alloreactivity model included information not only on the ligands involved, but also the donor KIR genotype.

This model, known as the KIR receptor-ligand model, suggested that absence of cognate ligand in the recipient for any donor-encoded KIR, regardless of previous licensing status, was suggestive of improved risk of relapse, and proposed to be a more accurate prognostic tool in HLA-mismatched transplants [266]. These results have also been repeated in multiple different transplantation scenarios [267-269].

More recent permutations of this theory have investigated certain polymorphisms of both the KIR and ligand; associating low functionality KIR receptor-ligand interactions with reduction in disease relapse. For instance, alleles of KIR2DL1 exhibit different degrees of functional strength according to a dimorphism at residue 245, responsible for efficient recruitment of ß-arrestin 2, as discussed in Section 1.03.08.02. This dimorphism has been shown to affect the outcomes of HCT for AML and ALL: donors who encode at least one highly functional KIR2DL1 allele correlate with less disease

progression and improved overall survival in an apparent dominant effect of this allelic polymorphism [270].

Conversely, a weaker interaction between KIR3DL1 and certain HLA-Bw4 has been proposed to reduce relapse in AML patients [271]. In this scenario, polymorphism between different HLA-Bw4 alleles, characterised by the presence of isoleucine or threonine at residue 80, corresponds with high and low affinity interactions, respectively. In the study by Marra *et al.* (2015) [271], patients who encoded the high affinity KIR3DL1-HLA-Bw4-80Isoleucine(I) receptor-ligand pair experienced higher rates of disease relapse following autologous HCT. This effect was compounded by copy number of HLA-Bw4-80I, such that each additional high affinity allele corresponded with an increase in relapse risk.

1.05.05.03 The donor KIR genotype model

Although findings from the KIR ligand-ligand and receptor-ligand models proposed methods to reduce relapse, studies often utilised haploidentical related HCT, and evidence from large cohorts of HLA-matched VUD HCT was lacking. As it is well-known that HLA matching is the most influential prognostic factor in HCT outcomes and, where possible, an HLA-matched donor is the primary source of HSCs, an approach to best utilise KIR-mediated immunity was much sought-after.

The Minnesota group have pioneered this effort and repeatedly implicated the KIR B haplotype, characterised by a higher proportion of activating KIR genes, in favourable outcomes in HLA-matched (and mismatched) transplants [83,272,273]. Although applicable to the broad KIR B haplotype, sub-analyses revealed that a reduction in relapse was predominantly associated with presence of the Cen-B motif in donors. This

led to the development of a scoring system. Donors encoding two or more KIR B motifs were separated according to the presence of the Cen-BB haplotype motif combination: Cen-BB donors were regarded as the "best" donors, whilst other donors encoding two or more KIR B motifs were denoted the "better" score. Donors encoding one or no KIR B motif received the "neutral" donor score. The "neutral", "better, "best" scoring system was found to be particularly effective when the recipient encodes at least one copy of an HLA-C1 ligand [272,273]. Interestingly, in a subset of patients who were HLA-C mismatched, the effect was enhanced irrespective of whether the HLA-C mismatch resulted in a KIR ligand mismatch. Although relapse and disease-free survival (DFS) were shown to improve with donor-encoded KIR B status, an increase in chronic, but not acute, GVHD was also observed [83].

The influence of the donor KIR B haplotype has been repeated elsewhere in a variety of scenarios including different diseases and HLA matching scenarios [263,274-276]. Importantly, when investigating KIR-mediated relapse reduction in HLA-matched transplants to treat acute leukaemias, the findings have corroborated most previous investigation to suggest that the beneficial effects are limited to AML and are not observed in ALL recipients. However, this remains disputed in the haploidentical, HLA-mismatched HCT scenario [277]. Additionally, as with the models before, investigations in separate cohorts have found opposing results, whereby presence of the KIR B haplotype in donors can have a negative impact on survival following HCT, a finding which may be linked to other transplant factors such as the use of TCD and/or low intensity conditioning regimens [278-280]. Indeed, the Minnesota group's own recent findings have strongly implicated the era of transplant with the relevance of donor Cen-B motif presence [281].

1.05.05.04 The KIR matching model

In addition to the relatively complex concepts mentioned above, a more simplistic model, closely resembling HLA matching, has also been proposed. This model, originally a comparison of the occurrence of GVHD between HLA-identical sibling and HLA-matched unrelated donor transplants compares the presence/absence status of each KIR gene between the donor and recipient [282]. The original findings suggested that mismatches in the GVH direction were detrimental in the unrelated setting, but beneficial when the patient and donor were related. Adaptations of the KIR matching model, investigating genotype mismatching (KIR AA *vs* BX) or separating mismatches according to the activating or inhibitory KIR functional groups, have also implicated genetic disparity at the KIR loci with increased GVHD [283,284].

The findings from each of the models outlined above have strongly implicated polymorphism within KIR haplotypes in affecting the outcomes of HCT. However, with each analysis, nuances within the findings have cast doubt over exactly how best to utilise KIR typing information to improve HCT. However, many of the analyses have concluded in a similar manner: in a system a complex as KIR, whereby each gene exhibits extensive allelic polymorphism as well as copy number variation, investigation into the encoded alleles of each KIR gene may help to better decipher the true role of KIR in influencing the outcomes of HCT.

1.06 Thesis scope

The obvious benefits that NGS offer have been noted in immunogenetics laboratories, leading to the rapid adoption of NGS by the immunogenetic field for routine HLA typing. In addition, the sequencing range, ability to phase distant polymorphisms and

high-throughput nature of long-read technologies make them an attractive choice for clinical laboratories [285]. In addition, HLA polymorphism is already well defined, and several PCR approaches to generate suitable sequencing amplicons for HLA previously designed, making the transition from older technologies to NGS relatively simple. By applying this style of sequencing to the KIR locus, I hope to better define what has already been shown to be a clinically relevant genetic system. To this end, I have designed a PCR-based long-read technology sequencing strategy for each of the inhibitory KIR loci as well as KIR2DS2, KIR2DS5, KIR3DS1 and KIR3DP1. This has been validated against a panel of previously-characterised cell lines from the International HLA and Immunogenetics Workshop (IHIW) cell lines [286].

In addition, I have applied a previously designed PCR-SSP genotyping strategy to determine the role of KIR presence/absence polymorphism in a large, retrospective cohort of UK AML patients. This cohort differs from many of the other larger cohorts in its predominant inclusion of alemtuzumab-based TCD transplants. By KIR genotyping this cohort, I have attempted to confirm the role of KIR in this scenario. In addition, it has assisted in guiding future analyses.

However, the complexity of KIR polymorphism is unlikely to be resolved simply by assessing the presence or absence of particular genes or gene motifs and, as such, requires a higher resolution of typing. Conversely, complete definition of each KIR locus at nucleotide resolution for a large cohort of patients and donors is beyond the realm of this project. As such, I have selected three key loci to study clinically: KIR2DL1, KIR2DL2/3 and KIR3DL1/S1. These loci have been relatively well-characterised previously: their ligands are known (and, conveniently, previously

sequenced at ultra-high resolution as part of a related project [216]), and the functionality of different alleles at each locus has been proposed. The remainder of the KIR locus not conferred by KIR2DL1, KIR2DL2, KIR2DL3, KIR3DL1 and KIR3DS1 undoubtedly holds many more, interesting secrets. By developing the PCR amplification and sequencing strategy for some of these other loci, I hope to encourage further investigation into their function and clinical role in the future. This is a complex and exciting topic of immunogenetics, where little relating to the influence of allelic KIR polymorphism on the outcomes of HCT has been described. I hope the remainder of this thesis begins to decipher some of this complexity and starts to answer some of the questions about the influence of allelic KIR polymorphism in HCT.

Chapter 2 Materials and methods

2.01 Study samples

2.01.01 Cell line samples

To validate the amplification and SMRT sequencing methods, 66 well-characterised B lymphoblastoid cell lines (BLCLs) available at the Anthony Nolan Research Institute were selected (see Supplementary Table A). Although some BLCLs were previously characterised at genomic sequence level for each of the KIR loci [87], the majority (80%) had only been studied at CDS level [94]. In addition, a further subset of BLCLs (6%) with only presence/absence data available were also studied. All BLCLs included originate from the IHIW cell line studies [286].

Vials of BLCLs frozen in liquid nitrogen were removed and thawed at room temperature by gradual addition and removal of Roswell Park Memorial Institute (RPMI) 1640 (Fisher Scientific, Loughborough, UK). Cells were then washed and pelleted in a 20 millilitre (mL) Sterilin tube (ThermoFisher Scientific, Waltham, MA, USA) by centrifugation at 1,600 revolutions per minute (rpm) for 3 minutes (mins). The supernatant was discarded and the wash step repeated in 10 mL of RPMI 1640. The supernatant was discarded and the cells resuspended in 20% foetal bovine serum (FBS, Fisher Scientific), 1% Penicillin-Streptomycin (Fisher Scientific) RPMI-1640-based media. Cell suspensions were incubated in T25 tissue culture flasks (Sarstedt, Nümbrecht, Germany) at 37°C in 5% CO_2 at approximately $1x10^6$ cells per mL. Once BLCLs entered exponential growth phase, FBS concentration was reduced to 10%. Viability and total cell count were determined by Trypan Blue (Sigma-Aldrich,

Gillingham, UK) staining cells in a 1:1 ratio and counting all live and dead cells within a 1 millimetre-squared (mm^2) region of a haemocytometer (Appleton Woods Ltd., Birmingham, UK). The percentage of living cells from the total number of cells counted was calculated to determine the viability, whilst total cell count accounted for total culture volume and dilution factor within the stain.

Once total cell count was sufficient to restore cryopreserved stock and provide at least $10x10^6$ cells for DNA extraction, aliquots of $5–10x10^6$ cells were pelleted by centrifugation at 1,600 rpm for 3 mins and either frozen at -20°C to be used as material for DNA extraction, or frozen in 10% dimethylsulphoxide (DMSO, VWR, Lutterworth, UK), 90% FBS freezing media for 24 hours at -80°C before transfer into liquid nitrogen.

2.01.02 Clinical study population

To determine the importance of KIR polymorphism in HCT outcomes, paired samples from patients and their respective VUD were studied. A long-running study exists at Anthony Nolan to collect these samples. Briefly, eligibility criteria for inclusion in this study are as follows: donors must be registered within the United Kingdom (UK) aligned registry and patients must have had their transplant at one of the UK hospitals registered within the study (Table 2.1). Although the wider study recruits patients being treated with HCT for all diseases, this study only investigates the effects of KIR polymorphism on the outcomes of AML and, as such, only utilises transplants where AML was the indication for HCT. All samples had previously undergone high resolution HLA typing from which the KIR ligand status (HLA-Bw4/C1/C2) was determined [216]. Clinical follow-up data, used to determine the effects of different

KIR parameters on the outcomes of HCT, were collected by dedicated data managers in collaboration with the British Society for Blood and Marrow Transplantation (BSBMT).

Table 2.1 Number of patients contributed from different transplant centres

Transplant Centre	Number of patients	Transplant Centre	Number of patients
Royal Marsden Hospital	53	Churchill Hospital	10
Nottingham City Hospital	46	Guy's Hospital	8
King's College Hospital	40	Royal Manchester Children's Hospital	6
The Christie Hospital	25	St James' University Hospital	5
Bristol Royal Hospital for Children	23	Birmingham Children's Hospital	5
Royal Free Hospital	21	Great Ormond Street Hospital	4
University College London Hospital	21	Glasgow Royal Hospital for Sick Children	4
Royal Liverpool University Hospital	20	Hammersmith Hospital	4
Glasgow Royal Infirmary	18	Beatson West of Scotland	4
Manchester Royal Infirmary	15	Royal Hallamshire Hospital	3
Southampton General	14	Alder Hey Children's Hospital	3
Queen Elizabeth Hospital, Birmingham	11	St. George's Hospital	3
St Bartholomew's Hospital	11	University Hospital of Wales	3
Freeman Hospital	11	Sheffield Children's Hospital	2
Birmingham Heartlands Hospital	10	Addenbrooke's Hospital	2

Blood samples in ethylenediaminetetraacetic acid (EDTA) vacutainers (Sarstedt) were drawn from the donor by the physician attending their medical assessment prior to donation, whilst samples from the patient were obtained either as EDTA blood prior to conditioning commencement or as buccal swabs (Fisher Scientific) if it was not possible to obtain blood samples. All samples were then sent to Anthony Nolan Research Institute, London, where they were processed. At the time that samples were obtained, all participants gave written, informed consent for the use of their samples in genetic analysis research studies. Clinical outcomes data were obtained by the British Society of Blood and Marrow Transplantation and kindly provided to Anthony Nolan Research

Institute. Ethical approval was obtained from the National Research Ethics Service (www.nres.nhs.uk, application number: MREC 01/8/31). The project was also approved by Anthony Nolan medical and scientific committees.

2.02 DNA extraction

To ensure reliable amplification of long targets, template DNA must be high quality and contain fragments with high molecular weight (MW). As such, DNA extraction protocols were selected based on the quality, not quantity of DNA obtained. Three different DNA extraction protocols were used for extraction of DNA in this study: two different salting-out procedures, and a semi-automated paramagnetic bead-based purification. These will be discussed in turn below. Each DNA sample was subjected to qualification and quantification as discussed in later sections of this chapter.

2.02.01 DNA extraction by salting out from blood and cell line samples

In a method similar to Miller *et al.* (1988) [287], salting out was used as a method of DNA extraction for both EDTA blood and cell line culture samples. Briefly, 1 mL of each sample was mixed with 14 mL red cell lysis buffer (RCLB: 10 mM Tris-hydrogen chloride (HCl), pH 8.0, ThermoFisher Scientific; 5 mM magnesium chloride ($MgCl_2$), VWR; 10 mM sodium chloride (NaCl), VWR) in a 15 mL Falcon tube (Sarstedt) and allowed to stand for 10 mins. Centrifugation at 3,200 rpm was then performed to pellet intact cells. The supernatants were discarded, cell pellets resuspended, and the RCLB wash steps repeated again until the pellets no longer contained red cells. Aliquots of 20 µL sodium dodecyl sulphate (10% SDS, Fisher Scientific), 30 µL proteinase K (10 mg/mL, Sigma-Aldrich), 80 µL of proteinase K buffer (10 mM Tris-HCl, pH 7.5; 10 mM EDTA, ThermoFisher Scientific; 50 mM NaCl) and 240 µL water were added to the cell pellets, followed by overnight incubation at 37°C. The following day, cell

lysates were transferred to 1.5 mL microcentrifuge tubes (Elkay, Basingstoke, UK) to which an aliquot of 100 µL of 5 M NaCl was added to complete the lysis of cells. The solution was vortexed for 15 seconds (secs) before centrifugation at 13,000 rpm for 5 mins. Each cell lysate supernatant was then added to a 15 mL Falcon tube containing 1 mL 100% ethanol (VWR) at -20°C. If DNA immediately precipitated and became visible, it was spooled onto a glass pipette (Fisher Scientific) and allowed to air dry for 20 mins before being rehydrated in at least 50 µL of water. If DNA precipitate did not immediately become visible, the mixture was kept at -20°C overnight, and centrifugation (13,500 rpm, 20 mins) was used to pellet DNA. Following supernatant discard, DNA pellets were air dried for at least 20 mins before being rehydrated in 50 µL of water.

2.02.02 DNA extraction by salting out from buccal swab samples

Buccal swab samples were collected by scraping a cytobrush along the upper and lower buccal gutter for approximately 30 secs. Gentra puregene technology (QIAGEN, Manchester, UK) was used for extraction of DNA from buccal swab. Unless stated otherwise, reagents and plasticware were part of the Gentra puregene kit. Briefly, two swabs from each individual were incubated together with 300 µL Cell Lysis Solution and 3 µL proteinase K (10 mg/mL) in a 1.5 mL microcentrifuge tube at 55°C overnight. The following morning, the swabs were removed and 1.5 µL RNase A solution was added to the tube, mixed by inversion, and the solution incubated at 37°C for 60 mins. Following this, an aliquot of 100 µL of protein precipitation solution was added to the solution. This was then incubated on ice for 5 mins before centrifugation at 13,500 rpm for 5 mins. The supernatant was then transferred to fresh 1.5 mL microcentrifuge tube with 300 µL isopropanol (VWR) and 1 µL glycogen (QIAGEN) to precipitate the DNA.

These solutions were gently mixed by inversion before undergoing centrifugation at 13,500 rpm for 5 mins. The supernatants were discarded, and pellets washed in 300 µL of 70% ethanol. The tubes then underwent repeat centrifugation, before the supernatant was discarded and the DNA pellets allowed to air dry for at least 20 mins. The DNA pellets were then rehydrated in 50 µL of water.

2.02.03 Paramagnetic bead-based DNA purification from blood samples

Although the quality of DNA obtained by salting out is often very high, it requires large amounts of "hands-on" time to complete. To reduce this, an automated paramagnetic bead-based DNA purification was used to extract DNA from EDTA blood samples. All reagents and plasticware were provided as part of the Maxwell Whole Blood DNA kits (AS1520, Promega, Madison, WI, USA). To extract DNA by this method, 300 µL of blood was mixed with 300 µL of Lysis Buffer and 30 µL of Proteinase K and incubated at 56°C for 20 mins. The cell lysate was then transferred to the first well of a Maxwell reagent cartridge and loaded onto the Maxwell® 16 automated robotic DNA extraction system (Promega). Extraction was performed as per manufacturer's guidelines. Briefly, this involved the mixing of paramagnetic beads with the lysed blood sample before transferal through a series of wash steps to purify the DNA. Final elution took place in 50 µL of Elution Buffer within a 0.6 mL microcentrifuge tube.

2.03 DNA quantification and fragment length determination

2.03.01 DNA fragment length estimation by agarose gel electrophoresis

DNA fragment size was assessed by agarose gel electrophoresis for two separate purposes in this study. First, freshly extracted DNA samples underwent agarose gel electrophoresis to determine DNA quality: highly fragmented or low MW DNA was

deemed to be of insufficient quality, and thus these samples were re-extracted. Contrarily, if high MW bands (>20 kbp) were observed in the absence of smaller, contaminating fragments, these samples were considered to be high quality and were used in downstream processes. The second use of agarose gel electrophoresis was to assess amplicon length following PCR-based target amplification.

To prepare agarose gels, powdered agarose (GENEFLOW Ltd., Lichfield, UK) was combined with the appropriate volume of 1% Tris-Borate-EDTA (TBE, Lonza, Slough, UK). Agarose gels to assess genomic DNA quality were made at 1% (mass per volume), whilst agarose gels to assess PCR product fragment length were made at 1.5-2% (mass per volume). Agarose was melted into solution using a microwave until completely dissolved. The agarose solution was then mechanically stirred until cool enough to safely combine with ethidium bromide stain (0.00005% volume per volume, Sigma-Aldrich). Stained agarose solution was then poured into sealed trays and combs were added. Agarose gels were allowed to set for approximately 30 mins before submerging in 1% TBE and removal of the combs. Gel combs, trays and tanks were all supplied by Fisher Scientific.

Samples were mixed with loading buffer (Bioline, London, UK) before being loaded into the wells of the gel. Lanes within each row were reserved for MW ladders (Figure 2.1) Electrophoresis was performed at 150 volt (V) until adequate fragment separation had occurred. Fragment size estimation could then be made against the relevant ladder bands by observing the gel in an Essential V6 UV transilluminator (Thistle Scientific, Glasgow, UK).

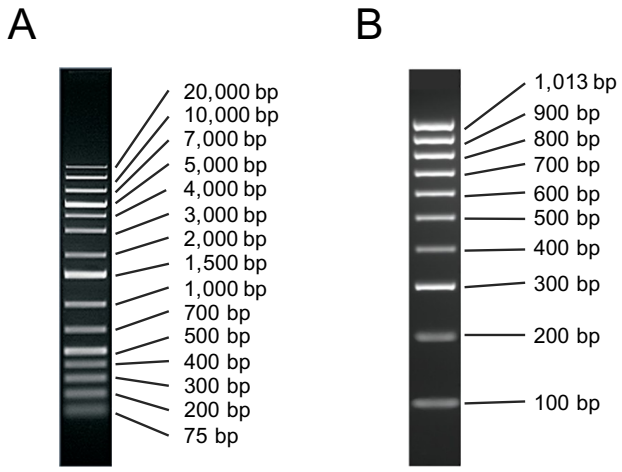


Figure 2.1 Ladders used in agarose gel electrophoresis to determine MW

A) GeneRuler™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific) was used to estimate fragment lengths of recently extracted genomic DNA and PCR amplicons of full length KIR genes. B) HyperLadder™ 100bp (Bioline) was used to estimate fragment lengths for all KIR presence/absence genotyping. Images adapted from manufacturer datasheets.

2.03.02 DNA quantification by NanoDrop™ spectrophotometry

To determine the concentration of DNA samples and to enable standardization of concentrations between different samples, DNA samples were analysed using the NanoDrop One platform (ThermoFisher Scientific). Aliquots of 1.5 µL of sample were loaded onto the pedestal and analysed to determine the absorbance of 220-340 nm wavelength light. In addition, NanoDrop One is capable of detecting several contaminants, including phenol and protein, and correcting the DNA concentration to give more accurate DNA concentration measurements.

As well as determining the concentration of DNA, spectrophotometry also measures DNA purity. Commonly quoted measures are the ratios of absorbance at 260/280 and 260/230 nm. Providing that the extracted DNA is pure, the value of both of these ratios

should be in excess of 1.6. As such, we used these values as a proxy for DNA quality in addition to the agarose gel electrophoresis results.

2.03.03 Fragment length and concentration determination by Fragment Analyzer

Although agarose gel electrophoresis is able to measure approximate DNA fragment length, it is very limited in the accuracy of estimation of individual DNA fragment concentrations. Additionally, photospectrometry is only capable of assessing overall DNA concentration of a sample. As such, to determine the concentrations of individual fragments within my PCR amplicons, the Fragment Analyzer platform (Advanced Analytical Technologies Inc., Milton Keynes, UK) was used to simultaneously assess DNA fragment lengths and concentrations across 48 wells. This parallel capillary-based gel electrophoresis system uses a fluorescent intercalating dye to bind DNA fragments that are separated over a large distance (33 cm) before excitation by light-emitting diode (LED) and detection by a CCD. This results in high resolution, accurate fragment sizing. In addition, the intensity of fluorescence accurately measures DNA fragment concentration.

Unless otherwise stated, all reagents were supplied as part of the Standard Sensitivity Large Fragment Analysis kit (DNF-492, Advanced Analytical Technologies Inc.). Briefly, 22 µL of diluent marker was mixed with 2 µL of PCR product in a 96-well plate (Fortitude, Wotton, UK). In one well of each plate, 22 µL of diluent marker was instead mixed with 2 µL of MW ladder. The plate was sealed using PCR plate seals (ThermoFisher Scientific) and mixed by vortexing vigorously for 5-10 secs. Sample plates then underwent centrifugation at 900 rpm for 3 secs. A single droplet of mineral oil (Sigma-Aldrich) was added to each well of the sample plate before being loaded on

the Fragment Analyzer platform. Sample injection was performed by 3 kV for 5 secs and separation occurred over 50 mins at 6 kV.

Analysis of Fragment Analyzer electropherograms was performed in ProSize software v2.0.0.51 (Advanced Analytical Technologies Inc.). Automatic identification of individual peaks was determined by those with at least 200 relative fluorescence units (RFU) and not occurring within less than 5 secs of each other. However, each electropherogram was inspected manually and corrections made according to operator preferences.

2.04 KIR genotyping by PCR-SSP

To determine the presence or absence status of each KIR gene and pseudogene, a PCR-SSP-based genotyping method was employed, as previously described [178]. Briefly, short fragments of KIR genes were amplified using primers specific to individual KIR loci. In addition, a control primer pair targeting a monomorphic region of the ubiquitous gene, HLA-DRA, was included in each reaction to detect false negative results. The primer sequences, reaction reagents and PCR cycling conditions are given in Table 2.2, Table 2.3 and Table 2.4, respectively. Each KIR locus utilised identical PCR conditions.

All primers were supplied at 100 µM concentration and desalt purity (Sigma-Aldrich). Primer dilutions were made in 0.6 mL microcentrifuge tubes (Sarstedt) with water. PCR chemistry (Ammonium [NH_4] Reaction Buffer, $MgCl_2$, dinucleotide triphosphates [dNTPs] and BIOTAQ DNA Polymerase) was supplied by Bioline. PCR steps were performed in 0.2 mL PCR tube strips (ThermoFisher Scientific) and took place within either Veriti 96-well (Applied Biosystems, Foster City, CA, USA), C1000 Touch (BIO-

RAD, Watford, UK) or PTC-200 (MJ Research, Quebec, Canada) thermal cyclers. All reactions were set up at 4°C and, following amplification, PCR products were stored at 4°C.

Table 2.2 KIR genotyping by PCR-SSP amplification primer specificities

KIR target	5′ primer		3′ primer	
	Sequence (5′-3′)	Conc. (µM)	Sequence (5′-3′)	Conc. (µM)
KIR2DL1	GTTGGTCAGATGTCATGTTTGAA	0.5	CCTGCCAGGTCTTGCG	0.5
KIR2DL2	AAACCTTCTCTCTCAGCCCA	2.5	GCCCTGCAGAGAACCTACA	2.5
KIR2DL3	AGACCCTCAGGAGGTGA	2.5	CAGGAGACAACTTTGGATCA	2.5
KIR2DL4	TCAGGACAAGCCCTTCTGC	2.5	GGACAGGGACCCCATCTTTC	2.5
KIR2DL5	ATCTATCCAGGGAGGGGAG	1.25	CATAGGGTGAGTCATGGAG	1.25
KIR2DS1	TCTCCATCAGTCGCATGAG / TCTCCATCAGTCGCATGAA	2.5	GGTCACTGGGAGCTGAC	2.5
KIR2DS2	TGCACAGAGAGGGGAAGTA	2.5	CCCTGCAAGGTCTTGCA	2.5
KIR2DS3	CTTGTCCTGCAGCTCCT	1.25	GCATCTGTAGGTTCCTCCT	1.25
KIR2DS4	GGTTCAGGCAGGAGAGAAT	2.5	CTGGAATGTTCCGTKGATG	2.5
KIR2DS5	AGAGAGGGGACGTTTAACC	2.5	CTGATAGGGGGAGTGAGT	2.5
KIR3DL1	CCATYGGTCCCATGATGCT / TCCATCGGTCCCATGATGTT	5.0	CCACGATGTCCAGGGGA	2.5
KIR3DL2	CATGAACGTAGGCTCCG	0.5	GACCACACGCAGGGCAG	0.2
KIR3DL3/ KIR3DX1	AATGTTGGTCAGATGTCAG / TTTCTGTGGGCCGTGCAA	5.0	GCYGACAACTCATAGGGTA / GTCACTGGGGGCTTATAG	5.0
KIR3DS1	CATCGGTTCCATGATGCG / CATCAGTTCCATGATGCG	2.5	CCACGATGTCCAGGGGA	2.5
KIR2DP1	CGACACTTTGCACCTCAC	0.625	GGGAGCTGACAACTGATG	0.625
KIR3DP1	GTGTGGTAGGAGCCTTAG / GTACGTCACCCTCCCATGATGTA	2.5	GAAAACGGTGTTTCGGAATAC	1.25
HLA-DRA Control	GAGGTAACTGTGCTCACGAACAGC	0.1	GGTCCATACCCCAGTGCTTGAGAAG / CACGTTCTCTGTAGTCTCTGGG[§]	0.1

[§]An alternative 3′ internal control primer was used in the KIR3DP1 amplification

Table 2.3 KIR genotyping by PCR-SSP reagents

Reagent (conc.)	Volume (µL)
Water	6.4
Primer mix	1.0
NH_4 Reaction Buffer (10X)	1.0
$MgCl_2$ (50 mM)	0.4
dNTPs (50 mM)	0.2
BIOTAQ DNA Polymerase (5 U/µL)	0.06
Template DNA (25-200 ng/µL)	1.0

Table 2.4 KIR genotyping by PCR-SSP cycling parameters

Step	Temperature (°C)	Duration (min:sec)	Repetitions
1	95	2:00	1
2	94	0:10	10
	65	0:40	
3	94	0:20	20
	61	0:20	
	72	0:30	
4	4	∞	1

Samples were considered to possess a particular KIR gene/pseudogene if a correct sized band was observed in agarose gel electrophoresis analysis of the corresponding PCR product (Table 2.5). In addition, variant typing for the commonly occurring insertion/deletion alleles of KIR2DS4 and KIR3DP1 was also possible by the same means. For example, if a single band (excluding the internal control), sized approximately 111 bp, was observed in the KIR2DS4 reaction, not only was it deduced that the sample encoded KIR2DS4, but also that it only encoded alleles carrying the deletion polymorphism, such as KIR2DS4*003, 004, 006, 007, 008, 009, 010, 013, 014 and 018. If, however, that single band was approximately 133 bp length, that sample must only encode alleles not carrying the deletion polymorphism. Should bands be observed at both 111 and 133 bp, a sample was denoted as heterozygous for the allele groups.

From the presence/absence profiles of each sample, it was possible to assign common KIR haplotype structures, specifically KIR AA, BX, Cen-A, Cen-B, Tel-A and Tel-B, as described in Section 1.03.02 [272]. The KIR AA genotype was defined by the presence of KIR3DL3, KIR2DL3, KIR2DP1, KIR2DL1, KIR3DP1, KIR2DL4, KIR3DL1, KIR2DS4 and KIR3DL2, or a lesser combination of these genes. Importantly, presence of any other KIR gene defined the KIR BX genotype. Cen and

Tel motif structure could also be determined, as described by Cooley *et al.* (2010) [273]. Briefly, the presence of Cen-A specific KIR genes in the absence of Cen-B specific KIR genes denotes the presence of the Cen-AA structure, whilst the inverse corresponds to the Cen-BB structure. Presence of both Cen-A and -B specific KIR genes implies the presence of the Cen-AB structure. A similar stratification based on KIR Tel-specific genes applies for the Tel motif structure. Finally, it was possible to designate donors with the "Neutral", "Better" and "Best" nominations proposed by Cooley *et al.* (2014) [272]. This suggests that donors encoding CenB/B are "Best", whilst those encoding two or more B motifs (but not CenB/B) are "Better" donors for AML patients than the remaining donors, who encode one or fewer B motifs and are designated, "Neutral".

Table 2.5 KIR genotyping expected fragment sizes

KIR target	Expected fragment size (bp)
KIR2DL1	142
KIR2DL2	142
KIR2DL3	156
KIR2DL4	131
KIR2DL5	147
KIR2DS1	96
KIR2DS2	110
KIR2DS3	158
KIR2DS4	111/133†
KIR2DS5	147
KIR3DL1	108/109†
KIR3DL2	131
KIR3DL3/KIR3DX1	196/88
KIR3DS1	107
KIR2DP1	141
KIR3DP1	279/398†
HLA-DRA Control	283/608§

§ The alternative 3′ primer used in the KIR3DP1 reaction resulted in a larger internal control amplicon length (608 bp).

† Varying fragment lengths caused by allelic variation.

2.05 Full length KIR gene PCR-based amplification

To better understand the impact of KIR genes on HCT outcomes, polymorphism at the allelic level must be characterised. To study KIR genes at this ultra-high resolution, I developed a PCR-based amplification strategy to be used in conjunction with SMRT DNA sequencing, the full details of which are given in Chapter 4. For enhanced targeting, KIR loci were targeted individually or as pairs. In addition, due to the high rate of recombination between alleles of different loci, a semi-generic amplification strategy was selected. Included in each PCR amplification was an internal control primer pair targeting a short region of the HLA-DRA gene.

Primer sequences, reaction reagents and PCR cycling conditions are given in Table 2.6, Table 2.7 and Table 2.8, respectively. Primer oligonucleotides were purified using a reverse-phase purification cartridge and supplied at 100 µM concentration in water. Dilutions to the required working concentration were made in 0.6 mL microcentrifuge tubes. PCR chemistry (PrimeSTAR GXL Buffer, dNTP Mixture and PrimeSTAR GXL DNA Polymerase) were supplied by Takara Bio Inc. (Paris, France). All reactions were set up at 4°C and all thermal cycling programmes required a manual ‘hot-start’. All PCR steps were performed in Veriti 96-well thermal cyclers. Following amplification, all PCR products were stored at 4°C.

Table 2.6 Full-length KIR gene PCR amplification primer specificities (unbarcoded)

Target	5′ primer		3′ primer	
	Sequence (5′-3′)	Conc. (µM)	Sequence (5′-3′)	Conc. (µM)
KIR2DL1	ACTCCAGAATTTACAGGTGG	12.5	CAGATGGGATTATATGGACATGGTAC	12.5
KIR2DL2/ KIR2DL3	CAGGTTCAAGCTATTCTGATGCC	1.25	CAGATGGGATTATATGGACATGGTAC	1.25
KIR2DL4	GGTCAATGTGTCAACTGCAC	10.0	GTGGGCAGGGGTCAAGTG / GCAATGCCGTTTACAATAGATACGC	10.0
KIR2DL5/ KIR3DP1	GGAGGTTGGATCTGAGACGTGTTG / GCACAGATTTTAGGCATCTTGTGTTCA	10.0	GTGGGCAGGGGTCAAGTG / GCAATGCCGTTTACAATAGATACGC	10.0
KIR2DS2	CAGGTTCAAGCTATTCTGATGCC	10.0	GTGGGCAGGGGTCAAGTG / GCAATGCCGTTTACAATAGATACGC	10.0
KIR2DS5	CCACCTCGGCTTCCAAACA	10.0	GTGGGCAGGGGTCAAGTG / GCAATGCCGTTTACAATAGATACGC	10.0
KIR3DL1/ KIR3DS1	GTCTCAAACTCCTGACCTCGGTTGATC ACT	10.0	GTGGGCAGGGGTCAAGTG / GCAATGCCGTTTACAATAGATACGC	2.0
KIR3DL2	GCACAGAATTCAATCACCTCATGTG	10.0	GTGGGCAGGGGTCAAGTG / GCAATGCCGTTTACAATAGATACGC	10.0
KIR3DL3	AGCATGTAAACTGCATGAGCC	10.0	GTGGGCAGGGGTCAAGTG / GCAATGCCGTTTACAATAGATACGC	10.0
HLA-DRA control	GAGGTAACTGTGCTCACGAACAGC	0.5	GGTCCATACCCCAGTGCTTGAGAAG	0.5

Table 2.7 Full-length KIR gene PCR amplification reagents

Reagent (conc.)	Volume (µL)
Water	9.7*
PrimeSTAR GXL Buffer, Mg^{2+} (5X)	5.0
dNTP Mixture (10 mM)	2.0
5′ target primer†	1.0
3′ target primer†	1.0
5′ control primer†	1.0
3′ control primer†	1.0
PrimeSTAR GXL DNA Polymerase (1.25 U/µL)	0.3*
Template DNA (25 ng/µL)	4.0

* The reactions targeting KIR2DL1 and KIR3DL1/KIR3DS1 used 0.5 µL PrimeSTAR GXL Taq and 9.5 µL water.

† Primer concentrations for each reaction are given Table 2.6.

Table 2.8 Full-length KIR gene PCR amplification cycling conditions

Target	Step	Temperature (°C)	Duration (min:sec)	Repetitions
KIR2DL1	1	95	2:00	1
	2	95	0:20	33
		59	0:20	
		68	10:00	
	3	68	10:00	1
		4	∞	
KIR2DL2/ KIR2DL3	1	95	2:00	1
	2	95	0:20	33
		69	10:00	
	3	69	10:00	1
		4	∞	
KIR2DL4	1	95	2:00	1
	2	95	0:25	34
		72	12:00	
	3	72	10:00	1
		4	∞	
KIR2DL5/ KIR3DP1	1	95	2:00	1
	2	95	0:25	32
		70	12:00	
	3	70	10:00	1
		4	∞	
KIR2DS2	1	95	2:00	1
	2	96	0:30	35
		72	12:00	
	3	72	10:00	1
		4	∞	
KIR2DS5	1	95	2:00	1
	2	95	0:25	11
		74	12:00	
	3	95	0:25	28
		73	12:00	
	4	73	10:00	1
		4	∞	
KIR3DL1/ KIR3DS1	1	95	2:00	1
	2	95	0:10	37
		73.6	10:00	
	3	73.6	10:00	1
		4	∞	
KIR3DL2	1	95	2:00	1
	2	95	0:25	35
		73	12:00	
	3	73	10:00	1
		4	∞	
KIR3DL3	1	95	2:00	1
	2	95	0:25	34
		72	12:00	
	3	72	10:00	1
		4	∞	

The amplification procedure above is optimized for use with DNA barcoded primers. This allows the incorporation of a unique nucleotide sequence tag onto amplicons from

specific samples, ultimately allowing sequences to be decoded and assigned to the correct samples downstream. The DNA barcodes used for full length KIR gene PCR amplification were each 16 nucleotides bound to the 5′ end of each KIR-specific primer sequence. In addition, an extra five nucleotides (GGTAG) was also included at the 5′ end of the barcode as a buffer region used to limit barcode nucleotide loss through DNA degradation occurring during PCR and library preparation. A list of barcode sequences used is given in Table 2.9. All barcode sequences were developed by Pacific Biosciences to have limited binding affinity to regions of the human genome.

Table 2.9 DNA barcode sequences

Barcode ID	Barcode sequence	Barcode ID	Barcode sequence	Barcode ID	Barcode sequence
002	CTATACATGACTCTGC	187	ATGATACACGCGCGAC	300	AGCACAGTCACATGTC
014	CGTCTATATACGTATA	192	TACATATGTCACGCGC	305	CTCTGCTCTGACTCTC
033	AGAGAGAGACATGCGC	204	ATACACTCATGTGCAC	307	TGTATGAGTGTCTGAC
038	TGCTCGCAGTATCACA	205	GCTACGCTATAGACAT	309	ACTGCGAGATACACAC
048	TCACACTCTAGAGCGA	212	CTAGTCTCTATCGCAT	326	CACTATACACTGCGCT
059	GTGCAGTGATCGATGA	219	CTCACACATACACGTC	329	GCGCTCTCTCACATAC
068	ACAGTATGATGTACTC	223	ATATGACATACACGCA	333	CTATACGTATATCTAT
087	ATAGCGACGCGATATA	233	ATCGTATAGTCATACA	338	CACTCGACTCTCGCGT
090	TGTCGTCTATCATGTA	239	TATACGAGATACGTGA	342	AGCACACATATAGCGC
094	TCTCGACTGCACATAT	245	ACAGTAGACTCTCAGA	350	TCTCTCTATCGCGCTC
098	TATACACACTCGCTCG	249	GTGACTCTATGCTATA	362	ACACGTGTGCTCTCTC
101	CTGTACTAGAGCGTCT	251	GATGAGTATAGACACA	369	ACATATACAGCGTATC
136	AGCACGTGTGTCGACA	260	GTGCACTCGCGCTCTC	381	CTCACACTCTCTCACA
139	GTCTCTCTCTCACGCA	265	CTGTGTATCTGTGTAC	409	TACTCACTGCGCTCAC
143	ACACACTCTATCAGAT	266	CGACGCACGATACTAT	410	ACACACACACTCTATA
153	TCTACTGCATGATGTC	273	ACTCTATGTCGATGTA	411	CTCTATATATCTCGTC
165	CTATCTAGCACTCACA	277	GTGTGTCTCGATGCGC	429	GATATACGCGAGAGAG
169	GATACTGACACACTAT	281	ACAGTACTAGTGCGAG	437	CGTGTCTCTCGATACA
178	ACACAGTAGAGCGAGC	289	CGTCTCTATCTCTCTA	443	TAGAGCGTGCATATAT
179	ACGACGCGCACTGACA	290	TACATGTGTCTATGTC	446	GATATATATGTGTGTA
180	CTCATAGCGTGTACTC	292	TATGTGTCTGCGCATA	449	CGCACACATAGATACA

During the optimisation of full length KIR gene PCR-based amplification, a range of alternative PCR chemistries were assessed for their ability to specifically amplify the target genes of interest. A list of these and their supplier is given in (Table 2.10). Further discussion may be found in Chapter 4, Section 4.02.

Table 2.10 Alternative PCR chemistries tested

PCR chemistry	Supplier	PCR chemistry	Supplier
Platinum SuperFi DNA Polymerase	Thermo Fisher Scientific	LA Taq DNA Polymerase	TaKaRa Bio Inc.
Platinum Taq DNA Polymerase high fidelity	Thermo Fisher Scientific	SpeedSTAR HS DNA Polymerase	TaKaRa Bio Inc.
AmpliTaq Gold DNA Polymerase	Thermo Fisher Scientific	Elongase Enzyme Mix	Fisher Scientific
Phusion high-fidelity DNA polymerase	Thermo Fisher Scientific	GoTaq Long	Promega
PCRBIO Ultra Polymerase	PCRBIO, London, UK	RANGER DNA Polymerase	Bioline
PCRBIO HiFi Polymerase	PCRBIO, London, UK	Taq98 for Hot Start PCR	Cambridge Biosciences, Cambridge, UK

2.06 Sanger sequencing

2.06.01 Sanger sequencing of KIR

To determine the correct KIR loci had been targeted by PCR amplification and to verify novel polymorphism identified during SMRT amplicon sequencing, Sanger sequencing of KIR exons and nearby nucleotides within the intervening introns was performed using protocols adapted from Hou *et al.* (2012) [181]. Briefly, full length amplification was performed as described in Section 2.05. Successful PCR products were purified using a commercially available column-based kit (GFX, Sigma-Aldrich). This involved adding 200 µL of GFX capture buffer to the PCR product, pipette-mixing and transferal to a fresh GFX column. The column was centrifuged at 13,500 rpm for 30 secs. To the column, 200 µL of GFX wash buffer was added and centrifugation repeated. Finally, the column was moved into a fresh 1.5 mL microcentrifuge tube, 50 µL of water added,

and centrifugation at 13,500 rpm performed for 60 secs. The product of this final elution was purified amplicon used as DNA template in downstream sequencing reactions.

The sequencing primers were identical to those published [181]. Chain-termination PCR was performed in 96-well PCR plates (BIOPlastics, Landgraaf, The Netherlands) using BigDye Terminator (BDT) v3.1 Cycle Sequencing kit (Applied Biosystems). Adapted reagent volumes and cycling conditions are shown in Table 2.11 and Table 2.12, respectively. Cycle-sequencing products were then ethanol-precipitated by the following protocol. To each product, 30 µL 100% ethanol and 2.5 µL 125 mM EDTA were added and the solution briefly vortexed to mix. The plate was incubated at room temperature for 15 mins before centrifugation at 3,200 rpm for 30 mins. Immediately after the spin finished, the supernatant was removed by inverted centrifugation on tissue at 900 rpm for 3 secs. The remaining pellet was washed in 30 µL of 70% ethanol and centrifuged at 3,200 rpm for 15 mins. Again, immediately following completion of the spin, supernatant was removed by inverted centrifugation at 900 rpm for 3 secs. Pellets were then allowed to air dry for 5 mins before rehydration in 10 µL Hi-Di Formamide solution (Applied Biosystems).

Table 2.11 Reagents for chain-termination cycle sequencing

Reagent (conc.)	Volume (µL)
Water	5.5
BDTv3.1 Sequencing Buffer (5X)	1.0
Primer (1.5 µM)	1.0
BDTv3.1 Ready Reaction Mix	1.0
Template DNA	1.5

Table 2.12 Thermal cycler programme for chain-termination cycle sequencing

Step	Temperature (°C)	Duration (min:sec)	Repetitions
1	96	1:00	1
2	96	0:10	25
	50	0:05	
	60	2:00	
3	4	∞	1

Sequencing of rehydrated cycle-sequencing products was performed using a programme designed to obtain maximum sequencing read length in a 96-capillary ABI 3730 DNA Analyzer (Applied Biosystems): run voltage and time were set to 8.5 V and 7,200 secs, respectively. Base-calling was performed using the default 3730BDTv3-KB-DeNovo_v5.2 parameters and electropherogram traces were analysed using FinchTV software (v1.5.0, Geospiza Inc., Seattle, WA, USA).

2.06.02 Microsatellite-homopolymer targeted sequencing

A large, variable-length microsatellite region (TA, approximately 50 copies) immediately followed by a variable-length extended homopolymer region (up to 30mer of T) characterises intron 5 of the KIR2DL2/2DL3 locus [98]. When attempting to sequence this challenging region of the gene using SMRT DNA sequencing technology, inconsistencies were observed between sequences of the same allele. As such, a targeted sequencing approach was designed to attempt to resolve the ambiguity and assist in establishing true polymorphism at the locus. Sequencing of a single region was sufficient to provide coverage of both repeat regions. Specific 5′ sequencing primers were designed to anneal 76 bp upstream of the start of the microsatellite region of KIR2DL2 and KIR2DL3, respectively. A 3′ sequencing primer with complementarity to both KIR2DL2 and KIR2DL3 annealed 137 bp downstream of the end of the homopolymer region.

In addition, both extended homopolymers (up to 24mer of T) and microsatellites (TA, 9 copies) are also encoded within intron 5 of KIR2DL1. However, these regions are separated by several hundred bases; prohibitive when attempting to sequence as a single region. As such, the two regions were sequenced independently using specific 5′ and 3′ sequencing primers (Table 2.13). The remainder of the Sanger sequencing protocol was identical to that described in Section 2.06.01, including chain-termination thermal cycling reagents and cycling conditions (Table 2.11 and Table 2.12).

Table 2.13 KIR2DL1/2/3 intron 5 sequencing primer specificities

Sequencing primer ID	Primer sequence (5′-3′)
KIR2DL2-F	ATTACAGGCATGAGCCACCACT
KIR2DL3-F	ATTACAGGCATGAGCCACAACG
KIR2DL2/3-R	GCAGCCACATGGATGGAACTG
KIR2DL1-F_Microsatellite	GCTGGTCTCGAACTCCTGAC
KIR2DL1-F_Homopolymer	TCTTGGCTACTGTGAACAGTGC
KIR2DL1-R_Microsatellite	TGCAGCCACATGGATGGAAC
KIR2DL1-R_Homopolymer	AGGTTGCAATGAGCCAAGG

2.07 Preparation of SMRT libraries

To prepare PCR amplicons for SMRT sequencing, they must be assembled into libraries. The optimised steps in this multistage process are outlined below. Unless otherwise stated, all reagents were supplied by Pacific Biosciences (Menlo Park, CA, USA). For details on the optimisation of library preparation, please see Section 4.02.10.

The initial step following PCR amplification is to qualify and quantify the amplicons using the Fragment Analyzer, as in Section 2.03.03. This step was referred to as quality control step 1 (QC1). Those PCR products containing amplicon of approximate length to the expected amplicon with concentration greater than 1.8 ng/µL and no contaminating smaller fragments were accepted to have amplified sufficiently.

To increase the efficiency of the SMRT sequencing process, PCR amplicons from multiple samples were combined into a single library (Figure 2.2). For the KIR2DL1, KIR2DL4, KIR2DL5/KIR3DP1, KIR2DS2, KIR2DS5, KIR3DL2 and KIR3DL3 amplicons, up to 20 different samples were combined into pools for each individual target. For the KIR2DL2/KIR2DL3 and KIR3DLS1 amplicons, the maximum multiplex level was reduced to 15 to improve read depth. Because the concentration of each amplicon is unique, the next step of library preparation is an equinanogram pooling step. Using the concentration of each amplicon as determined during QC1, the volume equating to an approximately equal number of amplicon molecules from each sample was calculated. The corresponding volume of each PCR product was then aliquoted into a single 1.5 mL Lo-Bind microcentrifuge tube (Eppendorf, Stevenage, UK).

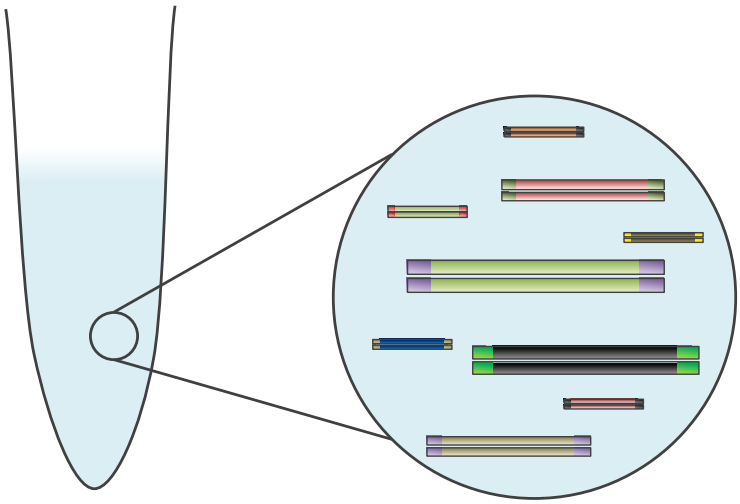


Figure 2.2 Cartoon representation of multiplexed, equinanogram amplicon pool

Within each equinanogram pool, amplicons from multiple different PCR reactions were pooled together. DNA barcodes (attached at both ends of the amplicon during the amplification PCR) allow the amplicons to be traceable throughout the sequencing stages.

Following equinanogram pooling, each pool underwent AMPure magnetic bead-based purification. Briefly, the volume of each pool was made up to the nearest 100 µL volume using Elution Buffer. A 0.4X volume of AMPure PB beads was added, gently mixed to homogeneity, and rotated at room temperature for 20 mins. Following rotation, the samples were briefly centrifuged to collect the liquid in the bottom of the tubes and placed in a magnetic rack (ThermoFisher Scientific). The magnetic beads, now bound to the amplicon, were collected on one side of each tube, and the supernatant removed. The beads were twice washed in 70% ethanol and allowed to air dry for 30 secs, before 38.5 µL of Elution Buffer was added. The samples were gently mixed to homogeneity before being rotated for 10 mins at room temperature. After the second rotation, samples were centrifuged briefly to collect the liquid in the bottom of the tubes and then placed in a magnetic rack. The magnetic beads (now unbound from purified amplicon) were collected on one side of each tube, and the supernatants removed and placed in fresh 1.5 mL Lo-Bind microcentrifuge tubes.

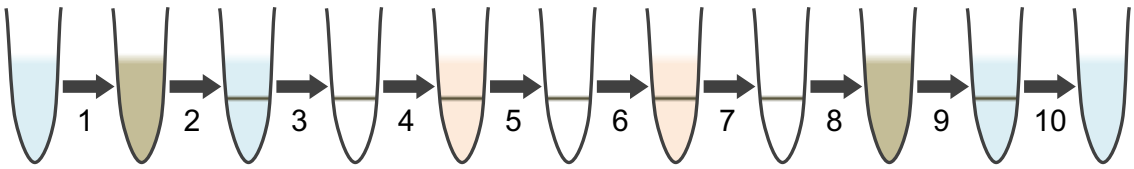


Figure 2.3 Cartoon representation of AMPure purification

To remove contaminates, such as short DNA fragments and library preparation enzymes, multi-step AMPure bead-based purification was used at several steps of the process. 1) AMPure purification beads (brown) were added to the sample (blue) at 0.4X concentration. The solution was mixed by rotation for 10 minutes. 2) The AMPure beads (now bound to desired product) were collected on the wall of the microcentrifuge tube. 3) The supernatant was removed. 4)-7) The AMPure beads were washed in 70% ethanol (pink). 8) Elution buffer was added to the AMPure beads and the solution was mixed by rotation. 9) The AMPure beads (now released from desired product) were collected on the walls of the microcentrifuge tube. 10) The supernatant (desired product) was removed and placed in fresh tube.

To ensure that AMPure purifications had been successful, 1.5 µL aliquots of each purified equimolar pool were quantified using NanoDrop photospectrometry (QC2). A minimum DNA concentration of 25 ng/µL was used to proceed to subsequent stages of the library preparation process, each of which were set up on ice.

Due to unavoidable DNA damage introduced during the PCR cycling, each equimolar pool underwent a DNA damage (Figure 2.4) and end repair step (Figure 2.5). This involved the addition of 5 µL DNA Damage Repair Buffer (10X), 0.5 µL NAD+ (100X), 5 µL ATP high (10 mM), 0.5 µL dNTP (10 mM) and 2 µL DNA Damage Repair Mix (25X) to each equimolar pool. The mixture was then gently mixed and incubated at 37°C for 90-150 mins, before cooling at 4°C for 1 minute. Subsequently, 2.5 µL of End Repair Mix (25X) was added to each tube, gently mixed, and incubated at 25°C for 5 mins. Following this incubation, to re-purify the equimolar pool, an AMPure bead-based purification was performed as described above, differing only in the final elution volume: 32.5 µL Elution Buffer. DNA quantification by NanoDrop photospectrometry was also repeated to ensure successful AMPure purification (QC3).

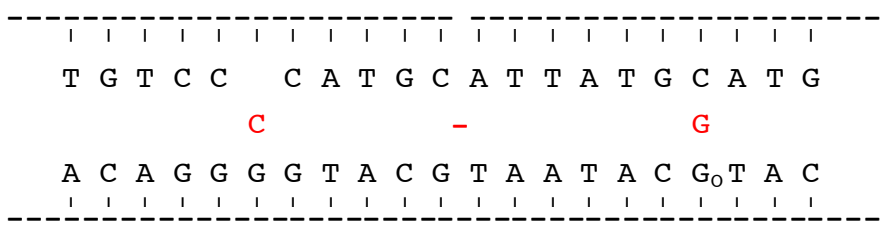


Figure 2.4 Diagrammatic representation of possible DNA damage repair

Several types of DNA damage were repaired during library preparation. From left to right: Abasic positions were replaced with nucleotides complimentary to the opposing strand. Next, nicks to the DNA backbone were repaired. Finally, oxidative damage could also be repaired. Repairs are shown in red.

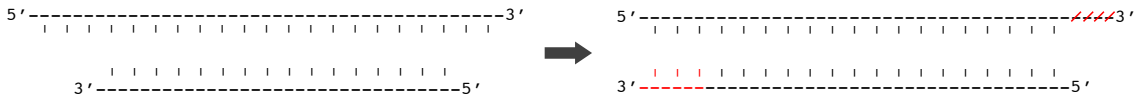


Figure 2.5 Diagrammatic representation of DNA end repair

To blunt the ends of repaired amplicons, DNA end repair was performed. Overhangs at the 5′ end were infilled, whilst 3′ overhangs were trimmed to match the length of the opposing strand.

To circularise the amplicons, thus allowing CCS, the repaired equimolar pools underwent a ligation reaction to combine each double-stranded amplicon molecule with two SMRTbell adapter molecules, one at each end (Figure 2.6). This was achieved by adding 2 µL of Blunt Adapter (25 µM) to the remaining repaired equimolar pool. This was then mixed before the addition of 2 µL adenosine triphosphate (ATP low, 1 mM), 4 µL Template Prep Buffer (10X) and 1 µL Ligase (30 U/µL). The solution was then mixed again and incubated overnight at 25°C.

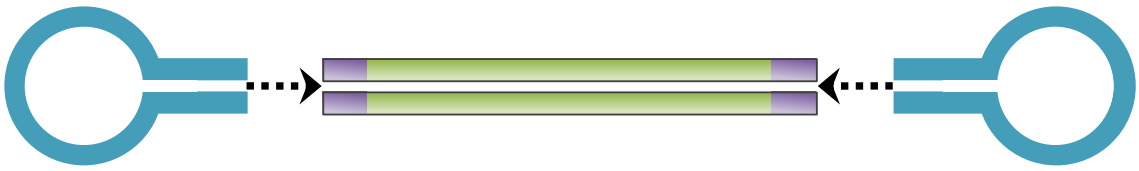


Figure 2.6 SMRTbell adapter molecule ligation

To circularise the repaired amplicons, hairpin SMRTbell adapters (blue) were ligated at both ends of the amplicon (green), preserving DNA barcode sequence (purple).

Following the overnight incubation, the ligase was removed from solution by AMPure purification, eluting into a final volume of 40 µL. The equimolar pools were then incubated at 37°C with 1 µL of Exonuclease (Exo) III and 1 µL of Exo VII for 60 mins to remove unligated amplicon and residual Blunt Adapter from the solution. One final AMPure selection, eluting into 12 µL of Elution Buffer, purified the final library. A 1 µL aliquot of each library was diluted in 4 µL of Elution Buffer and mixed. This

dilution was used as input DNA for Fragment Analyzer analysis and the final library fragment length and concentration determined for each library accordingly (QC4). Libraries were stored at 4°C until used in sequencing reactions.

2.08 Library-MagBead preparations and sequencing reactions

To anneal sequencing primer, sequencing polymerase and MagBeads to a library, a series of temperature-controlled incubations were undertaken, as outlined below. Due to the small volumes, and for the sake of accuracy, all reactions were set up on ice and any incubations at or above 20°C were performed in a Veriti 96-well thermal cycler. All reagents, unless otherwise stated, were supplied by Pacific Biosciences. All volumes of reagents required for the MagBead binding steps were calculated using Binding Calculator v2.3.1.1 (Pacific Biosciences', available at: https://github.com/PacificBiosciences/BindingCalculator).

A 1 µL aliquot of Sequencing Primer v2 was diluted in 32.3 µL of Elution Buffer within a 0.2 mL Lo-Bind microcentrifuge tube (Eppendorf, Stevenage, UK). This then underwent conditioning at 80°C for 2 mins before being cooled to 4°C. A lid temperature of 90°C was used for this reaction. A 10 µL aliquot of this dilution was then further diluted in 30 µL of Elution Buffer in a fresh 0.2 mL microcentrifuge tube. A 1 µL aliquot of this final dilution was added to a fresh 0.2 mL Lo-Bind microcentrifuge tube, alongside 0.9 µL of Primer Buffer v2 (10X). The size and concentration of library as measured in QC4 were used to calculate the required volume of library to add to the conditioned primer, using a primer:library molecular ratio of 5:1. The volume was then made up to 9 µL using water. The mixture was then incubated at 20°C for 30 mins to anneal the primer to the template using a lid temperature of 50°C.

Following primer annealing, polymerase molecules were bound. A polymerase:library molecular ratio of 10:1 was used. This was achieved by adding 2.85 µL of Binding Buffer v2, 1.5 µL of dNTP v2, 1.5 µL of dithiothreitol (DTT) and 0.15 µL of DNA Polymerase P6 to each primer-annealed library, followed by incubation at 30°C for 30 mins. Lid temperature for this reaction was also 50°C. Following polymerase annealing, a 0.63 µL aliquot of the library-primer-polymerase complex was diluted in 8.37 µL of MagBead Binding Buffer v2, resulting in a custom "concentration on plate" of 35 pM.

MagBeads required washing prior to use. This preparation involved adding a 35 µL aliquot of MagBeads to a clean 1.5 mL microcentrifuge tube and placing in a magnetic rack to collect the MagBeads on one side of the tube. The supernatant was removed and discarded, and MagBeads washed by gently mixing in 35 µL of MagBead Wash Buffer v2. The solution was centrifuged briefly to collect all liquid in the bottom of the tube, before being placed back on the magnetic rack. Again, the supernatant was discarded, and 35 µL of MagBead Binding Buffer v2 was added to beads. This was gently mixed, briefly centrifuged, placed on a magnetic rack and supernatant discarded. Following MagBead preparation, the 9 µL of diluted library complex was added to the MagBeads and mixed by slow rotation at 4°C for 60 mins. Following this cold incubation, the MagBead-library complex was briefly centrifuged and placed on a magnetic rack. The supernatant was discarded, and the bound library washed first in 18 µL of MagBead Binding Buffer v2, followed by 18 µL of MagBead Wash Buffer v2 before final addition of 45 µL of MagBead Binding Buffer v2. Each wash was interspersed with gentle mixing, brief centrifugation and removal of supernatant after collection of

MagBeads on a magnetic rack. The final volume of 45 µL of bound library complex was added to an individual well of a 96-well PCR plate (BIO-RAD).

The process above was performed for several libraries concomitantly, allowing up to 16 libraries to be sequenced within a single run. Each run was programmed using the RS Remote software. Sequencing steps were controlled by the RS II platform and utilised DNA Sequencing Kit 4.0 v2 (DNA Sequencing Reagent Plate 4.0 v2 and Oxygen Scavenging Enzyme), SMRT Cell Oil and SMRT Cell v3. OS Enzyme and SMRT Cell Oil tubes were recapped using a tube septum. Plate septa were used to seal the sample plate and DNA Sequencing Reagent Plate 4.0 v2. All tube and plate septa were pierced prior to use to ensure accurate robotic pipetting. Each library was sequenced on one SMRT Cell v3 for 6 hours using the one-cell-per-well (OCPW) and stage-start settings.

2.09 Bioinformatic sequence analysis

After each SMRT Cell completed sequencing and data transfer, primary analysis was performed using SMRT Analysis (version 5.0.1.9578) in collaboration with Anthony Nolan Research Institute's Bioinformatics department [288]. Briefly, SMRTbell adapter sequence was cleaved and removed from the remaining sequence, separating the polymerase reads into subreads. Each subread was then assessed for barcode sequence and grouped accordingly. Several filters were applied to these subreads in order to remove poor quality or incomplete sequences. Firstly, the minimum length (minLength) parameter for each gene was set to remove incomplete subreads (Table 2.14). In addition, a minimum signal:noise ratio (minSNR) was set at 3.75. Other filters were used at their default parameters. Each barcode group underwent iterative clustering to

define individual allele sequences within a barcode group, resulting in an unbiased, *de novo* allele sequence assembly.

Table 2.14 minLength parameters

KIR locus	minLength (kb)
KIR2DL1	10.0
KIR2DL2/3	12.0
KIR2DL4	8.6
KIR2DL5/ KIR3DP1	7.5
KIR2DS2	13.0
KIR2DS5	12.3
KIR3DL1/ KIR3DS1	13.0
KIR3DL2	13.4
KIR3DL3	9.6

Following clustering into individual alleles, AlleleTeapot (AT) software (version 2.0.0-23-g01794ad, Anthony Nolan, London, UK) was used to assign allele names to each consensus sequence, as described below. The consensus sequences aligned against a reference library of all known KIR alleles using BLAST-n v2.2.31 [289]. This identified the targeted gene and most homologous known allele. The consensus sequences were then trimmed to match the start and end co-ordinates of the reference allele assigned by the IPD-KIR Database [58,98]. Should an exact match to a known allele exist, this allele was reported with no discrepancies. When discrepancies were noted against the most homologous allele, co-ordinates and a description of the discrepancy were noted. Discrepancies were described against both the CDS and the full-length, genomic sequence of alleles where available. Values for the quality value (QV, a Phred-like quality score generated by Pacific Biosciences' SMRTAnalysis software reflecting the total probability that a base call is an insertion or substitution or is preceded by a

deletion), predicted sequence accuracy, sequence length and number of reads used in each allele analysis were also included in the output file.

2.10 Size selection

2.10.01 BluePippin size selection

Protocols utilising SMRT sequencing for genome assembly often employ a specialised size selection step to purify libraries by removal of shorter, contaminating fragments that would otherwise inhibit the sequencing reaction. This is often performed using BluePippin (Labtech, Heathfield, UK), an electrophoresis-based size selection platform. To assess whether this could be used to remove internal control fragments (and other small contaminating fragments), the BluePippin platform was trialled as per manufacturer's instructions. Unless otherwise stated, all reagents were supplied as part of the BLF7503 kit (Labtech).

Briefly, samples were purified by AMPure purification before final elution into 30 µL of elution buffer. To this, 10 µL of loading buffer was added and mixed by pipetting. The gel cassette was prepared by replenishing the electrophoresis buffers in both the sample and elution wells. Following this, 40 µL of buffer within the sample well was removed, allowing the addition of the 40 µL of sample/loading buffer. Run setup specified selection of fragments with length between 10 and 50 kbp. Following electrophoresis, samples were left for 30 mins prior to removal of the solution in the elution chamber. The eluate underwent AMPure purification before quantification and qualification on the Fragment Analyzer platform.

2.10.02 Genomic DNA Clean and Concentrator columns

An alternative method to remove short contaminating DNA fragments from SMRT DNA sequencing libraries utilises a spin column-based purification called Genomic DNA Clean and Concentrator (Zymo Research, Irvine, CA, USA). Unless otherwise stated, all reagents were supplied as part of the Genomic DNA Clean & Concentrator™-10 kit (Zymo Research). DNA purification was performed in accordance with manufacturer's instructions. Briefly, ChIP DNA Binding Buffer was mixed with the DNA sample (eluate from AMPure following amplicon pooling) in a 1.5 mL Lo-bind microcentrifuge tube (Eppendorf) at a 5:1 ratio. This solution was then transferred to a Zymo-Spin™ IC-XL Column, which was subsequently placed inside a collection tube. The tubes underwent centrifugation at 13,500 rpm for 30 secs. Flow-through was discarded and 200 µL DNA Wash Buffer was added to the column. The tubes underwent centrifugation at 13,500 rpm for 60 secs. The wash step was repeated. Following the second wash, the column was transferred to a clean 1.5 mL Lo-bind microcentrifuge tube and 37 µL of DNA Elution Buffer was added to the column. The column was incubated at room temperature for 60 secs prior to centrifugation at 13,500 rpm for 30 secs. The column was discarded and the eluted DNA was stored at 4°C.

2.10.03 Agarose gel size selection purification

Size selection was also performed using agarose gel electrophoresis to separate the desired amplicon from smaller contaminating fragments. Initial agarose gel electrophoresis was performed as described above (see Section 2.03.01). The desired bands were cut from the gel and DNA was purified using Zymoclean Gel DNA Recovery Kit (Zymo Research) as per the manufacturer's instructions. Unless otherwise stated, all reagents were supplied as part of this kit. Briefly, the gel cut-out was added to a clean 1.5 mL Lo-Bind microcentrifuge tube (Eppendorf) and mixed with 500 µL of

Agarose Dissolving Buffer. This was incubated at 45°C for 5-10 mins (until the agarose was dissolved). The solution was then transferred to a Zymo-spin column and the column placed inside a clean 1.5 mL Lo-Bind microcentrifuge tube. This underwent centrifugation at 13,500 rpm for 30 secs before the supernatant was discarded and 200 µL of DNA Wash Buffer was added to the column. This underwent centrifugation at 13,500 rpm for 30 secs, followed by a repeat of this wash step. To elute the purified DNA, 12 µL of water was added to the column, and centrifugation applied at 13,500 rpm for 60 secs. Eluted DNA was stored at 4°C.

2.11 Data storage and database design using FileMaker Pro

All steps of sample and library preparation were recorded and stored to assist in data consistency and analysis. As such a relational database was designed and built in FileMaker Pro (version 16.0.5.500, Santa Clara, CA, USA) to house this data. The main tables were as follows: Transplants, Samples, Traxis Tubes, Amplifications, Libraries and TGS Results. In addition, join tables (Traxis Locations, Pooling and TGS Runs) were also used to allow many:many relationships to be created. A range of short scripts and automated calculations were written to perform a variety of tasks. These included validation tests, assigning and recording DNA tube locations, selecting samples that required amplification, determination of optimal barcodes, assigning barcoded amplicons to libraries, monitoring residual amplicon and library volumes, coding sub-groups for statistical analysis and a variety of specific import, export and search functions. In addition, the ability for an operator to add commentary to the allele sequencing data generated was also designed into this database to further centralise data storage. This factor allowed rapid identification of samples requiring re-analysis and incorporation into future library preparation reactions.

2.12 Statistical analysis

All statistical analysis was performed in SPSS Statistics v25 (SPSS Inc., Chicago, IL, USA) and/or R v3.4.3 (http://www.r-project.org). Differences in overall survival (OS) and DFS were viewed by the Kaplan-Meier method [290] and compared using the log-rank test, whilst outcomes with competing risk (non-relapse mortality [NRM] and relapse) were assessed by cumulative incidence methodology [291,292]. The competing event for NRM outcome was relapse, whilst death without relapse was the competing event during relapse analysis. The incidence of grades 2-4 aGVHD was assessed by chi-squared test where the smallest group was composed of five or more individuals [293], or by Fisher's Exact test should the smallest grouping have been composed of fewer than five cases.

Multivariate analysis tests also differed accordingly to outcome assessed: OS and DFS were evaluated using Cox regression analysis [294], relapse and NRM by Fine and Gray methodology [295] and aGVHD by binary logistic regression. Forward stepwise selection of covariables utilised a $p \leq 0.05$ inclusion threshold to calculate significant contributors within multivariate models. Covariates were calculated for each sub-cohort requiring multivariate analysis. In both univariate and multivariate analysis, statistical significance was determined at $p \leq 0.05$, whilst a trend was denoted at $0.05 < p \leq 0.1$.

Chapter 3 The impact of donor and recipient KIR genotypes on the outcomes of HCT

3.01 Introduction

There has been substantial investigation into the impact of KIR on the outcomes of HCT around the world. However, many conflicting results have been obtained and, to complicate this further, many different models of KIR influence have been proposed. For example, research has often focussed solely on the mismatching of KIR ligands whilst relying on common KIR genotype information to assume the presence/absence status of the KIR genes whose ligands are under assessment [259,296]. Other models only investigate the KIR genotype of the donor [83,272,273], whilst yet another group of studies focus on matching the KIR genotypes of donors and recipients [282-284]. In addition, at the time of writing, nothing has been published on the effects of KIR in a UK cohort of AML recipients receiving HLA-matched VUD HCT. Moreover, common HCT protocols to treat AML in the UK differ from much of the rest of the world in its extensive utilisation of TCD with alemtuzumab as a GVHD prophylaxis [297]. As such, I first thought it prudent to apply several of the different existing models of KIR impact analysis to a UK AML cohort using lowest resolution KIR genotyping data obtained by PCR-SSP. This allowed the identification of the importance of KIR presence/absence genotyping in this unique UK population, as well as helping to direct clinical analyses based on KIR allele-level resolution typing (discussed in Chapter 6).

3.02 Models of data analysis

Due to the variety of previous methods used to analyse KIR data in HCT, I have opted to assess only those most relevant. Because the majority of the study cohort is well

HLA-matched, only 15 cases of the donor "missing self" scenario were observed. As such, it was inappropriate to assess the KIR ligand-ligand model proposed by Ruggeri *et al.* (2002) [259] for haploidentical transplantation. The three methods I have chosen are discussed fully in Chapter 1 and briefly outlined below:

1. The KIR receptor-ligand model: assesses the combination of KIR genes encoded by the donor with ligands encoded by the recipient. KIR-ligand combinations assessed include: KIR2DL1–HLA-C2, KIR2DL2–HLA-C1, KIR2DL3–HLA-C1, KIR3DL1–HLA-Bw4 and KIR2DS1–HLA-C2. In addition to this, the KIR3DL1–HLA-Bw4 combination was further discriminated according to HLA-Bw4 subtype at position 80: Isoleucine (80I) *vs* Threonine (80T).
2. The KIR receptor-receptor model: assesses the combination of KIR genes encoded by both patient and donor as an additional evaluation of genetic parity. Matching was assessed at both individual gene and gene motif level.
3. The donor KIR B motif model: assesses the impact of the donor KIR B haplotype content on the outcomes of HCT. Analysis assesses both centromeric and telomeric motifs as well as broad KIR haplotype. In addition, the influence of patient-encoded ligands was also assessed.

3.03 Models of clinical cohort

Several different subgroups of the cohort were analysed for two main reasons: to determine the baseline effect of KIR in my cohort, allowing the demonstration of possible benefits resulting from KIR allele typing (Chapter 6); and to reduce heterogeneity within the overall cohort. The findings from each cohort are discussed in turn in the following sections. The sub-cohorts are shown below:

1. Entire cohort: all 405 donor-recipient pairs.
2. HLA-C1 negative recipients.
3. HLA-C1 positive recipients who underwent HLA-C mismatched HCT.
4. Adult, HLA-matched, myeloablative cohort: adult patients (>18 years) who underwent MAC prior to HLA-matched transplantation.
5. Adult, HLA-matched, reduced intensity cohort: adult patients (>18 years) who underwent RIC prior to HLA-matched transplantation.

3.04 Results: transplantation summary

3.04.01 Transplantation characteristics

The donor, recipient and other transplant characteristics are summarised in Table 3.1. From this, it is apparent that the average donor age is quite high, with mean and median age of 35 years. As is standard practice in the UK, most recipients received transplants from male donors, and CMV matching also appears to have been a donor selection preference. Most donors were well HLA-matched to their recipients, with only 6% of transplants utilising a donor with two or more HLA mismatches. Although a relatively even divide of the cohort is found when investigating the HSC source (BM *vs* PBSC), this is highly era-dependent (Chi-squared test, $p<0.001$). As time progressed, the HSC source of preference shifted from BM (64% of transplants in 1997-2003) to PBSC in later eras (70% in 2004-2011). It is also apparent that, due to the retrospective nature of this study, some data has been difficult to obtain. For example, the TCD status of 11% of the transplants is unknown although, based on the majority of transplants with known TCD status, we may assume that many of the unknown cases also utilised alemtuzumab as a TCD agent. However, the remaining transplant conditions are well characterised.

Table 3.1 Transplant characteristics

Factor	Frequency (%)
Donor age, years	
<30	128 (32)
>30	274 (68)
Average (mean, median)	35, 35
Missing	3 (1)
Recipient age	
<40	196 (48)
>40	209 (52)
Average (mean, median)	40, 41
Donor gender	
Male	325 (80)
Female	80 (20)
TCD by alemtuzumab	
Yes	343 (85)
No	17 (4)
Missing	45 (11)
Donor-Recipient gender matching	
Female-Male	38 (9)
Male-Male	194 (48)
Female-Female	42 (10)
Male-Female	131 (32)
Conditioning regimen	
MAC	217 (54)
RIC	183 (45)
Missing	5 (1)
Previous autograft transplant	
0 previous autografts	381 (94)
1 or more previous autografts	24 (6)

Factor	Frequency (%)
EMBT disease risk score	
Good	179 (44)
Intermediate	174 (43)
Poor	47 (12)
Missing	5 (1)
Donor-Recipient CMV matching	
Matched	275 (68)
Mismatched	118 (29)
Missing	12 (3)
Recipient gender	
Male	232 (57)
Female	173 (43)
HSC source	
BM	175 (43)
PBSC	228 (56)
Missing	2 (1)
Transplant era	
1996-1999	35 (9)
2000-2003	127 (31)
2004-2007	131 (32)
2008-2011	112 (28)
HLA matching	
10/10 HLA-matched	287 (71)
9/10 HLA-matched	94 (23)
<9/10 HLA-matched	24 (6)

3.04.02 Outcomes summary

3.04.02.01 Endpoint analysis

A brief summary of timepoint outcomes data from all the patients in the cohort is given in Figure 3.1 and Table 3.2. From these, it is evident that many patients died in the year immediately following transplant. Relapse was marginally more probable than NRM. Although a large proportion of events occur in the 12 months after transplantation, the cumulative incidence of relapse continues to increase up to and beyond two years post-transplant. Final estimates of relapse and NRM incidence after 12 years are 42% and 29%, respectively.

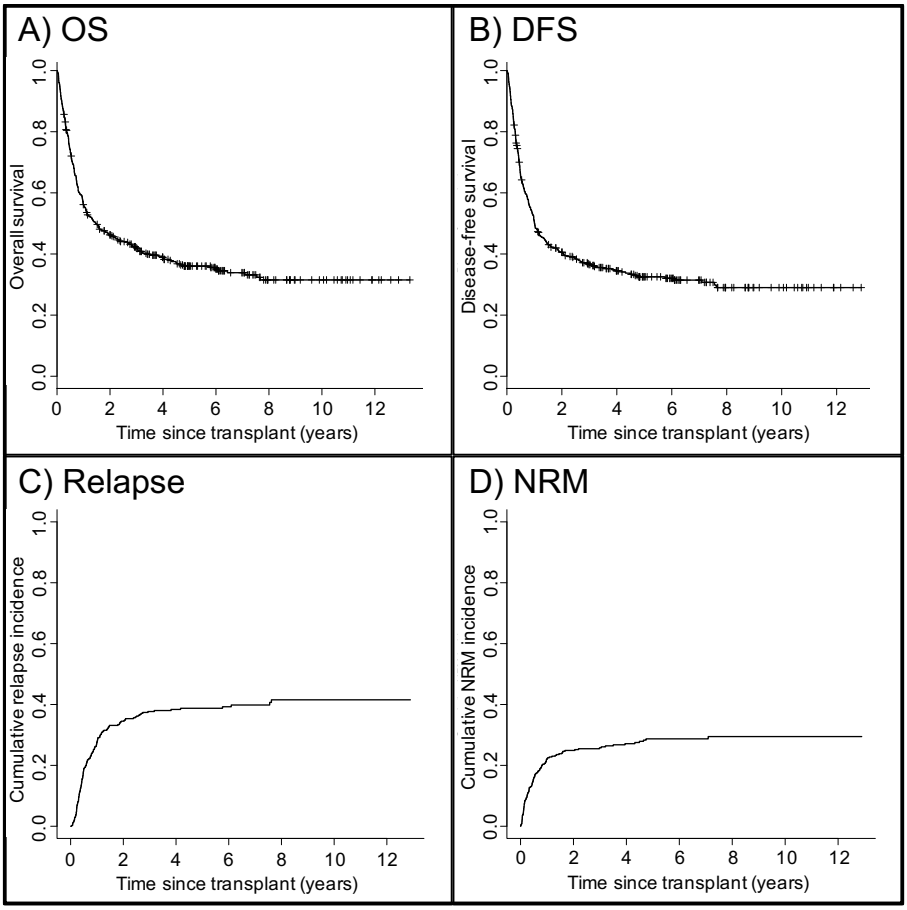


Figure 3.1 HCT outcomes summary from the entire cohort

Time-point analysis of OS (A), DFS (B), relapse (C) and NRM (D) for every recipient of the study cohort until final follow-up. These graphs demonstrate that both relapse and NRM incidence accumulate rapidly during the initial 18 months post-transplant. During this period, OS and DFS decrease accordingly. Check marks in A and B correspond to censorship without event.

Table 3.2 HCT outcomes summary

Event	Frequency (%)	Median survival (years)
Overall survival	31.5	1.46
Non-relapse mortality[§]	29.5	N/A
Relapse[§]	41.5	N/A
Disease-free survival[§]	29.0	1.00

[§] For 12 transplants, it was not possible to obtain relapse/NRM information – valid % presented.

Due to the vast majority of relapse and mortality occurring within the first five years post-transplant, further analyses of OS, DFS and relapse were limited to this time

period. In addition, due to the multifaceted nature of NRM and the lack of information relating to cGVHD, all further NRM analysis was limited to only the 12 months immediately following transplantation. Estimated probability of OS, DFS, relapse and NRM at their respective endpoints were 40%, 36%, 39% and 22%, respectively.

Although data relating to cGVHD was unavailable, information regarding the incidence and severity of aGVHD was available for the majority of patients. Of the 405 transplant recipients in the entire cohort, aGVHD data was available for 372 (92%). Of these recipients, 173 (47%) did not develop aGVHD, 113 (30%) developed grade 1 aGVHD and 86 (23%) developed more severe grade 2-4 aGVHD (Table 3.3). Due to the low frequency of the most severe grades of aGVHD, incidence of grade 2-4 aGVHD was grouped together. Incidence of aGVHD incidence was assessed by Chi-squared test (or Fisher's Exact test for small subgroups).

Table 3.3 aGVHD incidence

Grade	Frequency (valid %)
0	173 (47)
1	113 (30)
2	62 (17)
3	14 (4)
4	10 (3)
Missing	33 (0)

The impact of a range of transplant characteristics on the outcomes of HCT are summarised in Supplementary Table B and discussed further below.

3.04.02.02 Clinical factors affecting five year overall survival probability

The estimated five year overall survival probability in this AML cohort was approximately 40%. However, there were several clinical factors that significantly improved this statistic. Donors who were younger (<30 years, 48% *vs* 36% p=0.042), CMV-matched (44% *vs* 29%, p=0.002) or well HLA-matched (10/10 *vs* 9/10 *vs* <9/10, 43% *vs* 36% *vs* 8%, p<0.001) were associated with significantly improved OS at five years. In addition, patients with improved disease risk score (good *vs* intermediate *vs* poor, 43% *vs* 40% *vs* 21%, p=0.009) and no prior autograft history (41% *vs* 17%, p=0.01) also experienced significantly better five year OS probability.

3.04.02.03 Clinical factors affecting five year relapse probability

After accounting for the competing risk of non-relapse mortality, the estimated probability of relapse at five years was 39%. Similarly to the findings in OS analysis, both donor and recipient characteristics influenced the susceptibility of patients to the incidence of relapse. However, when investigating relapse, the only clinical factors demonstrated to have a beneficial effect were younger donor age (31% *vs* 42%, p=0.034) and patient's disease risk (Good *vs* Intermediate *vs* Poor; 38% *vs* 35% *vs* 59%; p=0.003). Interestingly, recipients with "good" disease risk were fractionally more likely to relapse than recipients with "intermediate" disease risk although, when considered independently of the poor disease risk group, this was not statistically significant (p>0.2). The highest five year relapse incidence, however, was still associated with the "poor" disease risk group.

3.04.02.04 Clinical factors affecting one year non-relapse mortality probability

The risk of NRM at one year was also significantly affected by several factors associated with the transplant itself. Unsurprisingly, the era of transplantation was associated with significant differences in the estimates of NRM incidence. Those transplants occurring between 1996 and 1999 experienced approximately doubled NRM rates at one year post-transplant when compared to any of the other time periods (1996-1999 *vs* 2000-2003 *vs* 2004-2007 *vs* 2008-2011, 41% *vs* 19% *vs* 21% *vs* 21%, p=0.02). Additionally, donor-recipient CMV and HLA-matching was also correlated with improved one year NRM estimates (CMV-matched *vs* CMV-mismatched, 19% *vs* 29%, p=0.034; 10/10 *vs* 9/10 *vs* <9/10 HLA-matched, 19% *vs* 23% *vs* 54%, p<0.001).

3.04.02.05 Clinical factors affecting five year disease-free survival

The factors affecting the estimated probability of five year DFS closely followed those which influenced OS. Donors who were younger (<30 years, 46% *vs* 31%, p=0.017), CMV-matched (40% *vs* 27%, p=0.008) or well HLA-matched (10/10 *vs* 9/10 *vs* <9/10, 39% *vs* 32% *vs* 8%, p<0.001) were associated with significantly improved DFS at five years. The autograft history of the recipient also remained significant (no previous autograft, 37% *vs* 13%, p=0.008), although the influence of disease risk score lost significance but remained a trend (p=0.052)

3.04.02.06 Clinical factors affecting aGVHD incidence

Interestingly, the factors associating with aGVHD incidence did not align with those relating to increased risk of NRM. Instead, use of PBSC cell source rather than BM was strongly associated with higher frequency of more severe grades of aGVHD (grades 2-4, 31% *vs* 19%; p=0.007). Unsurprisingly, the use of alemtuzumab as a TCD agent

drastically reduced severe aGVHD incidence (24% *vs* 65%; p<0.001), although small size within the non-alemtuzumab subgroup (n=17) limits the strength of this analysis. Donor age also correlated with incidence of severe aGVHD: use of younger donors was associated with significantly less grades 2-4 aGVHD (18% *vs* 29%, p=0.017). Finally, although only a statistical trend, HLA matching also correlated with reduced aGVHD severity (grade 2-4: 10/10 *vs* 9/10 *vs* <9/10, 22% *vs* 33% *vs* 35%, p=0.068)

3.04.02.07 Conditioning regimen alone does not influence outcomes

The total cohort was divided for some of the following analyses based on the conditioning regimen utilised during transplant preparation. To assess whether the conditioning regimen itself could be responsible for any differences between these sub-cohorts, each of the above outcomes were also stratified by conditioning regimen. This analysis concluded that the outcomes from MAC and RIC cohorts were not significantly different from one another.

3.05 Results: KIR genotyping summary

3.05.01 Success rate

Due to limited DNA availability for some transplant recipients, in combination with the poor quality of some older DNA samples, complete KIR genotyping was not possible in 100% of individuals. As such, of the 405 total transplant pairs included in the study, it was not possible to complete KIR genotyping for one donor and 19 recipient samples. This results in an overall success rate of 97.5% (790/810). As observed in previous studies [90], and as indicated during the initial testing period, it was apparent that the framework KIR genes (KIR2DL4, KIR3DL2, KIR3DL3 and KIR3DP1) were present in

100% of samples. As such, further testing of these four loci was terminated and not considered when assessing the completeness of KIR genotyping.

3.05.02 KIR genotype frequency comparisons

To compare the KIR genotyping results from my cohort to those previously published in a similar population, I utilised the data available from the Allele Frequency Net Database [www.allelefrequencies.net] to determine observed gene frequencies from a combined European caucasoid population [31]. The comparison between my cohort and the previously published datasets, including Chi-squared test p-values, is shown in Table 3.4, demonstrating concordance with published data, and thus confidence in the KIR genotyping results.

Table 3.4 KIR gene frequency comparison

KIR factor	Observed frequency (%)	Published frequency (%)	*p-value*
KIR2DL1 presence	97.7	95.8	0.45
KIR2DL2 presence	51.3	52.1	0.91
KIR2DL3 presence	91.6	89.6	0.63
KIR2DL5 presence	50.4	51.4	0.89
KIR2DS1 presence	38.3	39.9	0.82
KIR2DS2 presence	52.0	52.0	1.00
KIR2DS3 presence	27.4	30.2	0.66
KIR2DS4 presence	95.8	93.6	0.49
KIR2DS5 presence	30.9	30.4	0.94
KIR2DP1 presence	97.6	96.8	0.73
KIR3DL1 presence	96.1	94.0	0.49
KIR3DS1 presence	39.6	40.0	0.95

3.06 Results: cohort validations

3.06.01 Comparison of genotype frequencies between donors and recipients

To confirm that the frequency of individual KIR genes, motifs or genotypes was not significantly different between the donor and recipient groups, comparisons were made on each sample with KIR genotype information available (Table 3.5).

The results of this analysis reveal that there are no significant differences in the presence of KIR factors between the patient and donor cohorts, implying that any KIR-mediated impacts on HCT outcomes are unlikely to be related to this. This also helps to indicate that there is no apparent influence of KIR on susceptibility to AML oncogenesis in this cohort.

Table 3.5 A comparison of KIR factors between donor and recipient populations

KIR factor	Recipient (%)	Donor (%)	*p-value*
Individual KIR genes			
KIR2DL1 presence	377 (98)	395 (98)	0.92
KIR2DL2 presence	193 (50)	211 (52)	0.53
KIR2DL3 presence	357 (92)	367 (91)	0.40
KIR2DL5 presence	188 (49)	208 (51)	0.43
KIR2DS1 presence	139 (36)	162 (40)	0.24
KIR2DS2 presence	194 (50)	215 (53)	0.41
KIR2DS3 presence	98 (25)	116 (29)	0.29
KIR2DS4 presence	371 (96)	386 (96)	0.69
KIR2DS5 presence	119 (31)	124 (31)	0.97
KIR2DP1 presence	376 (97)	395 (98)	0.74
KIR3DL1 presence	373 (97)	386 (96)	0.43
KIR3DS1 presence	145 (38)	166 (41)	0.31
KIR genotype			
KIR AA	126 (33)	119 (29)	0.33
KIR BX	260 (67)	285 (71)	
Cen KIR motif			
Cen-AA	190 (49)	189 (47)	0.69
Cen-AB	166 (43)	178 (44)	
Cen-BB	30 (8)	37 (9)	
Tel KIR motif			
Tel-AA	235 (61)	230 (57)	0.53
Tel-AB	136 (35)	156 (39)	
Tel-BB	15 (4)	18 (4)	
HLA class I KIR-ligands			
C1 presence	361 (89)	359 (89)	0.82
C2 presence	228 (56)	229 (57)	0.94
Bw4 presence	263 (65)	264 (65)	0.94

3.06.02 Impact of recipient KIR factors on the outcomes of HCT

Having found no difference between the KIR genotypes of patients and donor groups, I thought it also prudent to assess whether the patient KIR genotype influenced the outcomes of transplantation. Although previous studies have demonstrated no influence, clarification in this separate cohort was desirable. The results from this are summarised in Table 3.6.

Table 3.6 Summary of univariate analysis p-values of the impact of recipient KIR factors on the outcomes of HCT

Recipient KIR factor	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
Individual KIR gene presence					
KIR2DL1	0.49	0.59	***0.010***	0.11	0.46
KIR2DL2	0.44	0.52	0.44	0.20	0.83
KIR2DL3	0.71	0.93	0.11	0.11	0.25
KIR2DL5	0.68	0.92	0.82	0.78	0.92
KIR2DS1	0.53	0.58	0.96	0.45	0.42
KIR2DS2	0.41	0.48	0.35	0.13	0.68
KIR2DS3	0.86	0.66	0.50	0.95	0.37
KIR2DS4	0.26	***0.042***	0.51	0.12	0.37
KIR2DS5	0.47	0.48	0.94	0.37	0.85
KIR2DP1	0.84	0.97	***0.034***	0.089	1.00
KIR3DL1	0.18	0.094	0.55	0.19	0.53
KIR3DS1	0.59	0.82	0.85	0.96	0.29
KIR genotype					
KIR AA *vs* KIR BX	0.17	0.25	0.75	0.34	0.49
Cen KIR motifs					
Cen-AA *vs* Cen-BX	0.40	0.49	0.31	0.11	0.61
Tel KIR motifs					
Tel-AA *vs* Tel-BX	0.62	0.69	0.67	0.98	0.24
HLA class I KIR ligands					
HLA-Bw4 presence	0.36	0.39	0.41	0.14	0.13
HLA-C1 presence	1.00	0.68	0.48	0.58	0.94
HLA-C2 presence	0.56	0.60	0.92	0.63	0.79

Statistically significant results are denoted by ***bold italics***.

In accordance with previous findings, patient KIR genotype and ligand type demonstrated very little correlation with HCT outcomes. However, several factors did appear to correlate with particular outcomes. Presence of patient-encoded KIR2DL1 and KIR2DP1 appeared to decrease the risk of relapse at five years (p=0.01 and p=0.034, respectively), although relatively low frequency of KIR2DL1 and KIR2DP1 absence (n=9 and n=10, respectively) may contribute to generation of false positive results. The apparent decrease in five year DFS probability resulting from absence of patient-encoded KIR2DS4 (n=15, p=0.042) may also result from this problem. As such, their validity is unreliable and this cohort may be assumed to confirm previous large cohort analysis that indicated that recipient KIR genotype does not significantly affect the outcomes of HCT to treat AML.

3.07 Results: Donor KIR-mediated influences on the entire AML cohort

Previous studies have, however, implicated particular donor-encoded KIR genotypes with both relapse and DFS outcomes. As such, it was appropriate to replicate these investigations in my study cohort.

3.07.01 The KIR receptor-ligand model

Several different missing ligand scenarios exist, accounting for the multitude of KIR-HLA combinations. Univariate analysis to assess the influence of each of these different scenarios is summarised in Table 3.7, and discussed in turn below.

Table 3.7 Univariate analysis p-values of HCT outcomes based on specific donor KIR, recipient ligand models

Donor KIR, Recipient Ligand	5 year OS (n=405)	5 year DFS (n=393)	5 year relapse (n=393)	1 year NRM (n=393)	aGVHD (n=384)
KIR2DL1^{+ve}, HLA-C2^{-ve}	0.52	0.56	0.83	0.83	0.84
KIR2DL2/3^{+ve}, HLA-C1^{-ve}	1.00	0.68	0.48	0.58	0.94
KIR3DL1^{+ve}, HLA-Bw4^{-ve}	0.59	0.67	0.31	0.20	***0.042***
KIR3DL1^{+ve}, HLA-Bw4-80I strategy 1^{-ve}	0.52	0.73	0.37	0.15	0.17
KIR3DL1^{+ve}, HLA-Bw4-80I strategy 2^{-ve}	0.54	0.65	0.73	0.83	0.092
Missing ligand frequency	0.42	0.62	0.59	0.26	0.13
KIR2DS1^{+ve}, HLA-C2^{-ve}	0.65	0.45	0.74	0.20	0.91

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table C.

3.07.01.01 Donor KIR2DL1 and recipient HLA-C2

When in the presence of its ligand, HLA-C2, KIR2DL1 is known to have a strongly inhibitory effect on the NK cell upon which it is expressed. As such, I hypothesised that the combination of donor-encoded KIR2DL1 and recipient-encoded HLA-C2 may correspond with an increased incidence of disease relapse, owing to the lack of alloreactivity against missing self. Following the model proposed by Leung *et al.* (2004) [266], it was also predicted that the NK cell response in the absence of any KIR2DL1

would also be inhibitory, regardless of ligand status. As such, the entire cohort was stratified by these factors, resulting in the formation of the two following subgroups:

1. Proposed inhibitory: Donor KIR2DL1^{+ve}, Recipient HLA-C2^{+ve} or Donor KIR2DL1^{-ve}.
2. Proposed reactive: Donor KIR2DL1^{+ve}, Recipient HLA-C2^{-ve}.

The univariate probability values from this analysis are given in Table 3.7. From this analysis, it is evident that the transplants which involved a proposed "inhibitory" combination did not result in significantly different HCT outcomes when compared to the proposed "activating" stratum, leading us to accept the null hypothesis that there is no significant difference between these two groups.

3.07.01.02 Donor KIR2DL2/3 and recipient HLA-C1

A similar analysis assessing the combination of donor-encoded KIR2DL2/3 and recipient-encoded HLA-C1 ligand was also performed. Due to the mutual exclusivity of KIR2DL2 and KIR2DL3 on a single haplotype, absence of KIR2DL2 confirms the presence of KIR2DL3, whose ligand is also HLA-C1. As such, it may be assumed that, in the presence of HLA-C1, an inhibitory KIR receptor-ligand combination is always formed. In addition, because both KIR2DL2 and KIR2DL3 are also frequently encoded by the same individual (but on separate haplotypes), the only scenario in which no inhibitory combination is formed is in the absence of recipient-encoded HLA-C1 (as previously determined in Table 3.6). Again, no significantly different outcomes were observed following this analysis (Table 3.7).

3.07.01.03 Donor KIR3DL1 and recipient HLA-Bw4

The univariate analysis to investigate the inhibitory combination of donor-encoded KIR3DL1 with recipient-encoded HLA-Bw4 revealed very little in the timepoint analyses (Table 3.7). However, a statistical significance was observed between the incidence of severe aGVHD and proposed NK cell inhibition via KIR3DL1: a decreased risk of grades 2-4 aGVHD was observed when donors encoded KIR3DL1 and the ligand, HLA-Bw4, was absent from the recipient (19% *vs* 29%, p=0.042). This factor persisted in a multivariate analysis model encompassing use of TCD, HSC source and donor age group (hazard ratio [HR]=0.54, 95% confidence interval [CI]=0.31-0.95, p=0.034, Table 3.8).

Table 3.8 Multivariate analysis assessing the impact of missing donor KIR3DL1 ligand on the incidence of grades 2-4 aGVHD

Variable	Grades 2-4 aGVHD			
	N	HR	95% CI	*p-value*
TCD by alemtuzumab				
Yes	320	1.00	-	-
No	17	***5.83***	***2.04-16.64***	***0.001***
HSC source				
BM	150	1.00	-	-
PBSC	187	***1.76***	***1.04-2.97***	***0.034***
Donor age group, years				
<30	107	1.00	-	-
>30	230	1.78	0.98-3.20	0.057
Missing donor KIR3DL1 ligand				
No missing ligand	215	1.00	-	-
Missing ligand	122	***0.54***	***0.31-0.95***	***0.034***

Statistically significant results are denoted by ***bold italics***.

The strength of KIR3DL1 interaction with HLA-Bw4 has been proposed to differ according to the sub-type of HLA-Bw4 allele: those encoding Bw4-80I are proposed to interact with more avidity. Although binding assays have revealed that residue 80 dimorphism alone is insufficient to indicate avidity strength [298], studies assessing HLA-Bw4 isoforms in both viral infection and HCT outcomes have observed

differences between the different isoform populations [271,299], thus warranting its investigation in this separate cohort.

To assess the impact of each isoform on the outcomes of HCT, the isoform structure of each recipient was determined. As the corresponding receptor, donor KIR3DL1 was also considered using the following subgrouping (strategy 1) to closely reproduce the study by Marra *et al.* (2015) [271]:

1. Donor KIR3DL1^{-ve} or Recipient HLA-Bw4^{-ve}
2. Donor KIR3DL1^{+ve}, Recipient HLA-Bw4-80I^{+ve}
3. Donor KIR3DL1^{+ve}, Recipient HLA-Bw4-80I^{-ve}

In addition, the recipients' HLA-Bw4-80I copy number was also determined and assessed in a separate sub-group analysis (strategy 2):

1. Donor KIR3DL1^{-ve} or Recipient HLA-Bw4-80I^{-ve}
2. Donor KIR3DL1^{+ve}, Recipient encodes one copy of HLA-Bw4-80I
3. Donor KIR3DL1^{+ve}, Recipient encodes two or more copies of HLA-Bw4-80I

In this cohort of donor-recipient pairs, neither of the strategies encompassing proposed ligand avidity resulted in any significant differences in any of the HCT outcomes studied (Table 3.7). However, a slight statistical trend was observed when comparing recipient HLA-Bw4-80I copy number (strategy 2). This trend implied that, with increasing HLA-Bw4-80I copies, likelihood of more severe aGVHD was increased (23% *vs* 29% *vs* 44%, p=0.092). However, when assessed by multivariate analysis, the HLA-Bw4-80I copy number model was no longer statistically significantly associated

with incidence of grades 2-4 aGVHD, although a trend was still observed for increased risk when multiple copies of HLA-Bw4-80I were encoded (HR=2.74, CI=0.89-8.46, p=0.079, Table 3.9).

Table 3.9 Multivariate analysis assessing the impact of missing donor KIR3DL1 ligand copy number on the incidence of grades 2-4 aGVHD

Variable	Grades 2-4 aGVHD			
	N	HR	95% CI	*p-value*
TCD by alemtuzumab				
Yes	320	1.00	-	-
No	17	***6.49***	***2.27-18.58***	***<0.001***
HSC source				
BM	150	1.00	-	-
PBSC	187	***1.71***	***1.02-2.88***	***0.044***
Donor age group, years				
<30	107	1.00	-	-
>30	230	1.79	0.99-3.23	0.054
Missing donor KIR3DL1 ligand strategy 2				
Donor KIR3DL1^{-ve} or recipient HLA-Bw4-80I^{-ve}	243	1.00	-	-
Donor KIR3DL1^{+ve}, recipient encodes 1 copy of HLA-Bw4-80I	80	1.47	0.82-2.63	0.194
Donor KIR3DL1^{+ve}, recipient encodes >1 copy of HLA-Bw4-80I	14	2.74	0.89-8.46	0.079

Statistically significant results are denoted by ***bold italics***.

3.07.01.04 Frequency of potentially alloreactive KIR receptor-ligand combinations

To account for the multiplicity of potential inhibitory KIR interactions, the total frequency of each recipient's missing ligands for inhibitory KIR genes present in the donor was calculated for each donor-recipient pair, giving a score of 0, 1 or 2. For this calculation, absence of donor KIR2DL1 or KIR3DL1 genes was assumed to be an inhibitory combination, regardless of ligand. Each subgroup was then assessed for the five HCT outcomes (Table 3.7). These analyses revealed that no significant difference was apparent for any of the outcomes when comparing the frequency of missing ligands for donor inhibitory KIR.

3.07.01.05 Donor KIR2DS1 and recipient HLA-C2

In addition to the inhibitory KIR receptor interactions detailed above, the activating KIR2DS1, in combination with its ligand, HLA-C2, has also been implicated in modulating the relapse risk following HCT to treat AML [300]. However, recipient homozygosity for the ligand has been demonstrated to increase the risk of relapse, whilst presence of at least one copy of HLA-C1, in combination with donor-encoded KIR2DS1, was shown to reduce relapse probability. As such, to assess the impact of KIR2DS1 and its ligand in my cohort, subgroups were developed as below:

1. Recipient HLA-C1^{-ve}
2. Donor KIR2DS1^{-ve}, Recipient HLA-C1^{+ve}
3. Donor KIR2DS1^{+ve}, Recipient HLA-C1^{+ve}

When each of the HCT outcomes available for analysis were assessed, it was not possible to detect significant differences between the subgroups, preventing our cohort from confirming the previous findings (Table 3.7).

3.07.02 The KIR matching model

The importance of genetic parity across the classical HLA genes has been well-documented. Less explored, however, is how matching of genes and alleles across the KIR locus affects the outcomes of allogeneic HCT. Because HLA and KIR reside on separate chromosomes, even the best HLA-matched donors can exhibit substantial genetic disparity when considering KIR genes. First studied as a comparison between HLA-identical (related) and HLA-matched (allogeneic) transplants, it was revealed that GVHD may result from KIR gene mismatching [282]. To evaluate the effects of KIR

matching in my cohort, I performed several different analyses based on previously published data [283,284]:

1. KIR genotype matching
2. Individual KIR gene matching
3. KIR gene function (inhibitory/activating) matching
4. KIR haplotype structure matching (AA *vs* BX)

3.07.02.01 KIR genotype matching

The presence/absence status of each non-framework KIR gene was assessed in the donor and recipient from each transplant pair and compared to one another. Should the status of the donor match that of the recipient for each of the KIR genes, their KIR genotypes were deemed to be matched. If the presence/absence status of one or more of the KIR genes were mismatched, the KIR genotype was denoted mismatched. A univariate analysis of the available HCT outcomes data between the genotype matched and mismatched groups revealed no significant differences, implying that the presence of any KIR gene mismatch is insufficient to affect the outcomes of HCT (Table 3.10). These analyses were also repeated to consider mismatches in the GVH direction only (e.g. KIR2DL1^{+ve} recipient and KIR2DL1^{-ve} donor constitutes a GVH mismatch but KIR2DL1^{-ve} recipient and KIR2DL1^{+ve} does not; Table 3.11). However, no significant differences were observed in this grouping either.

Table 3.10 Univariate analysis p-values of HCT outcomes based on KIR matching models

KIR matching model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
All loci matching					
KIR genotype matching	0.89	0.80	0.13	0.23	0.62
Individual KIR gene matching					
KIR2DL1 matching	0.052	***0.048***	***<0.001***	0.31	0.38
KIR2DL2 matching	0.93	0.92	0.74	0.81	0.64
KIR2DL3 matching	0.38	0.63	0.75	0.96	0.95
KIR2DL5 matching	0.74	0.70	0.59	0.36	0.10
KIR2DS1 matching	0.83	0.84	0.63	0.95	0.79
KIR2DS2 matching	0.54	0.74	0.79	0.94	0.60
KIR2DS3 matching	0.78	0.59	0.70	0.37	0.82
KIR2DS4 matching	0.77	0.46	0.34	0.91	0.75
KIR2DS5 matching	0.65	0.96	0.47	0.89	0.56
KIR3DL1 matching	0.91	0.56	0.26	0.82	0.85
KIR3DS1 matching	0.48	0.34	0.47	0.66	0.87
KIR gene function matching					
Inhibitory KIR gene matching	0.76	0.92	0.38	0.53	0.67
Activating KIR gene matching	0.89	0.83	0.11	0.19	0.74
KIR haplotype structure matching (KIR AA *vs* BX)					
KIR genotype matching	0.46	0.52	0.56	0.56	0.81

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table D.

3.07.02.02 Individual KIR gene matching

Despite no evidence of mismatching across the total KIR genotype having any detrimental effect on HCT outcomes, it was possible that the effects from individual KIR gene mismatches were being overlooked by grouping all mismatches together. Accordingly, mismatching at each non-framework KIR locus was assessed independently. Interestingly, mismatching at KIR2DL1 was associated with a decrease in both five year OS and DFS (KIR2DL1 matched *vs* mismatched; OS, 61% *vs* 82%, p=0.052; DFS, 65% *vs* 82%, p=0.048). However, these results should be interpreted with caution owing to the borderline significance values and the relatively small number of KIR2DL1 mismatched donor-recipient pairs (n=17, <5% of total cohort). Additionally, the final clinical follow-up in the KIR2DL1 mismatched group was at

2.75 years post-transplant. Despite this, the increased risk of relapse at this timepoint was highly significant (36% *vs* 88%, p<0.001, Figure 3.2). When the model was adjusted to account only for mismatches in the GVH direction, five year DFS and relapse both retained significance and OS at five years post-transplant also became a significant factor (Table 3.12).

Table 3.11 Univariate analysis p-values of HCT outcomes based on KIR matching models, GVH direction only

KIR GVH matching model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
All loci matching					
KIR genotype matching	0.33	0.27	0.68	0.15	0.36
Individual KIR gene matching					
KIR2DL1 matching	***0.022***	***0.007***	***0.022***	0.81	0.68
KIR2DL2 matching	0.87	0.99	0.88	0.97	0.63
KIR2DL3 matching	0.20	0.31	0.59	0.14	0.41
KIR2DL5 matching	0.46	0.48	0.71	0.63	0.66
KIR2DS1 matching	0.32	0.47	0.65	0.15	0.43
KIR2DS2 matching	0.79	0.90	0.82	0.93	0.73
KIR2DS3 matching	0.70	0.66	0.30	0.55	0.62
KIR2DS4 matching	0.38	0.57	0.37	0.13	0.57
KIR2DS5 matching	0.24	0.46	0.40	0.094	0.88
KIR3DL1 matching	0.39	0.58	0.36	0.14	0.57
KIR3DS1 matching	0.47	0.59	0.94	0.28	0.58
KIR gene function matching					
Inhibitory KIR gene matching	0.36	0.37	0.75	0.61	0.27
Activating KIR gene matching	0.75	0.55	0.97	0.70	0.45

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table D.

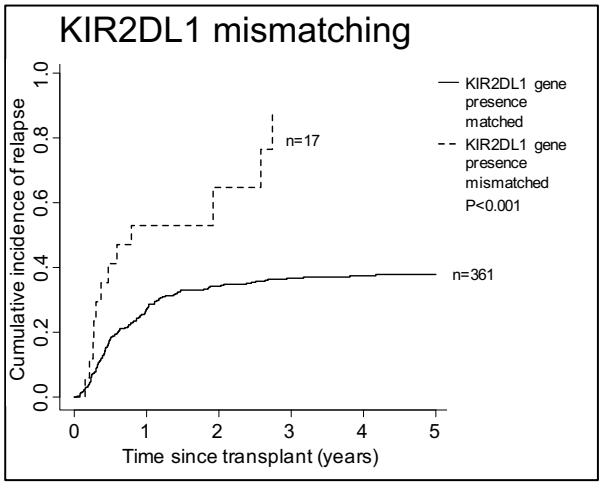


Figure 3.2 Probability of relapse comparing KIR2DL1 matching status

Recipients of transplants characterised by KIR2DL1 gene presence mismatch are more likely to experience relapse than recipients of KIR2DL1 matched transplants (p<0.001).

To further evaluate the validity of these observations, multivariate analysis was performed. When assessing the influence of KIR2DL1 gene mismatching in any direction, significance is lost for five year OS and DFS probability (Table 3.12). However, the presence of KIR2DL1 gene mismatches is still significantly prognostic of relapse at five years post-transplant (HR=2.53, CI=1.46-4.41, p=0.001, Table 3.13). Conversely, when investigating only those KIR2DL1 gene mismatches in the GVH direction, five year DFS probability remained significantly reduced in their presence (HR=2.38, CI=1.05-5.26, p=0.038, Figure 3.3, Table 3.14), whilst significance was lost when assessing relapse at five years post-transplant, although a trend was still observed (HR=2.41, CI=0.97-5.99, p=0.059, Table 3.15).

Table 3.12 Multivariate analysis assessing the impact of KIR2DL1 gene mismatching on five year OS and DFS probability

Variable	5 year OS				5 year DFS			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
HLA matching								
10/10	265	1.00	-	-		-	-	-
9/10	90	1.14	0.84-1.54	0.41	-	-	-	-
<9/10	23	***2.22***	***1.41-3.57***	***<0.001***	-	-	-	-
Donor age, years								
<30	-	-	-	-	110	1.00	-	-
>30	-	-	-	-	256	***1.39***	***1.04-1.89***	***0.026***
Previous autografts								
0	354	1.00	-	-	343	1.00	-	-
≥1	24	***1.64***	***1.03-2.56***	***0.038***	23	***1.69***	***1.06-2.70***	***0.027***
CMV matching								
Matched	268	1.00	-	-	261	1.00	-	-
Mismatched	110	***1.59***	***1.20-2.08***	***0.001***	105	***1.61***	***1.22-2.08***	***<0.001***
KIR2DL1 mismatching								
Matched	362	1.00	-	-	350	1.00	-	-
Mismatched	16	1.52	0.86-2.70	0.15	16	1.35	0.76-2.38	0.30

Statistically significant results are denoted by ***bold italics***.

Table 3.13 Multivariate analysis assessing the impact of KIR2DL1 gene mismatching on five year relapse risk

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Disease risk score				
Good	167	1.00	-	-
Intermediate	165	0.87	0.61-1.23	0.43
Poor	42	***2.18***	***1.31-3.63***	***0.003***
Donor-recipient gender matching				
Matched	218	1.00	-	-
Mismatched	156	1.31	0.92-1.85	0.13
KIR2DL1 mismatching				
Matched	357	1.00	-	-
Mismatched	17	***2.53***	***1.46-4.41***	***0.001***

Statistically significant results are denoted by ***bold italics***.

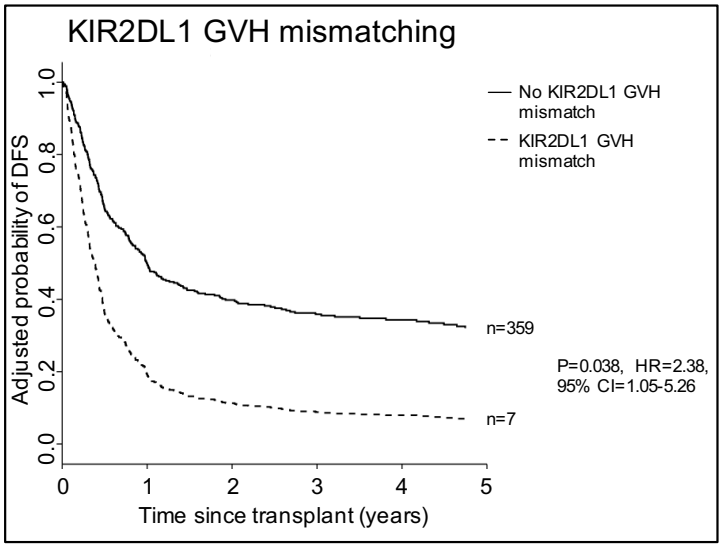


Figure 3.3 Cox regression analysis of DFS probability comparing KIR2DL1 GVH matching status

The adjusted probability of DFS at five years post-transplant is significantly reduced by GVH direction KIR2DL1 gene presence mismatching (p=0.038, HR=2.38, CI=1.05-5.26). Cox regression analysis adjusted for donor age group, previous autograft status, donor-recipient CMV status matching and KIR2DL1 mismatching in the GVH direction.

Table 3.14 Multivariate analysis assessing the impact of KIR2DL1 gene mismatching on five year OS and DFS, GVH direction only

Variable	5 year OS				5 year DFS			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
HLA matching								
10/10	265	1.00	-	-		-	-	-
9/10	90	1.15	0.85-1.56	0.37		-	-	-
<9/10	23	***2.22***	***1.39-3.57***	***0.001***		-	-	-
Donor age, years								
<30		-	-	-	110	1.00	-	-
>30		-	-	-	256	***1.41***	***1.04-1.89***	***0.023***
Previous autografts								
0	354	1.00	-	-	343	1.00	-	-
≥1	24	***1.67***	***1.04-2.63***	***0.032***	23	***1.72***	***1.08-2.70***	***0.021***
CMV matching								
Matched	268	1.00	-	-	261	1.00	-	-
Mismatched	110	***1.61***	***1.22-2.13***	***<0.001***	105	***1.61***	***1.23-2.13***	***<0.001***
KIR2DL1 GVH mismatching								
GVH Matched	371	1.00	-	-	359	1.00	-	-
GVH Mismatched	7	2.17	0.97-5.00	0.060	7	***2.38***	***1.05-5.26***	***0.038***

Statistically significant results are denoted by ***bold italics***.

Table 3.15 Multivariate analysis assessing the impact of KIR2DL1 gene mismatching on five year relapse risk, GVH direction only

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Disease risk score				
Good	167	1.00	-	-
Intermediate	165	0.89	0.63-1.27	0.53
Poor	42	***2.27***	***1.37-3.77***	***0.002***
Donor-recipient gender matching				
Matched	218	1.00	-	-
Mismatched	156	1.31	0.92-1.86	0.13
KIR2DL1 GVH mismatching				
GVH Matched	366	1.00	-	-
GVH Mismatched	8	2.41	0.97-5.99	0.059

Statistically significant results are denoted by ***bold italics***.

Mismatches at each of the other KIR loci did not relate to a statistically significant difference in any of the HCT outcomes assessed. This remained the case when considering only the mismatches in the GVH direction. In addition, no differences in severe aGVHD incidence or NRM probability at one year were observed for any of the conditions analysed.

3.07.02.03 Matching according to KIR gene function

In 2018, Sahin *et al.* [283] proposed that matching models of KIR genes could be refined according to the nature of the KIR gene function. As such, the presence/absence matching status for both the activating and inhibitory KIR gene subsets of my cohort was determined and analysed. However, no statistical significance was observed with either activating or inhibitory KIR gene matching for any of the HCT outcomes available (Table 3.10). When analysing only those mismatches in the GVH direction, the lack of significance persisted (Table 3.11).

3.07.02.04 KIR haplotype structure matching

Faridi *et al.* (2016) [284] postulated that an alternative KIR matching strategy, examining the genotypes assigned according to haplotype combinations (KIR AA *vs* BX), might be used to predict risk of cGVHD. Although the Faridi *et al.* (2016) study revealed no significant differences in the majority of HCT outcomes I am able to analyse (as there is no available cGVHD data for my cohort), a trend towards decreased aGVHD grades 2-4 was also noted by the authors. Unfortunately, analysis on my cohort was not able to confirm this trend ($p>0.1$). My findings did, however, support the lack of effect in OS, DFS and relapse (Table 3.10).

3.07.03 The donor KIR B content model

At present, it is widely regarded that the most appropriate method for assessing the impact of KIR in allogeneic HCT is the donor KIR B content model that utilises donor KIR genotyping data to calculate the frequency of B motifs in both the centromeric and telomeric sections of the KIR haplotype. Numerous iterations of this strategy have been developed by the Minnesota group [83,272,273].

The simplest version of this assesses only the donor's broad KIR genotype: KIR AA *vs* BX. When applied to my entire cohort, I was unable to confirm any differences between the outcomes of transplants utilising donors who encoded KIR BX and those who encoded KIR AA (Table 3.16).

Table 3.16 Univariate analysis p-values of HCT outcomes based on donor KIR B content models

KIR B content model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
KIR AA *vs* BX	0.47	0.87	0.24	0.47	0.17
Cen-AA *vs* BX	0.32	0.50	0.27	0.12	0.55
Tel-AA *vs* BX	0.41	0.17	0.54	0.075	0.37
Neutral *vs* Better *vs* Best	0.36	0.42	0.79	0.36	0.96
Neutral *vs* Better/Best	0.73	0.51	0.47	0.65	0.83

The number of transplants within each subgroup of each test is given in Supplementary Table E.

As such, the analysis model was further refined: stratifying transplants according to the presence of donor Cen and/or Tel-B motifs, and the "neutral", "better" and "best" KIR B content scores defined by Cooley *et al.* (2010) [273]. Following this stratification, it was still not possible to observe any significant differences between the outcomes. However, a slight decrease in NRM at one year post-transplant was noted when donors encoded at least one copy of the Tel-B motif, although this only reached a statistical trend (25% *vs* 17%, p=0.075). Furthermore, this lost any statistical significance following multivariate analysis (Table 3.17).

Table 3.17 Multivariate analysis assessing the impact of donor Tel-BX haplotype motif structure on one year NRM probability

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10	277	1.00	-	-
9/10	91	0.71	0.42-1.20	0.20
<9/10	24	***2.47***	***1.28-4.79***	***0.007***
Era				
1996-1999	34	1.00	-	-
2000-2003	122	***0.37***	***0.19-0.72***	***0.003***
2004-2007	129	***0.46***	***0.24-0.87***	***0.016***
2008-2011	107	***0.48***	***0.25-0.92***	***0.026***
Donor Tel motif structure				
Tel-AA	229	1.00	-	-
Tel-BX	163	0.76	0.47-1.21	0.25

Statistically significant results are denoted by ***bold italics***.

The most recent adaptation from the Minnesota group combined the transplants previously defined as "Better" or "Best" into a single category for analysis,

"Better/Best". However, when similar filtering parameters were applied, no differences were observed for any of the HCT outcomes assessed (Table 3.16).

3.08 Results: Donor KIR-mediated influences according to recipient HLA-C1 status

Analysis by Fischer *et al.* (2007) [301] has highlighted important differences between the outcomes of HCT in recipients that are HLA-C1^{+ve} or HLA-C1^{-ve}. Specifically, HLA-C1^{-ve} transplant recipients have demonstrated inferior outcomes, potentially as a result of delayed KIR2DL1-mediated immunity.

3.08.01 Comparison of HLA-C1 positive and HLA-C1 negative recipient outcomes

To assess whether any HLA-C1-dependent differences in HCT outcomes were replicated in this cohort, the cohort was stratified by HLA-C1 status. Forty-four transplants were performed in HLA-C1^{-ve} patients, whilst the remaining 361 recipients all encoded at least one HLA-C1^{+ve} allele. Univariate analysis revealed no significant differences within the five main HCT outcomes analysed in this study when comparing HLA-C1^{+ve} and HLA-C1^{-ve} recipients (Table 3.18).

Table 3.18 Univariate analysis p-values of HCT outcomes based on recipient HLA-C1 status

	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
HLA-C1^{+ve} *vs* C1^{-ve} recipients	1.00	0.68	0.48	0.58	0.94

3.08.02 Comparison of outcomes within the HLA-C1 negative sub-cohort

Despite no apparent differences in HCT outcomes between HLA-C1^{+ve} and HLA-C1^{-ve} recipients, KIR-mediated differences may be more apparent in the HLA-C1^{-ve} subgroup.

Neuchel *et al.* (2017) [302] implicated several activating KIR genes in alteration of outcomes in this stratum. Because of the relative infrequency of HLA-C1^{-ve} recipients within the broader cohort and the subsequent restricted cohort size (n=44), analysis of this subgroup was limited to only those tests that had been demonstrated to show significant effects in the Neuchel *et al.* (2017) study. Univariate analysis failed to reveal any significant differences in outcomes based on the presence or absence of individual activating KIR genes, although the presence of KIR2DS5 in the donor's genotype correlated with a trend towards increased relapse risk at five years post-transplant (p=0.079) and decreased five year DFS probability (p=0.09, Table 3.19), although neither of these findings retained any significance in multivariate analyses (p>0.1).

Table 3.19 Univariate analysis p-values of HCT outcomes based on donor activating KIR gene content in the HLA-C1^{-ve} recipient sub-cohort

Donor activating KIR presence	5 year OS	5 year DFS	5 year relapse	1 year NRM
KIR2DS1 present *vs* absent	0.40	0.40	0.20	0.69
KIR2DS2 present *vs* absent	0.43	0.69	0.94	0.38
KIR2DS5 present *vs* absent	0.73	0.090	0.079	0.98
KIR3DS1 present *vs* absent	0.11	0.40	0.20	0.69

3.08.03 Comparison of outcomes within the HLA-C1 positive recipient cohort

3.08.03.01 KIR genotype matching does not affect HCT outcomes

Faridi *et al.* (2016) [284] proposed that the occurrence of KIR genotype mismatching between donors and recipients led to increased cGVHD and subsequent reduced cGVHD and relapse free survival (cGRFS), particularly in HLA-C1^{+ve} recipients. As such, the HLA-C1^{+ve} subset of patients for whom donor-recipient KIR matching information was available (n=345) was assessed for each of the five HCT outcomes. The results indicated that none of these outcomes were significantly affected by KIR

genotype matching (Table 3.20). Unfortunately, no data was available relating to the incidence of cGVHD, meaning this outcome was not assessed.

Table 3.20 Univariate analysis p-values of HCT outcomes based on donor-recipient KIR genotype matching in the HLA-C1^{+ve} recipient sub-cohort

	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
KIR genotype matching	0.62	0.67	0.35	0.44	0.75

3.08.03.02 Donor KIR B motif content does not affect HCT outcomes following HLA-C mismatched transplants

In a recent study by the Minnesota group, it was demonstrated that recipients of HCT from donors encoding two or more KIR B motifs ('Better/Best' donors) experience significantly less relapse than recipients of HCT from donors with only one or no KIR B haplotype motifs ('Neutral' donors). Importantly, this reduction in relapse risk is more evident in HLA-C1^{+ve} recipients who undergo HLA-C mismatched HCT [272]. As such, a similar subgroup analysis was performed on this stratum of the overall cohort (n=56). Similarly to the HLA-C1^{-ve} cohort, analyses were restricted to only those tested in the original publication to minimise multiplicity of testing. Following univariate analysis, it was not possible to demonstrate any statistical indication of differences relating to KIR B content (Table 3.21), thus providing further indication that there is little significant difference in HCT outcomes when stratifying by recipient's HLA-C1 ligand status in this cohort.

Table 3.21 Univariate analysis p-values of HCT outcomes based on the donor neutral/better/best model in HLA-C1^{+ve} recipients of HLA-C mismatched transplants

KIR B content model	5 year OS	5 year DFS	5 year relapse	1 year NRM
Neutral *vs* Better/Best	0.52	0.33	0.51	0.24

3.09 Results: Donor KIR-mediated influences on the HLA-matched, RIC, adult AML sub-cohort

To reduce the heterogeneity of the entire cohort, and to investigate the influence of conditioning regimen on the outcomes of HCT, subsections of the entire cohort, restricted to only those transplants that were 10/10 HLA-matched and non-paediatric cases were studied (n=254). The HLA-matched, adult cohort was then further stratified according to conditioning regimen, and clinical analyses performed to assess the impact of each different donor KIR scenario. The results from the RIC sub-cohort (n=135) are reported below.

3.09.01 Transplantations characteristics

The transplantation characteristics for this subset of donor-recipient pairs is given in Supplementary Table F. Several interesting differences were noted between the adult, HLA-matched, RIC subgroup and the overall cohort. First, recipients in the RIC subgroup had increased likelihood of being older (>40 years, 90% *vs* 52%, p<0.001), a reflection of the more tolerable RIC regimens. In addition, more transplants utilised PBSC as an HSC source (70% *vs* 56%, p=0.005), possibly indicative of its increased use in later eras, which was also significantly different between the adult, HLA-matched, RIC subset and the overall cohort (p<0.001).

3.09.02 Outcomes summary

Estimated probability of OS, DFS, relapse and NRM at their designated endpoints were 38%, 32%, 42% and 16%, respectively. This low rate of NRM at one year post-transplant is significantly different from the remainder of transplants in the overall cohort (p=0.025) and represents the lower toxicity levels associated with RIC regimens.

When considering five year OS, DFS and relapse probabilities, no significant differences to the remainder of the overall cohort were observed.

In contrast to equivalent analysis on the overall cohort, differences within clinical factors had no statistically significant impact on the outcomes of HCT in the adult, HLA-matched RIC cohort (Supplementary Table B). However, several factors formed trends. Patient age group became a statistical trend in OS, with older recipients more likely suffer death by five years post-transplant (p=0.051). The impact of the use of alemtuzumab in this cohort was also a statistical trend, with improved five year OS probability in those transplants utilising alemtuzumab (p=0.076). When considering five year DFS probability, trends for beneficial outcome were observed in transplants featuring younger donors or patients (p=0.067 and p=0.097, respectively). Statistical trends towards benefits in aGVHD incidence were also observed, whereby use of alemtuzumab and BM-derived HSCs resulted in less aGVHD (p=0.093 and p=0.095, respectively). The probabilities of relapse at five years and NRM at one year post-transplant were unaffected by any clinical factors available for analysis.

3.09.03 The KIR receptor-ligand model

The impact of missing ligands was assessed in the adult, HLA-matched, RIC cohort, and the results are summarised in Table 3.22. Due to the low frequency of recipients encoding two or more copies of HLA-Bw4-80I (n=5), the KIR3DL1 missing ligand strategy 2 test was not performed in this restricted sub-cohort. The results from the remaining analyses reveal that no missing ligand scenario were associated with any difference in any of the HCT outcomes assessed.

Table 3.22 Univariate analysis p-values of HCT outcomes based on specific donor KIR, recipient ligand models in the adult, HLA-matched, RIC HCT sub-cohort

Donor KIR, Recipient Ligand	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
KIR2DL1^{+ve}, HLA-C2^{-ve}	0.71	0.98	0.73	0.54	0.87
KIR2DL2/3^{+ve}, HLA-C1^{-ve}	0.51	0.40	0.79	0.91	0.46
KIR3DL1^{+ve}, HLA-Bw4^{-ve}	0.67	0.80	0.92	0.93	0.68
KIR3DL1^{+ve}, HLA-Bw4-80I strategy 1^{-ve}	0.65	0.78	0.90	0.28	0.46
Missing ligand frequency	0.26	0.62	0.90	0.46	0.75
KIR2DS1^{+ve}, HLA-C2^{-ve}	0.63	0.30	0.70	0.78	0.72

The number of transplants within each subgroup of each test is given in Supplementary Table C.

3.09.04 The KIR matching model

When assessing the impact of different KIR matching scenarios on the impact of HCT outcomes in the adult, HLA-matched, RIC cohort, the frequency of KIR2DL1 mismatches between donors and recipients (n=5) prevented valid analysis. In addition, mismatches in the GVH direction for KIR2DS4 and KIR3DL1 were also very few (n=6 and n=6). As such, comparative tests were not performed for these scenarios. The results from the remaining tests are given in Table 3.23 and Table 3.24.

Table 3.23 Univariate analysis p-values of HCT outcomes based on KIR matching models in the adult, HLA-matched, RIC HCT sub-cohort

KIR matching model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
All loci matching					
KIR genotype matching	0.39	0.34	0.25	0.79	0.12
Individual KIR gene matching					
KIR2DL2 matching	0.72	0.62	0.78	0.92	0.20
KIR2DL3 matching	0.52	0.16	0.16	0.65	0.53
KIR2DL5 matching	0.82	0.82	0.96	0.63	0.13
KIR2DS1 matching	0.79	0.47	0.48	0.53	0.93
KIR2DS2 matching	0.86	0.87	0.80	0.92	0.20
KIR2DS3 matching	0.14	0.27	0.77	0.34	0.18
KIR2DS4 matching	0.39	0.27	0.32	0.70	0.21
KIR2DS5 matching	0.73	0.37	0.24	0.39	0.82
KIR3DL1 matching	0.38	0.26	0.31	0.69	0.21
KIR3DS1 matching	0.41	0.50	0.64	0.76	0.92
KIR gene function matching					
Inhibitory KIR gene matching	0.99	0.96	0.42	0.28	0.74
Activating KIR gene matching	0.40	0.37	0.21	0.64	0.13
KIR haplotype structure matching (KIR AA *vs* BX)					
KIR genotype matching	0.56	0.88	0.82	0.18	0.81

The number of transplants within each subgroup of each test is given in Supplementary Table D.

Table 3.24 Univariate analysis p-values of HCT outcomes based on KIR matching models in the adult, HLA-matched, RIC HCT sub-cohort, GVH direction only

KIR GVH matching model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
All loci matching					
KIR genotype matching	0.50	0.37	0.90	0.61	0.47
Individual KIR gene matching					
KIR2DL2 matching	0.39	0.20	0.91	0.068	0.59
KIR2DL3 matching	0.86	0.34	0.49	0.94	0.38
KIR2DL5 matching	0.40	0.12	0.092	0.73	1.00
KIR2DS1 matching	0.62	0.66	0.92	0.89	0.76
KIR2DS2 matching	0.51	0.30	0.82	***0.039***	0.78
KIR2DS3 matching	0.24	0.88	0.58	0.41	0.36
KIR2DS5 matching	0.62	0.71	0.99	0.69	0.78
KIR3DS1 matching	0.76	0.54	0.39	0.97	0.36
KIR gene function matching					
Inhibitory KIR gene matching	0.24	0.21	0.44	0.60	0.61
Activating KIR gene matching	0.66	0.28	0.70	0.76	0.12

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table D.

The results from this univariate analysis revealed only one statistically significant difference: mismatching of the KIR2DS2 gene in the GVH direction associated with increased probability of NRM at one year post-transplant (matched *vs* mismatched, 12% *vs* 29%, p=0.039, Figure 3.4). This complements the statistical trend observed when considering GVH direction KIR2DL2 mismatching (p=0.068), as these genes are in strong linkage disequilibrium. Multivariate analysis was not performed in this scenario as no clinical factors remained following forward stepwise selection of co-variables.

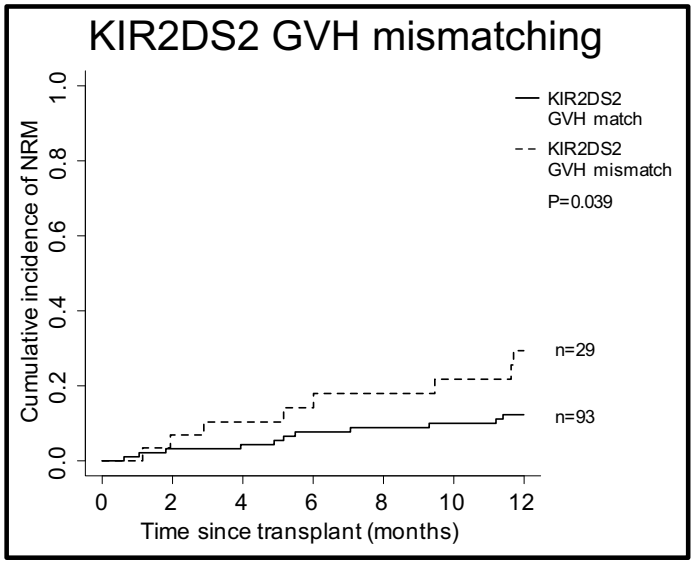


Figure 3.4 Probability of NRM comparing GVH direction KIR2DS2 matching status

Transplants featuring KIR2DS2 mismatching in the GVH direction are more likely to result in NRM at one year post-transplant (p=0.039).

3.09.05 The donor KIR B content model

3.09.05.01 Donor Cen-B motifs protect against occurrence of NRM

Although no significant observations were made when comparing different KIR subgroups on the probability of OS, DFS or relapse at 5 years post-transplant, the occurrence of NRM at one year post-transplant was significantly increased in recipients of transplants from donors encoding the Cen-AA haplotype motif structure (26% *vs* 6%, p=0.006, Figure 3.5, Table 3.25). This effect was particularly apparent after 5 months, supporting the observation that this KIR factor was not associated with aGVHD (p=NS, Table 3.25). A separate observation, investigating the influence of the "better/best" score on one year NRM risk, found borderline significance (p=0.045) implicating improved NRM probability in the better/best subgroup of recipients. However, a similar effect was not observed when considering the presence of the Tel-B motif (p=NS, Table 3.25), strongly suggesting that this effect is a result of centromeric factors. As no

clinical factors remained following forward stepwise selection of co-variables, multivariate regression analysis was not performed.

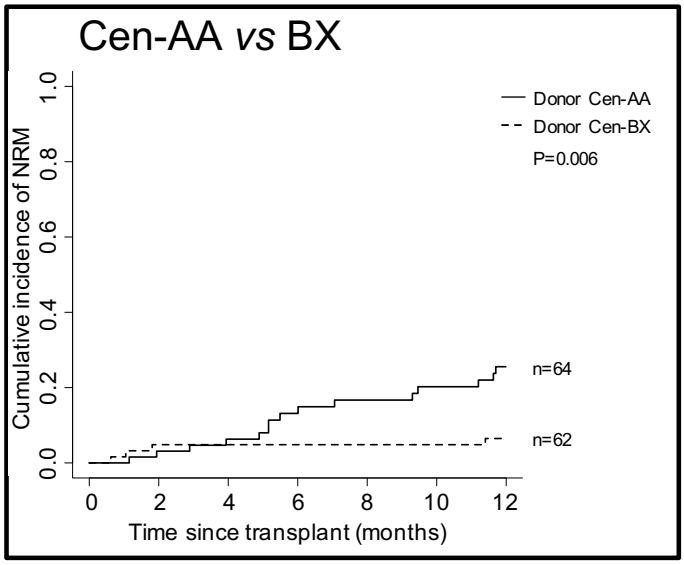


Figure 3.5 Probability of NRM comparing donor Cen-AA and Cen-BX haplotype motif structures

Univariate analysis demonstrates that the presence of at least one donor-encoded KIR B motif is protective against the risk of NRM at one year post-transplant (p=0.006).

When the causes of death were investigated further in the transplants resulting in NRM at the one year post-transplant timepoint, infection (n=13, 65%) was slightly more prevalent than GVHD (n=8, 40%), although in four cases (20%) both GVHD and infection were co-mortality factors. Several other factors also contributed to mortality in this sub-cohort, including cardiac toxicity (n=1), interstitial pneumonitis (n=1), rejection (n=1), renal failure (n=1) and multi-organ failure (n=4). Cause of death was unknown in two cases (10%). Due to this high heterogeneity within a small subset of transplants, it was not possible to further investigate the causes of NRM.

Table 3.25 Univariate analysis p-values of HCT outcomes based on donor KIR B content models in RIC, adult, HLA-matched sub-cohort

KIR B content model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
KIR AA *vs* BX	0.87	0.26	0.50	0.089	0.80
Cen-AA *vs* BX	0.60	0.21	0.80	***0.006***	0.50
Tel-AA *vs* BX	0.64	0.14	0.16	0.41	0.90
Neutral *vs* Better/Best	0.53	0.62	0.64	***0.045***	0.69

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table E.

3.10 Results: Donor KIR-mediated influences on the HLA-matched, MAC, adult AML sub-cohort

As mentioned in Section 3.03, clinical analyses were repeated on divisions of the entire cohort. The section below concerns the outcomes from the analysis of HLA-matched, adult transplants utilising MAC regimens (n=119).

3.10.01 Transplantations characteristics

The transplantation characteristics for this subset of donor-recipient pairs is given in Supplementary Table F. This highlights no substantial unexpected variation from the overall cohort statistics, with the exception of recipient age. This statistic appears to indicate that, despite removing the paediatric cases from this cohort, the proportion of younger recipients has increased (<40 years, 48% *vs* 71%, p<0.001). This likely reflects the preference of use of MAC for younger patients who can better tolerate its toxicity.

Several differences also arose between the MAC and RIC adult, HLA-matched subgroups. MAC subset recipients were more likely to be younger (p<0.001) and receive BM-derived HSCs (p=0.01) in earlier eras (p<0.001) than recipients in the adult, HLA-matched, RIC subset.

3.10.02 Outcomes summary

Estimated probability of OS, DFS, relapse and NRM at their designated endpoints were 39%, 36%, 35% and 23%, respectively, indicating that there are no statistically significant differences between this sub-cohort and the remainder of the transplants.

The impact of several clinical factors on the outcomes of HCT differed between the smaller MAC cohort and the entire cohort (Supplementary Table B). For five year OS, the younger recipient age group became a borderline significant factor indicative of improved OS ($p=0.049$), whilst significance was lost for donor age, recipient EMBT risk score and donor-recipient CMV matching. When considering both five year DFS and one year NRM probabilities, no clinical factors retained significance. Whilst EBMT risk score remained a significant predictor of AML relapse ($p<0.001$), the impact of donor age was lost, a characteristic also observed in incidence of aGVHD. However, the era of transplant became a borderline significant indicator of relapse at five years post-transplant ($p=0.049$).

3.10.03 The KIR receptor-ligand model

Each of the KIR and concomitant ligand combinations described in Section 3.07.01 were assessed in the adult, HLA-matched, MAC sub-cohort. Because the population of the "two or more copies" subgroup within the recipient HLA-Bw4-80I copy number test was very small (n=4), this test was not performed in this restricted sub-cohort. The findings from univariate analyses are summarised in Table 3.26.

3.10.03.01 Missing KIR3DL1 ligand associates with increased relapse risk

This analysis revealed a statistical trend towards difference in relapse risk at five years post-transplant between transplants featuring opposing missing KIR3DL1 ligand scenarios. However, in contrast to the hypothesis, the absence of ligand for donor KIR3DL1 corresponded with increased relapse risk (28% *vs* 47%, p=0.083, Figure 3.6). Although only borderline significance was denoted in univariate analysis, multivariate analysis adjusting for disease risk score suggested that the risk of disease relapse at five years post-transplant in the missing ligand scenario was statistically significant (HR=1.96, CI=1.07-3.59, p=0.03, Table 3.27). This increase in relapse risk was not associated with changes to the probability of OS, DFS or NRM at their designated endpoints.

Table 3.26 Univariate analysis p-values of HCT outcomes based on specific donor KIR, recipient ligand models in the adult, HLA-matched, MAC HCT sub-cohort

Donor KIR, Recipient Ligand	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
KIR2DL1^{+ve}, HLA-C2^{-ve}	0.46	0.51	0.64	0.83	0.28
KIR2DL2/3^{+ve}, HLA-C1^{-ve}	0.77	0.95	0.86	0.65	0.49
KIR3DL1^{+ve}, HLA-Bw4^{-ve}	0.43	0.54	0.083	0.67	0.25
KIR3DL1^{+ve}, HLA-Bw4-80I strategy 1^{-ve}	0.59	0.73	0.36	0.84	0.84
Missing ligand frequency	0.95	0.95	0.69	0.60	0.52
KIR2DS1^{+ve}, HLA-C2^{-ve}	0.88	0.95	0.98	0.56	0.76

The number of transplants within each subgroup of each test is given in Supplementary Table C.

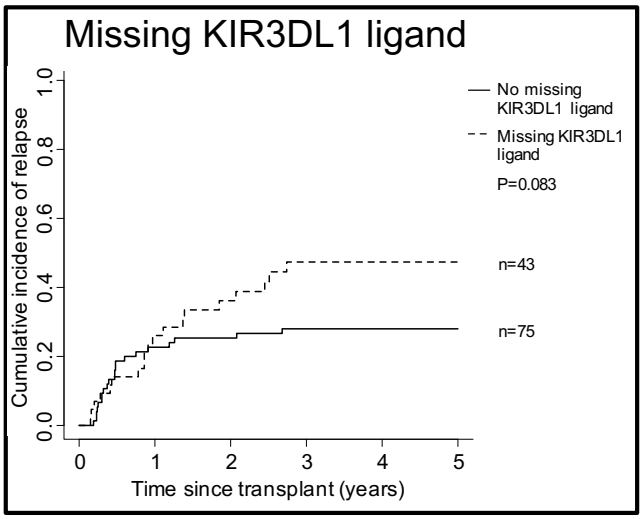


Figure 3.6 Probability of relapse comparing KIR3DL1 missing ligand status

Transplants in which donors encode KIR3DL1 but their concomitant recipient does not encode the HLA-Bw4 ligand are more likely to have experienced relapse at five years post-transplant than transplants that do not feature a missing ligand scenario.

Table 3.27 Multivariate analysis assessing the impact of KIR3DL1 missing ligand on five year relapse risk

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Disease risk score				
Good	50	1.00	-	-
Intermediate	52	1.06	0.52-2.17	0.87
Poor	15	***4.66***	***2.09-10.40***	***<0.001***
KIR3DL1 missing ligand				
No missing KIR3DL1 ligand	74	1.00	-	-
Missing KIR3DL1 ligand	43	***1.96***	***1.07-3.59***	***0.030***

Statistically significant results are denoted by ***bold italics***.

3.10.04 The KIR matching model

The results from the adult, HLA-matched, MAC sub-cohort analysis of KIR gene matching between donor and recipient are summarised in Table 3.28 and Table 3.29. Due to low frequency of KIR mismatching in the GVH direction for the KIR2DL1, KIR2DS4 and KIR3DL1 genes (n=4, n=6 and n=6, respectively), these loci were excluded from the analysis.

Table 3.28 Univariate analysis p-values of HCT outcomes based on KIR matching models in the adult, HLA-matched, MAC HCT sub-cohort

KIR matching model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
All loci matching					
KIR genotype matching	0.45	0.96	0.11	0.23	0.73
Individual KIR gene matching					
KIR2DL2 matching	0.50	0.50	0.37	0.85	0.69
KIR2DL3 matching	0.13	0.24	0.95	0.40	1.00
KIR2DL5 matching	0.56	0.88	0.97	0.88	0.77
KIR2DS1 matching	0.73	0.75	0.75	0.76	0.91
KIR2DS2 matching	0.54	0.52	0.48	0.78	0.62
KIR2DS3 matching	0.22	0.33	0.54	0.27	0.61
KIR2DS5 matching	0.52	0.51	0.95	0.58	0.24
KIR3DS1 matching	0.90	0.96	0.97	0.96	0.99
KIR gene function matching					
Inhibitory KIR gene matching	0.92	0.53	0.23	0.98	0.78
Activating KIR gene matching	0.45	0.96	0.11	0.23	0.73
KIR haplotype structure matching (KIR AA/BX)					
KIR genotype matching	0.31	0.26	0.54	0.14	0.93

The number of transplants within each subgroup of each test is given in Supplementary Table D.

Table 3.29 Univariate analysis p-values of HCT outcomes based on KIR matching models in the adult, HLA-matched, MAC HCT sub-cohort, GVH direction only

KIR GVH matching model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
All loci matching					
KIR genotype matching	0.31	0.56	***0.034***	***0.045***	0.65
Excluding KIR2DL2/S2	***0.004***	***0.018***	0.15	***<0.001***	0.57
Individual KIR gene matching					
KIR2DL2 matching	***0.030***	***0.041***	0.47	***0.018***	1.00
KIR2DL3 matching	***0.011***	***0.031***	0.91	0.11	0.29
KIR2DL5 matching	0.78	0.47	0.23	0.79	0.65
KIR2DS1 matching	0.44	0.63	0.72	0.18	0.90
KIR2DS2 matching	***0.030***	***0.041***	0.47	***0.018***	1.00
KIR2DS3 matching	0.42	0.43	0.15	0.95	0.78
KIR2DS5 matching	0.42	0.71	0.30	0.059	0.48
KIR3DS1 matching	0.60	0.92	0.52	0.33	0.82
KIR gene function matching					
Inhibitory KIR gene matching	0.75	0.94	0.11	0.56	0.91
Activating KIR gene matching	0.90	0.65	***0.037***	0.32	0.20

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table D.

3.10.04.01 KIR2DL2/S2/L3 mismatches in the graft-versus-host direction affect OS and DFS

When assessing overall KIR gene mismatching, no significant differences were observed for any of the mismatching scenarios tested (Table 3.28). However, when investigating only those mismatches in the GVH direction, several significant observations were made. For those transplants in which recipients encoded KIR2DL2 but the corresponding donor did not (GVH direction mismatch), the probability of five year OS was significantly increased (matched *vs* mismatched: 33% *vs* 60%, p=0.03, Figure 3.7A). Unsurprisingly, these results were identical when comparing KIR2DS2 GVH mismatching, due to the strong linkage disequilibrium between KIR2DL2 and KIR2DS2. Furthermore, presence of GVH direction KIR2DL3 mismatching had the opposite effect: a dramatic decrease in five year OS probability (42% *vs* 8%, p=0.011, Figure 3.7B). These results each correlated with significant changes to DFS probability at five years post-transplant (Table 3.29).

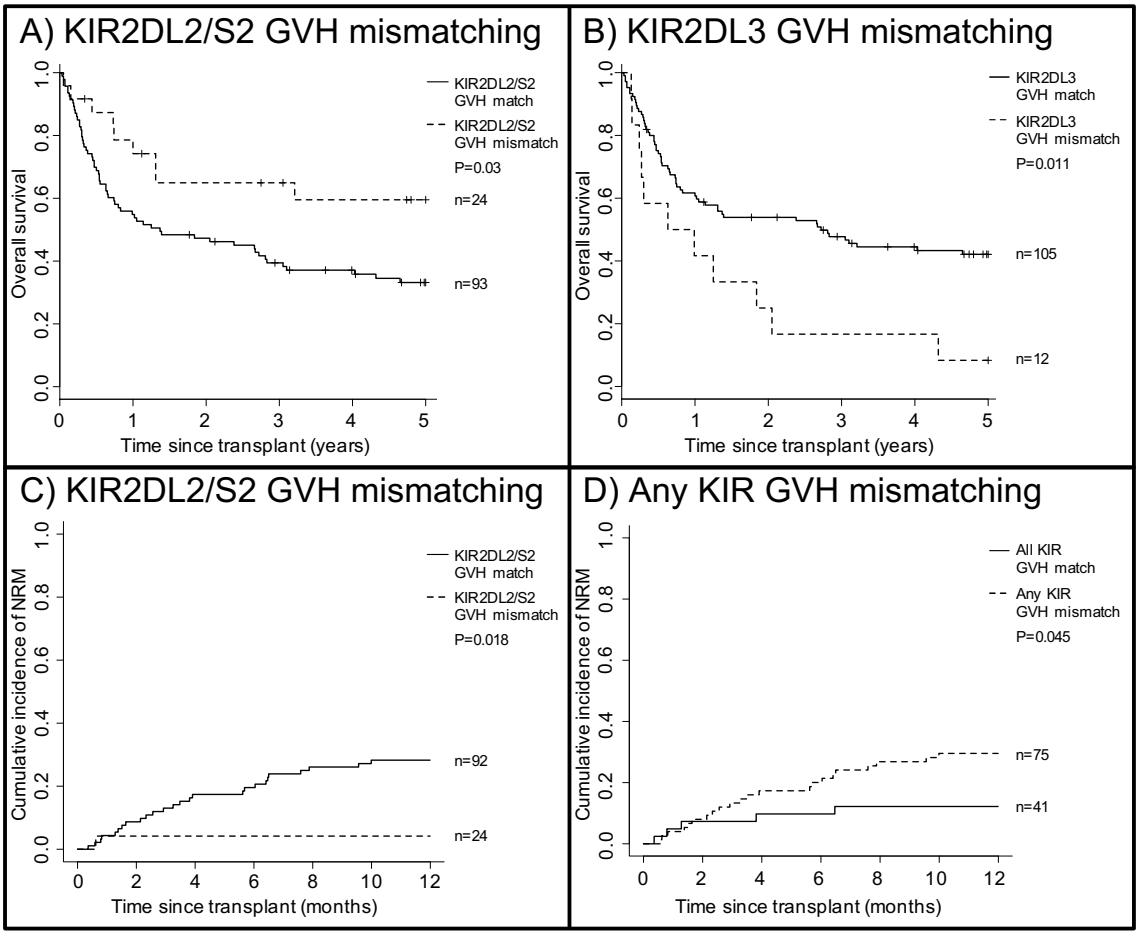


Figure 3.7 Probability of OS and NRM comparing GVH direction KIR gene mismatching

A) A Kaplan-Meier estimate of five year OS probability in transplants featuring KIR2DL2/S2 mismatching in the GVH direction demonstrates that mismatching for these loci is associated with improved OS. B) GVH direction mismatching has the opposite effect at the KIR2DL3 locus. C) The improved OS associated with KIR2DL2/S2 GVH mismatching is a result of decreased NRM at one year post-transplant. D) By contrast, when GVH KIR gene mismatches are considered from all the loci, a negative influence on one year NRM probability is observed.

The results from multivariate analyses of these factors are summarised in Table 3.30. Both KIR2DL2/S2 and KIR2DL3 GVH mismatching retain their significant effect on five year OS probability in a combined model, with each having an approximate doubling impact on the probability of OS (KIR2DL2/S2 mismatched, HR=0.49, CI=0.24-1.00, p=0.05; KIR2DL3 mismatched, HR=2.17, CI=1.12-4.17, p=0.021). When considering five year DFS probability, however, mismatching in the GVH

direction for these loci only depicts statistical trends with borderline significance (Table 3.30).

Table 3.30 Multivariate analysis assessing the impact of GVH direction KIR gene mismatching on five year OS and DFS probabilities

Variable	5 year OS				5 year DFS			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
Recipient age, years								
<40	84	1.00	-	-	83	1.00	-	-
>40	33	***1.75***	***1.06-2.94***	***0.028***	33	1.64	1.00-2.70	0.052
Previous autografts								
0	110	1.00	-	-	109	1.00	-	-
≥1	7	***3.45***	***1.43-8.33***	***0.006***	7	***2.86***	***1.20-6.67***	***0.017***
KIR2DL2/S2 GVH mismatching								
GVH Matched	93	1.00	-	-	92	1.00	-	-
GVH Mismatched	24	***0.49***	***0.24-1.00***	***0.050***	24	*0.52*	*0.26-1.03*	0.061
KIR2DL3 GVH mismatching								
GVH Matched	105	1.00	-	-	104	1.00	-	-
GVH Mismatched	12	***2.17***	***1.12-4.17***	***0.021***	12	*1.92*	*1.00-3.70*	0.053

Statistically significant results are denoted by ***bold italics***.

3.10.04.02 KIR gene mismatches in the graft-versus-host direction increase NRM

Although KIR2DL3 GVH mismatching was not associated with any other HCT outcomes, KIR2DL2/S2 GVH mismatching was prognostic of improved NRM risk at one year post-transplant (28% *vs* 4%, p=0.018, Figure 3.7C). In addition, a statistical trend was observed following the comparison of NRM at this timepoint between KIR2DS5 GVH direction matched and mismatched transplants, suggesting that those mismatched transplants had increased probability of NRM (p=0.059). In addition to these single gene mismatches, the presence of any mismatched KIR gene in the GVH direction also correlated with increased one year NRM risk (12% *vs* 30%, p=0.045, Figure 3.7D).

Due to the inclusion of KIR2DL2/S2 GVH mismatches within the "any GVH mismatches" variable, and because of their opposing influences, multivariate analysis featured a novel variable: transplants featuring GVH mismatches but excluding those transplants with KIR2DL2/S2 GVH mismatches. When considered in both univariate and multivariate analysis models, this was revealed as a significant predictor of OS at five years post-transplant (univariate analysis: 51% *vs* 23%, p=0.004, Figure 3.8A; multivariate analysis: HR=1.89, CI=1.10-3.23, p=0.021) and NRM at one year post-transplant (univariate analysis: 9% *vs* 41%, p<0.001, Figure 3.8B; multivariate analysis: HR=5.93, CI=1.72-20.47, p=0.005). Contrarily, presence of mismatches at the KIR2DL2/S2 loci in the GVH direction lost significance (Table 3.31).

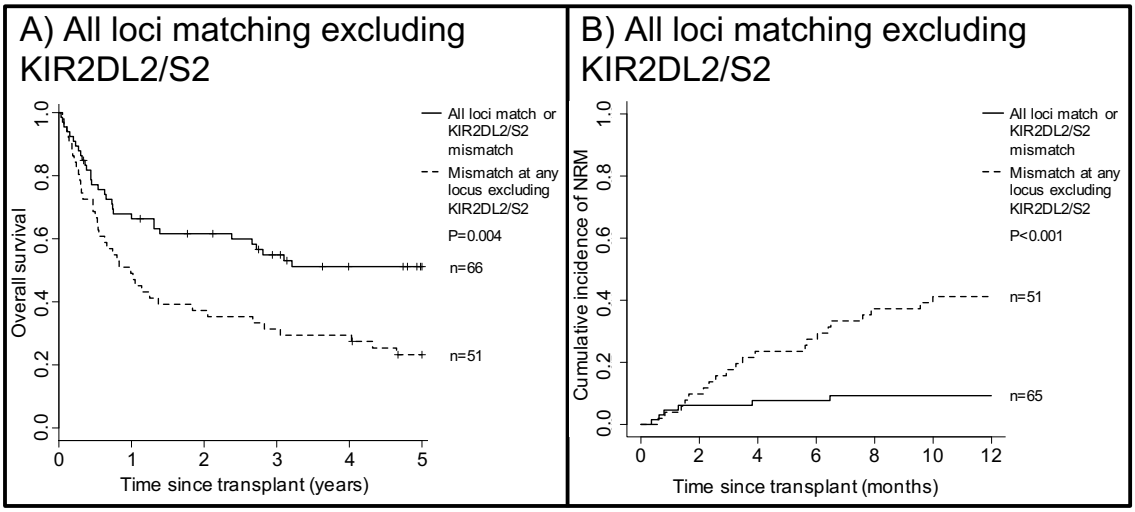


Figure 3.8 Probability of OS and NRM comparing a KIR mismatching model excluding KIR2DL2/S2 mismatches

The presence of GVH direction mismatching at any KIR locus, excluding transplants also featuring KIR2DL2/S2 GVH mismatch, leads to decreased five year OS (A) and increased one year NRM probabilities (B).

Table 3.31 Multivariate analysis assessing the impact of GVH direction KIR gene mismatching on five year OS and one year NRM probabilities

Variable	5 year OS				1 year NRM			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
Disease risk score[§]								
Good	-	-	-	-	50	1.00	-	-
Intermediate/Poor	-	-	-	-	65	0.55	0.24-1.25	0.15
Recipient age, years								
<40	84	1.00	-	-	83	1.00	-	-
>40	33	***1.69***	***1.01-2.78***	***0.045***	32	1.78	0.84-3.74	0.13
Previous autografts								
0	110	1.00	-	-	108	1.00	-	-
≥1	7	***4.00***	***1.64-10.00***	***0.002***	7	***8.39***	***1.98-35.63***	***0.004***
GVH mismatching at KIR2DL2/S2								
Absent	93	1.00	-	-	92	1.00	-	-
Present	24	0.63	0.29-1.37	0.24	23	0.49	0.05-5.07	0.55
GVH mismatching excluding KIR2DL2/S2 mismatches								
GVH Matched	66	1.00	-	-	64	1.00	-	-
GVH Mismatched	51	***1.89***	***1.10-3.23***	***0.021***	51	***5.93***	***1.72-20.47***	***0.005***

[§] Disease risk score grouping altered to permit multivariate analysis convergence.

Statistically significant results are denoted by ***bold italics***.

3.10.04.03 Activating KIR gene mismatching in the GVH direction confers benefits by reducing relapse risk

Although detrimental in terms of one year NRM risk, when investigating the probability of relapse at five years post-transplant, mismatching in the GVH direction for any KIR gene was responsible for a significantly improved relapse risk (matched *vs* mismatched, 45% *vs* 25%, p=0.034, Figure 3.9A). This observation may have been driven by mismatching between activating KIR genes, as improved relapse risk at this timepoint was also observed when comparing GVH mismatching at these loci alone (47% *vs* 27% p=0.037, Figure 3.9B), but not when comparing mismatching at the inhibitory loci (Table 3.29).

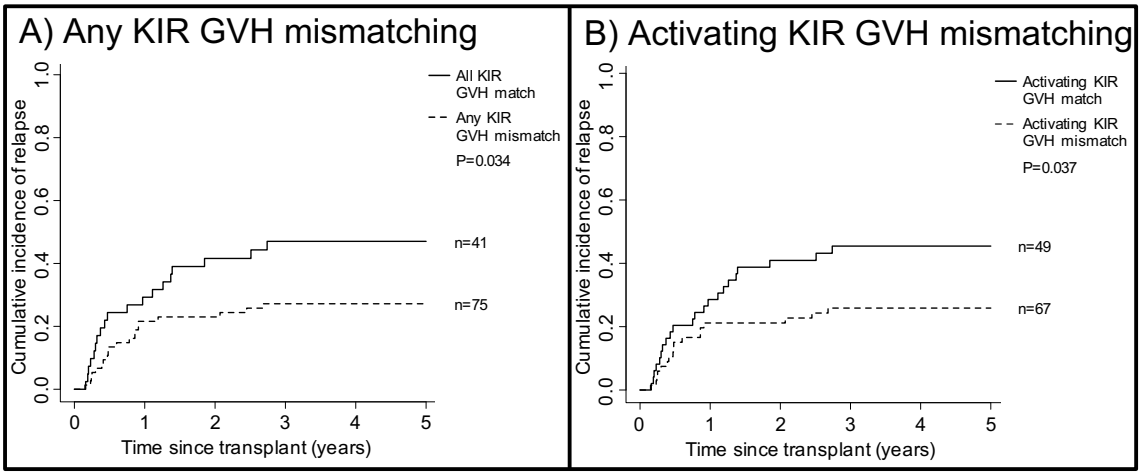


Figure 3.9 Probability of relapse comparing GVH direction KIR gene mismatching

A) Competing risk analysis reveals that five year relapse risk is affected by GVH mismatching. When present, GVH mismatching at any KIR locus results in decreased relapse risk at five years post-transplant. B) GVH mismatching at any activating KIR gene may have strong influence in this reduction in five year relapse risk.

Multivariate analysis considering the effects of GVH direction mismatching at any KIR locus on the probability of relapse at five years post-transplant confirmed the protective effect observed in univariate analysis (mismatched: HR=0.53, CI=0.28-0.99, p=0.047, Table 3.32). To investigate the influence of GVH direction mismatching at activating KIR loci in a multivariate model, a new variable was created to adjust for transplants featuring GVH-direction mismatches not in activating KIR genes. This analysis provided evidence that GVH mismatches at activating, not inhibitory, KIR loci are responsible for the decrease in five year relapse risk (mismatched: HR=0.49, CI=0.25-0.96, p=0.037, Table 3.33).

Table 3.32 Multivariate analysis assessing the impact of GVH direction KIR gene mismatching on five year relapse risk

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Disease risk score				
Good	50	1.00	-	-
Intermediate	51	1.33	0.65-2.73	0.43
Poor	14	***4.42***	***1.86-10.52***	***<0.001***
GVH mismatching at any locus				
GVH Matched	41	1.00	-	-
GVH Mismatched	74	***0.53***	***0.28-0.99***	***0.047***

Statistically significant results are denoted by ***bold italics***.

Table 3.33 Multivariate analysis assessing the impact of GVH direction activating KIR gene mismatching on five year relapse risk

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Disease risk score				
Good	50	1.00	-	-
Intermediate	51	1.36	0.66-2.81	0.41
Poor	14	***4.68***	***1.91-11.50***	***<0.001***
GVH mismatching excluding activating KIR loci mismatches				
Absent	107	1.00	-	-
Present	8	0.91	0.30-2.72	0.87
GVH mismatching at activating KIR locus only				
GVH Matched	49	1.00	-	-
GVH Mismatched	66	***0.49***	***0.25-0.96***	***0.037***

Statistically significant results are denoted by ***bold italics***.

3.10.05 The donor KIR B content model

3.10.05.01 Donor Cen-B motifs decrease five year OS and DFS probabilities

Analysis of the impact of donor-encoded KIR B content highlighted some interesting and unexpected results. In contradiction to previous findings, five year OS probability was improved by the absence of donor KIR B (49% *vs* 34%, p=0.06, Table 3.34). This influence became statistically significant only when considering the Cen-B motif (47% *vs* 31%, p=0.024, Figure 3.10A): no significant differences were noted between HCT outcomes when assessing donors' Tel-B content or "better/best" score. Accordingly, the B motif content of the centromeric portion of the haplotypes was further stratified,

revealing an apparent dose-dependent effect on OS probability at five years (Cen-AA *vs* Cen-AB *vs* Cen-BB, 48% *vs* 37% *vs* 8%, p=0.01, Figure 3.10B). In a multivariate analysis adjusting for recipient age group and history of previous autografts, donor Cen-BX genotype remained a significant risk factor for decreased OS (HR=1.89, CI=1.16-3.13, p=0.01, Table 3.35).

Table 3.34 Univariate analysis p-values of HCT outcomes based on donor KIR B content models in MAC, adult, HLA-matched sub-cohort

KIR B content model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD
KIR AA *vs* BX	0.060	0.087	0.60	***0.019***	0.35
Cen-AA *vs* BX	***0.024***	***0.045***	0.45	***0.001***	0.73
Tel-AA *vs* BX	0.42	0.47	0.77	0.13	0.90
Neutral *vs* Better/Best	0.57	0.66	0.77	0.83	0.61
Cen-AA *vs* AB *vs* BB	***0.010***	***0.031***	0.75	***0.005***	0.38

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table E.

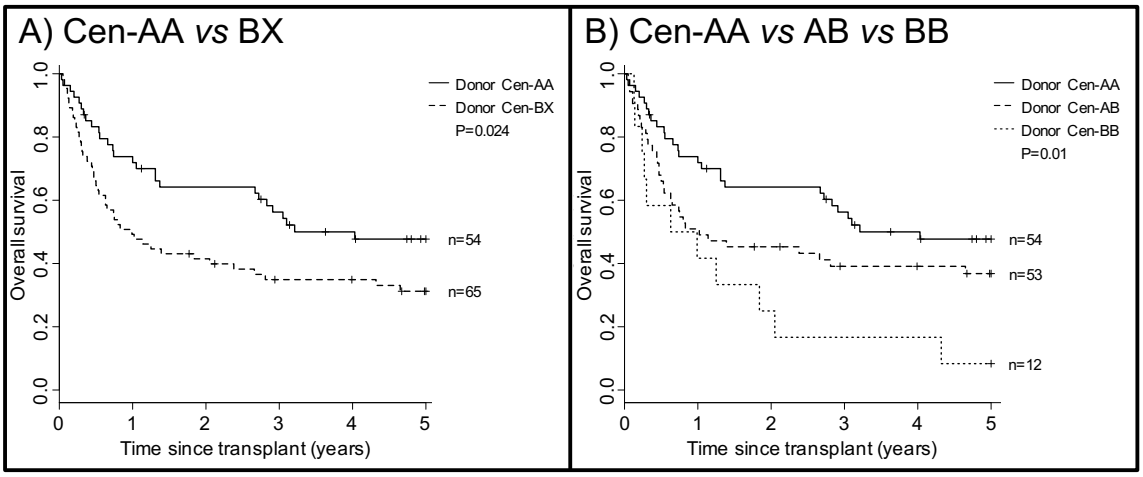


Figure 3.10 Probability of OS comparing donor centromeric haplotype motif content

A) Univariate analysis demonstrates that the presence of donor-encoded Cen-B motifs is detrimental to five year OS probability. B) When further stratified by Cen-B motif copy number, a dose effect is observed with each additional Cen-B motif.

The impact of donor Cen motif variation was preserved when assessing five year DFS (Cen-AA *vs* BX, 45% *vs* 29%, p=0.045; Cen-AA *vs* AB *vs* BB, 45% *vs* 34% *vs* 8%,

p=0.031, Table 3.34), which also persisted in multivariate analysis (Cen-BX, HR=1.75, CI=1.09-2.78, p=0.022, Table 3.35).

Table 3.35 Multivariate analysis assessing the impact of donor Cen-BX haplotype motif structure on five year OS and DFS

Variable	5 year OS				5 year DFS			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
Recipient age, years								
<40	85	1.00	-	-	84	1.00	-	-
>40	34	***1.92***	***1.15-3.13***	***0.012***	34	***1.75***	***1.06-2.94***	***0.027***
Previous autografts								
0	112	1.00	-	-	111	1.00	-	-
≥1	7	***3.03***	***1.30-7.14***	***0.010***	7	***2.63***	***1.12-6.25***	***0.026***
Donor Cen motif pattern								
Cen-AA	54	1.00	-	-	54	1.00	-	-
Cen-BX	65	***1.89***	***1.16-3.13***	***0.010***	64	***1.75***	***1.09-2.78***	***0.022***

Statistically significant results are denoted by ***bold italics***.

3.10.05.02 Donor Cen-B motifs increase NRM risk at one year post-transplant

When investigating the probable factors contributing to mortality, neither relapse risk nor aGVHD differed significantly according to donor Cen-B content (Table 3.34). However, NRM risk at one year was significantly different when comparing donor KIR AA to BX genotypes (9% *vs* 29%, p=0.019, Figure 3.11A) and when comparing donor Cen-AA to Cen-BX motifs (9% *vs* 34%, p=0.001, Figure 3.11B). Statistical significance was maintained in a multivariate analysis adjusting for recipient age group and previous autografting history (Cen-BX, HR=4.16, CI=1.58-11.00, p=0.004, Table 3.36). When the centromeric portion of the haplotype was stratified further, the copy number of donor-encoded Cen-B motifs also appeared to have an additive effect on the increase in NRM risk at one year post-transplant (9% *vs* 33% *vs* 42%, p=0.005, Figure 3.11C), a factor that also persisted in multivariate analysis (Table 3.37).

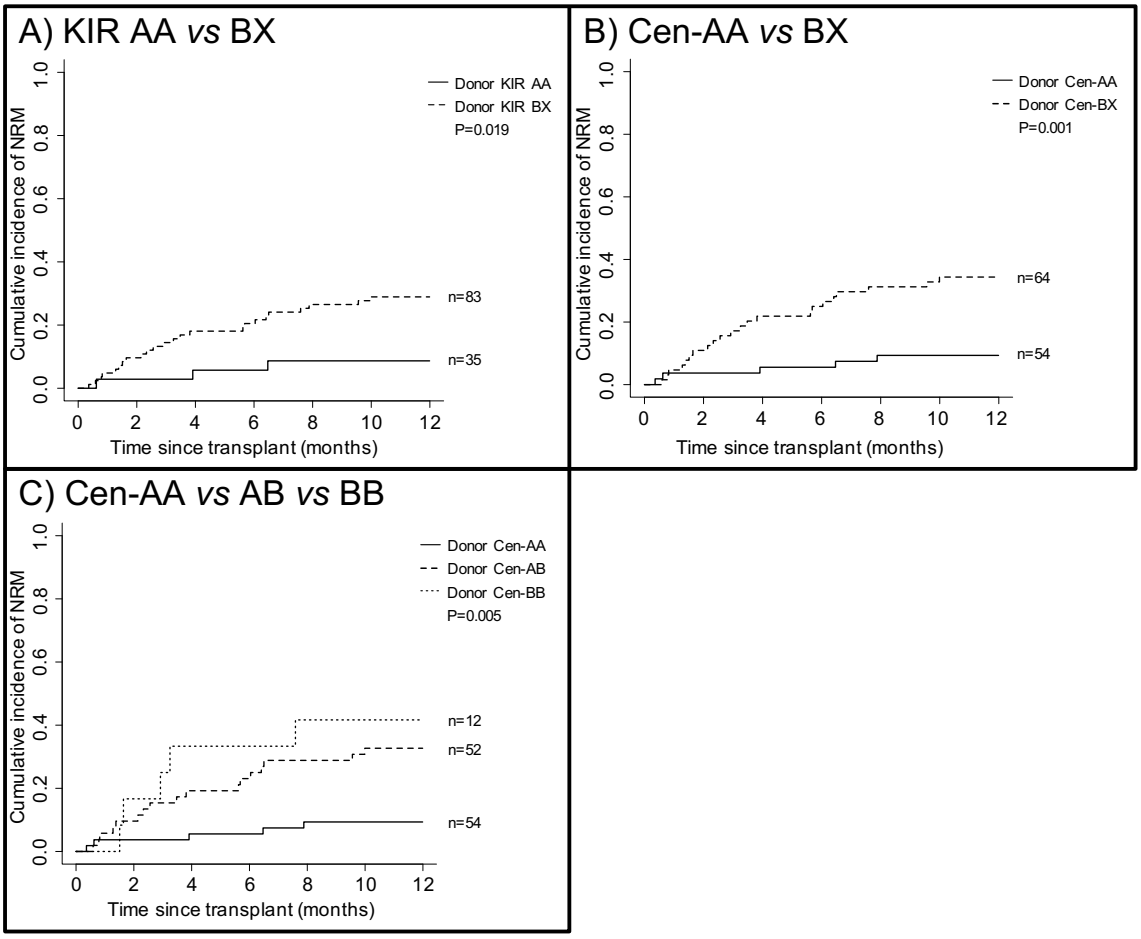


Figure 3.11 Probability of NRM comparing donor B haplotype motif content

A) Univariate analysis demonstrates that the presence of at least one donor-encoded KIR B motif is detrimental to risk of NRM at one year post-transplant. B) When stratified by Cen-B motif, it is evident that the centromeric region of the haplotype is responsible for the increased risk of NRM at this timepoint. C) When the centromeric motif is further stratified, by Cen-B copy number, a dose effect is observed.

Table 3.36 Multivariate analysis assessing the impact of donor Cen-BX haplotype motif structure on one year NRM and infectious mortality risk

Variable	1 year NRM				1 year infectious mortality			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
Recipient age, years								
<40	84	1.00	-	-	83	1.00	-	-
>40	34	1.81	0.82-4.01	0.15	32	2.28	0.91-5.69	0.078
Era								
1996-1999	-	-	-	-	14	1.00	-	-
2000-2003	-	-	-	-	43	1.15	0.15-8.99	0.89
2004-2007	-	-	-	-	38	5.27	0.84-32.90	0.075
2008-2011	-	-	-	-	20	0.73	0.05-9.93	0.82
Previous autografts								
0	111	1.00	-	-	-	-	-	-
≥1	7	2.45	0.55-10.92	0.24	-	-	-	-
Donor Cen motif pattern								
Cen-AA	54	1.00	-	-	54	1.00	-	-
Cen-BX	64	***4.16***	***1.58-11.00***	***0.004***	61	***5.50***	***1.49-20.32***	***0.011***

Statistically significant results are denoted by ***bold italics***.

3.10.05.03 Infection is the predominant cause of NRM

To determine the causes of death responsible for the increase in non-relapse mortality observed in the transplants utilising donors encoding Cen-B, deaths by infection, GVHD or other causes were analysed. Of the 27 deaths by NRM in the first year post-transplant, cause of death data was recorded for 24 (89%). Of these, one patient (4%) died with no evidence of infection or GVHD, one patient's (4%) cause of death was recorded as both GVHD and infection, four patients (17%) died of GVHD only, whilst the final 18 patients (75%) died of infection only. Of the patients whose cause of death included infection, there were 15 patients where the type of infection was also available. In 13 (87%) of these cases, viral infection was included as the cause of death.

Table 3.37 Multivariate analysis assessing the impact of donor Cen-B haplotype motif copy number on one year NRM and infectious mortality risk

Variable	1 year NRM				1 year infectious mortality			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
Recipient age, years								
<40	84	1.00	-	-	83	1.00	-	-
>40	34	1.81	0.82-3.99	0.14	32	2.26	0.88-5.78	0.090
Era								
1996-1999	-	-	-	-	14	1.00	-	-
2000-2003	-	-	-	-	43	1.15	0.15-9.02	0.89
2004-2007	-	-	-	-	38	5.21	0.84-32.45	0.077
2008-2011	-	-	-	-	20	0.74	0.05-9.94	0.82
Previous autografts								
0	111	1.00	-	-	-	-	-	-
≥1	7	2.63	0.58-11.89	0.21	-	-	-	-
Donor Cen motif pattern								
Cen-AA	54	1.00	-	-	54	1.00	-	-
Cen-AB	52	***3.85***	***1.40-10.57***	***0.009***	51	***5.38***	***1.41-20.62***	***0.014***
Cen-BB	12	***5.79***	***1.79-18.69***	***0.003***	10	***5.96***	***1.05-33.92***	***0.044***

Statistically significant results are denoted by ***bold italics***.

In a competing risk analysis of infectious mortality, whereby relapse or death by any non-infectious means were the competing risks, presence of at least one KIR B or Cen-B motif in the donor genotype was observed as a strong risk factor by one year post-transplant (KIR AA *vs* KIR BX, 3% *vs* 21%, p=0.014, Figure 3.12A; Cen-AA *vs* Cen-BX, 6% *vs* 25%, p=0.005, Figure 3.12B). Similarly to the analyses above, an additive effect was observed with each additional Cen-B motif encoded by the donor (Cen-AA *vs* Cen-AB *vs* Cen-BB: 6% *vs* 24% *vs* 30%, p=0.018, Figure 3.12C). Although the statistical significance was downgraded to a statistical trend upon multivariate analysis of KIR AA *vs* KIR BX (HR=6.39, CI=0.87-47.1, p=0.069), both stratifications of the centromeric motif resulted in statistically significant differences between the groups (Cen-BX, HR=5.5, CI=1.49-20.32, p=0.011; Cen-AB, HR=5.38, CI=1.41-20.62, p=0.014; Cen-BB, HR=5.96, CI=1.05-33.92, p=0.044; Table 3.36 and Table 3.37).

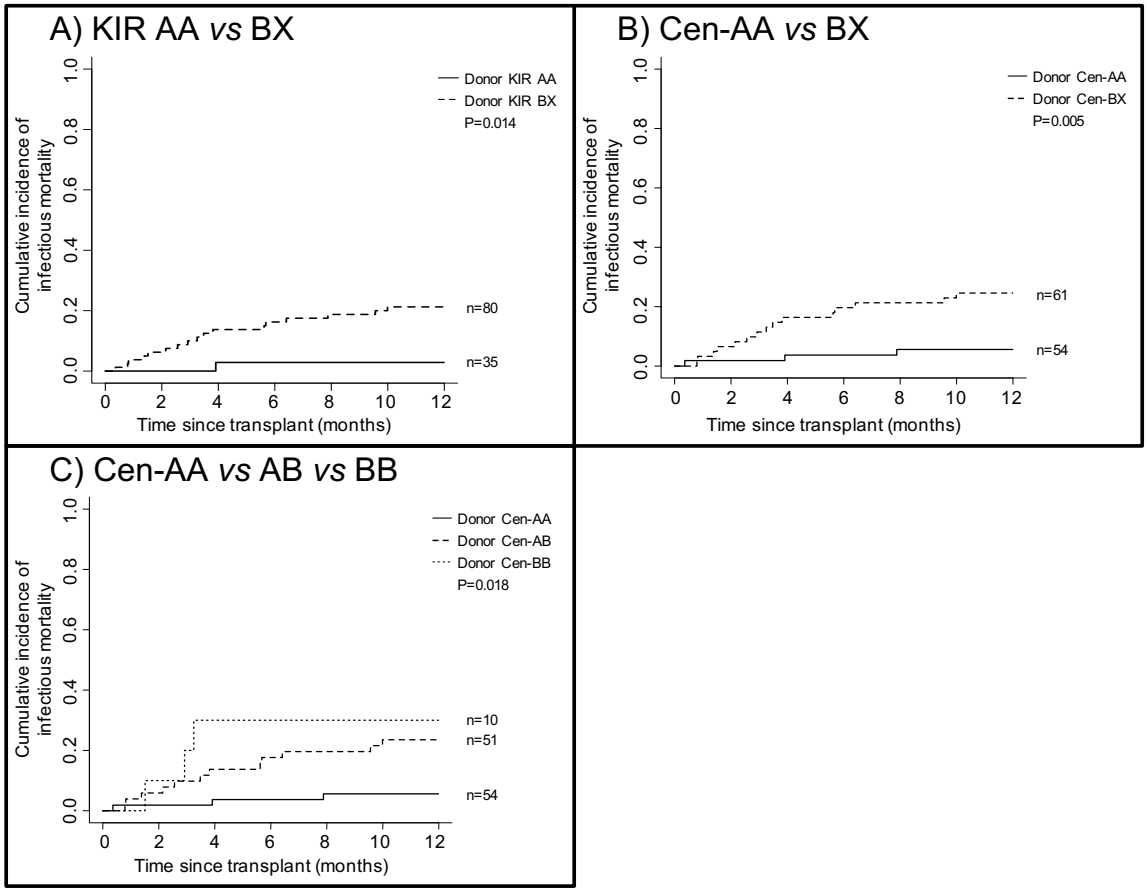


Figure 3.12 Probability of infectious mortality comparing donor B haplotype motif content

A) Univariate analysis demonstrates that the presence of at least one donor-encoded KIR B motif is detrimental to risk of infectious mortality at one year post-transplant. B) When stratified by Cen-B motif, it is evident that the centromeric region of the haplotype is associated with the increased risk of death by infection. C) When the centromeric motif is further stratified, by Cen-B copy number, a dose effect is observed.

3.11 Discussion

Genetic polymorphism at the KIR and corresponding HLA ligand loci has been implicated in many models of alloreactivity following VUD HCT. The original model of alloreactivity – based on missing self – was proposed following investigation in haploidentical HCT outcomes [259]. However, the high frequency of HLA matching in this cohort limited the occurrence of missing self. Indeed, only 30 transplants utilised KIR ligand mismatched donors. Of these, only 15 were mismatched in a donor KIR-

incompatible manner. However, an adaptation of this theory, the missing ligand hypothesis, as well as investigation into the influence of KIR receptor matching and donor KIR B haplotype motif content, were all explored in this novel UK AML cohort. The published findings from some of the many investigations of these different models are compared against key findings from my study in Table 3.38, Table 3.39 and Table 3.40.

That widely heterogeneous transplant practices exist both between and within individual centres, even when treating the same underlying disease, may explain the manifestation of multiple models to assess the influence of KIR on transplant outcomes. Conditioning regimen, HLA-matching, HSC source, TCD, donor source, HSC dose and a host of other factors all contribute to this heterogenous pool of transplants. As some of these factors can have dramatic effects on the immune reconstitution, risk of relapse or incidence of GVHD in recipients, what constitutes an effective KIR genotype strategy in one scenario may well be ineffective or detrimental in another. The inability to use the donor missing self theory in well-matched VUD transplants is a prime example of this. As such, the establishment of baseline characteristics within this novel UK AML cohort forms an important part of this thesis, especially as UK HCT practice is unique in its utilisation of alemtuzumab as the predominant TCD agent and represents a regional HCT population yet to be assessed for functionally relevant KIR polymorphism [297].

Table 3.38 Comparison of the results published on KIR missing/mismatched ligand models and their influence on HCT outcomes

Study	Donor source	Recipient diagnosis	TCD	Conditioning	OS	DFS	Relapse	NRM	aGVHD
Anthony Nolan Total	VUD	AML	Yes	Mixed	NS	NS	NS	NS	Reduced by KIR3DL1 missing ligand
Anthony Nolan Adult, HLA-M, RIC	VUD	AML	Yes	RIC	NS	NS	NS	NS	NS
Anthony Nolan Adult, HLA-M, MAC	VUD	AML	Yes	MAC	NS	NS	Increased by KIR3DL1 missing ligand	NS	NS
Ruggeri *et al.* 2002 [259]	Haplo	AML	Yes	-	-	-	Reduced by mismatched ligands	-	Reduced by mismatched ligands
Giebel *et al.* 2003 [303]	VUD	Mixed	Yes	MAC	Increased	Increased	Reduced	Reduced	NS
Leung *et al.* 2004 [266]	Haplo	Mixed	Yes	MAC	-	-	Reduced	-	-
Schaffer *et al.* 2004 [304]	VUD	Mixed	Yes	MAC	Reduced	-	NS	Increased	NS
Verheyden *et al.* 2005 [305]	10/10 related	AML CML ALL	Mixed	MAC	NS	-	NS	NS	Increased
Hsu *et al.* 2005 [306]	10/10 related	AML MDS	Yes	MAC	Increased	Increased	Reduced	-	-
Giebel *et al.* 2005 [307]	VUD	Mixed	Yes	MAC	Reduced by missing KIR2DL2/3 ligand	Reduced by missing KIR2DL2/3 ligand	Increased by missing KIR2DL2/3 ligand	NS	NS
Hsu *et al.* 2006 [308]	VUD	Mixed	No	MAC	-	-	Reduced	-	-
Kröger *et al.* 2006 [278]	VUD	AML	Yes	MAC	Reduced	Reduced	NS	Increased	-
Farag *et al.* 2006 [309]	VUD	AML MDS CML	Mixed	MAC	Reduced by mismatched ligands (*vs* HLA matched)	Reduced by mismatched ligands (*vs* HLA matched))	NS	Increased by mismatched ligands (*vs* HLA matched)	Severe grade aGVHD increased by mismatched ligands
Clausen *et al.* 2007 [310]	10/10 related	Mixed	No	Mixed	NS	Increased by missing ligand (with high NK cell dose)	Reduced	-	NS

Table 3.38 (continued)

Study	Donor source	Recipient diagnosis	TCD	Conditioning	OS	DFS	Relapse	NRM	aGVHD
Miller *et al.* 2007 [260]	VUD	AML CML MDS	No	-	NS	-	Reduced by missing ligands (early stage disease only)	-	Severe grade aGVHD increased by missing ligand
Sobecks *et al.* 2007 [311]	10/10 related	AML	No	MAC	Increased	Increased	Reduced	-	NS
Brunstein *et al.* 2009 [265]	CBU	Mixed	Mixed	RIC	Reduced	-	NS	Increased	Increased
Bjorklund *et al.* 2010 [312]	10/10 related	AML MDS	Mixed	Mixed	NS	NS	NS	NS	-
Clausen *et al.* 2010 [313]	10/10 related	Mixed	No	Mixed	Reduced	Reduced	Increased	NS	Increased
Venstrom *et al.* 2012 [300]	VUD	AML	Mixed	Mixed	Increased by KIR2DS1+ve HLA-C1+ve	Increased by KIR2DS1+ve HLA-C1+ve	Increased when KIR2DS1+ve and HLA-C2C2	-	NS
Marra *et al.* 2015 [271]	Auto	AML	-	MAC	-	-	Increased by missing/strong ligand	NS	-
Sobecks *et al.* 2015 [314]	VUD	AML MDS	Mixed	RIC	NS	NS	Reduced by KIR3DL1 missing ligand	Reduced by missing KIR2DS1 ligand	Increased by missing ligands (for KIR2DL1 or multiple KIR)
Faridi *et al.* 2016 [284]	VUD	Mixed	Yes	MAC	NS	Increased	Reduced	-	NS
Shimoni *et al.* 2017 [315]	Allo	AML MDS	Mixed	RTC	-	Increased by missing KIR2DL2/3 ligand	Reduced by missing KIR2DL2/3 ligand	NS	NS
Martinez-Losada *et al.* 2017 [267]	CBU	AML ALL	Yes	MAC	NS	NS	Reduced	NS	Reduced
Hoff *et al.* 2017 [296]	VUD	Mixed	Mixed	Mixed	NS	NS	NS	NS	NS
Neuchel et al, 2017 [302]	VUD	Mixed	No	Mixed	Reduced by missing KIR2DL2/3 ligand	Reduced by missing KIR2DL2/3 ligand	Increased by missing KIR2DL2/3 ligand	Increased by missing KIR2DL2/3 ligand	-

VUD=Volunteer unrelated donor. Haplo=haploidentical HCT. Auto=autograft HCT. Allo=allogeneic HCT. CBU=cord blood unit HCT. AML=Acute myeloid leukaemia. CML=Chronic myeloid leukaemia. ALL=Acute lymphoblastic leukaemia. MDS=myelodysplasia. TCD=T cell depletion. NS=Not significant. RTC=Reduced-toxicity myeloablative conditioning. Mixed status if highest frequency group <90% of transplants.

Table 3.39 Comparison of the results published on KIR matching models and their influence on HCT outcomes

Study	Donor source	Recipient diagnosis	TCD	Conditioning	OS	DFS	Relapse	NRM	aGVHD
Anthony Nolan Total	VUD	AML	Yes	Mixed	Increased by GVH KIR2DL1 matching	Increased by GVH KIR2DL1 matching	Reduced by KIR2DL1 matching	NS	NS
Anthony Nolan RIC	VUD	AML	Yes	RIC	NS	NS	NS	Increased by KIR2DS2 mismatching	NS
Anthony Nolan MAC	VUD	AML	Yes	MAC	Increased by GVH matching at all loci excluding KIR2DL2/S2	Increased by GVH matching at all loci excluding KIR2DL2/S2	Increased by GVH matching at activating KIR loci	Reduced by GVH matching at any locus excluding KIR2DL2/S2	NS
Gagne *et al.* 2002 [282]	VUD	AML CML ALL	-	-	-	-	-	-	Reduced by GVH KIR matching
Gagne *et al.* 2002 [282]	10/10 related	Mixed	-	-	-	-	-	-	Increased by GVH KIR matching
Verheyden *et al.* 2005 [305]	10/10 related	AML CML ALL	Mixed	MAC	-	-	NS	NS	NS
Symons *et al.* 2010 [279]	Haplo	Mixed	No	NMA	Reduced by inhibitory KIR matching	-	Increased by inhibitory KIR matching	NS	NS
Bao *et al.* 2016 [316]	VUD	AML MDS	Yes	MAC	NS	NS	-	NS	-
Faridi *et al.* 2016 [284]	10/10 allo	Mixed	Yes	MAC	NS	NS	NS	-	Reduced by KIR matching
Sahin *et al.* 2018 [283]	10/10 related	AML CML	No	Mixed	NS	Increased by activating KIR matching	Reduced by activating KIR matching	NS	cGVHD reduced by KIR matching (activating or inhibitory)

NMA=Non-myeloablative

Table 3.40 Comparison of the results published on donor KIR B content/activating KIR models and their influence on HCT outcomes

Study	Donor source	Recipient diagnosis	TCD	Conditioning	OS	DFS	Relapse	NRM	aGVHD
Anthony Nolan Total	VUD	AML	Yes	Mixed	NS	NS	NS	NS	NS
Anthony Nolan RIC	VUD	AML	Yes	RIC	NS	NS	NS	Reduced by donor Cen-B	NS
Anthony Nolan MAC	VUD	AML	Yes	MAC	Reduced by donor Cen-B	Reduced by donor Cen-B	NS	Increased by donor Cen-B	NS
Verheyden *et al.* 2005 [305]	10/10 related	AML CML ALL	Mixed	MAC	NS	-	Reduced by donor KIR2DS1+ve, KIR2DS2+ve	NS	NS
Kröger *et al.* 2006 [278]	VUD	AML	Yes	MAC	Reduced by donor KIR BX	Reduced by donor KIR BX	Increased by donor KIR BX	NS	NS
Clausen *et al.* 2007 [310]	10/10 related	Mixed	No	Mixed	Reduced by multiple donor activating KIR	-	-	Increased by multiple donor activating KIR	-
Cooley *et al.* 2009 [83]	VUD	AML	No	MAC	Increased by donor KIR BX	Increased by donor KIR BX	Reduced by donor KIR BX	Reduced by donor KIR BX	NS (but cGVHD increased by donor KIR BX)
Clausen *et al.* 2010 [313]	10/10 related	Mixed	No	Mixed	NS	NS	NS	NS	NS
Venstrom *et al.* 2010 [317]	VUD	Mixed	Mixed	MAC	Increased by donor KIR3DS1	-	NS	Reduced by donor KIR3DS1	Reduced by donor KIR3DS1
Stringaris *et al.* 2010 [318]	10/10 related	AML	Yes	MAC	Increased by donor Tel-BX	-	Reduced by donor Tel-BX	NS	-
Cooley *et al.* 2010 [273]	VUD	AML	No	MAC	Increased by donor Cen-BX	Increased by donor Cen-BX	Reduced by donor Cen-BX	NS	NS
Symons *et al.* 2010 [279]	Haplo	Mixed	No	NMA	Increased by donor KIR BX (KIR AA recipients only)	-	NS	Reduced by donor KIR BX (KIR AA recipients only)	NS
Venstrom *et al.* 2012 [300]	VUD	AML	Mixed	Mixed	Increased by donor KIR2DS1	Increased by donor KIR2DS1	Reduced by donor KIR2DS1	-	NS
Cooley *et al.* 2014 [272]	VUD	AML	No	MAC	Increased by "better/best" donor (C1+ve patients)	Increased by "better/best" donor (C1+ve patients	Reduced by "better/best" donor (C1+ve patients	NS	NS
Impola *et al.* 2014 [275]	10 /10 related	Mixed	No	Mixed	-	Increased by donor Cen-BX	Reduced by donor Cen-BX	-	-
Sobecks *et al.* 2015 [314]	VUD	AML MDS	Mixed	RIC	NS	NS	NS	NS	NS
Bao *et al.* 2016 [316]	VUD	AML MDS	Yes	MAC	Increased by donor KIR BX	Increased by donor KIR BX	NS	Reduced by donor KIR BX	NS
Hosokai *et al.* 2017 [319]	Allo	Mixed	No	Mixed	NS	NS	NS	NS	Increased by donor KIR BX (HLA-mismatched only)
Neuchel *et al.* 2017 [302]	VUD	Mixed	No	Mixed	Increased by donor Cen-BX (C2C2 recipients)	Increased by donor Cen-BX (C2C2 recipients)	Decreased by donor Tel-BX (C2C2 recipients)	Increased by donor Tel-BX (C2C2 recipients)	-
Heatley *et al.* 2018 [274]	10/10 related	AML	No	Mixed	Increased by donor Cen-BX	NS	NS	Reduced by donor Cen-BX	Reduced by donor Cen-BX
Escudero *et al.* 2018 [280]	10/10 related	AML ALL	Yes	RIC	Reduced by donor Cen-B	Reduced by "better/best" donor	Increased by donor Tel-B	-	-
Sahin *et al.* 2018 [283]	10/10 related	AML CML	No	Mixed	NS	NS	NS	Reduced by donor KIR BX	NS (but cGVHD increased by donor KIR BX)

When previously determined models were tested on the total cohort of AML patients included in my study, several interesting results were observed. First, very few significant findings were observed from the total cohort. Nonetheless, the incidence of grades 2-4 aGVHD were decreased in the presence of missing KIR3DL1 ligand, supportive of the theory proposed by Ruggeri *et al.* (2002) [259], but contradictory to a similar study in a T cell replete cohort [260], suggestive that TCD may be required for this effect to be observed. Interestingly the decrease in aGVHD was not associated with an appreciable difference in NRM risk at one year post-transplant. This could indicate that, although better at reducing GVHD, a missing ligand scenario may also represent a reduced ability for NK cells to recognize and kill infected cells and, as such, the net effect on the risk of NRM remains similar. No additive effect resulting from HLA-Bw4 dimorphism at residue 80 was observed, although this may result from this polymorphism not being the sole factor responsible for differential KIR3DL1 interaction strength [298].

There was no obvious difference observed for any HCT outcomes when evaluating missing HLA-C ligand scenarios. Although contradicting several studies that have observed detrimental outcomes in recipients with HLA-C1^{-ve} genotypes [302,307], my observations are in agreement with those from the recent, very large cohort study performed by Hoff *et al.* (2017) [296], where no effect of recipient HLA-C genotype was observed.

With the exception of a beneficial effect observed from a reduction in five year relapse risk resulting from a small subgroup of KIR2DL1 mismatched transplants, no other KIR-mediated differences in HCT outcomes were observed in the total cohort.

These findings may appear to suggest that the role of KIR genotype polymorphism is very limited. Indeed, several of studies on which these tests were based identified that only particular subsets of their cohort were responsible for the effect observed. However, even when equivalent subsets from this cohort were assessed, no significance was found. This may suggest that these models are not well suited to analysing this particular dataset. Alternatively, given the retrospective, multi-centre aspect of this study, the subgroup categories tested may still have represented a fairly heterogeneous group of transplants. As such, I performed analysis on refined sub-cohorts suited to this dataset. All paediatric cases were dismissed, as paediatric AML phenotypes are known to differ from the adult equivalents [320,321]. In addition, to remove the variation known to be caused by HLA mismatching, all HLA mismatched transplants were removed [205]. Although limiting the cohort size, removal of these factors was performed to maximise homogeneity. Selection of transplants known to have undergone TCD would have also been a desirable factor however, the incomplete nature of this variable would have resulted in unnecessary loss of transplants. Known T cell replete transplants make up a small minority (5%) of the adult, HLA-matched transplant pool in this study. One further stratification was made, separating the adult, HLA-matched sub-group according to conditioning regimen. As a key determinant of immune reconstitution and cause of subtle alteration in HCT outcomes, conditioning regimen forms an important difference in HCT protocol [322-324].

The results from these two subsets revealed key KIR-mediated influences on HCT outcomes. In the RIC subset, increased NRM risk at one year post-transplant was observed as a result of mismatching KIR2DS2, although this finding may simply reflect

KIR2DS2 acting as a minor histocompatibility antigen or may be a false-positive result associated with multiplicity of testing. A more striking observation, however, was the difference when assessing the influence of KIR B haplotype motifs. Previous publications on this factor have focussed on MAC transplants but, in a large cohort study performed by Sobecks *et al.* (2015) [314] investigating the influence in VUD, RIC transplants, no significant differences were observed. The analysis presented here, however, shows a highly significant reduction in NRM risk at one year post-transplant associated with the presence of donor-encoded Cen-B haplotype motifs. This was not correlated with a difference in any other outcome assessed and relapse, oft-reported by previous studies as the affected outcome, was near identical between transplants from donors with Cen-AA or Cen-BX haplotype motif structures.

Similar observations were made by Sahin *et al.* (2018) [283], who reported a donor KIR B motif-associated reduction in NRM risk, despite a significant increase in cGVHD probability. Although no discernible difference in the incidence of grades 2-4 aGVHD was observed in my analysis, a previous study in a T cell replete, HLA-identical, related donor setting explained a beneficial decrease in NRM risk associated with donor Cen B motif content with a reduction in aGVHD that corresponded with improved OS [274]. Because of the small size of sub-cohort, and the relative infrequency of NRM following RIC HCT, we were unable to investigate the cause of death in these patients. In an alternative study which utilised conditions similar to this sub-cohort of my analysis; TCD, RIC preparative regimen and HLA-matched donors (albeit in a related donor HCT setting that included some ALL patients), NRM was unaffected by donor KIR. Instead, donor-encoded KIR B motifs were associated with increased relapse, resulting in detrimental effects on OS and DFS [280].

As the findings from my RIC, TCD sub-cohort align better with findings from previous studies using MAC and T cell replete transplant protocols, it may seem unlikely that either of these factors influences the effect of donor KIR B content. However, the results from the adult, HLA-matched RIC sub-cohort directly contrast with the findings from the equivalent MAC sub-group of the overall cohort. In the MAC setting, donor-encoded Cen-B haplotype motifs increased the risk of NRM at one year post-transplant. In addition, the increased NRM risk associated with reduction in both five year OS and DFS probabilities. One possible explanation may relate to differences in toxicity that result following the different intensity conditioning regimens. MAC regimens are highly toxic, resulting in greater inflammation and a greater risk of NRM, particularly in elderly patients [325]. In recipients of grafts from donors with increased proportion of KIR A motifs, this increased inflammation may provide a stimulus responsible for activating NK cells otherwise inhibited by the high proportion of inhibitory KIR genes. Conversely, the response of NK cells from donors with increasing frequency of KIR B haplotype motifs may be symbolised by the hyporesponsiveness that is characteristic of education via activating KIR.

The detrimental effect of donor KIR B motifs is observed much less frequently in the current literature. Indeed, studies evaluating the influence of KIR B motifs encoded by both VUDs and HLA-identical, related donors in MAC, TCD cohorts have observed reduced rates of relapse or NRM and corresponding benefits in OS probability for AML recipients [316,318]. Studies in T cell replete cohorts also associate improved outcomes with donor-encoded KIR B motifs [83,272,273,302]. Interestingly, the findings of Bao *et al.* (2016) [316] revealed that although the overall donor KIR BX haplotype structure

conferred a benefit, the presence of the "BX1" genotype, characterised by the KIR3DL3, KIR2DL3, KIR2DP1, KIR2DL1, KIR3DP1, KIR2DL4, KIR3DL1, KIR3DS1, KIR2DL5A, KIR2DS5, KIR2DS1, KIR2DS4 and KIR3DL2 combination genotype, was associated with decreased OS and DFS. The frequency of this particular donor KIR genotype was too low to perform a meaningful analysis (n=17) and, as such, not likely to be responsible for the negative impact observed.

Instead, my findings more closely resemble those from earlier studies into the impact of donor-encoded activating KIR frequency. Clausen *et al.* (2007) [310] demonstrated increased NRM and reduced OS were associated with transplants from donors with increased numbers (>2) of unique activating KIR genes. Additionally, in a study into MAC, TCD, VUD HCT for AML patients, Kröger *et al.* (2006) [278] demonstrated an increase in relapse following transplants from donors encoding the KIR BX genotype or with elevated activating KIR frequency. Increased relapse risk translated into decreased OS and DFS probability.

Kröger *et al.* (2006) [278] also investigated the impact of missing ligands on HCT outcome in this relatively homogeneous cohort. Here, in agreement with a previous study in a similarly TCD cohort [304], missing donor KIR ligands were associated with negative outcomes: increased NRM risk, and reduced OS and DFS probabilities. Absence of HLA-Bw4 from recipients of KIR3DL1^{+ve} donor grafts also corresponded with a detrimental effect in my cohort; increased relapse risk at five years post-transplant, although no significant difference in probability of NRM, OS or DFS (or aGVHD) was apparent. An increased relapse risk without concurrent influence on other HCT outcomes has previously been demonstrated in a MAC, TCD cohort of VUD

transplants [326]. As such, the impact of missing KIR ligands, particularly in VUD HCT, is still under debate, with many studies showing opposing impacts, whilst others show no effect at all (Table 3.38).

Upon investigation of KIR gene matching between donors and recipients in the HLA-matched, adult, MAC sub-cohort, presence of mismatching at several KIR loci was indicative of significantly altered outcome, particularly in the GVH direction. In a univariate analysis, KIR2DL2 and KIR2DS2 GVH mismatching was associated with reduced NRM risk at one year post-transplant and improved five year OS and DFS probabilities, whilst the opposite appeared true for KIR2DL3. KIR2DL2/S2 GVH mismatching may be indicative of transplants featuring Cen-AA donors, shown above to have a beneficial impact on transplant outcomes in this sub-cohort. However, approximately half (28/52) of the Cen-AA donor transplants were not KIR2DL2/S2 mismatched, and yet strong significance was maintained. Furthermore, in an adjusted multivariate analysis model, GVH direction mismatches at any KIR loci other than KIR2DL2 or KIR2DS2 were an independent indicator of increased one year NRM risk and resultant inferior five year OS and DFS probabilities. This result has not been observed in prior publications (Table 3.39)

Relapse also appeared to be influenced by KIR gene matching. However, relapse risk did not appear to fluctuate by individual loci. Instead, mismatching for at least one activating KIR locus was associated with reduced relapse risk at five years post-MAC transplant. This may signify that KIR, particularly activating KIR, can act as minor histocompatibility antigen targets for T cell-mediated immunity. This finding is in direct contrast with the study by Sahin *et al.* (2018) [283] on a T cell replete cohort, in which

the opposite effect on relapse was observed. As incidence of cGVHD data was not available for analysis, we are unable to determine whether the reduced relapse risk was correlated with an increase in cGVHD, although no significant differences were observed in incidence of aGVHD.

No evidence of impact on HCT outcomes according to donor Tel-B content was observed in any HCT setting, conflicting with several large studies from the USA [273,300,317]. Additionally, and in contrast to many of the larger registry studies investigating the impact of KIR genotype polymorphism on HCT outcomes, relapse risk was not frequently affected by KIR polymorphism. Instead, the influence on NRM risk at one year post-transplant was more apparent. One possible explanation for these differences may be the use of alemtuzumab as the primary TCD agent in my cohort. TCD in any form, in addition to its role as prophylaxis against GVHD, has been demonstrated to improve the recovery of NK cells relative to T cells in the post-transplant setting [200]. Additionally, it has been demonstrated that, when compared to transplants which utilised ATG as the depleting agent, alemtuzumab resulted in less aGVHD, but increased risk of relapse [327,328]. That alemtuzumab has a long-lasting influence on depleting the number of circulating T cells supports the notion that T cells are responsible for reduction in relapse, and that KIR polymorphism may be liable for differences in the strength of the T cell response. As such, prolonged TCD in transplants utilising alemtuzumab may limit the impact of KIR polymorphism on T cell reactivity, as there are fewer T cells on which to act, consequently also removing its impact on relapse risk.

In addition, HCT utilising alemtuzumab protocols, when compared to equivalent T cell replete or ATG-depleted grafts, is also associated with an increased risk of viral complications [329-331]. As the vast majority of TCD performed in the previously published studies has involved ATG, the use of alemtuzumab in this study's transplant recipients may account for the significant findings associated with donor Cen-BX haplotype motif structure that have not frequently been observed previously. Alternatively, the multifactorial nature of NRM, and relative infrequency of infectious mortality (when compared to relapse-related death) makes investigation of the influence of KIR polymorphism on death involving infection more difficult. Indeed, in the adult, HLA-matched RIC cohort, the low incidence of NRM, and high heterogeneity within the NRM group's causes of death, prevented analysis. However, when specifically evaluated, Schaffer *et al.* (2004) [304] previously demonstrated similar increases in infectious mortality associated with KIR polymorphism.

One particularly puzzling finding of this study was the opposing impact that donor's centromeric motif structures had on the incidence of NRM by one year post-transplant in adult, HLA-matched recipients. In transplants that followed RIC regimens, protection was offered by the presence of at least one Cen-B motif in the donor's genotype. Conversely, in the MAC setting, this same donor factor was associated with an increased susceptibility to NRM at one year post-transplant which conveyed impaired probability of OS and DFS at five years post-transplant. This drastic effect of conditioning regimen on the role of KIR genotype polymorphism has not been thoroughly investigated. However, Brunstein *et al.* (2009) [265] demonstrated that increased KIR alloreactivity through ligand mismatching led to detrimental GVHD and NRM risk only in RIC, not MAC, CBU transplants. Another interesting observation

was made by Clausen *et al.* (2010) [313] in an HLA-matched, related donor scenario. When NK cell dose was higher in the RIC group of this study, the risk of relapse was reduced, corresponding with improved OS. However, the opposite was true in the equivalent MAC transplants, where increased NK cell count within the graft was associated with reduced OS probability. Interestingly, NK cell count has also been shown to affect the influence of the degree of NK cell-mediated modulation of antiviral T cells [332]. Although the retrospective nature of my study prevents an assessment of graft composition, the difference in effect of NK cells according to conditioning regimen may explain the differences observed between adult HLA-matched cohorts.

There are several hypotheses relating to how differing donor KIR and recipient HLA genotypes affect HCT outcome. When considering the missing KIR ligand model, it is assumed that newly reconstituted, uneducated NK cells identify the missing self scenario and mount an immune response. Reactivity is gradually reduced over the 100 days immediately after HCT as NK cells become tolerized [333].

Another theory assessing KIR genotype and NK cell reactivity relates to the presence of donor-encoded activating KIR genes. Donor-derived KIR2DS1^{+ve} NK cells have been shown to lyse leukaemic blasts via interaction with their HLA-C2 ligand [334]. A higher proportion of activating to inhibitory KIR genes may also correlate with a more reactive phenotype, as donor-derived NK cells are able to become activated more easily upon encountering their ligand expressed by residual leukaemic blasts. This theory supports the many studies showing reduced relapse rates in recipients of grafts from donors with increased numbers of activating KIR genes or higher B haplotype motif content [83,272,273,275,300,302,317,318]. This theory also supports the notion of

increased reactivity against DCs, as GVHD rates have rarely been observed to differ according to donor activating KIR content. Finally, although relapse was not an affected outcome in this particular analysis, increased NK cell reactivity against virally-infected cells may be responsible for the reduction of NRM observed in the adult, HLA-matched RIC sub-cohort [335].

By contrast, the presence of certain activating KIR genes has been linked with inferior viral disease outcome. The presence of KIR2DS2 and KIR2DS3 genes has been associated with increased chronicity of hepatitis C [336-338]. Khakoo *et al.* (2004) [336] proposed that, rather than being mediated directly by KIR2DS2, the effect was caused by the increased degree of inhibitive signalling offered by KIR2DL2 (in strong linkage disequilibrium with KIR2DS2) in comparison to KIR2DL3; the KIR gene occupying the same locus on the Cen-A haplotype motif. The increased avidity [105,174] and functionality [339] of KIR2DL2 over most KIR2DL3 alleles is generally attributed to this increased inhibitory capacity. Although the experiments by Bari *et al.* (2016) [339] demonstrated a stronger education effect from KIR2DL2 resulting in increased lysis of target cells lacking ligand, the complete loss of ligand is unlikely to occur naturally during infection, especially considering the targeted downregulation of HLA-A and -B only by certain viruses (HLA-C and -E expression is maintained) [29]. Instead, high avidity offered by KIR2DL2 may result in inhibitory contacts being made more frequently, resulting in less NK cell-mediated lysis.

Alternatively, activating KIR may be more directly responsible for downregulation of NK cell reactivity. Expression of KIR2DS1 on the surface of NK cells (and subsequent interaction with its ligand, HLA-C2) has been demonstrated to educate cognate cells not

concomitantly co-expressing the inhibitory KIR2DL1 receptor. However, rather than increasing the responsiveness to activating stimuli, education via the activating KIR2DS1 rendered NK cells hyporesponsive [340]. As such, an increased presence of activating KIR encoded by donors may in fact lead to a decrease in the reactivity of NK cells, potentially responsible for the increased risk of infectious mortality observed in the recipients of grafts from Cen-BX donors within the adult, HLA-matched MAC sub-cohort.

The target of NK cell reactivity is also variable. NK cell reactivity directed towards residual AML blasts may limit relapse directly. Additionally, NK cell recognition and direct lysis of virally infected cells may reduce infectious mortality. However, it has also been demonstrated that both cytotoxic and helper T cell subsets can be targeted for lysis by NK cells [332,341]. Conversely, inhibitory KIRs have also been shown to enhance T cell-mediated antiviral immunity [342]. Additionally, NK cells under inflammatory conditions have been shown to destroy naïve DCs via a TRAIL-mediated pathway [343]. This may result in a reduction of GVHD risk, due to a decrease in the number of DCs presenting alloreactive host peptides. Indeed, this may be responsible for the reduction in aGVHD detected in recipients with missing KIR3DL1 ligand from the overall cohort (Table 3.38). That this finding did not persist in either adult, HLA-matched cohort may reflect the lack of HLA mismatched alleles and corresponding decrease in alloreactivity. Interestingly, the targeting of DCs by NK cells has been shown to be upregulated by GCSF [344]. As this drug is commonly used in donors to mobilise HSCs to the peripheral blood prior to HSC collection, as well as to promote neutrophil reconstitution in recipients after their transplant [345], the potential for increased NK cell reactivity against DCs in these scenarios may be increased. The

observation that NK cells, under control of the polymorphic KIR gene system, can affect such a wide variety of immune system cells, demonstrates their importance as a cross-linking arm between adaptive and innate immunity, but also raises many questions about the mechanism by which KIR polymorphism modulates HCT outcomes.

There are several limitations to the analysis performed. By stratifying the cohort, sample size was inevitably significantly reduced. Despite this, more than 100 transplants exist within each of the adult, HLA-matched sub-cohorts. However, within the cohorts designed to assess the impact of HLA-C1 ligand encoded by the recipient, constricted cohort size may have underpowered the analysis, potentially explaining the lack of significance observed. Although this was implemented to maximise homogeneity within individual cohorts, the multi-centre aspect of this study also introduces the caveat of variability in transplant protocols between and within transplant centres. In addition, the retrospective nature of this analysis presented difficulties in obtaining complete clinical follow-up data resulting in completely missing variables, such as cGVHD and conditioning protocol composition; and incomplete factors, such as TCD and precise cause of death.

Additionally, as KIR3DL2 has been shown not to educate NK cells via its interaction with HLA-A3/11 ligand [340,346], it was not investigated within the missing ligand model. However, inhibitory signalling via this pairing may still have had bearing on the outcome of HCT. Furthermore, although predominantly expressed on the surface of NK cells, KIR are also expressed by certain subsets of T cell [347]. As a result of the relatively restricted recovery of T cells post-transplant in this TCD cohort, this has not

been considered in the current analysis. However, KIR-mediated influence on the activity level of T cells may also be relevant in HCT outcomes, particularly in differences observed following immune reconstitution. Finally, although theories have been generated based on KIR genetic data, no attempt has been made to validate these theories with cell-based, functional assays, a step that would be greatly beneficial in future studies.

The final limitation with the analysis of KIR genotype is inherent to all similar studies. This is that, although extensively polymorphic at the level of presence or absence of each KIR locus, each of the KIR loci also exhibits substantial allelic polymorphism. This type of polymorphism has been shown to modulate the impact of individual KIR genes and is the main motivation for the analysis in the following chapters.

3.12 Conclusions

The influence of KIR genotype polymorphism on the outcomes of HCT are a matter of considerable debate, with conflicting results from many different studies. This analysis, in a previously unstudied UK cohort, is no exception, with an apparently unique collection of findings. Despite this, the analysis succeeds in highlighting how transplant protocol, particularly preparative regimens, can have significant impacts on the influence of KIR polymorphism. Although the current dogma for utilisation of KIR genotyping data in donor selection for VUD HCT is to maximise KIR B motif content, I have demonstrated that caution should be taken when selecting such donors, particularly if the recipient is to undergo a MAC preparative regimen with *in vivo* alemtuzumab-based TCD. Instead, prospective studies controlling for TCD and conditioning intensity

are recommended to assist in deciphering the impact of KIR genotype within the VUD HCT setting.

Chapter 4 Design and optimisation of a full-length KIR allele sequencing strategy

4.01 Introduction

The impact of copy number variation (CNV) at each of the KIR loci has been shown to modulate NK cell activity and have functional relevance in both immunity and pregnancy. However, in addition to this variation, a further layer of complexity is added when considering that each of the KIR genes exhibits a high degree of allelic polymorphism. This has been demonstrated as having the potential to drastically alter the functionality of KIRs. Effects on ligand avidity [105,106,112,174,175,177,298,348,349], specificity [106,108,111,118,298,350], cell surface expression [139,153,156,157,348,349,351-356] and signalling strength [139,169,339,357] have all been observed when comparing different alleles of the same KIR gene. As such, understanding the allelic KIR environment, not just the presence or absence of individual KIR genes, may help to resolve conflicting reports regarding the influence of KIR in immunity.

The KIR genes are a highly homologous family of genes, sharing up to 98% sequence identity (comparison between genomic sequences of KIR2DL2*0010101 and KIR2DL3*0010101 alleles) [98]. As such, primer design for PCR-based targeting of individual KIR loci is difficult. Previous PCR-SSP KIR genotyping strategies have focussed on regions within exons that differ between the loci [178]. Alternatively, amplification of each exon, followed by Sanger sequencing, has been developed as a method to sequence the entire CDS [181]. More recently, PCR amplification using the highly conserved intergenic sequence as a primer binding site, followed by short-read

NGS, has allowed full length characterisation of KIR genes [182]. Short-read NGS has also been employed to sequence fragments of the KIR haplotype purified using specific oligo-nucleotide probes which are then reassembled into complete KIR haplotype sequences [94]. However, the complex nature of allelic polymorphism at each locus presents problems for any short-read sequencing method: phasing polymorphisms that are distant from each other within an allele sequence. This phase ambiguity has previously resulted in erroneous allele designation [358].

As such, I have designed a PCR-based strategy to amplify specific KIR genes (or pairs of genes), which are then prepared into SMRTbell libraries prior to sequencing on Pacific Biosciences' RS II platform. This platform allows the generation of continuous long reads, covering the entire length of KIR genes, preventing occurrence of any phase ambiguity. In addition, the high quality consensus sequences generated reduce the risk of false SNP detection. Optimisation of the PCR, library preparation and sequencing steps were all performed to maximise quality and efficiency of KIR allele designation, as described below. Optimisation of bioinformatic processes were performed in collaboration with Anthony Nolan Research Institute's Bioinformatics department and, as such, will not be discussed in detail.

4.02 Full-length KIR gene primer design and PCR optimisation

At the beginning of my project, there were no suitable published primer sequences targeting full-length KIR genes. As such, primer design was performed to determine optimal primer sequences to permit PCR-based amplification of a single DNA fragment encompassing an entire KIR gene. As the KIR genes exhibit such high sequence homology, primer design and optimisation of PCR protocols to achieve specific

amplification of only the intended target was a major obstacle and formed a large part of my project.

4.02.01 Data mining

To ensure that the entire CDS region and as many polymorphic locations as possible were within the amplified target, primers were designed to anneal within the UTR sequences at the beginning and end of each KIR gene. Although the official repository for KIR allele sequences, and the location of several useful tools for KIR sequence alignment, sequences deposited within the IPD-KIR Database at the time of primer design (early 2016) frequently did not have complete genomic sequence and were thus lacking the relevant UTR sequence data. Additionally, the published UTR sequences are trimmed to only align approximately 300 bp either side of the start/stop codon. As such, I sought to create my own alignments of KIR sequences.

Published sequencing data from whole genome sequencing projects, such as the 1000 genomes project, were not utilised for this effort as, although overall false detection rates of SNPs and indels are very low, low coverage results in errors in polymorphic loci such as KIR and HLA that are less likely to be detected [359]. Instead, data from projects aimed specifically at generating full KIR haplotype sequences were utilised [87,88]. Markers for each individual KIR gene were searched for within these haplotype sequences. From these, the start and stop codon for each gene within the haplotype were identified and the 1,000 bp of sequence upstream of the start codon, or downstream of the stop codon, were collected. This process was repeated for each haplotype sequence, developing a broad selection of extended UTR sequences for each individual KIR gene. The 3′ UTR of KIR3DP1 was the exception to this rule: instead of 1,000 bp of

sequence, 7,500 bp of sequence were extracted. This was to permit identification of primer binding sites that would allow equivalent length amplicon generation, discussed further in Section 4.02.07.02. Collection of KIR haplotype sequences published after the original data collection was maintained as a method to continually validate primer binding sites.

4.02.02 Primer binding site characterisation

Following initial data collection, alignments of both UTRs from each gene (excluding KIR3DP1 in the 3′ UTR alignment) were performed using the multiple alignment using fast fourier transform tool (MAFFT, EMBL-EBI, Cambridge, UK) [360]. Consensus sequences were then formed for each UTR alignment using Ambiguity Consensus Maker (HIV sequence database, Los Alamos, NM, USA) [361]. Variation in at least one sequence resulted in assignment of International Union of Pure and Applied Chemistry (IUPAC) degenerate code at that position [362].

The consensus sequences for each 5′ UTR were then aligned against each another using MAFFT. This alignment was performed to aid the selection of primer binding sites specific for every allele of a given locus, but without cross-reactivity against any allele of any other KIR locus. This alignment process was repeated for the 3′ UTR consensus sequences. The difficulty of designing primers specific for a single locus (as a result of the high degree of homogeneity between KIR sequences) is demonstrated in Figure 4.1. The annealing sites of the PCR primers used in the final strategy are given in Supplementary Figure A and Supplementary Figure B.

```
2DL1    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
2DL2    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
2DL3    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
2DL4    TATGTCCCCTTCACATGTTGTGGTCAATGTGTCAACTGCACGATCCGGGCCCCTCACCAC
2DL5A   TACGTCACCCTCCCGTGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
2DL5B   TACGTCACCCTCCCATGATGTAGTCAACATGTAAGCTGCATGGGCAGGGCGCCAAATAAC
2DP1    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
2DS1    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
2DS2    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
2DS3    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGAGCCAAATAAC
2DS4    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
2DS5    TACKTCACCCTCCCRTGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
3DL1    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCRAATAAC
3DL2    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
3DL3    TACGTCATCCTCCCATGATGTGGTCAGCATGTAAACTGCAT-----GAGCCCMTCACAAC
3DP1    TAYGTCACCCTCCCATGATGTRGTCAACATGTAARCTGCATGGGCAGGGCGCCAAATAAC
3DS1    TACGTCACCCTCCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAAC
```

Figure 4.1 Consensus sequence alignment of published KIR 5´ UTRs

A sequence alignment of consensus sequences for each KIR gene over a desirable primer binding location: approximately 100bp upstream of the start codon. However, high homogeneity between the KIR loci prevents design of primers specific for an individual locus. Nucleotides differing from the overall consensus are highlighted in red. ENA accession numbers of the sequences used to generate consensus sequences is given in Supplementary Table G.

4.02.03 Primer sequence optimisation by *in silico* characterisation

Where possible, several potential primer sequences were selected for each KIR locus. The selection of optimal primer sequences was based on an *in silico* analysis performed in OligoEvaluator™ (Sigma-Aldrich). Strong secondary structure was avoided where possible and approximately 50% GC content was desirable. Table 4.1 contains a summary of the final selected primer sequences and their characteristics.

In addition, each primer sequence underwent analysis by BLAST-n complementarity screen. Each primer was compared against the human genomic and transcript database. Oligos that were found to have significant complementarity to regions other than the

KIR loci within the human genome or transcriptome were excluded to avoid non-specificity.

Table 4.1 Selected primer *in silico* characterisation

Primer	Length (bp)	Melting temperature (°C)	GC content (%)	Predicted secondary structure
5′ primers				
KIR2DL1	20	58.1	45.0	Weak
KIR2DL2/3	23	65.8	47.8	Weak
KIR2DL4	20	61.8	50.0	Weak
KIR2DL5/3DP1	24	69.9	54.2	None
KIR2DL5_alt	27	69.5	40.7	Moderate
KIR2DS2	23	65.8	47.8	Weak
KIR2DS5	19	68.5	57.9	None
KIR3DL1/S1	30	73.9	50.0	Weak
KIR3DL2	25	68.5	44.0	Moderate
KIR3DL3	21	64.1	47.6	Moderate
3′ primers				
KIR2DL1/2/3	26	64.8	42.3	Very Weak
Generic	18	66.9	66.7	None
3DP1	25	67.7	44.0	Weak

4.02.04 Targeting KIR2DL1/2/3 with specific 5′ and 3′ primers

4.02.04.01 Amplification strategy

My initial strategy was focussed on sequencing the KIR2DL1, KIR2DL2 and KIR2DL3 genes only. However, due to the exceedingly high homogeneity between KIR2DL2 and KIR2DL3, it was not possible to design PCR primers to separate these genes. However, selection of primers that were adequately specific in both the 5′ and 3′ UTRs permitted targeted amplification of either KIR2DL1 or a combined amplification of KIR2DL2 and KIR2DL3 (herein referred to as KIR2DL2/3, primer sequences available in Chapter 2, Table 2.6). The KIR2DL1 5′ primer was specific for KIR2DL1, whilst the equivalent primer for the KIR2DL2/3 reaction shared 100% complementarity with KIR2DS2. The KIR2DL1 and KIR2DL2/3 amplifications shared a 3′ primer with specificity for these

three genes and KIR3DS1 only (Table 4.2). Accordingly, when used in the correct combinations, these primers ensured gene specificity despite having complementarity to other genes in single directions.

Table 4.2 Initial amplification strategy targeting KIR2DL1 and KIR2DL2/3

Amplification target	5´ primers		3´ primer	
	Binding site[§]	Specificity	Binding site[§]	Specificity
KIR2DL1	-561	KIR2DL1	14546	KIR2DL1, KIR2DL2/3, KIR3DS1
KIR2DL2/3	-476	KIR2DL2/3, KIR2DS2	14592	

[§] Binding sites are approximate locations based on the reference allele for each locus.

4.02.04.02 PCR optimisation

Initial PCR optimisation was attempted with the Phusion High-Fidelity DNA Polymerase (Table 4.3). An annealing temperature gradient PCR protocol was established, whereby the annealing temperature was varied in 3°C increments across the thermal cycler (57-72°C, Table 4.4). PCR products were separated by 1% agarose gel electrophoresis before analysis on a UV transilluminator. This revealed that, for KIR2DL1, specific amplification of a DNA fragment approximately matching the predicted amplicon length (~15 kbp) was achieved by utilising PCR with annealing temperature at 63°C (Figure 4.2A). However, the banding was very faint; indicative of low concentration product. Additionally, no bands of correct MW were observed at any annealing temperature for the KIR2DL2/3 PCR.

Table 4.3 Reagent composition for initial KIR2DL1 and KIR2DL2/3 amplification tests using Phusion DNA Polymerase

Reagent (conc.)	Volume (µL)
Water	9.4/8.8
† DMSO (100%)	0/0.6
† Phusion Buffer (5X, GC or HF)	4.0
dNTP Mixture (10 mM)	0.4
5´ primer (25 µM)	1.0
3´ primer (25 µM)	1.0
Phusion DNA Polymerase (0.2 U/µL)	0.2
§ Template DNA (25 ng/µL)	4.0

† PCR optimisation tests utilised varied DMSO and Phusion buffer compositions. Inclusion of 0.6 µL of DMSO required equivalent volume decrease of water.

§ The sample, HO104 (IHW9082), was used to test both KIR2DL1 and KIR2DL2/3 amplifications as it has been demonstrated to encode each KIR gene.

Table 4.4 PCR cycling condition for initial KIR2DL1 and KIR2DL2/3 amplification tests using Phusion DNA Polymerase

Step	Temperature (°C)	Duration (min:sec)	Repetitions
1	98	0:30	1
2	95	0:15	34
	57-72	0:30	
	72	13:00	
3	72	13:00	1
	4	∞	

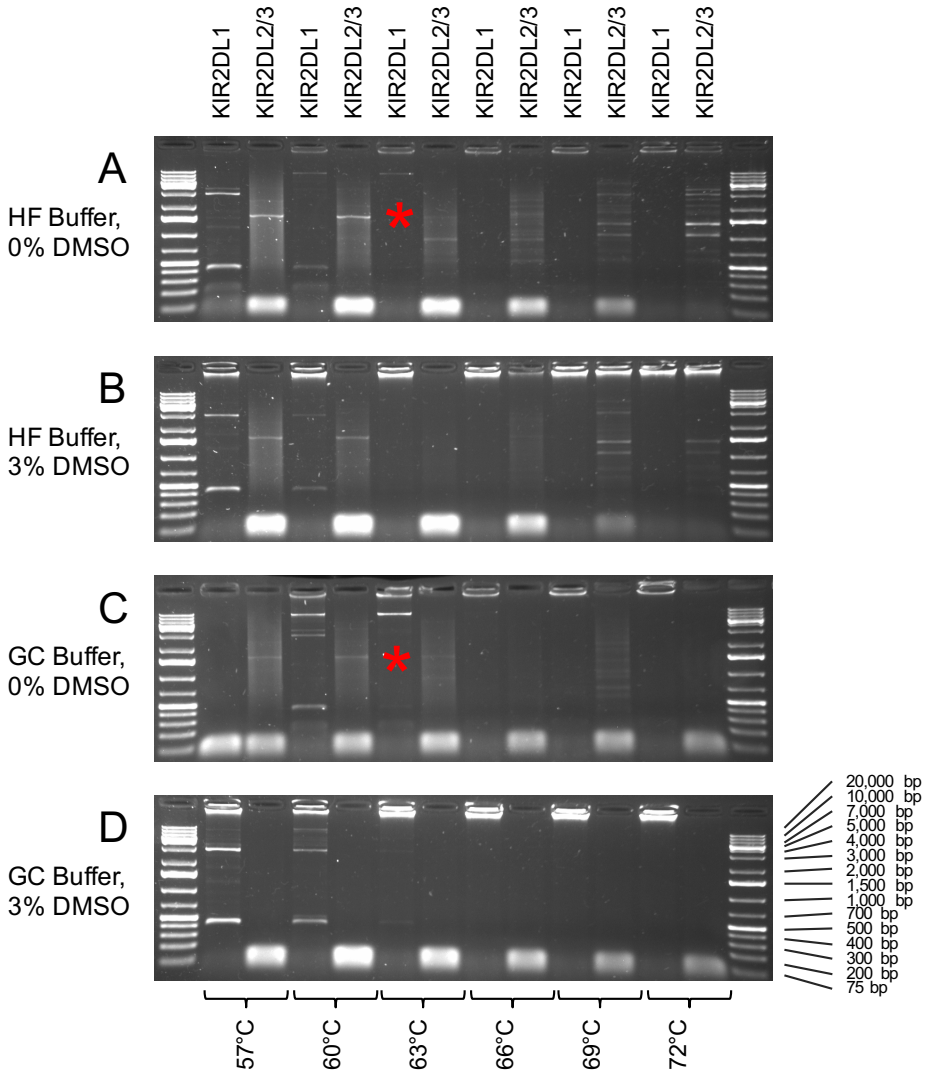


Figure 4.2 Agarose gel electrophoresis of KIR2DL1 and KIR2DL2/3 targeted PCR optimisation using Phusion DNA Polymerase

Annealing temperature gradient revealed differences in PCR specificity according to temperature, PCR buffer (A and B *vs* C and D) and inclusion of 3% DMSO (A and C *vs* B and D). Annealing at 63°C in the absence of DMSO was observed to facilitate specific amplification of a fragment ~15 kbp, presumed to be KIR2DL1 (denoted by a red star). GeneRuler™ 1 kb Plus DNA Ladder was run at either end of each group.

As such, the experiment was repeated in triplicate, changing a single variable in each replicate (Figure 4.2B-D):

i. Use of a PCR enhancer: 3% DMSO. DMSO assists in the denaturing of template DNA, but also lowers the annealing temperature of primers.
ii. Use of the lower fidelity, but improved performance, GC buffer.
iii. Use of both 3% DMSO and GC buffer.

This revealed that the KIR2DL1 amplicon concentration was improved at 63°C with the GC buffer, but that DMSO had a detrimental impact on all conditions. No specific KIR2DL2/3 amplification was observed. Throughout my project, several other PCR enhancers were tested, including betaine, DTT, thioglycerol and the enhancer supplied with the Platinum SuperFi DNA polymerase. In no instance was significant improvement observed.

As the initial temperature gradient experiment suggested that increased annealing temperatures resulted in a decreased proportion of shorter off-target fragments, a second temperature gradient, assessing a narrower range of the higher annealing temperatures (68-73°C) was performed to improve the specificity of the KIR2DL2/3 PCR amplification. This was combined with a $MgCl_2$ titration ($MgCl_2$ concentration was reduced by diluting the reaction buffer or increased by substituting water within the reaction for 50 mM $MgCl_2$). Each of these reactions also failed to specifically amplify KIR2DL2/3.

To maximise the likelihood of developing a successful KIR2DL2/3-specific PCR amplification, alternative thermostable DNA polymerases were tested. Of these, TaKaRa's PrimeSTAR GXL DNA polymerase was the most promising. High MW fragments were observed in initial temperature gradient reactions and, following

refinement by experimentation with touchdown PCR protocols and $MgCl_2$, DNA polymerase and primer concentration titrations, conditions amenable for PCR-based amplification of single DNA fragments with MW approximately equal to the predicted KIR2DL1 and KIR2DL2/3 amplicons were determined (Figure 4.3).

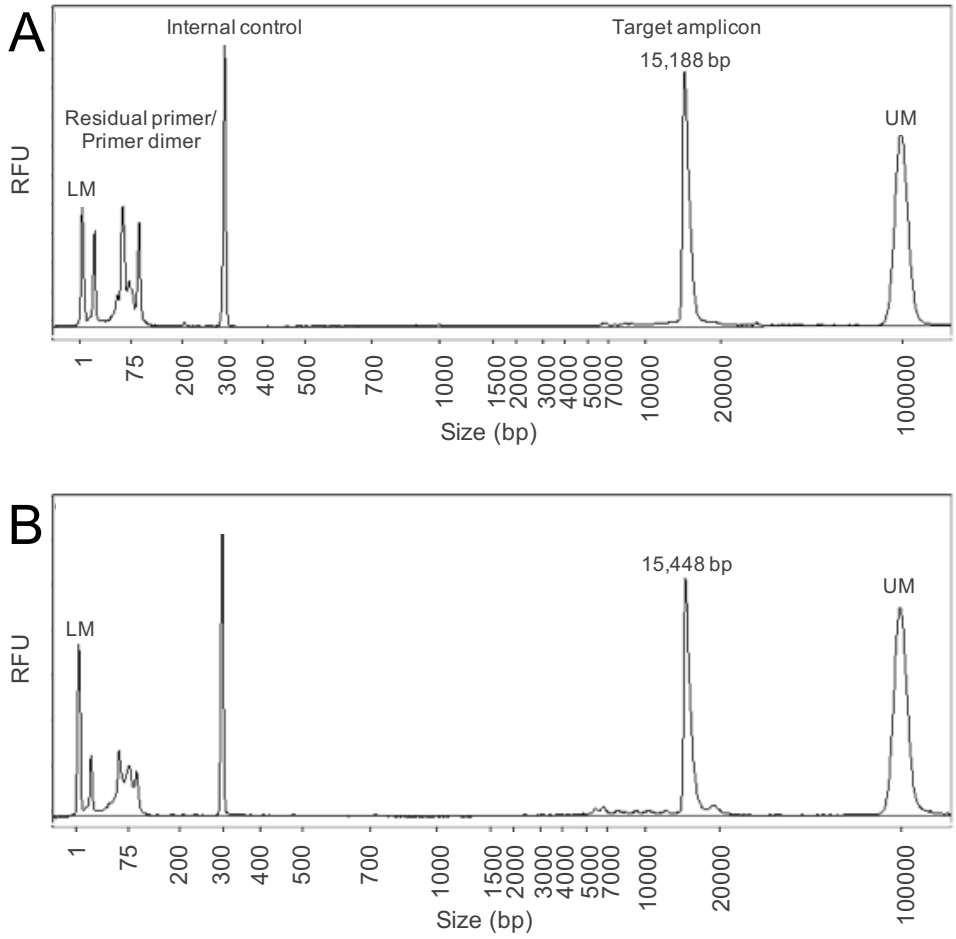


Figure 4.3 Fragment Analyzer analysis confirms specific amplification of fragments approximately equal to the predicted MW of KIR2DL1 and KIR2DL2/3 targets

Example electropherogram traces generated by the Fragment Analyzer platform reveal amplification of regions with similar MW to predicted amplicons. A) KIR2DL1 PCR generates a PCR product with an approximate fragment length of 15.2 kbp. B) KIR2DL2/3 PCR generates a PCR product with an approximate fragment length of 15.4 kbp. Lower marker (LM) and upper marker (UM) are annotated. Internal control PCR product (~300 bp) is clearly visible. Fragments of shorter size (<150 bp), corresponding to residual primer and primer dimer, are also observed.

4.02.05 Targeting KIR3DL1/S1 with specific 5′ and 3′ primers

4.02.05.01 Amplification strategy

Having created consensus sequence alignments for each of the KIR genes, I also attempted to design primers specific for the KIR3DL1 and KIR3DS1 genes. Specificity was achieved through the utilisation of specific 3′ UTR primer sequences. However, the high homogeneity between the genes within the 5′ UTR of the sequences prevented equivalent specificity here, resulting in 5′ primer design with equal binding potential to either KIR3DL1 or KIR3DS1. In addition, because of the observation of KIR3DL1 alleles formed through a recombination event with KIR3DL2 (e.g. KIR3DL1*059) [363], a 3′ primer combination merging specificity to both KIR3DL1 and KIR3DL2 was utilised to ensure amplification of all known KIR3DL1 alleles. Finally, to maximise the likelihood of discovering successful PCR primer sequences, several different options for each primer were designed (Table 4.5).

Table 4.5 Primers designed to specifically target KIR3DL1/S1

Primer specificity	Sequence (5′-3′)	Binding site§	Melting temperature (°C)
5′ primers			
3DL1/S1-F1	AGAGTCTCGCTCTGTTG	-931	53.4
3DL1/S1-F2	GAATAGCTGGCATTACAAG	-828	54.9
3DL1/S1-F3	GTCTCAAACTCCTGACCTCGGTTGATCACT	-727	73.9
3′ primers			
3DL1-R1	CCAGATTTGTGGCGTGAGGAG	13903	69.0
3DL1-R2	GTCCGAAGAAAGGTGA	14169	53.9
3DL1-R3	CAGTGTCACATTACCTGAAGCA	14279	62.7
3DL1-R4	GCGGTTTCTTTCAGCGAATACAGTGTCA	14293	73.5
3DL2-R1	AGTGTGATTGCAGCCTCAAGTAGGA	14030	68.9
3DL2-R2	AAGGTGGAACAGCAAGTGTG	14108	62.8
3DL2-R3	TTGAAGAGGAGAGAGCTACAC	14211	57.8
3DL2-R4	CCCAGTAGAGAACATATCA	14320	52.2
3DS1_R1	CAGATGGGATTATATGGACATGGTAC	14714	75.5
3DS1_R2	ATCAGAGTCCAGGGATGAGAACTCAGTGGG	14743	69.2

§ Binding sites are approximate locations based on the reference allele for each gene.

4.02.05.02 PCR optimisation

Initial optimisation of the PCR protocols targeting KIR3DL1 and KIR3DS1 employed a temperature gradient (57°C, 62°C and 67°C) to determine a narrower annealing temperature range. Each temperature was tested against 18 unique primer conditions. KIR3DL1 primer combinations were selected for similar melting temperatures between the different 3´ primers, whilst all primer combinations were assessed for KIR3DS1 (Table 4.6). PCR reagent concentrations and cycling conditions are given in Table 4.7 and Table 4.8, respectively.

Table 4.6 Primer conditions for initial KIR3DL1/3DS1 PCR optimisation

Target gene	Primer condition	5´ primer	3´ primer #1	3´ primer #2
KIR3DL1	1	3DL1/S1-F1	3DL1-R1	3DL2-R1
	2	3DL1/S1-F1	3DL1-R2	3DL2-R3
	3	3DL1/S1-F1	3DL1-R2	3DL2-R4
	4	3DL1/S1-F1	3DL1-R3	3DL2-R2
	5	3DL1/S1-F1	3DL1-R4	3DL2-R1
	6	3DL1/S1-F2	3DL1-R1	3DL2-R1
	7	3DL1/S1-F2	3DL1-R2	3DL2-R3
	8	3DL1/S1-F2	3DL1-R2	3DL2-R4
	9	3DL1/S1-F2	3DL1-R3	3DL2-R2
	10	3DL1/S1-F2	3DL1-R4	3DL2-R1
	11	3DL1/S1-F3	3DL1-R1	3DL2-R1
	12	3DL1/S1-F3	3DL1-R4	3DL2-R1
KIR3DS1	13	3DL1/S1-F1	3DS1_R1	-
	14	3DL1/S1-F1	3DS1_R2	-
	15	3DL1/S1-F2	3DS1_R1	-
	16	3DL1/S1-F2	3DS1_R2	-
	17	3DL1/S1-F3	3DS1_R1	-
	18	3DL1/S1-F3	3DS1_R2	-

Table 4.7 Reagent composition for KIR3DL1 and KIR3DS1 PCR optimisation

Reagent (conc.)	Volume (µL)
Water	12.1
PrimeSTAR GXL Buffer, Mg^{2+} (5X)	5.0
dNTP Mixture (10 mM)	2.0
5′ primer (25 µM)	0.5
3′ primer mix (12.5 µM each)	1.0
PrimeSTAR GXL DNA Polymerase (1.25 U/µL)	0.4
Template DNA[§] (25 ng/µL)	4.0

[§] GRC-212 cell line DNA was used as it encodes both KIR3DL1 and KIR3DS1 and had previously been shown to successfully amplify large DNA fragments by PCR.

Table 4.8 PCR cycling condition for KIR3DL1 and KIR3DS1 PCR optimisation

Step	Temperature (°C)	Duration (min:sec)	Repetitions
1	95	1:00	1
2	95	0:30	32
	57, 62 or 67	0:20	
	69	13:00	
3	69	13:00	1
	4	∞	

Following agarose gel electrophoresis of the PCR products, it was observed that most primer combinations did not result in amplification of a single target. However, for PCR targeting KIR3DL1 using primer conditions 8, 9 or 10, more refined banding was observed, particularly at higher annealing temperatures. Of these, use of primer condition 10 combined with 67°C annealing temperature resulted in a single band (Figure 4.4A). However, when assessed by the Fragment Analyzer, a technique with improved resolution compared to agarose gel electrophoresis, low concentrations of non-specific products were observed between 3,000 and 16,000 bp (Figure 4.4B).

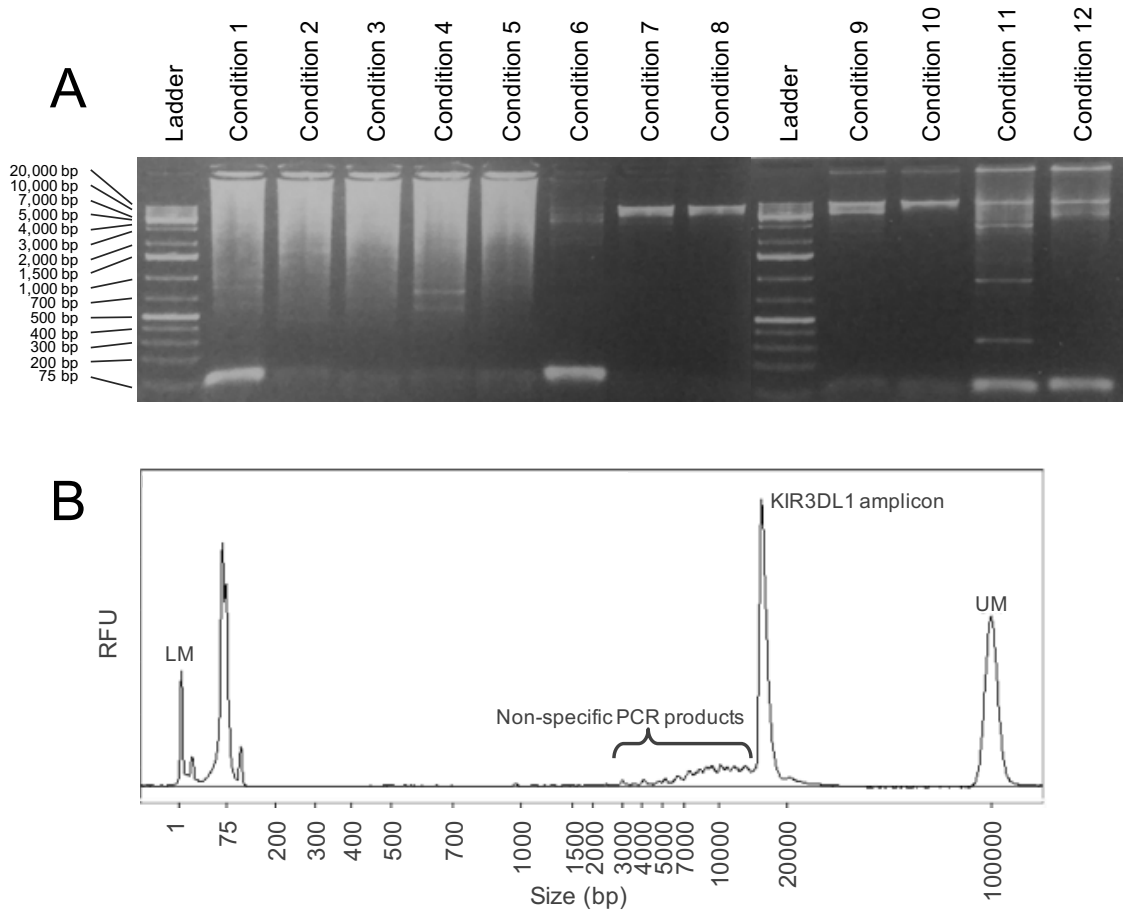


Figure 4.4 Agarose gel electrophoresis and Fragment Analyzer analysis of KIR3DL1 targeted PCR optimisation using multiple primer conditions

Multiple primer combinations (see Table 4.6) were tested to specifically amplify KIR3DL1 alleles. A) Of the 12 different conditions tested, agarose gel electrophoresis suggested that only condition 10 amplified a single fragment with MW approximately matching the predicted amplicon. Other conditions displayed off-target amplification characterised by smeared PCR products (conditions 1-5) or lower MW bands. B) However, analysis utilising higher resolution Fragment Analyzer technology revealed that non-specific PCR products are generated in condition 10, also.

The use of 67°C annealing temperature also appeared to result in the most specific KIR3DS1 amplification. Here, the only KIR3DS1-specific primer combinations were observed when the 5′ primer was KIR3DL1/S1-F3 (conditions 17 and 18). However, strong evidence of non-specific amplification in Fragment Analyzer analysis suggest that the attempts to specifically amplify KIR3DS1 failed (Figure 4.5).

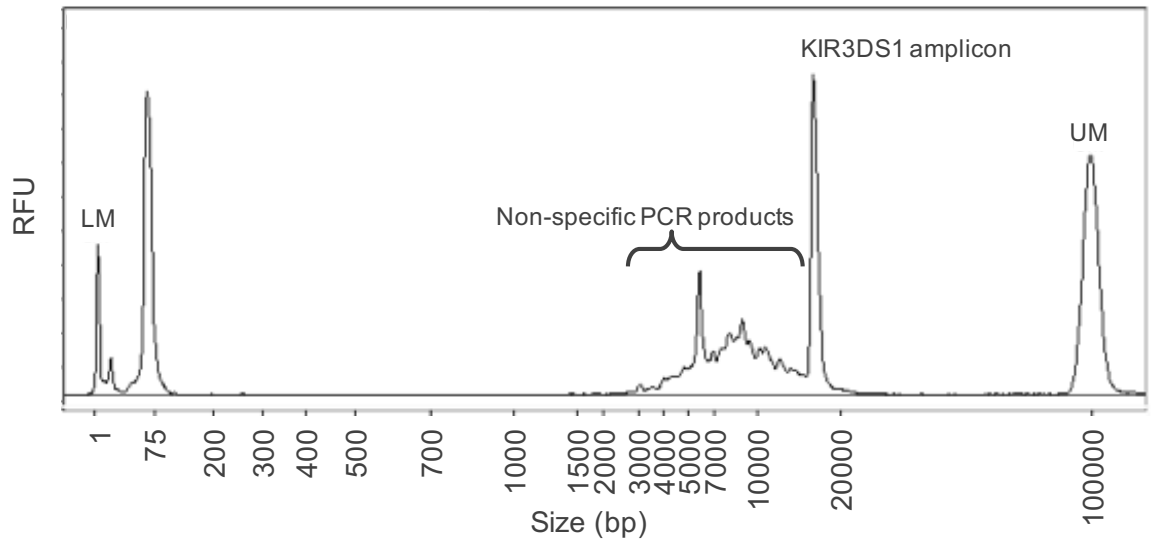


Figure 4.5 Fragment Analyzer analysis indicates non-specific PCR product generation from amplification targeting KIR3DS1

Attempts to specifically amplify KIR3DS1 alleles using primer condition 17 (and 18, not shown, Table 4.6) indicate that, although correctly sized fragments corresponding to KIR3DS1 alleles are amplified, a high concentration of smaller, non-specific products (3,000-13,000 bp) are also generated during PCR.

4.02.06 Targeting large groups of KIR genes within 'generic' reactions

Although attempts to better optimise the KIR3DL1 and KIR3DS1 PCR amplifications could have been made, upon further investigation of KIR gene and haplotype structure, the high frequency of recombination between the KIR loci prompted an alternative amplification strategy using 'generic' primers with shared specificity (i.e. will bind to multiple KIR genes).

4.02.06.01 Amplification strategy

Although highly homologous, very few regions of KIR sequences are completely conserved between all of the different genes. However, by using degenerate nucleotides within the sequences of just three primer sequences, it was possible to design a single PCR to theoretically target each of the 'long' KIR genes: KIR2DL1, KIR2DL2, KIR2DL3, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DS5, KIR2DP1,

KIR3DL1, KIR3DL2 and KIR3DS1 (Table 4.9). The approximate amplicon lengths of these products ranged between 12.6-16.6 kbp. This was predicted to be a sufficiently constricted fragment length range to permit loading within a single SMRT cell without causing an inhibitive degree of preferential loading of shorter fragments. The remaining, 'short' KIR genes, KIR2DL4, KIR3DL3, KIR2DL5 and KIR3DP1, could be selectively targeted using a second primer combination involving five unique primers. The predicted fragment length range for these PCR products was 9.5-12 kbp. Importantly, both PCR amplifications featured primers with broad specificity, theoretically permitting amplification of any novel hybrid allele, as well as all known KIR alleles (Table 4.10).

Table 4.9 Generic strategy primer sequences

PCR reaction	Primer ID	Sequence (5´-3´)
Long KIR	GenF1	AGGGMGCCRAATAACATCCTGTG
	GenR1	CTGMTGRCAGAAGRCTGA
	3DL1R4	GCGGTTTCTTTCAGCGAATACAGTGTCA
Short KIR	2DL4F2	GGTCAATGTGTCAACTGCAC
	3DL3F2	AGCATGTAAACTGCATGAGCC
	2DL5/3DP1F2	GGAGGTTGGATCTGAGACGTGTTG
	GenR2	GGTGGGCAGGGTCAAGTGA
	3DP1R3	CCTGTATGTTTATTGCAGCACTGTTTGC

Table 4.10 Generic strategy primer specificity chart

KIR gene	Long KIR PCR			Short KIR PCR				
	Gen F1	Gen R1	3DL1 R4	2DL4 F2	3DL3 F2	2DL5/3DP1 F2	Gen R2	3DP1 R2
KIR2DL1	✓	✓	✗	✗	✗	✗	✓	✗
KIR2DL2	✓	✓	✗	✗	✗	✗	✓	✗
KIR2DL3	✓	✓	✗	✗	✗	✗	✓	✗
KIR2DL4	✗	✓	✗	✓	✗	✗	✓	✗
KIR2DL5	✓	✗	✗	✗	✗	✓	✓	✗
KIR2DS1	✓	✓	✗	✗	✗	✗	✓	✗
KIR2DS2	✓	✓	✗	✗	✗	✗	✓	✗
KIR2DS3	✓	✓	✗	✗	✗	✗	✓	✗
KIR2DS4	✓	✓	✗	✗	✗	✗	✓	✗
KIR2DS5	✓	✓	✗	✗	✗	✗	✓	✗
KIR2DP1	✓	✓	✗	✗	✗	✗	✓	✗
KIR3DL1	✓	✗	✓	✗	✗	✗	✓	✗
KIR3DL2	✓	✓	✗	✗	✗	✗	✓	✗
KIR3DL3	✗	✓	✗	✗	✓	✗	✓	✗
KIR3DS1	✓	✓	✗	✗	✗	✗	✓	✗
KIR3DP1	✓	✗	✗	✗	✗	✓	✗	✓

4.02.06.02 PCR optimisation

Attempts to optimise the PCR conditions necessary for generic amplification of all KIRs within just two reactions focussed initially on an annealing temperature gradient (64, 66 and 68°C) and two separate primer balances:

i. Each primer at equal concentration,

ii. Equal concentration between total 5′ and 3′ primer pools.

The findings from the optimisation attempts of long KIR PCR were poor. Multiple short, contaminating fragments were observed and longer amplicon targets were not generated, suggestive of preferential amplification of shorter, off-target fragments (Figure 4.6). Although increased annealing temperatures did appear to correlate with improved balance of the longer fragments, the complex nature of these multi-primer

PCRs was deemed prohibitive to achieving equimolar ratios of each target within the time-frame available, and further optimisation efforts were ceased.

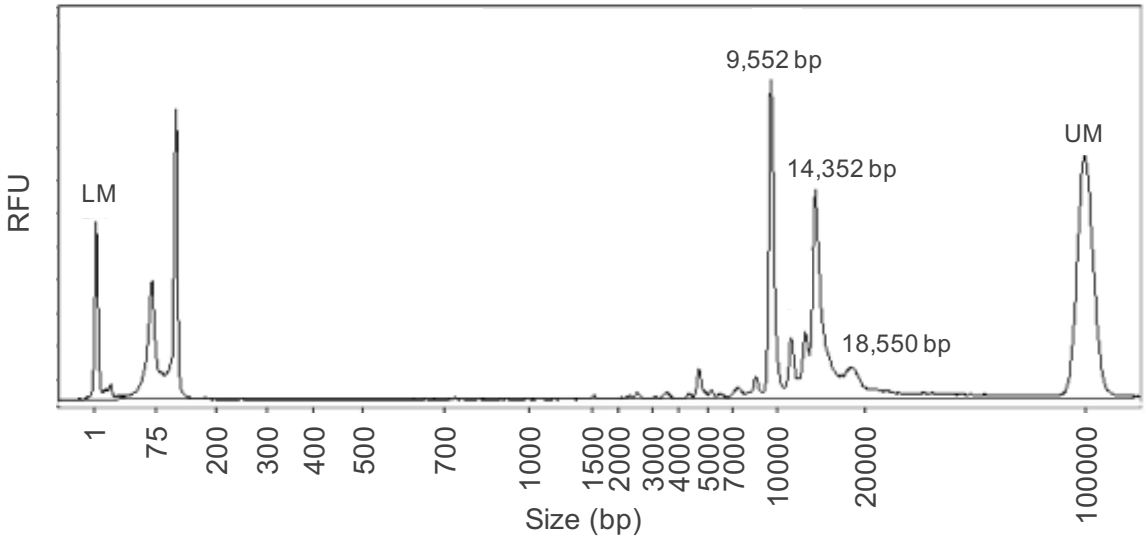


Figure 4.6 Fragment Analyzer analysis indicates poorly balanced amplification of multiple long KIR targets using generic primers

Attempts to amplify multiple long KIR genes using generic primers in both the 5′ and 3′ UTRs revealed preferential amplification of a short contaminating fragment. Although PCR products with approximate length of 14.4 kbp were generated, little indication of other fragment sizes (12-14 and 15-16 kbp) was evident.

The findings from the initial attempts at short KIR PCR optimisation were also disappointing. However, in a further round of development, the reaction was simplified by testing each 5′ primer separately from the others, but still in combination with the generic 3′ primer. This revealed a much more promising result and prompted the development of the ‘semi-generic’ PCR strategy, discussed below.

4.02.07 Generic 3′ primer introduces redundancy to targeted KIR amplification

4.02.07.01 Amplification strategy

Instead of attempting to optimise a fully generic PCR strategy, I adopted an alternative, semi-generic technique: single KIR genes (or pairs of genes) were targeted specifically in their 5′ UTR whilst a generic 3′ primer mix was utilised. All further KIR primer design was performed by this method (Table 4.11). The 5′ UTR of KIR2DL5 alleles features frequent polymorphism making design of an individual primer problematic. To ensure coverage of each different allele, two primers were designed. The first targets almost all KIR2DL5 alleles (as well as all KIR3DP1 alleles), whilst the second has complementarity to KIR2DL5*0020202, 003 and 00602 alleles. In addition, the first KIR2DL5 primer does not have specificity towards the KIR2DL5*0020106 and 0070102 alleles. However, although the published 5′ UTR sequence does not extend far enough to the alternative 5′ primer binding site, shared polymorphism with alleles of the second KIR2DL5 primer indicate that this primer may also be specific to KIR2DL5*0020106 and 0070102. As optimised protocols for KIR2DL1 and KIR2DL2/3 PCR-based amplification were previously developed (see Section 4.02.04), these genes were not optimised with the generic 3′ primer combination. Additionally, several activating KIR loci (KIR2DS1, KIR2DS3 and KIR2DS4) and KIR2DP1 do not yet have optimised amplification conditions. As such, these genes are not discussed further.

Table 4.11 Semi-generic amplification strategy

Amplification target	5′ primer		3′ primer mix	
	Binding site[§]	Specificity	Binding site[§]	Specificity
KIR2DL4	-97	KIR2DL4	10661	All KIR[†]
			-	KIR3DP1
KIR2DL5/3DP1	-262	KIR2DL5[‡]/3DP1	9160	All KIR[†]
	-289	KIR2DL5B*0020202 KIR2DL5B*003 KIR2DL5B*00602	10796	KIR3DP1
KIR2DS2	-476	KIR2DL2/3/S2	14049	All KIR[†]
			-	KIR3DP1
KIR2DS5	-733	KIR2DS5	14729	All KIR[†]
			-	KIR3DP1
KIR3DL1/S1	-757	KIR3DL1/S1	14029	All KIR[†]
			-	KIR3DP1
KIR3DL2	-290	KIR3DL2	16482	All KIR[†]
			-	KIR3DP1
KIR3DL3	-88	KIR3DL3	11872	All KIR[†]
			-	KIR3DP1

[§] Binding sites are approximate locations based on the reference allele for each gene.

[†] Binding specificity for all KIR excluding KIR3DP1.

[‡] Binding specificity for all KIR2DL5 excluding KIR2DL5*0020202, 0020106, 003, 00602 and 0070102.

- Not predicted to bind unless recombined with KIR3DP1.

4.02.07.02 Primer design in extended KIR3DP1 3′ UTR

To maximise the probability of detecting novel KIR alleles formed through recombination between different KIR genes, a generic 3′ primer mix was employed. This encompassed one oligo with specificity to all KIR genes, excluding KIR3DP1. A separate KIR3DP1 3′ primer was required due to the large deletion following exon 5 that defines KIR3DP1 and encompasses the region homologous to the other genes' primer binding region.

However, the desired amplicons from each individual targeted PCR must be of similar length to facilitate simple library preparation and balanced SMRT cell loading, a vital step in achieving balanced sequence representation during SMRT DNA sequencing. As such, the KIR3DP1 3′ sequences mined to generate the necessary consensus were extended to include 7,500 bp downstream of the end of exon 5 (equivalent to stop codon

for KIR3DP1). This allowed design of a primer that resides much further within the 3′ UTR than for other genes (approximately 5.4 kbp downstream of exon 5), but that also has a similar (albeit slightly larger) predicted amplicon size to its co-target, KIR2DL5.

4.02.07.03 PCR optimisation

As with other strategies, PCR optimisation began with a broad annealing temperature gradient PCR (57-72°C) to determine optimal ranges for more refined analysis. However, this was combined with a second variable: ramp rate. Ramp rate is the speed at which PCR thermal cyclers change between the different denaturing, annealing and extending temperature stages of PCR. The ramp rate between denaturing and annealing temperatures of the PCR cycle was reduced to 33% (1.3°C/sec) in a replicate of each reaction. This analysis revealed that, as expected, high annealing/extension temperatures often correlated with improved specificity to KIR targets. However, decreasing the ramp rate, hypothesised to increase specificity, failed to improve the results, and often resulted in decreased concentration of the target amplicon.

Further optimisation was achieved by refining the denaturation and annealing/extension temperatures and durations, as well as cycle number of each PCR stage. This, combined with observations from experiments to determine the optimal polymerase concentration and primer concentration ratios, resulted in the development of PCR protocols to specifically target each of the KIR genes in Table 4.11 (Figure 4.7). The optimised PCR protocols are given in Chapter 2, Section 2.05 and reproduced for convenience, below (Table 4.12 and Table 4.13).

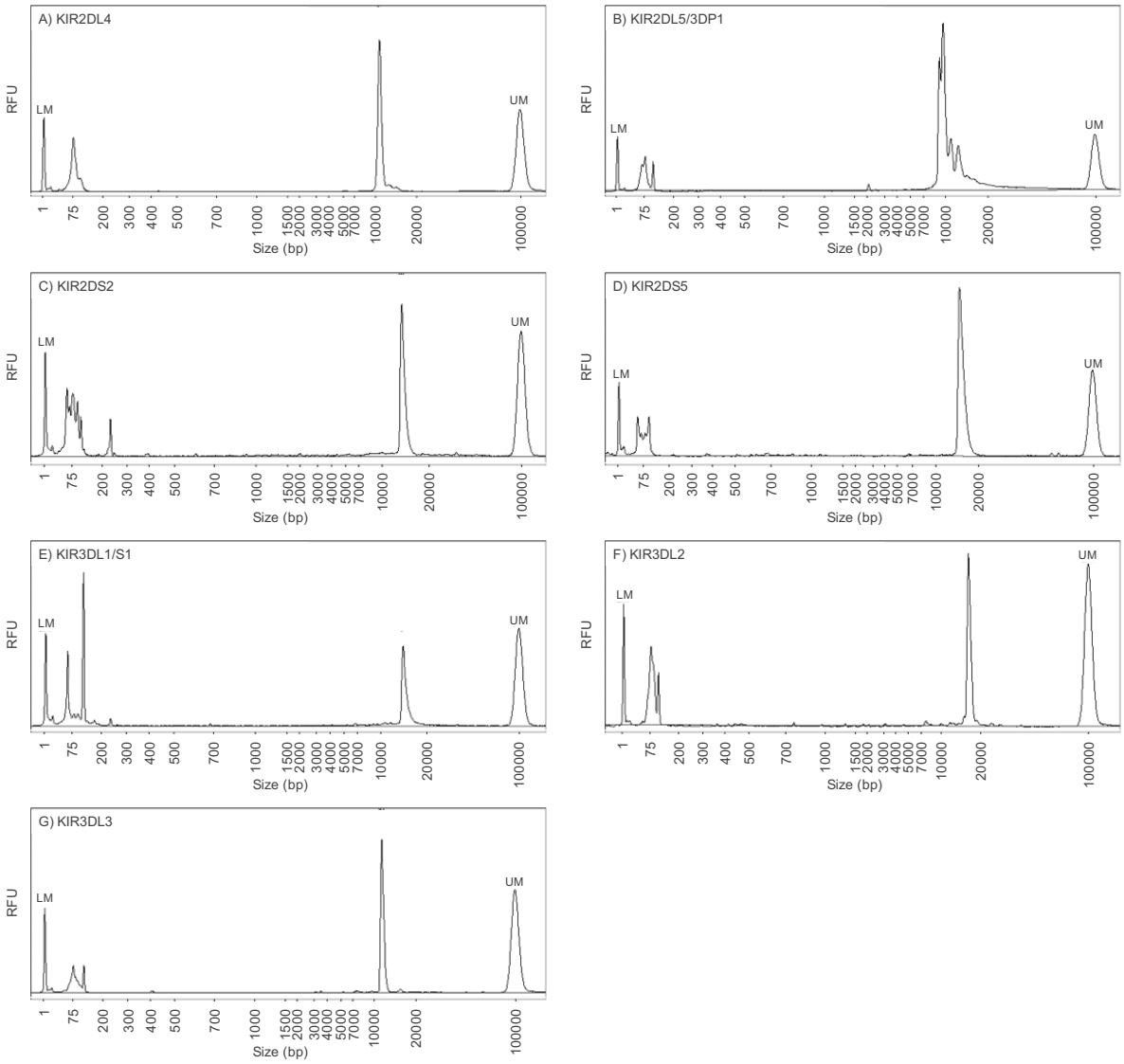


Figure 4.7 Fragment Analyzer analysis indicates targeted amplification of specific KIR targets using a semi-generic PCR strategy

Example electropherogram traces of products from the seven different optimised PCRs utilising a semi-generic KIR amplification strategy. DNA fragments with approximately equal size to the predicted KIR gene product were confirmed in each reaction targeting KIR2DL4 (A), KIR2DL5/3DP1 (B), KIR2DS2 (C), KIR2DS5 (D), KIR3DL1/S1 (E), KIR3DL2 (F) and KIR3DL3 (G). Approximate predicted amplicon sizes for each of these KIR targets were 10.8, 9.4/11.1, 14.5, 15.5, 14.8, 16.8 and 12.0 kbp, respectively.

Table 4.12 Full-length KIR gene PCR amplification cycling conditions

Target	Step	Temperature (°C)	Duration (min:sec)	Repetitions
KIR2DL1	1	95	2:00	1
	2	95	0:20	33
		59	0:20	
		68	10:00	
	3	68	10:00	1
		4	∞	
KIR2DL2/ KIR2DL3	1	95	2:00	1
	2	95	0:20	33
		69	10:00	
	3	69	10:00	1
		4	∞	
KIR2DL4	1	95	2:00	1
	2	95	0:25	34
		72	12:00	
	3	72	10:00	1
		4	∞	
KIR2DL5/ KIR3DP1	1	95	2:00	1
	2	95	0:25	32
		70	12:00	
	3	70	10:00	1
		4	∞	
KIR2DS2	1	95	2:00	1
	2	96	0:30	35
		72	12:00	
	3	72	10:00	1
		4	∞	
KIR2DS5	1	95	2:00	1
	2	95	0:25	11
		74	12:00	
	3	95	0:25	28
		73	12:00	
	4	73	10:00	1
		4	∞	
KIR3DL1/ KIR3DS1	1	95	2:00	1
	2	95	0:10	37
		73.6	10:00	
	3	73.6	10:00	1
		4	∞	
KIR3DL2	1	95	2:00	1
	2	95	0:25	35
		73	12:00	
	3	73	10:00	1
		4	∞	
KIR3DL3	1	95	2:00	1
	2	95	0:25	34
		72	12:00	
	3	72	10:00	1
		4	∞	

Table 4.13 Full-length KIR gene PCR amplification reagents

Reagent (conc.)	Volume (µL)
Water	9.7*
PrimeSTAR GXL Buffer, Mg^{2+} (5X)	5.0
dNTP Mixture (10 mM)	2.0
5′ target primer	1.0
3′ target primer	1.0
5′ control primer	1.0
3′ control primer	1.0
PrimeSTAR GXL DNA Polymerase (1.25 U/µL)	0.3*
Template DNA (25 ng/µL)	4.0

* The reactions targeting KIR2DL1 and KIR3DL1/KIR3DS1 used 0.5 µL PrimeSTAR GXL Taq and 9.5 µL water

4.02.08 High quality DNA controls

As a result of the variation in copy number of each KIR gene, selection of suitable control DNA samples was essential to PCR optimisation. Several IHIW cell lines have been characterised at high resolution for KIR genotype, including some with full haplotype sequences, and thus formed suitable controls [87,94]. A biobank of cryopreserved IHIW cell lines is maintained at Anthony Nolan Research Institute. From this biobank, required cells were selected and thawed. Following cell culture, aliquots were taken for DNA extraction. Alternatively, should a particular cell line with desired KIR genotype not feature within the Anthony Nolan biobank, prepared DNA samples were ordered from the International Histocompatibility Working Group (IHWG, Fred Hutchinson Cancer Research Centre, Seattle, WA, USA). All DNA samples were assessed by agarose gel electrophoresis to confirm the presence of high MW fragments from which long range PCR on large genes could be applied (Figure 4.8). To correlate each sample with its suitability to long range PCR, each sample was also subjected to full length HLA-DPB1 amplification by a published protocol [364]. This revealed that only those samples with the highest MW were amenable to long range PCR.

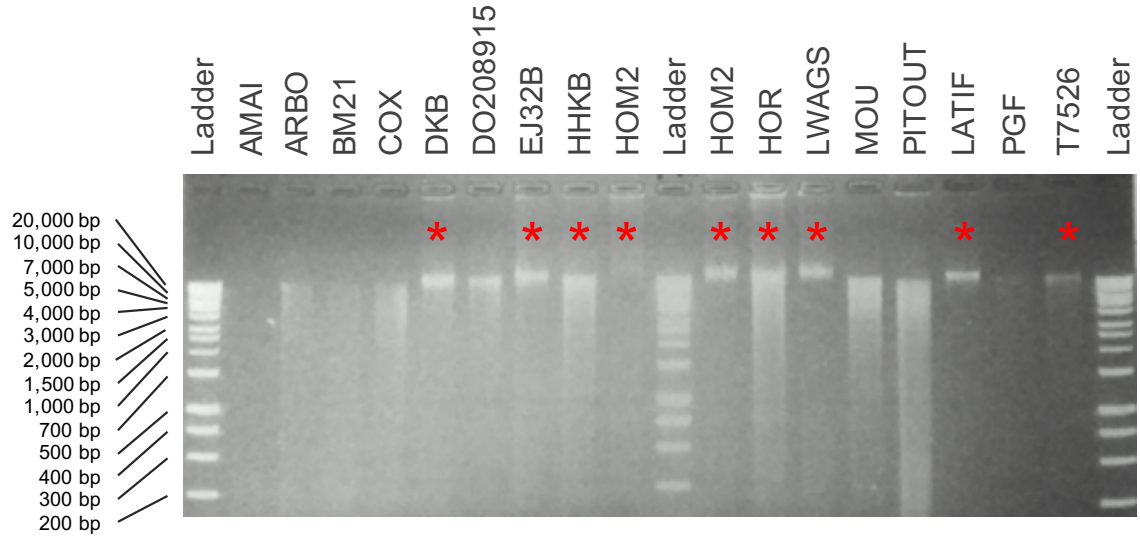


Figure 4.8 Only highest quality template DNA is suitable for long range PCR

Visualisation of an example agarose gel separation of template DNA. Asterisks (*) denote samples that successfully amplified using a published HLA-DPB1 PCR protocol. This demonstrates that only samples with the highest MW fragments are capable of amplifying long range PCR targets. GeneRuler™ 1 kb Plus DNA Ladder was run at either end of the gel. A table of IHIW names and their corresponding IHIW IDs is provided in Supplementary Table A.

4.02.09 Validation of full length KIR PCR amplification by DNA sequencing

Having achieved amplification of PCR products with fragment length matching the estimated amplicon size, to determine that these fragments were actually the desired target required a more detailed characterisation of the amplicon. This was achieved through DNA sequencing.

4.02.09.01 Sanger sequencing confirms amplification of the correct target genes

Confirmation of target specificity for the KIR2DL1 and KIR2DL2/3 was performed by Sanger sequencing of each exon [181] using a panel of IHIW cell lines with previous high resolution allele sequencing performed (Figure 4.9). This revealed that KIR2DL1

was amplified in one reaction and, in the other reaction, evidence of both KIR2DL2 and KIR2DL3 alleles were observed co-amplifying within the same PCR.

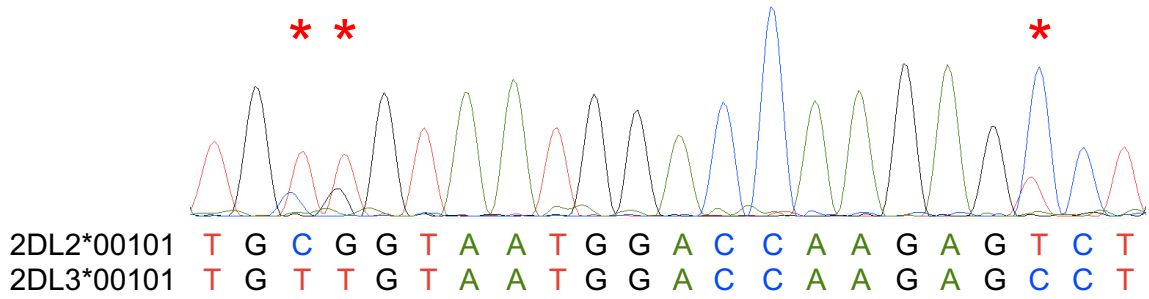


Figure 4.9 Sanger sequencing confirms co-amplification of KIR2DL2 and KIR2DL3

Heterozygous positions (denoted by a red star) within a Sanger sequencing electropherogram over a short region of exon 8 (CDS location using KIR2DL2*00101 as reference allele: 822-843) confirm that both KIR2DL2*00101 and KIR2DL3*00101 alleles of the RSH cell line amplify together within a single PCR.

4.02.09.02 SMRT DNA sequencing confirms specificity of KIR amplification

Although targeted Sanger sequencing determined that the correct genes were present in PCR products, it was not able to rule out co-amplification of other loci. As such, PCR products from individual KIR targets were prepared into libraries and sequenced by SMRT DNA sequencing. The results from these experiments confirmed that, not only were the PCR protocols able to amplify their target, but also that amplification of other genetic loci was not detectable. This suggests that the PCR amplifications are specific. The observed allele types are discussed further in Section 4.05.

4.02.09.03 SMRT DNA sequencing consistently generates accurate consensus sequences

To assess the reproducibility of SMRT DNA sequencing, a single library was prepared (as described in Chapter 2, Section 2.07) from KIR2DL1 amplicons of the same IHIW cell line DNA sample (BOB) using 11 unique barcodes (see Section 4.02.10). This

library was then sequenced on the RS II platform twice, resulting in generation of 22 sequences of the same KIR2DL1 allele with an average read depth of 141. Although the KIR genes in this sample had not been sequenced previously, it was possible to determine that the most homologous published allele sequence was KIR2DL1*0020102. When each of the 22 generated sequences were compared against each other, or against KIR2DL1*0020102, only ten discrepant locations were observed across the entire KIR allele sequence (including non-coding regions). When investigated in greater detail, it was determined that each of the discrepancies was characterised by indel mutations located within intronic homopolymers: genetic structures known to cause difficulty when sequencing by any technique [365]. Furthermore, in eight of the ten discrepant locations, only one of the 22 sequences was discrepant from the remainder, which were concordant with the KIR2DL1*0020102 reference allele. As such, the accuracy and consistency of SMRT DNA sequencing appears to be suitable to typing of polymorphic immunogenetic loci, such as KIR. However, limitations of sequencing through long homopolymer regions, an issue known to affect all current methods of DNA sequencing, results in inconsistent sequencing. As these regions are almost exclusively located within the intronic regions of the KIR genes, it was decided that, for clinical analyses, intronic variation would not be considered.

4.02.09.04 Sanger sequencing confirms substitution derived non-coding polymorphisms

For non-clinical analyses, however, polymorphism within the UTRs and introns of genes were investigated further. Non-coding SNP discrepancies were observed when comparing sequences obtained in this study to previously published allele types or

alleles assigned in haplotype sequencing studies [87]. To verify these discrepancies, an alternative methodology – Sanger sequencing – was employed in several different samples. However, due to the large introns, extended DNA repeat regions and limited published protocols for full length KIR allele sequencing, confirmatory Sanger sequencing of non-coding polymorphisms was limited to only those discrepancies in the regions immediately surrounding KIR exons. Without exception, discrepancies observed as substitution mutations were confirmed (Table 4.14).

As indel mutations were often located away from exon boundaries or in regions that did not permit conclusive Sanger sequencing (see below), it was only possible to perform verification Sanger sequencing on one instance of an indel mutation. This discrepancy, observed in the cell line, QBL, was a four bp (ATAG) deletion at gDNA position 4792 when compared to the most homologous published allele, KIR3DL2*0070102. However, Sanger sequencing revealed no evidence of this deletion and, as such, this discrepancy was rejected. This region of the KIR3DL2 gene is enriched in repeats of ATAG (and similar tetranucleotide repeats). Previous studies have also discovered erroneous indel sequence determination by SMRT DNA sequencing and other sequencing methodologies, particularly within repeat regions similar to this [365].

Table 4.14 Non-coding polymorphisms in KIR allele types from IHIW cell lines confirmed by Sanger sequencing

Sample ID	Most homologous published allele	Discrepancy description	Location
AKIBA	2DL5A*0010103	593G>C	Intron 1
AKIBA	2DL5A*0010103	637A>G	Intron 1
AKIBA	2DL5A*0010103	640A>T	Intron 1
AKIBA	2DL5A*0010103	2616A>T	Intron 3
BH	3DL2*0070102	3669A>G	Intron 4
BH	3DL2*0070102	5558A>G	Intron 5
BH	3DL2*0070102	8915A>C	Intron 6
CB6B	2DL5B*0020101	4838C>G	Intron 5
CB6B	3DL2*0070101	1966A>C	Intron 3
CB6B	3DL2*0070101	5558A>G	Intron 5
CB6B	3DL2*0070101	8915A>C	Intron 6
CB6B	3DL3*0030103	256A>G	Intron 1
DEM	3DL2*0070101	1966A>C	Intron 3
DEM	3DL2*0070101	5558A>G	Intron 5
DEM	3DL2*0070101	8915A>C	Intron 6
DEM	3DL3*0030103	256A>G	Intron 1
HO104	2DL5B*0020107	1521C>T	Intron 2
HO104	3DL2*0070102	3669A>G	Intron 4
HO104	3DL2*0070102	5558A>G	Intron 5
HO104	3DL2*0070102	8915A>C	Intron 6
HO104	3DL3*0030103	256A>G	Intron 1
PGF	3DL3*0030103	256A>G	Intron 1
QBL	3DL2*0070102	3669A>G	Intron 4
QBL	3DL3*0030103	256A>G	Intron 1
RSH	2DL2*0010101	5110G>T	Intron 4
RSH	2DL2*0010101	8869T>C	Intron 6
RSH	2DL2*0010101	12582G>C	Intron 6
RSH	2DL2*0010101	12633G>A	Intron 6
WT47	3DL2*0070101	1966A>C	Intron 3
WT47	3DL2*0070101	5558A>G	Intron 5
WT47	3DL2*0070101	8915A>C	Intron 6
WT47	3DL3*0030103	256A>G	Intron 1

This table describes only those polymorphisms for which it was possible to obtain confirmation by Sanger sequencing, not a complete list of all polymorphisms described in these samples. Discrepancy description combines the gDNA co-ordinate (based on the reference allele for the KIR locus) and SNP description. A table of IHIW names and their corresponding IHIW IDs is provided in Supplementary Table A.

4.02.09.05 Intron 5 is enriched for highly repetitive DNA sequence elements that are challenging to accurately define

Although examples of non-coding polymorphism were observed in each intron and both UTRs of the KIR genes, a high proportion were observed within intron 5. In particular, polymorphism was focussed in a region that contained a microsatellite repeat region

and/or an extended homopolymer. Both of these genetic features represent challenges for any sequencing technology due to their highly repetitive nature.

In an attempt to validate SMRT DNA sequencing over this region, Sanger sequencing primers were designed to cover the microsatellite/homopolymer regions and targeted Sanger sequencing was performed (as described in Chapter 2, Section 2.06.02). Although most sequencing primers were designed to have complete specificity to all known alleles of the individual KIR gene, the KIR2DL3-F primer was designed as an amplification refractory mutation system (ARMS) primer [366]. As such, this primer featured a deliberate mismatch four bp from the 3′ terminus. Where possible, primers were also designed to have similar melting temperatures, approximately 50% GC content and weak predicted secondary structure formation (Table 4.15).

Table 4.15 KIR2DL1 and KIR2DL2/3 repeat region sequencing primer characterisation

Sequencing primer ID	Melting temperature (ºC)	GC content (%)	Predicted secondary structure
KIR2DL2-F	66.7	50.0	Moderate
KIR2DL3-F	68.0	50.0	Moderate
KIR2DL2/3-R	70.0	57.2	Weak
KIR2DL1-F_Microsatellite	64.1	60.0	Weak
KIR2DL1-F_Homopolymer	64.3	50.0	Moderate
KIR2DL1-R_Microsatellite	69.5	55.0	Weak
KIR2DL1-R_Homopolymer	65.2	52.6	Very weak

The relatively short repeat regions of KIR2DL1 were sequenced concordantly between previous sequencing, SMRT DNA sequencing and Sanger sequencing (Figure 4.10A-B). The equivalent regions in KIR2DL2/3, which are significantly longer, were more commonly discrepant between previous and SMRT DNA sequencing. However, when these regions were characterised by Sanger sequencing, suspected polymerase slippage

was observed, prohibiting conclusive determination of repeat number (Figure 4.10C-D). As such, no further attempts at determining exact repeat copy number were performed, and polymorphism observed within this region was regarded with caution.

4.02.10 Barcoded primer optimisation

4.02.10.01 Barcoded DNA primers allow multiplexed KIR sequencing

In contrast to Sanger sequencing, the vast quantity of data that can be generated from a single NGS library is sufficient to sequence multiple amplicons from multiple samples with relative ease. However, the ability to differentiate these sequences from each other following a sequencing run requires the ability to demultiplex. Although several different methods exist to achieve this [183,367], I opted to utilise DNA barcoded primers. These adapt standard oligonucleotide primers by extending the 5′ terminus with a further 21 nucleotide bases. Of these, five form a buffer sequence, helping the remainder of the DNA barcode to resist degradation during the library preparation. Preservation of the other 16 nucleotides is vital as they form a molecular tag that is unique to any individual sample. During PCR reaction setup, each different sample is mixed with a different barcoded primer. This primer is incorporated into the amplicon and is ultimately sequenced. To demultiplex the sequencing data, generated sequence reads are separated according to the specific sequence of the 16 nucleotides at the beginning and end of each read: the DNA barcode. This sequence can then be used to trace back to the original sample to which it was added.

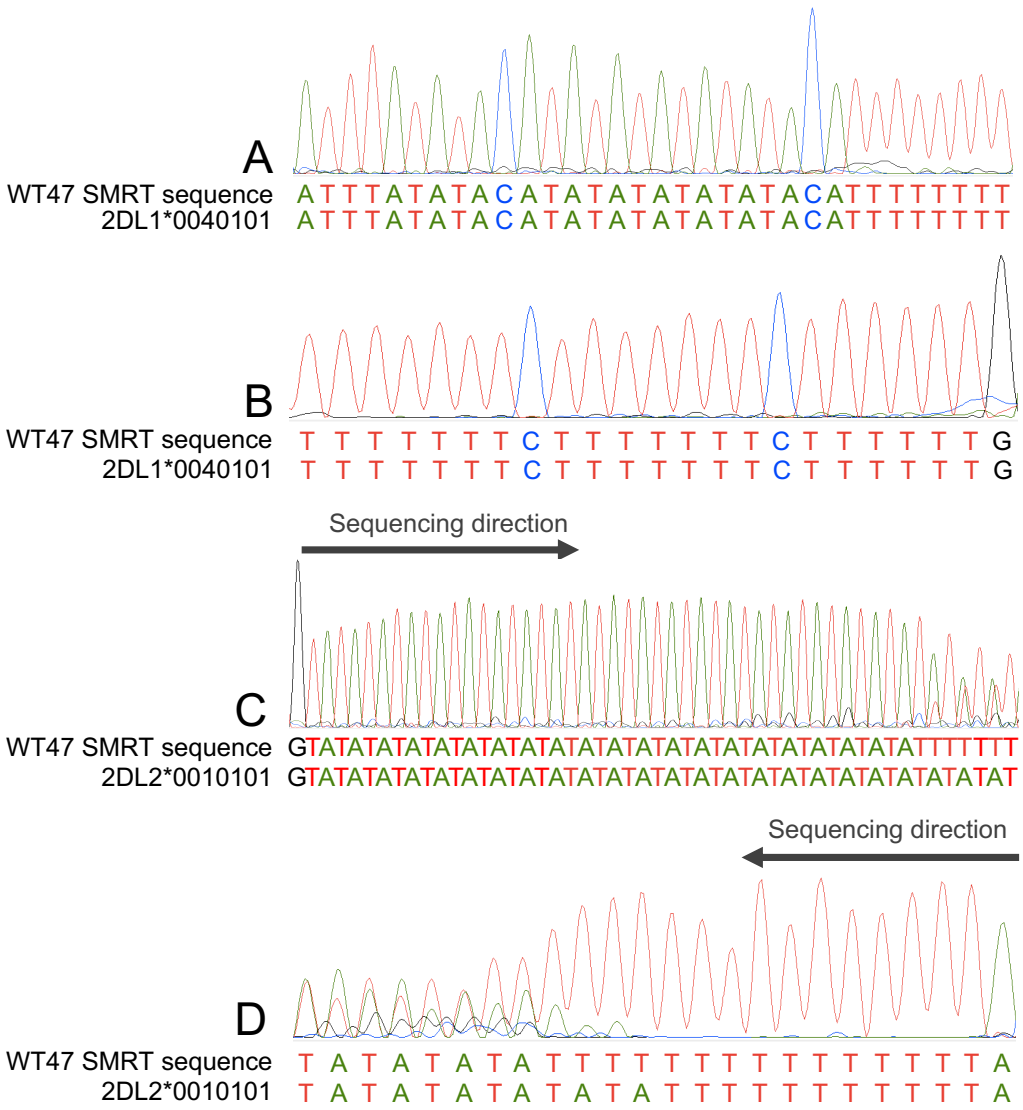


Figure 4.10 Sanger sequencing of intron 5 repeat regions in KIR2DL1 and KIR2DL2/3

Example electropherogram traces generated by Sanger sequencing of WT47 (homozygous) amplicons targeting the repeat regions found within intron 5 of some KIR gene sequences. A) The relatively short 'TA' dinucleotide repeat regions and 'T' 8mer of KIR2DL1*0040101 were accurately sequenced by both SMRT DNA sequencing and Sanger sequencing (region shown is gDNA positions 6380-6412). B) Concordance to previous sequencing was also observed in a separate repeat region of KIR2DL1 (gDNA 6782-6804). C) and D) The longer repeat regions of KIR2DL2*00101 were more challenging to sequence. Enzyme slippage is apparent towards the end of the sequencing reads of both the dinucleotide repeat (C, gDNA 6374-6423) and homopolymer regions (D, gDNA 6419-6438). An alignment of the sequence obtained by SMRT DNA sequencing against the previously published sequences for this region is provided for reference. A table of IHIW names and their corresponding IHIW IDs is provided in Supplementary Table A.

4.02.10.02 Barcoded primer selection

Pacific Biosciences have published 450 unique barcode sequences shown to have limited binding capacity within the human genome. However, predictions on multiplexing capacity implied a much smaller number of different amplicons could be multiplexed together within a single library (see Section 4.03.01), thus reducing the need for so many unique DNA barcoded primers. As such, primer sequences (based on binding site preference) were extended with each of these DNA barcodes and subjected to *in silico* analysis to select optimal DNA barcoded primers (as described in Section 4.02.03). As many of the PCR amplification reactions shared the same 3′ primer mix, the selection of DNA barcoded primers was based on the findings from all KIR loci to utilise a single set of barcodes. DNA barcoded primers with strong predicted secondary structure were avoided where possible. In addition, predicted melting temperatures were matched as closely as possible between genes to support robust PCR.

4.02.10.03 Differences in barcoded primer performance

The difference in yield from different barcodes was first noted in the initial trials of barcoded primers. As at least 50 ng of each amplicon was required for generation of a SMRT library, performance of each barcode was monitored throughout the course of my project and was used to guide barcode selection when setting up amplifications. As such, barcodes that performed well were often utilised far more frequently than poor performing barcodes (Table 4.16). However, as discussed in Section 4.02.08, DNA quality was also a key determinant in amplicon concentration and, as such, barcode variation is not solely responsible for the broad range in average amplicon concentration.

Table 4.16 The five best and worst performing barcoded primers for each clinically-tested KIR locus

KIR2DL1			KIR2DL2/3			KIR3DL1/S1		
Barcode	Frequency of use	Average conc. (ng/µL)	Barcode	Frequency of use	Average conc. (ng/µL)	Barcode	Frequency of use	Average conc. (ng/µL)
233	32	30.6	087	30	17.0	381	19	12.8
068	49	29.6	437	27	15.1	273	38	10.5
289	40	29.5	090	64	13.5	014	40	9.2
369	39	27.0	014	67	11.8	443	50	9.1
239	44	25.2	429	66	11.3	333	50	8.7
033	3	2.7	338	37	3.4	437	6	2.0
309	23	2.5	033	3	1.7	409	5	1.9
266	30	2.4	350	3	1.5	326	3	1.5
350	2	1.7	266	16	1.0	087	43	1.4
059	25	1.5	290	7	0.8	094	9	1.2

Averages were calculated from successful amplicons and excluded failed PCR attempts.

In addition to the different yield of targeted amplicon, certain barcoded primers resulted in formation of a large number of small, contaminating, non-specific PCR products (Figure 4.11). These barcodes were also deselected during future PCR setup. To determine whether something inherent to these barcodes sequences was causing mis-priming at the target region, four separate sequence aligners (Clustal Omega, T-Coffee, multiple sequence comparison by log-expectation [MUSCLE] and MAFFT, EBI, Cambridge) were used to align a range of different barcoded primer sequences (from both specific and non-specific subgroups) against the region surrounding the primer-annealing site. Despite each different aligner creating subtly different alignment results, no apparent differences were observed between the subgroups to indicate barcode sequence as a cause of mis-priming. Subtle differences in predicted primer melting temperature were also excluded from probable causes, as no consistency was observed.

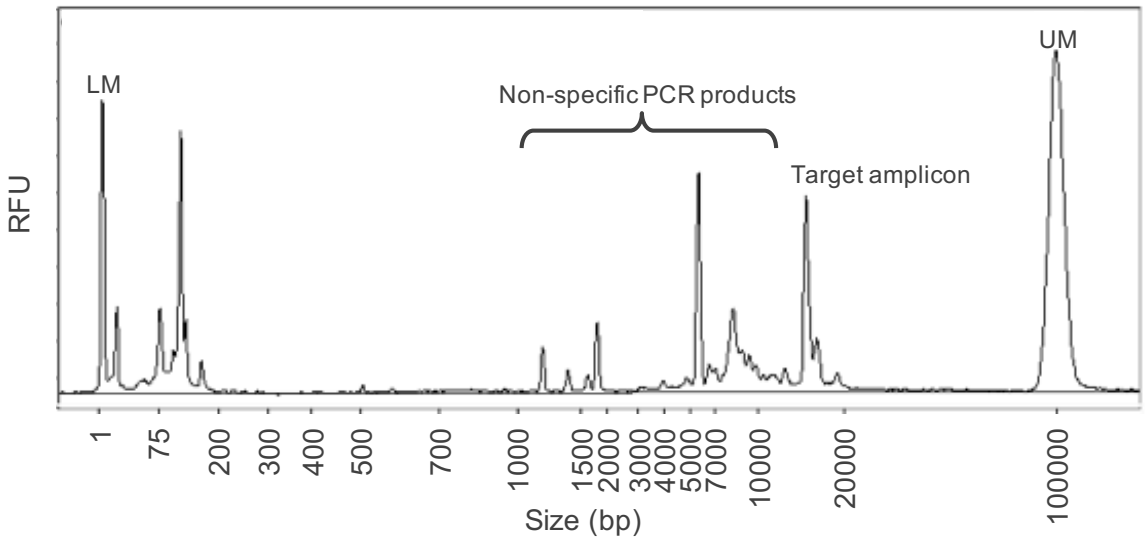


Figure 4.11 Barcode variation can result in production of non-specific PCR products

An example Fragment Analyzer electropherogram trace of a PCR product from a reaction targeting KIR3DL1/S1. Although fragments matching the target MW were generated, high concentration, short fragments from off-target amplification prohibit use of this PCR product. These non-specific products are generated during PCR amplification with only particular DNA barcodes, which were subsequently removed from further use.

4.02.11 Inclusion of an internal control PCR product

Internal controls are included into PCR reactions where a PCR target is not constitutively present. For example, the absence of some KIR genes from certain haplotypes prevents their amplification. However, this result can be differentiated from a failed amplification reaction by the presence of an internal control product.

When considering the use of an internal control within the PCR amplifications described in this chapter, the ability to detect the presence of high quality DNA (suitable for long range PCR) was also desirable. However, the internal control also required removal prior to sequencing to maximise the efficiency of KIR allele sequencing. As such, the internal control product was required to have a significantly different length to the desired products. Several different options were investigated:

i. Short fragments (individual HLA-DRA region, ~300 bp) that are easily removed by AMPure purification but which reveal little of the quality of template DNA;

ii. Intermediate length fragments (full length HLA-C gene, ~3.5 kbp) that are less easily separated from the main target product by AMPure purification, but could theoretically be simply removed by BluePippin size selection;

iii. Longer fragments (full length HLA-DPB1 gene, >13 kbp) that are more difficult to remove by conventional means, including BluePippin size selection, as they are much more similar in length to the main target product. In addition, fragments of similar size to the target PCR amplicon may be difficult to stratify during QC stages.

To assess whether size selection using a BluePippin platform was feasible to remove internal control fragments, several different trials were conducted (see Chapter 2, Section 2.10 for methodology). However, in all scenarios using PCR products or other downstream library products, the recovery yield was too low to permit downstream sequencing. As such, size selection by BluePippin was not explored further. This prevented the use of longer fragment internal controls, as size selection by AMPure purification is not sufficiently precise to remove such closely sized fragments.

Instead, the use of intermediate and short fragments as internal controls was investigated. Initial attempts to incorporate internal control primers within the existing PCR protocol utilised a titration to determine the optimal internal control primer concentration. This revealed that, although each internal control product could amplify

under the KIR PCR amplification protocols, the concentration of KIR target product was compromised. This was particularly evident when using the HLA-C internal control, such that the concentration of target amplicon was reduced by approximately 50% even at low internal control primer concentrations (Table 4.17, Figure 4.12A). Furthermore, attempts to specifically remove the HLA-C internal control fragments from PCR products by AMPure purification size selection also failed, as it was not possible to remove the control without concomitantly removing the target.

Table 4.17 HLA-C internal control primer concentration and corresponding PCR amplicon concentrations

HLA-C control primer (nM)	KIR2DL1			KIR2DL2/3		
	Amplicon	Conc. (pM)	Target amplicon reduction (%)§	Amplicon	Conc. (pM)	Target amplicon reduction (%)§
0	Control Target	0 474	-	Control Target	0 113	-
20	Control Target	130 247	47.9	Control Target	87 50	55.8
30	Control Target	263 188	60.3	Control Target	167 47	58.4
40	Control Target	315 99	79.1	Control Target	214 14	87.6
80	Control Target	620 0	100.0	Control Target	532 0	100.0
200	Control Target	1296 0	100.0	Control Target	998 0	100.0

§ Reduction in target amplicon concentration is based on the target amplicon concentration when compared to the equivalent concentration that resulted from PCRs using 0 nM internal control primer concentration.

However, when smaller internal control targets were utilised, the average reduction in target concentration (when using 40 nM final concentration internal control primer PCR) was limited to only 24.0% across all tested KIR loci (Figure 4.12B, Table 4.18). Interestingly, at 20 nM HLA-DRA internal control primer concentration, the average KIR target amplicon concentration increased by 27.9%. However, amplification of the internal control fragment was weak, preventing use of this lower internal control primer concentration in further experiments. AMPure-based size selection was appropriate when removing smaller internal control fragments. Size selection at 0.4X AMPure bead

concentration allowed retention of approximately 70% of the target amplicon, whilst completely removing smaller internal control products and residual primer. As such, 2 µL of a 0.5 µM primer pair solution, targeting a conserved region of the HLA-DRA gene, was included in each 25 µL PCR (final internal control primer concentration of 40 nM).

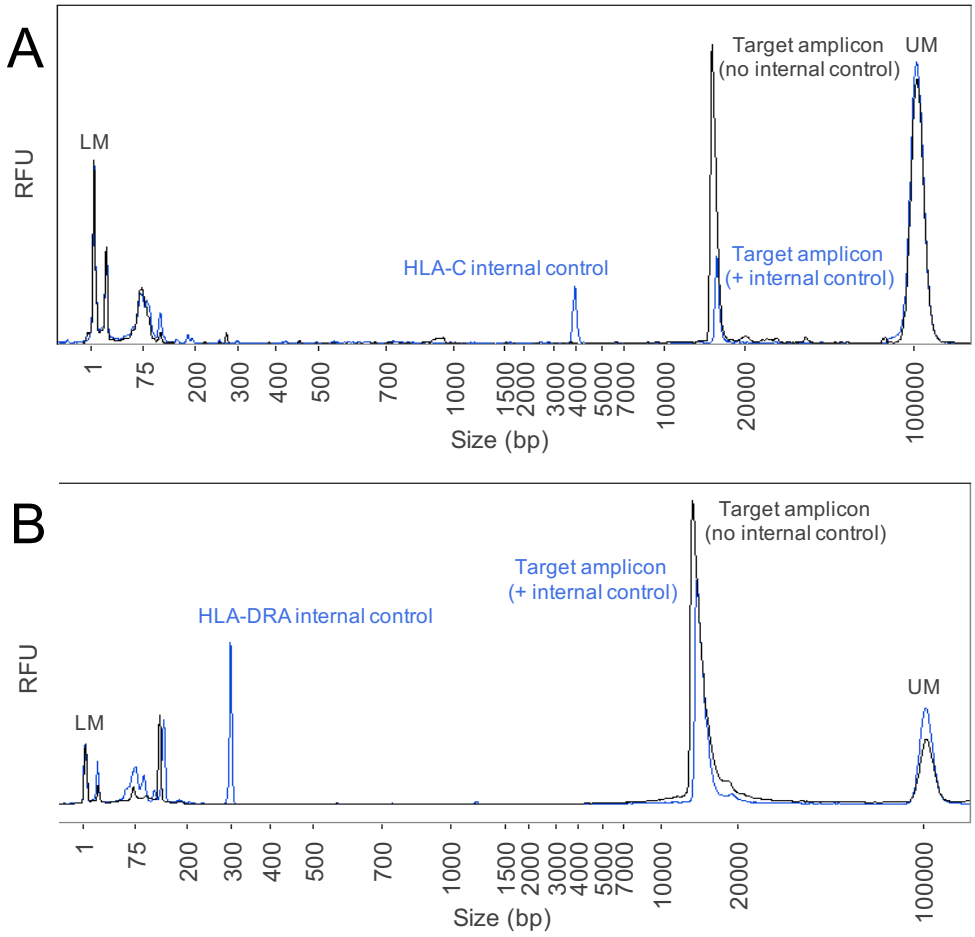


Figure 4.12 Inclusion of an HLA-C internal control primer pair severely impairs target amplicon concentration

Example Fragment Analyzer electropherogram trace overlays demonstrate that the size of internal control fragment affects the efficiency of target amplification. A) Comparison between PCR products from reactions using no internal control primer (black) and 40 nM HLA-C internal control primers (blue) demonstrates significant reduction in target amplicon concentration (no internal control=474 pM *vs* with internal control=99 pM). B) Equivalent reactions set up with the smaller, HLA-DRA fragment internal control primers (40 nM) reveal a much less substantial reduction in target amplicon concentration (no internal control [black]=3218 pM *vs* with internal control [blue]=2123 pM).

Table 4.18 HLA-DRA internal control primer concentration titration and average change in target amplicon concentration

HLA-DRA internal control primer conc. (nM)	Average target conc. (pM)	Target amplicon conc. change (%)[§]
0	1841	-
20	2354	+27.9
40	1399	-24.0
80	1278	-30.6

[§] Target amplicon concentration change is based on the average target amplicon concentration when compared to the equivalent average concentration that resulted from PCRs using 0 nM internal control primer concentration.

4.03 Optimisation of SMRT library preparation

In addition to the optimisation of amplification conditions detailed above, steps to optimise the manufacturer's protocol for library preparation were also undertaken.

4.03.01 Optimisation of multiplexing

The composition of libraries was evaluated to maximise the efficiency of each library. Initial libraries were formed from a combination of KIR2DL1 and KIR2DL2/3 amplicons from ten different DNA samples. Despite using two separate PCR targets, this was relatively simple to perform as similar length fragments load equally onto a SMRTcell during the sequencing stage. As such, straightforward equimolar representation of each amplicon within a library was sufficient to ensure no sequencing bias existed (average read depth per DNA barcode: KIR2DL1=119; KIR2DL2/L3=105).

However, as PCRs for additional KIR targets were optimised, large variation in amplicon length complicated processes for pooling multiple PCR targets, as well as analysing the sequencing results further downstream. As such, subsequent libraries consisted of single PCR targets only. By reducing the number of different KIR loci within individual libraries, inclusion of a greater number of different samples was

permitted. As such, amplicons from up to 20 different DNA samples were pooled into single target libraries. Following discovery of potential amplification imbalance in the PCRs targeting KIR2DL2/3, KIR3DL1/S1 and KIR2DL5/3DP1, multiplexing in these libraries was reduced to only 15 samples per library to maximise the likelihood of allele detection.

To determine the impact that this reduction in multiplexing had on sequence quality, the number of reads and quality scores for accepted KIR3DL1/S1 allele types were compared between samples from 15-plex and 20-plex libraries (Table 4.19). This revealed that, after adjusting for the number of productive ZMWs (P1 value), the number of reads contributing to each allele was increased by almost 40%. Furthermore, read QV and predicted accuracy were unaffected.

Table 4.19 Comparison of adjusted sequencing scores for alleles derived from 15-plex and 20-plex libraries

Sequencing measure	20-plex	15-plex	% change
Average read depth*	148	205	+38%
Average QV	92.60	92.98	+0.4%
Average predicted accuracy	1.00	1.00	No change

*Average read depth was determined following a calculation to adjust for the number of productive ZMWs (P1 value).

4.03.02 Limitation of DNA damage

Damage to DNA molecules, such as nicks or strand breaks, will result in premature termination of SMRT sequencing, as the DNA polymerase molecule is incapable of continuing. Additionally, long DNA molecules, such as those generated by PCR amplifying full length KIR alleles, are prone to becoming damaged during library

preparation. As such, to minimise the number of damaged DNA molecules present in a SMRT library, several stages of the library preparation method were optimised.

4.03.02.01 Reducing exposure of DNA to high temperatures

The exposure of DNA strands to high temperatures, such as those required for PCR, can result in damage to DNA. To reduce the risk of damage accumulating this way, a range of denaturation and strand extension durations were tested to determine the minimum exposure to high temperature during PCR. As such, PCR denaturation was reduced from 30 secs (to 10-25 secs) and from 98ºC to 95ºC. In addition, the thermal cycling extension period of the KIR genes studied in the clinical analysis (KIR2DL1, KIR2DL2/3 and KIR3DL1/S1) was reduced from 1 min/kb (manufacturer recommendation) to 0.67 min/kb to minimise extension duration (Figure 4.13).

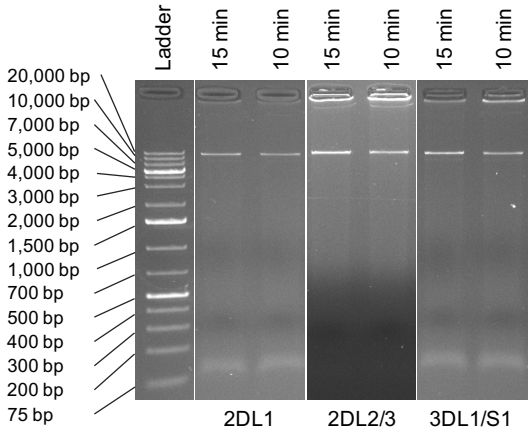


Figure 4.13 Reduction in PCR extension period has no detrimental effect on PCR amplification of KIR2DL1, KIR2DL2/3 or KIR3DL1/S1

Agarose gel electrophoresis reveals no significant impact on specificity of amplification by reducing the extension period of PCRs targeting KIR2DL1 (left), KIR2DL2/3 (middle) or KIR3DL1/S1 (right) to minimise exposure of long DNA fragments to high temperatures. In addition, amplicon concentration, assessed by band fluorescence intensity, does not appear to be significantly affected. Each PCR amplicon is approximately 15 kbp in length. A lane of GeneRuler™ 1 kb Plus DNA Ladder in included for sizing comparison.

In addition to alterations to the PCR methodology, a recommended 10 min incubation at 65ºC to facilitate exonuclease denaturation during library preparation (following ligation of SMRT bell adapters) was substituted with a step to remove exonuclease by AMPure purification [88].

4.03.02.02 Low intensity mixing during AMPure purification

The standard protocol for AMPure purification requires vortexing at 2,000 rpm for a total of 55 minutes over the course of a standard library preparation. This high intensity mixing conveys a high risk of shearing long DNA fragments. As such, each AMPure purification was adjusted to utilise 20 minute bead binding and 10 minute elution incubations on a slow rotator. In addition, only four, rather than five, AMPure purifications were performed per library. Although increasing the overall duration of library preparation by 65 minutes, the risk of excessive DNA damage was reduced.

4.03.02.03 Extending the duration of DNA damage repair

In addition to minimising damage caused, the repair of DNA damage was also optimised. Instead of performing enzymatic DNA damage repair over 20 mins, as recommended by the manufacturer's protocol, it was performed for 90-150 mins following PCR. Furthermore, in some libraries, a second DNA damage repair (>60 mins) was performed following exonuclease treatment to repair damage introduced during library preparation. However, the loss of yield associated with the additional AMPure purification (average recovery yield from each AMPure is approximately 70%) required to remove damage repair enzymes from the library negated the benefit of extra damage repair. As such, this was only performed when the initial library exceeded 2 µg.

4.03.03 Removal of small DNA fragments

In addition to damaged DNA, the efficiency of SMRT DNA sequencing is inhibited by shorter, off-target DNA fragments. Although the relative molarity of each individual fragment may be low, the total can contribute more than 40% of the total concentration of a library (Figure 4.14). To compound this, preferential loading of shorter fragments into the ZMWs of SMRT cells results in fewer ZMWs producing sequences of the desired KIR amplicon [368]. At the time of testing, four different methods were available to perform short fragment removal. These included using reduced AMPure bead concentration (to provide more stringent size selection), agarose gel size selection (Zymoclean Gel DNA Recovery with dual elution, Zymo Research, Irvine, CA, USA), BluePippin size selection and use of Genomic DNA Clean and Concentrator columns (Zymo Research). Unfortunately, all methods resulted in decreased library yield without corresponding benefit, and so were not continued (Table 4.20).

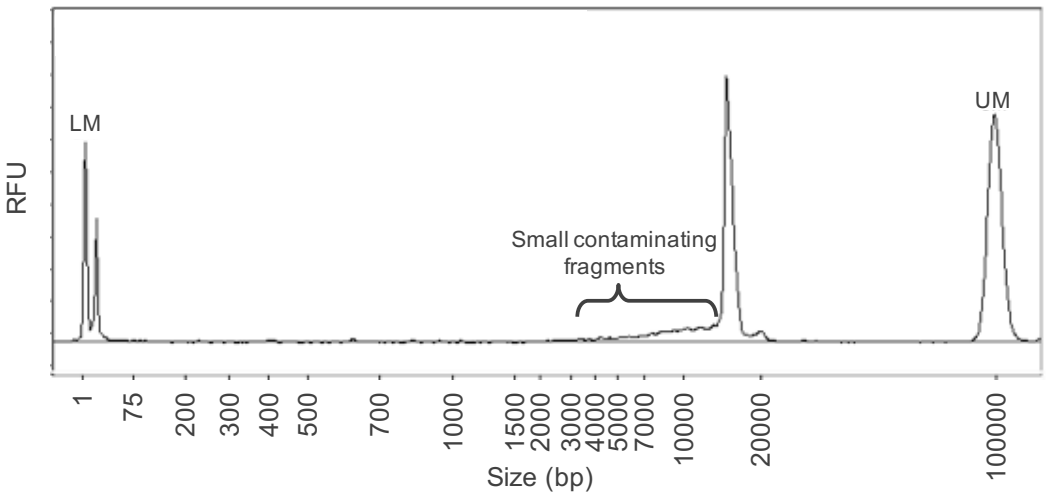


Figure 4.14 Short DNA fragments significantly contribute to final library quality

An example electropherogram trace of a complete library following attempted removal of small, contaminating fragments (by Genomic DNA Clean and Concentrator columns) is shown. Although a large peak, corresponding to target library is observed (4.3 ng/μL, 15,500 bp), shorter DNA fragments are also observed. Although no one fragment contributes a significant concentration, this collection contributes over 40% of the total molarity of the library.

Table 4.20 Recovery from various library size selection purification methods

Method	Contaminating fragments removed?	Specific target recovery (%)
AMPure size selection		
0.1x	Yes	<1%
0.2x	Yes	<1%
0.3x	Yes	1%
0.35x	No	62%
0.4x	No	70%
Agarose size selection		
First elution	Partially	27%
Second elution	Partially	13%
BluePippin size selection	Yes	<1%
Genomic DNA Clean columns	No	28%

4.04 Optimisation of SMRT DNA sequencing

Final optimisation of this full length KIR allele sequencing strategy involved improving the SMRT DNA sequencing stages.

4.04.01 Movie time

Although efforts to reduce DNA damage (described above in Section 4.03.02) were performed to minimise the risk of premature sequencing termination, a more direct method was also applied to increase average sequencing read length. Sequencing run time was extended from four to six hours per library to assess the impact on sequencing read depth. Assuming average sequencing rate of 3.4 bases/sec and average library length of 15 kbp, the duration of a complete circuit of a circularised SMRT library is approximately 2.5 hours. As such, a two hour increase in sequencing time was hypothesised to significantly increase average read length and consensus read depth.

An experiment using sequencing of the same library under multiple conditions confirmed a difference in average read length per ZMW, although the increase was relatively modest (11,180 *vs* 11,889 bases). Despite this, it was evident that increasing the run time dramatically increases the maximum sequencing read length (<48 kbp *vs* >65 kbp, Figure 4.15).

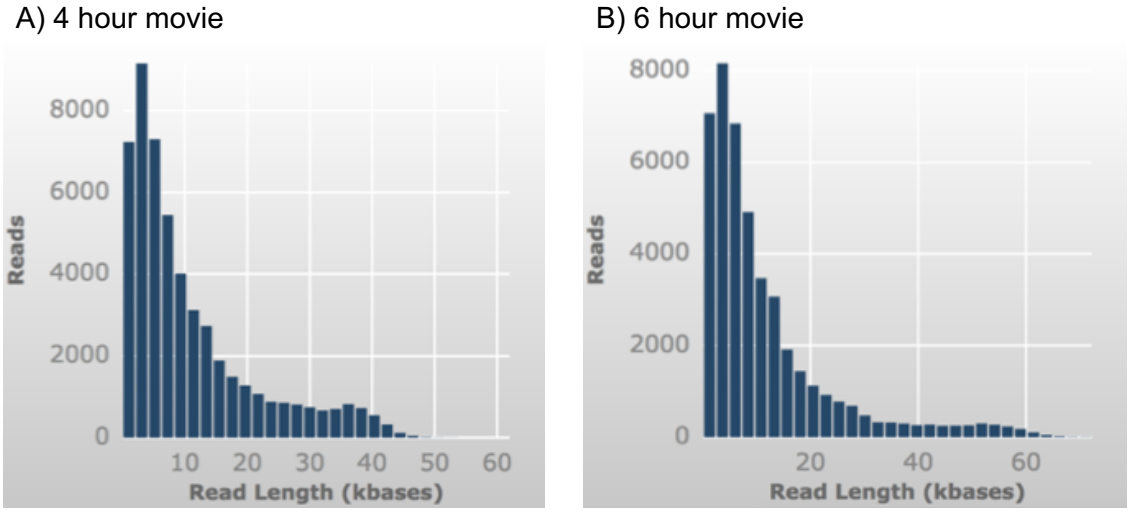


Figure 4.15 Longer movie times generate elongated sequencing reads

When comparing the length of sequencing reads and their frequency between 4 hour (A) and 6 hour (B) sequencing runs, it is possible to observe an increased number of very long (>45 kbp) sequencing reads only obtained with 6 hour movies. Graphics provided on PacBio Dashboard.

4.04.02 Primer concentration

In addition to sequencing run parameters, the chemistry of the sequencing reaction was also optimised. As a part of this, the sequencing primer, to which the sequencing polymerase anneals, underwent a concentration titration. Reduction in sequencing primer was performed as a theoretical method to reduce the proportion of library molecules with primer bound at both ends. As such, the manufacturer's default 20X concentration was tested against 5X and 3X sequencing primer concentrations. The initial results from this titration suggested that reduction in sequencing primer

concentration was only associated with a substantial reduction in the number of productive ZMWs (P1 values: 43%, 28% and 4%, respectively). However, when the number of complete passes of library (defined as a subread with detectable SMRTbell adapter at either end) was normalised against the number of ZMWs sequencing the target amplicon (only ZMWs with subread length equal to or exceeding the minLength parameter), the fraction of complete passes was increased with reduced primer concentration (Figure 4.16). To determine that this was not simply an effect of decreased P1 value, a 5X primer concentration was used with a range of library concentrations to increase the P1 value independently of sequencing primer concentration. This revealed no difference in the number of complete passes.

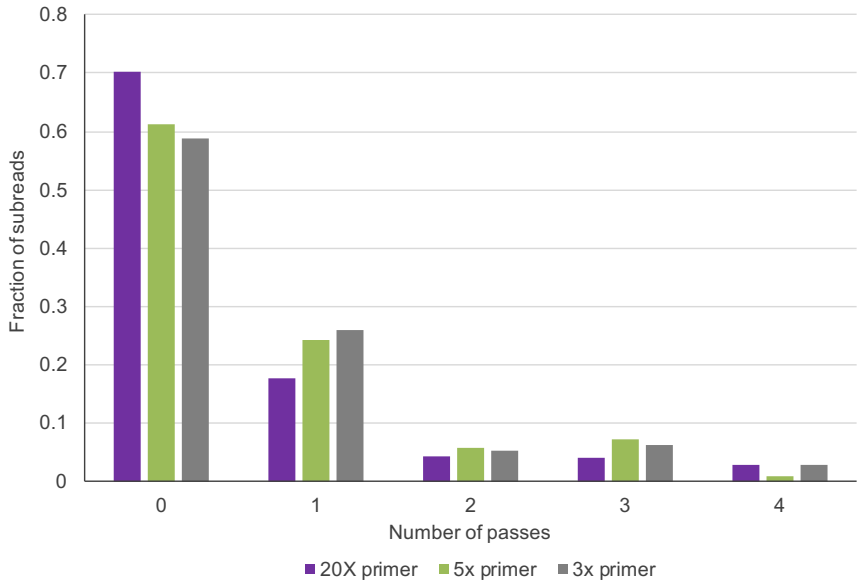


Figure 4.16 Number of complete passes increases with reduced sequencing primer concentration

The fraction of subreads that make complete passes of a library increases as the sequencing primer concentration decreases. It is also of note that this analysis revealed that approximately 60-70% of subreads fail to complete a full pass of KIR amplicons.

4.04.03 Library concentration

In addition to titrating the sequencing initiation primer concentration, the concentration of the library itself was also titrated. Due to the long fragment length of KIR libraries, an increased 'concentration on plate' was required as these fragments load less efficiently than shorter fragments. As such, sequencing runs utilising loading at 6x and 12.5x the default concentration (10 pM) were attempted. Although the average polymerase read lengths were not significantly affected, the quality of the reads was reduced (Table 4.21). This is most striking when comparing the proportion of high *versus* low quality bases within reads of a library (Figure 4.17). When higher loading concentrations were utilised, SMRTcells became overloaded and a much greater proportion of reads had an increased fraction of bases with low quality sequencing scores. As such, titrations using lower concentrations were performed. This revealed that the concentration to achieve optimal average read length and quality values, whilst retaining sufficient read depth, was 35 pM.

Table 4.21 Titration of library 'concentration on plate' in sequencing reactions

	125 pM	60 pM	40 pM	35 pM	20 pM
Average polymerase values					
Read length (bases)	11449	12285	11345	11655	10017
Quality	0.83	0.83	0.83	0.84	0.84
Loading evaluation					
P0 (% empty ZMW)	3	34	38	44	60
P1 (% productive ZMW)	65	54	50	47	35
P2 (% ≥2 molecules per ZMW)	32	12	12	9	5

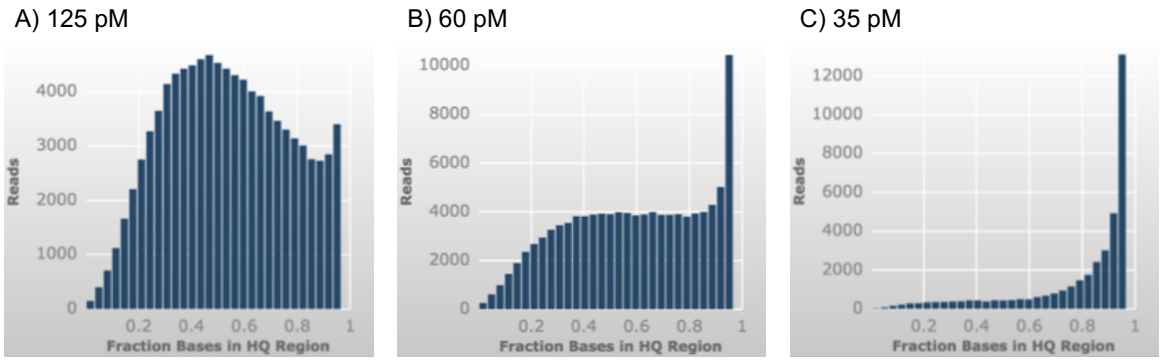


Figure 4.17 Titration of library 'concentration on plate' in sequencing reactions

Loading evaluation metrics (provided on PacBio Dashboard) reveal that, at lower library concentrations, SMRT DNA sequencing results in a higher proportion of bases within each read being within the high quality (HQ) regions.

4.04.04 SMRT cell variation

Unfortunately, considerable variation exists between the performance of individual SMRT cells. As such, sequencing of some libraries was repeated multiple times to establish adequate run characteristics (P1>30%, P2<10%). Although the variation existed between individual SMRT cells, a batch effect was also observed, such that SMRT cells from a certain batch production may be more likely to underload or overload or may have diminished ability to sequence longer fragments effectively. As a consequence, sequencing protocols were often prospectively adjusted according to batch number. For example, the concentration of libraries loaded onto SMRT cells from batches with a tendency to underload were deliberately increased to 45 pM, whilst the library concentration for overloading SMRT cell batches was 25 pM. Experiments to determine batch characteristics were performed in collaboration with Anthony Nolan Histocompatibility Laboratories.

4.05 SMRT DNA sequencing of KIR genes from cell line samples with known allele types

To fully validate the KIR sequencing strategy described above, KIR alleles from cell line DNA samples with known KIR allele types were sequenced [94,95]. A table of cell line names and their corresponding IHIW IDs is provided in Supplementary Table A. KIR allele types determined in previous studies are herein referred to as 'predicted KIR type'. KIR allele types generated as part of my study are referred to as 'observed KIR type'. As the majority of previously analysed samples had only been assessed at two field KIR allele resolution (corresponding to only the CDS), and to remove the complexity resulting from non-coding polymorphism, KIR allele types were abbreviated to two fields for this analysis. A summary of the KIR allele typing from KIR-characterised cell lines is given in Supplementary Table H. A total of 95 unique alleles were observed across the KIR loci tested, and allele typing concordance was 83.3% (349/419 alleles).

4.05.01 Validation of KIR2DL1 allele typing by SMRT DNA sequencing

For KIR2DL1, 41 different IHIW cell line samples with previous allele typing were tested. Of these, the allele types from 38 (93%) samples were concordant with previous analysis (Table 4.22). This corresponded to 64 of 68 (94%) individual alleles being typed concordantly. In addition to these samples with previous KIR allele typing, the IHIW cell line, BOB, was also included in the KIR2DL1 panel. Typing from this cell line revealed the presence of the common allele, KIR2DL1*00201.

However, in the cell line, BOLETH, previous typing had determined a KIR2DL1*001 + KIR2DL1*00302 heterozygote allele resolution genotype, whereas the

KIR2DL1*00201 + KIR2DL1*00302 genotype was observed following SMRT DNA sequencing in my study.

Table 4.22 KIR2DL1 allele concordance across tested IHIW cell lines

Sample	Predicted KIR2DL1 type	Observed KIR2DL1 type
AKIBA	2DL1*00302 + 2DL1*00302	2DL1*00302 + 2DL1*00302
AZH	2DL1*002 + 2DL1*00302	2DL1*00201 + 2DL1*00302
AZL	2DL1*00302	2DL1*00302
BM14	2DL1*00302 + 2DL1*00302	2DL1*00302 + 2DL1*00302
BM15	2DL1*00302 + 2DL1*00302	2DL1*00302 + 2DL1*00302
BOB	Unknown	2DL1*00201
BOLETH	2DL1*001 + 2DL1*00302	2DL1*00201 + 2DL1*00302
CB6B	2DL1*00401	2DL1*00401
DBB	2DL1*00302 + 2DL1*00401	2DL1*00302 + 2DL1*00401
DEM	2DL1*00302 + 2DL1*00401	2DL1*00302 + 2DL1*00401
DEU	2DL1*002	2DL1*00201
EA	2DL1*002	2DL1*00201
EK	2DL1*002 + 2DL1*00302	2DL1*00201 + 2DL1*00302
FH13	2DL1*00302 + 2DL1*00401	2DL1*00302 + 2DL1*00401
FH15	2DL1*00201 + 2DL1*00401	2DL1*00201 + 2DL1*00401
FH5	2DL1*00201 + 2DL1*00401	2DL1*00201 + 2DL1*00401
FH6	2DL1*00302	2DL1*00302
FPAF	2DL1*00401	2DL1*00401
G085	2DL1*00302	2DL1*00302
HO104	2DL1*00302 + 2DL1*00401	2DL1*00302 + 2DL1*00401
JBUSH	2DL1*002 + 2DL1*00302	2DL1*00201 + 2DL1*00302
JO528239	2DL1*00401 + 2DL1*032N	2DL1*00401 + 2DL1*032N
JESTHOM	2DL1*00401	2DL1*00401
JVM	2DL1*00302	2DL1*00302
LUCE	2DL1*00302	2DL1*00302
LZL	2DL1*00302	2DL1*00302
MANIKA	2DL1*00302 + 2DL1*00401	2DL1*00302 + 2DL1*00401
MOU	2DL1*00302 + 2DL1*00302	2DL1*00302 + 2DL1*00302
OMW	2DL1*00303 + 2DL1*012	2DL1*00303 + 2DL1*01201
PGF	2DL1*00302	2DL1*00302
PLH	2DL1*00302 + 2DL1*00302	2DL1*00302 + 2DL1*00302
QBL	2DL1*00302 + 2DL1*00401	2DL1*00302
RML	2DL1*00302	2DL1*00302
RSH	2DL1*00302 + 2DL1*01201	2DL1*00302 + 2DL1*01201
SLE005	2DL1*002 + 2DL1*00302	2DL1*00201 + 2DL1*00302
SPO010	2DL1*00302 + 2DL1*00302	2DL1*037 + 2DL1*037
STEINLIN	2DL1*00302	2DL1*00302
T7526	2DL1*00302 + 2DL1*00302	2DL1*00302 + 2DL1*00302
TISI	2DL1*00302 + 2DL1*00302	2DL1*00302 + 2DL1*00302
WJR076	2DL1*00302 + 2DL1*00401	2DL1*00302 + 2DL1*00401
WT47	2DL1*00401 + 2DL1*00401	2DL1*00401 + 2DL1*00401
WT51	2DL1*002 + 2DL1*00401	2DL1*00201 + 2DL1*00401

Red alleles are discordant between expected and observed allele types

In addition, the predicted KIR2DL1 allele types for the cell line, QBL, from previous analyses were KIR2DL1*00302 + KIR2DL1*00401. However, equivalent analysis using my SMRT DNA sequencing method was not able to identify the presence of a KIR2DL1*00401 sequence.

The final discrepant sample in the KIR2DL1 analysis was SPO010. This sample had previously been typed as homozygous for the KIR2DL1*00302 allele. However, in my analysis, a CDS SNP was observed, 797G>A, that converted the KIR2DL1*00302 allele into the previously undescribed KIR2DL1*037 allele. I also observed this allele in alternative samples from the HCT cohort, and published my findings as a New Allele Alert article [369].

4.05.02 Validation of KIR2DL2/3 allele typing by SMRT DNA sequencing

KIR2DL2 and KIR2DL3 sequencing was performed on amplicons from 34 IHIW cell lines, two of which (OLL and PE117) had not previously undergone published KIR allelic genotyping. In both of these cells, common KIR2DL3 allele types were observed (OLL, KIR2DL3*00101 homozygous; PE117, KIR2DL3*00101 + KIR2DL3*00201).

Of the 32 samples that had previously undergone KIR allele genotyping, 12 (37.5%) were discordant with the previous analysis (Table 4.23). In ten of these discrepant samples (BH, DEU, EA, FH13, FPAF, G085, GRC212, JO528239, SLE005 and WT51), allele dropout was detected when the observed KIR2DL2/3 allelic genotypes were compared to the predicted genotypes. As KIR2DL2*001, KIR2DL2*003 and KIR2DL3*002 were all observed to dropout on multiple occasions, and Sanger sequencing of KIR2DL2/3 amplicons detected unequal electropherogram peak heights

at heterozygous positions (Figure 4.9), amplification imbalance may account for this dropout. However, amplification imbalance was not a consistent observation, as the alleles frequently observed to dropout were also sequenced in combination with other allele types, as in DEM (KIR2DL2*001) PGF (KIR2DL2*003) and AZH (KIR2DL3*002).

Table 4.23 KIR2DL2/3 allele typing concordance across tested IHIW cell lines

Sample	Predicted KIR2DL2/3 type	Observed KIR2DL2/3 type
AZH	2DL3*001 + 2DL3*002	2DL3*00101 + 2DL3*00201
BH	2DL2*003 + 2DL3*001	2DL3*00101
BM14	2DL3*001 + 2DL3*001	2DL3*00101 + 2DL3*00101
BM15	2DL3*001 + 2DL3*001	2DL3*00101 + 2DL3*00101
BTB	2DL3*001 + 2DL3*030	2DL3*00101 + 2DL3*030
DEM	2DL2*001 + 2DL3*001	2DL2*00101 + 2DL3*00101
DEU	2DL2*001 + 2DL3*002	2DL2*00101
EA	2DL2*001 + 2DL3*002	2DL2*00101
EK	2DL3*001 + 2DL3*002	2DL3*00101 + 2DL3*00201
FH13	2DL2*00101 + 2DL3*00101	2DL3*00101
FH5	2DL2*00101 + 2DL3*00201	2DL2*00101 + 2DL3*00201
FH6	2DL2*00301 + 2DL3*00101	2DL2*00301 + 2DL3*00101
FPAF	2DL2*008 + 2DL2*009	2DL2*009
G085	2DL2*00101 + 2DL3*00101	2DL3*00101
GRC212	2DL2*00301 + 2DL3*00101	2DL3*00101
HO104	2DL2*001 + 2DL3*001	2DL2*00101 + 2DL3*00101
JO528239	2DL2*001 + 2DL3*001	2DL3*00101
KOSE	2DL2*003 + 2DL3*001	2DL2*00301 + 2DL3*00110
LO081785	2DL2*001 + 2DL2*003	2DL2*00101 + 2DL3*00101
LUCE	2DL2*00301 + 2DL3*00101	2DL2*00301 + 2DL3*00101
MANIKA	2DL2*001 + 2DL3*001	2DL2*00101 + 2DL3*00101
MOU	2DL3*001 + 2DL3*001	2DL3*00101 + 2DL3*00101
OLL	Unknown	2DL3*00101 + 2DL3*00101
PE117	Unknown	2DL3*00101 + 2DL3*00201
PGF	2DL2*003 + 2DL3*001	2DL2*00301 + 2DL3*00101
RSH	2DL2*001 + 2DL3*001	2DL2*00101 + 2DL3*00101
SCHU	2DL3*001 + 2DL3*001	2DL3*00101 + 2DL3*00101
SLE005	2DL3*001 + 2DL3*002	2DL3*00101
SPO010	2DL3*001 + 2DL3*001	2DL3*00101 + 2DL3*00101
T7526	2DL3*001 + 2DL3*001	2DL3*00101 + 2DL3*00101
WIN	2DL2*001 + 2DL2*001	2DL2*00101 + 2DL2*00301
WJR076	2DL2*001 + 2DL3*001	2DL2*00101 + 2DL3*00101
WT47	2DL2*001 + 2DL2*001	2DL2*00101 + 2DL2*00101
WT51	2DL2*001 + 2DL3*002	2DL3*00201

Red alleles are discordant between expected and observed allele types

In addition to samples affected by allele dropout, LO081785 and WIN were also discrepant to previous published analysis. For LO081785, the predicted allelic genotype was KIR2DL2*002 + KIR2DL2*003. However, the observed allelic KIR2DL2/3 genotype following SMRT DNA sequencing for LO081785 was KIR2DL2*00101 + KIR2DL3*00101. This was substantiated by KIR presence/absence genotyping suggesting that both KIR2DL2 and KIR2DL3 genes were present. In the final IHIW cell line sample with discrepant allelic genotyping, WIN, the expected genotype, KIR2DL2*001 (monotypic), differed to the observed genotype as the KIR2DL2*00301 allele was also detected (in addition to the expected KIR2DL2*001 allele).

4.05.03 Validation of KIR2DL4 allele typing by SMRT DNA sequencing

Twenty-five different IHIW cell lines underwent KIR2DL4 allelic genotyping as part of my study (Table 4.24). Of these, only one sample, EHM, had not previously undergone allelic KIR genotyping. This produced the common KIR genotype, KIR2DL4*00801 + KIR2DL4*00802, also found in the CALOGERO and KOSE cell lines.

In addition to EHM, 24 other IHIW cell lines (each with previous allelic KIR genotyping) were assessed. Within these, discrepancies resulted from SNPs only. In the majority of cases (AZL, JBUSH, MANIKA, WJR076 and WT51), a deletion of a single ‘A’ nucleotide within the long homopolymer at the end of exon 7 (802delA) was responsible for converting the expected allele into the observed novel allele sequence. As these discrepancies occur in repeat regions, it is likely that they are artefacts of SMRT DNA sequencing and, as such, may be discounted. However, this variation is also observed naturally in the sequences of KIR2DL4*007, 008, 009, 011, 013, 017,

019, 020 and 027 alleles. As such, particular care must be taken when analysing sequences at this position.

Table 4.24 KIR2DL4 allele concordance across tested IHIW cell lines

Sample	Predicted KIR2DL4 type	Observed KIR2DL4 type
AKIBA	2DL4*00102 + 2DL4*00501	2DL4*00102 + 2DL4*00501
AZL	2DL4*00102 + 2DL4*00501	2DL4*00102_802delA + 2DL4*00501
CALOGERO	2DL4*00801 + 2DL4*00802	2DL4*00801 + 2DL4*00802
CB6B	2DL4*00501 + 2DL4*00501	2DL4*00501 + 2DL4*00501
DEM	2DL4*00103 + 2DL4*00501	2DL4*00103 + 2DL4*00501
EHM	Unknown	2DL4*00801 + 2DL4*00802
HO104	2DL4*00102 + 2DL4*00501	2DL4*00102 + 2DL4*00501
JBUSH	2DL4*00102 + 2DL4*00103	2DL4*00102_802delA + 2DL4*00103
JO528239	2DL4*00501 + 2DL4*00801	2DL4*00501 + 2DL4*00801
JESTHOM	2DL4*00102 + 2DL4*00801	2DL4*00102 + 2DL4*00801
KOSE	2DL4*00801 + 2DL4*00802	2DL4*00801 + 2DL4*00802
LUCE	2DL4*00102 + 2DL4*00501	2DL4*00102 + 2DL4*00501
MANIKA	2DL4*00103 + 2DL4*00501	2DL4*00103_802delA + 2DL4*00501
MOU	2DL4*00801 + 2DL4*00801	2DL4*00801 + 2DL4*00801
OMW	2DL4*00103 + 2DL4*00501	2DL4*00103 + 2DL4*00501
PF04015	2DL4*011 + 2DL4*011	2DL4*011 + 2DL4*011
PGF	2DL4*00102 + 2DL4*00102	2DL4*00102 + 2DL4*00102
QBL	2DL4*00501 + 2DL4*00802	2DL4*00501 + 2DL4*00802
RML	2DL4*00103 + 2DL4*00501	2DL4*00102 + 2DL4*00501
RSH	2DL4*00103 + 2DL4*011	2DL4*00103 + 2DL4*011
STEINLIN	2DL4*00501 + 2DL4*011	2DL4*00501 + 2DL4*011
WJR076	2DL4*00102 + 2DL4*00501 + 2DL4*011	2DL4*00102_802delA + 2DL4*00501 + 2DL4*011
WT47	2DL4*00501 + 2DL4*00501	2DL4*00501 + 2DL4*00501
WT51	2DL4*00501 + 2DL4*00501	2DL4*00501_802delA + 2DL4*00501_802delA
YAR	2DL4*00102 + 2DL4*011	2DL4*00102 + 2DL4*011

Red alleles are discordant between expected and observed allele types

In the cell line RML, the expected CDS allele genotype (KIR2DL4*00103 + KIR2DL4*00501) was discrepant to the observed genotype (KIR2DL4*00102 + KIR2DL4*00501) by a SNP outside of this region that distinguishes the KIR2DL4*00102 and KIR2DL4*00103 alleles (CDS position 1023A>G, exon 9).

Also of note within this KIR2DL4 sample set was the presence of WJR076, a cell line with three different KIR2DL4 alleles predicted to arise following segment duplication within a haplotype. Presence of each different allele was detected, although the

KIR2DL4*00102 allele featured the 802delA discrepancy, described above. This highlights the ability of the SMRT long amplicon sequencing strategy to identify any number of unique alleles within a sample, although accurate copy number determination is not currently achievable.

4.05.04 Validation of KIR2DL5 allele typing by SMRT DNA sequencing

A total of 21 IHIW cell line samples underwent SMRT long amplicon sequencing for the KIR2DL5 locus, of which 17 were concordant with previously published results (Table 4.25). Two of the discrepant samples (CB6B and WT51) were both previously characterised as encoding three copies of the KIR2DL5 gene (KIR2DL5A*001 + KIR2DL5A*005 + KIR2DL5B*002). However, in both cases, it was not possible to observe the presence of the KIR2DL5A*005 allele. The CDS of the KIR2DL5A*005 allele differs from that of KIR2DL5B*002 by only a single SNP in exon 1 (16A>G), a polymorphism that is also shared by KIR2DL5A*00101.

Although unvalidated as a means of CNV determination in my study, it is also interesting to note that the measured read depth of the observed KIR2DL5A*00101 allele in CB6B (301 reads) was more than double that of the KIR2DL5B*002 allele (122 reads). A similar read ratio was also observed for the WT51 KIR2DL5 alleles, suggesting that future manifestations of bioinformatic analysis may be able to utilise this metric to predict CNV.

In another KIR2DL5 discrepant sample (OMW), previous typing had determined the presence of KIR2DL5A*01202. However, the results from my SMRT long amplicon sequencing strategy implied the presence of only the KIR2DL5*00603 allele. Although

highly homologous across the majority of the their CDS regions (also differing by the 16A>G SNP in exon 1), there are several other differences between these alleles within the 5′ UTR sequence. Unfortunately, intronic sequences are currently unavailable from the reference sequences of both alleles.

Table 4.25 KIR2DL5 allele concordance across tested IHIW cell lines

Sample	Predicted KIR2DL5 type	Observed KIR2DL5 type
AKIBA	2DL5A*001	2DL5A*00101
AZH	2DL5A*001	2DL5A*00101
AZL	2DL5A*001	2DL5A*00101
BOLETH	2DL5A*001	2DL5A*00101
CB6B	2DL5A*001 + 2DL5A*005 + 2DL5B*002	2DL5A*00101 + 2DL5B*00201
DBB	2DL5B*002	2DL5B*00201
DEM	2DL5A*001 + 2DL5B*002	2DL5A*00101 + 2DL5B*00201
EJ32B	2DL5A*001	2DL5A*00101
HO104	2DL5A*001 + 2DL5B*002	2DL5A*00101 + 2DL5B*00201
JO528239	2DL5A*005 + 2DL5B*002	2DL5A*00501 + 2DL5B*00201
JESTHOM	2DL5B*002	2DL5B*00201
LUCE	2DL5A*00101	2DL5A*00101
MANIKA	2DL5A*001 + 2DL5B*002	2DL5A*00501 + 2DL5B*00801
OMW	2DL5A*01202	2DL5B*00603
QBL	2DL5A*001 + 2DL5B*002	2DL5A*00101 + 2DL5B*00201
RML	2DL5A*001	2DL5A*00101
RSH	2DL5B*004	2DL5B*004
STEINLIN	2DL5A*001	2DL5A*00101
WJR076	2DL5B*002	2DL5B*00201
WT47	2DL5A*001 + 2DL5B*002	2DL5A*00101 + 2DL5B*00201
WT51	2DL5A*001 + 2DL5A*005 + 2DL5B*002	2DL5A*00101 + 2DL5B*00201

Red alleles are discordant between expected and observed allele types

The final discrepancy, observed in the MANIKA sample, characterises a common issue in typing polymorphic genes: phase ambiguity. When the exon sequences of the predicted and observed alleles are considered in isolation, it is impossible to distinguish the different allele combinations as the same pattern of six polymorphisms are observed across the entire CDS (Table 4.26) [94]. However, by sequencing the full length of the allele in phase (i.e. as a single long read), as achieved by long read sequencing technologies (including SMRT DNA sequencing), this ambiguity can be removed. In this instance, unambiguous allele typing requires phase to be established across the

large gap between the polymorphisms at CDS positions 16 and 75, corresponding to approximately 1,550 bp of full-length genomic allele sequence when the introns are included.

Table 4.26 Phase ambiguity between the predicted and observed KIR2DL5 alleles in the cell line, MANIKA

	Exon 1	Exon 3			Exon 4	
CDS position	**16**	**75**	**300**	**321**	**517**	**583**
KIR2DL5A*001	A	T	A	A	A	G
KIR2DL5B*002	G	A	G	G	G	A
KIR2DL5A*00501	A	A	G	G	G	A
KIR2DL5B*00801	G	T	A	A	A	G

4.05.05 Validation of KIR2DS2 allele typing by SMRT DNA sequencing

Allele sequencing of KIR2DS2 was performed for a single library of nine different samples (Table 4.27). In each case, the common CDS allele, KIR2DS2*00101 was observed. This was concordant with previous typing for each of the cell line samples. However, it was also possible to observe the presence of the KIR2DS2*00104 CDS allele in the sample, CB6B, in addition to the common KIR2DS2*00101 variant. As such, the high resolution allele typing offered by the SMRT DNA sequencing method designed as part of this project has increased the data available for each of the tested cell line samples.

Table 4.27 KIR2DS2 allele concordance across tested IHIW cell lines

Sample	Predicted KIR2DS2 type	Observed KIR2DS2 type
BH	2DS2*001	2DS2*00101
CB6B	2DS2*001 + 2DS2*001	2DS2*00101 + 2DS2*00104
DEM	2DS2*001	2DS2*00101
HO104	2DS2*001	2DS2*00101
PGF	2DS2*001	2DS2*00101
QBL	2DS2*001	2DS2*00101
RSH	2DS2*001	2DS2*00101
STEINLIN	2DS2*001	2DS2*00101
WT47	2DS2*001 + 2DS2*001	2DS2*00101 + 2DS2*00101

4.05.06 Validation of KIR2DS5 allele typing by SMRT DNA sequencing

The typing results for KIR2DS5 were less promising. Ten different IHIW cell lines underwent attempted KIR2DS5 sequencing (Table 4.28). In each case, successful PCR amplification of a DNA fragment matching the predicted length of the KIR2DS5 amplicon were obtained. However, following library preparation and sequencing, allele types were only generated for five different samples, representing a 50% failure rate. In each of the samples where KIR2DS5 allele sequences were observed, typing was concordant with previous studies. However, failure to detect KIR2DS5 alleles in the remaining samples (AKIBA, BH, QBL, STEINLIN and WT47) may represent limitations in the current bioinformatic analysis of this locus.

Table 4.28 KIR2DS5 allele concordance across tested IHIW cell lines

Sample	Predicted KIR2DS5 type	Observed KIR2DS5 type
AKIBA	2DS5*002	
BH	2DS5*002	
CB6B	2DS5*002	2DS5*00201
DEM	2DS5*002	2DS5*00201
FH8	2DS5*010	2DS5*010
HO104	2DS5*002	2DS5*00201
QBL	2DS5*002	
RSH	2DS5*006	2DS5*006
STEINLIN	2DS5*002	
WT47	2DS5*002	

Red alleles are discordant between expected and observed allele types

4.05.07 Validation of KIR3DL1/S1 allele typing by SMRT DNA sequencing

Allele sequencing of the KIR3DL1/S1 locus was performed on 33 different IHIW cell line samples (Table 4.29). Of these, the allele types obtained following SMRT DNA sequencing were concordant for 27 (82%) of the samples. In addition, the concordance of individual allele types was (90%). In one sample, WT24, the predicted allele type, KIR3DL1*008 + KIR3DS1*013 was radically different to the KIR3DL1*00701 +

KIR3DL1*01502 type that was observed. This may represent an example of experimenter error resulting in sample mix-up.

Table 4.29 KIR3DL1/S1 allele concordance across tested IHIW cell lines

Sample	Predicted KIR3DL1/S1 type	Observed KIR3DL1/S1 type
AKIBA	3DL1*01502 + 3DS1*013	3DL1*01502 + 3DS1*01301
AZH	3DL1*020 + 3DS1*050	3DL1*02001 + 3DS1*050
AZL	3DL1*01502 + 3DS1*013	3DL1*01502 + 3DS1*01301
BM14	3DL1*00501 + 3DL1*01502	3DL1*00501 + 3DL1*01502
BOLETH	3DL1*00401 + 3DS1*013	3DL1*00401 + 3DS1*01301_1106delA
BSM	3DL1*00401 + 3DL1*01502	3DL1*00401 + 3DL1*01502
CB6B	3DS1*013 + 3DS1*013	3DS1*01301 + 3DS1*01301
DEM	3DL1*020 + 3DS1*013	3DL1*02001 + 3DS1*01301
DEU	3DL1*001 + 3DL1*00501	3DL1*00101 + 3DL1*00501
DKB	3DL1*002 + 3DL1*020	3DL1*002 + 3DL1*02001
EA	3DL1*00401 + 3DL1*007	3DL1*00401 + 3DL1*00701
EK	3DL1*00501 + 3DS1*013	3DL1*00501
FPAF	3DL1*00401 + 3DS1*013	3DL1*00401 + 3DS1*01301
G085	3DL1*00501 + 3DL1*00701	3DL1*00501 + 3DL1*00701
HO104	3DL1*01502 + 3DS1*013	3DL1*01502 + 3DS1*01301
JVM	3DL1*001 + 3DL1*008	3DL1*00101 + 3DL1*008
KGU	3DL1*00501 + 3DL1*007	3DL1*00701
KOSE	3DL1*001 + 3DL1*00401	3DL1*00101 + 3DL1*00401
LBF	3DL1*002 + 3DL1*00501	3DL1*002 + 3DL1*00501
LUCE	3DL1*01502 + 3DS1*01301	3DL1*01502 + 3DS1*01301
MOU	3DL1*001 + 3DL1*00401	3DL1*00101 + 3DL1*00401
PGF	3DL1*002 + 3DL1*002	3DL1*002 + 3DL1*002
PLH	3DL1*001 + 3DL1*008	3DL1*00101 + 3DL1*008
QBL	3DL1*00401 + 3DS1*013	3DL1*00401 + 3DS1*01301
RSH	3DL1*00501 + 3DL1*01701	3DL1*00501 + 3DL1*01701
SLE005	3DL1*00501 + 3DL1*01502	3DL1*00501 + 3DL1*01502
SPO010	3DL1*00501 + 3DL1*00501	3DL1*00501 + 3DL1*00501
T7526	3DL1*01502 + 3DS1*013	3DL1*01502 + 3DS1*01301
T7527	3DL1*01502 + 3DS1*013	3DL1*01502 + 3DS1*01301_1106delA
TEM	3DL1*001 + 3DL1*007	3DL1*00101 + 3DL1*00701
WJR076	3DL1*002 + 3DL1*00501 + 3DS1*013	3DL1*002 + 3DL1*00501
WT24	3DL1*008 + 3DS1*013	3DL1*00701 + 3DL1*01502
WT51	3DS1*013 + 3DS1*013	3DS1*01301 + 3DS1*01301

Red alleles are discordant between expected and observed allele types

In four of the other discrepant samples, the difference between predicted and observed allele types was restricted to the KIR3DS1*013 allele. In two cell lines (EK and WJR076), it was not possible to observe the presence of KIR3DS1. As with KIR2DL2/3, Sanger sequencing was employed to assist in determining whether amplification imbalance was responsible for this apparent allele dropout. However,

approximately equal electropherogram peak height implies amplification of KIR3DL1 and KIR3DS1 is balanced (Figure 4.18). This may be indicative that further bioinformatic analysis optimisation is required.

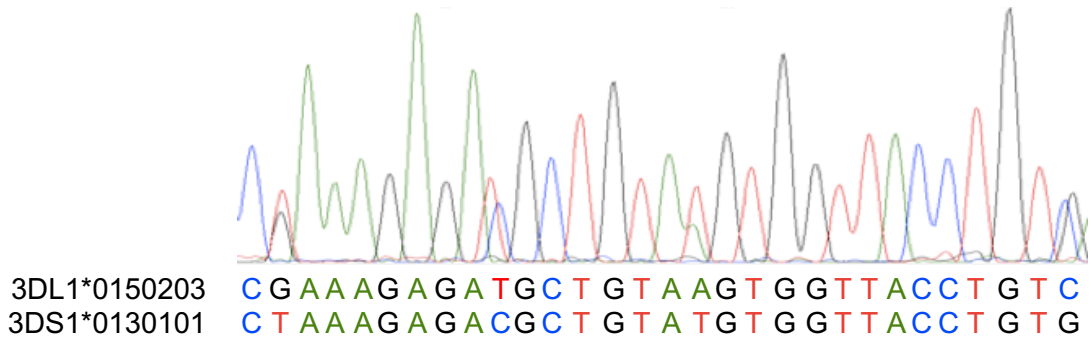


Figure 4.18 Sanger sequencing confirms amplification balance of KIR3DL1 and KIR3DS1 in the IHIW cell line, AKIBA

Heterozygous positions within a Sanger sequencing electropherogram over a short region of intron 6 (gDNA location using KIR3DL1*0150203 as reference allele: 12758-12787) confirm that both KIR3DL1 and KIR3DS1 alleles amplify with approximately equal concentration.

In the remaining two samples with KIR3DS1 discrepancy (BOLETH and T7527), a mutated version (1106delA) of the KIR3DS1*013 allele was observed. As the deletion of an ‘A’ at position 1106 of the CDS is located within the homopolymer region at the end of exon 7, it is possible that this represents sequencing error. However, as with KIR2DL4 above (Section 4.05.03, see comment on 802delA), naturally occurring examples of this deletion polymorphism have been observed previously in the KIR3DS1*010 allele.

The final KIR3DL1/S1 locus discrepancy was observed in the cell line sample, KGU. Here, the predicted KIR3DL1 allele type differed to that observed by the presence of the

KIR3DL1*005 allele. As such, this may also represent an allele dropout scenario, similar to KIR3DS1*013 above.

4.05.08 Validation of KIR3DL2 allele typing by SMRT DNA sequencing

Eleven different IHIW cell line samples underwent KIR3DL2 sequencing to determine allele types (Table 4.30). Discrepancy to previous KIR typing analysis was only observed in one sample, OMW, conferring a 95% allele concordance rate. The predicted KIR3DL2 allele type for the OMW sample was KIR3DL2*00902 + KIR3DL2*028, differing to the observed allele type following SMRT DNA sequencing by the replacement of the KIR3DL2*00902 allele with KIR3DL2*00201. These two alleles differ at three positions across the CDS (456T>C, 474G>T and 893A>G) and, although this discrepancy is unlikely to represent an error in the bioinformatic analysis performed in my study, the cause of this discrepancy is currently unknown. All other KIR3DL2 allele types were concordant between my study and previous analyses.

Table 4.30 KIR3DL2 allele concordance across tested IHIW cell lines

Sample	Predicted KIR3DL2 type	Observed KIR3DL2 type
AKIBA	3DL2*00201 + 3DL2*00701	3DL2*00201 + 3DL2*00701
BH	3DL2*00201 + 3DL2*00701	3DL2*00201 + 3DL2*00701
CB6B	3DL2*00701 + 3DL2*00701	3DL2*00701 + 3DL2*00701
DEM	3DL2*00701 + 3DL2*00902	3DL2*00701 + 3DL2*00902
HO104	3DL2*00201 + 3DL2*00701	3DL2*00201 + 3DL2*00701
MOU	3DL2*01004 + 3DL2*01101	3DL2*01004 + 3DL2*01101
OMW	3DL2*00902 + 3DL2*028	3DL2*00201 + 3DL2*028
PGF	3DL2*00201 + 3DL2*00201	3DL2*00201 + 3DL2*00201
QBL	3DL2*00501 + 3DL2*00701	3DL2*00501 + 3DL2*00701
RSH	3DL2*010 + 3DL2*023	3DL2*01001 + 3DL2*023
WT47	3DL2*00701 + 3DL2*00701	3DL2*00701 + 3DL2*00701

Red alleles are discordant between expected and observed allele types

4.05.09 Validation of KIR3DL3 allele typing by SMRT DNA sequencing

A total of 26 different IHIW cell lines underwent sequencing to determine KIR3DL3 allele types (Table 4.31). Of these, one sample, EHM, typed here as KIR3DL3*00207 +

KIR3DL3*01302, had not previously undergone KIR3DL3 typing and, accordingly, was not included in the concordance values. Of the remaining 25 samples, 8 samples exhibited discordance to previous analysis. In the majority of these samples (DEM, HO104, JHAF, JO528239, KOSE and MANIKA), CDS phase ambiguity, similar to that described for KIR2DL5 (Section 4.05.04, Table 4.26), may explain the discordance.

However, in the cell line sample, CALOGERO, the discrepancy was between the KIR3DL3*017 (predicted) and KIR3DL3*01001 (observed) alleles. These alleles differ at several positions throughout the CDS (69A>G, 168C>T, 229C>T, 502G>A, 869C>A and 961A>C).

Table 4.31 KIR3DL3 allele concordance across tested IHIW cell lines

Sample	Predicted KIR3DL3 type	Observed KIR3DL3 type
AKIBA	3DL3*01002 + 3DL3*01601	3DL3*01002 + 3DL3*01501_969C>T
AZL	3DL3*00402 + 3DL3*00802	3DL3*00402 + 3DL3*00802
CALOGERO	3DL3*00207 + 3DL3*017	3DL3*00207 + 3DL3*01001
CB6B	3DL3*00301 + 3DL3*01403	3DL3*00301 + 3DL3*01403
DBB	3DL3*00801 + 3DL3*020	3DL3*00801 + 3DL3*020
DEM	3DL3*00101 + 3DL3*01403	3DL3*00301 + 3DL3*01303
E481324	3DL3*00206 + 3DL3*01501	3DL3*00206 + 3DL3*01501
EHM	Unknown	3DL3*00207 + 3DL3*01302
HO104	3DL3*00902 + 3DL3*01402	3DL3*00209 + 3DL3*00301
JBUSH	3DL3*01102 + 3DL3*01302	3DL3*01102 + 3DL3*01302
JHAF	3DL3*00901 + 3DL3*026	3DL3*00101 + 3DL3*00801
JO528239	3DL3*00901 + 3DL3*01601	3DL3*00301 + 3DL3*01501
JESTHOM	3DL3*00301 + 3DL3*01404	3DL3*00301 + 3DL3*01404
KOSE	3DL3*00901 + 3DL3*028	3DL3*00301 + 3DL3*01001
MANIKA	3DL3*00903 + 3DL3*01405	3DL3*00210 + 3DL3*00301
MOU	3DL3*00207 + 3DL3*00801	3DL3*00207 + 3DL3*00801
OMW	3DL3*005 + 3DL3*01406	3DL3*005 + 3DL3*01406
PF04015	3DL3*01402 + 3DL3*01402	3DL3*01402 + 3DL3*01402
PGF	3DL3*00301 + 3DL3*00901	3DL3*00301 + 3DL3*00901
QBL	3DL3*00301 + 3DL3*00901	3DL3*00301 + 3DL3*00901
RML	3DL3*00402 + 3DL3*00902	3DL3*00402 + 3DL3*00902
RSH	3DL3*00402 + 3DL3*00901	3DL3*00402 + 3DL3*00901
STEINLIN	3DL3*00206 + 3DL3*007	3DL3*00206 + 3DL3*007
WT47	3DL3*00301 + 3DL3*00301	3DL3*00301 + 3DL3*00301
WT51	3DL3*00103 + 3DL3*036	3DL3*00103 + 3DL3*036
YAR	3DL3*00102 + 3DL3*044	3DL3*00102 + 3DL3*044

Red alleles are discordant between expected and observed allele types

The final KIR3DL3 discrepancy was observed in the cell line sample, AKIBA. Here, SMRT DNA sequencing as part of my study identified a novel CDS variant. This new allele has close CDS homology with KIR3DL3*01501 but featured a C>T substitution at CDS position 969. This was confirmed by repeat amplification and Sanger sequencing (Figure 4.19). Interestingly, this substitution was located in close proximity to two other polymorphisms (961C>T and 971T>C) that differentiate KIR3DL3*01501 from the allele predicted by previous analysis, KIR3DL3*01601. KIR3DL3 allele types from all other samples were concordant with previous analysis.

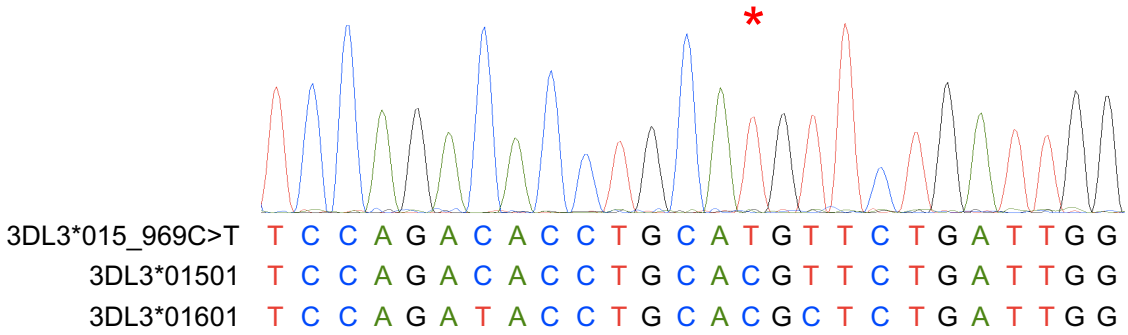


Figure 4.19 Sanger sequencing confirms presence of the novel allele, KIR3DL3*015_969C>T in the IHIW cell line, AKIBA

A Sanger sequencing electropherogram trace of a short region of exon 7 (CDS position 955-980) confirms the 969C>T substitution (denoted by a red star) that defines a novel KIR3DL3 allele in the cell line, AKIBA. A colour-coded sequence alignment of the novel allele, most homologous published allele (KIR3DL3*01501) and predicted allele (KIR3DL3*01601) sequences are also included.

4.05.10 Validation of KIR3DP1 allele typing by SMRT DNA sequencing

To assess the concordance of my KIR3DP1 allele sequencing strategy against the results of previous studies, 23 different IHIW cell samples were tested (Table 4.32). Of these, the cell line EHM had not previously been assessed. Typing of the EHM sample revealed the presence of a novel pseudo CDS (herein referred to simply as CDS) allele with homology to KIR3DP1*00302 but also featuring the substitution, A>C, at CDS

position 479. This was observed in combination with the wildtype KIR3DP1*00302 CDS allele.

Of the 22 samples that had undergone KIR3DP1 allele determination in previous studies, 12 were discordant. The discordance varied between different samples. In both AZL and RML samples, KIR3DP1*007 was predicted but not observed, perhaps indicative of an allelic sequencing imbalance originating during the initial amplification PCR. In addition, the observed AZL allele type featured a KIR3DP1*00302 allele, whereas the predicted allele type included a KIR3DP1*009 allele instead.

Table 4.32 KIR3DP1 allele concordance across tested IHIW cell lines

Sample	Predicted KIR3DP1 type	Observed KIR3DP1 type
AKIBA	3DP1*00302 + 3DP1*010	3DP1*00302 + 3DP1*010
AZH	3DP1*002 + 3DP1*015	3DP1*006_796C>T + 3DP1*00302
AZL	3DP1*007 + 3DP1*009	3DP1*00302
BOLETH	3DP1*00302 + 3DP1*006	3DP1*00302 + 3DP1*006
CB6B	3DP1*00301 + 3DP1*009	3DP1*00301 + 3DP1*00302
DBB	3DP1*00302 + 3DP1*00302	3DP1*00301 + 3DP1*00302
DEM	3DP1*00301 + 3DP1*00302	3DP1*00301 + 3DP1*00302
EHM	Unknown	3DP1*00302_479A>C + 3DP1*00302
EJ32B	3DP1*00302 + 3DP1*00302	3DP1*00302 + 3DP1*00302
HO104	3DP1*00301 + 3DP1*00302	3DP1*00301 + 3DP1*00302
JO528239	3DP1*00301 + 3DP1*00302	3DP1*00301 + 3DP1*00301
JESTHOM	3DP1*00301 + 3DP1*009	3DP1*00301 + 3DP1*00302
LUCE	3DP1*00303 + 3DP1*00901	3DP1*00303 + 3DP1*00901
MANIKA	3DP1*00302 + 3DP1*00302	3DP1*00301 + 3DP1*00302
MOU	3DP1*00302 + 3DP1*00302	3DP1*00302 + 3DP1*00302
OMW	3DP1*00302 + 3DP1*00302	3DP1*00302 + 3DP1*00302
PGF	3DP1*00302 + 3DP1*009	3DP1*00901
QBL	3DP1*00302 + 3DP1*00302	3DP1*00301 + 3DP1*00302
RML	3DP1*00302 + 3DP1*007	3DP1*00302
RSH	3DP1*00304 + 3DP1*008	3DP1*00304 + 3DP1*008
STEINLIN	3DP1*00302 + 3DP1*009	3DP1*00302
WJR076	3DP1*00302 + 3DP1*00302 + 3DP1*009	3DP1*00301 + 3DP1*00302
WT47	3DP1*00301 + 3DP1*00301	3DP1*00301 + 3DP1*00301
WT51	3DP1*00301 + 3DP1*006	3DP1*00301 + 3DP1*006

Red alleles are discordant between expected and observed allele types

The observation of KIR3DP1*00302 alleles in place of predicted KIR3DP1*009 alleles (and *vice versa*) was a recurring theme for several other IHIW cell line samples,

including CB6B, JESTHOM, PGF, STEINLIN and WJR076. Interestingly, the CDS of these alleles differs only in the presence (or absence) or exon 2, the sequence of which is identical to that of exon 2 in other KIR genes.

The presence or absence of exon 2 may also explain discrepant predicted and observed KIR3DP1 allele types for the AZH sample. The predicted KIR3DP1 allele type in this cell line was KIR3DP1*002 + KIR3DP1*015. However, the observed results in my own analysis were KIR3DP1*00302 in combination with a novel CDS variant with strong homology to KIR3PD1*006 (KIR3DP1*00601_796C>T). Interestingly, the novel variation is at the same location where the wildtype KIR3DP1*006 CDS differs to each of the other alleles in question. As such, were the presence of exon 2 to have been mis-assigned during previous published analysis, this sample discrepancy may also represent an example of phase ambiguity being corrected by full length sequencing capable of correctly determining the phase across the intervening region of over 1,800 bp (Table 4.33). Importantly, this analysis also identifies a novel CDS allele that, without correct phase information and full length *de novo* sequence formation, was omitted from previous analysis.

Table 4.33 Phase ambiguity between the predicted and observed KIR3DP1 alleles of the IHIW cell line, AZH

	Exon 2	Exon 4	Exon 5	
CDS position	**+/-**	**479**	**774**	**796**
KIR3DP1*002	+	A	C	T
KIR3DP1*015	-	C	G	T
KIR3DP1*00302	-	A	G	T
KIR3DP1*006_796C>T	-	C	C	T

The remaining discrepancies between the predicted and observed KIR3DP1 allele types are represented by cell lines DBB, JO528239, MANIKA and QBL. In the case of DBB,

MANIKA and QBL, previous typing indicated that the KIR3DP1 type was monotypic for the KIR3DP1*00302 allele. However, the observed allele genotypes also included presence of the KIR3DP1*00301 CDS allele. By contrast, the predicted typing of the cell line, JO528239, indicated heterotypic KIR3DP1*00301 + KIR3DP1*00302 allele type. In observations following SMRT DNA sequencing, however, two different alleles (differing outside of the CDS) of only the KIR3DP1*00301 CDS subgroup were observed in this sample. The CDS of the KIR3DP1*00301 and KIR3DP1*00302 alleles differ by a single substitution, 111C>T.

4.06 Discussion

Until recently, efforts to determine KIR allele sequences have been hampered by both genetic and technological difficulties. The highly complex nature of KIR genes, exhibiting extensive polymorphism whilst retaining high homology between different loci, has made PCR-based amplification of individual KIR genes challenging. This has limited investigation by Sanger sequencing. In addition, the long gene sequences introduce a high cost to traditional sequencing techniques. Revolutionary NGS technologies, allowing the simultaneous sequencing of many DNA fragments from multiple targets, permits the characterisation of complex gene systems and has dramatically reduced the cost per base of DNA sequencing, allowing more investigation into this highly complex locus at a high resolution.

Despite this, PCR-based methods to specifically amplify full length KIR alleles from individual KIR genes remains a difficult task. High homology in UTR sequences between the different KIR loci prevents design of primers with multiple distinguishing bases. Primers I have designed often rely on a SNP at the 3′ terminal base of the primer

to determine specificity. In addition, the frequent recombination events between different KIR loci resulted in a modified amplification strategy to utilise a reverse primer mix theoretically capable of targeting any KIR allele, thus permitting detection of novel alleles formed through intergenic recombination events.

Optimisation of PCR-based amplification of KIR genes utilised a variety of different PCR chemistries and each step of the PCR process was evaluated. Upon achieving amplification of fragments matching the predicted amplicon length for individual KIR loci, specificity of the amplification was confirmed either by Sanger or SMRT DNA sequencing. For each of the KIR gene amplifications tested, it was only possible to detect sequences of those genes that were targeted, confirming specific targeting during PCR.

In addition to optimisation of PCR protocols, the processes of library preparation also underwent alterations to enhance their efficacy with long fragment amplicon libraries. The most appropriate method to assess the effect of each alteration was often to run sequencing reactions. Unfortunately, considerable variation in the performance of individual SMRT cells may mask any beneficial or detrimental effects, thus preventing this being a reliable observation without many replicates. However, sequencing the same library many times has obvious prohibitive financial and temporal implications. As such, alterations were assessed by their effect on library yield or other obvious changes in QC measurements (e.g. the presence or absence of off-target fragments within Fragment Analyzer electropherogram traces). Any alteration that had a theoretical sequencing benefit without detriment observed at any of these stages was adopted in the optimised protocol. As such, less aggressive AMPure mixing, less

exposure to high temperatures and extended DNA damage repair reactions were all implemented. In addition, the degree of multiplexing was adjusted to minimise the risk of allele amplification imbalance resulting in dropout.

The sequencing reactions themselves also underwent optimisation. Six hour movie times were selected to utilise the extended maximum sequence length data generated when compared to four hour sequencing reactions. In addition, the sequencing primer concentration was reduced to limit the number of library molecules with primer bound at each end. This theoretically reduces the risk of two polymerase molecules becoming bound to the same library molecule, and thus limiting the risk of premature sequencing termination caused by polymerase collision. Recent improvements implemented by Pacific Biosciences have advanced this theory by utilising ligation of 'asymmetric adapters' at either end of the amplicon during library preparation stages, thus preventing polymerase binding at both ends of the library [370].

In addition, experiments to determine optimal library concentration showed a strong correlation between overloaded SMRT cells and poor quality sequences. Although variation in SMRT cell performance also impacts these features, the optimisation process established approximate values for underloading, overloading and standard loading SMRT cell batches.

Several limitations to the current KIR sequencing strategy exist. Despite theoretically allowing the amplification of any known or novel KIR allele, the PCR strategy is incomplete without PCR amplifications targeting all KIR genes. At present, KIR2DS1, KIR2DS3, KIR2DS4 and KIR2DP1 are not targeted by optimised PCRs. In addition,

the current amplification strategy for the KIR2DL1, KIR2DL2 and KIR2DL3 genes does not encompass the generic reverse primer mix. As such, further optimisation for these seven KIR genes is required before the amplification strategy is complete. Furthermore, due to lack of available material, the ability to target novel hybrid alleles has not yet been validated.

Another limitation of the current PCR strategy is that, despite every effort being made to select primer binding sites in conserved locations, polymorphism within these regions may exist, potentially preventing the amplification of alleles encoding these polymorphisms. This is compounded by the necessity for very high quality DNA to allow long-range PCR amplification. As well as preventing the use of some samples, particularly those that have been stored for long periods of time, poor quality DNA adds a second cause of amplification failure. As the PCR internal control is not able to confer information relating to the template DNA quality, false negative outcomes resulting from insufficient quality DNA or primer binding site polymorphism are indistinguishable.

Additionally, some DNA barcoded primers resulted in the production of off-target or low concentration amplicons. As such, trials of a greater number of alternative DNA barcodes would be beneficial. Alternatively, utilisation of different multiplexing strategies, such as Pacific Biosciences' barcoded adapters [367], would remove the need for barcoded PCR primers, thus potentially reducing heterogeneity of amplification outcomes.

This may also help to address another limitation of the current strategy, the degree of multiplexing. The ability to sequence more than 20 samples for an individual PCR target within a single library would be greatly beneficial not only in terms of financial benefits, but also in terms of time-saving. At present, the ability to increase the degree of multiplexing is limited by the technology available (RS II SMRT cells have only 150,000 ZMWs), but also the quality of library. Small, off-target products have resisted attempted removal methods, and contribute to the low proportion of sequencing reads being utilised for KIR typing (Figure 4.20).

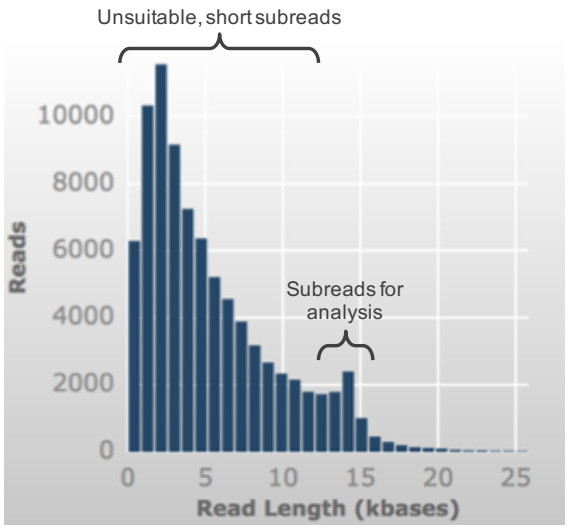


Figure 4.20 The majority of subreads are too short to analyse

This example subread analysis (provided on PacBio Dashboard) demonstrates that, despite the ability to generate full length KIR allele sequence subreads, the majority of SMRT DNA sequencing subreads do not reach sufficient length to pass the stringent minLength parameters required by the current analysis parameters.

Despite the difficulties in their design, optimised sequencing strategies for many of the KIR genes were formed. To validate them as a method to achieve allele level resolution KIR typing, a panel of IHIW cell lines were tested in an attempt to reproduce KIR allele

typing performed in previous publications [87,94]. A high concordance was achieved for many of the samples tested.

Of the discrepancies that were observed in the IHIW cell line typing evaluation, many may be potentially explained by limitations of the short-read sequencing and isolated exon bioinformatic analysis employed by the previously published studies. The inability of short-read technology to reliably establish polymorphism phase across long, homologous regions of the KIR genes was particularly evident in the evaluation of KIR typing concordance of KIR3DL3, where 24% of the allele genotypes appeared to be affected by this problem. In each instance, phase-dependent polymorphisms were separated by several kilobases of genomic sequence (Figure 4.21). A similar issue was observed with the discrepant KIR2DL5 typing of MANIKA.

Figure 4.21 Sequence alignment displaying correction of KIR3DL3 phase ambiguity

A sequence alignment of the alleles previously typed within the HO104 cell line (KIR3DL3*00902 and KIR3DL3*01402) against the alleles generated by SMRT DNA sequencing (KIR3DL3*00209 and KIR3DL3*00301). The gDNA numbering (in accordance with the reference allele, KIR3DL3*00101) is shown above. In this example, it is critical to resolve the phase between polymorphic nucleotides at gDNA positions 5205 and 10726 as no other polymorphisms within the CDS allow the unambiguous determination of correct allele combination.

Additionally, a similar bioinformatic issue arising from short read sequencing and allele assignment based only on individual exon sequences may account for the inability to

detect KIR2DL5A*005 following SMRT DNA sequencing in the CB6B and WT51 IHIW cell line samples that each exhibit three copies of the KIR2DL5 locus. Here, the shared exon 1 sequence of KIR2DL5A*001 and KIR2DL5A*005, and shared exons 2-9 sequences of KIR2DL5A*005 and KIR2DL5B*002 may make the predicted KIR2DL5A*00101 + KIR2DL5A*005 + KIR2DL5B*002 allele genotype indistinguishable from the observed KIR2DL5A*00101 + KIR2DL5B*002 allele genotype.

Examples of potential phase ambiguity also arise when considering exons from multiple KIR genes concomitantly within a single sequencing assay, as achieved by the capture and short-read approach used by Norman *et al.* (2016) [94]. The CDS of the KIR3DL1*00901 allele is identical to KIR3DP1*00302 with the exception of exon 2, which is deleted from KIR3DP1*00302. Furthermore, the sequence of exon 2 from KIR3DP1*00901 is identical to that from a number of other KIR alleles, including all or the majority of KIR2DL1, KIR2DL2, KIR2DL3, KIR2DS1, KIR2DS2, KIR2DP1 as well as a smaller fraction of KIR2DL5 alleles. The sequence homology extends into the surrounding introns [58]. As such, it is possible that the capture and short-read sequencing method that provided many of the predicted KIR allele types in IHIW cell line samples used in my project had falsely assigned exon 2 subread sequences from other genes to KIR3DP1, resulting in erroneous detection of KIR3DP1*009 instead of KIR3DP1*00302 [Personal communication: Prof P. Norman, 2019]. This may explain discordance between the predicted and observed KIR3DP1 allele types in several IHIW cell lines. It should also be noted, however, that KIR3DP1 alleles encoding exon 2 have been demonstrated to associate with haplotype structures missing the KIR2DL1 gene [371]. In each instance of apparent KIR3DP1*009 allele dropout, at least one haplotype

is predicted to not encode the KIR2DL1 gene. As such, further investigation is warranted to clarify these discrepancies.

The issue of incorrectly assigning exon 2 sequence to KIR3DP1 allele sequences may also account for the differences observed in the KIR3DP1 allele typing of the IHIW cell line, AZH. This sample, in which a novel CDS allele was detected by the SMRT DNA sequencing strategy, also highlights the importance of full-length, *de novo* sequence generation from which KIR allele types are derived. By forming the consensus sequence of each allele without any expected polymorphism pattern, and restricted only by a minimum expected sequence length, novel alleles, such as the KIR3DP1*006_796C>T, are less likely to be overlooked. These results highlight the value of using long-read sequencing technologies, such as SMRT DNA sequencing, to characterise full length amplicons. These findings also cast further doubt over the validity of some results from previous publications [94].

However, potential issues relating to amplification imbalance of particular alleles and/or bioinformatic analyses of SMRT DNA sequencing data may also highlight limitations of the KIR sequencing strategy I have designed. In particular, KIR2DL2/3 sequencing failed to generate sequence from KIR2DL2*001, KIR2DL2*003 and KIR2DL3*002 alleles on multiple occasions. In addition, KIR3DS1*013 was also observed to ‘dropout’ on a small number of occasions. Despite this, examples of correct sequencing of these alleles is observed in some other samples. As the affected samples were included in 20-plex libraries, it is likely that the increased multiplexing associated with these samples, and the concurrent decrease in average allele read depth, prevented effective stratification of each allele sequence. By reducing the level of multiplexing

within each library, it was possible to achieve an increased average read depth per allele, suggestive of decreased probability of allele dropout. In addition, by combining the results from low resolution KIR gene presence/absence analysis, the absence of an allele sequence from an expected KIR gene can often be detected and the sample scheduled for repeat sequencing. An additional method to assist in flagging of potential allele dropout is the accurate determination of CNV at each locus. Although the current KIR allele analysis pipeline is not able to perform this function, it was interesting to note that, in both the samples that had previously been determined to encode three alleles of KIR2DL5, one of the two detected alleles had significantly increased read depth compared to the other. As the predicted and observed allelic typing for these cell lines were discrepant, it is not possible to determine whether the read depth ratio in my strategy represents a correct estimation of allele copy balance. However, further investigation into this is warranted in the future.

The absence of KIR2DS5 typing results for 50% of samples tested represents a different form of allele dropout that appears unrelated to PCR amplification imbalance. Although the cause of this problem is still unknown, it is likely that bioinformatic analysis improvements to identify KIR2DS5 allele sequences are required.

In addition to allele dropout discrepancies, issues surrounding homopolymer polymorphism introduced erroneous exonic SNP detection. This was particularly problematic around the ‘A’ homopolymer at the end of exon 7. Although all expressed KIR genes contain this feature, it is extended in both KIR2DL4 and KIR3DS1 [58]. As longer homopolymers are more susceptible to indel errors during amplification and sequencing [365], this may be responsible for the increased frequency of discrepancies

at this location within these genes. Furthermore, polymorphism at this location is known to define a subset of KIR2DL4 alleles with failed expression and function. This lack of function is a consequence of the premature stop codon that results from the deletion mutation [159]. Variation at this position also has a significant effect on KIR3DS1*010, extending the mature protein by several residues [98]. As such, the correct characterisation of this homopolymer is essential to understand the function of KIR alleles. Future sequencing technologies (and the associated bioinformatic analyses) will be required to address this issue.

Despite sequencing errors within homopolymers, KIR allele typing performed as part of this study was able to detect and correct SNP errors in previously published typing results. When comparing the results of SMRT DNA sequencing against the study by Norman *et al.* (2016) [94], exonic SNP discrepancies, such as those observed in the KIR2DL1*037 sequence from the IHIW cell line, SPO010, or the novel KIR3DL3 allele sequence described in the IHIW cell line, AKIBA, were discovered and confirmed by Sanger sequencing [94,369]. Furthermore, although it was not possible to confirm each incidence of non-coding polymorphism that arose during IHIW cell line testing, Sanger sequencing was able to demonstrate that substitution mutations were often determined correctly by SMRT DNA sequencing. This was true in samples that had previously undergone full haplotype analysis (such as WT47, RSH and PGF) [95] as well as those samples featuring differences to their most homologous published allele. It was only possible to perform confirmatory Sanger sequencing on one of the many indel discrepancies. However, it is probable that, particularly in repeat regions such as homopolymers, indel mutations were less accurately determined by SMRT DNA sequencing than substitutions [365]. As repetitive DNA sequence regions are

mostly restricted to introns, clinical analyses focussing on the impact of allelic polymorphism on HCT outcomes (Chapter 6) will exclude these regions.

4.07 Conclusions

A method to successfully target individual (or pairs of) KIR genes using PCR-based amplification has been developed, followed by optimisation of library preparation and SMRT DNA sequencing. This allows the generation of full length, phased KIR allele sequences. In addition, an analysis pipeline has been established to report the most homologous known allele to each sequence, alongside any discrepancies to this sequence. As such, I have obtained a method to type KIR alleles unambiguously. By testing this protocol against a panel of samples with previously determined KIR types, I have confirmed the validity of this method, whilst identifying problems both within the previously determined KIR allele types and my own strategy. Although too few samples exist with published genomic (i.e. third field resolution) KIR allele types to conclude that this method is suitable at this level of resolution, the vast majority of individual nucleotides match those previously published. This supports the hypothesis that the method of PCR amplification and SMRT DNA sequencing described above is suitable for KIR allele typing.

Chapter 5 High resolution characterisation of KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 alleles in a UK population

5.01 Introduction

Hyperpolymorphism at the HLA loci has been well discussed and is predicted to encompass more than one million unique alleles per HLA class I locus within the human population [63,372]. The knowledge of HLA allelic polymorphism is largely the result of high throughput laboratories providing large amounts of HLA allele data for VUD stem cell registries across the world. However, as KIR genes are rarely even assessed for presence or absence prior to HCT, let alone allele type, the extent of investigation into KIR allelic polymorphism is behind that of HLA. Recent analysis on a very large scale does, however, predict that KIR genes are also exceptionally polymorphic [373]. In humans, genes of the KIR B haplotype group appear less susceptible to allelic polymorphism, but exhibit extensive CNV [95]. By contrast, the KIR A haplotype group is more conserved in CNV but demonstrates a higher degree of allelic polymorphism [88]. As such, although the majority of haplotypes may be restricted to a small number of CNV motifs, the number of unique haplotypes, when considering allelic variation at each locus, is considerably increased [88,94,165].

The official repository of KIR allele sequences, the IPD-KIR Database, currently houses almost 1,000 unique allele sequences. Notably, allelic polymorphism is most extensive at the inhibitory KIR loci, and particularly the KIR3DL molecules, where over 90 different protein variants have been described at each locus. Although this may pale in comparison to polymorphism at the some of the classical HLA loci, where thousands of

unique proteins have been discovered at individual loci, the number of samples from which KIR alleles have been sequenced is considerably smaller [98].

In addition to variable allelic content resulting in genotypic differences, the variegated expression of the different KIR genes on the surface of NK cells leads to phenotypic differences. Stochastic expression of KIR is a highly regulated cellular process controlled by three main promoter regions. At the 5′ end of each locus, approximately 1,000 bp upstream of the start codon of most KIR genes, resides a unidirectional distal promoter region responsible for generating transcripts thought to be responsible for enhancing access of transcription factors (TF) to the proximal promoter [374]. Alternatively, distal transcripts have been hypothesised to assist in the correct splicing of proximal promoter transcripts into translatable mRNA as, in samples with reduced distal transcript generation, alternatively spliced proximal transcripts were detected [375]. The distal promoter's own TF, c-myc, is thought to be the key enhancer of expression of the distal transcript, resulting in increased frequency of KIR expression [376].

The proximal promoter, located in closer proximity to coding region (between 20 and 300 bp upstream of the start codon) has several different TF binding sites, giving redundancy to KIR expression and, as such, enabling mature NK cells to maintain continuous KIR expression [377]. The proximal promoter region has bidirectional, probabilistic qualities, meaning that polymorphisms within the TF binding sites can determine the balance of sense and antisense transcription [378]. This, in turn, impacts the probability of expression of that particular KIR gene within the NK cell population. A high proportion of sense transcripts will, understandably, increase the KIR expression

level. By contrast, antisense transcripts form a region of dsRNA covering the proximal promoter site. Following studies in KIR2DL1 and KIR3DL1 promoter function, it has been demonstrated that this region of dsRNA is processed into a 28 bp P-element induced wimpy testis (PIWI)-interacting (pi)RNA. Presence of this piRNA correlates with local KIR gene silencing and is implicated in promoter methylation [379]. As such, gene silencing effects are long-lasting and may explain the variegated expression of KIR across the NK cell subset. In addition to antisense transcripts derived from the proximal promoter, antisense transcripts capable of decreasing KIR expression may also originate from promoter regions in intron 2 of KIR2DL1, KIR3DL1 and KIR3DS1. In these instances, the TF, myeloid zinc finger 1 (MZF-1), is directly involved in instigating antisense transcript generation, and may be assisted by the additional TFs, myc and CCAAT-enhancer-binding protein (C/EBP) [380].

More recently, a third promoter region, termed the intermediate promoter, was discovered first in KIR2DL1 [375] and, subsequently, separate intermediate promoters have been identified in groups of KIR genes with different KIR expression frequencies [166]. As suggested by its name, this promoter region resides between the distal and proximal promoters. Even in samples with decreased distal transcript expression and, as such, decreased sense transcription from the proximal promoter, translatable KIR mRNA transcription from the intermediate promoter is maintained, suggestive of another method of KIR expression redundancy [166]. As polymorphism both between and within promoters of individual KIR genes has been shown to influence expression of KIR by NK cells, further investigation into the polymorphism observed within the sequenced regions of the 5′ UTRs of KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 was warranted.

In this chapter, I have explored the extent of genetic polymorphism exhibited within each of the sequenced promoter regions of these genes, highlighting several interesting incidences where putative TF binding sites have been disrupted. I also present an in depth analysis of the KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 allele typing results from the UK donor-recipient population (405 donor-recipient pairs) of my study and discuss the large amount of novel polymorphism uncovered. Finally, I compare the allele and haplotype motif frequencies of this UK HCT cohort to a similar cohort from the USA, comprised of 506 individuals of European ancestry, the majority of whom were donors in HCT to treat haematological disorders (468 donors, 38 recipients) [165].

5.02 DNA sample availability

As a result of limited DNA availability in some instances, it was not possible to sequence the alleles of every donor and recipient in the UK clinical cohort (n=810). However, 772 (95%) samples were tested for at least one KIR locus. Donor samples accounted for 404 of the 772 total, whilst lack of available DNA sample resulted in only 368 (91%) recipient DNA samples being tested. Although two donor samples did not have sufficient quantity of DNA to test each clinical KIR locus, the remaining 770 samples were tested for each of the three KIR PCR amplification targets (KIR2DL1, KIR2DL2/3 and KIR3DL1/S1).

5.03 Results: KIR2DL1 analysis in a UK HCT population

5.03.01 KIR2DL1 amplification and sequencing success rate

Despite prior KIR presence/absence genotyping revealing that several samples did not possess the KIR2DL1 gene (See Section 3.05), full-length KIR2DL1 amplification was attempted in all available samples (n=771) to rule out the small possibility of novel polymorphism resulting in false detection of KIR2DL1 absence. In each case (n=16), amplification of large DNA fragments was not detected, although the internal control fragment was clearly visible, thus validating the previous typing results.

Of the remaining 755 samples, 738 (98%) successfully amplified the KIR2DL1 target gene. Donor samples accounted for 390 successful amplifications, whilst the remaining 348 were recipient KIR2DL1 allele genotypes. This corresponded to 99% and 97% amplification success rates for donor and recipient samples, respectively, and implies that sample type does not have a significant effect on KIR2DL1 typing efficiency. Five donor samples and 12 recipient samples failed to amplify for KIR2DL1. With only one exception, these samples failed amplification for the other KIR loci tested, suggestive that DNA quality was insufficient to permit long-range PCR.

5.03.02 KIR2DL1 CDS polymorphism

Each successful KIR2DL1 PCR amplification resulted in successful KIR2DL1 allele sequencing. From the 738 samples successfully allele typed for KIR2DL1, a total of 1,225 allele sequences were generated with a mean and median average read depth of 258 and 221, respectively. Although extensive non-coding variation exists (see Section 5.03.03), only 29 unique CDS sequences were observed. Of these, 21 were observed

three or fewer times. Following haplotype estimation to account for haplotypes missing the KIR2DL1 gene (see Section 5.07), approximate allele frequencies were calculated (Figure 5.1). This concluded that five common (>2% allele frequency) CDS allelic variants at the KIR2DL1 locus, including the absence of any KIR2DL1 from a haplotype, constitute over 95% of the total allele frequency. However, despite making up less than 5% of the total known allele frequency, over 80% of KIR2DL1 allelic polymorphism is encoded within rare CDS.

Nineteen samples generated results that required confirmation by repeated sequencing. Allele sequences were considered to be confirmed if two matching sequences were obtained from separate amplicons (either from the same or different samples). Alleles were re-amplified to remove the possibility that errors introduced during the original amplification were re-sequenced. Reasons for repeated sequencing included:

1. Generation of allele sequences encoding a novel CDS,
2. Generation of three or more allele sequences,
3. Generation of sequences for alleles of KIR genes believed to be absent (based on previous KIR genotyping),
4. Absence of sequencing results for alleles of KIR genes believed to be present (based on previous KIR genotyping).

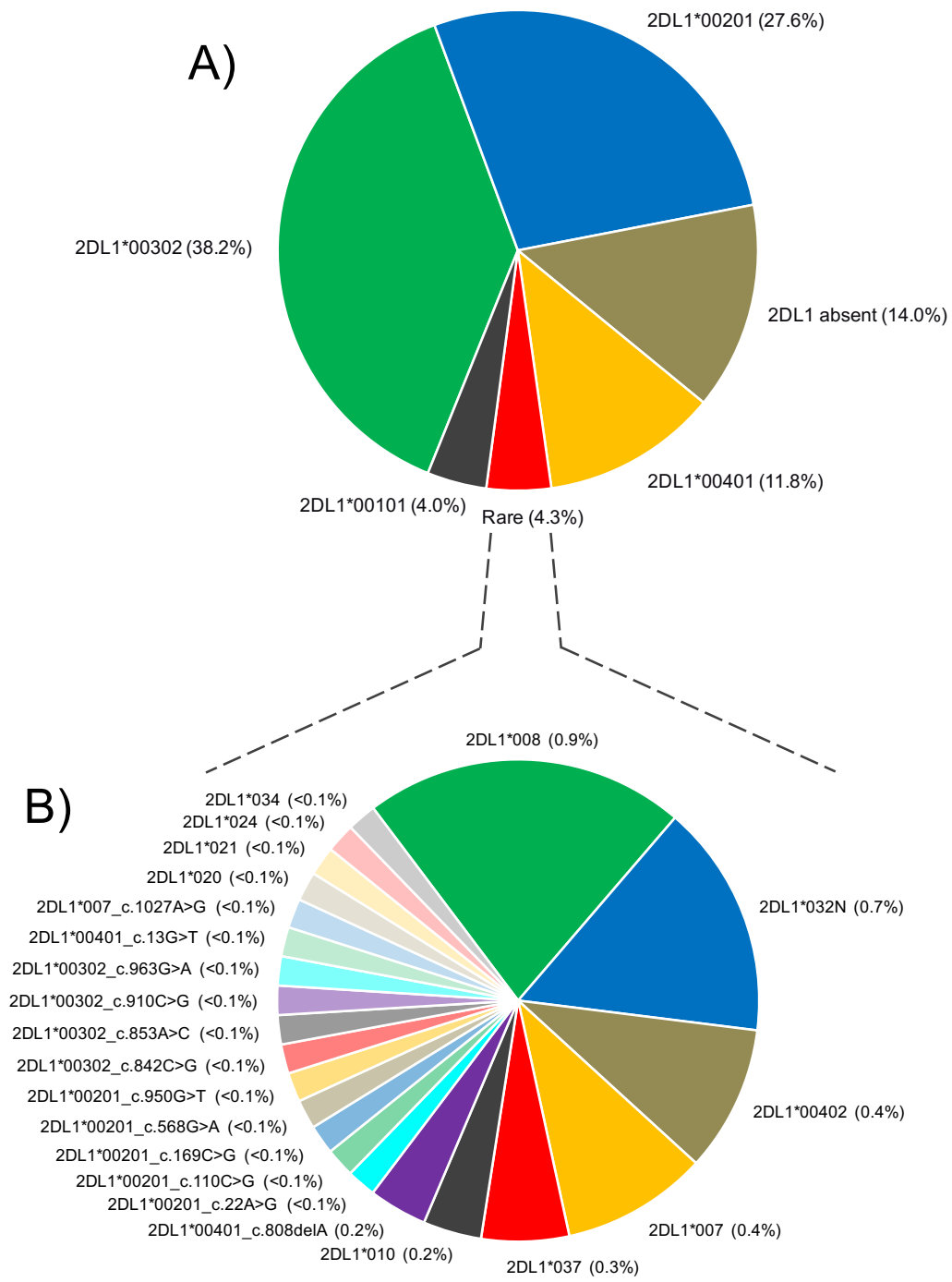


Figure 5.1 KIR2DL1 CDS allele frequency

Pie charts depict the frequency of KIR2DL1 alleles at CDS level resolution in a UK HCT donor-recipient population. A) Haplotypes most commonly contain one of five possible KIR2DL1 variants, including the absence of KIR2DL1. B) The rare allele group, consisting of many more unique alleles, comprises less than 5% of the total allele frequency. Samples required both complete KIR2DL1 and KIR2DL2/3 allele typing for haplotype estimation. As such, not all KIR2DL1 polymorphism is represented, as some samples were missing KIR2DL2/3 allele information. Novel CDS are described by combining their most homologous allele with the CDS mutation description.

In each of the 19 cases, the original sequencing observations were confirmed. This included 16 instances of novel CDS generation and four instances of unusual KIR2DL1 CNV (one sample possessed, in addition to two previously published alleles, a novel KIR2DL1 allele). From the 16 samples encoding novel alleles, 14 unique, novel CDS were determined, as two of the novel alleles were each detected in two separate samples. A summary of the KIR2DL1 CDS resolution typing results is given in Table 5.1.

Several of the novel variations may impact KIR2DL1 protein function, as key amino acid characteristics are changed. For example, one CDS variant with homology to KIR2DL1*00302, but characterised by a C>G polymorphism at cDNA position (c.) 910, converts the otherwise conserved glutamine at residue 283 into glutamate. Although structurally similar, this substitution, from a polar to a charged amino acid, is located within the first ITIM motif of the KIR2DL1 cytoplasmic tail. As such, it is possible that this polymorphism has an effect on the signalling capacity of this particular molecule. Additionally, the deletion of an ‘A’ at c.808 in a novel variant with homology to KIR2DL1*00401 is predicted to cause premature termination of the protein sequence. As such, this frameshift mutation may prevent correct folding and expression of the mature protein.

Table 5.1 Confirmed KIR2DL1 CDS polymorphism

Most homologous allele	CDS polymorphism	Codon	Amino acid polymorphism	Observation frequency
KIR2DL1*00101	-		-	62
KIR2DL1*00201	-		-	375
KIR2DL1*00302	-		-	531
KIR2DL1*00401	-		-	189
KIR2DL1*00402	-		-	7
KIR2DL1*007	-		-	7
KIR2DL1*008	-		-	14
KIR2DL1*010	-		-	2
KIR2DL1*01202	-		-	1
KIR2DL1*020	-		-	2
KIR2DL1*021	-		-	1
KIR2DL1*024	-		-	1
KIR2DL1*032N	-		-	13
KIR2DL1*034	-		-	1
KIR2DL1*037	-		-	3
KIR2DL1*00201	22A>G	-14	Met>Val	1
KIR2DL1*00201	110C>G	16	Pro>Arg	1
KIR2DL1*00201	169C>G	36	His>Asp	1
KIR2DL1*00201	568G>A	169	Ala>Thr	1
KIR2DL1*00201	856A>T	265	Thr>Ser	1
KIR2DL1*00201	950G>T	296	Arg>Leu	1
KIR2DL1*00302	421G>A	120	Ala>Thr	2
KIR2DL1*00302	842C>G	260	Ser>Cys	1
KIR2DL1*00302	853A>C	264	-	1
KIR2DL1*00302	910C>G	283	Gln>Glu	1
KIR2DL1*00302	963G>A	300	-	1
KIR2DL1*00401	13G>T	-17	Val>Phe	1
KIR2DL1*00401	808delA	249	Premature termination[§]	2
KIR2DL1*007	1027A>G	322	Lys>Glu	1

[§] Asn-Lys-Lys-Asn-Ala-Ala-Val>Thr-Lys-Lys-Met-Leu-Arg-STOP

5.03.03 Non-coding polymorphism within the published KIR2DL1 sequence

As discussed in Chapter 4, long homopolymers and other repeat regions are difficult to accurately characterise by any current sequencing methodology. The majority of these repeat regions reside within introns of the KIR genes, permitting accurate CDS determination by exclusion of non-coding sequence. However, many of the allele sequences generated in the UK HCT population differed from any of the existing published allele sequences by only non-coding substitutions. As such, I felt it prudent to evaluate the non-coding variation generated throughout the course of this project, whilst

simultaneously acknowledging that some variation, specifically that located in repeat regions, may be falsely attributed.

Of the 1,225 allele sequences generated from this UK HCT population, 287 unique KIR2DL1 sequences covering the published region (-268 to 14470) were generated. The frequency at which these genomic sequences were observed ranged from 120 occurrences (for KIR2DL1*0030204) to those observed only once. In fact, 64% (184/287) of the unique sequences generated were observed in only one of the samples. Although the majority of these unique sequences remain unconfirmed (having been sequenced from only a single amplicon), it was possible to confirm 120 of the 287 different unique genomic sequences (confirmation was obtained when two separate amplicons – either from the same or different samples – generated matching allele sequences across the full length of the gene). As many of these genomic sequences were observed in multiple samples, they account for 86% (1058/1225) of the total KIR2DL1 sequences generated. Owing to reliability problems in indel detection, only those sequences that differed to their most homologous allele by only substitutions have been submitted to the European nucleotide archive (ENA, https://www.ebi.ac.uk/ena). A complete list of sequence descriptions and, where available, ENA accession numbers, is given in Supplementary Table I.

A summary of the frequency of unique genomic allele observations is provided in Figure 5.2. Five common KIR2DL1*00201 alleles were each observed between 30 and 45 times within the overall 1225 sequences of KIR2DL1, and together comprise almost 15% of the total observed allele sequences. Four common alleles of KIR2DL1*00302 were also observed more than 30 times, with KIR2DL1*0030204 forming the most

commonly observed genomic allele type (observed 120 times). Only a single genomic allele of the KIR2DL1*00401 CDS subgroup (KIR2DL1*0040101) was observed at high frequency. Ten alleles were observed between 10 and 29 times, whilst 83 alleles were observed between 2 and 9 times. Alleles observed only once (n=184) comprise approximately 15% of the total.

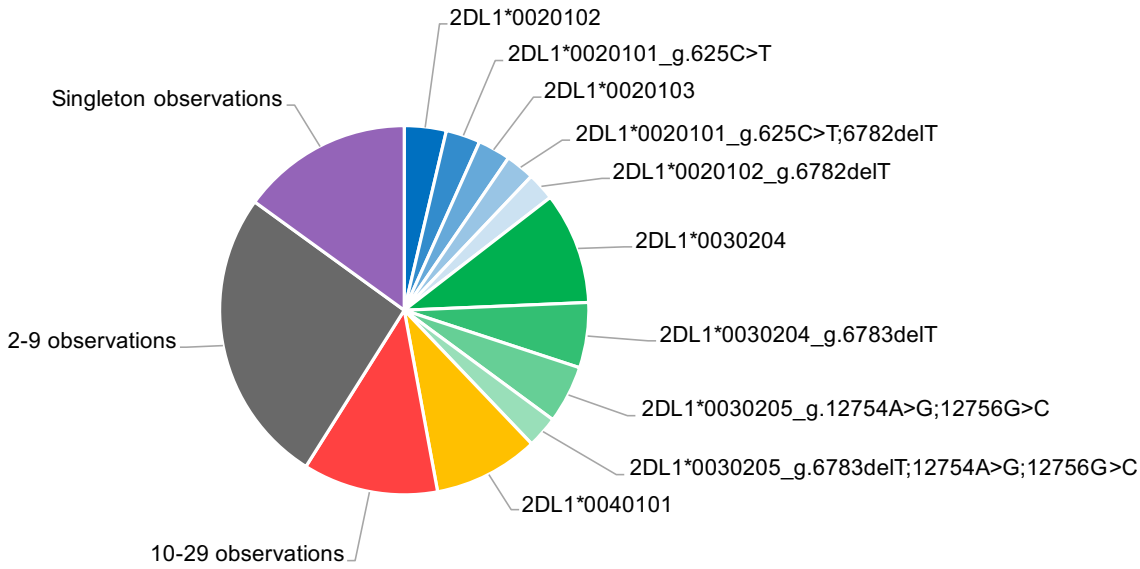


Figure 5.2 Common KIR2DL1 genomic allele observation frequency chart

This pie chart depicts the frequency of observations of unique genomic allele sequences. The details of the 10 most commonly observed KIR2DL1 genomic alleles (the most homologous published allele and, if present, discrepancies to it) are given. Members of the same CDS allele group are shaded accordingly (blue, KIR2DL1*00201; green, KIR2DL1*00302; yellow, KIR2DL1*00401). The remaining unique genomic alleles were grouped according to the frequency of observations.

As this analysis revealed potentially different patterns of non-coding polymorphism between members of the different KIR2DL1 CDS subgroups, non-coding polymorphism within each of the four common KIR2DL1 CDS allele groups was investigated in more detail (Figure 5.3). A broadly similar pattern of polymorphism frequency is observed for the KIR2DL1*00101, 00201 and 00302 CDS allele groups: multiple non-coding variants of each allele group are found at relatively high frequency

(>10% of the total observed for each group) and contribute to a large portion of the total frequency. In contrast, the stratification of the KIR2DL1*00401 CDS subgroup reveals that one allele, KIR2DL1*0040101 contributes 60% (113/189) of observations of this CDS from the entire group, making it the second most commonly observed sequence overall. For comparison, the next most frequent variant of the KIR2DL1*00401 CDS allele group, with homology to KIR2DL1*0040101 but also featuring a G>C substitution at gDNA co-ordinate (g.) 3951, was observed only 13 times (7%). In addition, the most frequently observed variant in all KIR2DL1 CDS allele groups, KIR2DL1*0030204, comprised only 23% of its CDS group. This highlights that the Cen-A-associated allele groups, KIR2DL1*00101, 00201 and 00302, exhibit a greater degree of frequent polymorphism than the KIR2DL1*00401 CDS group, located on the Cen-B haplotype motif.

In all the different KIR2DL1*00101 variants sequenced (n=16), 21 locations were polymorphic. Of these, 17 were characterised by substitutions, and only four by indels. Three of the indels were located in homopolymer or short microsatellite repeat regions (dinucleotide repeats) and variation was found in multiple samples. The fourth, however, was detected in only one sample and characterised by a long deletion at g.716-734 (intron 1, GATATGGGCCTGGATTGGC). Typically, a deletion of this nature is not a common error of sequencing. Thirty-eight other subunits with similar length and high sequence homology to the deleted region are located throughout intron 1 and, as such, this may be an example of naturally occurring repeat region polymorphism.

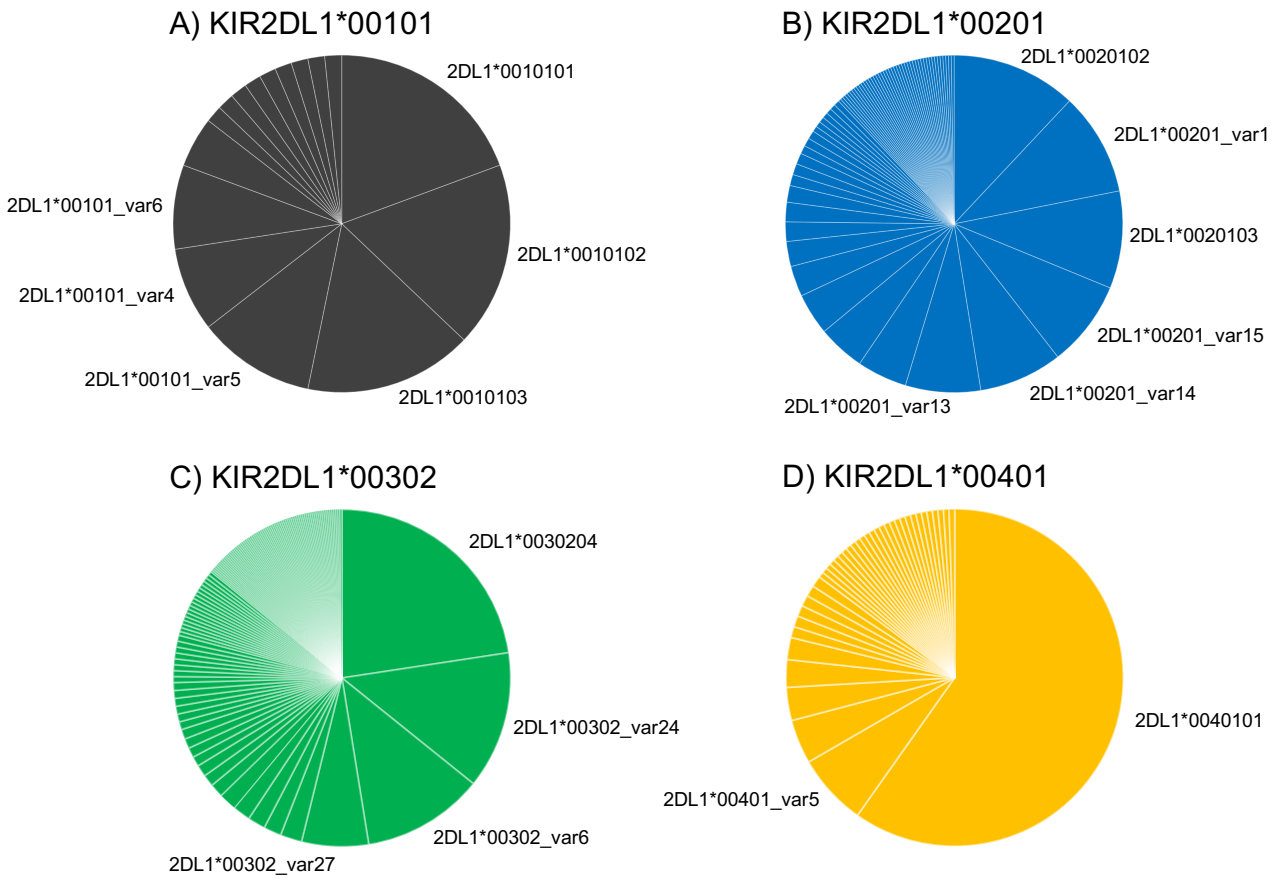


Figure 5.3 Non-coding polymorphism observations within the four common KIR2DL1 CDS allele groups

Pie charts for each common KIR2DL1 CDS allele group demonstrate the extent of non-coding polymorphism. In addition, variants of the KIR2DL1*00101, 00201 and 00302 groups (A-C), which locate on the Cen-A motif, have different patterns of polymorphism frequency compared to KIR2DL1*00401 group variants (D), which are located on the Cen-B motif, suggestive of increased frequency of allelic polymorphism within Cen-A alleles. Pie charts represent 62 KIR2DL1*00101 (16 unique), 375 KIR2DL1*00201 (78 unique), 531 KIR2DL1*00302 (128 unique) and 189 KIR2DL1*00401 (42 unique) sequences, respectively. Alleles contributing more than 5% of the CDS subgroup are labelled. For allele sequences that did not match any existing published sequences, a description (most homologous published allele and discrepancy description) is available in Supplementary Table I.

No such deletions were observed in KIR2DL1*00201 sequences, although six locations with variation in homopolymer or dinucleotide repeat regions were identified. However, an additional 22 polymorphic bases were found to differ by substitutions only. Interestingly, many of these polymorphisms were found as single differences to published sequences, such as an allele most homologous to KIR2DL1*0020101 but also featuring a g.625C>T substitution (this allele was observed 37 times), but also found in combination with other polymorphisms in other novel variants. The g.625C>T polymorphism, for example, was located in 23 alternative KIR2DL1*00201 allele sequences that were observed 100 times in total, as well as in the intronic sequence of several of the alleles featuring novel CDS. Within the 23 KIR2DL1*00201 alleles, several other polymorphisms were identified in multiple different sequences. As such, 72 different KIR2DL1*00201 genomic sequences were generated from combining just 28 different polymorphisms. Although the relatively small size and limited ethnic diversity of this UK HCT cohort is insufficient to find every missing evolutionary 'link', it does provide evidence to how the non-coding regions of KIR genes present in this population may have developed polymorphism over time.

Slightly different patterns of polymorphism accumulation are observed for KIR2DL1*00302 and 00401 subgroups, where 110 sites of substitution and seven indel positions combine to form 123 unique KIR2DL1*00302 sequences (including the previously published KIR2DL1*0030204, 0030205, 0030206, 0030208 and 0030209 alleles) and three indel sites and 25 substitution locations combine to form 28 different KIR2DL1*00401 allele sequences. Although this may suggest that these allele subgroups are composed of predominantly distinct polymorphisms, many of the sequences are formed of multiple unique substitutions, while others combine shared

polymorphisms more similarly to the KIR2DL1*00201 alleles described above. As such, it is likely that many more linking sequences exist in the broader population but have not been sequenced as part of this small cohort.

Aligning all the KIR2DL1 sequences against each other and comparison to the composite sequence formed permitted calculation of the number of polymorphic sites across the published sequence region. This revealed that 494 of a total 14,760 (3.3%) nucleotides were observed to be polymorphic. Of these, 487 (98.6%) represented bases that exhibited dimorphism, whilst the remaining seven represented trimorphic positions. Further investigation suggested that polymorphism was located throughout the gene without particular enrichment in exons (Figure 5.4). In fact, exon 2 was completely conserved in all 1,225 KIR2DL1 sequences. When compared to the composite, 62 (12.6%) of the polymorphic bases were characterised by only indel polymorphisms and all but one of the remaining polymorphic bases (87.2%) exhibited variation by substitution only. One base was shown to differ by either substitution or indel mutation.

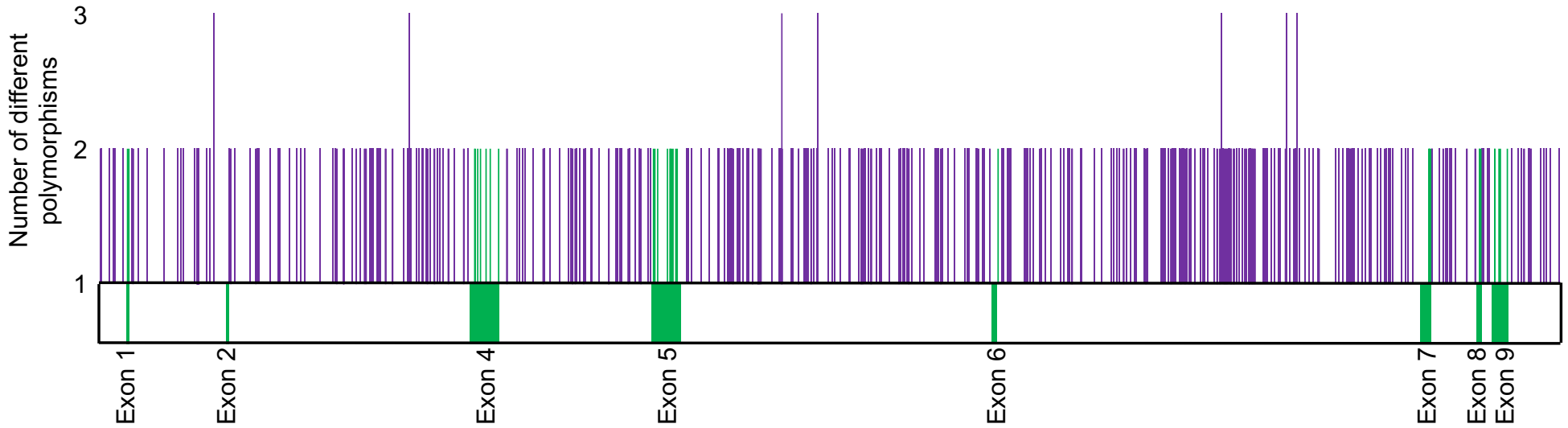


Figure 5.4 Location of polymorphism within the KIR2DL1 gene

A schematic representation of the location of polymorphisms within 1,225 KIR2DL1 sequences generated for a UK HCT cohort reveal relatively even distribution of polymorphism throughout the gene, without particular enrichment in exons (green) or introns (purple). Double height peaks represent trimorphic nucleotide positions. Beneath, in green, the locations of the exons are indicated.

5.04 Results: KIR2DL2/3 analysis in a UK HCT population

5.04.01 KIR2DL2/3 amplification and sequencing success rate

Owing to the lack of available DNA for some samples, the number of samples with attempted amplifications for the KIR2DL2/3 PCR target was 771. As one copy of KIR2DL2 or KIR2DL3 is located on each common KIR haplotype, it was predicted that all samples would generate correctly sized amplicon. However, successful amplification was only obtained in 712 samples, representing an amplification success rate of 92.3%.

All successful amplicons were pooled into libraries and sequenced. The results were then analysed, including a comparison to previous KIR presence/absence data. Following analysis, the success rate was assessed again. At this point, the sequencing status for KIR2DL2 and KIR2DL3 was complete for only 598 samples. This represents a sequencing success rate of 84% for all available amplicons or 77.6% from all available samples. The main causes of failure to complete sequencing for an individual sample included suspected KIR2DL2 dropout (based on discrepancy to previous KIR2DL2 presence/absence genotyping, n=73) failure to generate any KIR2DL2/3 allelic sequence (n=25) and suspected KIR2DL3 dropout (n=8). In each of these cases, repeat KIR presence/absence genotyping was performed and confirmed the original presence/absence genotyping findings. Unconfirmed novel CDS alleles (n=4) or unconfirmed uncommon CNV (three or more KIR2DL2/3 alleles, n=10) also accounted for failure to complete KIR2DL2/3 sequencing in several samples where template DNA stocks were exhausted. In several samples, multiple causes of failure were observed (e.g. suspected KIR2DL2 dropout and unexpected KIR2DL3 CNV). Of the 598

completed samples, 322 (53.8%) were donor samples and 276 (46.2%) were recipient samples.

5.04.02 KIR2DL2/3 CDS polymorphism

From the 598 samples with completed KIR2DL2/3 typing, 1,066 allele sequences were generated, of which 290 were KIR2DL2 alleles and the remaining 776 were KIR2DL3 sequences. The overall mean and median average read depth per allele was 196 and 182 reads, respectively, although the read depth averages of KIR2DL2 alleles (mean=162, median=132) were lower than KIR2DL3 alleles (mean=209, median=200). When compared to the results for KIR2DL1, a relatively small number of unique CDS were observed (n=17, Table 5.2). Only two different KIR2DL2 CDS alleles were observed: KIR2DL2*00101 and 00301. Seven confirmed novel KIR2DL3 CDS (confirmed by repeat sequencing from separate amplicons) were observed in addition to eight previously published KIR2DL3 CDS. Following haplotype estimation (see Section 5.07) accounting for homozygosity at the KIR2DL2/3 locus, five common (allele frequency >2%) KIR2DL2/3 allele CDS were observed, including both the KIR2DL2 alleles and KIR2DL3*00101, 00201 and 00501 (Figure 5.5). The remaining 12 CDS were classed as rare in this UK HCT population, each comprising less than 2% of the total allele frequency.

Each of the novel KIR2DL2/3 CDS variants were observed only once in this cohort. Although molecular assays to determine the functional impact of these novel polymorphisms was not performed, it may be predicted that one of these polymorphisms, observed in the KIR2DL3*00101_c.598G>A allele, may drastically reduce expression of the gene product as the equivalent Gly>Ser mutation in the closely

related KIR2DL1 gene (encoded by the allele, KIR2DL1*014), results in misfolding of the protein and intracellular retention [381].

Table 5.2 Confirmed KIR2DL2/3 CDS polymorphism

Most homologous allele	CDS polymorphism	Codon	Amino acid polymorphism	Observation frequency
KIR2DL2*00101	-	-	-	158
KIR2DL2*00301	-	-	-	132
KIR2DL3*00101	-	-	-	441
KIR2DL3*00110	-	-	-	17
KIR2DL3*00201	-	-	-	251
KIR2DL3*003	-	-	-	8
KIR2DL3*00501	-	-	-	49
KIR2DL3*010	-	-	-	1
KIR2DL3*01202	-	-	-	1
KIR2DL3*030	-	-	-	1
KIR2DL3*00101	142C>A	27	Gln>Lys	1
KIR2DL3*00101	274C>G	71	Gln>Glu	1
KIR2DL3*00101	549C>T	162	-	1
KIR2DL3*00101	598G>A	179	Gly>Ser	1
KIR2DL3*00201	171C>A	36	His>Gln	1
KIR2DL3*00201	618A>C	185	-	1
KIR2DL3*00501	709G>A	216	Glu>Lys	1

Uncommon CNV at the KIR2DL2/3 locus was confirmed in three separate samples. In two instances, two distinct KIR2DL3 allele sequences were observed alongside a single KIR2DL2 allele sequence. In the other, three distinct KIR2DL3 allele sequences were observed in combination with a single KIR2DL2 allele sequence. It was also possible to confirm the presence of KIR2DL2 allele sequence in two samples that had, by previous KIR presence/absence genotyping methods, tested negative for this gene. Repeat KIR presence/absence genotyping was performed on these samples, confirming the findings from the full length sequencing analysis. This indicates that false negative tests results were generated during the original genotyping assay.

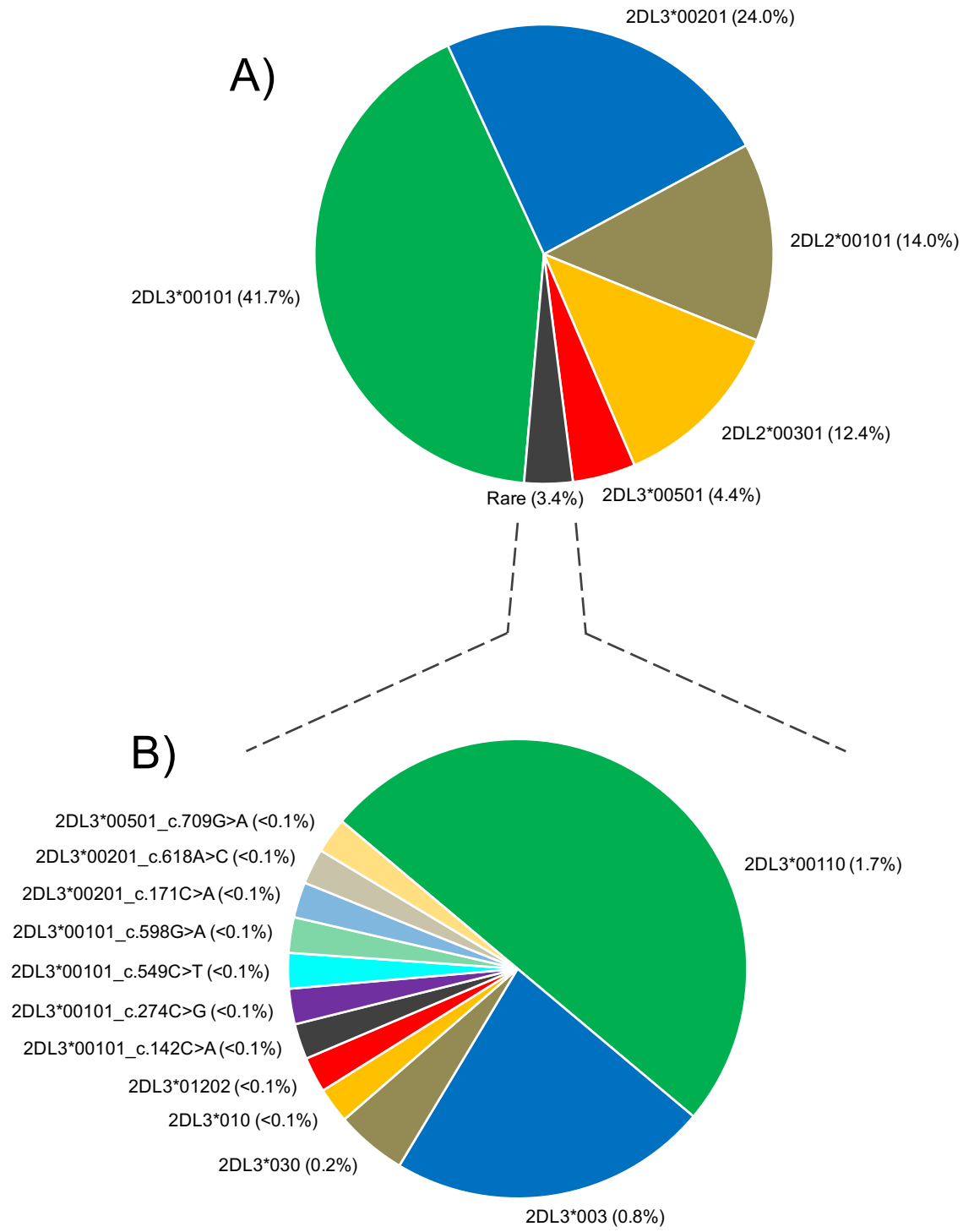


Figure 5.5 KIR2DL2/3 CDS allele frequency

Pie charts depict the frequency of KIR2DL2/3 alleles at CDS level resolution in a UK HCT donor-recipient population. A) Haplotypes most commonly contain one of five possible KIR2DL2/3 variants. B) The rare allele group, consisting of eight published and seven novel KIR2DL3 CDS alleles, comprises less than 4% of the total allele frequency. Novel CDS are described by combining their most homologous allele with the CDS mutation description.

5.04.03 KIR2DL2/3 non-coding polymorphism

In this cohort, genomic KIR2DL2/3 sequences that were observed only once make up 68% (730/1066) of the total sequences and 85% (730/855) of the total number of different allele sequences (Figure 5.6). Approximately 15% of genomic sequences were observed twice, and a further 15% were observed between three and seven times. The most frequent genomic sequence (a full-length extension of KIR2DL3*00501) was observed only 14 times. Unlike the breakdown of KIR2DL1 genomic alleles, where only a small number of frequently observed genomic allele sequences comprised the majority of their individual CDS allele group, most KIR2DL2/3 CDS allele groups are formed almost entirely from rarely observed variants (Figure 5.7).

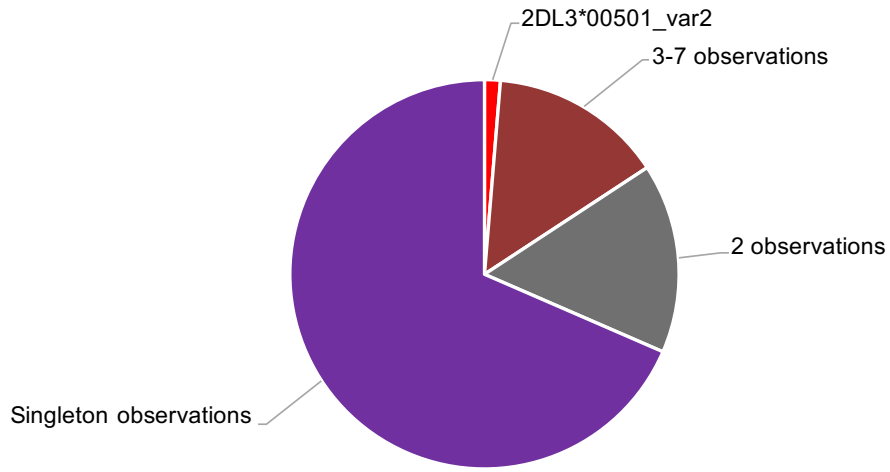


Figure 5.6 The majority of KIR2DL2/3 genomic alleles were observed only once

This pie chart depicts the frequency of observations of unique genomic allele sequences over the KIR2DL2/3 locus. Only one allele, a full-length extension of the existing published KIR2DL3*00501 sequence, is observed more than seven times. The remaining unique genomic alleles were grouped according to the frequency of observations.

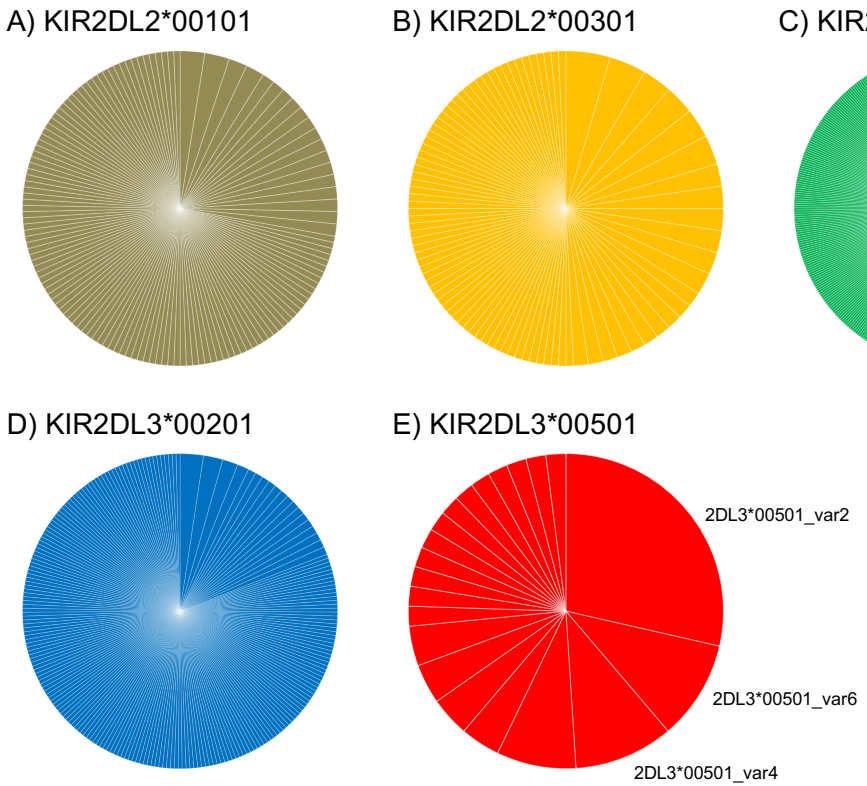


Figure 5.7 Non-coding polymorphism observations within the common KIR2DL2/3 CDS allele groups

Pie charts for each common KIR2DL2/3 CDS allele group demonstrate the extent of non-coding polymorphism. Pie charts represent A) 158 KIR2DL2*00101 (133 unique), B) 132 KIR2DL2*00201 (89 unique), C) 441 KIR2DL3*00101 (359 unique), D) 251 KIR2DL3*00201 (222 unique) and E) 49 KIR2DL3*00501 (21 unique) sequences, respectively. Alleles contributing more than 5% of the CDS subgroup are labelled. For allele sequences that did not match any existing published sequences, a description (most homologous published allele and discrepancy description) is available in Supplementary Table I.

The non-coding regions of KIR2DL2 and KIR2DL3 alleles underwent analysis to assess the location of polymorphism across the full length of the published KIR gene sequence (-300 to 14490). To achieve this, alleles from the different genes were separated and the sequences in each group aligned against each other (Figure 5.8). For KIR2DL2, a total of 203 polymorphic bases were observed (1.4% of the total nucleotide positions within the alignment), with 17 of these being characterised by trimorphism. Analysis of the type of polymorphism revealed 98 positions (48.3%) subject to only substitution polymorphism, 88 positions (43.3%) subject to only indel polymorphism and a further 17 positions (8.4%) where both indels and substitutions were observed. Although every effort was made to reduce errors of alignment by minimising the gap penalty, it is possible that some allele sequences were imperfectly aligned, potentially subtly affecting these polymorphism counts.

More polymorphism was observed within the KIR2DL3 alignment, revealing 605 polymorphic positions (4.1% of the total nucleotide positions within the alignment) including 11 trimorphic positions. Of these 605 positions, 340 (56.2%) were characterised by substitution variation only, 256 (42.3%) by indel variation only and nine (1.5%) by both indel and substitution variation.

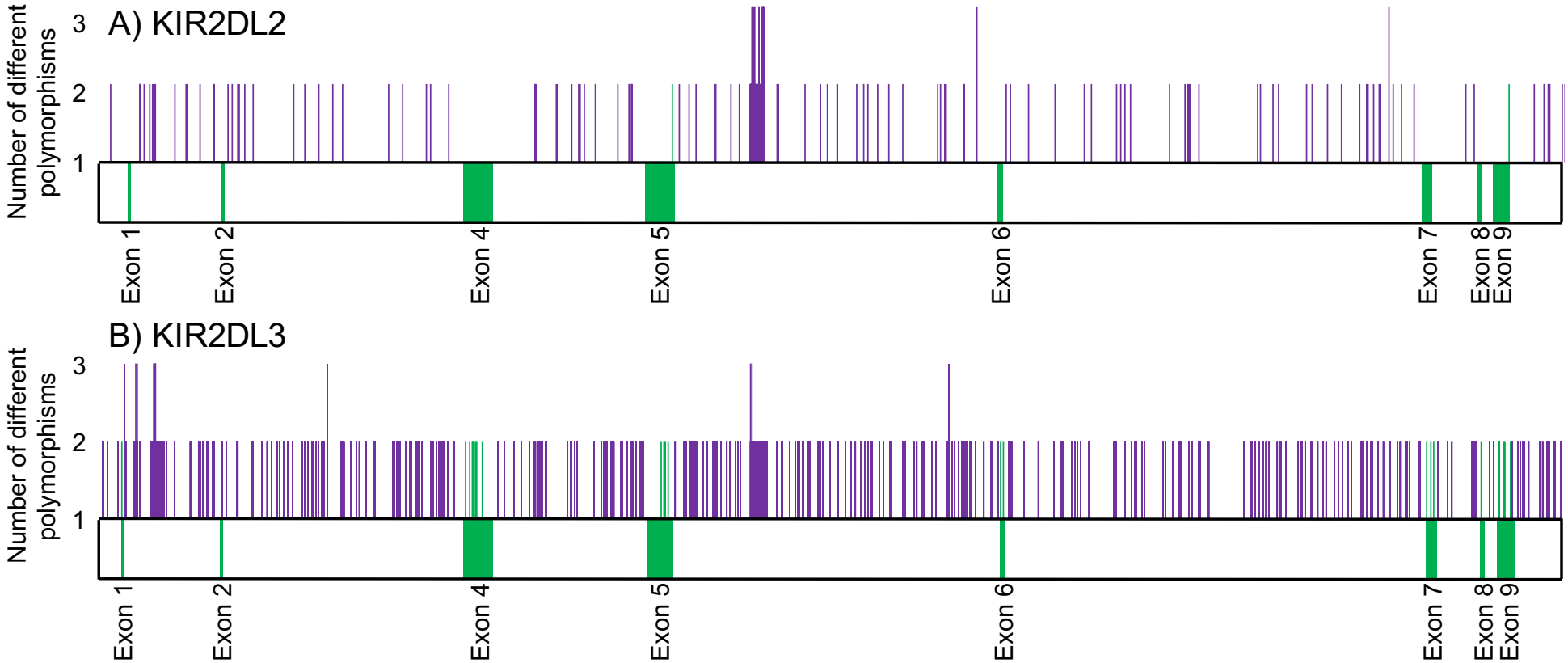


Figure 5.8 Location of polymorphism within the KIR2DL2 and KIR2DL3 genes

A schematic representation of the location of polymorphisms within 290 KIR2DL2 (A) and 776 KIR2DL3 (B) sequences generated for a UK HCT cohort. For KIR2DL2, several regions of conserved sequence are apparent across both the introns (purple) and exons (green), whilst only two polymorphic locations are found within KIR2DL2 exons. A greater frequency and relatively even distribution of polymorphism is observed for KIR2DL3. Double height peaks represent trimorphic nucleotide positions. Beneath, in green, the locations of the exons are indicated.

Both KIR2DL2 and KIR2DL3 contain a region of apparent high density polymorphism within the microsatellite-homopolymer region in intron 5. In depth analysis of the sequences revealed that the length of microsatellite region varied from 30 to 60 repeats, and the homopolymer length varied between 9 and 32 nucleotides. It was predicted that a high proportion of these polymorphisms may be unreliable indel mutations and, as such, a further sub-analysis was performed to exclude this region. In the residual KIR2DL2 sequence, only 125 polymorphic nucleotide positions were observed (0.9% of total sequence), of which 85 (68.0%) were characterised by substitution variation only, 38 (30.4%) by indel variation only and only two positions (1.6%) were characterised by both indels and substitutions. In a similar analysis for the KIR2DL3 alignment, 495 positions remained polymorphic, of which 338 (68.3%) were characterised by substitution variation only, 150 (30.3%) were characterised by indel variation only and seven positions (1.4%) by both indels and substitutions.

5.05 Results: KIR3DL1/S1 analysis in a UK HCT population

5.05.01 KIR3DL1/S1 amplification and sequencing success rate

As with the KIR2DL1 and KIR2DL2/3 loci, PCR amplification of the KIR3DL1/S1 target region was attempted in 771 samples. Similarly to KIR2DL2/3, as each KIR haplotype is predicted to encode at least one KIR3DL1/S1 allele, amplification of fragments corresponding to KIR3DL1/S1 alleles was expected for every sample. However, successful sequencing was only obtained in 711 samples, representing an amplification success rate of 92.2%. Of these 711 successful samples, 381 (53.6%) were from donors and 330 (46.4%) were from recipients.

Each successful amplicon was pooled within a library and sequenced. However, following analysis of the sequencing results, only 626 complete samples were obtained. This represents a sample sequencing success rate of 81.2% and amplicon sequencing success rate of 88.0%. The main cause of sequencing failure was suspected KIR3DS1 allele dropout (based on discrepancy to previous KIR3DS1 presence/absence genotyping, n=49). To confirm that errors were not made during the initial KIR3DS1 presence/absence assessment, each sample underwent repeat presence/absence genotyping, confirming the original findings. In addition, 17 samples generated sequences with high homology to existing KIR3DL1/S1 alleles but featuring homopolymer discrepancies within the CDS. Most commonly, these CDS were homologous to KIR3DS1*01301, but featured a single deletion within the homopolymer at the end of exon 7. As this polymorphism has been observed in an alternative KIR3DS1 allele (KIR3DS1*010), the deletion mutation generated may represent genuine polymorphism. However, despite being sequenced multiple times and from multiple samples, this discrepancy within a homopolymer may also represent a simple sequencing error. As such, the sequencing status of these samples could not be marked as complete. Further causes of sequencing failure included unconfirmed CNV (n=9) and suspected dropout (n=8) of KIR3DL1 alleles. Amplicons from three further samples failed to generate any identifiable KIR3DL1/S1 sequences. Finally, two novel CDS alleles (featuring only substitutions) were unable to be confirmed as a result of diminished DNA stock, each representing another sample without complete KIR3DL1/S1 typing.

From the 626 samples with complete KIR3DL1/S1 allele typing, 1,200 allele sequences were obtained. The majority of these were KIR3DL1 sequences (982/1200, 81.8%),

whilst the remainder were derived from KIR3DS1 amplicons. The mean and median average read depth for KIR3DL1 alleles was 99 and 89, respectively whilst, for KIR3DS1 alleles, equivalent averages were 98 and 82, respectively.

5.05.02 KIR3DL1/S1 CDS polymorphism

Despite only 626 samples having complete KIR3DL1/S1 typing, a high level of diversity was observed at this locus. In total, 26 unique CDS, including six novel CDS alleles were observed (Table 5.3). Nine of these were observed more than 25 times in this small population. Of the novel CDS, two particularly interesting variants were observed. The first, in a CDS otherwise homologous to KIR3DL1*00401, featured a three bp deletion (CDS position 605-607). Although not corresponding to an in-frame codon, the resulting deletion coded for only a single amino acid deletion. The remainder of the protein sequence remained unchanged.

In the second interesting novel allele, the termination codon for an allele with homology to KIR3DS1*01301 was shifted by an insertion of two nucleotides, AC, at CDS position 1144 in codon 361. This, in combination with a substitution (G>C) at CDS position 1155, resulted in the extension of the predicted protein sequence by six amino acid residues. This forms a novel method to extend the cytoplasmic tail of KIR3DS1 but utilises a similar codon framework to reach the same stop codon as KIR3DS1*010 (although KIR3DS1*010 reaches the same stop codon via a one bp deletion at the end of exon 7 rather than a two bp insertion in exon 8).

Table 5.3 Confirmed KIR3DL1/S1 CDS polymorphism

Most homologous allele	CDS polymorphism	Codon	Amino acid polymorphism	Observation frequency
KIR3DL1*00101	-	-	-	191
KIR3DL1*002	-	-	-	147
KIR3DL1*00401	-	-	-	179
KIR3DL1*00402	-	-	-	20
KIR3DL1*00501	-	-	-	189
KIR3DL1*00701	-	-	-	35
KIR3DL1*008	-	-	-	50
KIR3DL1*009	-	-	-	20
KIR3DL1*01501	-	-	-	1
KIR3DL1*01502	-	-	-	95
KIR3DL1*01701	-	-	-	2
KIR3DL1*019	-	-	-	11
KIR3DL1*02001	-	-	-	26
KIR3DL1*03101	-	-	-	1
KIR3DL1*033	-	-	-	1
KIR3DL1*053	-	-	-	4
KIR3DL1*099	-	-	-	3
KIR3DL1*109	-	-	-	1
KIR3DS1*01301	-	-	-	210
KIR3DS1*049N	-	-	-	7
KIR3DL1*00101	1149C>G;1152C>T	362, 363	-	1
KIR3DL1*00101	1303A>G	414	Lys>Glu	1
KIR3DL1*002	1195G>C	378	Ala>Pro	2
KIR3DL1*00401	605delCCT	181, 182	Single amino acid deletion[§]	1
KIR3DL1*008	1316A>T	418	Lys>Ile	1
KIR3DS1*01301	1144insAC;1155G>C	361-368	Cytoplasmic tail extension[†]	1

[§] Thr-Ser-Tyr>Thr-Tyr

[†] Gln-Lys-STOP>Gln-Asn-Ser-Glu-Gln-Pro-Gly-Phe-STOP

In addition to novel alleles, five instances of less common CNV at the KIR3DL1/S1 locus were observed. In each case, one allele of KIR3DS1 was found in combination with two different KIR3DL1 alleles.

As the KIR2DL1/2/3 and KIR3DL1/S1 loci are located on opposing sides of the recombination hotspot on the KIR haplotype, the haplotype estimation procedure (Section 5.07) did not include KIR3DL1/S1 allele typing. However, by assuming that samples where only one allele sequence was generated were homozygous for this locus (and found only once on each haplotype), it was possible to estimate the KIR3DL1/S1 CDS allele frequencies from samples with complete allele typing (Figure 5.9).

Unlike KIR2DL1 and KIR2DL2/3 CDS allele frequencies, where one allele is observed much more frequently than any other (KIR2DL1*00302 and KIR2DL3*00101), several KIR3DL1/S1 CDS alleles each share approximately 15% of the total allele frequency. Despite the KIR3DS1 gene being present in less than one third of the samples with complete KIR3DL1/S1 allele typing (206/626, 32.9%), KIR3DS1*01301 is the most common CDS allele with a frequency of 18.1%. The KIR3DL1*00501, 00101, 00401 and 002 CDS allele groups share approximately equal allele frequencies (16.2%, 15.9%, 14.7% and 12.1%, respectively). In addition, the allele frequencies of KIR3DL1*01502, 008, 00701 and 02001 alleles (7.7%, 4.1%, 2.8% and 2.1%, respectively) also suggest these are common alleles (>2% of total frequency).

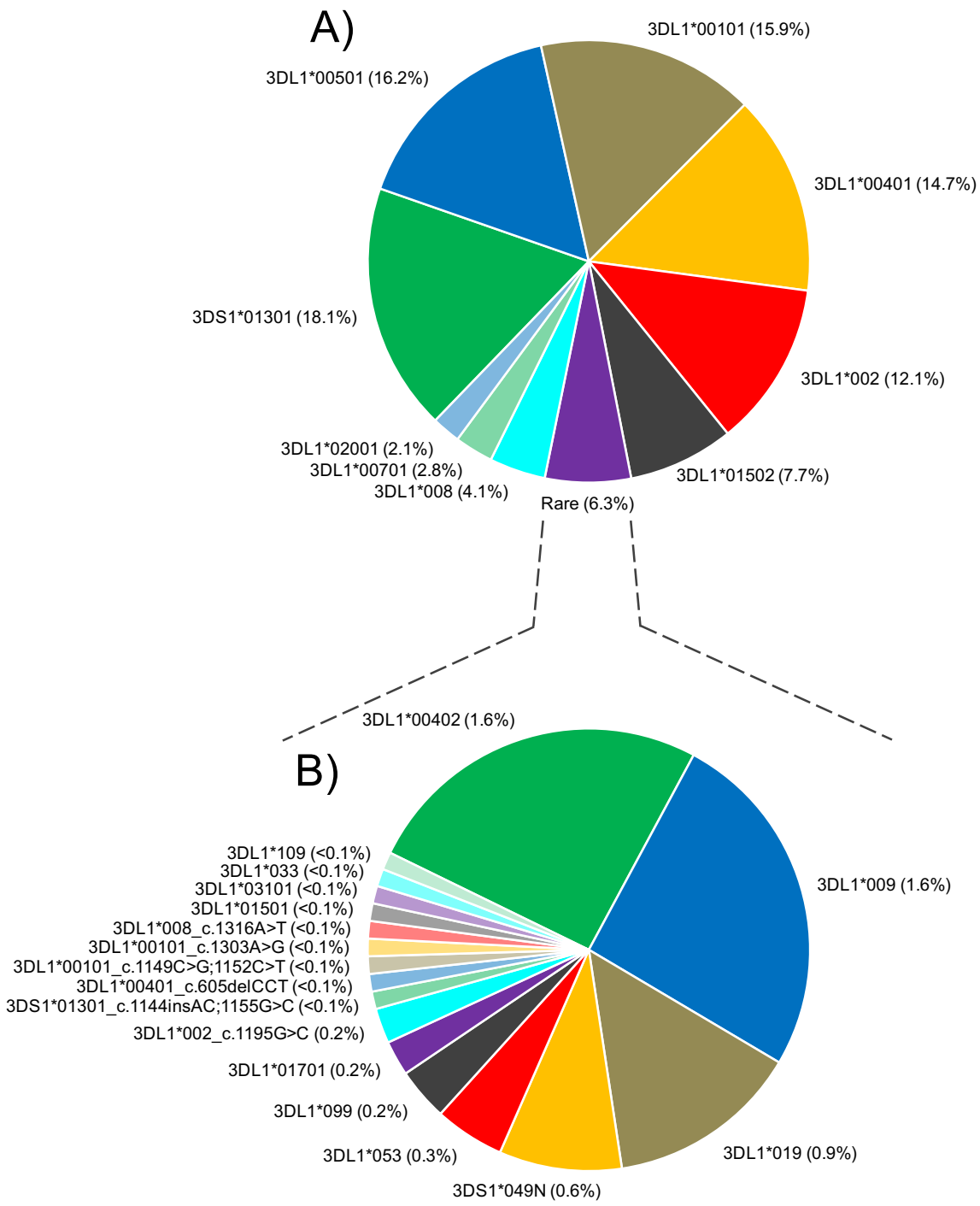


Figure 5.9 KIR3DL1/S1 CDS allele frequency

Pie charts depict the frequency of KIR3DL1/S1 alleles at CDS level resolution in a UK HCT donor-recipient population. A) Haplotypes most commonly contain one of nine common KIR3DL1/S1 variants. B) The rare allele group, consisting of 11 published and six novel KIR3DL1/S1 CDS alleles, comprises over 6% of the total allele frequency. Novel CDS are described by combining their most homologous allele with the CDS mutation description.

5.05.03 KIR3DL1/S1 non-coding polymorphism

The subgroup of each CDS allele group with more than 5% allele frequency, was further divided to investigate the degree of non-coding polymorphism within common KIR3DL1/S1 CDS alleles (Figure 5.10). This demonstrates that, for KIR3DL1*00101, many different non-coding polymorphisms are present at reasonably high frequency (>5% of subgroup). However, for each of the other common KIR3DL1/S1 CDS allele subgroups, a small number of different variants comprise a fairly large proportion of the total observations of that CDS. Indeed, for KIR3DL1*00501, one full length sequence, representing the KIR3DL1*0050101 variant, was generated in 95 different samples and makes up 50.3% of the total KIR3DL1*00501 observations. A summary of the 12 most frequently observed KIR3DL1/S1 genomic allele sequences, and their share of the overall observation frequency, is provided in Figure 5.11. Approximately a quarter of KIR3DL1/S1 genomic allele sequences were observed only once (285/1200=24%), although they account for over 75% of the different unique allele sequences (285/376).

As with the other KIR genes, both the KIR3DL1 and KIR3DS1 full length sequences underwent alignment to map the location of polymorphisms across the gene (Figure 5.12). This demonstrated that, for KIR3DL1, polymorphism is frequent throughout the gene: 810 polymorphic sites (5.6% of the total alignment), including 60 trimorphic and four tetramorphic (four or more variations at a single position) sites, were observed. The types of variation included 339 (41.9%) nucleotide sites that differed by substitutions only, 413 (51.0%) nucleotide sites that differed by indels only and 58 (7.2%) nucleotide sites where both indel and substitution variation was observed.

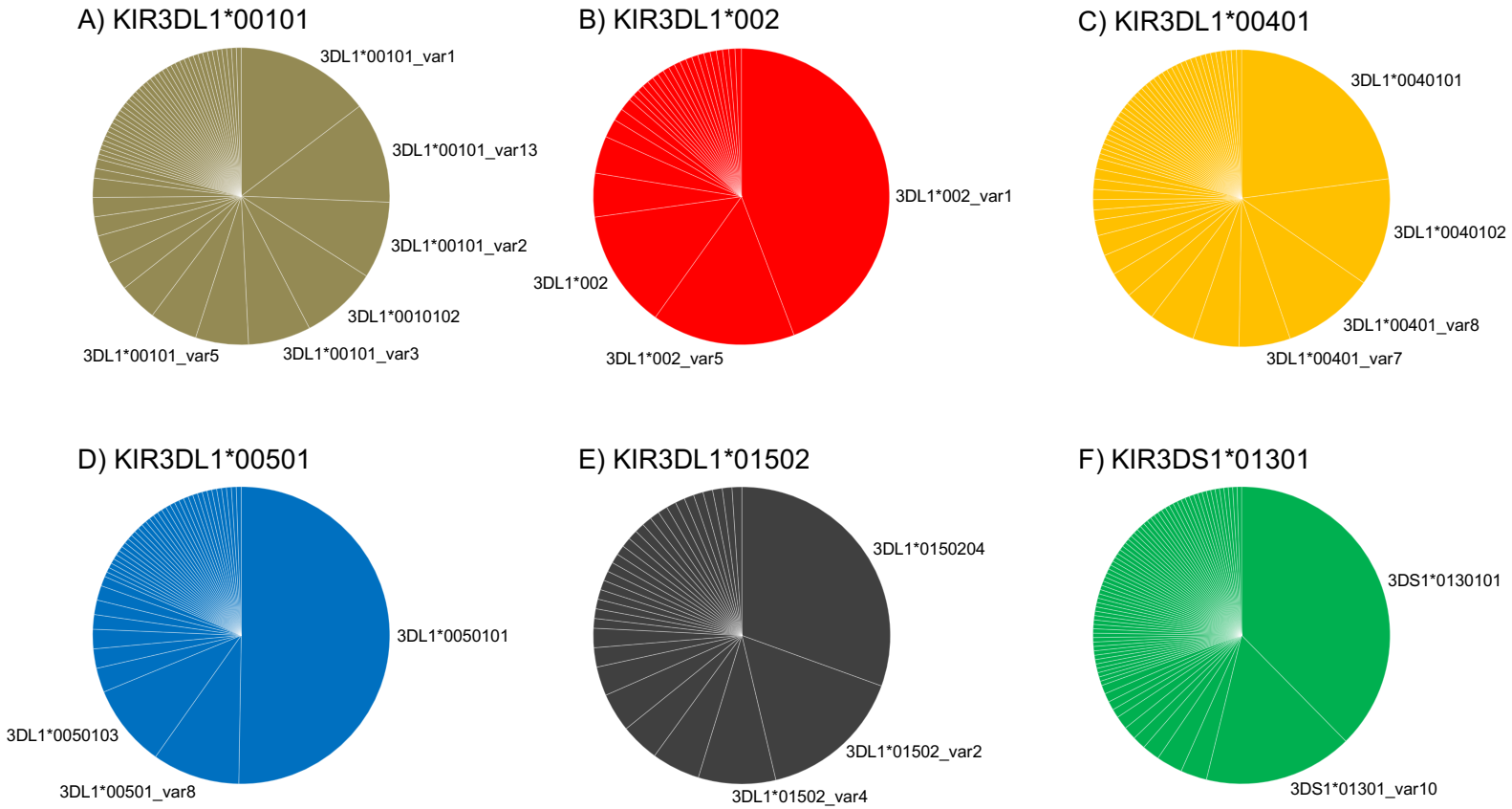


Figure 5.10 Non-coding polymorphism observations within six common KIR3DL1/S1 CDS allele groups

Pie charts for the six most common KIR3DL1/S1 CDS allele group demonstrate the extent of non-coding polymorphism. Pie charts represent A) 191 KIR3DL1*00101 (55 unique), B) 147 KIR3DL1*002 (28 unique), C) 179 KIR3DL1*00401 (54 unique), D) 189 KIR3DL1*00501 (45 unique), E) 95 KIR3DL1*01502 (32 unique) F) and 210 KIR3DS1*01301 (76 unique) sequences, respectively. Alleles contributing more than 5% of the CDS subgroup are labelled. For allele sequences that did not match any existing published sequences, a description (most homologous published allele and discrepancy description) is available in Supplementary Table I.

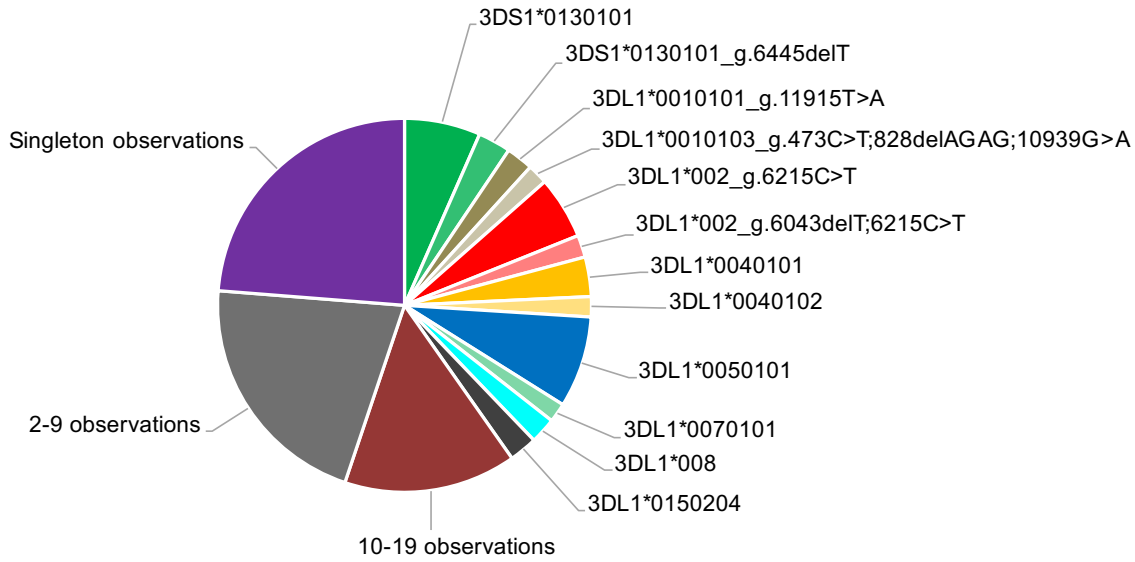


Figure 5.11 Common KIR3DL1/S1 genomic allele observation frequency chart

This pie chart depicts the frequency of observations of unique genomic allele sequences over the KIR3DL1/S1 locus. The details of the 10 most commonly observed KIR3DL1 genomic alleles (the most homologous published allele and, if present, discrepancies to it) are given. For KIR3DS1, the two most commonly observed variants are shown. Members of the same CDS allele group are shaded accordingly. The remaining unique genomic alleles were grouped according to the frequency of observations.

A similar alignment of KIR3DS1 sequences implied a much less even distribution of polymorphism across the gene, with most polymorphism focussed within one region in intron 3 (discussed below). Despite this, 723 polymorphic sites were observed, including 93 trimorphic and 12 tetramorphic positions. Polymorphism type was dominated by indel variation: 577 (79.8%) nucleotide sites were indel variants only, whilst a further 105 (14.5%) featured indels or substitutions. Only 41 (5.7%) nucleotide sites were characterised by substitution only.

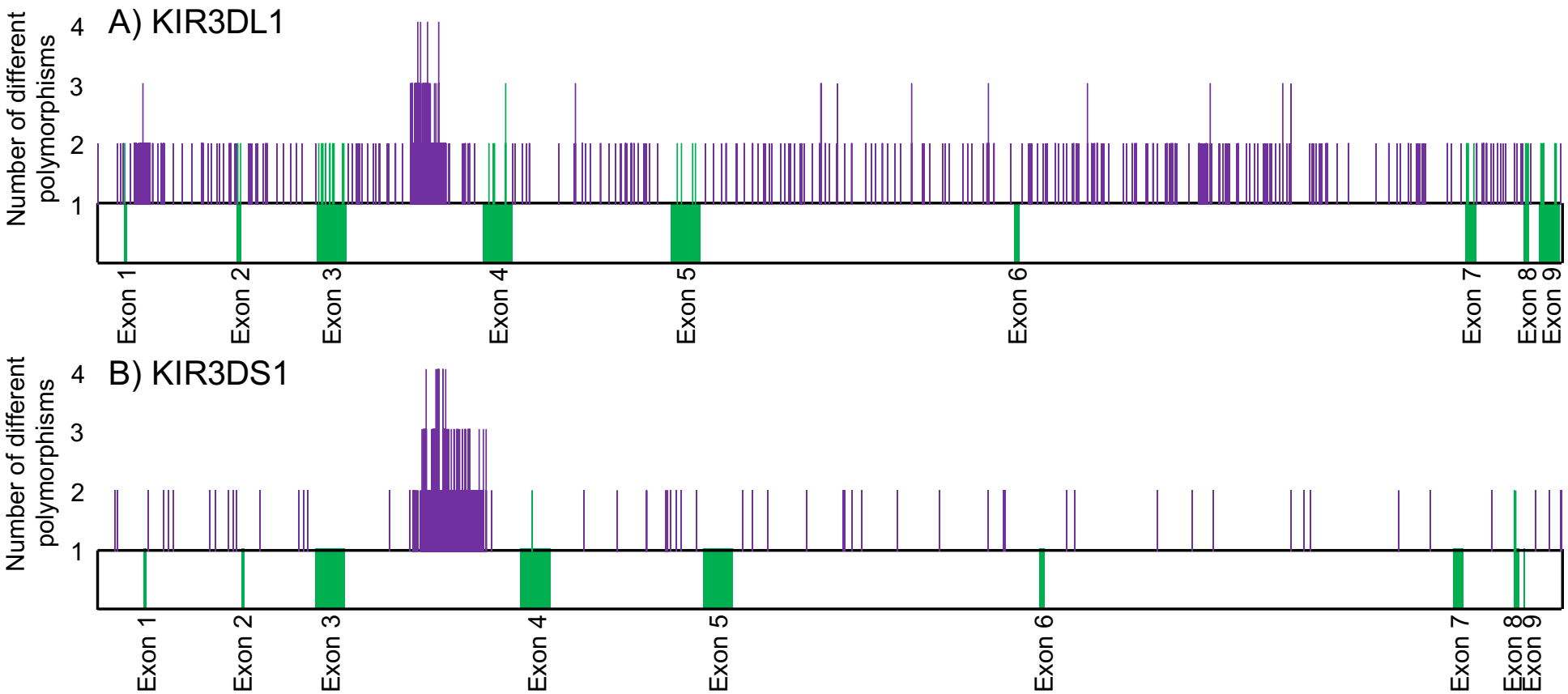


Figure 5.12 Location of polymorphism within the KIR3DL1 and KIR3DS1 genes

A schematic representation of the location of polymorphisms within 982 KIR3DL1 (A) and 218 KIR3DS1 (B) sequences generated for a UK HCT cohort. For KIR3DL1, polymorphism is distributed across the entire gene in both the introns (purple) and exons (green). Less polymorphism is observed for KIR3DS1 and appears highly concentrated in intron 3. Double and triple height peaks represent trimorphic and tetramorphic nucleotide positions, respectively. Beneath, in green, the locations of the exons are indicated.

In both KIR3DL1 and KIR3DS1 genes, large regions of polymorphism were observed in intron 3. In addition, a polymorphism-dense region is observed in intron 1 of KIR3DL1. When investigated in detail, these regions of sequence are characterised by many indel mutations. Interestingly, in all the previously published sequences, very little polymorphism is located in intron 3 of either KIR3DL1 or KIR3DS1 and, although two short indels are observed in the intron 1 region of currently published KIR3DL1 sequences, a great number of additional novel indels were observed in this UK HCT cohort. As such extensive polymorphism was not predicted in this region and, as we have demonstrated unreliability of SMRT DNA sequencing to correctly resolve indel mutations, a separate sub-analysis was performed to exclude these regions.

In this sub-analysis, the number of polymorphic sites in KIR3DL1 and KIR3DS1 were reduced to 412 and 69, respectively. In the KIR3DL1 alignment, 305 ‘substitution only’ sites remained, as well as 99 ‘indel only’ sites and 8 sites where both substitution and indels were observed. This represented a decrease of 10.0%, 76.0% and 86.2% in each group, respectively, and a decrease of 49.1% of the total polymorphism. For KIR3DS1, an even higher proportion of the observed polymorphism was lost upon performing sub-analysis excluding the indel-rich region of intron 3. Only 31 ‘substitution only’ and 38 ‘indel only’ polymorphic sites remained, and no nucleotide sites contributed both substitution and indel variation. This represents a decrease of 24.4%, 93.4% and 100% of each type of polymorphism, respectively, and a collective reduction in polymorphism of 90.5%. It seems unlikely that a single, small, non-coding region of any gene should contribute over 90% of the polymorphism observed, supporting the performance of a sub-analysis to remove this potentially unreliable sequence region.

5.06 Results: Polymorphism within KIR promoter regions

Several of the promoter regions are within the sequenced regions of the KIR genes studied in this clinical cohort. As such, polymorphism across the 5′ UTR and intron 2 antisense promoter binding site were investigated in more detail, and differences within putative TF binding sites highlighted (Figure 5.13). A phylogenetic comparison between promoter sequences from different alleles is displayed in Figure 5.14.

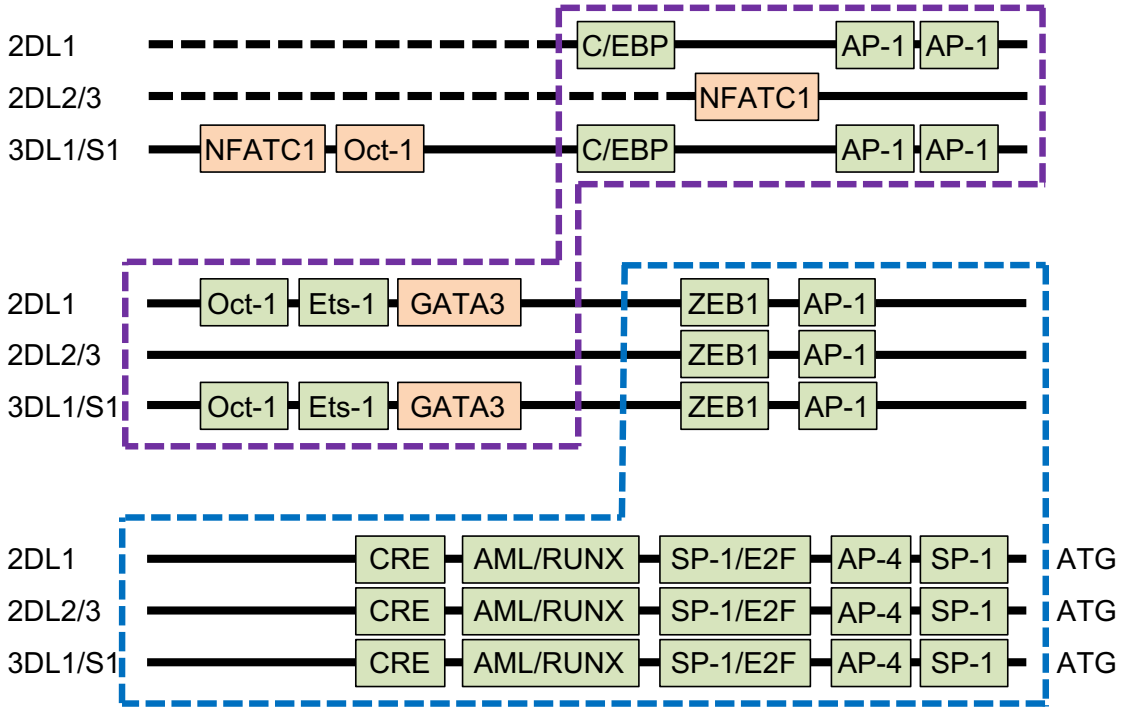


Figure 5.13 Putative TF binding site locations within the sequenced regions of KIR2DL1, KIR2DL2/3 and KIR3DL1/S1

A schematic representation of the 5′ UTR of KIR genes studied in the UK HCT cohort. TF binding sites within the KIR2DL1 and KIR3DL1/S1 intermediate promoter region (purple dashed area) and KIR2DL1, KIR2DL2/3, KIR3DL1/S1 proximal promoter region (blue dashed area) are shown. Putative TF binding sites are depicted by green boxes, whilst previously undetermined TF binding sites are demonstrated in peach-coloured boxes. Dashed black lines represent unsequenced regions of the 5′ UTR.

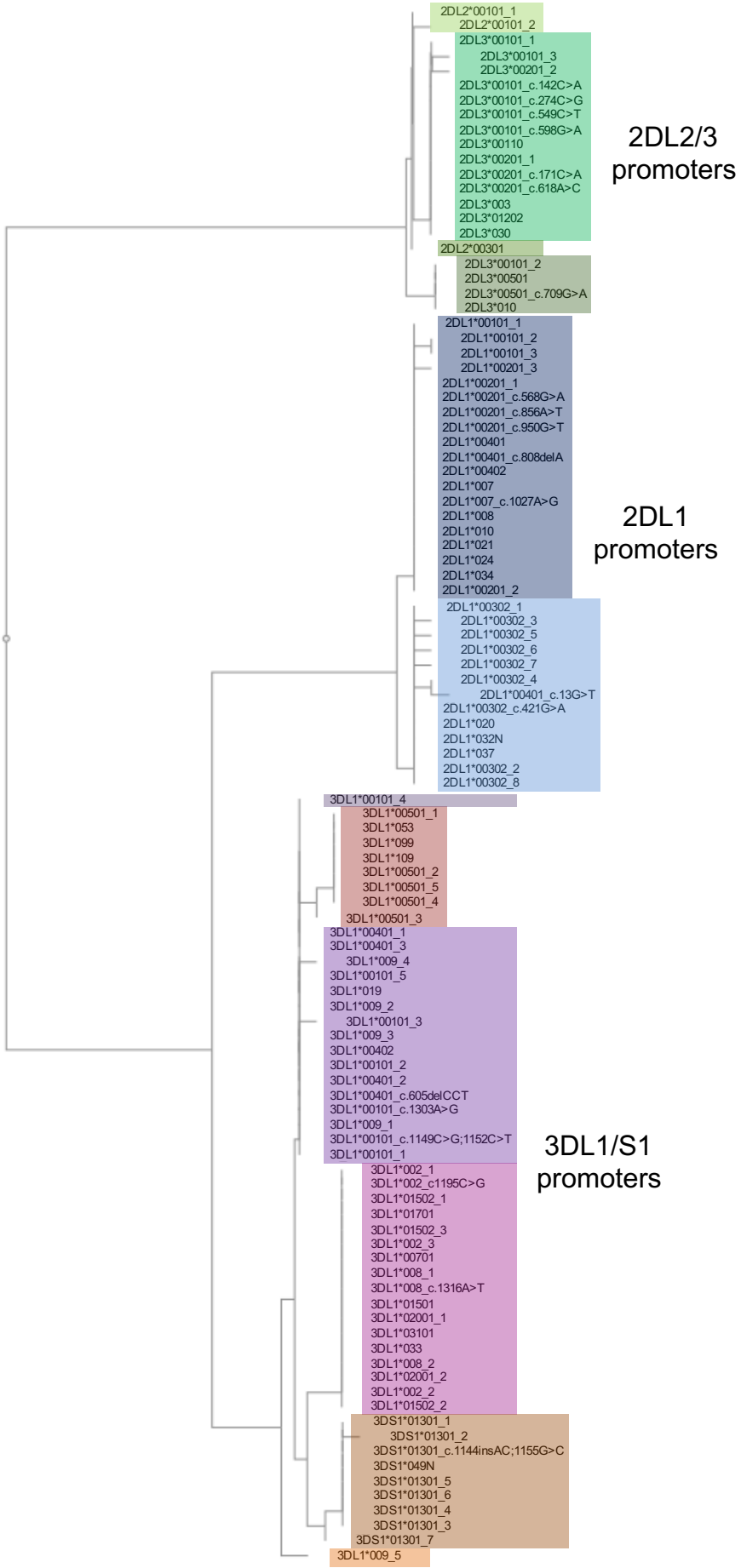


Figure 5.14 Phylogenetic tree analysis of distinct promoter regions from the each unique KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 CDS allele group

This analysis, assessing sequence homology across the sequenced regions of KIR 5′ UTR, clearly demonstrates that individual loci have specific promoter regions. Within each locus, further subgrouping is also apparent, indicated by different coloured boxes. This phylogenetic tree was generated using Archaeopteryx.js, version 1.8.1 [382].

5.06.01 KIR2DL1 5′ untranslated region (UTR) polymorphism

The KIR2DL1 5′ primer designed for this study (see Section 4.02) binds approximately 540 bp upstream of the start codon. Within the interspersing region, 18 different KIR2DL1 polymorphisms were observed. The most frequently observed 5′ UTR sequence was located upstream of most KIR2DL1*00201, 00401, 00402, 007, 008, 010, 021 and 034 alleles and, as such, appears to be widespread. KIR2DL1*00302 alleles appear to have a broader spectrum of different 5′ UTR sequences but exists predominantly with a 5′ UTR sequence shared with the KIR2DL1*020, 032N and 037 alleles. KIR2DL1*00101 alleles utilise a distinct promoter sequence that is not found upstream of any other KIR2DL1 allele, differing to other alleles by a G>A substitution 127 bp upstream of the start codon.

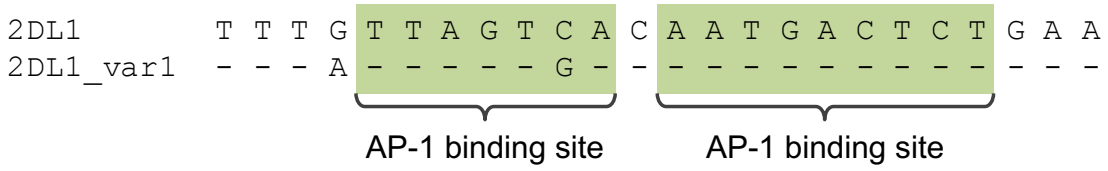


Figure 5.15 An AP-1 binding site is disrupted in most alleles of the KIR2DL1*00302 CDS subgroup

A sequence alignment of the region 415-392 bp upstream of the KIR2DL1 start codon reveals that the first of the dual AP-1 binding sites (green boxes) within the intermediate promoter exhibits polymorphism. This g.-406C>G substitution is observed upstream of most, but not all, KIR2DL1*00302 alleles and a single novel CDS variant with homology to KIR2DL1*00401. These alleles have been grouped here as "2DL1_var1". All other KIR2DL1 alleles encode the 'C' variant at this position. A second substitution, g.-412G>A, adjacent to the AP-1 binding site, is also associated with the 2DL1_var1 allele group. Dashes (-) represent homologous sequence.

Although much variation exists, many of the putative TF-binding sites remain conserved. However, within the KIR2DL1 intermediate promoter, a g.-406C>G substitution polymorphism within the AP-1 binding site is associated with most (but not all) KIR2DL1*00302 alleles (Figure 5.15) [166].

Interestingly, this polymorphism is also associated with the novel CDS allele, KIR2DL1*00401_c.13G>T. The coding variation, c.13G>T, is only observed in KIR2DL1*00201 alleles and, as such, this allele forms an interesting combination of KIR2DL1*00302 intermediate promoter and KIR2DL1*00201 exon 1, whilst the remainder of the CDS maintains strong homology to KIR2DL1*00401. This novel allele also contains a G>T substitution 351 bp upstream of the start codon, within Ets-1 binding site (g.-354 to -346 bp, Figure 5.16) [166]. As well as being shared with a small minority of KIR2DL1*00302 alleles, presence of thymidine at this position is conserved across each of the KIR2DL2/3 alleles sequenced. Furthermore, within the cyclic adenosine monophosphate (cAMP) response element (CRE) binding motif of the proximal promoter (g.-119 to -112), another mutation is observed in this interesting novel allele (Figure 5.17). Here, the g.-119G>A substitution represents another potentially functional variation within a promoter region. This polymorphism is also shared by a small minority of the KIR2DL3*00101 alleles sequenced.

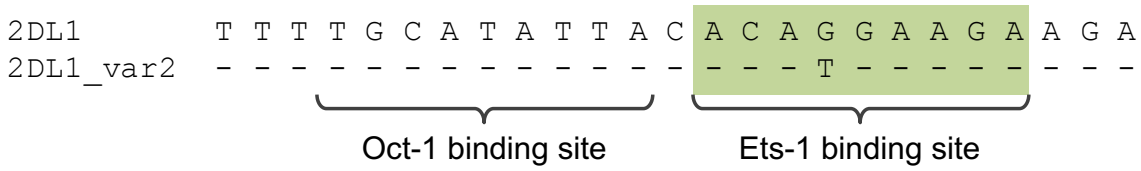


Figure 5.16 An Ets-1 binding site is disrupted in some alleles of the KIR2DL1*00302 CDS subgroup and a novel KIR2DL1 CDS

A sequence alignment of the region 367-343 bp upstream of the KIR2DL1 start codon reveals that the Ets-1 binding site (green box) of the intermediate promoter exhibits polymorphism. This g.-351G>T substitution is observed upstream of a minority of KIR2DL1*00302 alleles and a single novel CDS variant with homology to KIR2DL1*00401. These alleles have been grouped here as "2DL1_var2". All other KIR2DL1 alleles encode the 'G' variant at this position. The adjacent Oct-1 TF binding site is unaffected. Dashes (-) represent homologous sequence.

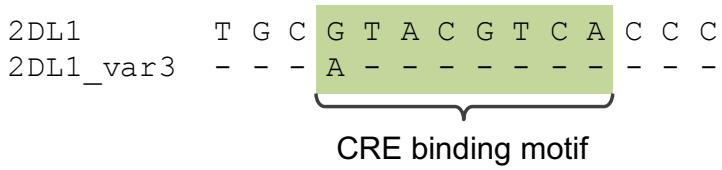


Figure 5.17 A CRE binding motif is disrupted upstream of a novel KIR2DL1 CDS

A sequence alignment of the region 122-109 bp upstream of the KIR2DL1 start codon reveals that the CRE binding motif (green box) of the proximal promoter exhibits polymorphism. Within the KIR2DL1 analysis, this g.-119G>A substitution was only observed upstream of a novel CDS variant with homology to KIR2DL1*00401. This allele has been labelled here as "2DL1_var3". All other KIR2DL1 alleles encode the 'G' variant at this position. Dashes (-) represent homologous sequence.

The final polymorphism within speculated promoter elements of KIR2DL1 alleles was observed in the 5′ UTR sequences of KIR2DL1*00302 alleles from 11 different samples. This substitution, g.-28G>A, within the SP-1 binding site, was not observed in any other sequence within this cohort but has previously been associated with KIR2DL3*0010105 and some KIR2DP1 allele sequences (Figure 5.18).

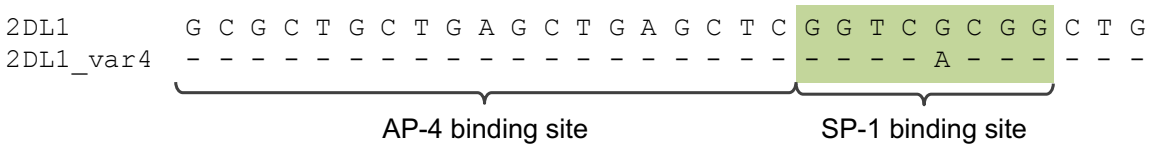


Figure 5.18 An SP-1 binding motif is disrupted upstream of some KIR2DL1*00302 alleles

A sequence alignment of the region 51-22 bp upstream of the KIR2DL1 start codon reveals that the SP-1 binding motif (green box) of the proximal promoter exhibits polymorphism. This g.-28G>A substitution was observed upstream of a small minority of alleles belonging to the KIR2DL1*00302 CDS subgroup. These alleles have been grouped here as "2DL1_var4". All other KIR2DL1 alleles encode the 'G' variant at this position. The adjacent AP-4 TF binding site is unaffected. Dashes (-) represent homologous sequence.

All other polymorphism within the KIR2DL1 5′ UTR sequences was located within regions not previously implicated as TF binding sites and, as such, is unlikely to influence KIR2DL1 expression.

5.06.02 KIR2DL2/3 5′ untranslated region (UTR) polymorphism

Despite exhibiting high levels of polymorphism in non-coding regions following the start codon, very little polymorphism was observed within the 5′ UTR of KIR2DL2 and KIR2DL3 allele sequences. In fact, the 453 bases preceding the start codon were characterised by only six different sequences. Two of these sequences associated only with KIR2DL2 alleles, with only the g.-137C>A substitution separating the common and rare sequence. The rarer sequence, observed only five times throughout the entire cohort, was only associated with a minority of the KIR2DL2*00101 CDS allele group. The remaining four 5′ UTR sequences each associated only with KIR2DL3 alleles. The majority of the alleles within the KIR2DL3*00101, 00110, 00201, 003, 01202, 030 and novel CDS groups shared the same sequence. However, alleles of the KIR2DL3*00501 and 010 CDS subgroups utilised a separate 5′ UTR sequence. Three differences exist

between these two relatively common sequences: G>C, A>G and C>A substitutions 441, 176 and 21 bp upstream of the start codon, respectively. Although a 5′ UTR sequence for KIR2DL3*010 has been published previously [58], the published sequence of the more common KIR2DL3*00501 allele was, at the time of writing, unavailable for this section of non-coding sequence (full length sequences generated as part of this project have since been submitted, accession number: LR593920). However, one of the differences that define these 5′ UTR sequences may have a potentially functional impact on KIR2DL3 expression. Although the g.-21C>A and g.-441G>C substitutions are not located within known TF binding sites, the g.-176A>G substitution is located within the putative YY-1 TF binding site of the KIR2DL3 proximal promoter and converts the sequence to more closely resemble KIR2DL1 alleles at this position (Figure 5.19).

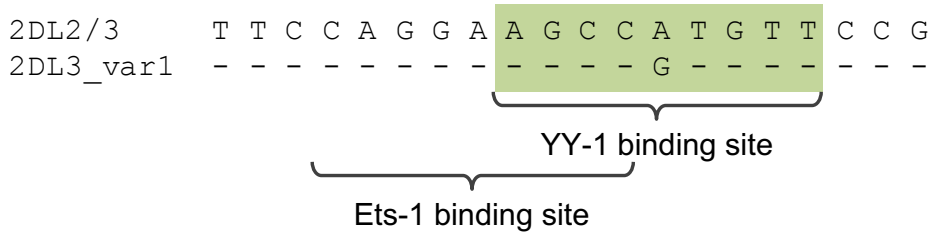


Figure 5.19 The YY-1 binding motif is disrupted upstream of some KIR2DL3 alleles

A sequence alignment of the region 188-169 bp upstream of the KIR2DL2/3 start codon reveals that the YY-1 binding motif (green box) of the proximal promoter exhibits polymorphism. This g.-176A>G substitution was observed upstream of each KIR2DL3*00501 and 010 allele, as well as a single allele of the KIR2DL3*00101 CDS subgroup and a novel CDS allele with homology to KIR2DL3*00501. These alleles have been grouped here as "2DL3_var1". All other KIR2DL2/3 alleles encode the 'A' variant at this position. The overlapping Ets-1 TF binding site is unaffected. Dashes (-) represent homologous sequence.

Of the two remaining KIR2DL3 5′ UTR sequences, associating with small minorities of KIR2DL3*00101 and 00201 alleles (n=3 and n=4, respectively) one differs from the common KIR2DL3*00201 5′ UTR sequence by a single substitution outside of putative TF binding sites (g.-154C>A). The other sequence, however, represents the published g.-119G>A substitution present in the KIR2DL3*0010106 allele. This substitution, located with the CRE binding element, was also discovered in a novel KIR2DL1 allele sequence (discussed above in Section 5.06.01, Figure 5.17).

5.06.03 KIR3DL1/S1 5′ untranslated region (UTR) polymorphism

The highest degree of 5′ UTR variation existed in the 749 bp upstream of the KIR3DL1/S1 start codon. Thirty-two different variations were observed, and seven of these were observed more than 25 times, suggestive of common frequency. The 5′ UTR sequence for most alleles of the KIR3DL1*00101 CDS subgroup were shared with KIR3DL1*009 alleles, whilst KIR3DL1*002 subgroup alleles shared a common 5′ UTR sequence with KIR3DL1*01502 and 01701 alleles. KIR3DL1*00401 and 00402 alleles had a 5′ UTR sequence common to KIR3DL1*019 alleles, and the KIR3DL1*00501 5′ UTR sequence was also observed upstream of KIR3DL1*053, 099 and 109 alleles. The KIR3DL1*00701, 008, 01501, 03101 and 033 allele subgroups also shared a 5′ UTR sequence. With only one exception, the 5′ UTR sequence upstream of KIR3DL1*02001 alleles was unique (in the one exception, it was homologous to the 5′ UTR sequence of the KIR3DL1*00701/008/01501/03101/033 group). In addition, the majority of different KIR3DS1 alleles shared the same 5′ UTR sequence, although several distinct, less frequent 5′ UTR sequences were observed in some KIR3DS1*01301 sequences. Each of these KIR3DS1 variants differed outside of predicted TF binding sites and, as such, are unlikely to affect KIR3DS1 expression.

As with the other KIR genes assessed in this HCT cohort, the unique sequences were aligned against each other to investigate whether differences between these sequences had a potential functional impact on expression. This revealed a total of 34 polymorphic positions, 22 of which were characterised by substitution only, 10 of which featured only indel mutations, whilst the remaining two could be characterised by either substitution or indel polymorphism.

Despite the KIR3DL1/S1 sequencing generating the longest 5′ UTR sequences, coverage of the distal promoter was not achieved. However, within the intermediate promoter region, one potentially important polymorphism was observed in a putative AP-1 binding site. In 15 of the 32 different KIR3DL1/S1 5′ UTR sequences, an insertion of an 'A' 409 bp upstream of the start codon (g.-409insA) was observed (Figure 5.20). This polymorphism is not located within a repetitive DNA sequence and has been characterised previously [166], reducing the probability that this resulted from a sequencing error. This insertion was observed upstream of the following KIR allele CDS groups: KIR3DL1*002, 00701, 008, 01501, 01502, 01701, 020, 03101, 033 and KIR3DS1*01301 and 049N. In addition, despite every other example of KIR3DL1*00501 alleles featuring the deleted variation of the AP-1 site, one KIR3DL1*00501 sequence is characterised by an insertion at this position. The remainder of the intermediate promoter TF binding sites appear conserved in this cohort.

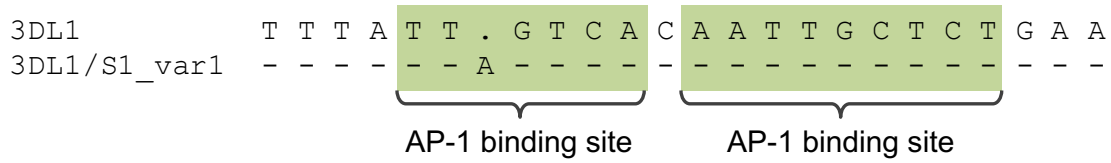


Figure 5.20 An AP-1 binding motif upstream of KIR3DL1/S1 alleles features indel polymorphism

A sequence alignment of the region 415-392 bp upstream of the KIR3DL1/S1 start codon reveals that the first of the AP-1 binding motifs (green boxes) of the intermediate promoter exhibits indel polymorphism. This g.-409insA polymorphism was observed upstream of every KIR3DS1 allele sequence and several different KIR3DL1 CDS allele subgroups. These alleles have been grouped here as "3DL1/S1_var1". All other KIR3DL1 alleles encode the deletion variant at this position. The adjacent AP-1 TF binding site is unaffected. Dots (.) represent deleted nucleotides. Dashes (-) represent homologous sequence.

Within the proximal promoter, the putative ZEB-1 TF binding site (g.-255 to g.-240) also features a previously published polymorphism: a G>A substitution 249 bp upstream of the start codon (Figure 5.21) [58]. This polymorphism is found upstream of KIR3DL1*002, 00701, 008, 01501, 01502, 01701, 02001, 03101 and 033 alleles.

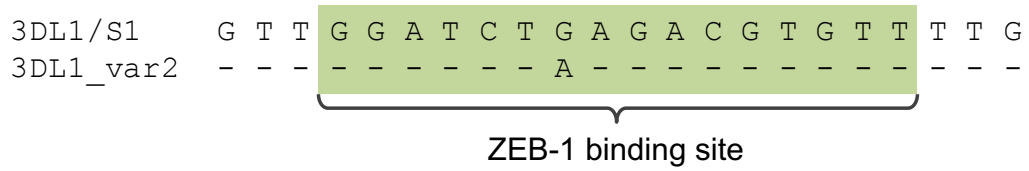


Figure 5.21 The ZEB-1 binding motif is disrupted upstream of some KIR3DL1 alleles

A sequence alignment of the region 258-237 bp upstream of the KIR3DL1/S1 start codon reveals that the ZEB-1 binding motif (green box) of the proximal promoter exhibits polymorphism. This g.-249G>A substitution was observed upstream of several different KIR3DL1 CDS allele groups. These alleles have been grouped here as "3DL1_var2". All other KIR3DL1/S1 alleles encode the 'G' variant at this position. Dashes (-) represent homologous sequence.

Also within the proximal promoter, every KIR3DS1 5′ UTR sequence, but no KIR3DL1 5′ UTR sequence, features a g.-173T>C substitution within the putative YY-1 binding site (Figure 5.22) [155].

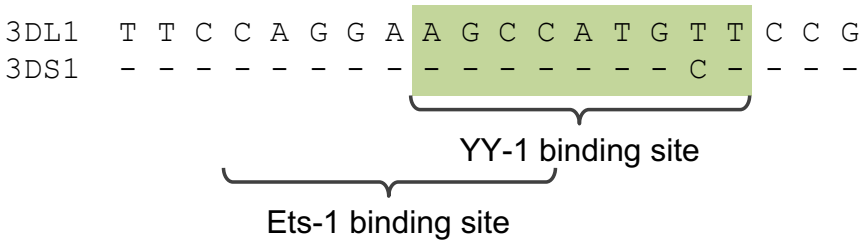


Figure 5.22 The YY-1 binding motif is disrupted upstream of some KIR2DL3 alleles

A sequence alignment of the region 188-169 bp upstream of the KIR3DL1/S1 start codon reveals that the YY-1 binding motif (green box) of the proximal promoter exhibits polymorphism. This g.-173T>C substitution is a key characteristic to stratify KIR3DL1 and KIR3DS1 allele promoter sequences. The overlapping Ets-1 TF binding site is unaffected. Dashes (-) represent homologous sequence.

Furthermore, the g.-65A>G substitution, located within the SP-1/E2F TF binding site, may also interfere with TF binding (Figure 5.23). This polymorphism is associated with almost every KIR3DL1*00501, 053, 099 and 109 allele. Interestingly, one occurrence of a KIR3DL1*00501 allele was observed that does not feature this substitution and so may exhibit cell surface expression more similar to alleles encoding the ‘A’ variant at this position.

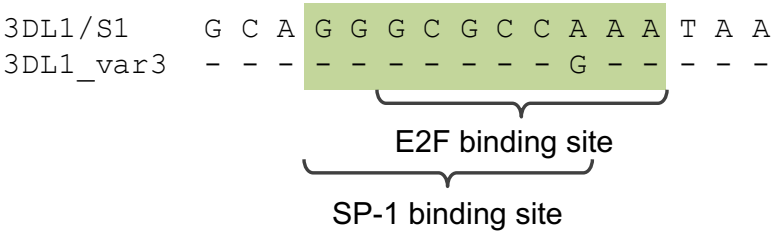


Figure 5.23 The SP-1/E2F binding motif is disrupted upstream of some KIR3DL1 alleles

A sequence alignment of the region 75-60 bp upstream of the KIR3DL1/S1 start codon reveals that the SP-1/E2F binding motifs (green box) of the proximal promoter exhibits polymorphism. This g.-65A>G substitution was observed upstream of several different KIR3DL1 CDS allele groups, including KIR3DL1*053, 099, 109 and almost all 00501 alleles. These alleles have been grouped here as “3DL1_var3”. All other KIR3DL1/S1 alleles encode the ‘A’ variant at this position. Dashes (-) represent homologous sequence.

The final polymorphism within the KIR3DL1/S1 putative TF binding sites is located in the SP-1 binding site (Figure 5.24). A substitution, g.-26G>A, is characteristic of every KIR3DL1*00101, 00401, 00402, 00501, 009, 019, 053, 099, 109 and KIR3DS1 5′ UTR, without exception.

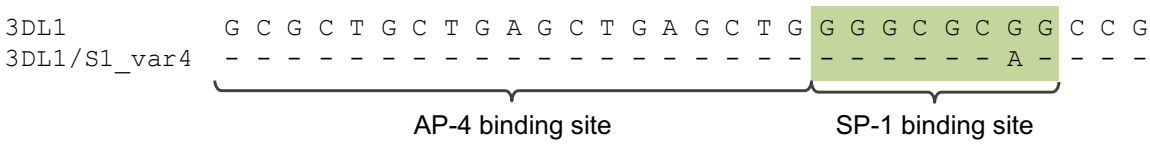


Figure 5.24 An SP-1 binding motif is disrupted upstream of some KIR3DL1/S1 alleles

A sequence alignment of the region 51-22 bp upstream of the KIR3DL1/S1 start codon reveals that the SP-1 binding motif (green box) of the proximal promoter exhibits polymorphism. This g.-26G>A substitution was observed upstream of a several different KIR3DL1/S1 CDS subgroups. These alleles have been grouped here as "3DL1/S1_var4". All other KIR3DL1 alleles encode the 'G' variant at this position. The adjacent AP-4 TF binding site is unaffected. Dashes (-) represent homologous sequence.

Although a functional assessment of the impact of each of these polymorphisms was beyond the scope of this thesis, the potential impact based on previous analysis is discussed in Section 5.08.

5.06.04 Unreported transcription factor (TF) binding sites

In addition to published TF binding sites, I also performed a search for novel TF binding elements introduced by novel polymorphism observed in this cohort. Interestingly, the TF, GATA3, was predicted to bind in a region adjacent to the Oct-1 and Ets-1 binding sites of the KIR2DL1 and KIR3DL1/S1 intermediate promoters (330 to 325 bp upstream of the start codon). Prior published investigations did not include data on this TF binding element. In addition, two putative binding sites were detected for nuclear factor of activated T cells, cytoplasmic 1 (NFATC1). The first was specific

to KIR2DL2/3 sequences, as the g.-425A>G substitution of all KIR2DL2/3 alleles (relative to KIR2DL1 and KIR3DL1 sequence at this position) is responsible for introducing the correct core binding element for NFATC1. This location, in close proximity to the dual AP-1 binding elements of the KIR2DL1 and KIR3DL1 proximal promoter, may imply a previous functional relevance. The second instance of the NFATC1 binding element was observed further upstream, approximately 700 bp before the start codon. As such, sequence was only available for KIR3DL1 (5′ UTR sequence for KIR2DL1 and KIR2DL2/3 alleles did not extend this far). In addition, a disrupted Oct-1 binding site was also observed in close proximity to this upstream NFATC1 binding element. No novel TF binding sites exhibited polymorphism in the sequences generated in this HCT cohort.

5.06.05 Putative intron 2 TF binding sites

The putative TF binding sites located in intron 2, responsible for generating antisense transcripts capable of reducing KIR cell surface expression [380], were investigated to determine the presence of polymorphism within and between the KIR2DL1, KIR2DL2/3 or KIR3DL1/S1 sequences generated for this HCT cohort. As such, the first 300 bp of intron 2 from every sequence were aligned and the MZF-1 and myc-C/EBP binding sites located.

Several key differences were noted between the different loci (Figure 5.25). In the core MZF-1 binding element, located approximately 100 bp after the end of exon 2, every KIR2DL1 features a G>A substitution compared to KIR2DL2/3 and KIR3DL1/S1. In addition, the adjacent nucleotide of KIR2DL2/3 sequences is substituted, also G>A.

Finally, the ‘GGGG’ that characterises the 3′ end of the binding site of KIR2DL1 and KIR2DL2/3 is altered in KIR3DL1/S1 sequences to ‘CGGA’.

Figure 5.25 Intron 2 TF binding element polymorphisms

Sequence alignments of the key TF binding sites in intron 2. A) The core MZF-1 binding site, approximately 100 bp downstream of the end of exon 2, features several inter- and intra-locus polymorphisms. In addition to the AG>GA substitutions, a variant of KIR2DL2 contains a deletion of G towards the 3′ end of the binding site. In addition, a substitution, G>T, characterises a variant of KIR3DL1. B) Polymorphism is also present across the myc-C/EBP binding site. In addition to inter-locus differences, a small subset of KIR2DL2 alleles feature a G>A substitution. Dashes (-) and dots (.) represent homologous and deleted bases, respectively.

Within the myc-C/EBP binding site, located approximately 130 bp further downstream, KIR2DL1 and KIR2DL2/3 differ only by a single G>T substitution 228 bp into intron 2. However, the KIR3DL1/S1 sequences each differ by substitutions adjacent on either side of the above SNP (T>G and G>A), in addition to an AG>GT dinucleotide substitution at the 3′ end of the myc binding element.

Interestingly, polymorphism also exists within the sequences of KIR2DL2/3 and KIR3DL1/S1 (Figure 5.25). One sequence of KIR3DL1 features a substitution in the

MZF-1 region (g.1176G>T) that has not been observed previously. Additionally, one sequence of a KIR2DL2 MZF-1 binding element is characterised by a deletion of a 'G'. However, the location of this deletion – within the short G homopolymer at the 3′ end of the binding site – raises doubt over the validity of the sequence. Finally, within the myc binding element further downstream, a single substitution within KIR2DL2 alleles (g.1215G>A) was observed eight times in this cohort.

5.07 Results: KIR2DL2/3~KIR2DL1 haplotype estimations

Strong linkage disequilibrium (LD) has been observed previously between the KIR2DL1 and KIR2DL2/3 loci [383]. However, allelic resolution KIR typing of a UK population has not been assessed for LD before and, as such, I sought to investigate this further in my cohort.

5.07.01 Data simplification

Although some efforts have been made to resolve KIR haplotypes at a genomic resolution (i.e. including CDS, introns and intergenic sequence) [87,88], the only analysis on a population scale is restricted to the first field of allelic nomenclature (that which corresponds to amino acid sequence) [165]. Furthermore, the vast array of KIR2DL2/3 non-coding polymorphism reduces individual allele frequencies substantially. As such, for the purposes of this analysis, high resolution allele types from my analysis were converted to CDS resolution (second field allelic nomenclature).

To further simplify the analysis, each haplotype was assumed to contain exactly one allele of each gene. As such, samples with three or more copies of either KIR2DL1 or KIR2DL2/3 were excluded from analysis. Furthermore, inference of KIR copy number

was required in samples without two allele types at each locus. As such, in samples with only one KIR2DL2/3 allele type, it was assumed that the sample was homozygous for this allele and a second allele type matching the first was inferred. For KIR2DL1, however, presence of only one allele type may reflect either a sample with homozygous KIR2DL1 alleles or a sample containing one haplotype that does not feature the KIR2DL1 locus. In these instances, a second KIR2DL1 allele type was generated to be either a repetition of the same allele (homozygous) or "KIR2DL1*absent".

5.07.02 Haplotype estimation and linkage disequilibrium (LD)

Calculation of probable diplotypes and LD analysis was performed in collaboration with Anthony Nolan Research Institute's Bioinformatic Department using in-house software (Cactus, Anthony Nolan, London, UK) that utilises an expectation-maximisation methodology [384]. As such, the likelihood of pairs of KIR2DL1 and KIR2DL2/3 alleles appearing in combination was assessed and the most probable option selected. A full list of the resulting haplotype frequencies are provided in Supplementary Table J.

As a comparison, an equivalent analysis was performed using the data available from the study by Vierra-Green *et al.* (2012) [165], herein referred to as the USA dataset. As a cohort of European-Americans, the ethnicity is reasonably comparable. However, to refine the USA dataset to better match the CNV constraints placed on my own HCT cohort, individual haplotypes were excluded when neither KIR2DL2 nor KIR2DL3 alleles were present or when two KIR2DL1 alleles were encoded within a single haplotype. From the refined USA dataset, allele and haplotype frequencies at first field KIR allele resolution were generated. Following this, commonly occurring alleles and haplotypes in the UK HCT cohort, accounting for more than 95% of alleles and 81% of

haplotypes, were compared to the findings from the USA dataset. The comparison revealed that highly similar allele and haplotype frequencies were obtained for each of the KIR alleles and haplotypes assessed, despite the different resolution of allele typing (Table 5.4). Although it was possible to distinguish the KIR2DL1*00401 and 00402 alleles in the UK HCT cohort, combined allele and haplotype frequencies are given in Table 5.4 to be concordant with those in the USA dataset.

Table 5.4 Comparison of common KIR2DL1 and KIR2DL2/3 allele and haplotype frequencies between the USA and UK HCT cohort datasets

	USA dataset frequency	UK HCT cohort frequency
KIR2DL1 alleles		
2DL1*003	0.372	0.384[§]
2DL1*002	0.232	0.276
2DL1*absent	0.192	0.140
2DL1*004	0.140	0.123[†]
2DL1*001	0.050	0.040
KIR2DL2/3 alleles		
2DL3*001	0.377	0.435[‡]
2DL3*002	0.241	0.241[±]
2DL2*001	0.174	0.140
2DL2*003	0.145	0.124
2DL3*005	0.045	0.044
Haplotypes		
2DL3*001~2DL1*003	0.366	0.373[§‡]
2DL3*002~2DL1*002	0.231	0.226[±]
2DL2*003~2DL1*absent	0.141	0.115
2DL2*001~2DL1*004	0.123	0.108[†]

[§] The UK HCT cohort KIR2DL1*003 allele/haplotype frequency encompasses the total frequency of this protein, including the synonymous KIR2DL1*00302, 00302_c.853A>C and 00302_c.963G>A CDS alleles.

[†] The UK HCT cohort KIR2DL1*004 allele/haplotype frequency encompasses the total frequency of this protein, including the synonymous KIR2DL1*00401 and 00402 CDS alleles.

[‡] The UK HCT cohort KIR2DL3*001 allele/haplotype frequency encompasses the total frequency of this protein, including the synonymous KIR2DL3*00101, 00110 and 00101_c.549C>T CDS alleles.

[±] The UK HCT cohort KIR2DL3*002 allele/haplotype frequency encompasses the total frequency of this protein, including the synonymous KIR2DL3*00201, and 00201_c.618A>C CDS alleles.

The coefficient of LD (D) and its normalised value (D′) were calculated. D′ values represent the probability that the two alleles associate non-randomly within a genome

such that pairs with a value of 0 are in perfect linkage equilibrium (i.e. randomly associated) and pairs with a value of 1 are in perfect LD (i.e. non-randomly associated). This revealed that, for each of the high frequency (>10%) predicted haplotypes, strong LD was observed (Table 5.5, a complete list of LD values for each predicted haplotype is included in Supplementary Table J). Furthermore, the global LD between the genes was also calculated. This revealed a D′ value of 0.870, further suggestive of strong LD between the KIR2DL1 and KIR2DL2/3 loci.

Table 5.5 Linkage disequilibrium analysis of the four most common KIR2DL2/3~2DL1 haplotypes

KIR2DL2/3 allele	KIR2DL1 allele	Coefficient of LD (D)	Normalised coefficient of LD (D′)
2DL3*00101	2DL1*00302	0.205	0.920
2DL3*00201	2DL1*00201	0.159	0.912
2DL2*00301	2DL1*absent	0.098	0.912
2DL2*00101	2DL1*00401	0.088	0.866

5.08 Discussion

The above analysis has highlighted the extraordinary allelic diversity encoded within three KIR loci: KIR2DL1, KIR2DL2/3 and KIR3DL1/S1. This was achieved by application of an ultra-high resolution typing strategy that utilised fully-phased allele sequencing across the entire length of each gene. Seventy-two unique CDS were observed across the three loci, corresponding to 63 distinct proteins and encompassing 28 novel CDS. The majority of novel CDS were characterised by non-synonymous mutations. Furthermore, many of the amino acid changes were located in regions of known importance or resulted in significant changes to residue characteristics. Further assays are warranted to investigate the impact these polymorphisms have on the functionality of the encoded proteins.

Although substantially polymorphic overall, the majority of KIR2DL1 and KIR2DL2/3 allele sequences were dominated by only two different CDS alleles (KIR2DL1*00302 and 00201; KIR2DL3*00101 and 00201). In each case, these alleles comprise over two thirds of the total allele frequency. However, for the KIR3DL1/S1 locus, no dominant allele is observed. Instead, five distinct CDS (KIR3DS1*01301, KIR3DL1*00501, 00101, 00401 and 002) each share between 12 and 18% of the total allele frequency. As more variants are observed more frequently, this finding supports the hypotheses that alleles of the KIR3DL1/S1 locus are more ancient and under a stronger influence of balancing selection [97,385,386].

The patterns of polymorphism support the hypothesis that KIR A haplotype motif structures exhibit more allelic polymorphism than KIR B haplotype motif structures [88]. More distinct KIR2DL3 (Cen-A associated) CDS were observed when compared to KIR2DL2 (Cen-B associated). Similarly, KIR3DL1 (Tel-A-associated) exhibited considerably more extensive CDS variation than KIR3DS1 (Tel-B-associated). When considering KIR2DL1, more CDS diversity was observed with KIR2DL1 alleles associated with the Cen-A motif than those associated with the Cen-B motif [348]. Furthermore, this may be extended to non-coding polymorphism. For example, within the non-coding polymorphism analysis on each common KIR2DL1 CDS subgroup, a broader range of genomic variants were observed with high frequency for the KIR2DL1*00101, 00201 and 00302 CDS subgroups (Cen-A associated) than for the KIR2DL1*00401 CDS subgroup, where only the KIR2DL1*0040101 allele was observed at high frequency. This provides compelling evidence that KIR-A haplotypes exhibit more allelic diversity than KIR-B haplotypes [87].

Although less allelic diversity was observed in genes associated with KIR B haplotype motifs, extensive CNV is a key feature of KIR B haplotype polymorphism. A series of well-defined KIR copy number haplotypes have been described following investigation into thousands of samples from a range of different ethnic origins [86]. Although it has been demonstrated that, in a UK cohort similar to that studied here, 11 common copy number haplotypes, comprised of only five different gene motifs, constitute almost 95% of the KIR haplotypes observed [90], there remains a very broad range of rarer haplotypes. The complexity of these is increased by recombination between different KIR genes causing gene duplication, deletion and hybrid gene formation. Although the SMRT DNA sequencing strategy used to sequence KIR alleles in this cohort is not able to quantify CNV, several samples with unusual CNV were, nonetheless, detected. This was possible as three or more different alleles were observed at a given locus. Unfortunately, as only individual loci were targeted, it was not possible to determine any more information from these samples. However, a more complete analysis, including sequencing of all KIR genes and the intergenic regions of the affected samples may help to decipher recombination events that resulted in the unusual CNV observed.

Although extensive CDS polymorphism was observed, it was within the non-coding regions that the most allelic diversity was detected. Over 1,500 different unique allele sequences were generated from only three loci of this relatively small cohort. As an interesting comparison, examination of the complete classical HLA class I (HLA-A, -B and -C) genotypes from a randomly selected population of 772 British/Irish North-West European Caucasoids on the Anthony Nolan VUD register, used as an equivalent population to the study cohort, resulted in only 314 unique alleles following full-length

sequencing (data not shown). Although HLA class I genes are considerably shorter (~3.5 kbp), comparison against these hyperpolymorphic loci demonstrates the enormous variation also exhibited by the three KIR loci tested in my clinical cohort. Interestingly, trimorphic and tetramorphic nucleotide positions were relatively infrequent, especially following subanalysis to remove problematic regions of repetitive DNA elements in KIR3DL1/S1 sequences. As such, the lack of CDS polymorphism may simply represent the relatively small sample size, when compared to the large scale studies performed on other polymorphic gene systems, such as HLA [63].

Although many of the genomic sequences differed by indel mutations, known to cause issues when sequencing by any current methodology, the majority of polymorphisms were characterised by substitutions. As demonstrated in Chapter 4, Section 4.02.09.04, SMRT DNA sequencing is able to accurately define these substitution polymorphisms. Following submission of only those sequences which differed to their most homologous published allele sequence by substitutions only, it was possible to increase the total number of alleles deposited in the IPD-KIR Database for the KIR2DL1 and KIR3DL1 genes by 100% and 28%, respectively. The percentage increase at the other tested loci was less substantial as sequences were more frequently affected by indel polymorphism and therefore not submitted to the IPD-KIR Database.

The majority of polymorphism within non-coding regions is not predicted to have functional relevance. However, polymorphism within splice sites and promoter regions can significantly affect the encoded protein molecule and its expression [378,387]. Although no polymorphism was observed in known KIR2DL1, KIR2DL2/3 or KIR3DL1/S1 splice sites, extensive polymorphism was observed across the 5′ UTR of

each of these loci. Although much of the polymorphism was not located within putative TF binding sites, many of the binding sites were disrupted. Several of these polymorphisms have previously been shown to affect KIR expression. For example, the g.-406G>C substitution, found upstream of most KIR2DL1*00302 alleles (and the interesting novel KIR2DL1*00401_c.13G>T) in the AP-1 binding site, has been demonstrated to reduce AP-1 binding and may be related to the decreased expression level of KIR2DL1*003 relative to KIR2DL1*002 [166,388]. However, the proportion of NK cells expressing KIR2DL1*003 is increased relative to some other KIR2DL1 alleles, perhaps indicating that the decreased AP-1 binding correlates with improved probability of expression (albeit at a lower level). Interestingly, a minority of KIR2DL1*003 alleles did not encode this substitution, introducing the possibility that not all KIR2DL1*003 alleles are expressed equally.

Several other polymorphisms within the TF binding sites upstream of KIR2DL1 were observed in the 5′ UTR sequence of the KIR2DL1*00401_c.13G>T allele. A CRE binding motif, located within the proximal promoter, is interrupted by a -119G>A substitution. This polymorphism, located at the 5′ end of the putative binding site, has not been observed in any previously published KIR2DL1 allele sequences. However, previous mutagenesis experimentation has demonstrated that interruption of the CRE binding element causes significant reduction in KIR transcription [377].

In addition, the -351G>T substitution, within the Ets-1 binding site, is otherwise associated with KIR2DL2/3 alleles, which notably utilise an alternative intermediate promoter region [166]. As such, it is likely that Ets-1 binding is interrupted, reducing transcription from this site. Characterisation further into the 5′ UTR may establish the

presence of an alternative intermediate promoter region as well as assist in establishing the origins of this potentially recombinant allele.

The SP-1 binding site located within the proximal promoter also exhibited polymorphism in both KIR2DL1 and KIR3DL1/S1 alleles. In KIR2DL1, the -28G>A substitution was observed whilst in some KIR3DL1/S1 allele sequences, the -26G>A substitution was noted. The functional impact of the -28G>A substitution observed in 11 KIR2DL1*00302 alleles is, as yet, unknown, although disruption of the same promoter element in KIR3DL1 and KIR3DS1 alleles has previously been speculated to reduce antisense transcription and thus have an upregulatory effect on cell surface KIR expression [378].

In studies investigating the effects of polymorphisms within the YY-1 TF binding site of KIR2DL1 and KIR3DL1/S1 alleles, intact binding sites, present in all KIR3DL1 alleles but not KIR2DL1 or KIR3DS1, was shown to reduce antisense transcription [378,389]. As such, the -176A>G substitution encoded by KIR2DL3*00501 and 010 alleles will likely disrupt the TF binding site, causing it to resemble the equivalent region in the KIR2DL1 proximal promoter. It is therefore likely to enhance antisense transcription, potentially reducing cell surface expression of these alleles.

The published indel polymorphism of KIR3DL1/S1 alleles 409 bp upstream of the start codon has previously been correlated with disruption of a putative AP-1 binding site in the intermediate promoter [166]. Although the majority of KIR3DL1*00501 alleles encode the deletion variant, and are characterised by reduced cell surface expression [388], it was possible to observe the insertion variation in a single example of a

KIR3DL1*00501 allele. This suggests that AP-1 binding upstream of this allele may be rescued. The same KIR3DL1*00501 allele also escapes a downregulatory polymorphism within the overlapping E2F and SP-1 binding site in the proximal promoter. Here, the -65A>G substitution associated with all other KIR3DL1*00501 alleles introduces an additional methylation site on the antisense DNA strand and has previously been demonstrated to reduce transcription [390]. In combination, these polymorphisms may explain the reduced frequency of expression of KIR3DL1*005 relative to many other KIR3DL1/S1 alleles [388] and further demonstrates that, alone, CDS genotyping is not always sufficient to determine certain phenotypic qualities.

Antisense transcription derived from a promoter within the intron 2 sequence of KIR2DL1 and KIR3DL1 has previously been implicated as an additional method by which NK cells silence particular genes. However, antisense transcripts from intron 2 of KIR2DL2/3 have not been detected thus far but have not been specifically searched for. In this study, I demonstrated that the core MZF-1 and upstream myc-C/EBP binding elements appear completely conserved in KIR2DL1, KIR2DL3 and KIR3DS1, whilst it was possible to detect rare polymorphisms within the MZF-1 binding region of both KIR2DL2 and KIR3DL1. The myc-C/EBP binding element of KIR2DL2 also featured polymorphism. Although mutations within the MZF-1 binding region have previously been demonstrated to disrupt TF binding in KIR2DL1 [380], the effects of these particular polymorphisms require further investigation.

In addition to investigation into the extent of allelic polymorphism, the allelic haplotype frequencies were also calculated. Because the methodology utilised in this study to determine KIR gene presence/absence status does not define CNV, and because existing

studies with considerably larger sample sizes have already defined CNV haplotype structures, I did not investigate CNV haplotype frequencies in this analysis. Instead, comparison between the KIR2DL2/3~KIR2DL1 allele-level haplotype data generated in this study and that previously published on a similarly sized European American cohort revealed remarkably similar population structures. This is a further validation of the KIR typing strategy described in Chapter 4.

With regards to the strong global LD observed between KIR2DL2/3 and KIR2DL1, the close proximity of these genes on a KIR haplotype is the likely cause of the high degree of LD. However, when the D′ values from less frequently observed haplotypes are investigated, many appear to feature lower values suggestive of random association. It is likely that this results from the low frequency of these haplotypes and the relatively small sample size afforded by this UK HCT cohort.

Several other limitations exist within the current analysis. Although the overall sequencing success rate was high, several samples resisted amplification. As long DNA fragments are required to permit long-range PCR amplification (see Chapter 4, Section 4.02.08), this may indicate that DNA samples that failed to amplify were of insufficient quality for this assay. Alternatively, unknown polymorphism within primer binding sites may also prevent efficient PCR amplification. Future investigation into methods utilising capture and enrichment of long DNA fragments encoding KIR genes, similar to the method described by Norman *et al.* (2016) [94], is warranted as a possible approach to alleviate these problems. However, a robust accompanying analysis that does not mis-assign homologous allele regions is also required.

In addition to samples that failed to amplify, several samples did not generate allele sequences for genes that, by previous KIR presence/absence genotyping, were known to be encoded by certain samples. This issue was also encountered when sequencing KIR alleles from IHIW cell line samples and is discussed further in Chapter 4, Section 4.06. The exclusion of these samples from the analysis may have resulted in a reduction in the total polymorphism observed or may potentially have slightly skewed some of the estimated allele/haplotype frequencies.

The other main limitation of this current analysis is the lack of protein/expression assays to determine the functional impact of the observed polymorphisms. Although much of the promoter polymorphism has been investigated previously, the novel mutations observed in this study have often not been investigated previously [101,106,156,174,348,381]. A thorough analysis of mature protein expression, ligand avidity and signalling capacity for each of the novel CDS alleles described may reveal some interesting results.

Finally, although requiring a larger dataset (and beyond the scope of this thesis), an evolutionary analysis of KIR sequences may uncover noteworthy information relating to the genetic selection of different KIR alleles and explain the vast amount of polymorphism observed at this locus.

5.09 Conclusions

By applying a full-length, unambiguous KIR allele sequencing strategy to a cohort of several hundred samples, an enormous amount of novel polymorphism has been observed in just three KIR loci. Although the majority of polymorphism resides within

non-coding regions that are not predicted to influence the function of the gene, differences across the promoter regions and CDS of each sequenced gene will undoubtedly affect the expression and function of these molecules. Whilst presence/absence genotyping assays are largely unable to decipher information relating to this functional polymorphism, I have been able to obtain much higher resolution allelic information that will guide an improved analysis into the role of KIR polymorphism on the outcomes of HCT in this UK cohort.

Chapter 6 The impact of allelic polymorphism at the KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 loci on the outcomes of HCT

6.01 Introduction

As discussed in Chapter 3, the impact of KIR polymorphism, in the form of presence or absence genotyping, has been extensively investigated in a number of different transplantation settings. In the concluding remarks of several previous publications, the role of allelic polymorphism, and it's potential to significantly alter the outcomes of HCT, is suggested as worthy of further investigation [182,270,272,339,391-397]. However, to date, only a few studies to investigate this have been undertaken [270,393]. Nonetheless, the group of Bari *et al.* [169,270,398], who characterised the impact of KIR2DL1 allelic polymorphism on the recruitment of SHP to the receptor molecule, demonstrated that the same polymorphism, a substitution of arginine for cysteine at residue 245 (R245C), when encoded by donors, relayed a detrimental impact on the risk of relapse and OS in paediatric allogeneic HCT outcomes. Interestingly, this donor polymorphism appeared to have equal impact regardless of recipient disease. However, the validity of this initial analysis remains under doubt following realisation that the genotyping methodology utilised had potential to cross-react with alternative KIR loci [399].

Furthermore, common polymorphism within other KIR genes has also been shown to have functional impact. Substitution of serine for leucine at residue 86 of KIR3DL1 has been shown to prevent cell surface expression with additive effects resulting from polymorphism at residue 182 [153]. As demonstrated in Chapter 5, Section 5.05.02,

these alleles, including KIR3DL1*004 and 019, constitute approximately 15% of the overall allele frequency at the KIR3DL1/S1 locus. As many haploidentical HCT donors are selected according to the presence of predicted KIR receptor-ligand alloreactivity, including that proposed from HCT involving KIR3DL1+ve donors and HLA-Bw4-ve recipients, the correct determination of expressed KIR alleles, not just those that are encoded as null alleles, may be a potential method to improve outcomes in these recipients [356].

Furthermore, the variable expression of KIR3DL1 alleles, and the differing avidity to HLA-Bw4-80T/I alleles, has prompted further investigation in the HLA-matched VUD HCT setting. Here, donor samples predicted to encode weak or no inhibitory KIR3DL1-HLA-Bw4 interactions were associated with a stronger cytotoxic response than those in which a strong inhibitory interaction was predicted. When translated into a clinical scenario, the effect correlated with reduced relapse and mortality [393].

When considering the KIR2DL2/3 locus, although polymorphism has been demonstrated to result in significantly different cytotoxic capabilities [339], current literature is lacking any in-depth analysis into the effects of allelic polymorphism on HCT outcomes. However, as has been demonstrated in Chapter 3 and by others [273,275,400], the centromeric gene content variation (i.e. Cen-A *vs* Cen-B) of donor genotype does influence the outcomes of HCT. As a key defining feature of centromeric motif content, the allelic variation at the KIR2DL2/3 locus is likely to be associated with distinct outcomes.

In this chapter, I investigate the effect of allelic variation at each of the KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 loci on the outcomes of HCT in a VUD, UK, AML cohort. I focus predominantly on donor KIR allele genotype but also utilise recipient KIR allele typing to explore the impact of GVH direction KIR allele mismatching on the outcomes of HCT. Allelic KIR genotypes of recipients were not explored independently of donor KIR genotype as there was little evidence from analysis of recipient KIR copy number genotypes to suggest that further investigation was warranted (Chapter 3, Section 3.06.02).

6.02 Analysis strategy

Previous analysis focussing only on the KIR presence/absence genotype of donors and recipients revealed many of the previously applied KIR analysis models were not well suited to this unique UK HCT cohort. To determine whether the models of KIR allelic analysis proposed by Bari *et al*. (2013) [270] or Boudreau *et al.* (2017) [393] were suitable in this cohort, equivalent analyses were performed.

In addition, other key defining polymorphisms of the alleles at each of the sequenced loci were assessed for their influence on HCT outcomes. Many of these have previously been associated with differences in KIR allele cell surface expression level [153,156,349,352,388], ligand affinity [105,106,112,174,175,298,339,348,349,381], functionality [152,339,357] and probability of expression [378,388-390]. Analysis was limited to only utilise KIR allele types at the most basic, protein-level resolution to simplify the findings and ensure sufficient population sizes for meaningful statistical analysis. As the absence of KIR2DL1 from a donor's phenotype (either by absence of the gene or presence of only null KIR2DL1 alleles) may mask the effects of allelic

polymorphism within the protein (see Section 6.03.01, below), KIR2DL1 univariate analyses were restricted to exclude such donors and all multivariate analysis adjusted for this variable, unless otherwise stated.

Further investigation utilised samples with complete KIR2DL1 and KIR2DL2/3 allele typing to assess the impact of predicted KIR2DL2/3~KIR2DL1 allelic haplotypes on HCT outcomes. As the four most common allelic haplotypes comprised over 80% of the total frequency, analysis was restricted to only those samples that encoded only common haplotypes.

Finally, the allele matching status at each sequenced locus was established to determine the effect of allele-resolution mismatching. As KIR allele matching is not performed routinely prior to HCT, and because of its high degree of polymorphism, it was only possible to investigate the impact of GVH direction KIR allele mismatching on the outcomes of HCT, as too few fully KIR allele matched transplants exist to provide a meaningful analysis.

As findings from the KIR presence/absence genotyping analysis had implicated strong opposing effects relating to different HCT preparative regimens in this UK cohort, KIR allele variables were also explored following stratification of conditioning. However, as it was not possible to complete allelic resolution KIR genotyping in all samples, sub-cohorts to investigate the impact of KIR2DL2/3 and KIR3DL1/S1 allelic polymorphism according to conditioning regimen were extended to also include all transplants within each conditioning strata (i.e. HLA mismatched and paediatric transplants were also assessed). This was performed to increase sub-cohort size, increasing the statistical

power of the analyses. Where appropriate, HLA matching and patient age group were included in multivariate analysis models to correct for variation resulting from these factors.

6.03 Results: an analysis of the effects of donor KIR2DL1 allelic polymorphism on HCT outcomes

6.03.01 Absence of expressed KIR2DL1

The absence of KIR2DL1 from an individual's phenotype is most frequently caused by complete absence of the gene. However, the KIR2DL1 negative phenotype is also possible when the gene is still present in the genotype, as the KIR2DL1*032N allele is present in almost 1% of haplotypes (see Chapter 5, Section 5.03.02). Therefore, particularly if only one KIR2DL1 positive haplotype is encoded (and it encodes only the KIR2DL1*032N allele) then it is possible to observe a KIR2DL1 negative phenotype in KIR2DL1 positive genotype individuals. This phenomenon was detected three times in my HCT cohort and, as such, accounts for approximately 14% (3/21) of KIR2DL1 negative phenotype individuals.

To assess whether the absence of KIR2DL1 from a donor's KIR phenotype was correlated with any HCT outcome, estimated probabilities of OS, DFS and relapse at five years, as well as NRM at one year post-transplant were calculated, stratifying on this variable. In addition, the incidence of grades 2-4 aGVHD were also assessed. Although the number of transplants involving KIR2DL1 negative phenotype donors was small (n=11), a strong detrimental effect on both five year OS and DFS was observed (OS: 18.2% *vs* 36.9%, p=0.033; DFS: 18.2% *vs* 33.4%, p=0.028, Table 6.1,

Figure 6.1), although no significant difference was observed for relapse, NRM or grades 2-4 aGVHD incidence.

Table 6.1 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL1 expression phenotype in the overall cohort

KIR2DL1 phenotype	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Expressed *vs* non-expressed	***0.033***	***0.028***	0.37	0.56	1.00

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table K.

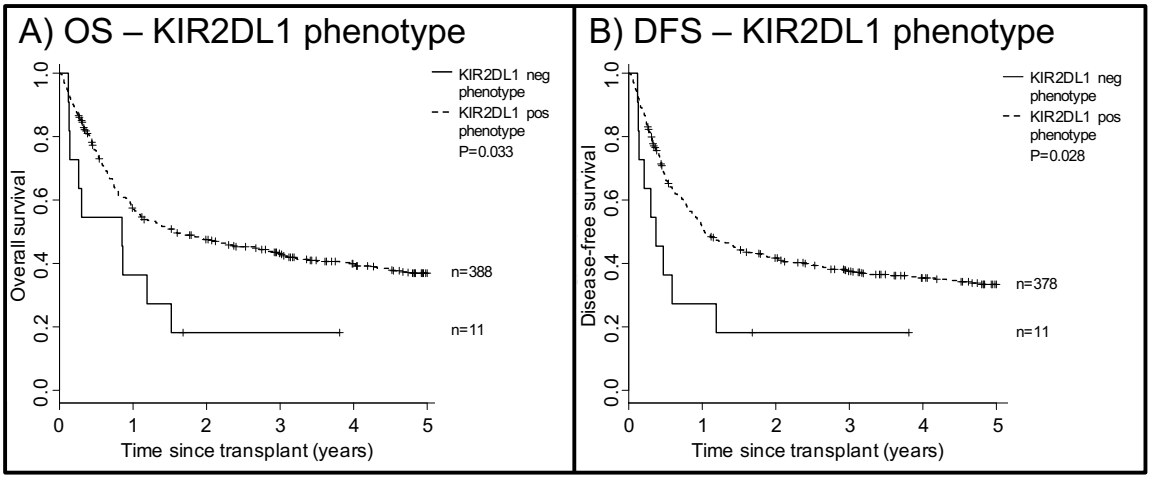


Figure 6.1 Probability of OS and DFS comparing donor KIR2DL1 expression phenotypes in the overall cohort

A) Univariate Kaplan-Meier analysis of the probability of OS at five years post-transplant reveals a significantly increased risk of mortality for recipients of transplants that involve donors with a KIR2DL1 negative phenotype compared to transplants where the donor expresses KIR2DL1 (p=0.033). B) The same transplant stratification is also associated with reduced DFS probability (p=0.028).

As this analysis suggested that donor KIR2DL1 phenotype may affect HCT outcomes, all further univariate analysis assessing the impact of KIR2DL1 allelic polymorphism excluded transplants involving KIR2DL1 negative phenotype donors.

6.03.02 The R245C polymorphism

6.03.02.01 The influence of donor KIR2DL1 R245C polymorphism within the overall cohort

The R245C substitution within the KIR2DL1 protein inhibits the ability of the receptor to recruit ß-arrestin and SHP-2 to the immunological synapse and, consequently, reduces the inhibitory signalling capacity of some KIR2DL1 alleles, including KIR2DL1*004 [169]. To investigate the impact of this polymorphism in the UK cohort, transplants were stratified according to the presence of either the R^{245} or C^{245} alleles encoded within the donor's genotype. Of the 18 transplants that were excluded from univariate analysis, nine involved donors who did not encode the KIR2DL1 gene, two involved donors whose KIR2DL1 allele genotype was monotypic for the KIR2DL1*032N allele, and six transplants involved donors without complete KIR2DL1 genotyping. The final excluded transplant featured a donor who encoded the KIR2DL1*037 allele. As this allele features the R245H amino acid substitution, and its effect on KIR2DL1 signalling is yet to be established [369], this transplant was also omitted from the current analysis.

The estimated probabilities of OS, DFS, relapse, NRM and aGVHD were calculated for the different groups of transplants according to R245C polymorphism (Table 6.2). This demonstrated that the presence of R^{245} KIR2DL1 alleles was not associated with any significant differences in any of the assessed outcomes. However, the presence C^{245} alleles was associated with a significant increase in NRM (27.6 *vs* 18.7%, p=0.038, Figure 6.2A). Interestingly, this difference in NRM did not correspond to any significant differences between the probabilities of OS or DFS at five years post-

transplant, but may be associated with a reduced risk of relapse, although this test did not reach statistical significance (32.7% *vs* 41.3%, p=0.10, Figure 6.2B).

Table 6.2 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL1 R245C polymorphism in the overall cohort

Donor KIR2DL1 R245C polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
R^{245} pos *vs* R^{245} neg	0.33	0.76	0.29	0.16	0.64
C^{245} pos *vs* C^{245} neg	0.32	0.70	0.10	***0.038***	0.62
R/C^{245} heterotypic *vs* R^{245} monotypic *vs* C^{245} monotypic	0.52	0.92	0.24	0.10	0.85

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table K.

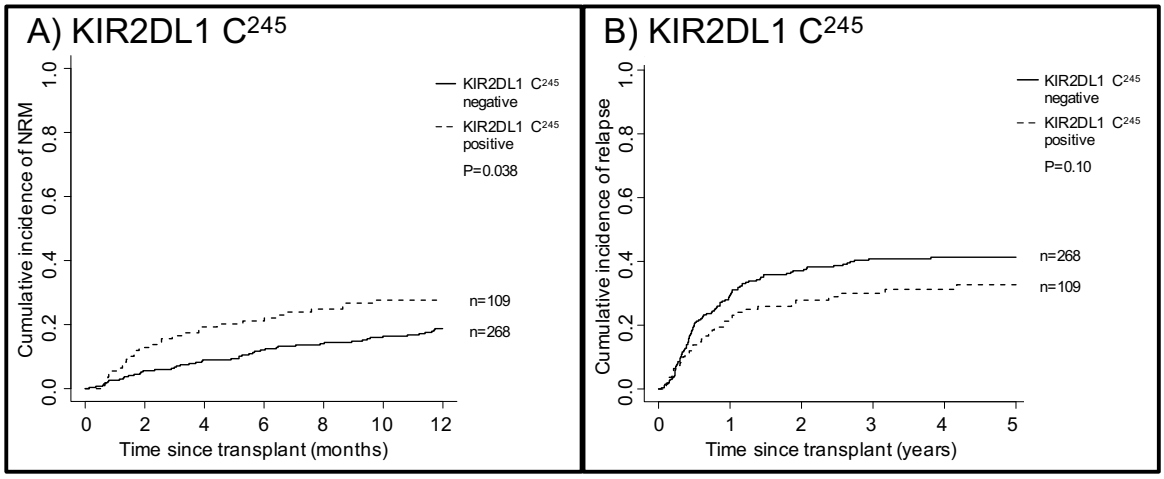


Figure 6.2 Probability of NRM and relapse comparing the presence of donor-encoded KIR2DL1 C^{245} alleles in the overall cohort

A) Univariate competing risk analysis of the cumulative incidence of NRM reveals a significantly increased risk associated with transplants that involve donors who encode at least one C^{245} positive allele of KIR2DL1 compared to transplants where the donor only encodes KIR2DL1 R^{245} alleles (p=0.038). B) The same transplant stratification may also be associated with reduced relapse risk but does not reach statistical significance (p=0.10).

To assess whether the observed effect on NRM withstood multivariate analysis, a competing risk regression analysis was performed (Table 6.3). This analysis included

adjustment for HLA mismatching and, as mentioned in Section 6.02, also included the absence of expressed KIR as category of KIR2DL1 allelic polymorphism. Multivariate analysis confirmed the significance of the increased risk of NRM at one year post-transplant associated with the C^{245} polymorphism (HR=1.71, CI=1.09-2.69, p=0.021).

Table 6.3 Multivariate analysis assessing the impact of donor KIR2DL1 C^{245} alleles on one year NRM risk in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA-matched	275	1.00	-	-
9/10 HLA-matched	90	1.30	0.77-2.18	0.33
<9/10 HLA-matched	23	***3.80***	***2.06-7.01***	***<0.001***
KIR2DL1 C^{245} presence				
Donor C^{245} neg	268	1.00	-	-
Donor C^{245} pos	109	***1.71***	***1.09-2.69***	***0.021***
Donor KIR2DL1 neg phenotype	11	1.30	0.40-4.29	0.66

Statistically significant results are denoted by ***bold italics***.

To assess whether the observed effect was influenced by heterotypic genotypes (i.e. donor encodes both C^{245} and R^{245}), transplants were stratified as such. Analysis of the same HCT outcomes revealed no significant difference between these three groups (Table 6.2).

For many of the samples with complete KIR2DL1 allele sequencing, KIR2DL2/3 allele typing was also complete, allowing the estimation of allelic-level haplotypes (Chapter 5, Section 5.07) and, importantly, probable copy number variation at both of the loci. HCT outcomes were assessed in cases in which it was possible to estimate the donor's KIR2DL1 copy number (n=310). As before, transplants involving donors with a KIR2DL1 negative phenotype or those involving KIR2DL1*037 positive donors were excluded from univariate analysis (Table 6.4).

Table 6.4 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL1 R245C CNV polymorphism in the overall cohort

Donor KIR2DL1 R245C CNV polymorphisms	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
0 *vs* 1 *vs* 2 R^{245} copies	0.90	0.90	0.23	0.12	0.72
0 *vs* 1 *vs* 2 C^{245} copies	0.54	0.69	0.20	0.060	0.25

The number of transplants within each subgroup of each test is given in Supplementary Table K.

This revealed only one statistical trend towards increased risk of NRM at one year post-transplant with each increasing copy of KIR2DL1 C^{245} positive allele (0 *vs* 1 *vs* 2 copies: 18.7% *vs* 28.9% *vs* 40%, p=0.060, Figure 6.3). Although only a statistical trend in univariate analysis, a multivariate analysis correcting for HLA-matching status indicated that recipients of transplants involving donors who encode one copy of a KIR2DL1 C^{245} allele were significantly more likely to suffer NRM at one year post-transplant (HR=2.01, CI=1.13-3.56, p=0.017, Table 6.5). Interestingly, although the HR associated with donors encoding two copies of KIR2DL1 C^{245} alleles was similarly increased, this was not a statistically significant finding, possibly owing to the small number of transplants in this subcategory.

Table 6.5 Multivariate analysis assessing the impact of donor KIR2DL1 C^{245} CNV polymorphism on one year NRM risk in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA-matched	214	1.00	-	-
9/10 HLA-matched	78	1.35	0.77-2.36	0.30
<9/10 HLA-matched	19	***4.25***	***2.09-8.67***	***<0.001***
KIR2DL1 C^{245} CNV				
0 donor KIR2DL1 C^{245} copies	228	1.00	-	-
1 donor KIR2DL1 C^{245} copies	66	***2.01***	***1.13-3.56***	***0.017***
2 donor KIR2DL1 C^{245} copies	10	1.99	0.85-4.68	0.11
Donor KIR2DL1 neg phenotype	7	1.76	0.48-6.47	0.39

Statistically significant results are denoted by ***bold italics***.

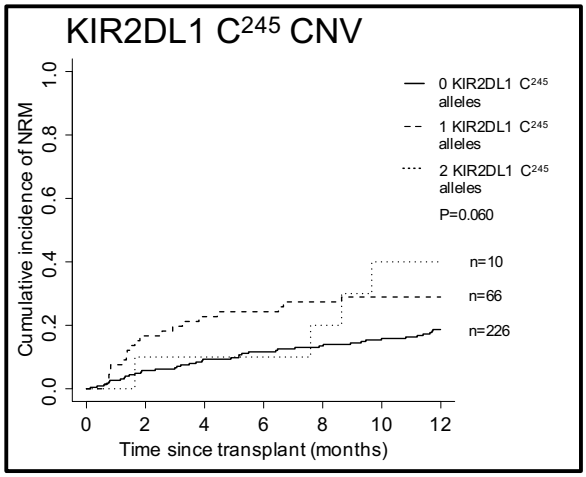


Figure 6.3 Probability of NRM comparing CNV of donor-encoded KIR2DL1 C^{245} alleles in the overall cohort

Univariate analysis of the cumulative incidence of NRM over the first 12 months post-transplant reveals a trend towards increased risk associated with each additional donor-encoded KIR2DL1 C^{245} allele (p=0.060).

6.03.02.02 The influence of donor KIR2DL1 R245C polymorphism within the RIC cohort

As the previous investigation into the role of KIR presence/absence polymorphism could be significantly altered by the conditioning regimen, analysis into the role of KIR allelic polymorphism also stratified the transplants according to this variable. Furthermore, as it was possible to complete the KIR2DL1 allele typing for the vast majority of transplants, the cohort was refined to only include adult, HLA-matched transplants.

When the donor KIR2DL1 R245C polymorphism was investigated in the RIC sub-cohort, its significance in increasing the risk of NRM at one year post transplant was lost (Table 6.6). Although a statistical trend was observed correlating the heterotypic residue 245 donor genotype with reduced incidence of grades 2-4 aGHVD (R/C^{245}

heterotypic=7.7%, R^{245} monotypic=20%, C^{245} monotypic=40%, p=0.076), some subgroups were composed of very small populations (<10 transplants) and, as such, this result may not be entirely reliable. As no clinical factors remained significant following forward stepwise selection of covariates, multivariate analysis was not performed.

Table 6.6 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL1 R245C polymorphism in the adult, HLA-matched, RIC cohort

Donor KIR2DL1 R245C polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
R^{245} pos *vs* R^{245} neg	0.53	0.37	0.56	0.95	0.095
C^{245} pos *vs* C^{245} neg	0.61	0.26	0.41	0.57	0.67
R/C^{245} heterotypic *vs* R^{245} monotypic *vs* C^{245} monotypic	0.58	0.47	0.69	0.83	0.076

The number of transplants within each subgroup of each test is given in Supplementary Table K.

To assess the impact of the KIR2DL1 R245C polymorphism CNV in the RIC HCT setting, all RIC transplants were assessed (the reduced number of samples in which it was possible to estimate KIR2DL1 allele copy number precluded further restriction to also exclude paediatric and HLA-mismatched transplants). Despite the wider RIC subcategory restrictions, the number of transplants involving donors who encoded two copies of KIR2DL1 C^{245} positive alleles was too small to perform a meaningful analysis (n=3). However, when investigating the CNV of donor-encoded KIR2DL1 R^{245} alleles, a statistical trend was observed linking each additional copy of R^{245} with a reduction in DFS probability (0 *vs* 1 *vs* 2 copies: 77.8% *vs* 30.4% *vs* 27.0%, p=0.070, Table 6.7, Figure 6.4). When the donor-encoded KIR2DL1 R^{245} allele copy number was assessed in a multivariate analysis, a significant reduction in DFS is observed when comparing no copies against two copies of R^{245} alleles (HR=4.55, CI=1.09-20.00, p=0.038, Table 6.8), although only a statistical trend was retained when comparing no copies with one copy of R^{245} alleles (HR=3.85, CI=0.90-16.67, p=0.069).

Table 6.7 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL1 R^{245} CNV polymorphism in the RIC cohort

Donor KIR2DL1 R^{245} CNV polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
0 *vs* 1 *vs* 2 R^{245} copies	0.37	0.070	0.17	0.80	0.40

The number of transplants within each subgroup of each test is given in Supplementary Table K.

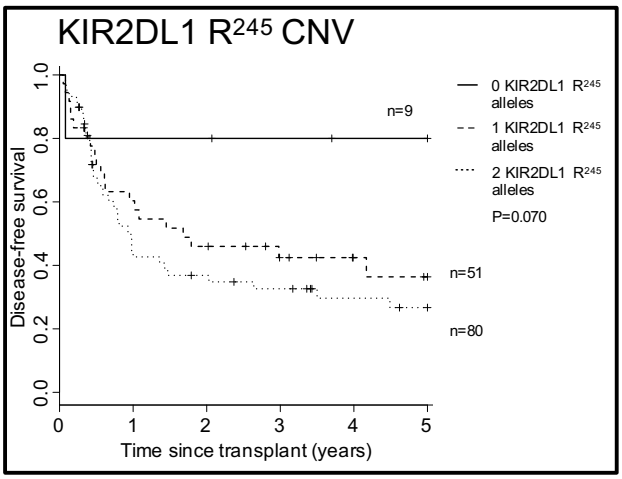


Figure 6.4 Probability of DFS comparing CNV of donor-encoded KIR2DL1 R^{245} alleles in the RIC cohort

Univariate analysis of the probability of DFS reveals a trend towards reduced DFS associated with each additional donor-encoded KIR2DL1 R^{245} allele (p=0.070).

Table 6.8 Multivariate analysis assessing the impact of donor KIR2DL1 R^{245} CNV polymorphism on five year DFS probability in the RIC cohort

Variable	5 year DFS			
	N	HR	95% CI	*p-value*
Previous autografts				
0	136	1.00	-	-
≥1	8	1.49	0.68-3.23	0.32
KIR2DL1 R^{245} CNV				
0 donor KIR2DL1 R^{245} copies	9	1.00	-	-
1 donor KIR2DL1 R^{245} copies	51	3.85	0.90-16.67	0.069
2 donor KIR2DL1 R^{245} copies	80	***4.55***	***1.09-20.00***	***0.038***
Donor KIR2DL1 neg phenotype	4	5.88	0.95-33.33	0.057

Statistically significant results are denoted by ***bold italics***.

6.03.02.03 The influence of donor KIR2DL1 R245C polymorphism within the MAC cohort

The KIR2DL1 R245C polymorphism was also investigated in the adult, HLA-matched MAC sub-cohort. Here, the detrimental effect of donor-encoded KIR2DL1 C^{245} alleles was most apparent. The probability of OS and DFS at five years post-transplant were significantly reduced when at least one KIR2DL1 C^{245} allele was encoded (OS: 21.6% *vs* 45.8%, p=0.004; DFS: 21.8% *vs* 43%, p=0.015, Figure 6.5, Table 6.9)

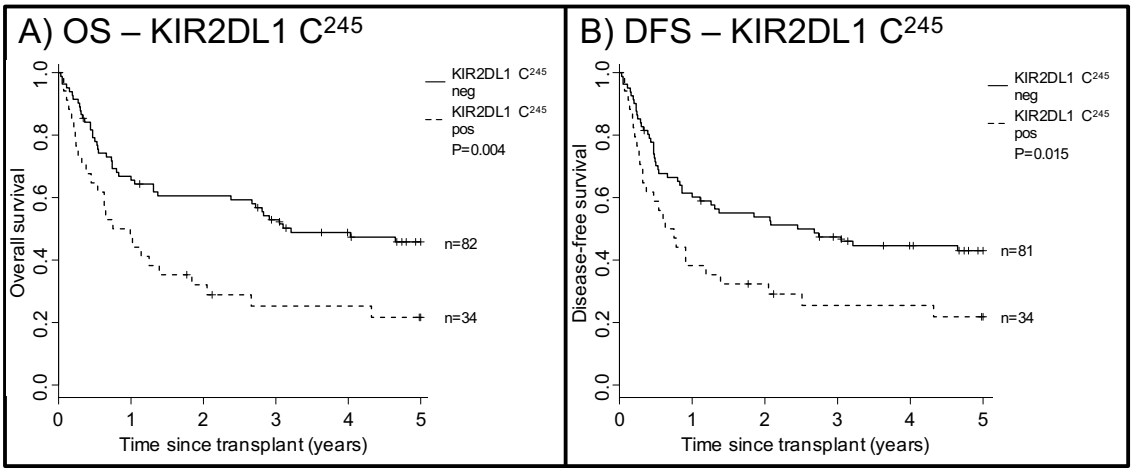


Figure 6.5 Probability of OS and DFS comparing the presence of donor-encoded KIR2DL1 C^{245} alleles in the adult, HLA-matched MAC cohort

A) Univariate Kaplan-Meier analysis of the probability of OS at five years post-transplant reveals a significantly increased risk of mortality for recipients of transplants that involve donors who encode at least one C^{245} positive allele of KIR2DL1 compared to transplants where the donor only encodes KIR2DL1 R^{245} alleles (p=0.004). B) The same transplant stratification is also associated with reduced DFS probability (p=0.015).

Table 6.9 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL1 R245C polymorphism in the adult, HLA-matched, MAC cohort

Donor KIR2DL1 R245C polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
R^{245} pos *vs* R^{245} neg	0.058	0.10	0.74	0.18	0.45
C^{245} pos *vs* C^{245} neg	***0.004***	***0.015***	0.44	0.092	0.75
R/C^{245} heterotypic *vs* R^{245} monotypic *vs* C^{245} monotypic	***0.015***	***0.047***	0.53	0.20	0.67

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table K.

To confirm the detrimental effect of donor-encoded KIR2DL1 C^{245} alleles in the adult, HLA-matched MAC HCT setting, multivariate analysis by Cox regression was performed (Table 6.10). This confirmed the findings from the univariate analysis for both OS and DFS at five years post-transplant (OS: HR=2.13, CI=1.28-3.45, p=0.003; DFS: HR=1.92, CI=1.16-3.13, p=0.010).

Table 6.10 Multivariate analysis assessing the impact of donor KIR2DL1 C^{245} alleles on five year OS and DFS in the adult, HLA-matched MAC cohort

Variable	5 year OS				5 year DFS			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
Recipient age, years								
<40	84	1.00	-	-	83	1.00	-	-
>40	34	***1.72***	***1.04-2.86***	***0.033***	34	***1.67***	***1.02-2.78***	***0.043***
Previous autografts								
0	111	1.00	-	-	110	1.00	-	-
≥1	7	***3.45***	***1.47-8.33***	***0.005***	7	***3.03***	***1.28-7.14***	***0.012***
KIR2DL1 C^{245} presence								
Donor C^{245} neg	82	1.00	-	-	81	1.00	-	-
Donor C^{245} pos	34	***2.13***	***1.28-3.45***	***0.003***	34	***1.92***	***1.16-3.13***	***0.010***
Donor KIR2DL1 neg phenotype	2	***14.29***	***3.33-100.00***	***<0.001***	2	***11.11***	***2.56-50.00***	***0.001***

Statistically significant results are denoted by ***bold italics***.

The difference in risk of NRM at one year was only a statistical trend in the MAC sub-cohort (C^{245} pos=32.4% *vs* C^{245} neg=18.7%, p=0.092, Table 6.9), despite being significant in the previous analysis performed on the overall cohort. When this factor

was assessed in a multivariate analysis scenario, this trend was lost (C^{245} positive donor: HR=1.78, CI=0.79-4.02, p=0.17).

The comparison of transplants involving R^{245} positive and negative donors revealed a statistical trend for overall survival (R^{245} pos=41.7% *vs* R^{245} neg=10.0%, p=0.058, Table 6.9). Multivariate analysis to assess this factor further revealed a statistically significant increase in OS in recipients of transplants from donors encoding at least one R^{245} positive KIR2DL1 allele (HR=0.47, CI=0.23-0.94, p=0.034, Table 6.11), although the small subgrouping populations may raise some doubt over the analysis validity.

Table 6.11 Multivariate analysis assessing the impact of donor KIR2DL1 R^{245} alleles on five year OS in the adult, HLA-matched MAC cohort

Variable	5 year OS			
	N	HR	95% CI	*p-value*
Recipient age, years				
<40	84	1.00	-	-
>40	34	***1.85***	***1.11-3.03***	***0.018***
Previous autografts				
0	111	1.00	-	-
≥1	7	***3.23***	***1.37-7.69***	***0.007***
KIR2DL1 R^{245} presence				
Donor R^{245} neg	10	1.00	-	-
Donor R^{245} pos	106	***0.47***	***0.23-0.94***	***0.034***
Donor KIR2DL1 neg phenotype	2	***5.88***	***1.19-33.33***	***0.030***

Statistically significant results are denoted by ***bold italics***.

As the effect of polymorphism at residue 245 of KIR2DL1 alleles appeared to be significant, comparison of transplant outcomes assessing both donor monotypic KIR2DL1 allele genotypes against the heterotypic genotype was performed (Table 6.9). This also suggested that OS and DFS were significantly influenced by this polymorphism and appeared to demonstrate a simultaneous advantageous effect of KIR2DL1 R^{245} alleles alongside detrimental effect of KIR2DL1 C^{245} alleles (Figure 6.6).

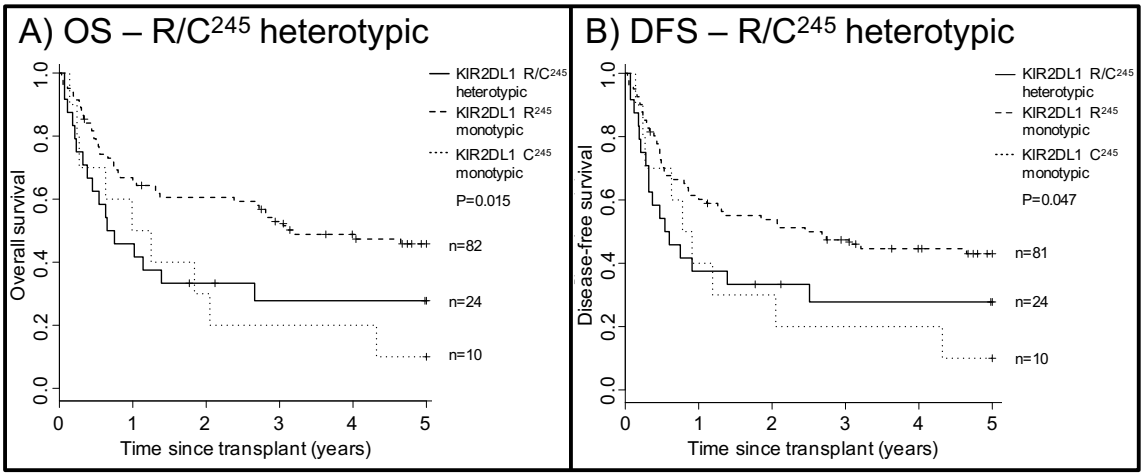


Figure 6.6 Probability of OS and DFS comparing the presence of donor-encoded monotypic or heterotypic KIR2DL1 R245C polymorphism in the adult, HLA-matched MAC cohort

A) Univariate Kaplan-Meier analysis of the probability of OS at five years post-transplant reveals a significant additive beneficial effect of donor-encoded KIR2DL1 R^{245} alleles such that recipients of transplants involving donors with increasing R^{245} KIR2DL1 allele copy number display stepwise improvements in five year OS probability (p=0.015). B) The same transplant stratification is also associated with reduced DFS probability (p=0.047).

Multivariate analysis was also applied to confirm these findings in a model adjusting for patient age group and history of previous autografts (Table 6.12). This revealed that, although the improved OS probability associated with transplants involving KIR2DL1 R^{245} monotypic donors was significant when compared against either KIR2DL1 R/C^{245} heterotypic donor transplants (HR=0.51, CI=0.29-0.91, p=0.023, Table 6.12) or KIR2DL1 C^{245} monotypic donor transplants (HR=0.39, CI=0.19-0.81, p=0.012), the difference between transplants involving R/C^{245} heterotypic and C^{245} monotypic donors was not statistically significant (HR=1.30, CI=0.58-2.94, p=0.52). Similar observations were also recorded for DFS at five years post-transplant (Table 6.12).

Table 6.12 Multivariate analysis assessing the impact of monotypic/heterotypic donor KIR2DL1 R245C allelic polymorphism on five year OS and DFS in the adult, HLA-matched MAC cohort

Variable	5 year OS				5 year DFS			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
Recipient age, years								
<40	84	1.00	-	-	83	1.00	-	-
>40	34	***1.75***	***1.05-2.94***	***0.030***	34	***1.69***	***1.02-2.78***	***0.040***
Previous autografts								
0	111	1.00	-	-	110	1.00	-	-
≥1	7	***3.57***	***1.49-8.33***	***0.004***	7	***3.03***	***1.30-7.14***	***0.011***
KIR2DL1 R245C polymorphism								
Donor R/C^{245} heterotypic	24	1.00	-	-	24	1.00	-	-
Donor R^{245} monotypic	82	***0.51***	***0.29-0.91***	***0.023***	81	***0.56***	***0.32-0.99***	***0.046***
Donor C^{245} monotypic	10	1.30	0.58-2.94	0.52	10	1.25	0.55-2.78	0.60
Donor KIR2DL1 neg phenotype	2	***7.69***	***1.67-33.33***	***0.009***	2	***6.25***	***1.37-33.33***	***0.018***

Statistically significant results are denoted by ***bold italics***.

Similarly to analysis in the overall and RIC cohorts, CNV of the R245C polymorphism was also investigated. As with the RIC cohort, the MAC cohort was extended for this analysis to include any MAC HCT (including paediatric and HLA-mismatched transplants).

Findings from the univariate analysis revealed a strong association between increasing KIR2DL1 R^{245} allele copy number and decreased risk of NRM at one year post-transplant (0 *vs* 1 *vs* 2 copies: 46.7%, *vs* 31.7% *vs* 16.4% p=0.017, Table 6.13, Figure 6.7). When tested in a multivariate analysis model, comparison of absence of KIR2DL1 R^{245} alleles against the presence of two copies of KIR2DL1 R^{245} alleles revealed a statistically significant reduction in the risk of NRM at one year post-transplant (HR=0.26, CI=0.11-0.58, p=0.001, Table 6.14). However, the difference in NRM risk when comparing against transplants involving donors with only a single copy of KIR2DL1 R^{245} was not statistically significant.

Table 6.13 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL1 R^{245} CNV polymorphism in the MAC cohort

Donor KIR2DL1 R^{245} CNV polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
0 *vs* 1 *vs* 2 R^{245} copies	0.15	0.091	0.80	***0.017***	0.28

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table K.

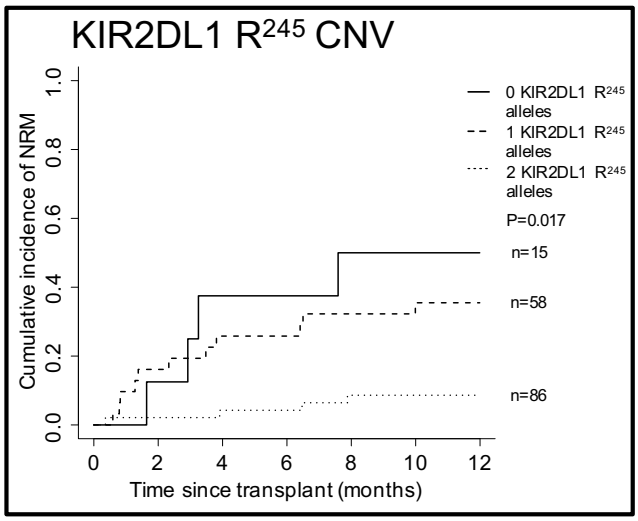


Figure 6.7 Probability of NRM comparing CNV of donor-encoded KIR2DL1 R^{245} alleles in the MAC cohort

Univariate analysis of the cumulative incidence of NRM reveals a statistically significant reduced risk associated with each additional donor-encoded KIR2DL1 R^{245} allele (p=0.017).

Table 6.14 Multivariate analysis assessing the impact of donor KIR2DL1 R^{245} CNV polymorphism on one year NRM in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
Recipient age, years				
<40	128	1.00	-	-
>40	36	***2.46***	***1.25-4.83***	***0.009***
Previous autografts				
0	158	1.00	-	-
≥1	6	***3.98***	***1.28-12.35***	***0.017***
HLA matching				
10/10 HLA matched	108	1.00	-	-
9/10 HLA matched	45	0.87	0.38-1.98	0.74
<9/10 HLA matched	11	***4.58***	***2.34-8.96***	***<0.001***
KIR2DL1 R^{245} CNV polymorphism				
0 donor KIR2DL1 R^{245} copies	15	1.00	-	-
1 donor KIR2DL1 R^{245} copies	58	0.73	0.34-1.54	0.41
2 donor KIR2DL1 R^{245} copies	86	***0.26***	***0.11-0.58***	***0.001***
Donor KIR2DL1 neg phenotype	5	0.78	0.22-2.78	0.70

Statistically significant results are denoted by ***bold italics***.

6.03.03 KIR2DL1 avidity polymorphisms

6.03.03.01 The influence of donor KIR2DL1 avidity polymorphism within the overall cohort

In addition to receptor signalling strength, KIR2DL1 allelic polymorphism has been shown to affect the avidity to which KIR2DL1 binds to its HLA-C2 ligand. As the genetic linkage between the 154, 163 and 182 polymorphisms is preserved amongst all alleles observed in this cohort, it was possible to differentiate allelic avidity to one of three factors: alleles encoding P^{114}~P^{154}~D^{163}~H^{182} were classified as having predicted high avidity, alleles encoding L^{114}~P^{154}~D^{163}~H^{182} were classified as having predicted intermediate avidity and alleles encoding P^{114}~T^{154}~N^{163}~R^{182} were classified as having predicted low avidity [348].

An analysis was performed to assess the influence of alleles with different ligand binding avidities as independent and combined variables. As with the R245C

polymorphism, transplants involving donors with KIR2DL1 null genotypes were excluded from univariate analysis. Initial investigation focussed simply on the presence or absence of each different avidity group (Table 6.15). This revealed that, although presence or absence of donor-encoded high avidity KIR2DL1 alleles did not correlate with any HCT outcome, presence of either intermediate or low avidity KIR2DL1 alleles did appear to influence the risk of NRM at one year post-transplant (intermediate avidity donor KIR2DL1 present *vs* absent: 18% *vs* 26.4%, p=0.049 Figure 6.8A; low avidity donor KIR2DL1 present *vs* absent: 27.4% *vs* 18.7%, p=0.043, Figure 6.8B).

Table 6.15 Univariate analysis p-values of HCT outcomes comparing the presence of different avidity donor KIR2DL1 alleles in the overall cohort

Predicted donor KIR2DL1 avidity	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
High avidity KIR2DL1 present	0.67	0.91	0.66	0.45	0.87
Intermediate avidity KIR2DL1 present	0.20	0.17	0.99	***0.049***	0.70
Low avidity KIR2DL1 present	0.38	0.78	0.088	***0.043***	0.48

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table K.

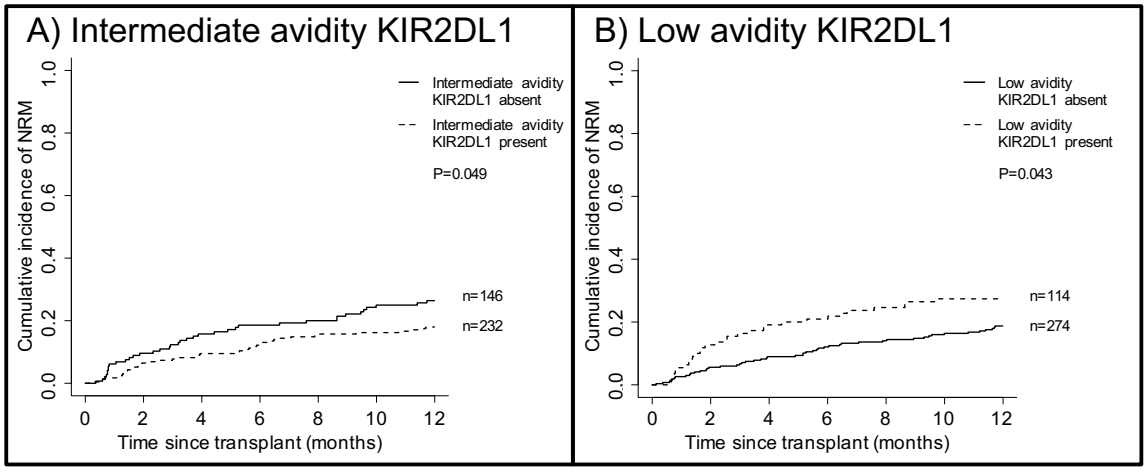


Figure 6.8 Probability of NRM comparing the presence of donor-encoded KIR2DL1 alleles with different ligand binding avidities in the overall cohort

A) Univariate analysis of NRM reveals a significantly decreased risk associated with transplants that involve donors who encode at least one intermediate avidity allele of KIR2DL1 compared to transplants where the donor encodes only low or high avidity KIR2DL1 alleles (p=0.049). B) When comparing the presence or absence of low avidity KIR2DL1 alleles, presence of low avidity alleles is associated with increased NRM risk (p=0.043).

When these findings were assessed in multivariate analyses that corrected for HLA matching, the presence of donor-encoded intermediate avidity KIR2DL1 alleles was no longer significantly associated with reduced risk of NRM, although it did persist as a statistical trend (HR=0.67, CI=0.43-1.04, p=0.075, Table 6.16). However, the detrimental effect observed in transplants involving donors encoding at least one low avidity KIR2DL1 allele remained statistically significant (HR=1.69, CI=1.08-2.67, p=0.023, Table 6.17).

Table 6.16 Multivariate analysis assessing the impact of donor KIR2DL1 intermediate ligand avidity on one year NRM risk in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA-matched	276	1.00	-	-
9/10 HLA-matched	90	1.28	0.76-2.15	0.35
<9/10 HLA-matched	23	***3.46***	***1.85-6.45***	***<0.001***
KIR2DL1 intermediate avidity alleles				
Donor intermediate avidity KIR2DL1 absent	146	1.00	-	-
Donor intermediate avidity KIR2DL1 present	232	0.67	0.43-1.04	0.075
Donor KIR2DL1 neg phenotype	11	1.11	0.27-2.97	0.86

Statistically significant results are denoted by ***bold italics***.

Table 6.17 Multivariate analysis assessing the impact of donor KIR2DL1 low ligand avidity on one year NRM risk in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA-matched	276	1.00	-	-
9/10 HLA-matched	90	1.30	0.77-2.19	0.32
<9/10 HLA-matched	23	***3.81***	***2.06-7.05***	***<0.001***
KIR2DL1 low avidity alleles				
Donor low avidity KIR2DL1 absent	268	1.00	-	-
Donor low avidity KIR2DL1 present	110	***1.69***	***1.08-2.67***	***0.023***
Donor KIR2DL1 neg phenotype	11	1.30	0.40-4.29	0.67

Statistically significant results are denoted by ***bold italics***.

Interestingly, a statistical trend observed in the univariate analysis of the effect of donor-encoded low avidity KIR2DL1 alleles (Table 6.15) also suggested that this factor may be associated with a protection against relapse at five years post-transplant (32.4% *vs* 41.3%, p=0.088). To further investigate this effect, multivariate analysis was performed. However a loss of significance was observed when the probability of relapse was adjusted for donor-recipient gender matching and disease risk score (Table 6.18).

Table 6.18 Multivariate analysis assessing the impact of donor KIR2DL1 low ligand avidity on five year relapse risk in the overall cohort

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Disease risk score				
Good	173	1.00	-	-
Intermediate	166	0.91	0.64-1.30	0.62
Poor	45	***2.08***	***1.25-3.44***	***0.005***
Donor-recipient gender matching				
Matched	227	1.00	-	-
Mismatched	162	0.78	0.55-1.10	0.16
KIR2DL1 low avidity alleles				
Donor low avidity KIR2DL1 absent	268	1.00	-	-
Donor low avidity KIR2DL1 present	110	0.77	0.52-1.14	0.19
Donor KIR2DL1 neg phenotype	11	1.08	0.39-3.01	0.88

Statistically significant results are denoted by ***bold italics***.

As the presence of low avidity KIR2DL1 alleles within donors conferred a detrimental effect on the outcomes of HCT, transplants were stratified by the presence of only one type of KIR2DL1 avidity allele. When compared against the remainder of transplants, these groups did not appear to have any significant effect (Table 6.19), although a statistical trend was observed correlating donor-encoded intermediate avidity KIR2DL1 alleles with reduced NRM at one year post-transplant (14.6% *vs* 23.6%, p=0.056). However, this also lost significance upon multivariate analysis (HR=0.63, CI=0.36-1.12, p=0.12).

Table 6.19 Univariate analysis p-values of HCT outcomes comparing the presence of different avidity donor KIR2DL1 alleles in the overall cohort

Predicted donor KIR2DL1 avidity	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Only high avidity KIR2DL1 present	0.56	0.26	0.12	0.75	0.45
Only intermediate avidity KIR2DL1 present	0.88	0.88	0.64	0.056	0.77
Only low avidity KIR2DL1 present	0.32	0.75	0.29	0.15	0.63

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table K.

To investigate whether the sum of avidities from each KIR2DL1 allele in an individual donor had an influence on the outcomes of HCT, a basic scoring system was assigned to the different allele avidities. Low, intermediate and high avidity alleles were scored 1, 2 and 3, respectively, and the total score for each donor sample calculated. As it was only possible to assign a score to donors with estimated haplotype information available, scores ranged from 0 to 6 (it was not possible to estimate haplotypes for donors encoding three or more alleles at any single locus). Each donor was then assigned to a high (total score≥4) or low (total score≤3) total avidity group. As in previous KIR2DL1 analysis, donors with a KIR2DL1 negative phenotype were excluded from univariate analysis. Despite previous analysis, in which the presence or absence of different avidity KIR2DL1 alleles indicated that NRM risk may differ, overall avidity (accounting for allelic copy number) was not statistically associated with any of the HCT outcomes assessed in the overall cohort (Table 6.20).

Table 6.20 Univariate analysis p-values of HCT outcomes comparing the effect of donor KIR2DL1 ligand avidity score in the overall cohort

Donor KIR2DL1 avidity group	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
High *vs* low total avidity	0.90	0.98	0.26	0.27	0.80

The number of transplants within each subgroup of each test is given in Supplementary Table K.

6.03.03.02 The influence of donor KIR2DL1 avidity polymorphism within the RIC cohort

Although the findings from the overall cohort suggested that the effects of ligand avidity were limited, tests were repeated on the sub-cohorts of adult, HLA-matched HCT stratified by conditioning regimen as previous analyses had demonstrated dramatic effects masked by this variable in the overall cohort. The results from the initial univariate analysis of the RIC cohort are summarised in Table 6.21.

Table 6.21 Univariate analysis p-values of HCT outcomes comparing the presence of different avidity donor KIR2DL1 alleles in the adult, HLA-matched RIC cohort

Predicted donor KIR2DL1 avidity	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
High avidity KIR2DL1 present	0.86	0.21	0.22	0.39	0.95
Only high avidity KIR2DL1 present	0.48	***0.045***	***0.024***	0.64	0.78
Intermediate avidity KIR2DL1 present	0.45	0.19	***0.045***	0.65	0.68
Only intermediate avidity KIR2DL1 present	0.77	0.95	0.97	0.22	0.43
Low avidity KIR2DL1 present	0.51	0.19	0.35	0.53	0.99
Only low avidity KIR2DL1 present	0.51	0.39	0.57	0.96	0.093

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table K.

Interestingly, although the presence of high avidity KIR2DL1 alleles within the donor genotype did not correlate with any outcomes in the overall cohort, when only this type of allele is encoded by donors involved in adult, HLA-matched RIC HCT, a statistically significant increase in relapse was observed (57.1% *vs* 36.2%, p=0.024). This increase in relapse associated with a significant decrease in DFS at five years post-transplant (23.8% *vs* 34.9%, p=0.045). As no clinical factors correlate with significant differences in relapse or DFS in this refined cohort, it was not possible to perform multivariate analyses.

The only other statistically significant finding from the univariate analysis was that, in transplants in which the donor encodes at least one intermediate avidity KIR2DL1 allele, relapse was significantly reduced (34.8% *vs* 51.4%, p=0.045). However, as all significance was lost when the donors that encoded only intermediate avidity alleles were compared against transplants from donors encoding at least one low or high

avidity KIR2DL1 allele, it is unlikely that the observed effect was related to the presence of intermediate alleles (Table 6.21).

Similarly to the overall cohort, when the overall avidity score was assessed in RIC transplants from which it was possible to estimate the donor's KIR2DL1 allele copy number (encompassing paediatric and HLA-mismatched transplants), no significantly different HCT outcomes were observed between high and low overall avidity donors groups (Table 6.22).

Table 6.22 Univariate analysis p-values of HCT outcomes comparing the effect of donor KIR2DL1 ligand avidity score in the RIC cohort

Donor KIR2DL1 avidity group	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
High *vs* low total avidity	0.33	0.19	0.13	0.94	1.00

The number of transplants within each subgroup of each test is given in Supplementary Table K.

6.03.03.03 The influence of donor KIR2DL1 avidity polymorphism within the MAC cohort

When equivalent analyses were performed in the MAC cohort, it was possible to observe a significant decrease in OS and DFS at five years post-transplant associated with the presence of donor-encoded low avidity KIR2DL1 alleles (OS: 21.6% *vs* 45.8%, p=0.004; DFS: 21.8% *vs* 43.0%, p=0.015; Table 6.23). In addition, a trend was observed linking this decrease in OS and DFS to an increase in NRM at one year post-transplant (32.4% *vs* 18.7%, p=0.092, Table 6.23). These results resemble those observed when comparing transplants involving C^{245} positive *vs* negative KIR2DL1 donors (Table 6.9, Table 6.10 and Figure 6.5), highlighting the genetic linkage (observed in the common KIR2DL1*004 allele) between this polymorphism and the avidity-determining polymorphisms. No statistically significant differences were

observed in any of the measured HCT outcomes when considering the presence of other avidity KIR2DL1 alleles.

Table 6.23 Univariate analysis p-values of HCT outcomes comparing the presence of different avidity donor KIR2DL1 alleles in the adult, HLA-matched MAC cohort

Predicted donor KIR2DL1 avidity	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
High avidity KIR2DL1 present	0.11	0.11	0.25	0.86	0.73
Only high avidity KIR2DL1 present	0.36	0.27	0.62	0.90	0.80
Intermediate avidity KIR2DL1 present	0.58	0.86	0.27	0.12	0.95
Only intermediate avidity KIR2DL1 present	0.62	0.84	0.99	0.36	0.44
Low avidity KIR2DL1 present	***0.004***	***0.015***	0.44	0.092	0.75
Only low avidity KIR2DL1 present	0.058	0.10	0.74	0.18	0.45

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table K.

Furthermore, when the overall avidity score (incorporating allelic copy number) was assessed in the broad MAC cohort, no statistically significant differences were observed for any of the assessed HCT outcomes (Table 6.24).

Table 6.24 Univariate analysis p-values of HCT outcomes comparing the effect of donor KIR2DL1 ligand avidity score in the MAC cohort

Donor KIR2DL1 avidity group	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
High *vs* low total avidity	0.44	0.25	0.97	0.17	0.75

The number of transplants within each subgroup of each test is given in Supplementary Table K.

6.03.04 The V-17F polymorphism

6.03.04.01 The influence of donor KIR2DL1 V-17F polymorphism within the overall cohort

Although polymorphism within the mature protein is sufficient to distinguish many of the different KIR2DL1 allelic variants, it is only possible to distinguish the common KIR2DL1*001 and KIR2DL1*002 alleles from each other by a V-17F amino acid substitution in the signal peptide. The F^{-17} variation of KIR2DL1*002 alleles is also shared with KIR2DL1*008, 011, 021, 034 and 004_c.13G>T alleles. To investigate the influence of this polymorphism, all transplants in which donors expressed KIR alleles were stratified by this variable. In addition, where possible, the copy number of V^{-17} and F^{-17} alleles were also correlated against HCT outcomes. The results from univariate analysis were unable to establish any significant effect on HCT outcomes in the overall cohort (Table 6.25).

Table 6.25 Univariate analysis p-values of HCT outcomes comparing the presence of donor KIR2DL1 V-17F polymorphism in the overall cohort

Donor KIR2DL1 V-17F polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
V^{-17} pos *vs* V^{-17} neg	0.98	0.56	0.16	0.81	0.74
F^{-17} pos *vs* F^{-17} neg	0.64	0.78	0.11	0.73	0.43
V/F^{-17} heterotypic *vs* V^{-17} monotypic *vs* F^{-17} monotypic	0.86	0.85	0.22	0.94	0.73
0 *vs* 1 *vs* 2 V^{-17} copies	0.43	0.65	0.47	0.81	0.69
0 *vs* 1 *vs* 2 F^{-17} copies	0.77	0.94	0.25	0.78	0.40

The number of transplants within each subgroup of each test is given in Supplementary Table K.

6.03.04.02 The influence of donor KIR2DL1 V-17F polymorphism within the RIC cohort

Although no statistically significant differences were observed in the overall cohort when investigating the V-17F polymorphism, a protective effect of the V^{-17} variant was observed when comparing the probabilities of relapse at five years post-transplant in the RIC cohort (36.4% *vs* 61.2%, p=0.021, Table 6.26). Although a statistical trend was also observed comparing V/F^{-17} heterotypic against each of the monotypic polymorphism genotypes (V/F^{-17} heterotypic *vs* V^{-17F} monotypic *vs* F^{-17} monotypic: 38.5% *vs* 35.7% *vs* 61.2%, p=0.063), it is arguable that the best model to assess the impact of this polymorphism was by utilising estimated allelic copy number. This demonstrated statistical significance when investigating copy number of either polymorphism (V^{-17} 0 *vs* 1 *vs* 2 copies: 63.1% *vs* 39.1% *vs* 32.1%, p=0.043; F^{-17} 0 *vs* 1 *vs* 2 copies: 29.7% *vs* 48.4% *vs* 69.7%, p=0.022, Table 6.26, Figure 6.9). As no clinical factors remained following forward stepwise selection of covariables in this RIC cohort, multivariate analysis was not performed.

Table 6.26 Univariate analysis p-values of HCT outcomes comparing the presence of donor KIR2DL1 V-17F polymorphism in the RIC cohort

Donor KIR2DL1 V-17F polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
V^{-17} pos *vs* V^{-17} neg	0.82	0.38	***0.021***	0.35	0.76
F^{-17} pos *vs* F^{-17} neg	0.66	0.46	0.14	0.98	0.52
V/F^{-17} heterotypic *vs* V^{-17} monotypic *vs* F^{-17} monotypic	0.91	0.64	0.063	0.56	0.82
0 *vs* 1 *vs* 2 V^{-17} copies	0.80	0.77	***0.043***	0.33	0.17
0 *vs* 1 *vs* 2 F^{-17} copies	0.93	0.55	***0.022***	0.69	0.38

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table K.

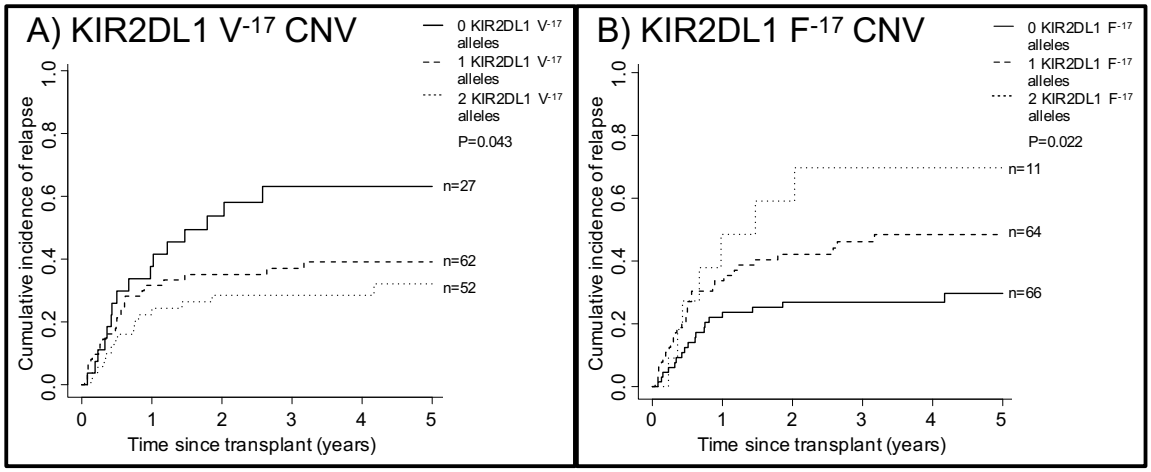


Figure 6.9 Probability of relapse comparing the copy number of donor-encoded KIR2DL1 V-17F alleles in the RIC cohort

A) Univariate analysis of relapse reveals a significantly decreased risk associated with each additional donor-encoded KIR2DL1 V^{-17} allele (p=0.043). B) When comparing copy number of donor-encoded KIR2DL1 F^{-17} alleles, an increase in relapse risk is observed with each additional allele (p=0.022).

6.03.04.03 The influence of donor KIR2DL1 V-17F polymorphism within the MAC cohort

When equivalent analysis is performed in the MAC cohort, the results more closely resemble those of the whole cohort: it is not possible to observe an effect of V-17F polymorphism on any of the measured HCT outcomes (Table 6.27).

Table 6.27 Univariate analysis p-values of HCT outcomes comparing the presence of donor KIR2DL1 V-17F polymorphism in the MAC cohort

Donor KIR2DL1 V-17F polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
V^{-17} pos *vs* V^{-17} neg	0.52	0.43	0.59	0.89	0.62
F^{-17} pos *vs* F^{-17} neg	0.25	0.28	0.82	0.72	0.53
V/F^{-17} heterotypic *vs* V^{-17} monotypic *vs* F^{-17} monotypic	0.52	0.54	0.87	0.87	0.81
0 *vs* 1 *vs* 2 V^{-17} copies	0.36	0.65	0.89	0.81	0.79
0 *vs* 1 *vs* 2 F^{-17} copies	0.65	0.72	1.00	0.95	0.86

The number of transplants within each subgroup of each test is given in Supplementary Table K.

6.04 Results: an analysis of the effects of donor KIR2DL2/3 allelic polymorphism on HCT outcomes

Many polymorphic residues distinguish different alleles of the KIR2DL2/3 locus, several of which have been directly demonstrated to influence the strength of ligand avidity and KIR2DL2/3-based NK cell cytotoxicity [105,174,339]. As well as representing a method to stratify the presence of KIR2DL2 and KIR2DL3 genes, the P16R and R148C polymorphisms also represent vital determinants of ligand avidity [105,174]. In addition, I have investigated a further contributing functional polymorphism, Q35E [339], as well as several other polymorphisms that distinguish alleles of this locus.

6.04.01 The P16R and R148C polymorphisms

As eluded to above, these polymorphisms, that are strongly associated with one another, can be used to distinguish the presence of KIR2DL2 (R^{16}~C^{148}) from KIR2DL3 (P^{16}~R^{148}) in this cohort. They have also been demonstrated to contribute to steric alteration of the D1-D2 interdomain hinge angle that results in increased HLA-C1 ligand binding [105].

6.04.01.01 The influence of donor KIR2DL2/3 P16R and R148C polymorphisms in the overall cohort

To investigate the impact of these polymorphisms in the HCT cohort, the presence of both KIR2DL2/3 P^{16}~R^{148} and R^{16}~C^{148} alleles was assessed. In addition, it was also possible to estimate the copy number of these allele types in all but two cases. This revealed that, in the presence of donor-encoded KIR2DL2/3 R^{16}~C^{148} alleles, recipients were significantly more likely to suffer NRM at one year post-transplant (17.8% *vs*

27.6%, p=0.028, Table 6.28, Figure 6.10A). Although a statistical trend was observed when comparing NRM risk according to donor KIR2DL2/3 R^{16}~C^{148} CNV, the difference between one and two copies was minimal (0 *vs* 1 *vs* 2 copies: 17.3% *vs* 27.3% *vs* 28.6%, p=0.071, Table 6.28, Figure 6.10B).

Table 6.28 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3 P^{16}~R^{148} polymorphisms in the overall cohort

Donor KIR2DL2/3 P16R~R148C polymorphisms	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
P^{16}~R^{148} pos *vs* P^{16}~R^{148} neg	0.83	0.88	0.39	0.41	0.86
R^{16}~C^{148} pos *vs* R^{16}~C^{148} neg	0.38	0.43	0.19	***0.028***	0.29
0 *vs* 1 *vs* 2 R^{16}~C^{148} copies	0.69	0.63	0.36	0.071	0.56

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table L.

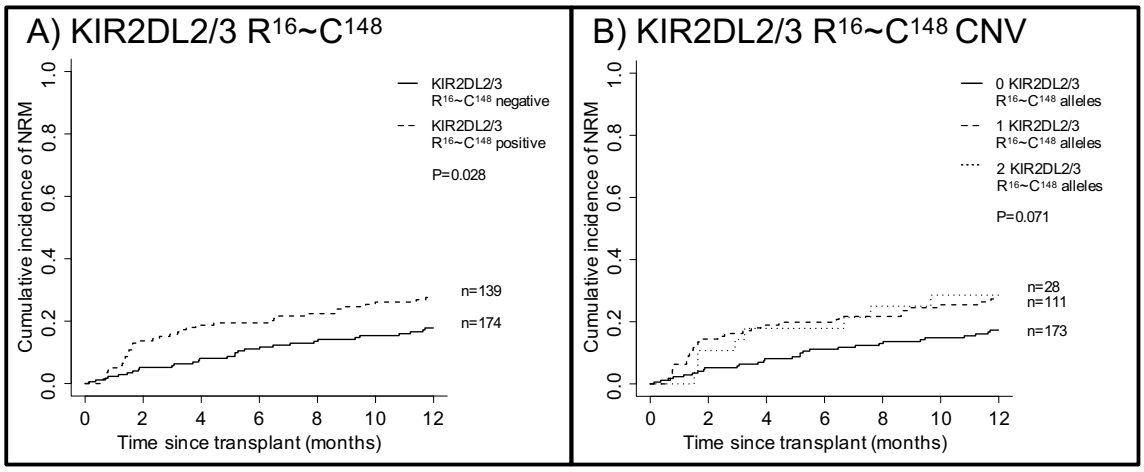


Figure 6.10 Probability of NRM comparing the presence and copy number of donor-encoded KIR2DL2/3 R^{16}~C^{148} alleles in the overall cohort

A) Univariate analysis of NRM at one year post-transplant reveals a significantly increased risk associated with donor-encoded KIR2DL2/3 R^{16}~C^{148} alleles (p=0.028). B) When comparing the copy number of these alleles, no dose effect is observed, although the difference between strata remains a trend (p=0.071).

In a multivariate analysis adjusting for HLA matching, the increase in NRM remained significantly different when considering the presence or absence of donor KIR2DL2/3 R^{16}~C^{148} (HR=1.75, CI=1.08-2.82, p=0.022, Table 6.29), but also gained significance when comparing transplants that involved donors who encode one copy of the allele group to transplants in which the donor did not encode any KIR2DL2/3 R^{16}~C^{148} alleles (HR: 1.76, CI=1.06-2.94, p=0.030, Table 6.30). Although only a statistical trend, multivariate analysis implied that the presence of two copies of donor-encoded KIR2DL2/3 R^{16}~C^{148} was correlated with further increased risk of NRM at one year post-transplant (HR=1.97, CI=0.93-4.19, p=0.077, Table 6.30), perhaps suggestive of a "dose" effect, as observed previously.

Table 6.29 Multivariate analysis assessing the influence of donor KIR2DL2/3 P16R and R148C polymorphism on the risk of NRM at one year post-transplant in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	216	1.00	-	-
9/10 HLA matched	78	1.28	0.73-2.22	0.39
<9/10 HLA matched	19	***3.81***	***1.93-7.53***	***<0.001***
Donor KIR2DL2/3 R^{16}~C^{148}				
KIR2DL2/3 R^{16}~C^{148} absent	174	1.00	-	-
KIR2DL2/3 R^{16}~C^{148} present	139	***1.75***	***1.08-2.82***	***0.022***

Statistically significant results are denoted by ***bold italics***.

Table 6.30 Multivariate analysis assessing the influence of donor KIR2DL2/3 R^{16}~C^{148} allele CNV on the risk of NRM at one year post-transplant in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	215	1.00	-	-
9/10 HLA matched	78	1.31	0.75-2.28	0.35
<9/10 HLA matched	19	***3.93***	***1.99-7.80***	***<0.001***
Donor KIR2DL2/3 R^{16}~C^{148} CNV				
0 copies	173	1.00	-	-
1 copy	111	***1.76***	***1.06-2.94***	***0.030***
2 copies	28	1.97	0.93-4.19	0.077

Statistically significant results are denoted by ***bold italics***.

6.04.01.02 The influence of donor KIR2DL2/3 P16R and R148C polymorphisms in the RIC cohort

To assess whether a stronger effect was observed within the different conditioning regimen cohorts, further analysis was performed. In the RIC cohort, only a statistical trend towards reduced DFS at five years post-transplant was noted in transplants involving donor-encoded KIR2DL2/3 P^{16}~R^{148} alleles (66.7% *vs* 28.5%, p=0.063, Table 6.31, Figure 6.11). Although a substantial difference in DFS probability, the lack of significance may result from the small population of donors who did not encode KIR2DL2/3 P^{16}~R^{148} alleles (n=12). As no clinical factors persisted during forward stepwise selection of covariates, multivariate analysis was not performed.

Table 6.31 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3 P^{16}~R^{148} polymorphisms in the RIC cohort

Donor KIR2DL2/3 P16R~R148C polymorphisms	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
P^{16}~R^{148} pos *vs* P^{16}~R^{148} neg	0.28	0.063	0.14	0.82	0.19
R^{16}~C^{148} pos *vs* R^{16}~C^{148} neg	0.48	0.24	0.25	0.75	0.90
0 *vs* 1 *vs* 2 R^{16}~C^{148} copies	0.53	0.16	0.25	0.98	0.42

The number of transplants within each subgroup of each test is given in Supplementary Table L.

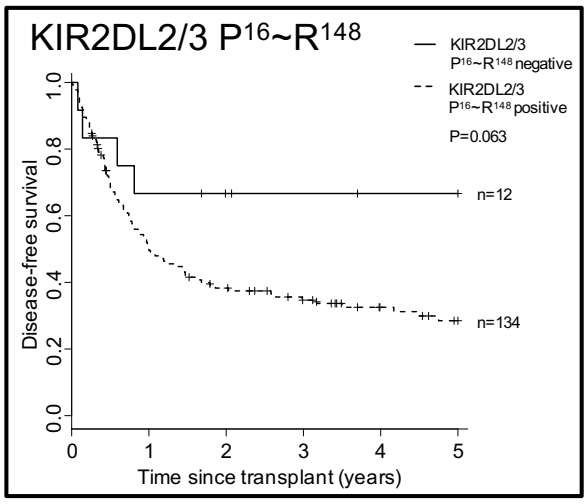


Figure 6.11 Probability of DFS comparing the presence of donor-encoded KIR2DL2/3 P^{16}~R^{148} alleles in the RIC cohort

Univariate analysis reveals a trend towards reduced probability of DFS at five years post-transplant associated with donor-encoded KIR2DL2/3 P^{16}~R^{148} alleles (p=0.063).

6.04.01.03 The influence of donor KIR2DL2/3 P16R and R148C polymorphisms in the MAC cohort

Although little of note was observed in the RIC cohort, when the equivalent analyses were performed in the MAC cohort, strong effects were seen correlating the presence of donor-encoded KIR2DL2/3 R^{16}~C^{148} alleles with increased NRM probability at one year post-transplant (15.9% *vs* 36.4%, p=0.003, Table 6.32, Figure 6.12A) although, as in the overall cohort, the difference between one and two copies of this allele group had a minor effect compared to the presence overall (0 *vs* 1 *vs* 2 copies: 15.9% *vs* 36.1% *vs* 37.5%, p=0.011, Table 6.32, Figure 6.12B).

Table 6.32 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3 P^{16}~R^{148} polymorphisms in the MAC cohort

Donor KIR2DL2/3 P16R~R148C polymorphisms	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
P^{16}~R^{148} pos *vs* P^{16}~R^{148} neg	0.17	0.11	0.78	0.28	0.56
R^{16}~C^{148} pos *vs* R^{16}~C^{148} neg	0.056	***0.028***	0.54	***0.003***	0.21
0 *vs* 1 *vs* 2 R^{16}~C^{148} copies	0.12	0.062	0.72	***0.011***	0.17

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table L.

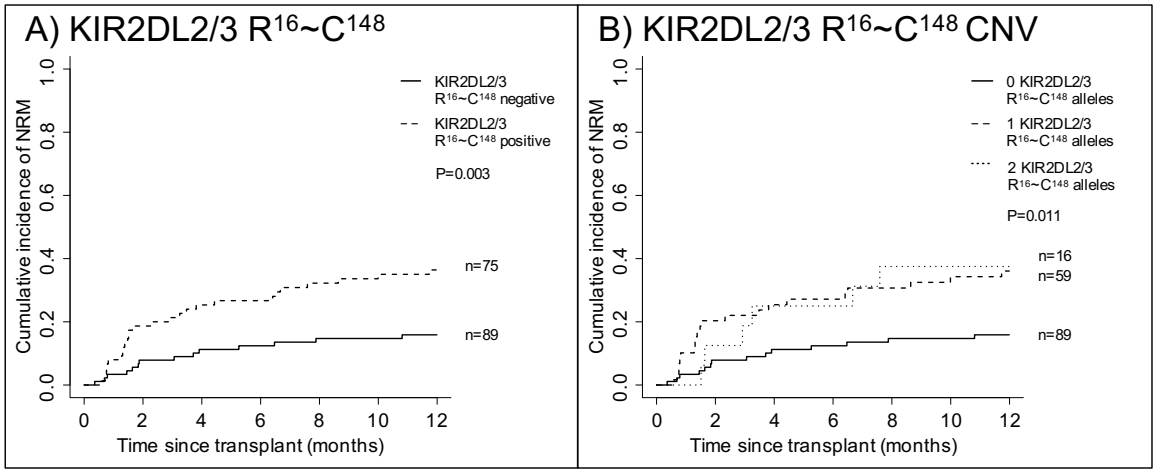


Figure 6.12 Probability of NRM comparing the presence and copy number of donor-encoded KIR2DL2/3 R^{16}~C^{148} alleles in the MAC cohort

A) Univariate analysis of NRM at one year post-transplant reveals a significantly increased risk associated with donor-encoded KIR2DL2/3 R^{16}~C^{148} alleles (p=0.003). B) When comparing the copy number these alleles, no dose effect is observed, although the difference between strata remains statistically significant (p=0.011).

In addition, a significant reduction in the probability of DFS (and a trend towards reduced OS) at five years post-transplant was also correlated with the presence of this allele group within the donor genotype (DFS: 41.1% *vs* 26.0%, p=0.028; OS: 42.4% *vs* 30.4%, p=0.056; Table 6.32, Figure 6.13).

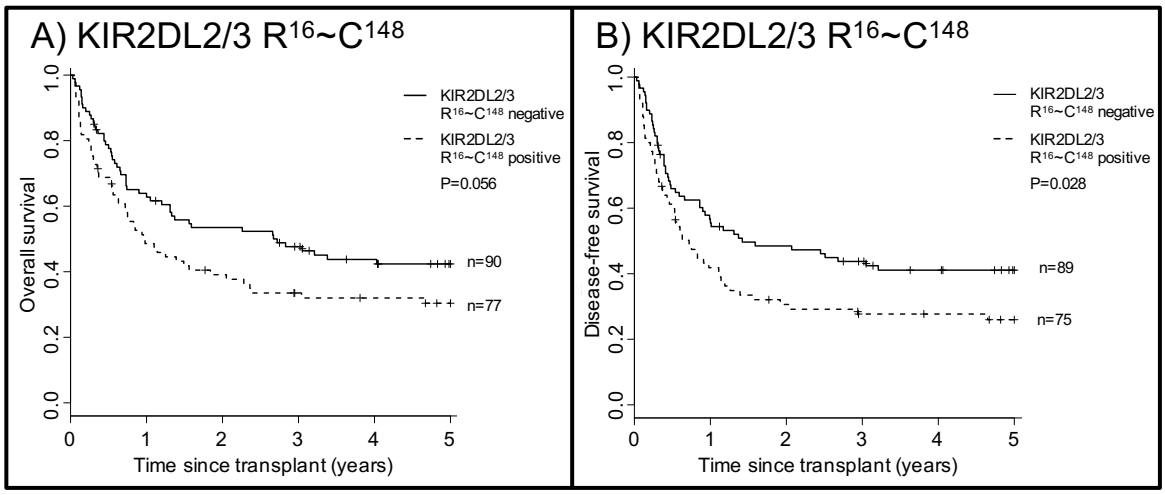


Figure 6.13 Probability of OS and DFS comparing the presence of donor-encoded KIR2DL2/3 R^{16}~C^{148} alleles in the MAC cohort

A) Univariate analysis of OS at five years post-transplant reveals a trend towards increased mortality risk associated with donor-encoded KIR2DL2/3 R^{16}~C^{148} alleles (p=0.056). B) When an equivalent comparison investigates DFS probability, the difference becomes statistically significant (p=0.028).

With the exception of DFS analysis, in which no clinical factors remained following forward stepwise covariate selection, each of the factors associated with significant differences or statistical trends from univariate analysis were also assessed in multivariate analyses. NRM risk at one year post-transplant was demonstrated to remain significantly increased following transplant involving donors encoding KIR2DL2/3 R^{16}~C^{148} alleles (HR=3.21, CI=1.65-6.27, p=0.001, Table 6.33). Furthermore, when the NRM analysis was further refined to include the CNV of this allele group, transplant subgroups involving donors with either one or two copies were both significantly different from the KIR2DL2/3 R^{16}~C^{148} subgroup (one copy: HR=3.07, CI=1.52-6.18, p=0.002; two copies: HR=3.89, CI=1.58-9.54, p=0.003, Table 6.34). However, no significant difference exists between these two KIR2DL2/3 R^{16}~C^{148} positive subgroups (HR: 1.26, CI=0.55-2.90, p=0.57).

Table 6.33 Multivariate analysis assessing the impact of donor KIR2DL2/3 R^{16}~C^{148} alleles on one year NRM risk in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	108	1.00	-	-
9/10 HLA matched	45	0.84	0.37-1.92	0.69
<9/10 HLA matched	11	***4.62***	***2.45-8.70***	***<0.001***
Recipient age, years				
<40	128	1.00	-	-
>40	36	***2.42***	***1.24-4.70***	***0.009***
Previous autograft history				
0	158	1.00	-	-
≥1	6	***3.71***	***1.21-11.38***	***0.022***
Donor KIR2DL2/3 R^{16}~C^{148}				
KIR2DL2/3 R^{16}~C^{148} absent	89	1.00	-	-
KIR2DL2/3 R^{16}~C^{148} present	75	***3.21***	***1.65-6.27***	***0.001***

Statistically significant results are denoted by ***bold italics***.

Table 6.34 Multivariate analysis assessing the impact of donor KIR2DL2/3 R^{16}~C^{148} allele CNV on one year NRM risk in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	108	1.00	-	-
9/10 HLA matched	45	0.87	038-1.98	0.73
<9/10 HLA matched	11	***4.82***	***2.44-9.53***	***<0.001***
Recipient age, years				
<40	128	1.00	-	-
>40	36	***2.42***	***1.25-4.66***	***0.008***
Previous autograft history				
0	158	1.00	-	-
≥1	6	***3.84***	***1.23-11.95***	***0.020***
Donor KIR2DL2/3 R^{16}~C^{148} CNV				
0 copies	89	1.00	-	-
1 copy	59	***3.07***	***1.52-6.18***	***0.002***
2 copies	11	***3.89***	***1.58-9.54***	***0.003***

Statistically significant results are denoted by ***bold italics***.

Although only a trend in univariate analysis, the difference in OS probability at five years post-transplant associated with the presence of at least one copy of donor-encoded KIR2DL2/3 R^{16}~C^{148}, became statistically significant during multivariate analysis (HR=1.49, CI=1.01-2.22, p=0.044, Table 6.35). As such, the impact of CNV on five year OS probability was also explored for this variable. In this analysis, however,

stratifying by CNV reveals that only the two copy donor genotype is correlated with a significant difference in five year OS probability (HR=1.39, CI=0.91-2.13, p=0.13; two copies: HR=2.04, CI=1.08-3.85, p=0.029; Table 6.36).

Table 6.35 Multivariate analysis assessing the impact of donor KIR2DL2/3 R^{16}~C^{148} alleles on five year OS probability in the MAC cohort

Variable	5 year OS			
	N	HR	95% CI	*p-value*
Previous autograft history				
0	161	1.00	-	-
≥1	6	***2.50***	***1.01-6.25***	***0.047***
HLA matching				
10/10 HLA matched	109	1.00	-	-
9/10 HLA matched	47	1.35	0.88-2.08	0.17
<9/10 HLA matched	11	***2.63***	***1.30-5.26***	***0.008***
Donor KIR2DL2/3 R^{16}~C^{148}				
KIR2DL2/3 R^{16}~C^{148} absent	90	1.00	-	-
KIR2DL2/3 R^{16}~C^{148} present	77	***1.49***	***1.01-2.22***	***0.044***

Statistically significant results are denoted by ***bold italics***.

Table 6.36 Multivariate analysis assessing the impact of donor KIR2DL2/3 R^{16}~C^{148} allele CNV on five year OS probability in the MAC cohort

Variable	5 year OS			
	N	HR	95% CI	*p-value*
Previous autograft history				
0	161	1.00	-	-
≥1	6	***2.63***	***1.05-6.25***	***0.039***
HLA matching				
10/10 HLA matched	109	1.00	-	-
9/10 HLA matched	47	1.41	0.91-2.22	0.12
<9/10 HLA matched	11	***2.78***	***1.35-5.56***	***0.005***
Donor KIR2DL2/3 R^{16}~C^{148} CNV				
0 copies	90	1.00	-	-
1 copy	61	1.39	0.91-2.13	0.13
2 copies	16	***2.04***	***1.08-3.85***	***0.029***

Statistically significant results are denoted by ***bold italics***.

6.04.02 The Q35E polymorphism

In addition to the P16R and R148C polymorphisms, a third polymorphism has been shown to result in significant functional heterogeneity amongst KIR2DL2/3 alleles: Q35E. This substitution functions to adjust the D1-D2 interdomain hinge angle in a

similar manner to the P16R and R148C substitutions common to KIR2DL2 but is also encoded by the KIR2DL3*005-like allele group. Importantly, as the KIR2DL3*005 allele associates with KIR2DL1*001 on the Cen-A haplotype motif (see Chapter 5, Section 5.07), this represents a polymorphism that introduces strong KIR receptors for both HLA-C1 (KIR2DL3*005) and HLA-C2 (KIR2DL1*001) within a single haplotype motif.

6.04.02.01 The influence of donor KIR2DL2/3 Q35E polymorphism in the overall cohort

The expression of KIR2DL2/3 E^{35} alleles, including the common KIR2DL2*001, 002 and KIR2DL3*005 alleles, has been demonstrated to increase cytotoxicity of NK cells relative to KIR2DL2/3 Q^{35} alleles [339]. The effects that this dimorphic polymorphism have on the outcomes of HCT were assessed. This revealed a strong correlation between a significantly increased risk of NRM at one year post-transplant and the presence of donor-encoded KIR2DL2/3 E^{35} alleles (15.3% *vs* 28.9%, p=0.004, Table 6.37, Figure 6.14A).

Table 6.37 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3 Q35E polymorphism in the overall cohort

Donor KIR2DL2/3 Q35E polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Q^{35} pos *vs* Q^{35} neg	0.63	0.97	0.24	0.15	0.74
E^{35} pos *vs* E^{35} neg	0.16	0.19	0.16	***0.004***	0.24
0 *vs* 1 *vs* 2 E^{35} copies	0.37	0.36	0.27	***0.010***	0.49

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table L.

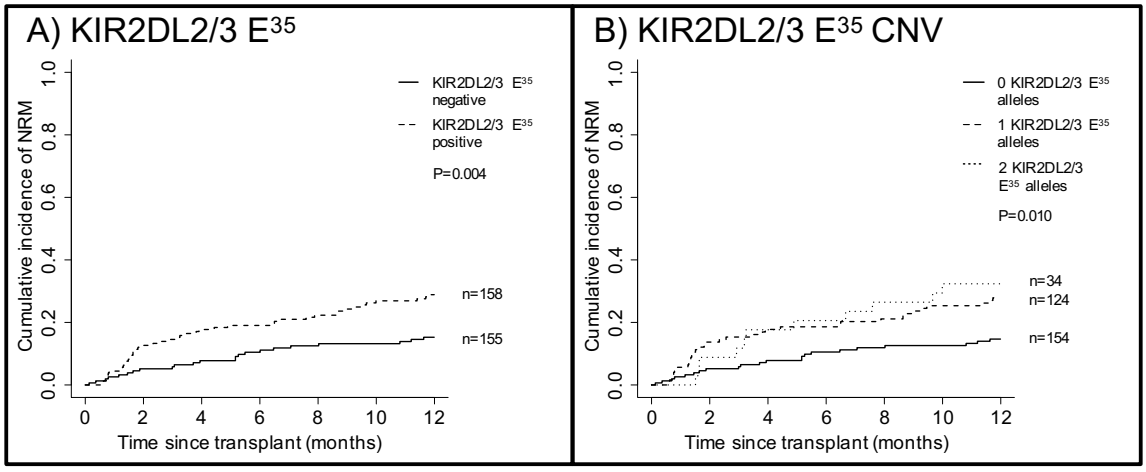


Figure 6.14 Probability of NRM comparing transplants involving donor-encoded KIR2DL2/3 Q35E polymorphism in the overall cohort

A) Univariate analysis of NRM reveals a significantly increased risk associated with donor-encoded KIR2DL2/3 E^{35} alleles (p=0.004). B) This remained significant when investigating the different copy number of these alleles and demonstrated an increase in NRM risk with each additional E^{35} allele (p=0.010).

KIR2DL2/3 allelic copy number was also estimated in all but one transplant with complete KIR2DL2/3 allele sequencing (the excluded transplant featured a donor with three copies of KIR2DL1 and, as such, was not included in haplotype estimation calculations). When the copy number of KIR2DL2/3 E^{35} alleles were investigated, this also demonstrated a significant risk associated with each additional KIR2DL2/3 E^{35} allele (0 *vs* 1 *vs* 2 copies: 14.7% *vs* 27.9% *vs* 32.4%, p=0.010, Figure 6.14B).

Multivariate analysis of both the presence and copy number of KIR2DL2/3 E^{35} alleles revealed a statistically significant increase in the risk of NRM associated with donor-encoded KIR2DL2/3 E^{35} alleles (KIR2DL2/3 E^{35} present: HR=2.16, CI=1.30-3.59, p=0.003, Table 6.38; 1 donor KIR2DL2/3 E^{35} copy: HR=2.13, CI=1.24-3.66, p=0.006; 2 donor KIR2DL2/3 E^{35} copies: HR=2.75, CI=1.36-5.55, p=0.005, Table 6.39). However, it was not possible to observe a significant difference between one and two

copies of the detrimental allele group, perhaps indicating an absence of dose effect (2 donor KIR2DL2/3 E^{35} copies: HR=1.29, CI=0.67-2.47, p=0.44).

Table 6.38 Multivariate analysis assessing the impact of donor KIR2DL2/3 E^{35} alleles on one year NRM risk in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	216	1.00	-	-
9/10 HLA matched	78	1.28	0.73-2.23	0.39
<9/10 HLA matched	19	***3.97***	***2.03-7.77***	***<0.001***
KIR2DL2/3 E^{35} alleles				
Donor KIR2DL2/3 E^{35} absent	155	1.00	-	-
Donor KIR2DL2/3 E^{35} present	158	***2.16***	***1.30-3.59***	***0.003***

Statistically significant results are denoted by ***bold italics***.

Table 6.39 Multivariate analysis assessing the impact of donor KIR2DL2/3 E^{35} allele CNV on one year NRM risk in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	216	1.00	-	-
9/10 HLA matched	78	1.32	0.76-2.31	0.33
<9/10 HLA matched	19	***4.20***	***2.14-8.23***	***<0.001***
KIR2DL2/3 E^{35} allele CNV				
0 donor KIR2DL2/3 E^{35} copies	154	1.00	-	-
1 donor KIR2DL2/3 E^{35} copy	124	***2.13***	***1.24-3.66***	***0.006***
2 donor KIR2DL2/3 E^{35} copies	34	***2.75***	***1.36-5.55***	***0.005***

Statistically significant results are denoted by ***bold italics***.

6.04.02.02 The influence of donor KIR2DL2/3 Q35E polymorphism in the RIC cohort

As eluded to in previous sections, the number of samples in which it was possible to estimate allelic copy number for the KIR2DL1 and KIR2DL2/3 loci was mainly limited by incomplete KIR2DL2/3 allele typing. As such, all analysis on KIR2DL2/3 allelic polymorphism stratified by conditioning regimen includes paediatric and HLA-mismatched transplants.

When the role of donor KIR2DL2/3 Q35E polymorphism was assessed in the RIC cohort, a statistically significant decrease in grades 2-4 aGVHD was observed in association with donor-encoded KIR2DL2/3 Q^{35} alleles (43.8% *vs* 21.2%, p=0.046, Table 6.40). In addition, a statistical trend was observed correlating presence of donor KIR2DL2/3 Q^{35} alleles with increased risk of relapse at five years post-transplant (13.3% *vs* 43.3%, p=0.053, Table 6.40, Figure 6.15). Forward stepwise selection failed to identify any clinical co-variables for multivariate analysis of aGVHD or relapse and, as such, these tests were not performed.

Table 6.40 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3 Q35E polymorphism in the RIC cohort

Donor KIR2DL2/3 Q35E polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Q^{35} pos *vs* Q^{35} neg	0.47	0.13	0.053	0.42	***0.046***
E^{35} pos *vs* E^{35} neg	0.93	0.64	0.33	0.74	0.50
0 *vs* 1 *vs* 2 E^{35} copies	0.78	0.32	0.14	0.66	0.15

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table L.

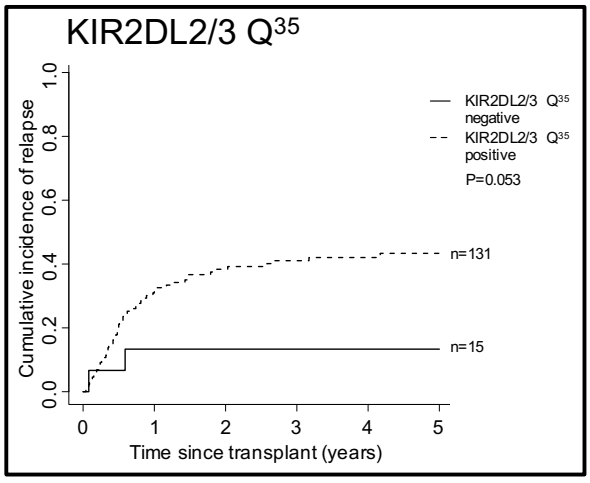


Figure 6.15 Probability of relapse comparing the presence of donor-encoded KIR2DL2/3 Q^{35} alleles in the RIC cohort

Univariate analysis of relapse reveals a statistical trend towards increased risk associated with donor-encoded KIR2DL2/3 Q^{35} alleles (p=0.053).

6.04.02.03 The influence of donor KIR2DL2/3 Q35E polymorphism in the MAC cohort

When the Q35E polymorphism was investigated in the MAC HCT setting, several significant findings were observed. First, the presence of any donor-encoded KIR2DL2/3 E^{35} alleles was associated with decreased OS and DFS at five years post-transplant (OS: 42.9% *vs* 30.8%, p=0.039, DFS: 41.5% *vs* 26.8%, p=0.024, Table 6.41, Figure 6.16). The same factor was also demonstrated to increase NRM risk at one year post-transplant (7.7% *vs* 37.2%, p=0.001, Table 6.41, Figure 6.17A). Although NRM was significantly different in the copy number analysis, both transplant categories incorporating KIR2DL2/3 E^{35} positive donors had similar NRM at the one year timepoint (0 *vs* 1 *vs* 2 copies, 13.5% *vs* 37.2% *vs* 36.8%, p=0.003, Figure 6.17B).

Table 6.41 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3 Q35E polymorphism in the MAC cohort

Donor KIR2DL2/3 Q35E polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Q^{35} pos *vs* Q^{35} neg	0.17	0.13	0.75	0.28	0.28
E^{35} pos *vs* E^{35} neg	***0.039***	***0.024***	0.36	***0.001***	0.32
0 *vs* 1 *vs* 2 E^{35} copies	0.094	0.060	0.52	***0.003***	0.12

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table L.

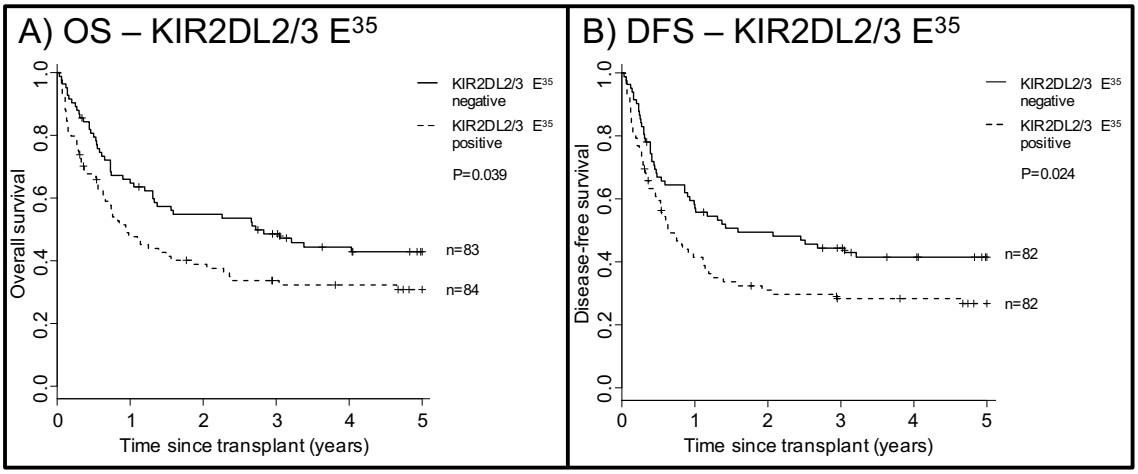


Figure 6.16 Probability of OS and DFS comparing the presence of donor-encoded KIR2DL2/3 E^{35} in the MAC cohort

A) Univariate analysis reveals a significantly reduced probability of OS associated with donor-encoded KIR2DL2/3 E^{35} alleles (p=0.039). B) Investigation into the effect of the same polymorphism on DFS probability also demonstrated a significant reduction associated with the presence of donor-encoded KIR2DL2/3 E^{35} alleles (p=0.024).

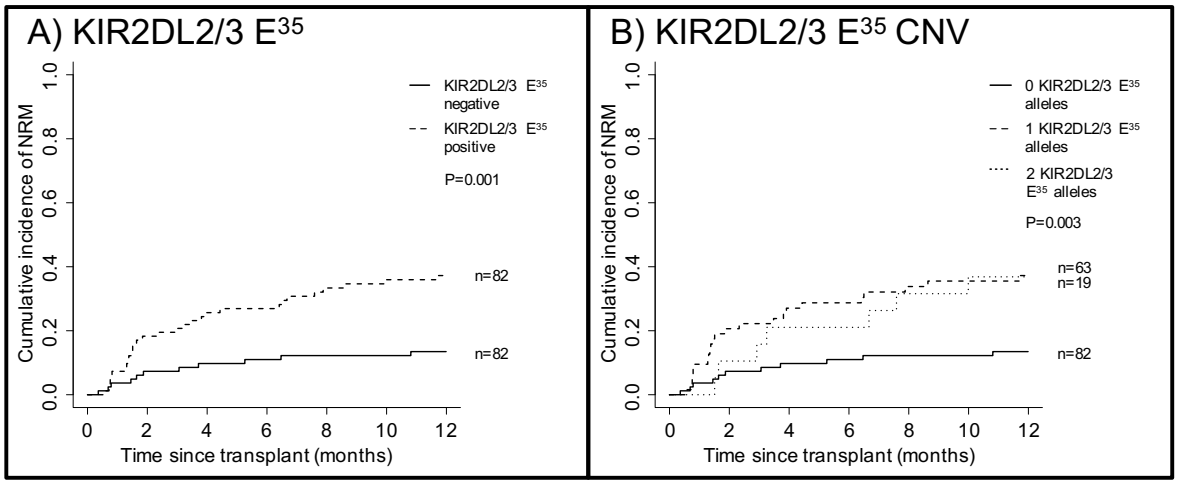


Figure 6.17 Probability of NRM comparing the donor-encoded KIR2DL2/3 E35Q polymorphism in the MAC cohort

A) Univariate analysis reveals a significantly increased probability of NRM associated with donor-encoded KIR2DL2/3 E^{35} alleles (p=0.001). B) Investigation into the effect of KIR2DL2/3 E^{35} allelic copy number on NRM risk also demonstrated a significant increase associated with the presence of donor-encoded KIR2DL2/3 E^{35} alleles, although a dose effect was not observed (p=0.003).

To further investigate the effect of donor-encoded KIR2DL2/3 E^{35} alleles on OS, DFS and NRM in the MAC cohort, multivariate analyses were performed. When investigating NRM, the findings were similar to those observed when investigating the overall cohort. The presence of donor KIR2DL2/3 E^{35} alleles was associated with significantly increased risk of NRM (HR=4.02, CI=1.91-8.47, CI<0.001, Table 6.42) and, although the risk of NRM was significantly increased when either 1 or 2 donor-encoded copies of this allele group were estimated (Table 6.43), the difference between them was not significant (HR=1.23, CI=0.56-2.67, p=0.61).

Table 6.42 Multivariate analysis assessing the impact of donor KIR2DL2/3 E^{35} alleles on one year NRM risk in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	108	1.00	-	-
9/10 HLA matched	45	0.84	0.37-1.92	0.67
<9/10 HLA matched	11	***4.66***	***2.58-8.43***	***<0.001***
Recipient age, years				
<40	128	1.00	-	-
>40	36	***2.61***	***1.35-5.04***	***0.004***
Previous autografts				
0	158	1.00	-	-
≥1	6	***4.04***	***1.28-12.80***	***0.018***
KIR2DL2/3 E^{35} alleles				
Donor KIR2DL2/3 E^{35} absent	82	1.00	-	-
Donor KIR2DL2/3 E^{35} present	82	***4.02***	***1.91-8.47***	***<0.001***

Statistically significant results are denoted by ***bold italics***.

Table 6.43 Multivariate analysis assessing the impact of donor KIR2DL2/3 E^{35} allele CNV on one year NRM risk in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	108	1.00	-	-
9/10 HLA matched	45	0.86	0.37-2.00	0.73
<9/10 HLA matched	11	***4.88***	***2.54-9.4***	***<0.001***
Recipient age, years				
<40	128	1.00	-	-
>40	36	***2.63***	***1.37-5.05***	***0.004***
Previous autografts				
0	158	1.00	-	-
≥1	6	***4.19***	***1.30-13.49***	***0.017***
KIR2DL2/3 E^{35} allele CNV				
0 donor KIR2DL2/3 E^{35} copies	82	1.00	-	-
1 donor KIR2DL2/3 E^{35} copy	63	***3.86***	***1.78-8.34***	***<0.001***
2 donor KIR2DL2/3 E^{35} copies	19	***4.73***	***1.85-12.11***	***0.001***

Statistically significant results are denoted by ***bold italics***.

However, in multivariate analysis assessing the probability of OS and DFS at five years post-transplant, the difference between donor-encoded KIR2DL2/3 E^{35} copy number was more apparent. In both OS and DFS, the presence of KIR2DL2/3 E^{35} was correlated with significantly reduced survival (OS: HR=1.54, CI=1.04-2.33, p=0.031; DFS: HR=1.59, CI=1.08-2.38, p=0.020; Table 6.44). When the copy number of these alleles were investigated, only transplants involving donors who encoded two KIR2DL2/3 E^{35} alleles continued to be significantly associated with a detrimental OS and DFS probability (OS: 1 copy HR=1.43, CI=0.93-2.17, p=0.10; 2 copies HR=2.08, CI=1.12-3.85, p=0.020; DFS: 1 copy HR=1.45, CI=0.95-2.22, p=0.080; 2 copies HR=2.13, CI=1.18-3.85, p=0.013; Table 6.45).

Table 6.44 Multivariate analysis assessing the impact of donor KIR2DL2/3 E^{35} alleles on five year OS and DFS in the MAC cohort

Variable	5 year OS				5 year DFS			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
HLA matching								
10/10 HLA matched	109	1.00	-	-	108	1.00	-	-
9/10 HLA matched	47	1.35	0.87-2.08	0.18	45	1.35	0.88-2.08	0.16
<9/10 HLA matched	11	***2.63***	***1.30-5.26***	***0.004***	11	***2.22***	***1.11-4.55***	***0.025***
Previous autografts								
0	161	1.00	-	-	158	1.00	-	-
≥1	6	***2.56***	***1.03-6.25***	***0.043***	6	***2.50***	***1.00-6.25***	***0.049***
KIR2DL2/3 E^{35} presence								
KIR2DL2/3 E^{35} negative	83	1.00	-	-	82	1.00	-	-
KIR2DL2/3 E^{35} positive	84	***1.54***	***1.04-2.33***	***0.031***	82	***1.59***	***1.08-2.38***	***0.020***

Statistically significant results are denoted by ***bold italics***.

Table 6.45 Multivariate analysis assessing the impact of donor KIR2DL2/3 E^{35} allele CNV on five year OS and DFS in the MAC cohort

Variable	5 year OS				5 year DFS			
	N	HR	95% CI	*p-value*	N	HR	95% CI	*p-value*
HLA matching								
10/10 HLA matched	109	1.00	-	-	108	1.00	-	-
9/10 HLA matched	47	1.41	0.91-2.22	0.13	45	1.45	0.93-2.08	0.11
<9/10 HLA matched	11	***2.78***	***1.35-5.56***	***0.005***	11	***2.38***	***1.16-4.76***	***0.017***
Previous autografts								
0	161	1.00	-	-	158	1.00	-	-
≥1	6	***2.63***	***1.08-6.67***	***0.035***	6	***2.56***	***1.04-6.67***	***0.040***
KIR2DL2/3 E^{35} CNV								
0 donor KIR2DL2/3 E^{35} copies	83	1.00	-	-	82	1.00	-	-
1 donor KIR2DL2/3 E^{35} copy	65	1.43	0.93-2.17	0.10	63	1.45	0.95-2.22	0.080
2 donor KIR2DL2/3 E^{35} copies	19	***2.08***	***1.12-3.85***	***0.020***	19	***2.13***	***1.18-3.85***	***0.013***

Statistically significant results are denoted by ***bold italics***.

6.04.03 KIR2DL2/3 allele-specific polymorphisms

Although not always having been previously associated with functional impact, several polymorphisms that distinguish different common alleles of the KIR2DL2/3 locus were investigated for their effect on HCT outcomes (for more information on the frequencies of common alleles, see Chapter 5). These include: H50R (R^{50} observed in KIR2DL3*005, allele frequency ~4.5%), T200I (I^{200} observed in KIR2DL2*001, allele

frequency ~14%), P208L (L^{208} observed in KIR2DL3*002, allele frequency ~24%) and R297H (H^{297} observed in KIR2DL3*002 and KIR2DL3*005, combined allele frequency ~29%). Although presence of the R^{50} polymorphism was assessed, the absence of the more common H^{50} allele was not investigated as it was too rarely observed to be missing from a donor's genotype (n=2). Accordingly, CNV of the H50R polymorphism was also not assessed due to low incidence of more than one KIR2DL2/3 R^{50} allelic copy.

6.04.03.01 The influence of donor KIR2DL2/3 allele-specific polymorphisms in the overall cohort

The presence and copy number of these polymorphisms were assessed in the overall UK cohort of HCT recipients to treat AML. Findings from this analysis revealed only one statistical trend towards increased NRM at one year post-transplant in the presence of donor-encoded KIR2DL2/3 R^{50} alleles (20.7% *vs* 41.0%, p=0.057, Table 6.46, Figure 6.18). When this was assessed in a multivariate analysis setting, statistical significance was achieved (HR=2.15, CI=1.15-4.03, p=0.017, Table 6.47).

Table 6.46 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3 allele-specific polymorphisms in the overall cohort

Donor KIR2DL2/3 allele-specific polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
R^{50} pos *vs* R^{50} neg	0.32	0.41	0.41	0.057	0.80
T^{200} pos *vs* T^{200} neg	0.58	0.34	0.42	0.12	0.12
I^{200} pos *vs* I^{200} neg	0.89	0.75	0.30	0.36	0.66
0 *vs* 1 *vs* 2 I^{200} copies	0.84	0.53	0.50	0.24	0.14
P^{208} pos *vs* P^{208} neg	0.63	0.68	0.63	0.83	0.70
L^{208} pos *vs* L^{208} neg	0.78	0.90	0.59	0.97	0.43
0 *vs* 1 *vs* 2 L^{208} copies	0.89	0.92	0.67	0.96	0.73
R^{297} pos *vs* R^{297} neg	0.70	0.53	0.70	0.94	0.63
H^{297} pos *vs* H^{297} neg	0.98	0.96	0.86	0.34	0.60
0 *vs* 1 *vs* 2 H^{297} copies	0.90	0.79	0.87	0.55	0.83

The number of transplants within each subgroup of each test is given in Supplementary Table L.

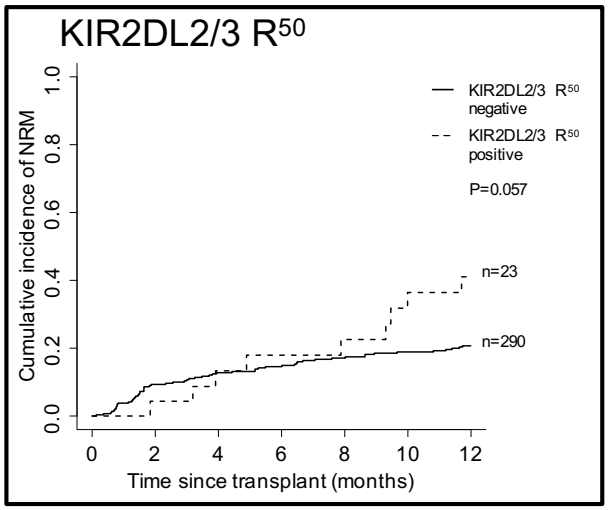


Figure 6.18 Probability of NRM investigating the presence of donor-encoded KIR2DL2/3 R^{50} alleles in the overall cohort

Univariate analysis reveals a statistical trend towards increased probability of NRM associated with the presence of donor-encoded KIR2DL2/3 R^{50} alleles (p=0.057).

Table 6.47 Multivariate analysis assessing the impact of donor KIR2DL2/3 R^{50} allele CNV on one year NRM risk in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	216	1.00	-	-
9/10 HLA matched	78	1.34	0.77-2.34	0.31
<9/10 HLA matched	19	***3.89***	***1.96-7.72***	***<0.001***
KIR2DL2/3 R^{50} allele CNV				
Donor KIR2DL2/3 R^{50} negative	290	1.00	-	-
Donor KIR2DL2/3 R^{50} positive	23	***2.15***	***1.15-4.03***	***0.017***

Statistically significant results are denoted by ***bold italics***.

6.04.03.02 The influence of donor KIR2DL2/3 allele-specific polymorphisms in the RIC cohort

Allele-specific polymorphisms were also assessed in the RIC sub-cohort. However, due to small population of the subcategories (Supplementary Table L), the influence that the presence or absence of the KIR2DL2/3 T^{200} polymorphism had on HCT outcomes was not assessed. Accordingly, the effect of polymorphism CNV was also not investigated

at this residue. The remaining analysis revealed no significant differences between any of the polymorphisms, although a trend was observed associating an increase in DFS at five years after transplants involving donors who encoded KIR2DL2/3 I^{200} alleles compared to transplants in which the donor did not encode a KIR2DL2/3 I^{200} allele (n=101 *vs* 45, 26.5% *vs* 42.5%, p=0.096, Table 6.48). Furthermore, a second statistical trend was observed suggestive of increased relapse risk in recipients of transplants involving donors encoding at least one KIR2DL2/3 L^{208} allele (34.5% *vs* 48.5%, p=0.069, Table 6.48). As no clinical factors were indicated as significant within the RIC cohort stepwise covariable selection, multivariate analysis was not performed.

Table 6.48 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3 allele-specific polymorphisms in the RIC cohort

Donor KIR2DL2/3 allele-specific polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
R^{50} pos *vs* R^{50} neg	0.22	0.21	0.92	0.12	0.19
I^{200} pos *vs* I^{200} neg	0.40	0.096	0.37	0.56	0.73
P^{208} pos *vs* P^{208} neg	0.74	0.83	0.66	0.60	0.69
L^{208} pos *vs* L^{208} neg	0.84	0.44	0.069	0.81	0.19
0 *vs* 1 *vs* 2 L^{208} copies	0.92	0.71	0.21	0.88	0.45
R^{297} pos *vs* R^{297} neg	0.47	0.11	0.27	0.75	0.73
H^{297} pos *vs* H^{297} neg	0.65	0.38	0.25	0.54	0.80
0 *vs* 1 *vs* 2 H^{297} copies	0.72	0.25	0.41	0.72	0.84

The number of transplants within each subgroup of each test is given in Supplementary Table L.

6.04.03.03 The influence of donor KIR2DL2/3 allele-specific polymorphisms in the MAC cohort

Allele-specific polymorphisms were also investigated in the MAC sub-cohort. Again, small subcategories prevented analysis into the influence of some donor KIR2DL2/3 T200I polymorphism on HCT outcomes. The univariate analysis performed demonstrated a single statistical trend towards increased NRM risk in transplants involving donors who encode at least one KIR2DL2/3 I^{200} allele (21.6% *vs* 37.4%,

p=0.069, Table 6.49). However, when analysed in a multivariate analysis, this factor became statistically significant (HR=2.20, CI=1.17-4.14, p=0.014, Table 6.50).

Table 6.49 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3 allele-specific polymorphisms in the MAC cohort

Donor KIR2DL2/3 allele-specific polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
R^{50} pos *vs* R^{50} neg	0.87	0.91	0.22	0.20	0.44
I^{200} pos *vs* I^{200} neg	0.21	0.19	0.63	0.069	0.71
P^{208} pos *vs* P^{208} neg	0.72	0.53	0.43	0.96	0.96
L^{208} pos *vs* L^{208} neg	0.63	0.49	0.33	0.70	0.90
0 *vs* 1 *vs* 2 L^{208} copies	0.88	0.74	0.57	0.90	0.99
R^{297} pos *vs* R^{297} neg	0.94	0.73	0.73	0.84	0.91
H^{297} pos *vs* H^{297} neg	0.72	0.50	0.19	0.39	0.66
0 *vs* 1 *vs* 2 H^{297} copies	0.93	0.79	0.41	0.58	0.91

The number of transplants within each subgroup of each test is given in Supplementary Table L.

Table 6.50 Multivariate analysis assessing the impact of the presence of donor KIR2DL2/3 I^{200} alleles on one year NRM risk in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	108	1.00	-	-
9/10 HLA matched	45	0.96	0.44-2.10	0.92
<9/10 HLA matched	11	***4.56***	***2.26-9.19***	***<0.001***
Recipient age, years				
<40	128	1.00	-	-
>40	36	***2.11***	***1.07-4.15***	***0.031***
Previous autografts				
0	158	1.00	-	-
≥1	6	***3.98***	***1.35-11.74***	***0.012***
Donor KIR2DL2/3 I^{200} presence				
KIR2DL2/3 I^{200} absent	126	1.00	-	-
KIR2DL2/3 I^{200} present	38	***2.20***	***1.17-4.14***	***0.014***

Statistically significant results are denoted by ***bold italics***.

6.05 Results: an analysis of the effects of donor KIR3DL1/S1 allelic polymorphism on HCT outcomes

Perhaps the best characterisation of how allelic KIR polymorphism can have functional implications has been achieved at the KIR3DL1/S1 locus, where extensive analysis has

revealed significant differences in both receptor expression and ligand avidity and their effect on NK cells [152,298,346,351,352]. As such, distinct hierarchies of KIR3DL1 expression have been established for many of the common alleles and has led to development of algorithms to predict KIR3DL1 expression based on the pattern of polymorphism observed [154]. Furthermore, polymorphism within the HLA-Bw4 ligand of KIR3DL1 has been demonstrated to alter the response elicited by KIR3DL1-expressing NK cells. As such, in this analysis I have investigated not only the different expression level grouping of KIR3DL1/S1 alleles, but also developed a scoring system based on measured KIR3DL1 ligand avidity between particular KIR and HLA ligand alleles [298].

6.05.01 KIR3DL1 expression polymorphism

6.05.01.01 The influence of the predicted donor KIR3DL1 expression in the overall cohort

As the best characterised functional diversity of KIR3DL1 alleles, cell surface expression was the first factor investigated in the UK HCT cohort. Initial investigation simply assessed the donor KIR3DL1 expression phenotype whereby transplants characterised by the presence of expressed KIR3DL1 alleles within the donor genotypes were compared against those in which the donor encoded only KIR3DS1 or null KIR3DL1 alleles. This analysis revealed no significant differences between any of the outcomes (Table 6.51) and, as such, analysis of allelic polymorphism at the KIR3DL1/S1 locus continued using all transplants.

Table 6.51 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 expression in the overall cohort

Donor KIR3DL1 expression	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
KIR3DL1 expression phenotype	0.77	0.79	0.28	0.15	0.71
High KIR3DL1 expression present	0.72	0.75	0.44	0.23	0.73
Low KIR3DL1 expression present	0.57	0.39	0.56	0.19	0.69
Null KIR3DL1 expression present	0.70	0.54	0.51	0.40	0.17
KIR3DL1 total expression null *vs* low *vs* high	0.88	0.91	0.40	0.13	0.51

The number of transplants within each subgroup of each test is given in Supplementary Table M.

Further analysis stratified expressed KIR3DL1 alleles by low and high expression. In addition, a rudimentary overall expression level was developed whereby information relating to the predicted number of allele copies and their expression level were combined to score the overall expression of a given genotype (each KIR3DS1 or KIR3DL1 null expression allele contributed 0 points, each low KIR3DL1 expression allele contributed 1 point and each high KIR3DL1 expression allele contributed 2 points). The total score was then categorised into one of three groups: i) null (score=0), ii) low KIR3DL1 expression (score≤2) or iii) high KIR3DL1 expression (score≥3). None of these analyses revealed any statistically significant differences between the groups (Table 6.51).

6.05.01.02 The influence of the predicted donor KIR3DL1 expression in the RIC cohort

When the presence of different predicted KIR3DL1 expression allele groups were assessed in the RIC cohort (including paediatric and HLA-mismatched transplants), no significant differences were observed for any of the assessed HCT outcomes (Table 6.52).

Table 6.52 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 expression in the RIC cohort

Donor KIR3DL1 expression	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
High KIR3DL1 expression present	0.86	0.55	0.67	0.77	0.18
Low KIR3DL1 expression present	0.99	0.31	0.65	0.12	0.49
Null KIR3DL1 expression present	0.43	0.84	0.93	0.64	0.22
KIR3DL1 total expression null *vs* low *vs* high	0.92	0.71	0.91	0.45	0.12

The number of transplants within each subgroup of each test is given in Supplementary Table M.

6.05.01.03 The influence of the predicted donor KIR3DL1 expression in the MAC cohort

Although no differences were observed when assessing the presence of different predicted donor KIR3DL1 expression allele groups in the overall or RIC cohorts, analysis in the MAC cohort (including paediatric and HLA-mismatched) revealed a statistical trend towards increased NRM in transplants involving donors with at least one predicted high expression KIR3DL1 allele within their genotype (14.0% *vs* 26.9%, p=0.064, Table 6.53). However, in a multivariate analysis, this factor remained only a statistical trend (HR=2.34, CI=0.99-5.55, p=0.054, Table 6.54). Given the lack of significant effects with this variable, it should be interpreted with caution.

Table 6.53 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 expression in the MAC cohort

Donor KIR3DL1 expression	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
High KIR3DL1 expression present	0.55	0.46	0.40	0.064	0.098
Low KIR3DL1 expression present	0.44	0.77	0.75	0.69	0.24
Null KIR3DL1 expression present	0.31	0.44	0.21	0.10	0.39
KIR3DL1 total expression null *vs* low *vs* high	0.85	0.89	0.26	0.15	0.17

The number of transplants within each subgroup of each test is given in Supplementary Table M.

Table 6.54 Multivariate analysis assessing the influence of predicted donor KIR3DL1 high expression alleles on risk of NRM at one year post-transplant in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	125	1.00	-	-
9/10 HLA matched	34	0.73	0.27-1.97	0.53
<9/10 HLA matched	11	***3.55***	***1.69-7.46***	***0.001***
Recipient age, years				
<40	126	1.00	-	-
>40	44	***2.10***	***1.06-4.19***	***0.033***
Predicted high expression KIR3DL1 allele				
Donor predicted high expression KIR3DL1 absent	57	1.00	-	-
Donor predicted high expression KIR3DL1 present	113	2.34	0.99-5.55	0.054

Statistically significant results are denoted by ***bold italics***.

6.05.02 The 'MSK' model

Much of the investigation into the effects of KIR polymorphism on HCT outcomes has involved research performed at the Memorial Sloan Kettering (MSK) Cancer Centre. Amongst this is a model of analysis involving polymorphism at the KIR3DL1 locus as well as the HLA-Bw4 ligand [393]. This model proposes that the level of NK cell inhibition granted by high and low expression KIR3DL1 alleles is dependent on the HLA-Bw4-80T/I status of the ligand, such that high expression KIR3DL1 alleles offer greater inhibition in combination with HLA-Bw4-80I ligands, whilst low expression KIR3DL1 alleles offer greater inhibition in combination with HLA-Bw4-80T ligands. Strong inhibition mediated by donor KIR3DL1 was demonstrated to result in an increased risk of relapse [393].

6.05.02.01 The influence of the MSK model in the overall cohort

To assess whether the presence of proposed strong, weak or non-inhibitory KIR3DL1 and ligand combinations influence the outcomes of HCT in my UK cohort, these factors were investigated as per the original publication. However, to incorporate HLA-mismatched transplants, two additional applications of the model were performed that

differed according to whether the donor or recipient HLA ligand information was utilised. In addition, comparison was made between each of the different inhibition categories, and between the strong and weak/non-inhibitory combined group (as per original publication).

Table 6.55 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 inhibitory strength in the overall cohort

MSK model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Strong *vs* weak *vs* non-inhibitory KIR3DL1					
10/10 HLA-matched only	0.82	0.91	0.52	0.44	0.28
Any HLA-matching & donor HLA-Bw4 ligand	0.64	0.60	0.38	***0.049***	0.15
Any HLA-matching & recipient HLA-Bw4 ligand	0.86	0.87	0.27	0.056	0.22
Strong *vs* weak/non-inhibitory KIR3DL1					
10/10 HLA-matched only	0.78	0.99	0.52	0.41	0.16
Any HLA-matching & donor HLA-Bw4 ligand	0.89	0.72	0.42	0.74	0.075
Any HLA-matching & recipient HLA-Bw4 ligand	0.90	0.93	0.51	0.66	0.10

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table M.

Interestingly, all analyses on the 10/10 HLA-matched cohort (i.e. comparable to the original publication) were unable to determine any significant differences between the groups for any of the assessed HCT outcomes. However, when HLA-mismatched transplants were also included, it was possible to observe a difference in NRM at one year post-transplant with borderline significance associated with the MSK score (weak *vs* strong *vs* non-inhibitory: 26.5% *vs* 20.9% *vs* 13.7%, p=0.049, Table 6.55). As only six of the HLA-mismatched transplants involved a change in the HLA-Bw4 ligand status, similar results were obtained when comparing the models utilising donor or recipient ligand information (although the borderline significant difference observed

when considering donor ligand became only a trend when recipient ligand was considered). When this factor was assessed in a multivariate model adjusting for HLA-matching and donor-recipient CMV matching, no significant difference was observed between the weak (HR=1.18, CI=0.63-2.20, p=0.60) or non-inhibitory (HR=0.70, CI=0.36-1.36, p=0.29) groups when compared against the strong inhibitory group (Table 6.56).

Table 6.56 Multivariate analysis assessing the influence of predicted donor KIR3DL1 inhibitory strength on one year NRM risk in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	230	1.00	-	-
9/10 HLA matched	65	1.33	0.72-2.47	0.31
<9/10 HLA matched	20	***3.46***	***1.71-6.99***	***0.002***
Donor-recipient CMV matching				
Matched	223	1.00	-	-
Mismatched	92	1.51	0.88-2.60	0.14
Predicted inhibitory strength				
Strong KIR3DL1 inhibition	90	1.00	-	-
Weak KIR3DL1inhibiton	95	1.18	0.63-2.20	0.60
Non-inhibitory KIR3DL1	130	0.70	0.36-1.36	0.29

Statistically significant results are denoted by ***bold italics***.

When comparing the strong predicted inhibition against the weak/non-inhibitory category, a trend was observed towards increased grades 2-4 aGVHD associated with strong predicted inhibition (20.1% *vs* 29.3%, p=0.075). When multivariate analysis by binary logistic regression was performed to adjust for stem cell source, donor age group and T cell depletion, the statistical trend status was lost (HR=1.59, CI=0.86-2.95, p=0.14, Table 6.57).

Table 6.57 Multivariate analysis assessing the influence of predicted donor KIR3DL1 inhibitory strength on incidence of grades 2-4 aGVHD in the overall cohort

Variable	Grades 2-4 aGVHD			
	N	HR	95% CI	*p-value*
HSC source				
BM	119	1.00	-	-
PBSC	159	***1.87***	***1.01-3.47***	***0.047***
Donor age, years				
<30	89	1.00	-	-
>30	189	***2.77***	***1.32-5.82***	***0.007***
T cell depletion				
Yes	264	1.00	-	-
No	14	***6.93***	***2.14-22.41***	***0.001***
Predicted inhibitory strength				
Weak/non-inhibitory KIR3DL1	196	1.00	-	-
Strong KIR3DL1 inhibition	82	1.59	0.86-2.95	0.14

Statistically significant results are denoted by ***bold italics***.

6.05.02.02 The influence of the MSK model in the RIC cohort

The MSK model of predicted KIR3DL1 inhibitory strength was also tested in the RIC cohort (all RIC transplants, including paediatric HCT). Here, the increased incidence of grades 2-4 aGVHD observed in the overall cohort became highly statistically significant in each of the different tested scenarios, although the strongest significance was observed when comparing the strong predicted KIR3DL1 inhibition against the combined weak/non-inhibitory predicted KIR3DL1 interactions that utilised donor HLA ligand information (strong *vs* weak/non-inhibitory KIR3DL1: 39.0% *vs* 16.0%, p=0.003, Table 6.58). As no clinical factors persisted in the forward stepwise selection of co-variables, multivariate analysis of grades 2-4 aGVHD incidence could not be performed.

Table 6.58 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 inhibitory strength in the RIC cohort

MSK model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Strong *vs* weak *vs* non-inhibitory KIR3DL1					
10/10 HLA-matched only	0.87	0.92	0.85	0.76	***0.021***
Any HLA-matching & donor HLA-Bw4 ligand	0.60	0.71	0.75	0.087	***0.004***
Any HLA-matching & recipient HLA-Bw4 ligand	0.80	0.93	0.58	0.12	***0.007***
Strong *vs* weak/non-inhibitory KIR3DL1					
10/10 HLA-matched only	0.65	0.88	0.76	0.90	***0.004***
Any HLA-matching & donor HLA-Bw4 ligand	0.86	0.75	0.90	0.46	***0.003***
Any HLA-matching & recipient HLA-Bw4 ligand	0.87	0.92	0.79	0.62	***0.003***

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table M.

6.05.02.03 The influence of the MSK model in the MAC cohort

The MSK model of predicted KIR3DL1 inhibitory strength was also tested in the MAC cohort (all MAC transplants, including paediatric HCT). No significant differences were observed in any of the tests (Table 6.59).

Table 6.59 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 inhibitory strength in the MAC cohort

MSK model	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Strong *vs* weak *vs* non-inhibitory KIR3DL1					
10/10 HLA-matched only	0.88	1.00	0.60	0.55	0.88
Any HLA-matching & donor HLA-Bw4 ligand	0.89	0.74	0.33	0.22	0.91
Any HLA-matching & recipient HLA-Bw4 ligand	0.97	0.87	0.34	0.27	0.87
Strong *vs* weak/non-inhibitory KIR3DL1					
10/10 HLA-matched only	0.92	0.97	0.45	0.37	0.69
Any HLA-matching & donor HLA-Bw4 ligand	0.99	0.81	0.18	0.31	0.82
Any HLA-matching & recipient HLA-Bw4 ligand	0.97	0.85	0.24	0.38	0.74

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table M.

6.05.03 The KIR3DL1 G238R and I320V polymorphisms

Two amino acid polymorphisms have been shown to increase KIR3DL1 functionality: G238R and V320I [152]. Each of the KIR3DL1 R^{238} variants, a group containing only the KIR3DL1*002, 054 and 002_c.1195G>C alleles, also encode the I^{320} variant and, as such, are predicted to have the highest functional strength. By contrast, several expressed alleles (including KIR3DL1*007 and many rarer alleles) encode G^{238}~V^{320} and are predicted to have weaker inhibitory function. The remaining alleles each encode G^{238}~I^{320} and have been assigned intermediate functional status. To investigate the impact of these dimorphisms on HCT outcomes, the presence of the different variants at each position were assessed independently within donor genotypes. In addition, presence of the intermediate functionality alleles was also assessed.

6.05.03.01 The influence of donor KIR3DL1 G238R and I320V polymorphisms in the overall cohort

When these polymorphisms were assessed in the overall cohort, no donors were observed to encode more than a single copy of the V^{320} polymorphism and, as such, the effect of CNV for this and the predicted low functionality (G^{238}~V^{320}) KIR3DL1 alleles were not performed. Similarly, only five donors encoded multiple copies of KIR3DL1 R^{238} alleles, resulting in the effect of copy number of this and the predicted high functionality (R^{238}~I^{320}) KIR3DL1 alleles not being tested. The results from the analysis assessing the effect of the remaining polymorphisms on HCT outcomes are summarised in Table 6.60.

Table 6.60 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 functionality polymorphism in the overall cohort

Donor KIR3DL1 functionality polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Expressed KIR3DL1 G^{238} pos *vs* neg	0.61	0.74	0.56	0.99	0.87
Expressed KIR3DL1 R^{238} pos *vs* neg	0.12	0.38	***0.021***	***0.001***	0.43
Expressed KIR3DL1 I^{320} pos *vs* neg	0.53	0.66	0.59	0.36	0.49
Expressed KIR3DL1 V^{320} pos *vs* neg	0.40	0.28	0.89	0.11	0.063
Expressed intermediate functionality KIR3DL1 pos *vs* neg	0.79	0.82	0.93	0.55	0.71

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table M.

This analysis revealed a strong correlation between the risk of NRM at one year post-transplant associated with presence of at least one expressed KIR3DL1 R^{238} allele (15.4% *vs* 32.4%, p=0.001, Table 6.60, Figure 6.19A). This factor also conveyed a significantly reduced risk of relapse at five years post-transplant (43.4% *vs* 29.0%,

p=0.021, Table 6.60, Figure 6.19B), perhaps explaining the lack of significant difference between OS probabilities in this cohort.

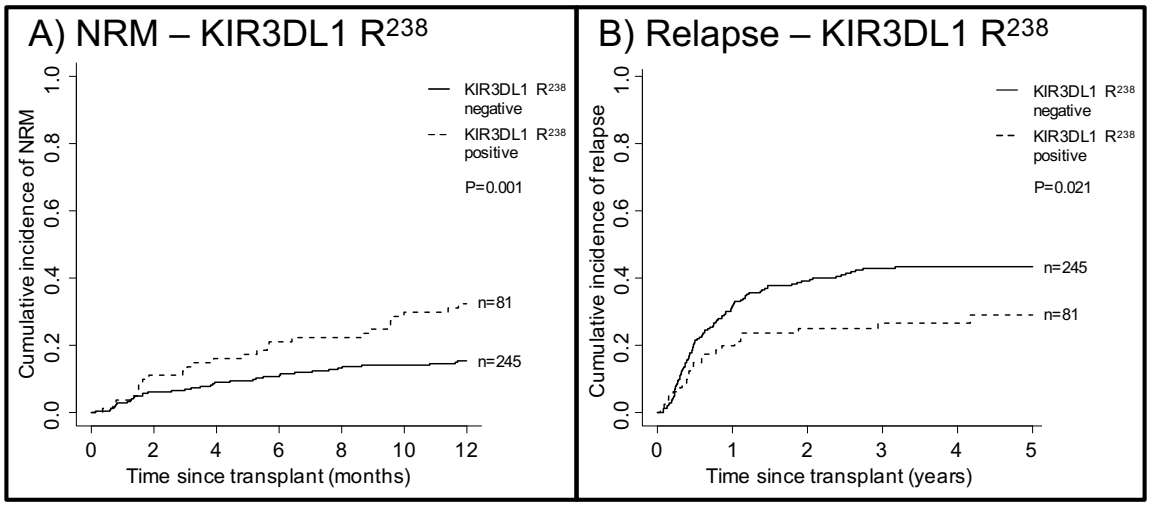


Figure 6.19 Probability of NRM and relapse investigating the presence of donor-encoded KIR3DL1 R^{238} alleles in the overall cohort

A) Univariate analysis reveals significantly increased probability of NRM associated with the presence of donor-encoded KIR3DL1 R^{238} alleles (p=0.001). B) Presence of the same polymorphism was also associated with a significantly reduced risk of relapse (p=0.021).

To assess these factors whilst adjusting for the impact of clinical factors, multivariate analysis was performed. For both NRM and relapse, the effect of donor-encoded KIR3DL1 R^{238} alleles retained statistical significance (NRM: HR=2.39, I=1.33-3.98, p=0.001, Table 6.61; Relapse: HR=0.61, CI=0.38-0.96, p=0.034, Table 6.62). No other predicted KIR3DL1 allele functionality polymorphism was associated with significantly different HCT outcomes in the overall cohort.

Table 6.61 Multivariate analysis assessing the influence of donor KIR3DL1 R^{238} alleles on the risk of NRM at one year post-transplant in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	230	1.00	-	-
9/10 HLA matched	65	1.44	0.78-2.66	0.25
<9/10 HLA matched	20	***3.21***	***1.62-6.37***	***0.001***
Donor-recipient CMV matching				
Matched	223	1.00	-	-
Mismatched	92	***1.76***	***1.06-2.92***	***0.029***
KIR3DL1 G238R polymorphism				
Donor KIR3DL1 R^{238} neg	238	1.00	-	-
Donor KIR3DL1 R^{238} pos	77	***2.39***	***1.44-3.98***	***0.001***

Statistically significant results are denoted by ***bold italics***.

Table 6.62 Multivariate analysis assessing the influence of donor KIR3DL1 R^{238} alleles on the risk of relapse at five years post-transplant in the overall cohort

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Previous autografts				
0	312	1.00	-	-
≥1	14	***2.17***	***1.26-3.75***	***0.005***
Donor-recipient gender matching				
Matched	185	1.00	-	-
Mismatched	141	0.79	0.56-1.13	0.20
KIR3DL1 G238R polymorphism				
Donor KIR3DL1 R^{238} neg	245	1.00	-	-
Donor KIR3DL1 R^{238} pos	81	***0.61***	***0.38-0.96***	***0.034***

Statistically significant results are denoted by ***bold italics***.

In addition to these highly significant findings, it was also possible to observe a statistical trend towards increased grades 2-4 aGVHD in recipients of HCT from donors who encoded weak functionality KIR3DL1 V^{320} alleles (21.7% *vs* 41.2%, p=0.063). However, in a multivariate analysis adjusting for stem cell source, donor age group and use of T cell depletion, this trend status was lost (HR=2.41, CI=0.73-7.94, p=0.15, Table 6.63).

Table 6.63 Multivariate analysis assessing the influence of donor KIR3DL1 V^{320} alleles on incidence of grades 2-4 aGVHD in the overall cohort

Variable	Grades 2-4 aGVHD			
	N	HR	95% CI	*p-value*
HSC source				
BM	119	1.00	-	-
PBSC	159	***1.93***	***1.04-3.60***	***0.037***
Donor age, years				
<30	89	1.00	-	-
>30	189	***2.97***	***1.42-6.19***	***0.004***
T cell depletion				
Yes	264	1.00	-	-
No	14	***6.44***	***1.95-21.26***	***0.002***
Predicted inhibitory strength				
Donor KIR3DL1 V^{320} neg	264	1.00	-	-
Donor KIR3DL1 V^{320} pos	14	2.41	0.73-7.94	0.15

Statistically significant results are denoted by ***bold italics***.

6.05.03.02 The influence of donor KIR3DL1 G238R and I320V polymorphisms in the RIC cohort

To assess whether the findings from the analysis on the overall cohort were consistent within the RIC and MAC only transplant cohorts, further analysis was performed. The results from the univariate analysis in the RIC cohort are below (Table 6.64).

Table 6.64 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 functionality polymorphism in the RIC cohort

Donor KIR3DL1 functionality polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Expressed KIR3DL1 G^{238} pos *vs* neg	0.51	0.92	0.95	0.85	0.10
Expressed KIR3DL1 R^{238} pos *vs* neg	0.077	0.60	0.17	0.15	0.61
Expressed KIR3DL1 I^{320} pos *vs* neg	0.85	0.76	0.52	0.81	0.16
Expressed KIR3DL1 V^{320} pos *vs* neg	0.20	***0.005***	0.16	0.11	0.31
Expressed intermediate functionality KIR3DL1 pos *vs* neg	0.59	0.75	0.93	0.62	0.069

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table M.

This analysis revealed that four of the five recipients of transplants from donors who expressed one KIR3DL1 V^{320} allele died within the first year post-transplant, whilst the fifth recipient relapsed within the same time frame. Although this conveyed a statistically significant p-value (DFS: 0% *vs* 30.3%, p=0.005, Figure 6.20), the small number of transplants of this type within the RIC cohort (n=5) place some doubt over the validity of this finding, and prevented a meaningful multivariate analysis.

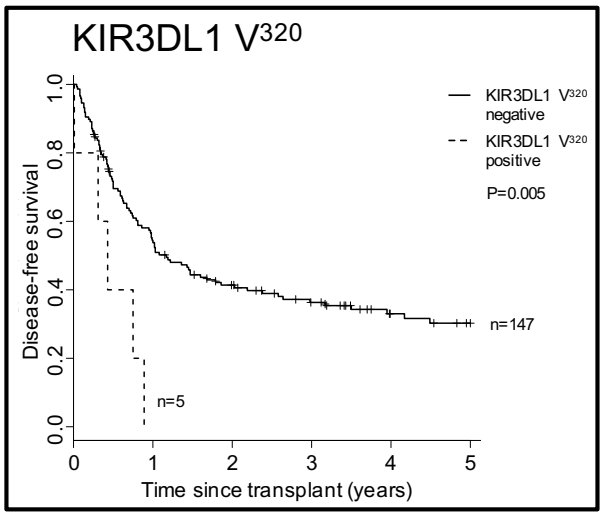


Figure 6.20 Probability of DFS investigating the presence of donor-encoded KIR3DL1 V^{320} alleles in the RIC cohort

Univariate analysis reveals greatly increased risk of mortality or relapse associated with the presence of donor-encoded KIR3DL1 V^{320} alleles (p=0.005). However, the small number of transplants involving these donors raises doubt over the validity of this finding.

In addition to the finding associated with functionally weak KIR3DL1 above, the presence of high functionality KIR3DL1 alleles (R^{238}) within the donor genotype was associated with a statistical trend towards reduced OS at five years post-transplant (39.7% *vs* 10.1%, p=0.077, Table 6.64, Figure 6.21). However, this did not withstand a multivariate analysis adjusting for EBMT disease risk score and donor-recipient CMV matching (HR=1.45, CI=0.87-2.38, p=0.15, Table 6.65).

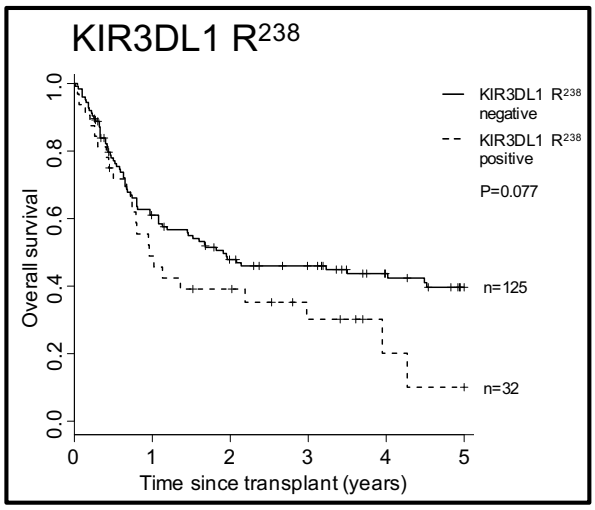


Figure 6.21 Probability of OS investigating the presence of donor-encoded KIR3DL1 R^{238} alleles in the RIC cohort

Univariate analysis reveals a statistical trend towards reduced probability of OS associated with the presence of donor-encoded KIR3DL1 R^{238} alleles (p=0.077).

Table 6.65 Multivariate analysis assessing the influence of donor KIR3DL1 R^{238} alleles on the probability of OS at five years post-transplant in the RIC cohort

Variable	5 year OS			
	N	HR	95% CI	*p-value*
Disease risk score				
Good	69	1.00	-	-
Intermediate	59	***1.67***	***1.05-2.70***	***0.030***
Poor	19	***1.92***	***1.02-3.70***	***0.042***
Donor-recipient CMV matching				
Matched	98	1.00	-	-
Mismatched	49	***1.67***	***1.09-2.56***	***0.020***
KIR3DL1 G238R polymorphism				
Donor KIR3DL1 R^{238} neg	117	1.00	-	-
Donor KIR3DL1 R^{238} pos	30	1.45	0.87-2.38	0.15

Statistically significant results are denoted by ***bold italics***.

A second statistical trend was observed, correlating the presence of donor-encoded KIR3DL1 G^{238}~I^{320} (intermediate functionality) alleles with increased grades 2-4 aGVHD incidence (11.1% *vs* 26.1%, p=0.069). However, as no clinical factors remained following forward stepwise selection of co-variates, no multivariate analysis was performed.

6.05.03.03 The influence of donor KIR3DL1 G238R and I320V polymorphisms in the MAC cohort

As the strong differences observed in the overall cohort were not confirmed in the RIC cohort, I hypothesised that presence of KIR3DL1 R^{238} alleles represented another variable that was dependent on particular conditioning regimen. As such, analysis was performed on the MAC cohort to confirm this hypothesis (Table 6.66).

Table 6.66 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 functionality polymorphism in the MAC cohort

Donor KIR3DL1 functionality polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Expressed KIR3DL1 G^{238} pos *vs* neg	0.99	0.78	0.54	0.90	0.099
Expressed KIR3DL1 R^{238} pos *vs* neg	0.54	0.48	***0.049***	***0.004***	0.55
Expressed KIR3DL1 I^{320} pos *vs* neg	0.48	0.49	0.79	0.31	***0.028***
Expressed KIR3DL1 V^{320} pos *vs* neg	0.74	0.94	0.84	0.45	0.13
Expressed intermediate functionality KIR3DL1 pos *vs* neg	0.85	1.00	1.00	0.77	***0.032***

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table M.

This univariate analysis confirmed the effects associated with KIR3DL1 R^{238} alleles in both relapse and NRM at five and one year post-transplant, respectively (relapse: 39.5% *vs* 23.3%, p=0.049; NRM: 16.6% *vs* 37.5%, p=0.004; Table 6.66, Figure 6.22). In addition, multivariate analysis also confirms each of these factors as correlating with significantly different HCT outcomes (relapse: HR=0.51, CI=0.26-0.99, p=0.048, Table 6.67; NRM: HR=2.86, CI=1.47-5.54, p=0.002, Table 6.68).

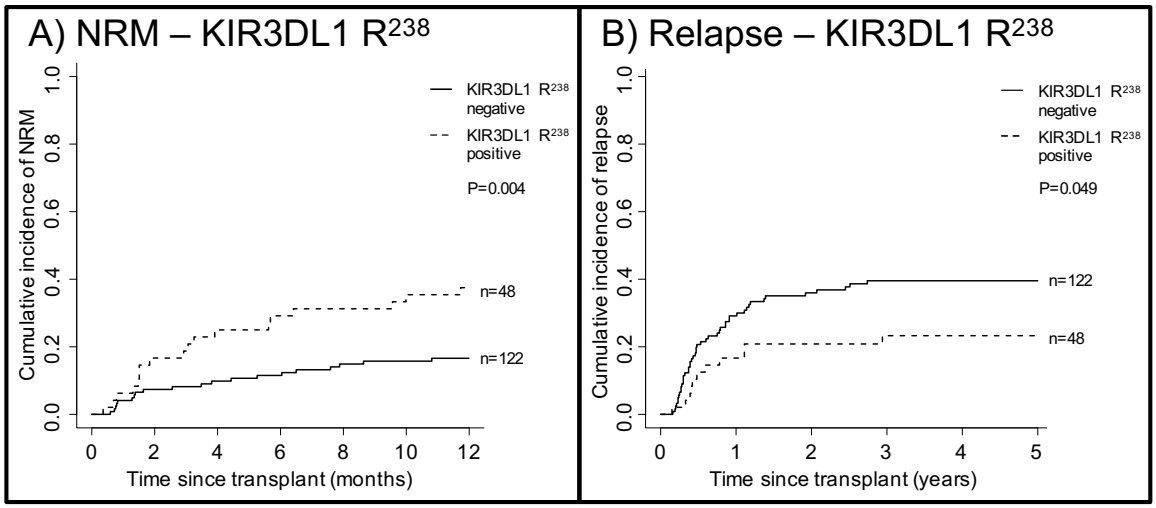


Figure 6.22 Probability of NRM and relapse investigating the presence of donor-encoded KIR3DL1 R^{238} alleles in the MAC cohort

A) Similarly to analysis in the overall cohort, univariate analysis reveals significantly increased probability of NRM associated with the presence of donor-encoded KIR3DL1 R^{238} alleles (p=0.004). B) Presence of the same polymorphism was also associated with a significant protection against relapse (p=0.049).

Table 6.67 Multivariate analysis assessing the influence of donor KIR3DL1 R^{238} alleles on the risk of relapse at five years post-transplant in the MAC cohort

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Disease risk score				
Good	69	1.00	-	-
Intermediate	80	1.28	0.72-2.30	0.40
Poor	20	***2.98***	***1.46-6.10***	***0.003***
KIR3DL1 G238R polymorphism				
Donor KIR3DL1 R^{238} neg	121	1.00	-	-
Donor KIR3DL1 R^{238} pos	48	***0.51***	***0.26-0.99***	***0.048***

Statistically significant results are denoted by ***bold italics***.

Table 6.68 Multivariate analysis assessing the influence of donor KIR3DL1 R^{238} alleles on the risk of NRM at one year post-transplant in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	125	1.00	-	-
9/10 HLA matched	34	0.68	0.25-1.87	0.45
<9/10 HLA matched	11	***3.26***	***1.45-7.32***	***0.004***
Recipient age, years				
<30	126	1.00	-	-
>30	44	***2.23***	***1.12-4.47***	***0.023***
KIR3DL1 G238R polymorphism				
Donor KIR3DL1 R^{238} neg	122	1.00	-	-
Donor KIR3DL1 R^{238} pos	48	***2.86***	***1.47-5.54***	***0.002***

Statistically significant results are denoted by ***bold italics***.

As well as the effects associated with the presence of donor-encoded KIR3DL1 R^{238} alleles, a statistically significant decrease in incidence of grades 2-4 aGVHD was observed following transplants involving donor-encoded KIR3DL1 I^{320} alleles (38.7% *vs* 20.1%, p=0.028). As samples expressing intermediate functionality alleles comprises a large proportion of the KIR3DL1 I^{320} subset, a weaker, yet still statistically significant, difference is also observed for this cohort stratification (34.0% *vs* 18.8%, p=0.032). However, when the expression of any KIR3DL1 I^{320} allele was assessed by multivariate analysis, statistical significance was lost, but remained a trend (HR=0.42, CI=0.16-1.11, p=0.080, Table 6.69).

Table 6.69 Multivariate analysis assessing the influence of donor KIR3DL1 I^{320} alleles on incidence of grades 2-4 aGVHD in the MAC cohort

Variable	Grades 2-4 aGVHD			
	N	HR	95% CI	*p-value*
HSC source				
BM	70	1.00	-	-
PBSC	74	***2.44***	***1.03-5.82***	***0.044***
Donor age, years				
<30	45	1.00	-	-
>30	99	***4.53***	***1.42-14.45***	***0.011***
T cell depletion				
Yes	137	1.00	-	-
No	7	***14.00***	***1.97-99.72***	***0.008***
KIR3DL1 I320V polymorphism				
Donor KIR3DL1 I^{320} negative	28	1.00	-	-
Donor KIR3DL1 I^{320} positive	116	0.42	0.16-1.11	0.080

Statistically significant results are denoted by ***bold italics***.

6.05.04 The KIR3DL1 P182S and W283L polymorphisms

As in the KIR molecules with two extracellular domains, ligand avidity of KIR3DL1 molecules is believed to be controlled, at least in part, by the interdomain hinge angle between the D1 and D2 domains. This, in turn, is believed to be impacted by the genetically linked P182S and W283L polymorphisms of the KIR3DL1 molecule, such that KIR3DL1 S^{182}~L^{283} alleles have higher avidity than P^{182}~W^{283} alleles [108,350]. Polymorphism at these residues in transplant donors was therefore investigated as a potential functional determinant of differential HCT outcomes.

6.05.04.01 The influence of donor P182S and W283L polymorphism in the overall cohort

To assess the significance of this polymorphism, presence, absence and estimated CNV of these dichotomous polymorphisms were investigated. As the high avidity, S^{182}~L^{283} allele group includes null expression alleles, only expressed copies of this allele group were considered to be S^{182}~L^{283} positive in this analysis. The univariate analysis testing

each of the polymorphism combinations against each of the different HCT outcomes revealed no significant differences in the overall cohort (Table 6.70).

Table 6.70 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 avidity in the overall cohort

Donor KIR3DL1 avidity polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Expressed high avidity KIR3DL1 pos *vs* neg	0.78	0.62	0.60	0.61	0.56
Expressed low avidity KIR3DL1 pos *vs* neg	0.88	0.79	0.32	0.12	0.88
Expressed high avidity KIR3DL1 CNV	0.92	0.87	0.64	0.78	0.91
Expressed low avidity KIR3DL1 CNV	0.90	0.96	0.61	0.27	0.27

The number of transplants within each subgroup of each test is given in Supplementary Table M.

6.05.04.02 The influence of donor P182S and W283L polymorphism in the RIC cohort

When the same tests were applied to the RIC cohort, a highly significant increase in grades 2-4 aGVHD was observed in transplants involving donors who expressed two low avidity KIR3DL1 alleles (0 *vs* 1 *vs* 2 copies, 16.3%, 16.4%, 45.2%, p=0.003, Table 6.71). As no clinical variables remained following forward stepwise selection of covariates, multivariate analysis was not performed.

Table 6.71 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 avidity in the RIC cohort

Donor KIR3DL1 avidity polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Expressed high avidity KIR3DL1 pos *vs* neg	0.76	0.66	0.55	0.35	0.24
Expressed low avidity KIR3DL1 pos *vs* neg	0.79	0.73	0.98	0.64	0.25
Expressed high avidity KIR3DL1 CNV	0.76	0.39	0.30	0.64	0.62
Expressed low avidity KIR3DL1 CNV	0.42	0.62	0.92	0.47	***0.003***

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table M.

6.05.04.03 The influence of donor P182S and W283L polymorphism in the MAC cohort

In the MAC cohort, presence of low avidity KIR3DL1 alleles was not associated with an increase in aGVHD, but was observed to correlate with a significant increase in the risk of NRM at one year post-transplant (10.2% *vs* 27.6%, p=0.018, Table 6.72, Figure 6.23). This was supported by multivariate analysis, where significance was retained (HR=3.44, CI=1.22-9.71, p=0.020, Table 6.73). Furthermore, when the copy number of low avidity donor KIR3DL1 alleles was assessed, a statistical trend was observed (0 *vs* 1 *vs* 2 copies, 10.2%, 28.4%, 25.9%, p=0.058, Table 6.72). Although multivariate analysis confirmed that both one and two copies of low avidity alleles within the genotype resulted in significantly increased NRM at one year post-transplant (1 copy: HR=3.49, CI=1.21-10.02, p=0.020; 2 copies: HR=3.31, CI=1.04-10.55, p=0.043, Table 6.74), these results were not significantly different from one another (HR=0.95, CI=0.47-1.93, p=0.88), discrediting theories relating to dose effect.

Table 6.72 Univariate analysis p-values of HCT outcomes comparing predicted donor KIR3DL1 avidity in the MAC cohort

Donor KIR3DL1 avidity polymorphism	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
Expressed high avidity KIR3DL1 pos *vs* neg	0.53	0.79	0.83	0.99	0.73
Expressed low avidity KIR3DL1 pos *vs* neg	0.67	0.61	0.14	***0.018***	0.20
Expressed high avidity KIR3DL1 CNV	0.57	0.67	0.96	0.79	0.94
Expressed low avidity KIR3DL1 CNV	0.82	0.63	0.32	0.058	0.36

Statistically significant results are denoted by ***bold italics***. The number of transplants within each subgroup of each test is given in Supplementary Table M.

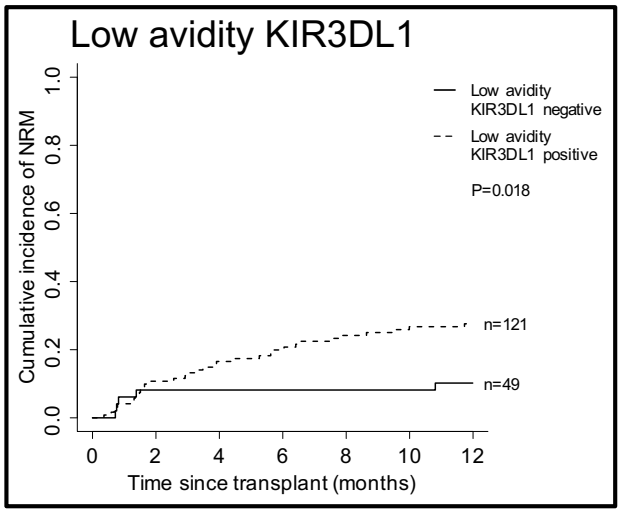


Figure 6.23 Probability of NRM investigating the presence of low avidity donor KIR3DL1 alleles in the MAC cohort

Univariate analysis reveals significantly increased probability of NRM associated with the presence of low avidity donor KIR3DL1 alleles (p=0.018).

Table 6.73 Multivariate analysis assessing the influence of low avidity donor KIR3DL1 alleles on the risk of NRM at one year post-transplant in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	125	1.00	-	-
9/10 HLA matched	34	0.74	0.27-2.01	0.55
<9/10 HLA matched	11	***3.77***	***1.85-7.68***	***<0.001***
Recipient age, years				
<30	126	1.00	-	-
>30	44	***2.19***	***1.12-4.26***	***0.021***
Low avidity donor KIR3DL1				
Low avidity donor KIR3DL1 neg	49	1.00	-	-
Low avidity donor KIR3DL1 pos	121	***3.44***	***1.22-9.71***	***0.020***

Statistically significant results are denoted by ***bold italics***.

Table 6.74 Multivariate analysis assessing the influence of low avidity donor KIR3DL1 CNV on the risk of NRM at one year post-transplant in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	125	1.00	-	-
9/10 HLA matched	34	0.73	0.27-2.01	0.55
<9/10 HLA matched	11	***3.77***	***1.85-7.66***	***<0.001***
Recipient age, years				
<30	126	1.00	-	-
>30	44	***2.18***	***1.11-4.25***	***0.023***
Low avidity donor KIR3DL1 CNV				
0 copies	49	1.00	-	-
1 copy	82	***3.49***	***1.21-10.02***	***0.020***
2 copies	39	***3.31***	***1.04-10.55***	***0.043***

Statistically significant results are denoted by ***bold italics***.

6.05.05 The KIR3DL1-HLA avidity score

When it is considered that polymorphism within the receptor, its ligand and even the peptide bound to the ligand can each affect the KIR-HLA binding avidity, the above analysis may appear crude [121,175,350,401]. Recent analysis, however, has begun to quantify and catalogue the avidity between different allelic variants of KIR and HLA [298]. Although, at present, the avidity of only three different KIR3DL1 alleles, KIR3DL1*001, 005 and 015, have been published, a wide panel of potential HLA ligands were assessed. By extracting the binding score as normalised to the best binding KIR3DL1-HLA ligand combination (KIR3DL1*005 and HLA-B*57:01), it was possible to generate a proxy avidity score for each different donor KIR-recipient ligand combination (Supplementary Table N).

Analysis was limited to those transplants in which the donor KIR3DL1/S1 allele genotype was comprised only of KIR3DL1*001, 005 and 015 alleles (as they have quantified interaction scores), unexpressed KIR3DL1 (as they have an interaction score of 0) and KIR3DS1 (as this is not currently thought to interact with classical HLA

molecules). Furthermore, although a large panel of HLA ligands were assessed in the original publication, several HLA-Bw4 alleles observed in the recipients of the UK cohort were not quantified. As such, 11 samples, encoding HLA-B*13:02 (seven samples), HLA-B*51:09 (two samples), HLA-B*44:04 (one sample) and a novel allele (since named, HLA-B*52:80, one sample), were excluded from analysis. As KIR3DL1-mediated immunity is traditionally only associated with HLA-Bw4 ligands, avidity scores with non-HLA-Bw4 alleles were not considered.

The total score for every different donor KIR3DL1-recipient HLA-Bw4 combination was assigned to each transplant permitted by these strict criteria. A model was then used to compare HCT outcomes from donor-recipient pairs with no KIR3DL1-HLA-Bw4 avidity (score=0, n=97) against pairs with a low score (<50, n=29) against pairs with a high score (≥50, n=55). The rationale behind this model was that even a small amount of inhibition would be sufficient to separate the effect of no avidity from low avidity KIR3DL1 interaction, whilst the effect of considerably increased avidity would be further distinguishable. When this scoring system was applied to the overall cohort in a univariate analysis, no significant differences in any of the outcomes were observed (Table 6.75). However, a statistical trend towards reduced NRM was observed for low interaction score pairs (no *vs* low *vs* high score: 12.2% *vs* 3.6% *vs* 20.2%, p=0.092), whilst a statistical trend was also observed correlating increased risk of grades 2-4 aGVHD incidence in transplants with high KIR3DL1-HLA-Bw4 interaction score (15.5% *vs* 15.4% *vs* 31.4%, p=0.083). However, neither of these statistical trends remained following multivariate analysis (Table 6.76 and Table 6.77). As the number of transplants for which it was possible to analyse KIR3DL1 interaction score was

impaired by the lack of available data on other KIR3DL1 alleles, analysis stratifying by conditioning regimen was not performed for this variable.

Table 6.75 Univariate analysis p-values of HCT outcomes comparing KIR3DL1 interaction score in the overall cohort

KIR3DL1 interaction score	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
No *vs* low *vs* high interaction score	0.23	0.25	0.88	0.092	0.083

The number of transplants within each subgroup of each test is given in Supplementary Table M.

Table 6.76 Multivariate analysis assessing the influence of KIR3DL1 interaction score on the risk of NRM at one year post-transplant in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
Donor-recipient CMV matching				
Matched	121	1.00	-	-
Mismatched	50	***3.28***	***1.41-7.65***	***0.006***
KIR3DL1 interaction score group				
Low interaction (score<50)	90	1.00	-	-
No interaction (score=0)	29	3.23	0.41-25.70	0.27
High interaction (score≥50)	52	5.21	0.67-40.75	0.12

Statistically significant results are denoted by ***bold italics***.

Table 6.77 Multivariate analysis assessing the influence of KIR3DL1 interaction score on the risk of grades 2-4 aGVHD incidence in the overall cohort

Variable	Grade 2-4 aGVHD incidence			
	N	HR	95% CI	*p-value*
T cell depletion				
Yes	143	1.00	-	-
No	9	***9.07***	***2.05-40.08***	***0.004***
KIR3DL1 interaction score group				
No interaction (score=0)	84	1.00	-	-
Low interaction (score<50)	23	0.98	0.25-3.82	0.97
High interaction (score≥50)	45	1.98	0.79-4.97	0.15

Statistically significant results are denoted by ***bold italics***.

6.06 Results: Combined analysis adjusting for multiple donor KIR allelic influences

Although several of the findings above are statistically significant within the parameters of their own locus, and by multivariate analysis adjusting for clinical cofactors, none have been corrected for multiple different KIR loci. As such, I sought to identify the impact of multi-locus allelic polymorphism, first by comparing the presence or absence of predicted KIR2DL2/3~KIR2DL1 CDS haplotypes, then by comparing the impact of several different key polymorphisms at each locus within multivariate analyses.

6.06.01 Estimated KIR2DL2/3~KIR2DL1 haplotype polymorphism

To assess the impact of the predicted donor KIR2DL2/3~KIR2DL1 haplotypes on HCT outcomes, transplants were stratified by the presence of the four most common haplotypes. These haplotypes account for over 80% of the total haplotype frequency (Chapter 5, Section 5.07). The results of this analysis in the overall, RIC and MAC cohorts are summarised in Table 6.78, Table 6.79 and Table 6.80, respectively.

Table 6.78 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3~KIR2DL1 haplotypes in the overall cohort

Donor KIR2DL2/3~KIR2DL1 estimated haplotype	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
2DL3*00101~2DL1*00302 haplotype present	0.31	0.26	0.69	***0.031***	0.93
2DL3*00201~2DL1*00201 haplotype present	0.92	0.86	0.65	0.56	0.40
2DL2*00301~2DL1*absent haplotype present	0.74	0.76	0.45	0.33	0.078
2DL2*00101~2DL1*00401 haplotype present	0.14	0.39	0.24	***0.024***	0.37

The number of transplants within each subgroup of each test is given in Supplementary Table O.

Table 6.79 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3~KIR2DL1 haplotypes in the RIC cohort

Donor KIR2DL2/3~KIR2DL1 estimated haplotype	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
2DL3*00101~2DL1*00302 haplotype present	0.86	0.75	0.32	0.62	0.56
2DL3*00201~2DL1*00201 haplotype present	0.66	0.41	0.096	0.80	0.19
2DL2*00301~2DL1*absent haplotype present	0.50	0.37	0.41	0.43	0.38
2DL2*00101~2DL1*00401 haplotype present	0.92	0.38	0.35	0.62	0.91

The number of transplants within each subgroup of each test is given in Supplementary Table O.

Table 6.80 Univariate analysis p-values of HCT outcomes comparing donor KIR2DL2/3~KIR2DL1 haplotypes in the MAC cohort

Donor KIR2DL2/3~KIR2DL1 estimated haplotype	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
2DL3*00101~2DL1*00302 haplotype	0.11	0.14	0.25	***0.003***	0.65
2DL3*00201~2DL1*00201 haplotype	0.85	0.72	0.34	0.50	0.94
2DL2*00301~2DL1*absent haplotype	0.35	0.26	0.77	0.10	0.13
2DL2*00101~2DL1*00401 haplotype	***0.023***	***0.034***	0.55	***0.013***	0.18

The number of transplants within each subgroup of each test is given in Supplementary Table O.

Unsurprisingly, these findings align closely with those investigating individual loci, owing to the strong LD displayed between the KIR2DL2/3 and KIR2DL1 loci. For example, presence of the KIR2DL3*00101~KIR2DL1*00302 haplotype was associated with significantly reduced NRM at one year-post-transplant in both the overall and MAC cohorts (overall cohort: 27.6% *vs* 17.4%, p=0.031; MAC cohort: 36.2% *vs* 16.0%, p=0.003), reflecting the presence of a neutral KIR2DL3 ($Q^{35}/T^{200}/P^{208}/R^{297}$) allele in combination with a KIR2DL1 allele encoding both L^{114} and R^{245} (these amino acid changes were both demonstrated to reduce NRM risk – see Sections 6.03.02 and 6.03.03). In addition, a strong detrimental effect was observed in recipients of HCT from donors encoding at least one copy of the KIR2DL2*00101~KIR2DL1*00401

haplotype, particularly in recipients whose pre-transplant conditioning regimen was myeloablative (OS: 40.1% *vs* 22.2%, p=0.023; DFS: 37.6% *vs* 19.1%; p=0.034, NRM: 21.1% *vs* 43.3%, p=0.013), reflecting the negative effects associated with both KIR2DL2/3 E^{35} and KIR2DL1 C^{245} variants. Finally, although only a statistical trend in univariate analysis, the presence of KIR2DL3*00201~KIR2DL1*00201 donor haplotypes was correlated with increased risk of relapse at five years post-transplant in RIC HCT recipients (35.6% *vs* 46.7%, p=0.096).

Interestingly, however, this analysis also revealed that not all KIR2DL1 R^{245} or KIR2DL2/3 E^{35} allele-containing haplotypes are equal. Despite both the KIR2DL3*00101~KIR2DL1*00302 and KIR2DL3*00201~KIR2DL1*00201 haplotypes being KIR2DL1 R^{245} positive Cen-A haplotypes with neutral KIR2DL3 Q^{35} alleles, one year NRM risk was only significantly reduced in the KIR2DL1*00302 donor transplants. Similarly, despite both the KIR2DL2*00301~KIR2DL1*absent and KIR2DL2*00101~KIR2DL1*00401 haplotypes being Cen-B haplotypes containing KIR2DL2/3 E^{35} alleles, significant detrimental effects in the overall and MAC cohorts were limited to transplants involving at least one donor-encoded KIR2DL2*00101~KIR2DL1*00401 haplotype.

Overall these findings were indicative of KIR2DL1 L^{114}~R^{245} alleles having a protective effect in the overall and MAC cohorts, whilst KIR2DL1 C^{245} alleles had a detrimental effect. The statistically significant effect of KIR2DL2/3 Q35E polymorphism observed was therefore hypothesised to primarily be a result of genetic association with these KIR2DL1 alleles.

6.06.02 Multivariate analyses of HCT outcomes adjusting for both centromeric and telomeric allelic polymorphism

Although the KIR2DL2/3 and KIR2DL1 loci are undeniably genetically linked, a much weaker linkage exists between the centromeric and telomeric motifs of the KIR haplotype. As such, multivariate analyses of several of the key KIR allele tests were performed to adjust for allelic factors in both of these regions as well as several clinical covariates surviving forward-stepwise selection. As centromeric allele polymorphism at the KIR2DL1 locus appears to elicit a stronger effect than KIR2DL2/3 polymorphism, only KIR2DL1 and KIR3DL1 allelic variation was considered. For the purposes of this analysis, only transplants in which the donor's KIR2DL1 and KIR3DL1 allele sequencing was complete were considered. Furthermore, the single transplant involving a donor encoding the KIR2DL1*037 allele (H^{245}) was excluded for reasons discussed in Section 6.03.02.

As both the presence of KIR2DL1 L^{114}~R^{245} alleles and absence of KIR3DL1 R^{238} alleles were associated with reduced NRM risk at one year post-transplant in the overall and MAC cohorts, these factors were assessed by multivariate analysis in their respective cohorts (Table 6.81 and Table 6.82). In the overall cohort, only the presence of KIR3DL1 R^{238} alleles remained statistically significant within the multivariate analysis model (HR=2.32, CI=1.40-3.86, p=0.001). However, in the MAC cohort, both the KIR2DL1 and KIR3DL1 allelic polymorphism factors retained statistical significance status (KIR2DL1 R^{245} presence: HR=0.43, CI=0.21-0.89, p=0.023; KIR3DL1 R^{238} presence: HR=2.83, CI=1.41-5.65, p=0.003).

Table 6.81 Multivariate analysis assessing the impact of influential donor KIR2DL1 and KIR3DL1 allelic polymorphism on the risk of NRM at one year post-transplant in the overall cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	229	1.00	-	-
9/10 HLA matched	65	1.39	0.75-2.57	0.30
<9/10 HLA matched	20	***3.17***	***1.56-6.43***	***0.001***
Donor-recipient CMV matching				
Matched	222	1.00	-	-
Mismatched	92	***1.69***	***1.01-2.82***	***0.045***
KIR2DL1 L^{114}~R^{245} allele presence				
Donor KIR2DL1 L^{114}~R^{245} negative	120	1.00	-	-
Donor KIR2DL1 L^{114}~R^{245} positive	185	0.76	0.45-1.28	0.30
Donor KIR2DL1 negative phenotype	9	0.68	0.21-2.21	0.52
KIR3DL1 R^{238} allele presence				
Donor KIR3DL1 R^{238} negative	237	1.00	-	-
Donor KIR3DL1 R^{238} positive	77	***2.32***	***1.40-3.86***	***0.001***

Statistically significant results are denoted by ***bold italics***.

Table 6.82 Multivariate analysis assessing the influence of key donor KIR2DL1 and KIR3DL1 allelic polymorphism on the risk of NRM at one year post-transplant in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	125	1.00	-	-
9/10 HLA matched	34	0.68	0.25-1.84	0.45
<9/10 HLA matched	11	***2.59***	***1.09-6.15***	***0.030***
Recipient age, years				
<40	126	1.00	-	-
>40	44	***2.66***	***1.27-5.58***	***0.010***
KIR2DL1 L^{114}~R^{245} allele presence				
Donor KIR2DL1 L^{114}~R^{245} negative	68	1.00	-	-
Donor KIR2DL1 L^{114}~R^{245} positive	97	***0.43***	***0.21-0.89***	***0.023***
Donor KIR2DL1 negative phenotype	5	0.97	0.28-3.36	0.96
KIR3DL1 R^{238} allele presence				
Donor KIR3DL1 R^{238} negative	122	1.00	-	-
Donor KIR3DL1 R^{238} positive	48	***2.83***	***1.41-5.65***	***0.003***

Statistically significant results are denoted by ***bold italics***.

In addition to an increased risk of NRM at one year post-transplant, the presence of donor-encoded KIR3DL1 R^{238} alleles was also associated with a significant reduction in relapse risk at five years post-transplant. To assess whether this withstood multivariate analysis adjusting for factors including the KIR2DL1 R^{245} polymorphism, this analysis was also performed in both the overall and MAC cohorts (Table 6.83 and Table 6.84).

Unsurprisingly, significance was retained in both cohorts (overall cohort: HR=0.58, CI=0.36-0.93, p=0.024; MAC cohort: HR=0.49, CI=0.24-1.00, p=0.049).

Table 6.83 Multivariate analysis assessing the influence of key donor KIR2DL1 and KIR3DL1 allelic polymorphism on the risk of relapse at five years post-transplant in the overall cohort

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Previous autograft history				
0	311	1.00	-	-
≥1	14	***1.93***	***1.10-3.39***	***0.022***
KIR2DL1 L^{114}~R^{245} allele presence				
Donor KIR2DL1 L^{114}~R^{245} negative	122	1.00	-	-
Donor KIR2DL1 L^{114}~R^{245} positive	193	0.85	0.59-1.22	0.37
Donor KIR2DL1 negative phenotype	10	1.20	0.40-3.62	0.75
KIR3DL1 R^{238} allele presence				
Donor KIR3DL1 R^{238} negative	244	1.00	-	-
Donor KIR3DL1 R^{238} positive	81	***0.58***	***0.36-0.93***	***0.024***

Statistically significant results are denoted by ***bold italics***.

Table 6.84 Multivariate analysis assessing the influence of key donor KIR2DL1 and KIR3DL1 allelic polymorphism on the risk of relapse at five years post-transplant in the MAC cohort

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Disease risk score				
Good	69	1.00	-	-
Intermediate	80	1.31	0.73-2.36	0.36
Poor	20	***2.96***	***1.42-6.17***	***0.004***
KIR2DL1 L^{114}~R^{245} allele presence				
Donor KIR2DL1 L^{114}~R^{245} negative	68	1.00	-	-
Donor KIR2DL1 L^{114}~R^{245} positive	96	1.09	0.63-1.87	0.76
Donor KIR2DL1 negative phenotype	5	1.70	0.35-8.19	0.51
KIR3DL1 R^{238} allele presence				
Donor KIR3DL1 R^{238} negative	121	1.00	-	-
Donor KIR3DL1 R^{238} positive	48	***0.49***	***0.24-1.00***	***0.049***

Statistically significant results are denoted by ***bold italics***.

The effects of independent allelic polymorphism at the KIR2DL1 and KIR2DL2/3 loci within the RIC cohort were less conclusive and compounded by an apparent lack of significant findings when considering the four most common centromeric haplotype motifs. However, a strong detrimental effect on relapse risk was observed in transplants involving donors who encoded the KIR2DL1 F^{-17} allele whilst presence of donor-

encoded KIR2DL2/3 Q^{35} alleles was associated with beneficial risk reduction in grades 2-4 aGVHD incidence but detrimental when considering five year relapse risk. When considering KIR3DL1 allelic polymorphism, a highly significant increase in grades 2-4 aGVHD risk was observed to be associated with donors encoding two copies of expressed, low avidity KIR3DL1 alleles. As such, these three KIR factors were assessed in a multivariate analysis investigating the probability of the incidence of grades 2-4 aGVHD and five year relapse risk following RIC HCT.

Table 6.85 Multivariate analysis assessing the influence of key donor KIR2DL1, KIR2DL2/3 and KIR3DL1 allelic polymorphism on the risk of relapse at five years post-transplant in the RIC cohort

Variable	5 year relapse			
	N	HR	95% CI	*p-value*
Previous autograft history				
0	123	1.00	-	-
≥1	7	0.82	0.12-5.46	0.84
KIR2DL1 F^{-17} allele presence				
Donor KIR2DL1 F^{-17} negative	59	1.00	-	-
Donor KIR2DL1 F^{-17} positive	67	0.90	0.29-2.79	0.86
Donor KIR2DL1 negative phenotype	4	0.97	0.11-8.71	0.98
KIR2DL2/3 Q^{35} allele presence				
Donor KIR2DL2/3 Q^{35} negative	14	1.00	-	-
Donor KIR2DL2/3 Q^{35} positive	116	1.95	0.52-7.33	0.33
KIR3DL1 low avidity allele CNV				
0 copies	38	1.00	-	-
1 copy	65	1.33	0.49-3.60	0.58
2 copies	27	1.71	0.50-5.79	0.39

Statistically significant results are denoted by ***bold italics***.

Table 6.86 Multivariate analysis assessing the influence of key donor KIR2DL1, KIR2DL2/3 and KIR3DL1 allelic polymorphism on the risk of grades 2-4 aGVHD in the RIC cohort

Variable	Grades 2-4 aGVHD			
	N	HR	95% CI	*p-value*
KIR2DL1 F^{-17} allele presence				
Donor KIR2DL1 F^{-17} negative	59	1.00	-	-
Donor KIR2DL1 F^{-17} positive	62	1.02	0.38-2.72	0.97
Donor KIR2DL1 negative phenotype	3	1.42	0.10-20.07	0.80
KIR2DL2/3 Q^{35} allele presence				
Donor KIR2DL2/3 Q^{35} negative	15	1.00	-	-
Donor KIR2DL2/3 Q^{35} positive	109	0.39	0.10-1.50	0.17
KIR3DL1 low avidity allele CNV				
0 copies	36	1.00	-	-
1 copy	63	1.12	0.37-3.40	0.84
2 copies	25	***4.00***	***1.17-13.63***	***0.027***

Statistically significant results are denoted by ***bold italics***.

Interestingly, despite several strong significance values in the independent univariate analysis, no KIR allelic polymorphism factors were associated with relapse risk at five years post-transplant in this multivariate analysis (Table 6.85). However, within the aGVHD multivariate analysis, despite not being possible to determine a difference associated with either the KIR2DL1 or KIR2DL2/3 allelic polymorphisms, these findings suggest that HCT was significantly more prone to result in grades 2-4 aGVHD when two copies of low avidity KIR3DL1 alleles were expressed by the donor (HR=4.00, CI=1.17-13.63, p=0.027, Table 6.86).

6.07 Results: an analysis of the effects of donor-recipient allelic matching at the KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 loci on HCT outcomes

In addition to the impact of donor KIR allelic polymorphism on HCT outcomes, investigation into the effect of allelic KIR matching between donor and recipient was also performed, extending the previous investigations into KIR genotype/haplotype matching performed in previous publications [282-284] and above (see Chapter 3, Section 3.07.02). As such, transplant pairs were coded as either allele matched, GVH

direction allele mismatched, HVG direction allele mismatched or bidirectional allele mismatched for each of the loci featuring complete allele sequencing between both donor and recipient. As this stratification resulted in some prohibitively small subgroup population sizes, comparison was made between pairs featuring mismatches in the GVH direction *vs* pairs without KIR allele GVH mismatching. In addition, the total number of different GVH direction mismatched alleles across the three loci were calculated and coded according to those transplants with three or fewer GVH mismatched alleles *vs* those transplants with four or more. Similarly to each of the previous analyses, the separate analyses were also performed in each of the cohorts stratified by conditioning regimen.

6.07.01 KIR allele matching in the overall cohort

In total, the number of transplants in which both the donor and recipient's allele typing were complete (a "pair") was 352, 229 and 245 for the KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 loci, respectively. In addition, when all loci were considered, the number of pairs with complete typing was 164. Univariate analysis assessing the impact of KIR allele matching for each of the different HCT outcomes was performed on these groups of pairs (Table 6.87).

Table 6.87 Univariate analysis p-values of HCT outcomes comparing donor-recipient KIR allele matching status in the overall cohort

Donor-recipient KIR allele GVH mismatching	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
KIR2DL1 alleles	0.67	0.53	0.30	0.15	0.072
KIR2DL2/3 alleles	0.23	0.18	0.46	0.68	0.71
KIR3DL1/S1 alleles	0.22	0.35	0.86	0.61	0.28
≤3 *vs* ≥4 allele mismatches	0.22	0.43	0.77	0.71	0.52

The number of transplants within each subgroup of each test is given in Supplementary Table P.

In the overall cohort, very little difference between any of the test conditions was observed, although a statistical trend towards increased grades 2-4 aGVHD was observed when comparing transplant pairs with *vs* without GVH direction KIR2DL1 allele mismatches (29.0% *vs* 19.4%, p=0.072, Table 6.87). When assessed in a multivariate analysis adjusting for TCD, donor age and stem cell source variables, the significance of this statistic remained only as a trend (HR=1.84, CI=0.97-3.49, p=0.061, Table 6.88).

Table 6.88 Multivariate analysis assessing the influence allelic KIR2DL1 GVH mismatching on the risk of grades 2-4 aGVHD in the overall cohort

Variable	Grades 2-4 aGVHD			
	N	HR	95% CI	*p-value*
T cell depletion				
Yes	278	1.00	-	-
No	15	***7.01***	***2.20-22.29***	***0.001***
Donor age, years				
<30	94	1.00	-	-
>30	199	***2.12***	***1.13-3.96***	***0.019***
HSC source				
BM	126	1.00	-	-
PBSC	167	1.71	0.98-3.00	0.060
KIR2DL1 allele GVH matching				
Matched	86	1.00	-	-
Mismatched	207	1.84	0.97-3.49	0.061

Statistically significant results are denoted by ***bold italics***.

6.07.02 KIR allele matching in the RIC cohort

Although little was observed in the overall cohort, stratification by conditioning regimen was also performed. In the RIC cohort, although no evidence of the trend towards increased aGVHD in KIR2DL1 GVH mismatched transplants (as observed in the overall cohort) persisted, a strong increase in both OS and DFS at five years post-transplant were observed when transplants featured a total of four or more GVH direction mismatches across the three loci tested (OS: 16.8% *vs* 60.0%, p=0.003; DFS: 6.9% *vs* 53.4%, p=0.008, Table 6.89, Figure 6.24). When assessed in multivariate

analysis scenarios, the difference in both OS and DFS remained statistically significant (OS: HR=0.41, CI=0.20-0.84, p=0.014, Table 6.90; DFS: HR=0.47, CI=0.25-0.89, p=0.021, Table 6.91).

Table 6.89 Univariate analysis p-values of HCT outcomes comparing donor-recipient KIR allele matching status in the RIC cohort

Donor-recipient KIR allele GVH mismatching	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
KIR2DL1 alleles	0.70	0.84	0.97	0.71	0.79
KIR2DL2/3 alleles	0.66	0.68	0.88	0.55	0.77
KIR3DL1/S1 alleles	0.27	0.43	0.93	0.57	0.45
≤3 *vs* ≥4 allele mismatches	***0.003***	***0.008***	0.26	0.15	0.61

The number of transplants within each subgroup of each test is given in Supplementary Table P.

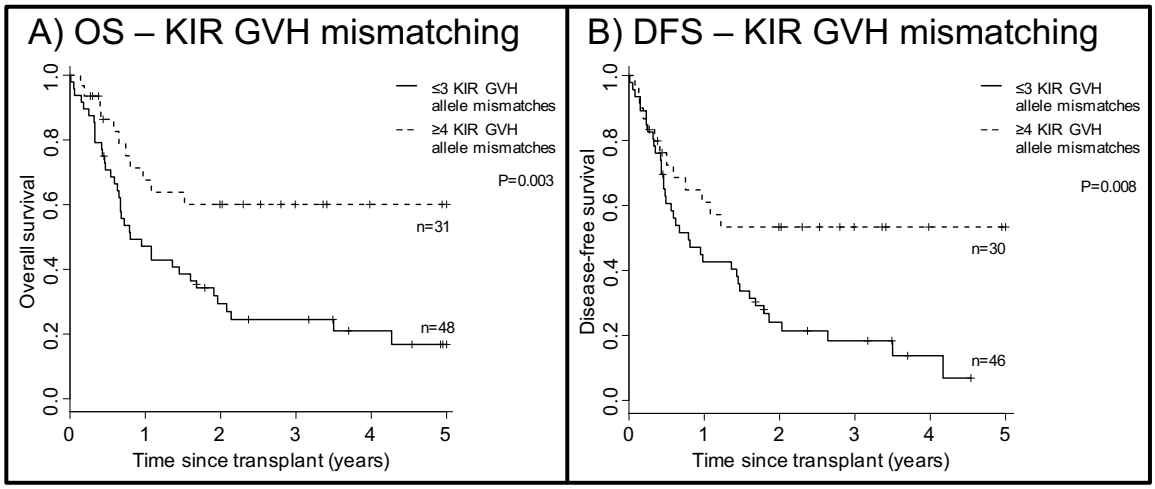


Figure 6.24 Probability of OS and DFS investigating the frequency of KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 GVH mismatches in the overall cohort

A) Univariate analysis assessing the probability of OS at five years post-transplant reveals significantly increased mortality associated with transplants involving three of fewer GVH direction allelic mismatches (p=0.003). B) The same stratification is also associated with significantly different DFS probability (p=0.008).

Table 6.90 Multivariate analysis assessing the influence of GVH direction allelic KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 mismatching frequency on the probability of OS at five years post-transplant in the RIC cohort

Variable	5 year OS			
	N	HR	95% CI	*p-value*
Donor-recipient CMV matching				
Matched	40	1.00	-	-
Mismatched	29	***2.44***	***1.32-4.55***	***0.005***
T cell depletion				
Yes	66	1.00	-	-
No	3	0.26	0.03-1.92	0.19
KIR allele GVH mismatching frequency				
≤3	43	1.00	-	-
≥4	26	***0.41***	***0.20-0.84***	***0.014***

Statistically significant results are denoted by ***bold italics***.

Table 6.91 Multivariate analysis assessing the influence of GVH direction allelic KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 mismatching frequency on the probability of DFS at five years post-transplant in the RIC cohort

Variable	5 year DFS			
	N	HR	95% CI	*p-value*
Donor-recipient gender matching				
Matched	42	1.00	-	-
Mismatched	34	***2.44***	***1.05-5.56***	***0.037***
Recipient gender				
Male	38	1.00	-	-
Female	38	0.56	0.24-1.30	0.17
KIR allele GVH mismatching frequency				
≤3	46	1.00	-	-
≥4	30	***0.47***	***0.25-0.89***	***0.021***

Statistically significant results are denoted by ***bold italics***.

Interestingly, despite having a strong impact on survival statistics, increased total number of KIR allele mismatches did not correlate with significantly different relapse (at five years post-transplant), NRM (at one year post-transplant) or grades 2-4 aGVHD risk, possibly due to the limited size of the available cohort.

6.07.03 KIR allele matching in the MAC cohort

Univariate analysis was performed to assess the impact of KIR allele GVH mismatching in the MAC cohort. Here, a highly significant difference in grades 2-4 aGVHD

incidence was observed to correlate with the presence of KIR2DL1 GVH mismatching (12.8% *vs* 34.9%, p=0.004, Table 6.92), as previously observed as a trend in the overall cohort. This increase in aGVHD also correlated with an increase in NRM probability at one year post-transplant (14.3% *vs* 30.3%, p=0.038, Table 6.92, Figure 6.25). Following multivariate analyses, performed to assess the influence of this factor whilst adjusting for other clinical variables, both outcomes remained significantly different when stratified by allelic KIR2DL1 GVH mismatching (NRM: HR=2.30, CI=1.01-5.26, p=0.047, Table 6.93; aGVHD: HR=8.41, CI=2.29-30.83, p=0.001, Table 6.94).

Table 6.92 Univariate analysis p-values of HCT outcomes comparing donor-recipient KIR allele matching status in the MAC cohort

Donor-recipient KIR allele GVH mismatching	5 year OS	5 year DFS	5 year relapse	1 year NRM	aGVHD (grade 2-4)
KIR2DL1 alleles	0.53	0.49	0.11	***0.038***	***0.004***
KIR2DL2/3 alleles	0.072	0.070	0.33	0.31	0.72
KIR3DL1/S1 alleles	0.59	0.55	0.62	0.99	1.00
≤3 *vs* ≥4 allele mismatches	0.36	0.27	0.52	0.63	0.17

The number of transplants within each subgroup of each test is given in Supplementary Table P.

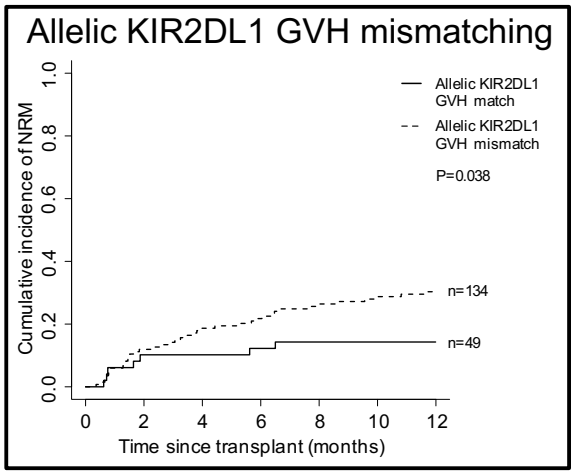


Figure 6.25 Probability of NRM investigating the presence of allelic KIR2DL1 GVH direction mismatches in the MAC cohort

Univariate analysis assessing the probability of NRM at one year post-transplant reveals significantly increased NRM risk associated with transplants involving allelic KIR2DL1 GVH direction mismatches (p=0.038).

Table 6.93 Multivariate analysis assessing the influence of allelic KIR2DL1 GVH direction mismatches on the risk of NRM at one year post-transplant in the MAC cohort

Variable	1 year NRM			
	N	HR	95% CI	*p-value*
HLA matching				
10/10 HLA matched	127	1.00	-	-
9/10 HLA matched	43	1.01	0.46-2.19	0.99
<9/10 HLA matched	13	***3.46***	***1.80-6.67***	***<0.001***
Previous autograft history				
0	175	1.00	-	-
≥1	8	2.84	0.95-8.48	0.061
Allelic KIR2DL1 GVH matching				
Matched	49	1.00	-	-
Mismatched	134	***2.30***	***1.01-5.26***	***0.047***

Statistically significant results are denoted by ***bold italics***.

Table 6.94 Multivariate analysis assessing the influence of allelic KIR2DL1 GVH direction mismatches on the risk of aGVHD in the MAC cohort

Variable	Grades 2-4 aGVHD			
	N	HR	95% CI	*p-value*
T cell depletion				
Yes	143	1.00	-	-
No	8	***40.54***	***4.95-332.34***	***0.001***
Donor age, years				
<30	44	1.00	-	-
>30	107	***4.90***	***1.69-14.15***	***0.003***
HSC source				
BM	78	1.00	-	-
PBSC	73	***2.23***	***1.00-4.95***	***0.049***
Allelic KIR2DL1 GVH matching				
Matched	41	1.00	-	-
Mismatched	110	***8.41***	***2.29-30.83***	***0.001***

Statistically significant results are denoted by ***bold italics***.

In addition to statistically significant differences, trends towards improved OS and DFS were observed for recipients of transplants matched for KIR2DL2/3 alleles in the GVH direction (OS: 31.8% *vs* 48.6%, p=0.072; DFS: 30.9% *vs* 49.4%, p=0.070, Table 6.92, Figure 6.26). However, when either of these findings were assessed by multivariate analysis, it was not possible to observe any significant differences resulting from KIR2DL2/3 GVH direction mismatching (OS: HR=1.54, CI=0.83-2.78, p=0.17; DFS: HR=1.52, CI=0.83-2.78, p=0.18).

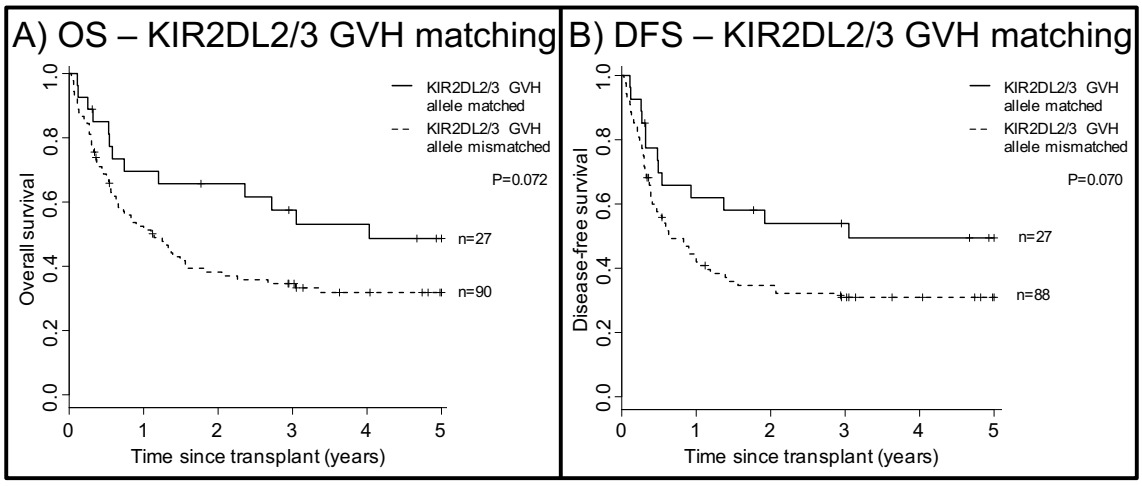


Figure 6.26 Probability of OS and DFS investigating the presence of allelic KIR2DL2/3 GVH direction mismatches in the MAC cohort

A) Univariate analysis assessing the probability of OS at five years post-transplant reveals a trend towards increased mortality risk associated with transplants involving allelic KIR2DL1 GVH direction mismatches (p=0.072). B) Equivalent analysis of DFS resulted in a similar finding (p=0.070)

6.08 Discussion

The field of research into the clinical effects of KIR allelic polymorphism has largely been hampered by the deficiency of high resolution typing methods. Although many interesting studies have been published investigating the presence or absence of particular KIR genes, motifs or haplotypes within an individual's genotype, the great extent of allelic variability that exists beyond presence/absence genotypes has been a bottleneck for further development. By sequencing just three KIR loci – KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 – I have already been able to detect differences within several outcomes of HCT that correlate with particular alleles or allele groups.

Previous publications had already highlighted the impact that some allelic polymorphisms at single loci have on HCT outcome. Amongst these are the studies by Bari *et al.* [169,270], in which the researchers first identified the differential KIR2DL1

signalling effects resulting from the R245C polymorphism, before applying this typing technique to the donors from a cohort of predominantly MAC paediatric HCT recipients and demonstrating that the presence of the 'CC' donor genotype (i.e. R^{245} negative), was associated with increased relapse and reduced OS. As the C^{245} allele is associated with weaker signalling, this study indicated that strong inhibitory signalling was associated with reduced relapse risk, perhaps suggesting that a strong licensing effect was responsible for the relapse protection. In their study, the presence of at least one copy of a KIR2DL1 R^{245} allele was sufficient to elicit this beneficial response.

To assess the applicability of this model in my HCT study cohort, the stratification of transplants by donor KIR2DL1 R245C genotype was reproduced. As the hypothesis that different KIR alleles may result in different NK cell licensing was being tested, and because absence of KIR2DL1 from the phenotype of donors was correlated with significantly different OS and DFS probability at five years post-transplant (when compared to KIR2DL1 positive donor phenotype), donors whose predicted NK cell phenotype was KIR2DL1 negative were removed from univariate analysis. The results from these analyses revealed that, indeed, the KIR2DL1 R245C polymorphism was sufficient to elicit a difference in HCT outcomes. As in the previous study, presence of donor KIR2DL1 R^{245} alleles was associated with beneficial OS and DFS probability in the HLA-matched MAC cohort. However, when comparing monotypic and heterotypic genotypes in a univariate analysis, it was possible to observe a diluted effect of this positive outcome in the presence of C^{245} alleles (heterozygotes), whilst C^{245} monotypic donors had the worst probability of five year OS and DFS of the different subgroups, although the latter two groups were not significantly different in a multivariate analysis. This differs subtly to the previous publication which found that at least one copy of R^{245}

was sufficient to offer protection, whilst my own study finds that at least one copy of C^{245} is sufficient to elicit detriment. Interestingly, however, no correlation with either NRM or relapse was observed in this initial analysis. By contrast, when the allelic CNV of donor KIR2DL1 R^{245} alleles was investigated, a substantial reduction in NRM risk was highly correlated with transplants involving donors who encoded two copies of KIR2DL1 R^{245} alleles, helping to confirm the finding from Chapter 3 that demonstrated reduced NRM risk with each copy of donor-encoded Cen-A motif in the MAC HCT setting (as R^{245} alleles are strongly associated with the Cen-A haplotype motif [165]). This finding also highlights the key difference between the monotypic R^{245} genotype (which can represent genotypes including the Cen-A/B genotype, assuming the encompassed Cen-B haplotype does not encode the KIR2DL1 gene), and samples with two copies of R^{245}, which are almost exclusively expected to encode the Cen-A/A genotype.

In addition to the potential errors introduced during genotyping of the previously published cohort [399], one potential reason for the differences in the findings of my study and that of Bari *et al.* (2013) [270] may be the TCD status of transplants. Approximately half of the published cohort underwent TCD, compared to 95% of my cohort (assessed in those recipients for which TCD data was available). As discussed in Chapter 3, prolonged TCD resulting from alemtuzumab-based transplant regimens may mitigate the effect of KIR polymorphism on relapse risk [327,328]. Furthermore, the published cohort was strictly paediatric recipients, potentially differing in additional conditioning protocols and even the disease itself [320,321]. These differences could feasibly have also resulted in considerable differences in HCT outcomes.

When the effect of KIR2DL1 R245C polymorphism was assessed in the HLA-matched RIC cohort, it was not possible to observe the same effects. Overall survival was not significantly different in any of the test conditions. Furthermore, the presence of multiple copies of donor KIR2DL1 R^{245} alleles was correlated with decreased DFS probability at five years post-transplant, supporting the previous observation that the presence of donor-encoded Cen-B haplotype motifs may protect against occurrence of NRM (see Chapter 3, Section 3.09.05). This further highlights the differences in the effects of KIR when stratifying transplants by conditioning regimen and may explain the lack of significant findings in the overall cohort. One hypothesis as to the cause of this variance may be that the strong inhibitory effects offered by KIR2DL1 R^{245} alleles augment alternative responses in the different inflammatory environments resulting from the distinct conditioning regimens. High toxicity, MAC results in increased inflammation that may provide sufficient stimulus to promote NK cells lacking sufficiently strong KIR2DL1-mediated inhibition to become reactive. These activated NK cells may target and lyse dendritic cells [343], recovering T cell subsets [332,341] or other immune cells that would otherwise offer protection against infection. However, in the RIC HCT scenario, the relatively low level inflammation may be insufficient to promote equivalent NK cell activation at this stage even in cells without strong KIR2DL1-mediated inhibition. As such, the strong inhibition offered by KIR2DL1 R^{245} alleles may limit alloreactivity against residual leukaemic cells following RIC HCT. Further hypotheses relating to the differences in KIR and, by association, NK cell function that result from different conditioning regimens are elaborated in more detail in Chapter 3, Section 3.11.

As the high resolution KIR allele typing generated sequence data across the entire length of the gene, it was also possible to investigate other polymorphisms at the KIR2DL1 locus. Differential avidity, attributed to a region of four polymorphisms – P114L, P154T, D163N and H182R [348] – was also explored. As a fairly strong LD exists between the latter three of these polymorphisms, it was possible to stratify the cohort according to three groups: low, intermediate and high avidity. Within the overall cohort, little effect of donor KIR2DL1 allele avidity was observed. However, similarly to other tests, this may simply represent the effect of conditioning regimen on transplants. Accordingly, the exclusive presence of high avidity KIR2DL1 alleles within the donor genotype correlated with an increase in relapse risk and corresponding decrease in DFS at the five year timepoint after RIC transplant. These high avidity alleles, including the common KIR2DL1*002 and 001 alleles, also predominantly encode R^{245}. This supports the hypothesis that the strongest inhibitory KIR2DL1 alleles, with strong avidity and strong signalling, may be responsible for hyporesponsivity against leukaemic cells, resulting in increased relapse risk.

Amongst the high expression KIR2DL1 alleles, the common alleles may only be stratified by a substitution within the signal peptide that does not form part of the mature receptor molecule at amino acid position -17. When RIC transplants were stratified by the presence or CNV of the different donor-encoded KIR2DL1 V-17F variants, it was possible to observe a strong correlation between donor F^{-17} alleles and increased relapse risk. It is possible that this substitution affects the function of the signal peptide, and that this has a detrimental effect in RIC HCT recipients [402]. Further assays, following the traffic of molecules with differing signal peptides, are required to determine the validity of this hypothesis.

An alternative hypothesis, however, is that alleles exhibiting the KIR2DL1 F^{-17} polymorphism are in strong linkage disequilibrium with particular alleles of other KIR loci, and that it is these alleles that are responsible for the detrimental effect. To explore this hypothesis, transplants were stratified by the presence of the four most common KIR2DL2/3~KIR2DL1 haplotype motifs within the donor genotype. However, the KIR2DL3*00201~KIR2DL1*00201 haplotype, on which the KIR2DL1 F^{-17} variant is most commonly located, was only correlated with a weak statistical trend towards increased relapse in RIC HCT recipients, perhaps suggestive that the KIR2DL3*00201 allele may, at least partially, negate this increased relapse risk. Alternatively, it may imply that other KIR2DL2/3 alleles that associate with KIR2DL1 F^{-17} alleles (mainly KIR2DL3*00101) may be associated with the increased risk of relapse, although this does not appear to be the case for transplants involving donors encoding the KIR2DL3*00101~KIR2DL1*00302 haplotype.

To better determine the role of KIR2DL2/3 allelic polymorphism, a similar set of analyses were performed, stratifying transplants according to the presence, absence or CNV of commonly-occurring polymorphisms. At the time of writing, no published analysis has focussed on the effect of high resolution KIR2DL2/3 allelic polymorphism on HCT outcomes. As such, I investigated the P16R, Q35E and R148C polymorphisms that have each been associated with KIR2DL2/3 functionality, as well as H50R, I200T, P208L and R297H amino acid substitutions that can be used to differentiate several different alleles.

The presence of E^{35} and combination of R^{16} and C^{148} have previously been established as significant contributing factors associated with enhanced ligand binding avidity and inhibitory functionality. In both cases, no residues are located at the ligand binding site, but, instead, an increase in the D1-D2 interdomain hinge angle is proposed to be responsible for the increased avidity exhibited by KIR2DL2/3 E^{35} and R^{16}~C^{148} alleles [105,174]. In the case of E^{35} alleles, increased stability has been hypothesised to originate from electrostatic interactions with H^{55} [339], whilst hydrogen bonding interactions between R^{16} and either A^{145} or E^{147} are proposed to stabilise the loops either side of the hinge. The presence of C^{148} is thought to bring these latter residues within hydrogen-bonding proximity of R^{16} [105].

Although not conserved across all published KIR2DL2/3 alleles [98], residues 16 and 148 in this cohort were observed in perfect linkage, such that R^{16}~C^{148} and P^{16}~R^{148} were the only observed genotypes at these residues. The presence of these genotypes also corresponded with the presence of KIR2DL2 and KIR2DL3 alleles, respectively. Supporting the observations in Chapter 3, presence of the R^{16}~C^{148} donor genotype correlated with a significantly increased NRM risk that conveyed a detrimental effect on OS and DFS probability in MAC HCT recipients. Perhaps more interestingly, however, was that a significant difference was also observed when, instead of comparing just KIR2DL2 R^{16}~C^{148} against KIR2DL3 P^{16}~R^{148} alleles, the presence of donor-encoded KIR2DL2/3 E^{35} alleles were compared against KIR2DL2/3 Q^{35} alleles. This suggests that the contribution of residues at positions 16, 35 and 148 are vital to the function of KIR2DL2/3 molecules and reduces the likelihood that the observed effect is related to some other intrinsic value of KIR2DL2, but not KIR2DL3, molecules, since both KIR2DL2 and KIR2DL3 can encode E^{35} variants.

That KIR2DL2/3 alleles with strong avidity (and increased killing capacity) are correlated with increased NRM and reduced survival in the MAC cohort supports the observation from Chapter 3 in which the presence of donor-encoded Cen-B haplotype motifs, often distinguishable by the presence of KIR2DL2 alleles, correlated with detrimental effects, and is discussed in depth in Chapter 3, Section 3.11. However, the finding that any high avidity KIR2DL2/3, including Cen-A associated KIR2DL3 E^{35} alleles, are correlated with this detrimental outcome demonstrates an improved resolution of analysis compared to KIR gene presence/absence genotyping alone, and may add to the functional understanding of the detrimental effects observed in MAC HCT.

Other polymorphisms shown to have an impact on the function of KIR2DL2/3, such as R41T (that leads to intracellular retention of KIR2DL2*004 alleles) [156], were not present in this relatively restricted cohort and, as such, were not investigated. However, several other polymorphic residues across the KIR2DL2/3 peptide sequence were investigated for a potential impact on HCT outcomes. Although several weak statistical trends were observed linking these allele-specific polymorphisms with particular HCT outcomes, it is likely that this represents genetic linkage between these polymorphisms and the functional polymorphisms discussed above. For example, the trend towards increased relapse in the RIC cohort associated with KIR2DL2/3 L^{208} alleles may simply reflect the LD between these alleles (including KIR2DL3*002) and KIR2DL1 F^{-17} alleles (including KIR2DL1*002). As the difference within relapse was more significant when assessing KIR2DL1 F^{-17}, this casts doubt over the significance of the L^{208} analysis. Furthermore, in CNV analysis, a highly significant ‘dose’ effect was observed

for donor-encoded KIR2DL1 F^{-17} variants that was not present when investigating donor KIR2DL2/3 L^{208} copy number, despite utilising identical samples. A second example, the significantly different NRM risk that correlated with the presence of donor-encoded KIR2DL2/3 I^{200} alleles in MAC recipients, may simply reflect that these alleles form a large part (52.4%) of the KIR2DL2/3 E^{35} allele group, which, together, have a much stronger significance in NRM. That five year OS and DFS probability were unaffected by the presence of donor KIR2DL2/3 I^{200}, and that CNV had no observable influence, supports this hypothesis.

Such strong LD between the KIR2DL2/3 and KIR2DL1 loci prevents assessment of the contribution of each of the loci involved, especially in small populations such as this HCT cohort. Furthermore, this LD often prevents the use of variables corresponding to individual loci within multivariate regression analyses. However, when the presence of at least one copy of the four most common haplotypes in the donor genotype was assessed, some effects attributed to particular polymorphisms only appeared to be associated with particular KIR haplotypes. For example, the KIR2DL1 R^{245} allele group shown to be protective in MAC transplants is shared by both KIR2DL1*002 and 003. However, only the haplotype including the KIR2DL1*003 allele (L^{114}~R^{245}) was associated with significantly reduced NRM risk in the overall and MAC cohorts. Furthermore, two separate Cen-B haplotype motifs were assessed. The first does not encode the KIR2DL1 gene whilst the other features the KIR2DL1*004 allele. Despite both haplotype motifs encoding a KIR2DL2 R^{16}~C^{148} allele, only the KIR2DL1*004 haplotype is correlated with the significant detrimental effects observed to correlate with KIR2DL2 alleles. This demonstrates two important factors: i) the distinction of Cen-A from Cen-B is not sufficient to distinguish function – significant variation exists

beyond this stratification; ii) the effect of KIR2DL1 polymorphism may outweigh the effect of KIR2DL2/3 polymorphism. This latter observation is especially interesting as KIR2DL1 appears to reconstitute slower than KIR2DL2/3 following HCT [403], although a recent published analysis demonstrates that alloreactive KIR2DL2/3 positive, but not KIR2DL1 positive, NK cells may be more susceptible to post-transplantation immunosuppression [404]. Unfortunately, as data relating to post-transplantation immunosuppression regimens was not available, this factor cannot be investigated within this analysis.

Interestingly, in the RIC cohort, the hypothesised increase in relapse risk that correlated with KIR2DL1 F^{-17} alleles was observed only as a weak statistical trend in the univariate analysis of common haplotypes. This may represent a scenario where the beneficial effect of the commonly associated KIR2DL3*002 allele is counterbalancing the negative effect of KIR2DL1*002. This would be in agreement with the hypothesis of Hilton *et al.* (2015) [348], in which the authors discuss the functional 'balance' created by allelic polymorphism at these loci.

Although the strong LD between the two loci presents challenges when attempting to stratify the effects of KIR2DL1 and KIR2DL2/3 allelic polymorphisms, no such issue exists when assessing KIR3DL1/S1, as this locus resides on the telomeric motif of the KIR haplotype, separated by a recombination hotspot. Analysis of this region, dominated by the MSK group, has focussed on the presence or absence of KIR3DL1/S1 and KIR2DS1 genes [300,317]. However, more recent analysis has also investigated the impact of allelic polymorphism at the KIR3DL1 locus [393]. Specifically, the predicted expression of KIR3DL1 (determined by five SNPs across the CDS [154]), and strength

of HLA-B ligand (distinguishing HLA-Bw6, HLA-Bw4-80I and HLA-Bw4-80T) was used to predict three different inhibitory statuses: strong, weak and non-inhibitory. This analysis by the MSK group revealed a survival and relapse risk advantage to recipients of transplants to treat AML involving HLA-matched donors with predicted low or non-inhibitory KIR3DL1-HLA-Bw4 combinations. However, when equivalent analyses were performed in the overall and MAC cohorts of my study, no significant differences pertaining to predicted inhibitory status were observed, despite the majority of the MSK study cohort also undergoing MAC HCT.

By contrast, strong effects of predicted inhibitory status were observed within the RIC cohort. Here, the presence of strong donor KIR3DL1 inhibition, when compared against the combined weak/non-inhibitory group, was significantly correlated with increased incidence of grades 2-4 aGVHD. Assuming that the inhibitory status prediction models proposed by this group are also appropriate to the HCT cohort in my study, this would suggest that high inhibition may be associated with increased severe aGVHD. This supports a hypothesis that NK cells characterised by strong KIR3DL1 inhibition are less capable of targeting DC cells that would otherwise present peptides to alloreactive T cells. That both T cell depletion and predicted KIR3DL1 inhibitory strength were both significant factors in a multivariate analysis of this model supports this hypothesis. Furthermore, the increase in grades 2-4 aGVHD was not associated with a statistical difference in NRM at one year post-transplant. This may also imply that the increase of aGVHD is balanced by a protection against infection, again supporting the hypothesis that DCs are targeted, as these would also be involved in presentation of viral antigens to cytotoxic T cells. However, as this effect is likely to occur within the first 100 days post-transplant (as this is the period in which aGVHD is most probable) and is restricted

to the RIC recipients, this hypothesis relies on an assumption that a sufficient recovery of alloreactive T cells exists, despite T cell depletion, and that T cell reconstitution is quicker in this scenario than an equivalent MAC HCT setting [405].

In an attempt to identify alternative methods by which KIR3DL1 allelic polymorphism may influence HCT outcomes, a more simplistic approach to donor KIR3DL1 expression was taken, excluding information relating to HLA ligand. In no scenario was it possible to observe a significant difference, although a trend towards increased NRM probability did exist in recipients of MAC HCT from donors encoding at least one high expression KIR3DL1 allele (high expression KIR3DL1 allele defined by the same criteria as Boudreau *et al.* (2014) [154]).

Instead, the functional strength of KIR3DL1 alleles, predicted by polymorphic residues at positions 238 and 320 [152], was tested. This revealed that the presence of at least one donor-encoded high strength KIR3DL1 allele (KIR3DL1 R^{238}) was sufficient to significantly reduce the risk of relapse five years after transplant, but concomitantly increase the probability of NRM at the one year post-transplant timepoint, although this effect appears to be most substantial following MAC transplantation. This may suggest that the degree of NK cell licensing operates on similar scale to the functionality of KIR3DL1 alleles, although this directly contrasts with the hypothesis of Boudreau *et al.* (2017) [393] who suggest that high functionality KIR-HLA-Bw4 interactions result in increased inhibition. A further noteworthy observation was that the balance between relapse and NRM is reminiscent of the beneficial reduction in relapse that associates with increased GVHD risk following HLA mismatched transplants [406]. However,

multivariate analysis that adjusted for HLA mismatching implied that both factors remained significant effectors of NRM in this cohort.

In addition to expression and functionality, allelic polymorphism at two positions within the KIR3DL1 gene, residues 182 and 283, have been demonstrated to influence avidity [108,350]. When this basic KIR3DL1 avidity factor was explored, low avidity alleles, defined as P^{182}~W^{283}, were detrimental in both RIC and MAC, but in subtly different ways. Following MAC HCT recipients, the risk of NRM was increased (without significant increase in grades 2-4 aGVHD) whereas in recipients of RIC HCT, the risk of grades 2-4 aGVHD was increased (without concurrent increase in NRM). As GVHD is a significant cause of NRM, the incidence of more severe aGVHD without concomitant NRM suggests that, following RIC HCT, KIR allelic polymorphism has opposing impacts on the risk of both aGVHD and alternative causes of NRM, such as infection. This data also supports a hypothesis in which dendritic cells, capable of presenting both viral and alloreactive antigens to T cells, are targeted by NK cells [343]. Unfortunately, due to low subgroup population sizes, it was not possible to perform a meaningful analysis stratifying NRM according to cause of death.

To further investigate these findings, HCT outcomes analysis was performed using a more refined binding interaction avidity assessment. Recent investigation has explored interaction affinities between three different KIR3DL1 alleles (KIR3DL1*001, 005 and 015) and a range of HLA ligands [298]. As both the KIR and HLA alleles have been determined at high resolution for this cohort, it was possible to devise a scoring system to estimate the overall binding affinity between KIR3DL1 and HLA for each sample in which the KIR3DL1 allele genotype was restricted to these three common alleles (or

included alleles known to be retained intracellularly). The scoring system utilised the normalised binding affinity % score calculated in the original study [298].

There was potential for results to be skewed during analysis featuring this score as all non-expressed KIR3DL1 alleles and KIR3DS1 alleles were included, but only a proportion of the expressed KIR3DL1 alleles were analysed. This analysis warrants repeating with complete predicted binding scores. Furthermore, equivalent analysis investigating avidity differences resulting from polymorphism within HLA-C and the KIR2DL1 and KIR2DL2/3 receptors would also be of benefit. Finally, this analysis assumed that KIR3DL1*004 and 019 alleles are completely intracellularly retained, with no signalling capacity [153]. However, separate analysis has revealed that a small proportion of KIR3DL1*004 is correctly folded, transported to the membrane and is capable of relaying an inhibitory signal upon binding HLA-Bw4 [351]. Although the low cell surface expression of these predominantly intracellularly retained alleles results in relatively low overall avidity between NK cells and their target cells (and corresponding weak inhibitory signalling), the assignment of a score of zero may not be appropriate. This further supports future analysis of this nature encompassing avidity data from all possible KIR receptor and ligand combinations.

To investigate the combined role of polymorphism within both the centromeric and telomeric haplotype motifs, multivariate analysis models assessed the impact of individual KIR polymorphism factors. As the KIR2DL1 and KIR2DL2/3 loci are in strong LD, analysis that included factors from both of these loci independently of one another was often precluded. In these circumstances, only KIR2DL1 factors were utilised for this analysis as the influence of these polymorphisms appeared to be

stronger than KIR2DL2/3 polymorphisms. In addition, to reduce the multiplicity of testing, only the tests resulting in the most significant of HCT outcome differences in univariate analyses were repeated in this manner. For instance, many different factors appeared to influence the probability of NRM at one year post-transplant, particularly in the overall and MAC cohorts. However, multi-locus multivariate analysis focussed only on the presence of donor-encoded KIR2DL1 L^{114}~R^{245} and KIR3DL1 R^{238} alleles. This revealed that, although significance was only retained for the KIR3DL1 allelic polymorphism in the overall cohort, both factors retained their significant effects in the MAC cohort. That these allele groups, that are both predicted to have high inhibitory function and intermediate avidity [152,169,298,348], correlate with opposing effects on NRM probability is perplexing, although it may be related to the timing of individual KIR gene reconstitution and the risks associated at different timepoints [404]. One hypothesis may be that, at early timepoints when KIR3DL1 has reconstituted but KIR2DL1 has not, NK cells with highly inhibitory KIR3DL1 alleles are less able to provide direct immunity against infections or are alloreactive against the new host. However, at later timepoints, following more complete KIR reconstitution (including KIR2DL1), acquisition of tolerance [333] and likely weaning of immunosuppressants, NK cells licensed by the highly inhibitory KIR2DL1 alleles are able recognize viral downregulation of HLA-C ligands and mount a strong immune response.

Interestingly, despite both KIR2DL1 and KIR3DL1 polymorphisms being correlated with significantly different NRM probability, only the KIR3DL1 polymorphism was associated with an apparent influence on relapse risk in the overall and MAC cohorts. This confirms the findings from univariate analysis. As it may be hypothesised that the effects of KIR polymorphism are most apparent in the presence of their ligand, this

finding may reflect the higher frequency of the KIR3DL1 ligand (HLA-Bw4) when compared with the KIR2DL1 ligand (HLA-C2). Unfortunately, small subgroup populations prevented a meaningful analysis that adjusted for recipients' ligand genotype.

Overall, few KIR-related factors were correlated with differential risk of relapse in the univariate analysis of the MAC cohort. However, in the RIC transplantation setting, presence of KIR2DL1 F^{-17} alleles within the donor genotype were associated with increased probability of relapse, and this finding appeared to feature a strong copy number effect. However, when the presence of donor-encoded KIR2DL1 F^{-17} alleles was assessed in a multivariate analysis including influential KIR2DL2/3 and KIR3DL1 polymorphisms, the association with increased relapse risk was lost. As this polymorphism is within the leader peptide that is cleaved from the mature receptor molecule, it is unlikely that this polymorphism was directly responsible for the increased relapse risk observed during multivariate analysis. It is possible that, as suggested by multivariate analysis, this effect is not accurate and instead represents a coincidentally significant difference observed in univariate analysis. This outcome becomes more probable as a result of multiplicity of testing.

However, when considering the incidence of grades 2-4 aGVHD in a similar multivariate analysis scenario to assess the impact of allelic KIR polymorphism on RIC HCT outcomes, the presence of two copies of low avidity KIR3DL1 alleles in the donor phenotype remained significantly associated with a detrimental aGVHD risk. This may suggest that, in the RIC HCT scenario, allelic polymorphism within the telomeric motif is more influential than centromeric polymorphism.

The discussion above focusses entirely on polymorphism within the donor KIR genotype. However, as with other genetic loci, it was possible to demonstrate an effect of allelic matching between donor and recipient KIR genotypes that potentially influences the outcomes of HCT. As KIR genotyping to this resolution is not routinely performed prior to HCT, and certainly does not feature in current donor selection algorithms, the frequency of allelic matches for these polymorphic genes is very low, precluding a meaningful analysis on allelic matching alone. However, when donors and recipients are mismatched in the GVH direction only, several interesting findings were observed. Firstly, following RIC HCT, a high frequency of GVH mismatches (four or more across the three tested loci) associated with increased OS and DFS at the five year post-transplant timepoint. This contradicts previous KIR genotype matching studies [283,284] as well as the standard model employed for matching HLA (in which increased number of mismatches correlates with reduced survival [216]), but may reflect beneficial recognition of foreign KIR alleles as minor histocompatibility antigens that promote alloreactivity against cancerous or infected cells without stimulating a strong GVHD response [201,282,407].

It was not possible to observe the same effect in MAC HCT, further demonstrating the difference between these two cohorts differentiated by conditioning. Instead, allelic KIR2DL1 GVH direction mismatching, but not equivalent mismatching at the KIR2DL2/3 or KIR3DL1/S1 loci, resulted in significantly increased NRM and aGVHD. One hypothesis may be that this is representative of the delayed reconstitution of KIR2DL1 relative to the other tested KIR loci [404]. This delay may postpone significant KIR2DL1 expression until after T cell recovery, preventing tolerization.

Delayed immune reconstitution following MAC HCT relative to equivalent RIC regimens may explain the different allelic KIR mismatch effects [324].

There are several issues with the current analysis. Perhaps most obviously, a great number of different tests were performed. As each additional test increases the risk of generating a type I error (leading to incorrect rejection of the null hypothesis), it is possible that some of the significant differences reported above are coincidental. Statistical techniques, such as Bonferroni's correction, may adjust probability values to account for this risk. However, as application of correction methods reduces the statistical power to reject an incorrect null hypothesis [408], and as the cohort assessed in this study is already underpowered by the small population size and sheer complexity of the KIR locus, statistical correction methods were not employed. However, future investigation, focussing on the influential polymorphisms identified in this analysis and utilising larger cohorts, may benefit from a more rigorous statistical analysis strategy that employs a correction technique.

Furthermore, it was not possible to generate a complete allelic genotype for all donor and recipient samples (see Chapter 5). This may have had several negative effects. Firstly, the reduced statistical power of smaller populations increases the probability of type II errors, in which false negatives are generated. This was minimised by performing multivariate analysis on each statistical trend as well as each significant difference. In addition, as many samples had partially complete allelic KIR genotyping, cohorts for analysis of each different locus differed slightly. This also reduced the population when multiple loci were considered together, as complete genotyping at each

locus is required. A similar reduction in population occurred when matching was considered, as complete genotyping was required for both donor and recipient samples.

As a result of issues relating to sample size, analysis of allelic KIR polymorphism was frequently performed upon less stringently filtered sub-cohorts. As such, most analysis stratified by conditioning regimen included all transplants matching that characterisation, including both adult and paediatric recipients, and HLA matched and mismatched transplants. As it is generally agreed that HCT for paediatric AML differs to equivalent adult HCT [320,321], and that HLA-mismatching can have detrimental effects on HCT outcome [205,216], this may have altered the observations. In an attempt to account for this, multivariate analysis included clinical factors to adjust probability values where appropriate. In addition, although the most significant differences in HCT outcomes were often observed to relate to a recipient's risk of NRM at one year post-transplant, the small sub-group population sizes available in this cohort prohibited additional analysis to stratify the causes of death in these cases.

Furthermore, KIR ligands were not frequently accounted for, particularly for KIR2DL1 and KIR2DL2/3 alleles. Although this decision is supported by the lack of significant findings during presence/absence genotyping (Chapter 3), a larger follow-up cohort may benefit from investigating the impact of KIR allelic polymorphism in the context of HLA ligand polymorphism. However, it should also be noted that alleles within the different ligand groups do not relay equal avidity with their KIR receptors. For example, HLA-C*03:02, 03:03, 03:04, 07:02 and 16:01 are all HLA-C1 ligands corresponding with strong binding to at least one KIR2DL2/3 allele. However, HLA-C*01:02, 08:01, 12:03 and 14:02, also each HLA-C1 ligands, are much weaker KIR binders [174].

Future analysis should account for this intra-ligand affinity variability in a similar method to that available for KIR3DL1-HLA-Bw4 interaction [298].

The final limitation of the current analysis is that only three KIR loci, accounting for five KIR genes, were investigated. As these are the best characterised of the KIR genes, and have more detailed analysis of allelic polymorphism than the other KIR loci, KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 were the practical choice to begin analysis in this time-restricted PhD project. However, future analyses should investigate allelic polymorphism at additional KIR loci, particularly for donor-recipient KIR allele matching studies.

Although several limitations exist, there are also some interesting impacts of polymorphism that, if verified in a suitable confirmatory cohort, could potentially improve donor selection to reduce adverse events post-transplant. For example, for patients in which a MAC regimen is indicated, the risk of NRM may be reduced by selection of donors encoding at least one KIR2DL1 L^{114}~R^{245} allele and/or avoidance of donor-encoded KIR3DL1 R^{238} alleles. However, for MAC patients with a high risk of relapse, the importance of KIR3DL1 R^{238} avoidance may be negated by its beneficial effect on the probability of relapse. In addition, where possible, GVH direction KIR allele mismatches should be avoided. By contrast, in patients for whom a RIC regimen is proposed, deliberate selection of these mismatches may be recommended.

6.09 Conclusions

To conclude, I have demonstrated that it is possible to apply high resolution allelic KIR genotyping to a retrospective cohort of donors and their respective recipients of HCT to

treat AML at UK hospitals. Although analysis was focussed on the influence of donor KIR allele genotype, the importance of both donor and recipient KIR genotypes was highlighted by significant differences in HCT outcomes based on GVH direction KIR allele mismatches. Similarly to observations in the KIR presence/absence genotyping, conditioning regimen has a large impact on the influence of allelic KIR polymorphism. However, by improving the resolution of KIR genotyping, it was possible to demonstrate that not all Cen-A, nor Cen-B nor Tel-A, haplotype motifs are equal, giving the potential to refine donor selection strategies that are beginning to encompass KIR genotype information [409]. Further investigation to confirm the findings from this study and extend them to other KIR loci is warranted.

Chapter 7 Conclusions

The use of allogeneic HCT remains the mainstay of treatment for high risk AML patients. However, despite efforts to reduce post-transplant complications (such as relapse, infection and GVHD) by clinical immunomodulatory techniques (including conditioning, DLI and TCD), selection of the most immunocompatible donors remains the primary method to improve the survival and quality-of-life of recipients, post-transplant [410]. Although immunocompatibility encompasses a range of factors, including information pertaining to CMV status of both donor and recipient, HLA matching is the most important determinant of HCT outcome. However, genetic polymorphism at other loci are also undoubtedly correlated with differential outcomes [411-414].

In this study, I have investigated the extent of polymorphism within a UK cohort across the KIR haplotype, an extremely polymorphic region of the human genome that exhibits extensive gene copy number and allelic variation. The receptor molecules that KIR genes encode elicit either inhibitory or activating responses within the NK cell on which they are expressed, making them important modulators of the immune response. To determine the relevance of KIR polymorphism on HCT outcomes, I have utilised the genotyping data of both donors and recipients and correlated genetic polymorphisms with clinical follow-up data. Although extensive investigation has been performed previously to investigate the importance of different donor gene and haplotype motifs on VUD HCT outcomes [83,84,272,273,300,314,316,317], very little has been previously published specifically utilising a UK cohort. As HCT practice in the UK differs from many other transplant centres worldwide, in its wide use of alemtuzumab

as a TCD agent, my cohort represents an important transplant protocol variation in which the effects of KIR required investigation.

Although current VUD selection strategies do not routinely encompass KIR genotype polymorphism within the donor or recipient, recent efforts have been made to incorporate preferential selection of donors with a "Better/Best" genotype, as defined by Cooley et al. (2014) [272,409]. However, when I investigated the impact of this donor genotype in a cohort of 405 donor-recipient UK HCT pairs, it did not have a uniform beneficial effect. Instead, it appears that, in my cohort at least, the effect of the KIR BX genotype is strongly related to the conditioning regimen employed. Transplants in which a MAC regimen precedes donor cell infusion were, by contrast, correlated with strongly detrimental effects, demonstrated by a reduction in OS and DFS probabilities associated with a significantly increased risk of infectious mortality. The distinctive use of TCD by alemtuzumab as GVHD prophylaxis between the two studies may be a crucial contributing factor to this apparent contradictory finding and warrants further prospective studies to investigate the role of TCD and conditioning on the influence of KIR polymorphism.

As the results of my study appear to contradict the current dogma of VUD selection based on KIR parameters, a more detailed analysis of KIR polymorphism was undertaken. To do this, I first needed to design and develop a KIR allele typing strategy. Full gene amplification targeting individual (or pairs of) KIR genes by PCR was devised utilising semi-generic primer pairs that should, theoretically, allow amplification of all currently published allele sequences as well as novel sequences formed through inter-gene recombination. Amplicons were then prepared into libraries

and sequenced using SMRT DNA sequencing technology. This TGS technique was selected as the very long, high quality DNA sequencing reads generated allowed the unambiguous determination of KIR allelic genotype. The methodology was validated against a panel of cell line samples with known KIR allelic genotypic and revealed a high degree of concordance with the expected genotypes. In some instances, errors in the previously published KIR genotypes, often resulting from phase ambiguity, were corrected, demonstrating the suitability of my strategy for high resolution allele typing of the extremely polymorphic KIR genes.

Following validation of this technique, genotyping of the KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 loci was performed for my cohort to determine the importance of allelic KIR polymorphism on the outcomes of HCT. This analysis also revealed key distinctions between the impact of KIR polymorphism apparently dependent on the conditioning method employed as part of the preparative regimen. In another discovery of interest, stratification of individual KIR haplotype motifs by their allelic variation demonstrated key distinctions that are only apparent at the higher level of allelic resolution. This reveals that donor selection models based on gene content haplotypes may be further refined by allelic typing to improve recipient outcomes.

Furthermore, the findings indicate that donor selection based on allelic KIR genotyping information may be more feasible than previously thought. Although this study utilised high resolution KIR allele sequencing using a TGS platform, my results suggest that genotyping only a small number of amino acids at each locus is sufficient to significantly improve HCT outcomes. This would be easily achievable through cheaper and higher throughput genotyping assays such as PCR-SSP, that are already available to

a much wider range of histocompatibility laboratories worldwide. By immediately employing these techniques within prospective clinical trials, the results of my study may be validated and beneficial effects realised very quickly. Selection criteria for adult recipients undergoing TCD HCT may be as follows:

- For patients in which MAC is preferred, and who have a low relapse risk, the increased detrimental NRM risk associated with donor-encoded KIR3DL1 R^{238} alleles may outweigh its beneficial effect on relapse risk. As such, in these instances, donor selection strategies should prioritise donors encoding KIR2DL1 L^{114}~R^{245} and KIR3DL1 G^{238} alleles whilst avoiding KIR2DL2/3 E^{35} alleles. In addition, should the recipients' own KIR2DL1 type allow it, KIR2DL1 allele-matching may be of added benefit.
- However, for patients with high relapse risk and who can tolerate MAC regimens, donors encoding KIR2DL1 L^{114}~R^{245} and/or KIR3DL1 R^{238} alleles should be preferentially sought, in addition to avoidance of KIR2DL2/3 E^{35}. Where possible, KIR2DL1 allele matching should also be prioritised.
- In patients in which RIC is proposed, however, KIR allele mismatching is encouraged for the beneficial increased survival probability it relays. In addition, multiple donor-encoded P^{182}~W^{283} (low avidity) KIR3DL1 alleles may also be encouraged. Furthermore, donors encoding multiple copies of KIR2DL1 R^{245} alleles should be avoided, especially if these alleles also encode F^{-17}.

To broadly assess the likelihood of being able to select donors with preferable allelic KIR genotypes, population analysis on the CDS genotypes determined from the UK HCT cohort was performed. This revealed that, for both KIR2DL1 and KIR2DL2/3 loci, few common alleles encompass the majority of the allele frequency, although large

numbers of alleles exist at low frequency for KIR2DL1. By contrast, the KIR3DL1/S1 locus is composed of a larger number of more frequent variants, in addition to many rare alleles.

Although refined strategies employing greater resolution allelic typing are often thought to confound donor selection, I suggest this is not necessarily the case. For instance, KIR2DL1*003-like alleles, characterised by L^{114}~R^{245} and predicted to reduce NRM risk in MAC HCT recipients, share almost 40% of the overall KIR2DL1 allele frequency. As such, it is likely that over 60% of the population encode at least one copy of this beneficial allele group. Furthermore, the KIR3DL1*002-like allele group, encoding R^{238} and demonstrated to increase NRM risk but decrease relapse risk following MAC HCT, is predicted to be present in almost a quarter of the UK population, and the presence of KIR2DL2/3 E^{35}, detrimental when encoded by donors for MAC recipients, may be avoided in approximately half of the UK population. As such, selection to include or avoid these donor features from a relatively small donor option pool remains eminently feasible. Similarly, although selection of donors with completely matched allelic KIR genotypes remains highly improbable, it is much easier for GVH direction mismatches to be deliberately selected for or against, depending on the proposed conditioning regimen. Furthermore, although undoubtedly controversial, data from my study suggests that single HLA mismatches may often be better tolerated than some donor KIR factors in the particular transplant conditions assessed. Given the abundance of evidence supporting the detrimental effect of single HLA mismatches [205,216,415], this observation would require significant validation before it could be applied to influence clinical decisions. However, if confirmed in larger, contemporary

cohorts, this factor may also be utilised to increase the available pool in which to search for the optimal donor.

The influence of the KIR2DL1 R245C polymorphism on MAC HCT outcomes is striking and has a dominant effect over the result of KIR2DL2/3 polymorphism. Presence of KIR2DL1 C^{245} alleles conveys a significant increase in NRM risk. Given the high incidence of infection-related mortality in this sub-cohort, I hypothesize that the differences relate to variable levels of antiviral immunity offered by NK cells expressing the different KIR allotypes.

The "tunable rheostat" model of NK cell licensing suggests that multiple strong inhibitory interactions result in increased NK cell immunological potential [124]. KIR2DL1 C^{245} alleles have been demonstrated to have lower expression [388,416] and promote less downstream signalling [169]. As such, these alleles are likely to confer a weaker NK cell phenotype than those expressing KIR2DL1 R^{245} alleles. When placed in a setting of viral immunity (e.g. CMV reactivation induced by immunosuppression), weaker NK cell licensing may result in a decreased capacity to lyse infected cells that have downregulated HLA class I expression as a means of T cell evasion. This may result in the increased probability of NRM observed in recipients receiving HCT from donors encoding KIR2DL1 C^{245} alleles.

However, the presence of KIR3DL1 R^{238} alleles is also associated with increased NRM probability, despite encoding a stronger inhibitory receptor (compared to G^{238} alleles) [152]. This may indicate that, instead of targeting infected cells with downregulated HLA class I expression, NK cells expressing this allele may instead elicit an increased

immune response against antiviral $CD4^+$ T cells [332,341] or dendritic cells [343]. By eliminating these immune cells, infection may persist and present clinically as increased non-relapse mortality.

Future investigation could utilise several different assays to test these hypotheses. First, the immunological capacity of NK cell phenotypes representing those recovering following highly lymphodepleting TCD, MAC HCT could be assessed upon exposure to a range of different target cells including HLA class I deficient cell lines, $CD4^+$ T cells, dendritic cells, CMV infected target cells or AML blasts. This might encompass measurements such as IFN-γ secretion or CD107a expression. To account for other polymorphisms across the entire KIR haplotype, comparison of allele types could utilise NK cell lines and site-directed mutagenesis to specifically alter only the residues in question. Furthermore, to validate the specificity of these findings to HLA class I downregulation, targeted overexpression of HLA-C1, HLA-C2 and HLA-Bw4 ligands may be used as "rescue" experiments in which killing is not hypothesized.

Recreation of the specific conditions that occur after HCT is difficult to achieve within the laboratory setting. As such, further experiments to assess the influence of these polymorphisms could utilise ongoing HCT in the form of a clinical trial. Donors encoding each combination of KIR2DL1 R245C and KIR3DL1 G238R variants could be selected for patients undergoing HCT to treat AML, and the incidence of viral infection and NRM monitored post-transplant. By assessing in a prospective manner, it would be possible to control for a wide range of factors (e.g. donor age, CMV-matching, exact conditioning regimen, cell dose, etc.). Furthermore, post-transplant samples could be collected, from which immune cell populations can be characterised

and recovering NK cells could be isolated. KIR allotype-specific subsets could then be assessed for immunological capacity against infected cells and other immune cell subtypes (similarly to that described above). In addition, the appearance of individual KIR allotypes at different timepoints may offer further insight into the cause of NRM associated with these particular KIR genotypes.

Overall, although limited in some respects, my project has laid foundations for future studies to further explore allelic KIR polymorphism and has uncovered several key observations. For immediate clinical impact, simplified genotyping strategies assessing the limited number of KIR2DL1, KIR2DL2/3 and KIR3DL1/S1 amino acid polymorphisms discussed above may add great benefit to donor selection strategies. However, research utilising high resolution KIR allele sequencing should also continue as the complexity of the KIR locus, reinforced by this study, demonstrates the need for unambiguous allele sequencing to confidently determine allele type. At present, this may only be achieved through full-length, phased sequencing analysis. Although further follow-up studies that investigate allelic polymorphism in all KIR genes in larger, more contemporary cohorts are certainly warranted to better refine these donor selection models and gain deeper insight into KIR biology, this thesis has demonstrated the extent and importance of the next layer of polymorphism at the clinically-relevant KIR locus.

References

[1] Davis MM, Bjorkman PJ "T-cell antigen receptor genes and T-cell recognition" (1988) *Nature* **334 (6181)**, p.395-402

[2] Dreyer WJ, Bennett JC "The molecular basis of antibody formation: a paradox" (1965) *Proc Natl Acad Sci U S A* **54 (3)**, p.864-869

[3] Bassing CH, Swat W, Alt FW "The mechanism and regulation of chromosomal V(D)J recombination" (2002) *Cell* **109 Suppl**, p.S45-55

[4] Starr TK, Jameson SC, Hogquist KA "Positive and negative selection of T cells" (2003) *Annu Rev Immunol* **21**, p.139-176

[5] Abbas AK, Murphy KM, Sher A "Functional diversity of helper T lymphocytes" (1996) *Nature* **383 (6603)**, p.787-793

[6] Bossi G, Griffiths GM "CTL secretory lysosomes: biogenesis and secretion of a harmful organelle" (2005) *Semin Immunol* **17 (1)**, p.87-94

[7] Anel A, Buferne M, Boyer C, Schmitt-Verhulst AM, Golstein P "T cell receptor-induced Fas ligand expression in cytotoxic T lymphocyte clones is blocked by protein tyrosine kinase inhibitors and cyclosporin A" (1994) *Eur J Immunol* **24 (10)**, p.2469-2476

[8] Sallusto F, Lenig D, Forster R, Lipp M, Lanzavecchia A "Two subsets of memory T lymphocytes with distinct homing potentials and effector functions" (1999) *Nature* **401 (6754)**, p.708-712

[9] Brack C, Hirama M, Lenhard-Schuller R, Tonegawa S "A complete immunoglobulin gene is created by somatic recombination" (1978) *Cell* **15 (1)**, p.1-14

[10] Cooper MD "The early history of B cells" (2015) *Nat Rev Immunol* **15 (3)**, p.191-197

[11] McHeyzer-Williams LJ, McHeyzer-Williams MG "Antigen-specific memory B cell development" (2005) *Annu Rev Immunol* **23**, p.487-513

[12] Nemazee D "Mechanisms of central tolerance for B cells" (2017) *Nat Rev Immunol* **17 (5)**, p.281-294

[13] Janeway CA, Jr. "Approaching the asymptote? Evolution and revolution in immunology" (1989) *Cold Spring Harb Symp Quant Biol* **54 Pt 1**, p.1-13

[14] Venge P, Stromberg A, Braconier JH, Roxin LE, Olsson I "Neutrophil and eosinophil granulocytes in bacterial infection: sequential studies of cellular and serum levels of granule proteins" (1978) *Br J Haematol* **38 (4)**, p.475-483

[15] Mukai K, Tsai M, Starkl P, Marichal T, Galli SJ "IgE and mast cells in host defense against parasites and venoms" (2016) *Semin Immunopathol* **38 (5)**, p.581-603

[16] Banchereau J, Steinman RM "Dendritic cells and the control of immunity" (1998) *Nature* **392 (6673)**, p.245-252

[17] Young JD, Hengartner H, Podack ER, Cohn ZA "Purification and characterization of a cytolytic pore-forming protein from granules of cloned lymphocytes with natural killer activity" (1986) *Cell* **44 (6)**, p.849-859

[18] Thomas DA, Du C, Xu M, Wang X, Ley TJ "DFF45/ICAD can be directly processed by granzyme B during the induction of apoptosis" (2000) *Immunity* **12 (6)**, p.621-632

[19] Lettau M, Schmidt H, Kabelitz D, Janssen O "Secretory lysosomes and their cargo in T and NK cells" (2007) *Immunol Lett* **108 (1)**, p.10-19

[20] Muppidi JR, Lobito AA, Ramaswamy M *et al.* "Homotypic FADD interactions through a conserved RXDLL motif are required for death receptor-induced apoptosis" (2006) *Cell Death Differ* **13 (10)**, p.1641-1650

[21] Chavez-Galan L, Arenas-Del Angel MC, Zenteno E, Chavez R, Lascurain R "Cell death mechanisms induced by cytotoxic lymphocytes" (2009) *Cell Mol Immunol* **6 (1)**, p.15-25

[22] Smyth MJ, Cretney E, Kelly JM *et al.* "Activation of NK cell cytotoxicity" (2005) *Mol Immunol* **42 (4)**, p.501-510

[23] Dupont PJ, Warrens AN "Fas ligand exerts its pro-inflammatory effects via neutrophil recruitment but not activation" (2007) *Immunology* **120 (1)**, p.133-139

[24] Deauvieau F, Ollion V, Doffin AC *et al.* "Human natural killer cells promote cross-presentation of tumor cell-derived antigens by dendritic cells" (2015) *Int J Cancer* **136 (5)**, p.1085-1094

[25] Martin-Fontecha A, Thomsen LL, Brett S *et al.* "Induced recruitment of NK cells to lymph nodes provides IFN-gamma for T(H)1 priming" (2004) *Nat Immunol* **5 (12)**, p.1260-1265

[26] Pallmer K, Oxenius A "Recognition and Regulation of T Cells by NK Cells" (2016) *Front Immunol* **7**, p.251

[27] Wang W, Erbe AK, Hank JA, Morris ZS, Sondel PM "NK Cell-Mediated Antibody-Dependent Cellular Cytotoxicity in Cancer Immunotherapy" (2015) *Front Immunol* **6**, p.368

[28] Ljunggren HG, Karre K "In search of the 'missing self': MHC molecules and NK cell recognition" (1990) *Immunol Today* **11 (7)**, p.237-244

[29] Cohen GB, Gandhi RT, Davis DM *et al.* "The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells" (1999) *Immunity* **10 (6)**, p.661-671

[30] Bubenik J "Tumour MHC class I downregulation and immunotherapy (Review)" (2003) *Oncol Rep* **10 (6)**, p.2005-2008

[31] Gonzalez-Galarza FF, Takeshita LY, Santos EJ *et al.* "Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations" (2015) *Nucleic Acids Res* **43 (Database issue)**, p.D784-788

[32] Kim S, Poursine-Laurent J, Truscott SM *et al.* "Licensing of natural killer cells by host major histocompatibility complex class I molecules" (2005) *Nature* **436 (7051)**, p.709-713

[33] Anfossi N, Andre P, Guia S *et al.* "Human NK cell education by inhibitory receptors for MHC class I" (2006) *Immunity* **25 (2)**, p.331-342

[34] Joncker NT, Shifrin N, Delebecque F, Raulet DH "Mature natural killer cells reset their responsiveness when exposed to an altered MHC environment" (2010) *J Exp Med* **207 (10)**, p.2065-2072

[35] Elliott JM, Wahle JA, Yokoyama WM "MHC class I-deficient natural killer cells acquire a licensed phenotype after transfer into an MHC class I-sufficient environment" (2010) *J Exp Med* **207 (10)**, p.2073-2079

[36] Bessoles S, Angelov GS, Back J, Leclercq G, Vivier E, Held W "Education of murine NK cells requires both cis and trans recognition of MHC class I molecules" (2013) *J Immunol* **191 (10)**, p.5044-5051

[37] Li H, Ivarsson MA, Walker-Sperling VE *et al.* "Identification of an elaborate NK-specific system regulating HLA-C expression" (2018) *PLoS Genet* **14 (1)**, p.e1007163

[38] Fang F, Xiao W, Tian Z "NK cell-based immunotherapy for cancer" (2017) *Semin Immunol* **31**, p.37-54

[39] North J, Bakhsh I, Marden C *et al.* "Tumor-primed human natural killer cells lyse NK-resistant tumor targets: evidence of a two-stage process in resting NK cell activation" (2007) *J Immunol* **178 (1)**, p.85-94

[40] Bryceson YT, March ME, Ljunggren HG, Long EO "Synergy among receptors on resting NK cells for the activation of natural cytotoxicity and cytokine secretion" (2006) *Blood* **107 (1)**, p.159-166

[41] Foley B, Cooley S, Verneris MR *et al.* "Cytomegalovirus reactivation after allogeneic transplantation promotes a lasting increase in educated NKG2C+ natural killer cells with potent function" (2012) *Blood* **119 (11)**, p.2665-2674

[42] O'Sullivan TE, Sun JC, Lanier LL "Natural Killer Cell Memory" (2015) *Immunity* **43 (4)**, p.634-645

[43] Reeves RK, Li H, Jost S *et al.* "Antigen-specific NK cell memory in rhesus macaques" (2015) *Nat Immunol* **16 (9)**, p.927-932

[44] Horton R, Wilming L, Rand V *et al.* "Gene map of the extended human MHC" (2004) *Nat Rev Genet* **5 (12)**, p.889-899

[45] Horton R, Gibson R, Coggill P *et al.* "Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project" (2008) *Immunogenetics* **60 (1)**, p.1-18

[46] Trolle T, McMurtrey CP, Sidney J *et al.* "The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference" (2016) *J Immunol* **196 (4)**, p.1480-1487

[47] Haworth KB, Leddon JL, Chen CY, Horwitz EM, Mackall CL, Cripe TP "Going back to class I: MHC and immunotherapies for childhood cancer" (2015) *Pediatr Blood Cancer* **62 (4)**, p.571-576

[48] Bjorkman PJ, Strominger JL, Wiley DC "Crystallization and X-ray diffraction studies on the histocompatibility antigens HLA-A2 and HLA-A28 from human cell membranes" (1985) *J Mol Biol* **186 (1)**, p.205-210

[49] Strominger JL "Structure of class I and class II HLA antigens" (1987) *Br Med Bull* **43 (1)**, p.81-93

[50] Greinix HT, Fae I, Schneider B *et al.* "Impact of HLA class I high-resolution mismatches on chronic graft-versus-host disease and survival of patients given hematopoietic stem cell grafts from unrelated donors" (2005) *Bone Marrow Transplant* **35 (1)**, p.57-62

[51] Horowitz A, Djaoud Z, Nemat-Gorgani N *et al.* "Class I HLA haplotypes form two schools that educate NK cells in different ways" (2016) *Sci Immunol* **1 (3)**, p.eaag1672

[52] Garcia-Beltran WF, Holzemer A, Martrus G *et al.* "Open conformers of HLA-F are high-affinity ligands of the activating NK-cell receptor KIR3DS1" (2016) *Nat Immunol* **17 (9)**, p.1067-1074

[53] Burian A, Wang KL, Finton KA *et al.* "HLA-F and MHC-I Open Conformers Bind Natural Killer Cell Ig-Like Receptor KIR3DS1" (2016) *PLoS One* **11 (9)**, p.e0163297

[54] Kiani Z, Dupuy FP, Bruneau J *et al.* "HLA-F on HLA-Null 721.221 Cells Activates Primary NK Cells Expressing the Activating Killer Ig-like Receptor KIR3DS1" (2018) *J Immunol* **201 (1)**, p.113-123

[55] Rajagopalan S, Long EO "KIR2DL4 (CD158d): An activation receptor for HLA-G" (2012) *Front Immunol* **3**, p.258

[56] Rajagopalan S "HLA-G-mediated NK cell senescence promotes vascular remodeling: implications for reproduction" (2014) *Cell Mol Immunol* **11 (5)**, p.460-466

[57] Cox ST, Danby R, Hernandez D, Madrigal JA, Saudemont A "Functional Characterisation and Analysis of the Soluble NKG2D Ligand Repertoire Detected in Umbilical Cord Blood Plasma" (2018) *Front Immunol* **9**, p.1282

[58] Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE "The IPD and IMGT/HLA database: allele variant databases" (2015) *Nucleic Acids Res* **43 (Database issue)**, p.D423-431

[59] Habets T, Hepkema BG, Kouprie N *et al.* "The prevalence of antibodies against the HLA-DRB3 protein in kidney transplantation and the correlation with HLA expression" (2018) *PLoS One* **13 (9)**, p.e0203381

[60] Brown JH, Jardetzky TS, Gorga JC *et al.* "Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1" (1993) *Nature* **364 (6432)**, p.33-39

[61] O'Brien C, Flower DR, Feighery C "Peptide length significantly influences in vitro affinity for MHC class II molecules" (2008) *Immunome Res* **4**, p.6

[62] Beck S, Geraghty D, Inoko H, Rowen L, Trowsdale J, Campbell D "Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium" (1999) *Nature* **401 (6756)**, p.921-923

[63] Robinson J, Guethlein LA, Cereb N *et al.* "Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles" (2017) *PLoS Genet* **13 (6)**, p.e1006862

[64] Parham P, Ohta T "Population biology of antigen presentation by MHC class I molecules" (1996) *Science* **272 (5258)**, p.67-74

[65] Apps R, Qi Y, Carlson JM *et al.* "Influence of HLA-C expression level on HIV control" (2013) *Science* **340 (6128)**, p.87-91

[66] Herzenberg LA, Cole LJ "Presence of Donor Specific Gamma-Globulins in Sera of Allogeneic Mouse Radiation Chimeras" (1964) *Nature* **202**, p.352-353

[67] Carrillo-Bustamante P, Kesmir C, de Boer RJ "The evolution of natural killer cell receptors" (2016) *Immunogenetics* **68 (1)**, p.3-18

[68] Biassoni R, Malnati MS "Human Natural Killer Receptors, Co-Receptors, and Their Ligands" (2018) *Curr Protoc Immunol* **121 (1)**, p.e47

[69] Borrego F, Masilamani M, Marusina AI, Tang X, Coligan JE "The CD94/NKG2 family of receptors: from molecules and cells to clinical relevance" (2006) *Immunol Res* **35 (3)**, p.263-278

[70] Lazetic S, Chang C, Houchins JP, Lanier LL, Phillips JH "Human natural killer cell receptors involved in MHC class I recognition are disulfide-linked heterodimers of CD94 and NKG2 subunits" (1996) *J Immunol* **157 (11)**, p.4741-4745

[71] Barrow AD, Trowsdale J "The extended human leukocyte receptor complex: diverse ways of modulating immune responses" (2008) *Immunol Rev* **224**, p.98-123

[72] Horton R, Coggill P, Miretti MM *et al.* "The LRC haplotype project: a resource for killer immunoglobulin-like receptor-linked association studies" (2006) *Tissue Antigens* **68 (5)**, p.450-452

[73] Kelley J, Walter L, Trowsdale J "Comparative genomics of natural killer cell receptor gene clusters" (2005) *PLoS Genet* **1 (2)**, p.129-139

[74] Kim T, Vidal GS, Djurisic M *et al.* "Human LilrB2 is a beta-amyloid receptor and its murine homolog PirB regulates synaptic plasticity in an Alzheimer's model" (2013) *Science* **341 (6152)**, p.1399-1404

[75] Hirayasu K, Arase H "Functional and genetic diversity of leukocyte immunoglobulin-like receptor and implication for disease associations" (2015) *J Hum Genet* **60 (11)**, p.703-708

[76] Lebbink RJ, van den Berg MC, de Ruiter T *et al.* "The soluble leukocyte-associated Ig-like receptor (LAIR)-2 antagonizes the collagen/LAIR-1 inhibitory immune interaction" (2008) *J Immunol* **180 (3)**, p.1662-1669

[77] Verheyden S, Bernier M, Demanet C "Identification of natural killer cell receptor phenotypes associated with leukemia" (2004) *Leukemia* **18 (12)**, p.2002-2007

[78] Zhang Y, Wang B, Ye S *et al.* "Killer cell immunoglobulin-like receptor gene polymorphisms in patients with leukemia: possible association with susceptibility to the disease" (2010) *Leuk Res* **34 (1)**, p.55-58

[79] Martin MP, Naranbhai V, Shea PR *et al.* "Killer cell immunoglobulin-like receptor 3DL1 variation modifies HLA-B*57 protection against HIV-1" (2018) *J Clin Invest*,

[80] Kusnierczyk P, Mozer-Lisewska I, Zwolinska K *et al.* "Contribution of genes for killer cell immunoglobulin-like receptors (KIR) to the susceptibility to chronic hepatitis C virus infection and to viremia" (2015) *Hum Immunol* **76 (2-3)**, p.102-108

[81] Hiby SE, Walker JJ, O'Shaughnessy K M *et al.* "Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success" (2004) *J Exp Med* **200 (8)**, p.957-965

[82] Kennedy PR, Chazara O, Gardner L *et al.* "Activating KIR2DS4 Is Expressed by Uterine NK Cells and Contributes to Successful Pregnancy" (2016) *J Immunol* **197 (11)**, p.4292-4300

[83] Cooley S, Trachtenberg E, Bergemann TL *et al.* "Donors with group B KIR haplotypes improve relapse-free survival after unrelated hematopoietic cell transplantation for acute myelogenous leukemia" (2009) *Blood* **113 (3)**, p.726-732

[84] Sobecks RM, Gallagher MM, Askar M *et al.* "Influence Of Killer Immunoglobulin-Like Receptor (KIR) and HLA Genotypes On Outcomes After Reduced-Intensity Conditioning Allogeneic Hematopoietic Stem Cell Transplantation For Patients With AML and MDS: A Report From The Center For International Blood and Marrow Transplant Research Immunobiology Working Committee" (2013) *Blood* **122 (21)**, p.159

[85] Marsh SGE, Parham P, Dupont B *et al.* "Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002" (2003) *Tissue Antigens* **62 (1)**, p.79-86

[86] Vierra-Green C, Roe D, Jayaraman J *et al.* "Estimating KIR Haplotype Frequencies on a Cohort of 10,000 Individuals: A Comprehensive Study on Population Variations, Typing Resolutions, and Reference Haplotypes" (2016) *PLoS One* **11 (10)**, p.e0163973

[87] Pyo CW, Guethlein LA, Vu Q *et al.* "Different patterns of evolution in the centromeric and telomeric regions of group A and B haplotypes of the human killer cell Ig-like receptor locus" (2010) *PLoS One* **5 (12)**, p.e15115

[88] Roe D, Vierra-Green C, Pyo CW *et al.* "Revealing complete complex KIR haplotypes phased by long-read sequencing technology" (2017) *Genes Immun* **18 (3)**, p.127-134

[89] Vendelbosch S, de Boer M, van Leeuwen K *et al.* "Novel insights in the genomic organization and hotspots of recombination in the human KIR locus through analysis of intergenic regions" (2015) *Genes Immun* **16 (2)**, p.103-111

[90] Jiang W, Johnson C, Jayaraman J *et al.* "Copy number variation leads to considerable diversity for B but not A haplotypes of the human KIR genes encoding NK cell receptors" (2012) *Genome Res* **22 (10)**, p.1845-1854

[91] Jiang W, Johnson C, Simecek N *et al.* "qKAT: a high-throughput qPCR method for KIR gene copy number and haplotype determination" (2016) *Genome Med* **8 (1)**, p.99

[92] Roberts Ch, Jiang W, Jayaraman J, Trowsdale J, Holland MJ, Traherne JA "Killer-cell Immunoglobulin-like Receptor gene linkage and copy number variation analysis by droplet digital PCR" (2014) *Genome Med* **6 (3)**, p.20

[93] Gomez-Lozano N, Gardiner CM, Parham P, Vilches C "Some human KIR haplotypes contain two KIR2DL5 genes: KIR2DL5A and KIR2DL5B" (2002) *Immunogenetics* **54 (5)**, p.314-319

[94] Norman PJ, Hollenbach JA, Nemat-Gorgani N *et al.* "Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing" (2016) *Am J Hum Genet* **99 (2)**, p.375-391

[95] Pyo CW, Wang R, Vu Q *et al.* "Recombinant structures expand and contract inter and intragenic diversification at the KIR locus" (2013) *BMC Genomics* **14**, p.89

[96] Hans JB, Vigilant L "Discovery of gorilla MHC-C expressing C1 ligand for KIR" (2018) *Immunogenetics* **70 (5)**, p.293-304

[97] Guethlein LA, Norman PJ, Hilton HG, Parham P "Co-evolution of MHC class I and variable NK cell receptors in placental mammals" (2015) *Immunol Rev* **267 (1)**, p.259-282

[98] Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SGE "IPD--the Immuno Polymorphism Database" (2013) *Nucleic Acids Res* **41 (Database issue)**, p.D1234-1240

[99] Vilches C, Pando MJ, Parham P "Genes encoding human killer-cell Ig-like receptors with D1 and D2 extracellular domains all contain untranslated pseudoexons encoding a third Ig-like domain" (2000) *Immunogenetics* **51 (8-9)**, p.639-646

[100] Trundley AE, Hiby SE, Chang C *et al.* "Molecular characterization of KIR3DL3" (2006) *Immunogenetics* **57 (12)**, p.904-916

[101] Biassoni R, Pessino A, Malaspina A *et al.* "Role of amino acid position 70 in the binding affinity of p50.1 and p58.1 receptors for HLA-Cw4 molecules" (1997) *Eur J Immunol* **27 (12)**, p.3095-3099

[102] Stewart CA, Laugier-Anfossi F, Vely F *et al.* "Recognition of peptide-MHC class I complexes by activating killer immunoglobulin-like receptors" (2005) *Proc Natl Acad Sci U S A* **102 (37)**, p.13224-13229

[103] Colonna M, Brooks EG, Falco M, Ferrara GB, Strominger JL "Generation of allospecific natural killer cells by stimulation across a polymorphism of HLA-C" (1993) *Science* **260 (5111)**, p.1121-1124

[104] Colonna M, Borsellino G, Falco M, Ferrara GB, Strominger JL "HLA-C is the inhibitory ligand that determines dominant resistance to lysis by NK1- and NK2-specific natural killer cells" (1993) *Proc Natl Acad Sci U S A* **90 (24)**, p.12000-12004

[105] Moesta AK, Norman PJ, Yawata M, Yawata N, Gleimer M, Parham P "Synergistic polymorphism at two positions distal to the ligand-binding site makes KIR2DL2 a stronger receptor for HLA-C than KIR2DL3" (2008) *J Immunol* **180 (6)**, p.3969-3979

[106] Hilton HG, Vago L, Older Aguilar AM *et al.* "Mutation at positively selected positions in the binding site for HLA-C shows that KIR2DL1 is a more refined but less adaptable NK cell receptor than KIR2DL3" (2012) *J Immunol* **189 (3)**, p.1418-1430

[107] Cella M, Longo A, Ferrara GB, Strominger JL, Colonna M "NK3-specific natural killer cells are selectively inhibited by Bw4-positive HLA alleles with isoleucine 80" (1994) *J Exp Med* **180 (4)**, p.1235-1242

[108] Thananchai H, Gillespie G, Martin MP *et al.* "Cutting Edge: Allele-specific and peptide-dependent interactions between KIR3DL1 and HLA-A and HLA-B" (2007) *J Immunol* **178 (1)**, p.33-37

[109] Erbe AK, Wang W, Reville PK *et al.* "HLA-Bw4-I-80 Isoform Differentially Influences Clinical Outcome As Compared to HLA-Bw4-T-80 and HLA-A-Bw4 Isoforms in Rituximab or Dinutuximab-Based Cancer Immunotherapy" (2017) *Front Immunol* **8**, p.675

[110] Gumperz JE, Litwin V, Phillips JH, Lanier LL, Parham P "The Bw4 public epitope of HLA-B molecules confers reactivity with natural killer cell clones that express NKB1, a putative HLA receptor" (1995) *J Exp Med* **181 (3)**, p.1133-1144

[111] O'Connor GM, Vivian JP, Gostick E *et al.* "Peptide-Dependent Recognition of HLA-B*57:01 by KIR3DS1" (2015) *J Virol* **89 (10)**, p.5213-5221

[112] O'Connor GM, Yamada E, Rampersaud A, Thomas R, Carrington M, McVicar DW "Analysis of binding of KIR3DS1*014 to HLA suggests distinct evolutionary history of KIR3DS1" (2011) *J Immunol* **187 (5)**, p.2162-2171

[113] Vales-Gomez M, Reyburn HT, Erskine RA, Strominger J "Differential binding to HLA-C of p50-activating and p58-inhibitory natural killer cell receptors" (1998) *Proc Natl Acad Sci U S A* **95 (24)**, p.14326-14331

[114] Winter CC, Gumperz JE, Parham P, Long EO, Wagtmann N "Direct binding and functional transfer of NK cell inhibitory receptors reveal novel patterns of HLA-C allotype recognition" (1998) *J Immunol* **161 (2)**, p.571-577

[115] Thiruchelvam-Kyle L, Hoelsbrekken SE, Saether PC *et al.* "The Activating Human NK Cell Receptor KIR2DS2 Recognizes a beta2-Microglobulin-Independent Ligand on Cancer Cells" (2017) *J Immunol* **198 (7)**, p.2556-2567

[116] Liu J, Xiao Z, Ko HL, Shen M, Ren EC "Activating killer cell immunoglobulin-like receptor 2DS2 binds to HLA-A*11" (2014) *Proc Natl Acad Sci U S A* **111 (7)**, p.2662-2667

[117] Graef T, Moesta AK, Norman PJ *et al.* "KIR2DS4 is a product of gene conversion with KIR3DL2 that introduced specificity for HLA-A*11 while diminishing avidity for HLA-C" (2009) *J Exp Med* **206 (11)**, p.2557-2572

[118] Blokhuis JH, Hilton HG, Guethlein LA *et al.* "KIR2DS5 allotypes that recognize the C2 epitope of HLA-C are common among Africans and absent from Europeans" (2017) *Immun Inflamm Dis* **5 (4)**, p.461-468

[119] Wong-Baeza I, Ridley A, Shaw J *et al.* "KIR3DL2 binds to HLA-B27 dimers and free H chains more strongly than other HLA class I and promotes the expansion of T cells in ankylosing spondylitis" (2013) *J Immunol* **190 (7)**, p.3216-3224

[120] Pende D, Biassoni R, Cantoni C *et al.* "The natural killer cell receptor specific for HLA-A allotypes: a novel member of the p58/p70 family of inhibitory receptors that is characterized by three immunoglobulin-like domains and is expressed as a 140-kD disulphide-linked dimer" (1996) *J Exp Med* **184 (2)**, p.505-518

[121] Hansasuta P, Dong T, Thananchai H *et al.* "Recognition of HLA-A3 and HLA-A11 by KIR3DL2 is peptide-specific" (2004) *Eur J Immunol* **34 (6)**, p.1673-1679

[122] Ponte M, Cantoni C, Biassoni R *et al.* "Inhibitory receptors sensing HLA-G1 molecules in pregnancy: decidua-associated natural killer cells express LIR-1 and CD94/NKG2A and acquire p49, an HLA-G1-specific receptor" (1999) *Proc Natl Acad Sci U S A* **96 (10)**, p.5674-5679

[123] Rajagopalan S, Bryceson YT, Kuppusamy SP *et al.* "Activation of NK cells by an endocytosed receptor for soluble HLA-G" (2006) *PLoS Biol* **4 (1)**, p.e9

[124] Brodin P, Karre K, Hoglund P "NK cell education: not an on-off switch but a tunable rheostat" (2009) *Trends Immunol* **30 (4)**, p.143-149

[125] Sim MJ, Stowell J, Sergeant R, Altmann DM, Long EO, Boyton RJ "KIR2DL3 and KIR2DL1 show similar impact on licensing of human NK cells" (2016) *Eur J Immunol* **46 (1)**, p.185-191

[126] Nowak J, Koscinska K, Mika-Witkowska R *et al.* "Role of Donor Activating KIR-HLA Ligand-Mediated NK Cell Education Status in Control of Malignancy in Hematopoietic Cell Transplant Recipients" (2015) *Biol Blood Marrow Transplant* **21 (5)**, p.829-839

[127] Fernandez NC, Treiner E, Vance RE, Jamieson AM, Lemieux S, Raulet DH "A subset of natural killer cells achieves self-tolerance without expressing inhibitory receptors specific for self-MHC molecules" (2005) *Blood* **105 (11)**, p.4416-4423

[128] Freund J, May RM, Yang E *et al.* "Activating Receptor Signals Drive Receptor Diversity in Developing Natural Killer Cells" (2016) *PLoS Biol* **14 (8)**, p.e1002526

[129] Garrido F, Aptsiauri N, Doorduijn EM, Garcia Lora AM, van Hall T "The urgent need to recover MHC class I in cancers for effective immunotherapy" (2016) *Curr Opin Immunol* **39**, p.44-51

[130] Schwartz O, Marechal V, Le Gall S, Lemonnier F, Heard JM "Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein" (1996) *Nat Med* **2 (3)**, p.338-342

[131] Wilkinson GW, Tomasec P, Stanton RJ *et al.* "Modulation of natural killer cells by human cytomegalovirus" (2008) *J Clin Virol* **41 (3)**, p.206-212

[132] Oliveira LM, Portela P, Merzoni J *et al.* "Reduced frequency of two activating KIR genes in patients with sepsis" (2017) *Hum Immunol* **78 (4)**, p.363-369

[133] Tomblyn M, Young JA, Haagenson MD *et al.* "Decreased infections in recipients of unrelated donor hematopoietic cell transplantation from donors with an activating KIR genotype" (2010) *Biol Blood Marrow Transplant* **16 (8)**, p.1155-1161

[134] Moffett A, Loke C "Immunology of placentation in eutherian mammals" (2006) *Nat Rev Immunol* **6 (8)**, p.584-594

[135] Hiby SE, Regan L, Lo W, Farrell L, Carrington M, Moffett A "Association of maternal killer-cell immunoglobulin-like receptors and parental HLA-C genotypes with recurrent miscarriage" (2008) *Hum Reprod* **23 (4)**, p.972-976

[136] Hiby SE, Apps R, Chazara O *et al.* "Maternal KIR in combination with paternal HLA-C2 regulate human birth weight" (2014) *J Immunol* **192 (11)**, p.5069-5073

[137] Penman BS, Moffett A, Chazara O, Gupta S, Parham P "Reproduction, infection and killer-cell immunoglobulin-like receptor haplotype evolution" (2016) *Immunogenetics* **68 (10)**, p.755-764

[138] Nakimuli A, Chazara O, Hiby SE *et al.* "A KIR B centromeric region present in Africans but not Europeans protects pregnant women from pre-eclampsia" (2015) *Proc Natl Acad Sci U S A* **112 (3)**, p.845-850

[139] Oszmiana A, Williamson DJ, Cordoba SP *et al.* "The Size of Activating and Inhibitory Killer Ig-like Receptor Nanoclusters Is Controlled by the Transmembrane Sequence and Affects Signaling" (2016) *Cell Rep* **15 (9)**, p.1957-1972

[140] Kumar S, Rajagopalan S, Sarkar P *et al.* "Zinc-Induced Polymerization of Killer-Cell Ig-like Receptor into Filaments Promotes Its Inhibitory Function at Cytotoxic Immunological Synapses" (2016) *Mol Cell* **62 (1)**, p.21-33

[141] Faure M, Barber DF, Takahashi SM, Jin T, Long EO "Spontaneous clustering and tyrosine phosphorylation of NK cell inhibitory receptor induced by ligand binding" (2003) *J Immunol* **170 (12)**, p.6107-6114

[142] Burshtyn DN, Lam AS, Weston M, Gupta N, Warmerdam PA, Long EO "Conserved residues amino-terminal of cytoplasmic tyrosines contribute to the SHP-1-mediated inhibitory function of killer cell Ig-like receptors" (1999) *J Immunol* **162 (2)**, p.897-902

[143] Yusa S, Catina TL, Campbell KS "SHP-1- and phosphotyrosine-independent inhibitory signaling by a killer cell Ig-like receptor cytoplasmic domain in human NK cells" (2002) *J Immunol* **168 (10)**, p.5047-5057

[144] Purdy AK, Campbell KS "SHP-2 expression negatively regulates NK cell function" (2009) *J Immunol* **183 (11)**, p.7234-7243

[145] Stebbins CC, Watzl C, Billadeau DD, Leibson PJ, Burshtyn DN, Long EO "Vav1 dephosphorylation by the tyrosine phosphatase SHP-1 as a mechanism for inhibition of cellular cytotoxicity" (2003) *Mol Cell Biol* **23 (17)**, p.6291-6299

[146] Lanier LL "Up on the tightrope: natural killer cell activation and inhibition" (2008) *Nat Immunol* **9 (5)**, p.495-502

[147] Kikuchi-Maki A, Catina TL, Campbell KS "Cutting edge: KIR2DL4 transduces signals into human NK cells through association with the Fc receptor gamma protein" (2005) *J Immunol* **174 (7)**, p.3859-3863

[148] Faure M, Long EO "KIR2DL4 (CD158d), an NK cell-activating receptor with inhibitory potential" (2002) *J Immunol* **168 (12)**, p.6208-6214

[149] Horowitz A, Strauss-Albee DM, Leipold M *et al.* "Genetic and environmental determinants of human NK cell diversity revealed by mass cytometry" (2013) *Sci Transl Med* **5 (208)**, p.208ra145

[150] Manser AR, Weinhold S, Uhrberg M "Human KIR repertoires: shaped by genetic diversity and evolution" (2015) *Immunol Rev* **267 (1)**, p.178-196

[151] Zhao X, Weinhold S, Brands J *et al.* "NK cell development in a human stem cell niche: KIR expression occurs independently of the presence of HLA class I ligands" (2018) *Blood Adv* **2 (19)**, p.2452-2461

[152] Carr WH, Pando MJ, Parham P "KIR3DL1 polymorphisms that affect NK cell inhibition by HLA-Bw4 ligand" (2005) *J Immunol* **175 (8)**, p.5222-5229

[153] Pando MJ, Gardiner CM, Gleimer M, McQueen KL, Parham P "The protein made from a common allele of KIR3DL1 (3DL1*004) is poorly expressed at cell surfaces due to substitution at positions 86 in Ig domain 0 and 182 in Ig domain 1" (2003) *J Immunol* **171 (12)**, p.6640-6649

[154] Boudreau JE, Le Luduec JB, Hsu KC "Development of a novel multiplex PCR assay to detect functional subtypes of KIR3DL1 alleles" (2014) *PLoS One* **9 (6)**, p.e99543

[155] Thomas R, Yamada E, Alter G *et al.* "Novel KIR3DL1 alleles and their expression levels on NK cells: convergent evolution of KIR3DL1 phenotype variation?" (2008) *J Immunol* **180 (10)**, p.6743-6750

[156] VandenBussche CJ, Dakshanamurthy S, Posch PE, Hurley CK "A single polymorphism disrupts the killer Ig-like receptor 2DL2/2DL3 D1 domain" (2006) *J Immunol* **177 (8)**, p.5347-5357

[157] Cisneros E, Estefania E, Vilches C "Allelic Polymorphism Determines Surface Expression or Intracellular Retention of the Human NK Cell Receptor KIR2DL5A (CD158f)" (2016) *Front Immunol* **7**, p.698

[158] Steiner NK, Dakshanamurthy S, VandenBussche CJ, Hurley CK "Extracellular domain alterations impact surface expression of stimulatory natural killer cell receptor KIR2DS5" (2008) *Immunogenetics* **60 (11)**, p.655-667

[159] Goodridge JP, Witt CS, Christiansen FT, Warren HS "KIR2DL4 (CD158d) genotype influences expression and function in NK cells" (2003) *J Immunol* **171 (4)**, p.1768-1774

[160] Witt CS, Martin A, Christiansen FT "Detection of KIR2DL4 alleles by sequencing and SSCP reveals a common allele with a shortened cytoplasmic tail" (2000) *Tissue Antigens* **56 (3)**, p.248-257

[161] Crum KA, Logue SE, Curran MD, Middleton D "Development of a PCR-SSOP approach capable of defining the natural killer cell inhibitory receptor (KIR) gene sequence repertoires" (2000) *Tissue Antigens* **56 (4)**, p.313-326

[162] Wu X, Yao Y, Bao X *et al.* "KIR2DS4 and Its Variant KIR1D Are Associated with Acute Graft-versus-Host Disease, Cytomegalovirus, and Overall Survival after Sibling-Related HLA-Matched Transplantation in Patients with Donors with KIR Gene Haplotype A" (2016) *Biol Blood Marrow Transplant* **22 (2)**, p.220-225

[163] Robinson J, Waller MJ, Stoehr P, Marsh SGE "IPD--the Immuno Polymorphism Database" (2005) *Nucleic Acids Res* **33 (Database issue)**, p.D523-526

[164] Gedil MA, Steiner NK, Hurley CK "Genomic characterization of KIR2DL4 in families and unrelated individuals reveals extensive diversity in exon and intron sequences including a common frameshift variation occurring in several alleles" (2005) *Tissue Antigens* **65 (5)**, p.402-418

[165] Vierra-Green C, Roe D, Hou L *et al.* "Allele-level haplotype frequencies and pairwise linkage disequilibrium for 14 KIR loci in 506 European-American individuals" (2012) *PLoS One* **7 (11)**, p.e47491

[166] Li H, Wright PW, McCullen M, Anderson SK "Characterization of KIR intermediate promoters reveals four promoter types associated with distinct expression patterns of KIR subtypes" (2016) *Genes Immun* **17 (1)**, p.66-74

[167] Gomez-Lozano N, Trompeter HI, de Pablo R, Estefania E, Uhrberg M, Vilches C "Epigenetic silencing of potentially functional KIR2DL5 alleles: Implications for the acquisition of KIR repertoires by NK cells" (2007) *Eur J Immunol* **37 (7)**, p.1954-1965

[168] Chan HW, Kurago ZB, Stewart CA *et al.* "DNA methylation maintains allele-specific KIR gene expression in human natural killer cells" (2003) *J Exp Med* **197 (2)**, p.245-255

[169] Bari R, Bell T, Leung WH *et al.* "Significant functional heterogeneity among KIR2DL1 alleles and a pivotal role of arginine 245" (2009) *Blood* **114 (25)**, p.5182-5190

[170] Traherne JA, Martin M, Ward R *et al.* "Mechanisms of copy number variation and hybrid gene formation in the KIR immune gene complex" (2010) *Hum Mol Genet* **19 (5)**, p.737-751

[171] Beziat V, Traherne JA, Liu LL *et al.* "Influence of KIR gene copy number on natural killer cell education" (2013) *Blood* **121 (23)**, p.4703-4707

[172] Gomez-Lozano N, Estefania E, Williams F *et al.* "The silent KIR3DP1 gene (CD158c) is transcribed and might encode a secreted receptor in a minority of humans, in whom the KIR3DP1, KIR2DL4 and KIR3DL1/KIR3DS1 genes are duplicated" (2005) *Eur J Immunol* **35 (1)**, p.16-24

[173] Boyington JC, Motyka SA, Schuck P, Brooks AG, Sun PD "Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand" (2000) *Nature* **405 (6786)**, p.537-543

[174] Frazier WR, Steiner N, Hou L, Dakshanamurthy S, Hurley CK "Allelic variation in KIR2DL3 generates a KIR2DL2-like receptor with increased binding to its HLA-C ligand" (2013) *J Immunol* **190 (12)**, p.6198-6208

[175] Sharma D, Bastard K, Guethlein LA *et al.* "Dimorphic motifs in D0 and D1+D2 domains of killer cell Ig-like receptor 3DL1 combine to form receptors with high, moderate, and no avidity for the complex of a peptide derived from HIV and HLA-A*2402" (2009) *J Immunol* **183 (7)**, p.4569-4582

[176] Hilton HG, Blokhuis JH, Guethlein LA, Norman PJ, Parham P "Resurrecting KIR2DP1: A Key Intermediate in the Evolution of Human Inhibitory NK Cell Receptors That Recognize HLA-C" (2017) *J Immunol* **198 (5)**, p.1961-1973

[177] Martin MP, Naranbhai V, Shea PR *et al.* "Killer cell immunoglobulin-like receptor 3DL1 variation modifies HLA-B*57 protection against HIV-1" (2018) *J Clin Invest* **128 (5)**, p.1903-1912

[178] Vilches C, Castano J, Gomez-Lozano N, Estefania E "Facilitation of KIR genotyping by a PCR-SSP method that amplifies short DNA fragments" (2007) *Tissue Antigens* **70 (5)**, p.415-422

[179] Nong T, Saito K, Blair L, Tarsitani C, Lee JH "KIR genotyping by reverse sequence-specific oligonucleotide methodology" (2007) *Tissue Antigens* **69 Suppl 1**, p.92-95

[180] Schofl G, Lang K, Quenzel P *et al.* "2.7 million samples genotyped for HLA by next generation sequencing: lessons learned" (2017) *BMC Genomics* **18 (1)**, p.161

[181] Hou L, Chen M, Steiner N, Kariyawasam K, Ng J, Hurley CK "Killer cell immunoglobulin-like receptors (KIR) typing by DNA sequencing" (2012) *Methods Mol Biol* **882**, p.431-468

[182] Maniangou B, Legrand N, Alizadeh M *et al.* "Killer Immunoglobulin-Like Receptor Allele Determination Using Next-Generation Sequencing Technology" (2017) *Front Immunol* **8**, p.547

[183] Bultitude WP, Mayor NP, Robinson J, Madrigal A, Marsh SGE "Further optimisation of SMRT DNA sequencing for HLA typing in a high throughput clinical laboratory" (2015) *International Journal of Immunogenetics* **42 (5)**, p.389

[184] Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW "A flexible and efficient template format for circular consensus sequencing and SNP detection" (2010) *Nucleic Acids Res* **38 (15)**, p.e159

[185] Larkin J, Henley RY, Jadhav V, Korlach J, Wanunu M "Length-independent DNA packing into nanopore zero-mode waveguides for low-input DNA sequencing" (2017) *Nat Nanotechnol* **12 (12)**, p.1169-1175

[186] Eid J, Fehr A, Gray J *et al.* "Real-time DNA sequencing from single polymerase molecules" (2009) *Science* **323 (5910)**, p.133-138

[187] Eid J, Murphy D, Otto G, Turner S, "System for the mitigation of photodamage in analytical reactions", ed: Google Patents, 2011.

[188] Bowman BN, Marks P, Hepler NL *et al.*, "Long Amplicon Analysis: Highly accurate, full-length, phased, allele-resolved gene sequences from multiplexed SMRT Sequencing data", presented at the Pacific Biosciences European User Group Meeting, 2014.

[189] Ardui S, Ameur A, Vermeesch JR, Hestand MS "Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics" (2018) *Nucleic Acids Res* **46 (5)**, p.2159-2168

[190] de la Morena MT, Gatti RA "A history of bone marrow transplantation" (2011) *Hematol Oncol Clin North Am* **25 (1)**, p.1-15

[191] Gyurkocza B, Sandmaier BM "Conditioning regimens for hematopoietic cell transplantation: one size does not fit all" (2014) *Blood* **124 (3)**, p.344-353

[192] Gratwohl A, Carreras E, "Principles of conditioning", in *The EBMT Handbook on Haematopoietic Stem Cell Transplantation*, J. F. Apperley, E. Carreras, E. Gluckman, and T. Masszi, Eds. 6th ed. Genoa: Forum Service Editore, 2012, p.122-137.

[193] D'Souza A, Fretham C. Current Uses and Outcomes of Hematopoietic Cell Transplantation (HCT): CIBMTR Summary Slides [Online]. Available: http://www.cibmtr.org

[194] Bacigalupo A, Ballen K, Rizzo D *et al.* "Defining the intensity of conditioning regimens: working definitions" (2009) *Biol Blood Marrow Transplant* **15 (12)**, p.1628-1633

[195] Malard F, Labopin M, Cho C *et al.* "Ex vivo and in vivo T cell-depleted allogeneic stem cell transplantation in patients with acute myeloid leukemia in first complete remission resulted in similar overall survival: on behalf of the ALWP of the EBMT and the MSKCC" (2018) *J Hematol Oncol* **11 (1)**, p.127

[196] Bacigalupo A, Oneto R, Lamparelli T *et al.* "Pre-emptive therapy of acute graft-versus-host disease: a pilot study with antithymocyte globulin (ATG)" (2001) *Bone Marrow Transplant* **28 (12)**, p.1093-1096

[197] Mohty M "Mechanisms of action of antithymocyte globulin: T-cell depletion and beyond" (2007) *Leukemia* **21 (7)**, p.1387-1394

[198] Stauch D, Dernier A, Sarmiento Marchese E *et al.* "Targeting of natural killer cells by rabbit antithymocyte globulin and campath-1H: similar effects independent of specificity" (2009) *PLoS One* **4 (3)**, p.e4709

[199] Lowdell MW, Craston R, Ray N, Koh M, Galatowicz G, Prentice HG "The effect of T cell depletion with Campath-1M on immune reconstitution after

chemotherapy and allogeneic bone marrow transplant as treatment for leukaemia" (1998) *Bone Marrow Transplant* **21 (7)**, p.679-686

[200] Cooley S, McCullar V, Wangen R *et al.* "KIR reconstitution is altered by T cells in the graft and correlates with clinical outcomes after unrelated donor transplantation" (2005) *Blood* **106 (13)**, p.4370-4376

[201] Randolph SS, Gooley TA, Warren EH, Appelbaum FR, Riddell SR "Female donors contribute to a selective graft-versus-leukemia effect in male recipients of HLA-matched, related hematopoietic stem cell transplants" (2004) *Blood* **103 (1)**, p.347-352

[202] Kollman C, Howe CW, Anasetti C *et al.* "Donor characteristics as risk factors in recipients after transplantation of bone marrow from unrelated donors: the effect of donor age" (2001) *Blood* **98 (7)**, p.2043-2051

[203] Kimura F, Sato K, Kobayashi S *et al.* "Impact of AB0-blood group incompatibility on the outcome of recipients of bone marrow transplants from unrelated donors in the Japan Marrow Donor Program" (2008) *Haematologica* **93 (11)**, p.1686-1693

[204] Collins NH, Gee AP, Durett AG *et al.* "The effect of the composition of unrelated donor bone marrow and peripheral blood progenitor cell grafts on transplantation outcomes" (2010) *Biol Blood Marrow Transplant* **16 (2)**, p.253-262

[205] Lee SJ, Klein J, Haagenson M *et al.* "High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation" (2007) *Blood* **110 (13)**, p.4576-4583

[206] Kollman C, Spellman SR, Zhang MJ *et al.* "The effect of donor characteristics on survival after unrelated donor transplantation for hematologic malignancy" (2016) *Blood* **127 (2)**, p.260-267

[207] Boeckh M, Nichols WG "The impact of cytomegalovirus serostatus of donor and recipient before hematopoietic stem cell transplantation in the era of antiviral prophylaxis and preemptive therapy" (2004) *Blood* **103 (6)**, p.2003-2008

[208] Thomas ED, Lochte HL, Jr., Lu WC, Ferrebee JW "Intravenous infusion of bone marrow in patients receiving radiation and chemotherapy" (1957) *N Engl J Med* **257 (11)**, p.491-496

[209] Thomas ED, Lochte HL, Jr., Cannon JH, Sahler OD, Ferrebee JW "Supralethal whole body irradiation and isologous marrow transplantation in man" (1959) *J Clin Invest* **38**, p.1709-1716

[210] Mathe G, Amiel JL, Schwarzenberg L *et al.* "Successful Allogenic Bone Marrow Transplantation in Man: Chimerism, Induced Specific Tolerance and Possible Anti-Leukemic Effects" (1965) *Blood* **25**, p.179-196

[211] Meuwissen HJ, Gatti RA, Terasaki PI, Hong R, Good RA "Treatment of lymphopenic hypogammaglobulinemia and bone-marrow aplasia by

transplantation of allogeneic marrow. Crucial role of histocompatiility matching" (1969) *N Engl J Med* **281 (13)**, p.691-697

[212] Thomas ED, Buckner CD, Banaji M *et al.* "One hundred patients with acute leukemia treated by chemotherapy, total body irradiation, and allogeneic marrow transplantation" (1977) *Blood* **49 (4)**, p.511-533

[213] Saber W, Opie S, Rizzo JD, Zhang MJ, Horowitz MM, Schriber J "Outcomes after matched unrelated donor versus identical sibling hematopoietic cell transplantation in adults with acute myelogenous leukemia" (2012) *Blood* **119 (17)**, p.3908-3916

[214] Shaw BE, Arguello R, Garcia-Sepulveda CA, Madrigal JA "The impact of HLA genotyping on survival following unrelated donor haematopoietic stem cell transplantation" (2010) *Br J Haematol* **150 (3)**, p.251-258

[215] Morishima Y, Kashiwase K, Matsuo K *et al.* "Biological significance of HLA locus matching in unrelated donor bone marrow transplantation" (2015) *Blood* **125 (7)**, p.1189-1197

[216] Mayor NP, Hayhurst JD, Turner TR *et al.* "Recipients Receiving Better HLA-Matched Hematopoietic Cell Transplantation Grafts, Uncovered by a Novel HLA Typing Method, Have Superior Survival: A Retrospective Study" (2019) *Biol Blood Marrow Transplant* **25 (3)**, p.443-450

[217] Shaw BE, Marsh SGE, Mayor NP, Russell NH, Madrigal JA "HLA-DPB1 matching status has significant implications for recipients of unrelated donor stem cell transplants" (2006) *Blood* **107 (3)**, p.1220-1226

[218] Shaw BE, Gooley TA, Malkki M *et al.* "The importance of HLA-DPB1 in unrelated donor hematopoietic cell transplantation" (2007) *Blood* **110 (13)**, p.4560-4566

[219] Rutten CE, van Luxemburg-Heijs SA, Halkes CJ *et al.* "Patient HLA-DP-specific CD4+ T cells from HLA-DPB1-mismatched donor lymphocyte infusion can induce graft-versus-leukemia reactivity in the presence or absence of graft-versus-host disease" (2013) *Biol Blood Marrow Transplant* **19 (1)**, p.40-48

[220] Crivello P, Heinold A, Rebmann V *et al.* "Functional distance between recipient and donor HLA-DPB1 determines nonpermissive mismatches in unrelated HCT" (2016) *Blood* **128 (1)**, p.120-129

[221] Fleischhauer K, Shaw BE, Gooley T *et al.* "Effect of T-cell-epitope matching at HLA-DPB1 in recipients of unrelated-donor haemopoietic-cell transplantation: a retrospective study" (2012) *Lancet Oncol* **13 (4)**, p.366-374

[222] Fernandez-Vina MA, Wang T, Lee SJ *et al.* "Identification of a permissible HLA mismatch in hematopoietic stem cell transplantation" (2014) *Blood* **123 (8)**, p.1270-1278

[223] Petersdorf EW, Gooley TA, Malkki M *et al.* "HLA-C expression levels define permissible mismatches in hematopoietic cell transplantation" (2014) *Blood* **124 (26)**, p.3996-4003

[224] Rocha V, Gluckman E, Registry E-N, Group EBaMT "Improving outcomes of cord blood transplantation: HLA matching, cell dose and other graft- and transplantation-related factors" (2009) *Br J Haematol* **147 (2)**, p.262-274

[225] Besse K, Maiers M, Confer D, Albrecht M "On Modeling Human Leukocyte Antigen-Identical Sibling Match Probability for Allogeneic Hematopoietic Cell Transplantation: Estimating the Need for an Unrelated Donor Source" (2016) *Biol Blood Marrow Transplant* **22 (3)**, p.410-417

[226] Kernan NA, Flomenberg N, Dupont B, O'Reilly RJ "Graft rejection in recipients of T-cell-depleted HLA-nonidentical marrow transplants for leukemia. Identification of host-derived antidonor allocytotoxic T lymphocytes" (1987) *Transplantation* **43 (6)**, p.842-847

[227] Anasetti C, Beatty PG, Storb R *et al.* "Effect of HLA incompatibility on graft-versus-host disease, relapse, and survival after marrow transplantation for patients with leukemia or lymphoma" (1990) *Hum Immunol* **29 (2)**, p.79-91

[228] McCurdy SR, Zhang MJ, St Martin A *et al.* "Effect of donor characteristics on haploidentical transplantation with posttransplantation cyclophosphamide" (2018) *Blood Adv* **2 (3)**, p.299-307

[229] Lorentino F, Labopin M, Fleischhauer K *et al.* "The impact of HLA matching on outcomes of unmanipulated haploidentical HSCT is modulated by GVHD prophylaxis" (2017) *Blood Adv* **1 (11)**, p.669-680

[230] WMDA. (2019, 20/08/2019). *Total Number of Donors and Cord blood units*. Available: https://statistics.wmda.info/

[231] Nolan A. (2018, 20/11/2018). *Facts and Stats*. Available: https://www.anthonynolan.org/facts-and-stats

[232] Heemskerk MB, van Walraven SM, Cornelissen JJ *et al.* "How to improve the search for an unrelated haematopoietic stem cell donor. Faster is better than more!" (2005) *Bone Marrow Transplant* **35 (7)**, p.645-652

[233] Craddock C, Labopin M, Pillai S *et al.* "Factors predicting outcome after unrelated donor stem cell transplantation in primary refractory acute myeloid leukaemia" (2011) *Leukemia* **25 (5)**, p.808-813

[234] Kekre N, Antin JH "Hematopoietic stem cell transplantation donor sources in the 21st century: choosing the ideal donor when a perfect match does not exist" (2014) *Blood* **124 (3)**, p.334-343

[235] Ballen KK "Is there a best graft source of transplantation in acute myeloid leukemia?" (2015) *Best Pract Res Clin Haematol* **28 (2-3)**, p.147-154

[236] Gratwohl A, Baldomero H, Gratwohl M *et al.* "Quantitative and qualitative differences in use and trends of hematopoietic stem cell transplantation: a Global Observational Study" (2013) *Haematologica* **98 (8)**, p.1282-1290

[237] Ruggeri A, Labopin M, Bacigalupo A *et al.* "Bone marrow versus mobilized peripheral blood stem cells in haploidentical transplants using posttransplantation cyclophosphamide" (2018) *Cancer* **124 (7)**, p.1428-1437

[238] Eapen M, Le Rademacher J, Antin JH *et al.* "Effect of stem cell source on outcomes after unrelated donor transplantation in severe aplastic anemia" (2011) *Blood* **118 (9)**, p.2618-2621

[239] Chen YB, Wang T, Hemmer MT *et al.* "GvHD after umbilical cord blood transplantation for acute leukemia: an analysis of risk factors and effect on outcomes" (2017) *Bone Marrow Transplant* **52 (3)**, p.400-408

[240] Sauter C, Abboud M, Jia X *et al.* "Serious infection risk and immune recovery after double-unit cord blood transplantation without antithymocyte globulin" (2011) *Biol Blood Marrow Transplant* **17 (10)**, p.1460-1471

[241] Lucchini G, Perales MA, Veys P "Immune reconstitution after cord blood transplantation: peculiarities, clinical implications and management strategies" (2015) *Cytotherapy* **17 (6)**, p.711-722

[242] Labopin M, Ruggeri A, Gorin NC *et al.* "Cost-effectiveness and clinical outcomes of double versus single cord blood transplantation in adults with acute leukemia in France" (2014) *Haematologica* **99 (3)**, p.535-540

[243] Solh M, Zhang X, Connor K *et al.* "Donor Type and Disease Risk Predict the Success of Allogeneic Hematopoietic Cell Transplantation: A Single-Center Analysis of 613 Adult Hematopoietic Cell Transplantation Recipients Using a Modified Composite Endpoint" (2017) *Biol Blood Marrow Transplant* **23 (12)**, p.2192-2198

[244] Jagasia M, Arora M, Flowers ME *et al.* "Risk factors for acute GVHD and survival after hematopoietic cell transplantation" (2012) *Blood* **119 (1)**, p.296-307

[245] Jacobsohn DA, Vogelsang GB "Acute graft versus host disease" (2007) *Orphanet J Rare Dis* **2**, p.35

[246] Glucksberg H, Storb R, Fefer A *et al.* "Clinical manifestations of graft-versus-host disease in human recipients of marrow from HL-A-matched sibling donors" (1974) *Transplantation* **18 (4)**, p.295-304

[247] Arora M, Cutler CS, Jagasia MH *et al.* "Late Acute and Chronic Graft-versus-Host Disease after Allogeneic Hematopoietic Cell Transplantation" (2016) *Biol Blood Marrow Transplant* **22 (3)**, p.449-455

[248] Filipovich AH, Weisdorf D, Pavletic S *et al.* "National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. Diagnosis and staging working group report" (2005) *Biol Blood Marrow Transplant* **11 (12)**, p.945-956

[249] Jagasia MH, Greinix HT, Arora M *et al.* "National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging Working Group report" (2015) *Biol Blood Marrow Transplant* **21 (3)**, p.389-401 e381

[250] MacMillan ML, DeFor TE, Weisdorf DJ "What predicts high risk acute graft-versus-host disease (GVHD) at onset?: identification of those at highest risk by a novel acute GVHD risk score" (2012) *Br J Haematol* **157 (6)**, p.732-741

[251] Kolb HJ "Graft-versus-leukemia effects of transplantation and donor lymphocytes" (2008) *Blood* **112 (12)**, p.4371-4383

[252] Apperley JF, Carreras E, Gluckman E *et al.*, "The EBMT Handbook on Haematopoietic Stem Cell Transplantation", 6th ed. Published: Genoa: Forum Service Editore, 2012

[253] Gratwohl A, Baldomero H, Passweg J "Hematopoietic stem cell transplantation activity in Europe" (2013) *Curr Opin Hematol* **20 (6)**, p.485-493

[254] Pavlu J, Apperley J, "Chronic myeloid leukaemia and the myeloproliferative disorders", in *The EBMT Handbook on Haematopoietic Stem Cell Transplantation*, J. F. Apperley, E. Carreras, E. Gluckman, and T. Masszi, Eds. 6th ed. Genoa: Forum Service Editore, 2012, p.316-329.

[255] Mohty M, "Acute Myeloid Leukaemia", in *The EBMT Handbook on Haematopoietic Stem Cell Transplantation*, J. F. Apperley, E. Carreras, E. Gluckman, and T. Masszi, Eds. 6th ed. Genoa: Forum Service Editore, 2012, p.316-329.

[256] Mardis ER, Ding L, Dooling DJ *et al.* "Recurring mutations found by sequencing an acute myeloid leukemia genome" (2009) *N Engl J Med* **361 (11)**, p.1058-1066

[257] Reusing SB, Manser AR, Enczmann J *et al.* "Selective downregulation of HLA-C and HLA-E in childhood acute lymphoblastic leukaemia" (2016) *Br J Haematol* **174 (3)**, p.477-480

[258] Stringaris K, Sekine T, Khoder A *et al.* "Leukemia-induced phenotypic and functional defects in natural killer cells predict failure to achieve remission in acute myeloid leukemia" (2014) *Haematologica* **99 (5)**, p.836-847

[259] Ruggeri L, Capanni M, Urbani E *et al.* "Effectiveness of donor natural killer cell alloreactivity in mismatched hematopoietic transplants" (2002) *Science* **295 (5562)**, p.2097-2100

[260] Miller JS, Cooley S, Parham P *et al.* "Missing KIR ligands are associated with less relapse and increased graft-versus-host disease (GVHD) following unrelated donor allogeneic HCT" (2007) *Blood* **109 (11)**, p.5058-5061

[261] Weisdorf D, Cooley S, Devine S *et al.* "T cell-depleted partial matched unrelated donor transplant for advanced myeloid malignancy: KIR ligand mismatch and outcome" (2012) *Biol Blood Marrow Transplant* **18 (6)**, p.937-943

[262] Shimoni A, Labopin M, Lorentino F *et al.* "Killer cell immunoglobulin-like receptor ligand mismatching and outcome after haploidentical transplantation with post-transplant cyclophosphamide" (2019) *Leukemia* **33 (1)**, p.230-239

[263] Michaelis SU, Mezger M, Bornhauser M *et al.* "KIR haplotype B donors but not KIR-ligand mismatch result in a reduced incidence of relapse after haploidentical transplantation using reduced intensity conditioning and CD3/CD19-depleted grafts" (2014) *Ann Hematol* **93 (9)**, p.1579-1586

[264] Wanquet A, Bramanti S, Harbi S *et al.* "Killer Cell Immunoglobulin-Like Receptor-Ligand Mismatch in Donor versus Recipient Direction Provides Better Graft-versus-Tumor Effect in Patients with Hematologic Malignancies Undergoing Allogeneic T Cell-Replete Haploidentical Transplantation Followed by Post-Transplant Cyclophosphamide" (2018) *Biol Blood Marrow Transplant* **24 (3)**, p.549-554

[265] Brunstein CG, Wagner JE, Weisdorf DJ *et al.* "Negative effect of KIR alloreactivity in recipients of umbilical cord blood transplant depends on transplantation conditioning intensity" (2009) *Blood* **113 (22)**, p.5628-5634

[266] Leung W, Iyengar R, Turner V *et al.* "Determinants of antileukemia effects of allogeneic NK cells" (2004) *J Immunol* **172 (1)**, p.644-650

[267] Martinez-Losada C, Martin C, Gonzalez R, Manzanares B, Garcia-Torres E, Herrera C "Patients Lacking a KIR-Ligand of HLA Group C1 or C2 Have a Better Outcome after Umbilical Cord Blood Transplantation" (2017) *Front Immunol* **8**, p.810

[268] Park S, Kim K, Jang JH *et al.* "KIR alloreactivity based on the receptor-ligand model is associated with improved clinical outcomes of allogeneic hematopoietic stem cell transplantation: Result of single center prospective study" (2015) *Hum Immunol* **76 (9)**, p.636-643

[269] Solomon SR, Aubrey MT, Zhang X *et al.* "Selecting the Best Donor for Haploidentical Transplant: Impact of HLA, Killer Cell Immunoglobulin-Like Receptor Genotyping, and Other Clinical Variables" (2018) *Biol Blood Marrow Transplant* **24 (4)**, p.789-798

[270] Bari R, Rujkijyanont P, Sullivan E *et al.* "Effect of donor KIR2DL1 allelic polymorphism on the outcome of pediatric allogeneic hematopoietic stem-cell transplantation" (2013) *J Clin Oncol* **31 (30)**, p.3782-3790

[271] Marra J, Greene J, Hwang J *et al.* "KIR and HLA genotypes predictive of low-affinity interactions are associated with lower relapse in autologous hematopoietic cell transplantation for acute myeloid leukemia" (2015) *J Immunol* **194 (9)**, p.4222-4230

[272] Cooley S, Weisdorf DJ, Guethlein LA *et al.* "Donor killer cell Ig-like receptor B haplotypes, recipient HLA-C1, and HLA-C mismatch enhance the clinical benefit of unrelated transplantation for acute myelogenous leukemia" (2014) *J Immunol* **192 (10)**, p.4592-4600

[273] Cooley S, Weisdorf DJ, Guethlein LA *et al.* "Donor selection for natural killer cell receptor genes leads to superior survival after unrelated transplantation for acute myelogenous leukemia" (2010) *Blood* **116 (14)**, p.2411-2419

[274] Heatley SL, Mullighan CG, Doherty K *et al.* "Activating KIR Haplotype Influences Clinical Outcome Following HLA-Matched Sibling Hematopoietic Stem Cell Transplantation" (2018) *HLA*,

[275] Impola U, Turpeinen H, Alakulppi N *et al.* "Donor Haplotype B of NK KIR Receptor Reduces the Relapse Risk in HLA-Identical Sibling Hematopoietic Stem Cell Transplantation of AML Patients" (2014) *Front Immunol* **5**, p.405

[276] Bachanova V, Weisdorf DJ, Wang T *et al.* "Donor KIR B Genotype Improves Progression-Free Survival of Non-Hodgkin Lymphoma Patients Receiving Unrelated Donor Transplantation" (2016) *Biol Blood Marrow Transplant* **22 (9)**, p.1602-1607

[277] Oevermann L, Michaelis SU, Mezger M *et al.* "KIR B haplotype donors confer a reduced risk for relapse after haploidentical transplantation in children with ALL" (2014) *Blood* **124 (17)**, p.2744-2747

[278] Kroger N, Binder T, Zabelina T *et al.* "Low number of donor activating killer immunoglobulin-like receptors (KIR) genes but not KIR-ligand mismatch prevents relapse and improves disease-free survival in leukemia patients after in vivo T-cell depleted unrelated stem cell transplantation" (2006) *Transplantation* **82 (8)**, p.1024-1030

[279] Symons HJ, Leffell MS, Rossiter ND, Zahurak M, Jones RJ, Fuchs EJ "Improved survival with inhibitory killer immunoglobulin receptor (KIR) gene mismatches and KIR haplotype B donors after nonmyeloablative, HLA-haploidentical bone marrow transplantation" (2010) *Biol Blood Marrow Transplant* **16 (4)**, p.533-542

[280] Escudero A, Martinez-Romera I, Fernandez L *et al.* "Donor KIR Genotype Impacts on Clinical Outcome after T Cell-Depleted HLA Matched Related Allogeneic Transplantation for High-Risk Pediatric Leukemia Patients" (2018) *Biol Blood Marrow Transplant* **24 (12)**, p.2493-2500

[281] Cooley S, Parham P, Spellman S *et al.* "Unrelated KIR B haplotype donors confer protection against relapse from acute myeloid leukemia after transplantation with reduced intensity conditioning in the modern era" (2019) *HLA* **94 (2)**, p.124

[282] Gagne K, Brizard G, Gueglio B *et al.* "Relevance of KIR gene polymorphisms in bone marrow transplantation outcome" (2002) *Hum Immunol* **63 (4)**, p.271-280

[283] Sahin U, Dalva K, Gungor F, Ustun C, Beksac M "Donor-recipient killer immunoglobulin like receptor (KIR) genotype matching has a protective effect on chronic graft versus host disease and relapse incidence following HLA-identical sibling hematopoietic stem cell transplantation" (2018) *Ann Hematol* **97 (6)**, p.1027-1039

[284] Faridi RM, Kemp TJ, Dharmani-Khan P *et al.* "Donor-Recipient Matching for KIR Genotypes Reduces Chronic GVHD and Missing Inhibitory KIR Ligands

Protect against Relapse after Myeloablative, HLA Matched Hematopoietic Cell Transplantation" (2016) *PLoS One* **11 (6)**, p.e0158242

[285] Mayor NP, Robinson J, McWhinnie AJ *et al.* "HLA Typing for the Next Generation" (2015) *PLoS One* **10 (5)**, p.e0127153

[286] Yang SY, Milford E, Hammerling U, Dupont B, "Description of the reference panel of B-lymphoblastoid cell lines for factors of the HLA system: The B-cell line panel designed for the Tenth Histocompatibility Workshop", in *Immunobiology of HLA Vol. I : Histocompatibility Testing*, vol. Vol. I : Histocompatibility Testing, B. Dupont, Ed. Berlin Heidelberg: Springer-Verlag, 1989.

[287] Miller SA, Dykes DD, Polesky HF "A simple salting out procedure for extracting DNA from human nucleated cells" (1988) *Nucleic Acids Res* **16 (3)**, p.1215

[288] Robinson J, Guijarro C, Leen G *et al.* "Addressing the bioinformatics challenges of high throughput HLA typing using SMRT® DNA sequencing" (2016) *HLA* **87**, p.271

[289] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ "Basic local alignment search tool" (1990) *J Mol Biol* **215 (3)**, p.403-410

[290] Kaplan EL, Meier P "Nonparametric Estimation from Incomplete Observations" (1958) *Journal of the American Statistical Association* **53 (282)**, p.457-481

[291] Gray RJ "A class of K-sample tests for comparing the cumulative incidence of a competing risk" (1988) *The Annals of Statistics* **16 (3)**, p.1141-1154

[292] Gooley TA, Leisenring W, Crowley J, Storer BE "Estimation of failure probabilities in the presence of competing risks: new representations of old estimators" (1999) *Stat Med* **18 (6)**, p.695-706

[293] Przepiorka D, Weisdorf D, Martin P *et al.* "1994 Consensus Conference on Acute GVHD Grading" (1995) *Bone Marrow Transplant* **15 (6)**, p.825-828

[294] Cox DR "Regression Models and Life-Tables" (1972) *Journal of the Royal Statistical Society. Series B (Methodological)* **34 (2)**, p.187-220

[295] Fine JP, Gray RJ "A Proportional Hazards Model for the Subdistribution of a Competing Risk" (1999) *Journal of the American Statistical Association* **94 (446)**, p.496-509

[296] Hoff GA, Fischer JC, Hsu K *et al.* "Recipient HLA-C Haplotypes and microRNA 148a/b Binding Sites Have No Impact on Allogeneic Hematopoietic Cell Transplantation Outcomes" (2017) *Biol Blood Marrow Transplant* **23 (1)**, p.153-160

[297] Michonneau D, Socié G, "GVHD Prophylaxis (Immunosuppression)", in *The EBMT Handbook*, E. Carreras, C. Dufour, M. Mohty, and N. Kroger, Eds. Cham, Switzerland: Springer Nature Switzerland AG, 2018, p.177-182.

[298] Saunders PM, Pymm P, Pietra G *et al.* "Killer cell immunoglobulin-like receptor 3DL1 polymorphism defines distinct hierarchies of HLA class I recognition" (2016) *J Exp Med* **213 (5)**, p.791-807

[299] Jiang Y, Chen O, Cui C *et al.* "KIR3DS1/L1 and HLA-Bw4-80I are associated with HIV disease progression among HIV typical progressors and long-term nonprogressors" (2013) *BMC Infect Dis* **13**, p.405

[300] Venstrom JM, Pittari G, Gooley TA *et al.* "HLA-C-dependent prevention of leukemia relapse by donor activating KIR2DS1" (2012) *N Engl J Med* **367 (9)**, p.805-816

[301] Fischer JC, Ottinger H, Ferencik S *et al.* "Relevance of C1 and C2 epitopes for hemopoietic stem cell transplantation: role for sequential acquisition of HLA-C-specific inhibitory killer Ig-like receptor" (2007) *J Immunol* **178 (6)**, p.3918-3923

[302] Neuchel C, Furst D, Niederwieser D *et al.* "Impact of Donor Activating KIR Genes on HSCT Outcome in C1-Ligand Negative Myeloid Disease Patients Transplanted with Unrelated Donors-A Retrospective Study" (2017) *PLoS One* **12 (1)**, p.e0169512

[303] Giebel S, Locatelli F, Lamparelli T *et al.* "Survival advantage with KIR ligand incompatibility in hematopoietic stem cell transplantation from unrelated donors" (2003) *Blood* **102 (3)**, p.814-819

[304] Schaffer M, Malmberg KJ, Ringden O, Ljunggren HG, Remberger M "Increased infection-related mortality in KIR-ligand-mismatched unrelated allogeneic hematopoietic stem-cell transplantation" (2004) *Transplantation* **78 (7)**, p.1081-1085

[305] Verheyden S, Schots R, Duquet W, Demanet C "A defined donor activating natural killer cell receptor genotype protects against leukemic relapse after related HLA-identical hematopoietic stem cell transplantation" (2005) *Leukemia* **19 (8)**, p.1446-1451

[306] Hsu KC, Keever-Taylor CA, Wilton A *et al.* "Improved outcome in HLA-identical sibling hematopoietic stem-cell transplantation for acute myelogenous leukemia predicted by KIR and HLA genotypes" (2005) *Blood* **105 (12)**, p.4878-4884

[307] Giebel S, Locatelli F, Wojnar J *et al.* "Homozygosity for human leucocyte antigen-C ligands of KIR2DL1 is associated with increased risk of relapse after human leucocyte antigen-C-matched unrelated donor haematopoietic stem cell transplantation" (2005) *Br J Haematol* **131 (4)**, p.483-486

[308] Hsu KC, Gooley T, Malkki M *et al.* "KIR ligands and prediction of relapse after unrelated donor hematopoietic cell transplantation for hematologic malignancy" (2006) *Biol Blood Marrow Transplant* **12 (8)**, p.828-836

[309] Farag SS, Bacigalupo A, Eapen M *et al.* "The effect of KIR ligand incompatibility on the outcome of unrelated donor transplantation: a report from the center for international blood and marrow transplant research, the European blood and

marrow transplant registry, and the Dutch registry" (2006) *Biol Blood Marrow Transplant* **12 (8)**, p.876-884

[310] Clausen J, Wolf D, Petzer AL *et al.* "Impact of natural killer cell dose and donor killer-cell immunoglobulin-like receptor (KIR) genotype on outcome following human leucocyte antigen-identical haematopoietic stem cell transplantation" (2007) *Clin Exp Immunol* **148 (3)**, p.520-528

[311] Sobecks RM, Ball EJ, Maciejewski JP *et al.* "Survival of AML patients receiving HLA-matched sibling donor allogeneic bone marrow transplantation correlates with HLA-Cw ligand groups for killer immunoglobulin-like receptors" (2007) *Bone Marrow Transplant* **39 (7)**, p.417-424

[312] Bjorklund AT, Schaffer M, Fauriat C *et al.* "NK cells expressing inhibitory KIR for non-self-ligands remain tolerant in HLA-matched sibling stem cell transplantation" (2010) *Blood* **115 (13)**, p.2686-2694

[313] Clausen J, Kircher B, Auberger J *et al.* "The role of missing killer cell immunoglobulin-like receptor ligands in T cell replete peripheral blood stem cell transplantation from HLA-identical siblings" (2010) *Biol Blood Marrow Transplant* **16 (2)**, p.273-280

[314] Sobecks RM, Wang T, Askar M *et al.* "Impact of KIR and HLA Genotypes on Outcomes after Reduced-Intensity Conditioning Hematopoietic Cell Transplantation" (2015) *Biol Blood Marrow Transplant* **21 (9)**, p.1589-1596

[315] Shimoni A, Vago L, Bernardi M *et al.* "Missing HLA C group 1 ligand in patients with AML and MDS is associated with reduced risk of relapse and better survival after allogeneic stem cell transplantation with fludarabine and treosulfan reduced toxicity conditioning" (2017) *Am J Hematol* **92 (10)**, p.1011-1019

[316] Bao X, Wang M, Zhou H *et al.* "Donor Killer Immunoglobulin-Like Receptor Profile Bx1 Imparts a Negative Effect and Centromeric B-Specific Gene Motifs Render a Positive Effect on Standard-Risk Acute Myeloid Leukemia/Myelodysplastic Syndrome Patient Survival after Unrelated Donor Hematopoietic Stem Cell Transplantation" (2016) *Biol Blood Marrow Transplant* **22 (2)**, p.232-239

[317] Venstrom JM, Gooley TA, Spellman S *et al.* "Donor activating KIR3DS1 is associated with decreased acute GVHD in unrelated allogeneic hematopoietic stem cell transplantation" (2010) *Blood* **115 (15)**, p.3162-3165

[318] Stringaris K, Adams S, Uribe M *et al.* "Donor KIR Genes 2DL5A, 2DS1 and 3DS1 are associated with a reduced rate of leukemia relapse after HLA-identical sibling stem cell transplantation for acute myeloid leukemia but not other hematologic malignancies" (2010) *Biol Blood Marrow Transplant* **16 (9)**, p.1257-1264

[319] Hosokai R, Masuko M, Shibasaki Y, Saitoh A, Furukawa T, Imai C "Donor Killer Immunoglobulin-Like Receptor Haplotype B/x Induces Severe Acute Graft-versus-Host Disease in the Presence of Human Leukocyte Antigen Mismatch in T

Cell-Replete Hematopoietic Cell Transplantation" (2017) *Biol Blood Marrow Transplant* **23 (4)**, p.606-611

[320] Farrar JE, Bolouri H, Ries RE *et al.* "Marked Differences in the Genomic Landscape of Pediatric Compared to Adult Acute Myeloid Leukemia: A Report from the Children's Oncology Group and NCI/COG Therapeutically Applicable Research to Generate Effective Treatments (TARGET) Initiative" (2016) *Blood* **128 (22)**, p.595

[321] Creutzig U, Zimmermann M, Reinhardt D *et al.* "Changes in cytogenetics and molecular genetics in acute myeloid leukemia from childhood to adult age groups" (2016) *Cancer* **122 (24)**, p.3821-3830

[322] Ringden O, Labopin M, Ehninger G *et al.* "Reduced intensity conditioning compared with myeloablative conditioning using unrelated donor transplants in patients with acute myeloid leukemia" (2009) *J Clin Oncol* **27 (27)**, p.4570-4577

[323] Shimoni A, Labopin M, Savani B *et al.* "Long-term survival and late events after allogeneic stem cell transplantation from HLA-matched siblings for acute myeloid leukemia with myeloablative compared to reduced-intensity conditioning: a report on behalf of the acute leukemia working party of European group for blood and marrow transplantation" (2016) *J Hematol Oncol* **9 (1)**, p.118

[324] Jimenez M, Ercilla G, Martinez C "Immune reconstitution after allogeneic stem cell transplantation with reduced-intensity conditioning regimens" (2007) *Leukemia* **21 (8)**, p.1628-1637

[325] Gilleece MH, Labopin M, Savani BN *et al.* "Allogeneic haemopoietic transplantation for acute myeloid leukaemia in second complete remission: a registry report by the Acute Leukaemia Working Party of the EBMT" (2019) *Leukemia*,

[326] Bornhauser M, Schwerdtfeger R, Martin H, Frank KH, Theuser C, Ehninger G "Role of KIR ligand incompatibility in hematopoietic stem cell transplantation using unrelated donors" (2004) *Blood* **103 (7)**, p.2860-2861; author reply 2862

[327] Kroger N, Shaw B, Iacobelli S *et al.* "Comparison between antithymocyte globulin and alemtuzumab and the possible impact of KIR-ligand mismatch after dose-reduced conditioning and unrelated stem cell transplantation in patients with multiple myeloma" (2005) *Br J Haematol* **129 (5)**, p.631-643

[328] Robin M, Raj K, Chevret S *et al.* "Alemtuzumab vs anti-thymocyte globulin in patients transplanted from an unrelated donor after a reduced intensity conditioning" (2018) *Eur J Haematol* **101 (4)**, p.466-474

[329] Veys P, Wynn RF, Ahn KW *et al.* "Impact of immune modulation with in vivo T-cell depletion and myleoablative total body irradiation conditioning on outcomes after unrelated donor transplantation for childhood acute lymphoblastic leukemia" (2012) *Blood* **119 (25)**, p.6155-6161

[330] Chakrabarti S, Mackinnon S, Chopra R *et al.* "High incidence of cytomegalovirus infection after nonmyeloablative stem cell transplantation: potential role of

Campath-1H in delaying immune reconstitution" (2002) *Blood* **99 (12)**, p.4357-4363

[331] Juliusson G, Theorin N, Karlsson K, Frodin U, Malm C "Subcutaneous alemtuzumab vs ATG in adjusted conditioning for allogeneic transplantation: influence of Campath dose on lymphoid recovery, mixed chimerism and survival" (2006) *Bone Marrow Transplant* **37 (5)**, p.503-510

[332] Waggoner SN, Cornberg M, Selin LK, Welsh RM "Natural killer cells act as rheostats modulating antiviral T cells" (2011) *Nature* **481 (7381)**, p.394-398

[333] Yu J, Venstrom JM, Liu XR *et al.* "Breaking tolerance to self, circulating natural killer cells expressing inhibitory KIR for non-self HLA exhibit effector function after T cell-depleted allogeneic hematopoietic cell transplantation" (2009) *Blood* **113 (16)**, p.3875-3884

[334] Pende D, Marcenaro S, Falco M *et al.* "Anti-leukemia activity of alloreactive NK cells in KIR ligand-mismatched haploidentical HSCT for pediatric patients: evaluation of the functional role of activating KIR and redefinition of inhibitory KIR specificity" (2009) *Blood* **113 (13)**, p.3119-3129

[335] Beziat V, Liu LL, Malmberg JA *et al.* "NK cell responses to cytomegalovirus infection lead to stable imprints in the human KIR repertoire and involve activating KIRs" (2013) *Blood* **121 (14)**, p.2678-2688

[336] Khakoo SI, Thio CL, Martin MP *et al.* "HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection" (2004) *Science* **305 (5685)**, p.872-874

[337] de Vasconcelos JM, de Jesus Maues Pereira Moia L, Amaral Ido S *et al.* "Association of killer cell immunoglobulin-like receptor polymorphisms with chronic hepatitis C and responses to therapy in Brazil" (2013) *Genet Mol Biol* **36 (1)**, p.22-27

[338] Dring MM, Morrison MH, McSharry BP *et al.* "Innate immune genes synergize to predict increased risk of chronic disease in hepatitis C virus infection" (2011) *Proc Natl Acad Sci U S A* **108 (14)**, p.5736-5741

[339] Bari R, Thapa R, Bao J, Li Y, Zheng J, Leung W "KIR2DL2/2DL3-E(35) alleles are functionally stronger than -Q(35) alleles" (2016) *Sci Rep* **6**, p.23689

[340] Fauriat C, Ivarsson MA, Ljunggren HG, Malmberg KJ, Michaelsson J "Education of human natural killer cells by activating killer cell immunoglobulin-like receptors" (2010) *Blood* **115 (6)**, p.1166-1174

[341] Peppa D, Gill US, Reynolds G *et al.* "Up-regulation of a death receptor renders antiviral T cells susceptible to NK cell-mediated deletion" (2013) *J Exp Med* **210 (1)**, p.99-114

[342] Boelen L, Debebe B, Silveira M *et al.* "Inhibitory killer cell immunoglobulin-like receptors strengthen CD8(+) T cell-mediated control of HIV-1, HCV, and HTLV-1" (2018) *Sci Immunol* **3 (29)**,

[343] Hayakawa Y, Screpanti V, Yagita H *et al.* "NK cell TRAIL eliminates immature dendritic cells in vivo and limits dendritic cell vaccination efficacy" (2004) *J Immunol* **172 (1)**, p.123-129

[344] Smith LE, Olszewski MA, Georgoudaki AM *et al.* "Sensitivity of dendritic cells to NK-mediated lysis depends on the inflammatory environment and is modulated by CD54/CD226-driven interactions" (2016) *J Leukoc Biol* **100 (4)**, p.781-789

[345] Battiwalla M, McCarthy PL "Filgrastim support in allogeneic HSCT for myeloid malignancies: a review of the role of G-CSF and the implications for current practice" (2009) *Bone Marrow Transplant* **43 (5)**, p.351-356

[346] Yawata M, Yawata N, Draghi M, Partheniou F, Little AM, Parham P "MHC class I-specific inhibitory receptors and their ligands structure diverse human NK-cell repertoires toward a balance of missing self-response" (2008) *Blood* **112 (6)**, p.2369-2380

[347] Bjorkstrom NK, Beziat V, Cichocki F *et al.* "CD8 T cells express randomly selected KIRs with distinct specificities compared with NK cells" (2012) *Blood* **120 (17)**, p.3455-3465

[348] Hilton HG, Guethlein LA, Goyos A *et al.* "Polymorphic HLA-C Receptors Balance the Functional Characteristics of KIR Haplotypes" (2015) *J Immunol* **195 (7)**, p.3160-3170

[349] Mulrooney TJ, Zhang AC, Goldgur Y, Boudreau JE, Hsu KC "KIR3DS1-Specific D0 Domain Polymorphisms Disrupt KIR3DL1 Surface Expression and HLA Binding" (2015) *J Immunol* **195 (3)**, p.1242-1250

[350] O'Connor GM, Vivian JP, Widjaja JM *et al.* "Mutational and structural analysis of KIR3DL1 reveals a lineage-defining allotypic dimorphism that impacts both HLA and peptide sensitivity" (2014) *J Immunol* **192 (6)**, p.2875-2884

[351] Taner SB, Pando MJ, Roberts A *et al.* "Interactions of NK cell receptor KIR3DL1*004 with chaperones and conformation-specific antibody reveal a functional folded state as well as predominant intracellular retention" (2011) *J Immunol* **186 (1)**, p.62-72

[352] Yawata M, Yawata N, Draghi M, Little AM, Partheniou F, Parham P "Roles for HLA and KIR polymorphisms in natural killer cell repertoire selection and modulation of effector function" (2006) *J Exp Med* **203 (3)**, p.633-645

[353] Gardiner CM, Guethlein LA, Shilling HG *et al.* "Different NK cell surface phenotypes defined by the DX9 antibody are due to KIR3DL1 gene polymorphism" (2001) *J Immunol* **166 (5)**, p.2992-3001

[354] Tao SD, He YM, Ying YL, He J, Zhu FM, Lv HJ "KIR3DL1 genetic diversity and phenotypic variation in the Chinese Han population" (2014) *Genes Immun* **15 (1)**, p.8-15

[355] Mulrooney TJ, Hou L, Steiner NK *et al.* "Promoter variants of KIR2DL5 add to diversity and may impact gene expression" (2008) *Immunogenetics* **60 (6)**, p.287-294

[356] Alicata C, Pende D, Meazza R *et al.* "Hematopoietic stem cell transplantation: Improving alloreactive Bw4 donor selection by genotyping codon 86 of KIR3DL1/S1" (2016) *Eur J Immunol* **46 (6)**, p.1511-1517

[357] Kumar S, Sarkar P, Sim MJ, Rajagopalan S, Vogel SS, Long EO "A single amino acid change in inhibitory killer cell Ig-like receptor results in constitutive receptor self-association and phosphorylation" (2015) *J Immunol* **194 (2)**, p.817-826

[358] Misra MK, Augusto DG, Martin GM *et al.* "Report from the Killer-cell Immunoglobulin-like Receptors (KIR) component of the 17th International HLA and Immunogenetics Workshop" (2018) *Hum Immunol* **79 (12)**, p.825-833

[359] Genomes Project C, Auton A, Brooks LD *et al.* "A global reference for human genetic variation" (2015) *Nature* **526 (7571)**, p.68-74

[360] Katoh K, Standley DM "MAFFT multiple sequence alignment software version 7: improvements in performance and usability" (2013) *Mol Biol Evol* **30 (4)**, p.772-780

[361] Kuiken C, Korber B, Shafer RW "HIV sequence databases" (2003) *AIDS Rev* **5 (1)**, p.52-61

[362] Cornish-Bowden A "Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984" (1985) *Nucleic Acids Res* **13 (9)**, p.3021-3030

[363] Shilling HG, Guethlein LA, Cheng NW *et al.* "Allelic polymorphism synergizes with variable gene content to individualize human KIR genotype" (2002) *J Immunol* **168 (5)**, p.2307-2315

[364] Hosomichi K, Jinam TA, Mitsunaga S, Nakaoka H, Inoue I "Phase-defined complete sequencing of the HLA genes by next-generation sequencing" (2013) *BMC Genomics* **14**, p.355

[365] Laehnemann D, Borkhardt A, McHardy AC "Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction" (2016) *Brief Bioinform* **17 (1)**, p.154-179

[366] Lo YM "The amplification refractory mutation system" (1998) *Methods Mol Med* **16**, p.61-69

[367] Ranade S, Lee W, Harting J *et al.*, "Multiplexing Human HLA Class I & II Genotyping with DNA Barcode Adapters for High Throughput Research", presented at the Advances in Genome Biology & Technology Conference 2015, Marco Island, FL, USA, 2015.

[368] Quail MA, Smith M, Coupland P *et al.* "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers" (2012) *BMC Genomics* **13**, p.341

[369] Bultitude WP, Gymer AW, Robinson J, Mayor NP, Marsh SGE "The novel KIR2DL1 allele, KIR2DL1*037, defined in the cell line SPO010 (IHW9036)" (2018) *HLA* **91 (6)**, p.547-548

[370] Biosciences P, "SMRTbell Library Preparation & SMRT Sequencing Workflow Updates", presented at the PacBio Americas User Group Meeting Sample Prep Workshop, Baltimore, MD, USA, 2017.

[371] Bono M, Pende D, Bertaina A *et al.* "Analysis of KIR3DP1 Polymorphism Provides Relevant Information on Centromeric KIR Gene Content" (2018) *J Immunol* **201 (5)**, p.1460-1467

[372] Klitz W, Hedrick P, Louis EJ "New reservoirs of HLA alleles: pools of rare variants enhance immune defense" (2012) *Trends Genet* **28 (10)**, p.480-486

[373] Wagner I, Schefzyk D, Pruschke J *et al.* "Allele-Level KIR Genotyping of More Than a Million Samples: Workflow, Algorithm, and Observations" (2018) *Front Immunol* **9**, p.2843

[374] Stulberg MJ, Wright PW, Dang H, Hanson RJ, Miller JS, Anderson SK "Identification of distal KIR promoters and transcripts" (2007) *Genes Immun* **8 (2)**, p.124-130

[375] Wright PW, Li H, Huehn A *et al.* "Characterization of a weakly expressed KIR2DL1 variant reveals a novel upstream promoter that controls KIR expression" (2014) *Genes Immun* **15 (7)**, p.440-448

[376] Cichocki F, Hanson RJ, Lenvik T *et al.* "The transcription factor c-Myc enhances KIR gene transcription through direct binding to an upstream distal promoter element" (2009) *Blood* **113 (14)**, p.3245-3253

[377] Presnell SR, Zhang L, Ramilo CA, Chan HW, Lutz CT "Functional redundancy of transcription factor-binding sites in the killer cell Ig-like receptor (KIR) gene promoter" (2006) *Int Immunol* **18 (8)**, p.1221-1232

[378] Li H, Pascal V, Martin MP, Carrington M, Anderson SK "Genetic control of variegated KIR gene expression: polymorphisms of the bi-directional KIR3DL1 promoter are associated with distinct frequencies of gene expression" (2008) *PLoS Genet* **4 (11)**, p.e1000254

[379] Cichocki F, Lenvik T, Sharma N, Yun G, Anderson SK, Miller JS "Cutting Edge: KIR Antisense Transcripts Are Processed into a 28-Base PIWI-Like RNA in Human NK Cells" (2010) **185 (4)**, p.2009-2012

[380] Wright PW, Huehn A, Cichocki F *et al.* "Identification of a KIR antisense lncRNA expressed by progenitor cells" (2013) *Genes Immun* **14 (7)**, p.427-433

[381] Hilton HG, Norman PJ, Nemat-Gorgani N *et al.* "Loss and Gain of Natural Killer Cell Receptor Function in an African Hunter-Gatherer Population" (2015) *PLoS Genet* **11 (8)**, p.e1005439

[382] Han MV, Zmasek CM "phyloXML: XML for evolutionary biology and comparative genomics" (2009) *BMC Bioinformatics* **10**, p.356

[383] Gourraud PA, Meenagh A, Cambon-Thomsen A, Middleton D "Linkage disequilibrium organization of the human KIR superlocus: implications for KIR data analyses" (2010) *Immunogenetics* **62 (11-12)**, p.729-740

[384] Maldonado Torres H, Robinson J, Madrigal JA, Marsh SGE "Cactus, a population genetics analysis environment" (2005) *Tissue Antigens* **66 (5)**, p.486

[385] Hurst LD "Fundamental concepts in genetics: genetics and the understanding of selection" (2009) *Nat Rev Genet* **10 (2)**, p.83-93

[386] Norman PJ, Hollenbach JA, Nemat-Gorgani N *et al.* "Co-evolution of human leukocyte antigen (HLA) class I ligands with killer-cell immunoglobulin-like receptors (KIR) in a genetically diverse population of sub-Saharan Africans" (2013) *PLoS Genet* **9 (10)**, p.e1003938

[387] Bruijnesteijn J, van der Wiel MKH, de Groot N *et al.* "Extensive Alternative Splicing of KIR Transcripts" (2018) *Front Immunol* **9**, p.2846

[388] Dunphy SE, Guinan KJ, Chorcora CN *et al.* "2DL1, 2DL2 and 2DL3 all contribute to KIR phenotype variability on human NK cells" (2015) *Genes Immun* **16 (5)**, p.301-310

[389] Davies GE, Locke SM, Wright PW *et al.* "Identification of bidirectional promoters in the human KIR genes" (2007) *Genes Immun* **8 (3)**, p.245-253

[390] Gao XN, Yu L "E2F1 contributes to the transcriptional activation of the KIR3DL1 gene" (2008) *Biochem Biophys Res Commun* **370 (3)**, p.399-403

[391] Shaffer BC, Hsu KC "How important is NK alloreactivity and KIR in allogeneic transplantation?" (2016) *Best Pract Res Clin Haematol* **29 (4)**, p.351-358

[392] Boudreau JE, Hsu KC "Natural Killer Cell Education and the Response to Infection and Cancer Therapy: Stay Tuned" (2018) *Trends Immunol* **39 (3)**, p.222-239

[393] Boudreau JE, Giglio F, Gooley TA *et al.* "KIR3DL1/HLA-B Subtypes Govern Acute Myelogenous Leukemia Relapse After Hematopoietic Cell Transplantation" (2017) *J Clin Oncol* **35 (20)**, p.2268-2278

[394] Forlenza CJ, Boudreau JE, Zheng J *et al.* "KIR3DL1 Allelic Polymorphism and HLA-B Epitopes Modulate Response to Anti-GD2 Monoclonal Antibody in Patients With Neuroblastoma" (2016) *J Clin Oncol* **34 (21)**, p.2443-2451

[395] Schellekens J, Gagne K, Marsh SGE "Natural killer cells and killer-cell immunoglobulin-like receptor polymorphisms: their role in hematopoietic stem cell transplantation" (2014) *Methods Mol Biol* **1109**, p.139-158

[396] Maniangou B, Retiere C, Gagne K "Next-generation sequencing technology a new tool for killer cell immunoglobulin-like receptor allele typing in hematopoietic stem cell transplantation" (2018) *Transfus Clin Biol* **25 (1)**, p.87-89

[397] Gagne K, Willem C, Legrand N *et al.* "Both the nature of KIR3DL1 alleles and the KIR3DL1/S1 allele combination affect the KIR3DL1 NK-cell repertoire in the French population" (2013) *Eur J Immunol* **43 (4)**, p.1085-1098

[398] Bari R, Leung M, Turner VE *et al.* "Molecular determinant-based typing of KIR alleles and KIR ligands" (2011) *Clin Immunol* **138 (3)**, p.274-281

[399] Bari R, Schell S, Tuggle M, Leung W "An Improved Method With High Specificity for KIR2DL1 Functional Allele Typing" (2015) *Lab Med* **46 (3)**, p.207-213

[400] Babor F, Peters C, Manser AR *et al.* "Presence of centromeric but absence of telomeric group B KIR haplotypes in stem cell donors improve leukaemia control after HSCT for childhood ALL" (2019) *Bone Marrow Transplant*,

[401] Sim MJW, Rajagopalan S, Altmann DM, Boyton RJ, Sun PD, Long EO "Human NK cell receptor KIR2DS4 detects a conserved bacterial epitope presented by HLA-C" (2019) *Proc Natl Acad Sci U S A*,

[402] Jarjanazi H, Savas S, Pabalan N, Dennis JW, Ozcelik H "Biological implications of SNPs in signal peptide domains of human proteins" (2008) *Proteins* **70 (2)**, p.394-403

[403] Stern M, de Angelis C, Urbani E *et al.* "Natural killer-cell KIR repertoire reconstitution after haploidentical SCT" (2010) *Bone Marrow Transplant* **45 (11)**, p.1607-1610

[404] Willem C, Makanga DR, Guillaume T *et al.* "Impact of KIR/HLA Incompatibilities on NK Cell Reconstitution and Clinical Outcome after T Cell-Replete Haploidentical Hematopoietic Stem Cell Transplantation with Posttransplant Cyclophosphamide" (2019) *J Immunol* **202 (7)**, p.2141-2152

[405] Jimenez M, Martinez C, Ercilla G *et al.* "Reduced-intensity conditioning regimen preserves thymic function in the early period after hematopoietic stem cell transplantation" (2005) *Exp Hematol* **33 (10)**, p.1240-1248

[406] van Besien K "Allogeneic transplantation for AML and MDS: GVL versus GVHD and disease recurrence" (2013) *Hematology Am Soc Hematol Educ Program* **2013**, p.56-62

[407] Recker K, Fuerst D, Schrezenmeier H, Mytilineos J "Effects of Minor Histocompatibility Antigen Mismatches on Clinical Outcome After Unrelated Hematopoietic Stem Cell Transplantation (HSCT) in 320 Recipients and Their

Unrelated Donors – a Multivariate Analysis" (2011) *Biology of Blood and Marrow Transplantation* **17 (2)**, p.S155

[408] Nakagawa S "A farewell to Bonferroni: the problems of low statistical power and publication bias" (2004) *Behavioral Ecology* **15 (6)**, p.1044-1045

[409] Weisdorf D, Cooley S, Wang T *et al.* "KIR Donor Selection: Feasibility in Identifying better Donors" (2019) *Biol Blood Marrow Transplant* **25 (1)**, p.e28-e32

[410] Cornelissen JJ, Blaise D "Hematopoietic stem cell transplantation for patients with AML in first complete remission" (2016) *Blood* **127 (1)**, p.62-70

[411] Arrieta-Bolanos E, Mayor NP, Marsh SG *et al.* "Polymorphism in TGFB1 is associated with worse non-relapse mortality and overall survival after stem cell transplantation with unrelated donors" (2016) *Haematologica* **101 (3)**, p.382-390

[412] Mayor NP, Shaw BE, Madrigal JA, Marsh SG "NOD2 Polymorphisms and Their Impact on Haematopoietic Stem Cell Transplant Outcome" (2012) *Bone Marrow Res* **2012**, p.180391

[413] Dickinson AM, Norden J "Non-HLA genomics: does it have a role in predicting haematopoietic stem cell transplantation outcome?" (2015) *Int J Immunogenet* **42 (4)**, p.229-238

[414] Balavarca Y, Pearce K, Norden J *et al.* "Predicting survival using clinical risk scores and non-HLA immunogenetics" (2015) *Bone Marrow Transplant* **50 (11)**, p.1445-1452

[415] Verneris MR, Lee SJ, Ahn KW *et al.* "HLA Mismatch Is Associated with Worse Outcomes after Unrelated Donor Reduced-Intensity Conditioning Hematopoietic Cell Transplantation: An Analysis from the Center for International Blood and Marrow Transplant Research" (2015) *Biol Blood Marrow Transplant* **21 (10)**, p.1783-1789

[416] Huhn O, Chazara O, Ivarsson MA *et al.* "High-Resolution Genetic and Phenotypic Analysis of KIR2DL1 Alleles and Their Association with Pre-Eclampsia" (2018) *J Immunol* **201 (9)**, p.2593-2601

Appendices

Supplementary Table A. IHIW Cell line names and IDs

Cell line name	Cell line ID	Cell line name	Cell line ID
AKIBA	IHW9286	JTHOM	IHW9004
AZH	IHW9293	JVM	IHW9039
AZL	IHW9064	KGU	IHW9309
BH	IHW9046	KOSE	IHW9056
BM14	IHW9033	LBF	IHW9048
BM15	IHW9040	LO081785	IHW9018
BOB	IHW9089	LUCE	IHW9393
BOLETH	IHW9031	LZL	IHW9099
BSM	IHW9032	MANIKA	IHW9106
BTB	IHW9067	MOU	IHW9050
CALOGERO	IHW9084	OLL	IHW9100
CB6B	IHW9060	OMW	IHW9058
DBB	IHW9052	PE117	IHW9028
DEM	IHW9007	PF04015	IHW9088
DEU	IHW9025	PGF	IHW9318
DKB	IHW9075	PLH	IHW9047
E481324	IHW9011	QBL	IHW9020
EA	IHW9081	RML	IHW9016
EHM	IHW9080	RSH	IHW9021
EJ32B	IHW9085	SCHU	IHW9013
EK	IHW9054	SLE005	IHW9059
FH13	IHW9387	SPO010	IHW9036
FH15	IHW9389	STEINLIN	IHW9087
FH5	IHW9377	T7526	IHW9076
FH6	IHW9380	T7527	IHW9077
FH8	IHW9382	TEM	IHW9057
FPAF	IHW9105	TISI	IHW9042
G085	IHW9263	WIN	IHW9095
GRC212	IHW9364	WJR076	IHW9012
HO104	IHW9082	WT24	IHW9015
JBUSH	IHW9035	WT47	IHW9063
JHAF	IHW9030	WT51	IHW9029
JO528239	IHW9041	YAR	IHW9026

Supplementary Table B. Univariate analysis p-values of HCT outcomes based on clinical factors

Clinical factor	5 year OS			5 year DFS			5 year relapse			1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
Donor age	***0.042***	0.13	0.67	***0.017***	0.067	0.37	***0.034***	0.35	0.12	0.80	0.34	0.36	***0.017***	0.74	0.089
Recipient age	0.13	0.051	***0.049***	0.36	0.097	0.083	0.59	0.48	0.79	0.73	0.49	0.097	0.68	0.70	0.45
Donor gender	0.25	0.25	0.99	0.12	0.71	0.53	0.59	0.42	0.66	0.099	0.64	0.49	0.28	0.35	0.74
Recipient gender	0.38	0.63	0.97	1.00	0.73	0.59	0.17	0.84	0.47	0.30	0.99	0.51	0.65	0.28	0.66
Donor-recipient gender matching	0.24	0.55	0.69	0.25	0.66	0.87	0.56	0.81	0.60	0.16	0.96	0.54	0.29	0.49	0.83
Donor-recipient CMV matching	***0.002***	0.12	0.17	***0.008***	0.20	0.15	0.71	0.79	0.33	***0.034***	0.25	0.52	0.32	0.18	0.79
Era of transplant	0.87	0.52	0.45	0.30	0.37	0.60	0.20	0.80	***0.049***	***0.020***	0.21	0.11	0.29	0.56	0.56
Use of alemtuzumab	0.19	0.076	0.28	0.20	0.14	0.22	0.22	0.61	0.46	0.82	0.29	0.63	***<0.001***	0.093	***0.028***
EBMT disease risk score	***0.009***	0.19	0.52	0.052	0.63	0.52	***0.003***	0.35	***<0.001***	0.29	0.69	0.080	0.44	0.82	1.00
HSC source	0.99	0.57	0.13	0.39	0.99	0.49	0.23	0.82	0.59	0.94	0.92	0.41	***0.007***	0.095	***0.026***
Previous autograft history	***0.010***	0.68	***0.028***	***0.008***	0.70	0.063	0.23	0.30	0.62	0.086	0.84	0.18	1.00	1.00	1.00
HLA-matching	***<0.001***	-	-	***<0.001***	-	-	0.53	-	-	***<0.001***	-	-	0.068	-	-
Conditioning regimen	0.97	-	-	0.97	-	-	0.33	-	-	0.11	-	-	0.59	-	-

Statistically significant results are denoted by ***bold italics***.

Supplementary Table C. Subgroup population sizes for missing ligand analysis on the overall and adult, HLA-matched RIC and MAC cohorts

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
Missing KIR2DL1 ligand									
No missing KIR2DL1 ligand	233	78	63	227	74	62	220	73	60
Missing KIR2DL1 ligand	172	57	56	166	53	56	164	54	55
Missing KIR2DL2/3 ligand									
No missing KIR2DL2/3 ligand	361	120	105	349	112	104	344	115	102
Missing KIR2DL2/3 ligand	44	15	14	44	15	14	40	12	13
Missing KIR3DL1 ligand									
No missing KIR3DL1 ligand	267	78	76	259	73	75	249	71	73
Missing KIR3DL1 ligand	137	56	43	133	53	43	134	55	42
KIR3DL1 ligand strategy 1									
Donor KIR3DL1^{-ve} or Recipient Bw4^{-ve}	155	63	46	150	59	46	151	61	45
Donor KIR3DL1^{+ve} & Recipient Bw4-80I^{+ve}	118	31	30	114	29	30	111	29	27
Donor KIR3DL1^{+ve} & Recipient Bw4-80I^{-ve}	131	40	43	128	38	42	121	36	43
KIR3DL1 ligand strategy 2									
Donor KIR3DL1^{-ve} or Recipient HLA-Bw4-80I^{-ve}	286	103	89	278	97	88	272	97	88
Donor KIR3DL1^{+ve} & Recipient 1 copy HLA-Bw4-80I	100	26	26	96	24	26	93	24	23
Donor KIR3DL1^{+ve} & Recipient >1 copy HLA-Bw4-80I	18	5	4	18	5	4	18	5	4
Missing inhibitory KIR ligand frequency									
0 missing inhibitory KIR ligands	149	45	39	145	43	38	140	43	37
1 missing inhibitory KIR ligand	157	50	47	151	45	47	148	45	46
2 missing inhibitory KIR ligands	98	39	33	96	38	33	95	38	32
Missing KIR2DS1 ligand									
Recipient HLA-C1^{-ve}	44	15	14	44	15	14	40	12	13
Recipient HLA-C1^{+ve} & Donor KIR2DS1^{+ve}	217	67	69	215	66	68	210	65	69
Recipient HLA-C1^{+ve} & Donor KIR2DS1^{-ve}	144	53	36	134	46	36	134	50	33

Supplementary Table D. Subgroup population sizes for KIR gene matching analysis on the overall and adult, HLA-matched RIC and MAC cohorts

Test	5 year OS						5 year DFS, 5 year relapse and 1 year NRM						aGVHD					
	Entire cohort		RIC cohort		MAC cohort		Entire cohort		RIC cohort		MAC cohort		Entire cohort		RIC cohort		MAC cohort	
	Any	GVH only	Any	GVH only	Any	GVH only	Any	GVH only	Any	GVH only	Any	GVH only	Any	GVH only	Any	GVH only	Any	GVH only
All loci matching																		
All KIR genes matched	56	168	23	66	12	42	55	164	23	63	11	41	53	158	20	62	12	39
>1 KIR gene mismatched	328	217	104	61	105	75	318	210	97	57	105	75	311	207	100	58	101	74
Individual KIR gene matching																		
KIR2DL1 matched	372	381	125	127	113	115	361	370	118	120	112	114	353	362	118	120	109	111
KIR2DL1 mismatched	17	8	5	3	4	2	17	8	5	3	4	2	16	7	5	3	4	2
KIR2DL2 matched	202	299	71	98	58	93	196	290	67	91	57	92	192	287	67	94	57	90
KIR2DL2 mismatched	186	89	58	31	59	24	181	87	55	31	59	24	176	81	55	28	56	23
KIR2DL3 matched	328	356	110	120	97	105	320	347	105	115	96	104	310	337	103	113	93	101
KIR2DL3 mismatched	60	32	19	9	20	12	57	30	17	7	20	12	58	31	19	9	20	12
KIR2DL5 matched	186	293	67	106	56	82	180	283	65	100	55	81	178	278	61	99	55	78
KIR2DL5 mismatched	203	96	63	24	61	35	198	95	68	23	61	35	191	91	62	24	58	35
KIR2DS1 matched	196	298	64	108	52	79	193	288	64	102	51	78	189	281	60	101	51	75
KIR2DS1 mismatched	191	89	65	21	65	38	183	88	58	20	65	38	178	86	62	21	62	38
KIR2DS2 matched	201	301	71	100	57	93	195	292	67	93	56	92	191	289	67	96	56	90
KIR2DS2 mismatched	187	87	58	29	60	24	182	85	55	29	60	24	177	79	55	26	57	23
KIR2DS3 matched	227	314	82	108	72	97	223	307	79	102	71	96	215	300	77	103	69	93
KIR2DS3 mismatched	161	74	47	21	45	20	154	70	43	20	45	20	153	68	45	19	44	20
KIR2DS4 matched	355	370	117	121	111	114	344	359	110	114	110	113	335	350	110	114	107	110
KIR2DS4 mismatched	30	16	10	6	6	3	30	16	10	6	6	3	30	16	10	6	6	3
KIR2DS5 matched	199	292	67	104	55	80	195	282	66	98	54	79	193	275	63	97	55	76
KIR2DS5 mismatched	188	95	61	24	62	37	181	94	55	23	62	37	174	92	58	24	58	37
KIR3DL1 matched	358	371	118	122	111	114	347	360	111	115	110	113	338	351	111	115	107	110
KIR3DL1 mismatched	29	16	10	6	6	3	29	16	10	6	6	3	29	16	10	6	6	3
KIR3DS1 matched	198	299	69	110	53	80	195	288	68	103	52	79	190	282	65	103	52	76
KIR3DS1 mismatched	189	89	60	19	64	37	181	89	54	19	64	37	177	86	57	19	61	37
KIR gene function																		
Inhibitory KIR matched	84	199	35	74	21	54	81	193	33	70	20	53	80	189	32	70	21	51
Inhibitory KIR mismatched	303	188	93	54	96	63	295	183	88	51	96	63	287	178	89	51	92	62
Activating KIR matched	58	191	25	72	12	50	57	185	25	67	11	49	55	180	22	68	12	47
Activating KIR mismatched	326	194	102	55	105	67	316	189	95	53	105	67	309	185	98	52	101	66
KIR haplotype structure																		
Matched	219	-	75	-	62	-	211	-	71	-	61	-	208	-	69	-	60	-
Mismatched	170	-	55	-	55	-	167	-	52	-	55	-	161	-	54	-	53	-

Supplementary Table E. Subgroup population sizes for donor KIR analysis on the overall and adult, HLA-matched RIC and MAC cohorts

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD			Infectious mortality		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
Donor KIR haplotype structure												
KIR AA	119	41	35	119	41	35	111	38	35	117	40	35
KIR BX	286	94	84	274	86	83	273	89	80	269	85	80
Donor centromeric motif structure												
Cen-AA	189	65	54	186	64	54	174	58	52	182	62	54
Cen-AB	178	57	53	171	52	52	173	56	51	170	52	51
Cen-BB	37	12	12	35	10	12	36	12	12	33	10	10
Donor telomeric motif structure												
Tel-AA	230	70	74	229	70	73	222	67	74	226	69	72
Tel-AB	156	57	42	146	50	42	144	53	38	142	49	40
Tel-BB	18	7	3	17	6	3	17	6	3	17	6	3
Donor KIR B content score												
Neutral	276	88	86	272	87	85	261	81	84	267	85	84
Better	91	34	21	85	29	21	86	33	19	85	29	21
Best	37	12	12	35	10	12	36	12	12	33	10	10

Supplementary Table F. Comparison of transplant characteristics between cohorts

Factor	Entire cohort (%)	RIC, HLA-matched adults (%)	MAC, HLA-matched, adults (%)
Donor age			
<30 years	128 (32)	52 (39)	39 (33)
>30 years	274 (68)	82 (61)	80 (67)
Missing	3 (1)	1 (1)	0 (0)
Recipient age[1,2,3]			
<40 years	196 (48)	13 (10)	85 (71)
>40 years	209 (52)	122 (90)	34 (29)
Donor gender			
Male	325 (80)	116 (86)	102 (86)
Female	80 (20)	19 (14)	17 (14)
TCD by alemtuzumab			
Yes	343 (85)	117 (87)	97 (82)
No	17 (4)	6 (4)	6 (5)
Missing	45 (11)	12 (9)	16 (13)
Donor-recipient gender matching			
Female-Male	38 (9)	7 (5)	10 (8)
Male-Male	194 (48)	69 (51)	63 (53)
Female-Female	42 (10)	12 (9)	7 (6)
Male-Female	131 (32)	47 (35)	39 (33)
Conditioning regimen[4,5,6]			
MAC	217 (54)	0 (0)	119 (100)
RIC	183 (45)	135 (100)	0 (0)
Missing	5 (1)	0 (0)	0 (0)
EMBT disease risk score			
Good	179 (44)	68 (50)	51 (43)
Intermediate	174 (43)	51 (38)	52 (44)
Poor	47 (12)	15 (11)	15 (13)
Missing	5 (1)	1 (1)	1 (1)
Donor-Recipient CMV matching			
Matched	275 (68)	93 (69)	91 (76)
Mismatched	118 (29)	37 (27)	26 (22)
Missing	12 (3)	5 (4)	2 (2)
Previous autograft transplant			
0 previous autografts	381 (94)	126 (93)	112 (94)
1 or more previous autografts	24 (6)	9 (7)	7 (6)
Recipient gender			
Male	232 (57)	76 (56)	73 (61)
Female	173 (43)	59 (44)	46 (39)
HSC source[7,8]			
BM	175 (43)	40 (30)	54 (45)
PBSC	228 (56)	94 (70)	65 (55)
Missing	2 (1)	1 (1)	0 (0)
Transplant era[9,10]			
1996-1999	35 (9)	1 (1)	15 (13)
2000-2003	127 (31)	26 (19)	44 (37)
2004-2007	131 (32)	51 (38)	39 (33)
2008-2011	112 (28)	57 (42)	21 (18)
HLA matching[11,12]			
10/10 HLA-matched	287 (71)	135 (100)	119 (100)
9/10 HLA-matched	94 (23)	0 (0)	0 (0)
<9/10 HLA-matched	24 (6)	0 (0)	0 (0)

[1,4,11] Statistically significant difference between entire cohort and MAC, HLA-matched, adult cohort.

[2,5,7,9,12] Statistically significant difference between entire cohort and RIC, HLA-matched, adult cohort.

[3,6,8,10] Statistically significant difference between MAC, HLA-matched, adult cohort and RIC, HLA-matched, adult cohort

[1] $p<0.001$, [2] $p<0.001$, [3] $p<0.001$, [4] $p<0.001$, [5] $p<0.001$, [6] $p<0.001$, [7] $p=0.006$, [8] $p=0.011$, [9] $p<0.001$, [10] $p<0.001$, [11] $p<0.001$, [12] $p<0.001$.

Supplementary Table G. ENA accession numbers of sequences used to create consensus sequences for primer design

KIR2DL1	KIR2DL2	KIR2DL3	KIR2DL4	KIR2DL5	KIR2DP1	KIR2DS1	KIR2DS2	KIR2DS3	KIR2DS4	KIR2DS5	KIR3DL1	KIR3DL2	KIR3DL3	KIR3DP1	KIR3DS1
AC011501	AL133414	AC011501	AC011501	AL133414	AC011501	AL133414	AL133414	AY320039	AC011501	AL133414	AC011501	AC011501	AC006293	AL133414	CU459006
AY320039	AY320039	CU041368	GU182338	AY320039	AY320039	AY320039	CU464060	GU182339	CU151839	AY320039	CU151839	AY320039	AL133414	CU041368	GU182347
CU041368	CU464060	CU459007	GU182340	AY320039	CU041368	CU459006	CU467816	GU182347	CU464062	CU459006	EU267269	CU151839	AY320039	CU459007	GU182352
CU459007	CU467816	GU182338	GU182341	GU182339	CU459007	GU182339	GU182339	GU182351	GU182338	GU182343	GU182338	CU464063	GU182338	GU182338	GU182359
GU182338	GU182339	GU182340	GU182342	GU182339	GU182338	GU182359	GU182341	GU182359	GU182340	GU182345	GU182340	GU182338	GU182339	GU182341	GU182361
GU182342	GU182341	GU182342	GU182343	GU182345	GU182340	KP420440	GU182345	KP420440	GU182341	GU182355	GU182341	GU182339	GU182340	GU182342	KP420440
GU182343	GU182345	GU182343	GU182345	GU182347	GU182342	KP420441	GU182349		GU182343	GU182357	GU182342	GU182340	GU182341	GU182343	
GU182344	GU182347	GU182344	GU182346	GU182352	GU182343	KP420443	GU182351		GU182346	KP420441	GU182343	GU182341	GU182342	GU182344	
GU182345	GU182349	GU182346	GU182349	GU182355	GU182344		GU182353		GU182349		GU182344	GU182342	GU182344	GU182345	
GU182346	GU182351	GU182348	GU182354	GU182357	GU182345		GU182355		GU182350		GU182348	GU182343	GU182345	GU182347	
GU182348	GU182353	GU182350	GU182355	GU182359	GU182346		GU182357		GU182351		GU182349	GU182346	GU182346	GU182348	
GU182350	GU182355	GU182352	GU182356	KP420440	GU182347		GU182359		GU182353		GU182350	GU182348	GU182349	GU182349	
GU182351	GU182357	GU182354	GU182358	KP420440	GU182348		KP420437		GU182356		GU182351	GU182349	GU182350	GU182350	
GU182352	GU182359	GU182356	GU182360	KP420441	GU182350		KP420440		GU182357		GU182353	GU182350	GU182351	GU182351	
GU182354	KP420440	GU182358	GU182361		GU182351		KP420441		GU182358		GU182354	GU182352	GU182352	GU182352	
GU182356	KP420441	GU182360	KP420437		GU182352		KP420444		GU182360		GU182356	GU182353	GU182353	GU182353	
GU182357	KP420444	KP420438	KP420438		GU182354				KP420437		GU182358	GU182354	GU182354	GU182354	
GU182358		KP420439	KP420439		GU182356				KP420438		KP420437	GU182355	GU182355	GU182355	
GU182360		KP420442	KP420440		GU182357				KP420439		KP420438	GU182356	GU182356	GU182356	
GU182361		KP420443	KP420442		GU182358				KP420442		KP420439	GU182358	GU182357	GU182357	
KP420437		KP420445	KP420444		GU182361				KP420444		KP420444	GU182362	GU182358	GU182359	
KP420438		KP420446	KP420445		KP420437				KP420445		KP420444	KP420437	GU182359	GU182361	
KP420439			KP420446		KP420438				KP420446		KP420445	KP420438	GU182360	KP420437	
KP420440					KP420439						KP420446	KP420439	KP420437	KP420438	
KP420442					KP420440							KP420440	KP420438	KP420439	
KP420445					KP420442							KP420441	KP420439	KP420440	
KP420446					KP420443							KP420442	KP420440	KP420442	
					KP420445							KP420443	KP420441	KP420444	
					KP420446							KP420444	KP420442	KP420445	
												KP420445	KP420443	KP420446	
												KP420446	KP420445		
													KP420445		
													KP420446		

Supplementary Table H. Summary of IHIW cell line KIR CDS allele typing concordance

Locus	No. samples	Concordant samples (%)	Concordant alleles (%)	Unique alleles observed
KIR2DL1	41	38 (93)	64 (94)	7
KIR2DL2/3	32	20 (63)	52 (81)	7
KIR2DL4	24	18 (75)	42 (86)	10
KIR2DL5	21	17 (81)	26 (84)	6
KIR2DS2	9	9 (100)	11 (100)	2
KIR2DS5	10	5 (50)	5 (50)	3
KIR3DL1/S1	33	27 (82)	60 (90)	12
KIR3DL2	11	10 (91)	21 (95)	9
KIR3DL3	25	17 (68)	36 (72)	29
KIR3DP1	23	11 (48)	32 (68)	10

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL1*00101	2DL1*0010101		2DL1*0010101	Confirmed	2DL1*0010101 (LT984761)
2DL1*00101	2DL1*0010101	6782delT	2DL1*00101_var4	Confirmed	N/A
2DL1*00101	2DL1*0010101	6384insTA	2DL1*00101_unconfirmed_var1	Unconfirmed	N/A
2DL1*00101	2DL1*0010102		2DL1*0010102	Confirmed	2DL1*0010102 (LT984768)
2DL1*00101	2DL1*0010102	6782delT	2DL1*00101_var5	Confirmed	N/A
2DL1*00101	2DL1*0010102	4027A>G	2DL1*00101_unconfirmed_var3	Unconfirmed	N/A
2DL1*00101	2DL1*0010102	6048delT;6782delT;7909G>A	2DL1*00101_unconfirmed_var7	Unconfirmed	N/A
2DL1*00101	2DL1*0010102	6048delT;6405delT;6782delT	2DL1*00101_unconfirmed_var8	Unconfirmed	N/A
2DL1*00101	2DL1*0010103		2DL1*0010103	Confirmed	2DL1*0010103 (LT984767)
2DL1*00101	2DL1*0010103	6782delT	2DL1*00101_var6	Confirmed	N/A
2DL1*00101	2DL1*0010103	5298T>C;6782delT	2DL1*00101_var7	Confirmed	N/A
2DL1*00101	2DL1*0010103	5298T>C	2DL1*00101_unconfirmed_var2	Unconfirmed	N/A
2DL1*00101	2DL1*0010103	10297C>T	2DL1*00101_unconfirmed_var4	Unconfirmed	N/A
2DL1*00101	2DL1*0010103	-127A>G;2288G>C;2853T>C;6854G>C;10551C>A;11771T>C;11948A>G;12360G>C;13181T>C;14191A>G	2DL1*00101_unconfirmed_var5	Unconfirmed	N/A
2DL1*00101	2DL1*0010103	3255C>G	2DL1*00101_unconfirmed_var6	Unconfirmed	N/A
2DL1*00101	2DL1*0010103	716delGATATGGGCCTGGATTGGC;6782delT	2DL1*00101_unconfirmed_var9	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T	2DL1*00201_var1	Confirmed	2DL1*0020105 (LR593931)
2DL1*00201	2DL1*0020101	625C>T;1333G>A;3045T>C	2DL1*00201_var5	Confirmed	2DL1*0020116 (LR593981)
2DL1*00201	2DL1*0020101	625C>T;3045T>C	2DL1*00201_var6	Confirmed	2DL1*0020107 (LR593941)
2DL1*00201	2DL1*0020101	625C>T;6613C>G	2DL1*00201_var9	Confirmed	2DL1*0020113 (LR593936)
2DL1*00201	2DL1*0020101	625C>T;1333G>A;3045T>C;6782delT	2DL1*00201_var13	Confirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;6782delT	2DL1*00201_var15	Confirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;3446G>C;6782delT	2DL1*00201_var17	Confirmed	N/A
2DL1*00201	2DL1*0020101	214G>C;625C>T;6782delT	2DL1*00201_var22	Confirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;6384T>A;6388insAC;13294C>G	2DL1*00201_unconfirmed_var1	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;1333G>A;3045T>C;11784C>T;11785C>G	2DL1*00201_unconfirmed_var4	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;13191G>C	2DL1*00201_unconfirmed_var5	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;11850C>G	2DL1*00201_unconfirmed_var9	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;7580T>A	2DL1*00201_unconfirmed_var11	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;1333G>A;3045T>C;12041C>G	2DL1*00201_unconfirmed_var13	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;13294C>G	2DL1*00201_unconfirmed_var14	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	214G>C;625C>T	2DL1*00201_unconfirmed_var16	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;7214T>C	2DL1*00201_unconfirmed_var18	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;14087A>C;14112T>C;14122A>T	2DL1*00201_unconfirmed_var19	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;3045T>C;6782delT	2DL1*00201_unconfirmed_var22	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;6782delT;7214T>C	2DL1*00201_unconfirmed_var25	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;6048delT	2DL1*00201_unconfirmed_var26	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	-264delG;625C>T;1333G>A;3045T>C	2DL1*00201_unconfirmed_var34	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;6405delT;6782delT	2DL1*00201_unconfirmed_var40	Unconfirmed	N/A
2DL1*00201	2DL1*0020101	625C>T;5195delGA;6782delT	2DL1*00201_unconfirmed_var43	Unconfirmed	N/A
2DL1*00201	2DL1*0020102		2DL1*0020102	Confirmed	N/A
2DL1*00201	2DL1*0020102	9100A>C	2DL1*00201_var2	Confirmed	2DL1*0020108 (LR593954)
2DL1*00201	2DL1*0020102	5710C>T	2DL1*00201_var3	Confirmed	2DL1*0020106 (LR593933)
2DL1*00201	2DL1*0020102	10294C>T;12712A>T	2DL1*00201_var4	Confirmed	2DL1*0020114 (LR593947)
2DL1*00201	2DL1*0020102	8606G>A	2DL1*00201_var7	Confirmed	2DL1*0020115 (LR593974)
2DL1*00201	2DL1*0020102	6101C>A;10294C>T;12712A>T	2DL1*00201_var8	Confirmed	2DL1*0020110 (LR593963)
2DL1*00201	2DL1*0020102	3174G>C;10294C>T;12712A>T	2DL1*00201_var10	Confirmed	2DL1*0020109 (LR593958)
2DL1*00201	2DL1*0020102	5892C>T	2DL1*00201_var11	Confirmed	2DL1*0020112 (LR593897)
2DL1*00201	2DL1*0020102	6782delT	2DL1*00201_var14	Confirmed	N/A
2DL1*00201	2DL1*0020102	6782delT;10294C>T;12712A>T	2DL1*00201_var16	Confirmed	N/A
2DL1*00201	2DL1*0020102	5710C>T;6782delT	2DL1*00201_var18	Confirmed	N/A
2DL1*00201	2DL1*0020102	5892C>T;6782delT	2DL1*00201_var19	Confirmed	N/A
2DL1*00201	2DL1*0020102	3973A>G;6782delT	2DL1*00201_var20	Confirmed	N/A
2DL1*00201	2DL1*0020102	2688T>G	2DL1*00201_unconfirmed_var3	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	6101C>A;7438T>C;10294C>T;12712A>T	2DL1*00201_unconfirmed_var6	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	2865T>G	2DL1*00201_unconfirmed_var7	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL1*00201	2DL1*0020102	2126A>G	2DL1*00201_unconfirmed_var8	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	8860G>A;10294C>T;12712A>T	2DL1*00201_unconfirmed_var10	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	4963G>T;9100A>C	2DL1*00201_unconfirmed_var12	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	13292G>A	2DL1*00201_unconfirmed_var15	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	10076G>T	2DL1*00201_unconfirmed_var17	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	3174G>C;6782delT;10294C>T;12712A>T	2DL1*00201_unconfirmed_var21	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	6782delT;10076G>T	2DL1*00201_unconfirmed_var23	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	3409C>T;6782delT;12712A>T	2DL1*00201_unconfirmed_var24	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	5806G>A;6782delT	2DL1*00201_unconfirmed_var27	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	6782delT;7309G>A	2DL1*00201_unconfirmed_var28	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	575G>A;2123C>G;6782delT;10294C>T;12712A>T	2DL1*00201_unconfirmed_var29	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	6101C>A;6782delT;10294C>T;12712A>T	2DL1*00201_unconfirmed_var33	Unconfirmed	N/A
2DL1*00201	2DL1*0020102	5892C>T;6048delT;6782delT	2DL1*00201_unconfirmed_var41	Unconfirmed	N/A
2DL1*00201	2DL1*0020103		2DL1*0020103	Confirmed	N/A
2DL1*00201	2DL1*0020103	7846G>C	2DL1*00201_var12	Confirmed	2DL1*0020104 (LR593907)
2DL1*00201	2DL1*0020103	1056C>T	2DL1*00201_var21	Confirmed	2DL1*0020111 (LR593986)
2DL1*00201	2DL1*0020103	6783delT	2DL1*00201_var23	Confirmed	N/A
2DL1*00201	2DL1*0020103	6783delT;7846G>C	2DL1*00201_var24	Confirmed	N/A
2DL1*00201	2DL1*0020103	1056C>T;6783delT	2DL1*00201_var25	Confirmed	N/A
2DL1*00201	2DL1*0020103	1056C>T;6385insTA	2DL1*00201_unconfirmed_var2	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	1548G>A	2DL1*00201_unconfirmed_var20	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	2618G>C;7846G>C	2DL1*00201_unconfirmed_var30	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	7151T>A	2DL1*00201_unconfirmed_var31	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	3842A>C;7846G>C	2DL1*00201_unconfirmed_var32	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	1056C>T;11700A>G	2DL1*00201_unconfirmed_var35	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	6918A>G	2DL1*00201_unconfirmed_var36	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	543A>G;1056C>T;6783delT;7426A>G	2DL1*00201_unconfirmed_var38	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	4223T>C;6783delT;7846G>C	2DL1*00201_unconfirmed_var39	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	-262delG	2DL1*00201_unconfirmed_var42	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	4215delAT;7846G>C	2DL1*00201_unconfirmed_var44	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	4215delAT;6783delT;7846G>C	2DL1*00201_unconfirmed_var45	Unconfirmed	N/A
2DL1*00201	2DL1*0020103	5195delGA;6783delT	2DL1*00201_unconfirmed_var46	Unconfirmed	N/A
2DL1*00201_c.110C>G	2DL1*0020101	625C>T;3522C>G	2DL1*00201_c.110C>G_var1	Confirmed	2DL1*0020101 (LR593949)
2DL1*00201_c.169C>G	2DL1*0020101	625C>T;1333G>A;3045T>C;3581C>G	2DL1*00201_c.169C>G_var2	Confirmed	2DL1*0020101 (LR593917)
2DL1*00201_c.22A>G	2DL1*0020102	22A>G	2DL1*00201_c.22A>G_var3	Confirmed	2DL1*0020102 (LR593914)
2DL1*00201_c.568G>A	2DL1*0020101	625C>T;1333G>A;3045T>C;5509G>A	2DL1*00201_c.568G>A_var4	Confirmed	2DL1*0020101 (LR593946)
2DL1*00201_c.856A>T	2DL1*0020102	6101C>A;10294C>T;12712A>T;13672A>T	2DL1*00201_c.856A>T_var5	Confirmed	2DL1*0020102 (LR593934)
2DL1*00201_c.950G>T	2DL1*0020102	13864G>T	2DL1*00201_c.950G>T_var14	Confirmed	2DL1*0020102 (LR593966)
2DL1*00302	2DL1*0030201	6783delT	2DL1*00302_unconfirmed_var62	Unconfirmed	N/A
2DL1*00302	2DL1*0030203	8710A>G	2DL1*00302_var35	Confirmed	2DL1*0030221 (LR593912)
2DL1*00302	2DL1*0030204		2DL1*0030204	Confirmed	N/A
2DL1*00302	2DL1*0030204	4121C>T	2DL1*00302_var9	Confirmed	2DL1*0030220 (LR593891)
2DL1*00302	2DL1*0030204	1962A>T;8653T>C;9465G>A;9523G>A;9558C>T;9561G>A;14122T>A	2DL1*00302_var11	Confirmed	2DL1*0030225 (LR593964)
2DL1*00302	2DL1*0030204	12917T>A	2DL1*00302_var14	Confirmed	2DL1*0030218 (LR594002)
2DL1*00302	2DL1*0030204	2546T>C	2DL1*00302_var18	Confirmed	2DL1*0030227 (LR593989)
2DL1*00302	2DL1*0030204	1962A>T;4990C>T;7438T>C;8653T>C;9465G>A;9558C>T;14122T>A	2DL1*00302_var20	Confirmed	2DL1*0030222 (LR593918)
2DL1*00302	2DL1*0030204	7708T>C;10492A>C	2DL1*00302_var21	Confirmed	2DL1*0030204 (LR594003)
2DL1*00302	2DL1*0030204	6783delT	2DL1*00302_var24	Confirmed	N/A
2DL1*00302	2DL1*0030204	3842A>C;6783delT;8200G>A	2DL1*00302_var28	Confirmed	N/A
2DL1*00302	2DL1*0030204	6783delT;12917T>A	2DL1*00302_var30	Confirmed	N/A
2DL1*00302	2DL1*0030204	6783delT;7708T>C;10492A>C	2DL1*00302_var31	Confirmed	N/A
2DL1*00302	2DL1*0030204	1962A>T;6783delT;7302T>G;8653T>C;9465G>A;9523G>A;9558C>T;9561G>A;14122T>A	2DL1*00302_var33	Confirmed	N/A
2DL1*00302	2DL1*0030204	1962A>T;6783delT;8653T>C;9465G>A;9523G>A;9558C>T;9561G>A;14122T>A	2DL1*00302_var36	Confirmed	N/A
2DL1*00302	2DL1*0030204	2118C>T;6783delT;7708T>C;10309G>C;10492A>C	2DL1*00302_var42	Confirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL1*00302	2DL1*0030204	6405delT;6783delT	2DL1*00302_var45	Confirmed	N/A
2DL1*00302	2DL1*0030204	4870delGATA	2DL1*00302_var46	Confirmed	N/A
2DL1*00302	2DL1*0030204	6385insTA;14122T>A	2DL1*00302_unconfirmed_var1	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6196G>A	2DL1*00302_unconfirmed_var3	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6048delT;6385insTA;6783delT;14122T>A	2DL1*00302_unconfirmed_var5	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	11493G>A	2DL1*00302_unconfirmed_var6	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	3067C>T;3842A>C;8200G>A	2DL1*00302_unconfirmed_var7	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	3842A>C;8200G>A	2DL1*00302_unconfirmed_var11	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	1962A>T;6384T>A;8653T>C;9465G>A;9523G>A;9558C>T;9561G>A;14122T>A	2DL1*00302_unconfirmed_var12	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	9231C>T	2DL1*00302_unconfirmed_var13	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6085A>C	2DL1*00302_unconfirmed_var16	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	13704C>T	2DL1*00302_unconfirmed_var17	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	12981G>A	2DL1*00302_unconfirmed_var19	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	1962A>T;4547C>G;8653T>C;9465G>A;9523G>A;9558C>T;9561G>A;14122T>A	2DL1*00302_unconfirmed_var22	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	2093A>G;9070A>G	2DL1*00302_unconfirmed_var25	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	815T>C	2DL1*00302_unconfirmed_var26	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	7417T>G	2DL1*00302_unconfirmed_var27	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	2865T>A;2872A>G	2DL1*00302_unconfirmed_var28	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6328G>C	2DL1*00302_unconfirmed_var29	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	-169G>A	2DL1*00302_unconfirmed_var32	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	5084C>G	2DL1*00302_unconfirmed_var33	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	2476A>G	2DL1*00302_unconfirmed_var34	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	2118C>T;4685G>A;7708T>C;10309G>C;10492A>C;11814G>C	2DL1*00302_unconfirmed_var35	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6326C>T;6783delT	2DL1*00302_unconfirmed_var37	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	1100G>C;4573T>A;5969A>C;6783delT	2DL1*00302_unconfirmed_var39	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6519T>G;6783delT;12474C>G;12477T>C;13746G>A	2DL1*00302_unconfirmed_var40	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	1962A>T;4990C>T;6783delT;7438T>C;8653T>C;9465G>A;9558C>T;14122T>A	2DL1*00302_unconfirmed_var41	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	1962A>T;6783delT;8653T>C;8667G>C;9465G>A;9523G>A;9558C>T;9561G>A;14122T>A	2DL1*00302_unconfirmed_var42	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	1962A>T;4990C>T;6783delT;7438T>C;8653T>C;9465G>A;9558C>T;13538C>G;13540A>T;14122T>A	2DL1*00302_unconfirmed_var47	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	1962A>T;6783delT;8653T>C;9465G>A;9523G>A;9558C>T;9561G>A;10999T>G;14122T>A	2DL1*00302_unconfirmed_var48	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6783delT;10788G>A	2DL1*00302_unconfirmed_var49	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	5674G>T;6783delT	2DL1*00302_unconfirmed_var56	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	3043C>T;6783delT	2DL1*00302_unconfirmed_var59	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	3067C>T;3842A>C;6783delT;8200G>A	2DL1*00302_unconfirmed_var60	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6783delT;7417T>G	2DL1*00302_unconfirmed_var61	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6783delT;10208G>A	2DL1*00302_unconfirmed_var64	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6405delT;6783delT;13704C>T	2DL1*00302_unconfirmed_var67	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6048delT;6783delT	2DL1*00302_unconfirmed_var68	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	6048delT;6783delT;7417T>G	2DL1*00302_unconfirmed_var70	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	5195delGA	2DL1*00302_unconfirmed_var71	Unconfirmed	N/A
2DL1*00302	2DL1*0030204	4870delGATA;6783delT	2DL1*00302_unconfirmed_var75	Unconfirmed	N/A
2DL1*00302	2DL1*0030205		2DL1*0030205	Confirmed	N/A
2DL1*00302	2DL1*0030205	6782insT;12754A>G;12756G>C	2DL1*00302_var1	Confirmed	N/A
2DL1*00302	2DL1*0030205	4946G>A	2DL1*00302_var2	Confirmed	2DL1*0030219 (LR593889)
2DL1*00302	2DL1*0030205	2372T>G;6384T>A;11898G>A	2DL1*00302_var4	Confirmed	2DL1*0030223 (LR593940)
2DL1*00302	2DL1*0030205	5138G>A;6980C>T;12561C>T	2DL1*00302_var5	Confirmed	2DL1*0030216 (LR593996)
2DL1*00302	2DL1*0030205	12754A>G;12756G>C	2DL1*00302_var6	Confirmed	2DL1*0030212 (LR593911)
2DL1*00302	2DL1*0030205	5138G>A;12561C>T	2DL1*00302_var7	Confirmed	2DL1*0030211 (LR593909)
2DL1*00302	2DL1*0030205	3114C>A	2DL1*00302_var8	Confirmed	2DL1*0030229 (LR593997)
2DL1*00302	2DL1*0030205	2372T>G;11898G>A	2DL1*00302_var10	Confirmed	2DL1*0030231 (LR594001)
2DL1*00302	2DL1*0030205	5138G>A;8057A>C;12561C>T	2DL1*00302_var12	Confirmed	2DL1*0030205 (LR594005)
2DL1*00302	2DL1*0030205	2372T>G;6186G>A;6519T>G;11898G>A	2DL1*00302_var16	Confirmed	2DL1*0030224 (LR593953)
2DL1*00302	2DL1*0030205	-28G>A;4946G>A	2DL1*00302_var17	Confirmed	2DL1*0030230 (LR593999)

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL1*00302	2DL1*0030205	2372T>G;10793A>T;11898G>A	2DL1*00302_var19	Confirmed	2DL1*0030226 (LR593987)
2DL1*00302	2DL1*0030205	2372T>G;6186G>A;6519T>G;10452A>T;11898G>A	2DL1*00302_var22	Confirmed	2DL1*0030228 (LR593995)
2DL1*00302	2DL1*0030205	2853C>T	2DL1*00302_var23	Confirmed	2DL1*0030215 (LR593988)
2DL1*00302	2DL1*0030205	-28G>A;4946G>A;6783delT	2DL1*00302_var25	Confirmed	N/A
2DL1*00302	2DL1*0030205	2372T>G;6186G>A;6519T>G;6783delT;11898G>A	2DL1*00302_var26	Confirmed	N/A
2DL1*00302	2DL1*0030205	6783delT;12754A>G;12756G>C	2DL1*00302_var27	Confirmed	N/A
2DL1*00302	2DL1*0030205	5138G>A;6783delT;12561C>T	2DL1*00302_var29	Confirmed	N/A
2DL1*00302	2DL1*0030205	4946G>A;6783delT	2DL1*00302_var37	Confirmed	N/A
2DL1*00302	2DL1*0030205	2372T>G;6384T>A;6783delT;11898G>A	2DL1*00302_var38	Confirmed	N/A
2DL1*00302	2DL1*0030205	2372T>G;6783delT;10793A>T;11898G>A	2DL1*00302_var40	Confirmed	N/A
2DL1*00302	2DL1*0030205	3114C>A;6783delT	2DL1*00302_var41	Confirmed	N/A
2DL1*00302	2DL1*0030205	4946G>A;12754A>G;12756G>C	2DL1*00302_unconfirmed_var2	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	6121C>T;12754A>G;12756G>C	2DL1*00302_unconfirmed_var4	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	12754A>G;12756G>C;12788A>G	2DL1*00302_unconfirmed_var8	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	4946G>A;6172A>G	2DL1*00302_unconfirmed_var9	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	4946G>A;8598C>T	2DL1*00302_unconfirmed_var14	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	10883G>C;12754A>G;12756G>C	2DL1*00302_unconfirmed_var15	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	2559G>A;12754A>G;12756G>C	2DL1*00302_unconfirmed_var21	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	2990C>G;12754A>G;12756G>C	2DL1*00302_unconfirmed_var23	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	10691A>G;12376A>G	2DL1*00302_unconfirmed_var24	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	375T>C;11194T>A	2DL1*00302_unconfirmed_var30	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	-28G>A;1642G>A;4946G>A;6783delT	2DL1*00302_unconfirmed_var38	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	1806G>T;6783delT;12754A>G;12756G>C	2DL1*00302_unconfirmed_var45	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	2372T>G;6186G>A;6519T>G;6783delT;7431G>A;11898G>A	2DL1*00302_unconfirmed_var46	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	2372T>G;6186G>A;6519T>G;6783delT;10452A>T;11898G>A	2DL1*00302_unconfirmed_var50	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	4946G>A;6172A>G;6783delT	2DL1*00302_unconfirmed_var51	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	6405delT;12754A>G;12756G>C	2DL1*00302_unconfirmed_var53	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	5138G>A;6783delT;6980C>T;12561C>T	2DL1*00302_unconfirmed_var54	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	6783delT;10883G>C;11591A>C;12754A>G;12756G>C	2DL1*00302_unconfirmed_var55	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	2372T>G;6783delT;11898G>A	2DL1*00302_unconfirmed_var57	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	6783delT;9448T>C;11504T>C;12754A>G;12756G>C	2DL1*00302_unconfirmed_var58	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	4287C>G;6783delT;12754A>G;12756G>C	2DL1*00302_unconfirmed_var63	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	2372T>G;6186G>A;6519T>G;6783delT;11898G>A;13418C>G	2DL1*00302_unconfirmed_var65	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	5195delGA;12754A>G;12756G>C	2DL1*00302_unconfirmed_var72	Unconfirmed	N/A
2DL1*00302	2DL1*0030205	62G>A;1309delCA;6783delT	2DL1*00302_unconfirmed_var74	Unconfirmed	N/A
2DL1*00302	2DL1*0030206		2DL1*0030206	Confirmed	N/A
2DL1*00302	2DL1*0030206	5002G>A;12359G>C	2DL1*00302_var3	Confirmed	2DL1*0030214 (LR593971)
2DL1*00302	2DL1*0030206	7527C>T	2DL1*00302_var13	Confirmed	2DL1*0030213 (LR593950)
2DL1*00302	2DL1*0030206	1642G>A;6967T>C	2DL1*00302_var15	Confirmed	2DL1*0030217 (LR593998)
2DL1*00302	2DL1*0030206	6783delT	2DL1*00302_var32	Confirmed	N/A
2DL1*00302	2DL1*0030206	1642G>A;6783delT;6967T>C	2DL1*00302_var39	Confirmed	N/A
2DL1*00302	2DL1*0030206	5002G>A	2DL1*00302_unconfirmed_var31	Unconfirmed	N/A
2DL1*00302	2DL1*0030206	6405delT	2DL1*00302_unconfirmed_var43	Unconfirmed	N/A
2DL1*00302	2DL1*0030206	687C>G;890C>T;6783delT	2DL1*00302_unconfirmed_var44	Unconfirmed	N/A
2DL1*00302	2DL1*0030206	6783delT;10741C>T;10745G>T	2DL1*00302_unconfirmed_var52	Unconfirmed	N/A
2DL1*00302	2DL1*0030206	6048delT;6783delT	2DL1*00302_unconfirmed_var69	Unconfirmed	N/A
2DL1*00302	2DL1*0030207	7928G>C;9292G>C	2DL1*00302_unconfirmed_var18	Unconfirmed	N/A
2DL1*00302	2DL1*0030208		2DL1*0030208	Confirmed	N/A
2DL1*00302	2DL1*0030208	6783delT	2DL1*00302_var43	Confirmed	N/A
2DL1*00302	2DL1*0030208	2620G>T	2DL1*00302_unconfirmed_var10	Unconfirmed	N/A
2DL1*00302	2DL1*0030208	523G>A	2DL1*00302_unconfirmed_var20	Unconfirmed	N/A
2DL1*00302	2DL1*0030208	12545C>A	2DL1*00302_unconfirmed_var36	Unconfirmed	N/A
2DL1*00302	2DL1*0030209		2DL1*0030209	Confirmed	N/A
2DL1*00302	2DL1*0030209	6783delT	2DL1*00302_var34	Confirmed	N/A
2DL1*00302_c.421G>A	2DL1*0030204	5362G>A;12345C>T;12917T>A	2DL1*00302_c.421G>A_var6	Confirmed	2DL1*0030204 (LR593916)
2DL1*00302_c.842C>G	2DL1*0030206	13658C>G	2DL1*00302_c.842C>G_var7	Confirmed	2DL1*0030206 (LR593972)
2DL1*00302_c.853A>C	2DL1*0030204	12917T>A;13669A>C	2DL1*00302_c.853A>C_var8	Confirmed	2DL1*0030204 (LR593960)

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL1*00302_c.910C>G	2DL1*0030205	6783delT;7603A>G;12754A>G;12756G>C;13824C>G	2DL1*00302_c.910C>G_var9	Confirmed	N/A
2DL1*00302_c.963G>A	2DL1*0030204	7708T>C;10492A>C;13877G>A	2DL1*00302_c.963G>A_var10	Confirmed	2DL1*0030204 (LR593943)
2DL1*00401	2DL1*0040101		2DL1*0040101	Confirmed	N/A
2DL1*00401	2DL1*0040101	10079C>T	2DL1*00401_var1	Confirmed	2DL1*0040106 (LR593965)
2DL1*00401	2DL1*0040101	12702G>A	2DL1*00401_var2	Confirmed	2DL1*0040114 (LR593942)
2DL1*00401	2DL1*0040101	6736G>A	2DL1*00401_var3	Confirmed	2DL1*0040103 (LR593923)
2DL1*00401	2DL1*0040101	3951G>C;5991C>G	2DL1*00401_var4	Confirmed	2DL1*0040111 (LR593895)
2DL1*00401	2DL1*0040101	3951G>C	2DL1*00401_var5	Confirmed	2DL1*0040109 (LR594004)
2DL1*00401	2DL1*0040101	1452A>G	2DL1*00401_var6	Confirmed	2DL1*0040107 (LR593980)
2DL1*00401	2DL1*0040101	10789C>A	2DL1*00401_var7	Confirmed	2DL1*0040105 (LR593952)
2DL1*00401	2DL1*0040101	11646T>C	2DL1*00401_var8	Confirmed	2DL1*0040104 (LR593927)
2DL1*00401	2DL1*0040101	12217C>G	2DL1*00401_var9	Confirmed	2DL1*0040113 (LR593930)
2DL1*00401	2DL1*0040101	12633G>A	2DL1*00401_var10	Confirmed	2DL1*0040115 (LR593983)
2DL1*00401	2DL1*0040101	1533T>C	2DL1*00401_var11	Confirmed	2DL1*0040108 (LR593992)
2DL1*00401	2DL1*0040101	6383delTA;6389C>T;6391T>C	2DL1*00401_var12	Confirmed	N/A
2DL1*00401	2DL1*0040101	3412C>G	2DL1*00401_var14	Confirmed	2DL1*0040110 (LR594006)
2DL1*00401	2DL1*0040101	6388insAC;6389C>T	2DL1*00401_unconfirmed_var1	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	6405insT	2DL1*00401_unconfirmed_var2	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	7430C>T;9560G>T	2DL1*00401_unconfirmed_var3	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	8505T>C	2DL1*00401_unconfirmed_var5	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	12756G>C	2DL1*00401_unconfirmed_var6	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	5674T>G;12633G>A	2DL1*00401_unconfirmed_var7	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	4616G>C;4617A>T	2DL1*00401_unconfirmed_var8	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	11375C>G	2DL1*00401_unconfirmed_var9	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	8187T>C	2DL1*00401_unconfirmed_var10	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	8502A>G	2DL1*00401_unconfirmed_var11	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	6873A>C	2DL1*00401_unconfirmed_var12	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	4462C>A	2DL1*00401_unconfirmed_var13	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	3121G>A;12702G>A	2DL1*00401_unconfirmed_var14	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	7430C>T;9505C>A	2DL1*00401_unconfirmed_var15	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	13382A>C	2DL1*00401_unconfirmed_var16	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	7430C>T;10639C>G	2DL1*00401_unconfirmed_var17	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	11743C>T	2DL1*00401_unconfirmed_var18	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	5674T>G	2DL1*00401_unconfirmed_var19	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	6405delT;10789C>A	2DL1*00401_unconfirmed_var20	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	-260delG	2DL1*00401_unconfirmed_var21	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	6048delT;6405delT	2DL1*00401_unconfirmed_var22	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	6383delTA;6389C>T;6391T>C;7430C>T;10551G>T;12875C>T	2DL1*00401_unconfirmed_var23	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	5195delGA;10789C>A	2DL1*00401_unconfirmed_var24	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	6383delTA;6389C>T;6391T>C;8474G>A	2DL1*00401_unconfirmed_var25	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	5195delGA;6736G>A	2DL1*00401_unconfirmed_var26	Unconfirmed	N/A
2DL1*00401	2DL1*0040101	3306T>G;6612delTCTC	2DL1*00401_unconfirmed_var27	Unconfirmed	N/A
2DL1*00401_c.13G>T	2DL1*0040102	-197A>G;-154A>C;-119G>A;13G>T;880A>G;1250G>A;1716T>C;1760A>C;1864A>G;3138C>A;4793G>A;5195delG;5195delA;6329A>G;8221C>T;10190G>A;11061G>T;11475C>G;11631C>G;11791T>G;12917G>T;13753A>C;13995A>G;14232T>A	2DL1*00401_c.13G>T_var11	Confirmed	N/A
2DL1*00401_c.808delA	2DL1*0040101	3951G>C;5991C>G;13162delA	2DL1*00401_c.808delA_var12	Confirmed	2DL1*0040101 (LR593969)
2DL1*00402	2DL1*00402		2DL1*00402	Confirmed	2DL1*00402 (LT984784)
2DL1*00402	2DL1*00402	13995G>A	2DL1*00402_unconfirmed_var1	Unconfirmed	N/A
2DL1*007	2DL1*007	1642A>G;14164A>G	2DL1*007_var1	Confirmed	2DL1*007 (LR593967)
2DL1*007	2DL1*007	121A>G;1064delA;1642A>G;7133G>A;12217G>C	2DL1*007_var2	Confirmed	N/A
2DL1*007	2DL1*007	121A>G;890C>A;1064delA;1642A>G;7133G>A;12217G>C	2DL1*007_unconfirmed_var1	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL1*007_c.1027A>G	2DL1*007	6126C>G;13941A>G	2DL1*007_c.1027A>G_var13	Confirmed	2DL1*007 (LR593937)
2DL1*008	2DL1*008		2DL1*008	Confirmed	2DL1*008 (LT984781)
2DL1*008	2DL1*008	6782delT	2DL1*008_var2	Confirmed	N/A
2DL1*008	2DL1*008	6405delT;6782delT	2DL1*008_unconfirmed_var1	Unconfirmed	N/A
2DL1*010	2DL1*010	66T>C;68A>C;1321T>C;1452G>A;2194A>G;2198A>C;2200T>C;2201insTG;2201A>T;2202A>G;2325C>A;2326A>G;2469A>G;2497T>G;2686A>G;2965G>C;3195T>C;3412G>C;4462A>C;4533G>T;4541A>G;4699C>A;4775G>C;5066C>G;5068C>G;5661G>A;6388delA;6388delC;6389T>C;6736G>A;6946G>C;7438C>T;8175G>A;8502G>A;8911C>A;8930T>C;9135A>G;10110C>G;10321A>G;10533G>A;10590C>T;10597G>A;10649C>T;10675G>A;10686A>G;10693C>T;10716delG;10749G>A;10853G>T;10883G>C;10888G>A;11045C>G;11120C>G;11504T>C;11521T>C;11814A>G;11982C>T;12045C>T;12320A>C;12363insTCC;12394G>C;12439T>G;12505C>T;12561T>C;12663C>T;12948C>T;13254T>A;13327C>T;13339G>A;13382C>A;13623A>G;13627C>T;13693G>A;13768G>C;14292C>A;14309G>C;14335C>T;14386C>T;14468T>C	2DL1*010_unconfirmed_var1	Unconfirmed	N/A
2DL1*010	2DL1*010	6388delA;6388delC;6389T>C;8725T>C	2DL1*010_unconfirmed_var2	Unconfirmed	N/A
2DL1*01202	None available		2DL1*01202_unconfirmed_var1	Unconfirmed	N/A
2DL1*020	None available		2DL1*020_unconfirmed_var1	Unconfirmed	N/A
2DL1*020	None available		2DL1*020_unconfirmed_var2	Unconfirmed	N/A
2DL1*021	None available		2DL1*021_unconfirmed_var1	Unconfirmed	N/A
2DL1*024	None available		2DL1*024_unconfirmed_var1	Unconfirmed	N/A
2DL1*032N	2DL1*032N		2DL1*032N	Confirmed	2DL1*032N (LT984782)
2DL1*032N	2DL1*032N	6782delT	2DL1*032N_var2	Confirmed	N/A
2DL1*032N	2DL1*032N	10882G>C	2DL1*032N_unconfirmed_var1	Unconfirmed	N/A
2DL1*032N	2DL1*032N	6782delT;11565T>C	2DL1*032N_unconfirmed_var2	Unconfirmed	N/A
2DL1*034	None available		2DL1*034_unconfirmed_var1	Unconfirmed	N/A
2DL1*037	2DL1*037		2DL1*037	Confirmed	2DL1*037 (LT984773)
2DL1*037	2DL1*037	6783delT	2DL1*037_unconfirmed_var1	Unconfirmed	N/A
2DL2*00101	2DL2*0010101	6422delTATATAT	2DL2*00101_unconfirmed_var45	Unconfirmed	N/A
2DL2*00101	2DL2*0010101	5966delT;6376delTATATA	2DL2*00101_unconfirmed_var46	Unconfirmed	N/A
2DL2*00101	2DL2*0010102	6423A>T;6424delT;6424delT;6574G>A	2DL2*00101_unconfirmed_var1	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6423delATA;6427A>T;9989T>G	2DL2*00101_var2	Confirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6423delATA;6427A>T;9989T>G	2DL2*00101_var5	Confirmed	N/A
2DL2*00101	2DL2*0010103	6421delATATA;6427A>T;9989T>G	2DL2*00101_var6	Confirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6421delATATA;6427A>T;9989T>G	2DL2*00101_var8	Confirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6423delATATA;9989T>G;12647G>A	2DL2*00101_var9	Confirmed	N/A
2DL2*00101	2DL2*0010103	6421delATATATA;9989T>G	2DL2*00101_var10	Confirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6419delATATATA;6427A>T;9989T>G;11563T>C	2DL2*00101_var11	Confirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6421delATATATA;9989T>G	2DL2*00101_var12	Confirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6419delATATATA;6427A>T;9989T>G	2DL2*00101_var13	Confirmed	N/A
2DL2*00101	2DL2*0010103	-137C>A;5966delT;6377A>G;6421delATATATA;9989T>G;11563T>C	2DL2*00101_var14	Confirmed	N/A
2DL2*00101	2DL2*0010103	6419delATATATATA;9989T>G;11563T>C	2DL2*00101_var16	Confirmed	N/A
2DL2*00101	2DL2*0010103	6419delATATATATA;9989T>G	2DL2*00101_var17	Confirmed	N/A
2DL2*00101	2DL2*0010103	6417delATATATATATA;9989T>G;11563T>C	2DL2*00101_var19	Confirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6417delATATATATATA;9989T>G;11563T>C;11903G>A	2DL2*00101_var20	Confirmed	N/A
2DL2*00101	2DL2*0010103	6425delA;6427A>T;9989T>G	2DL2*00101_unconfirmed_var5	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	4595G>A;5966delT;6425delA;6427A>T;9989T>G	2DL2*00101_unconfirmed_var7	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6376delTA;6425A>T;6427A>T;8913A>G;9989T>G;11563T>C	2DL2*00101_unconfirmed_var11	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	3277A>T;5966delT;6423delATA;6427A>T;8913A>G;9989T>G;11563T>C	2DL2*00101_unconfirmed_var14	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6419delATATA;6425A>T;6427A>T;7071C>A;9989T>G	2DL2*00101_unconfirmed_var16	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5601T>G;6421delATATA;6427A>T;9989T>G;10824G>A	2DL2*00101_unconfirmed_var17	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	1216G>A;6423delATATA;9989T>G;10132C>T;11563T>C;12775C>T	2DL2*00101_unconfirmed_var19	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6311delAT;6423delATA;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var21	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL2*00101	2DL2*0010103	4146insAT;4149A>G;4160G>C;6421delATATATA;9989T>G;11563T>C;13518G>A	2DL2*00101_unconfirmed_var22	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	4146insAT;4149A>G;4160G>C;6376delTATATA;6427A>T;8576T>C;9989T>G;10029delA;11563T>C;13518G>A	2DL2*00101_unconfirmed_var23	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6419delATATA;6425A>T;6427A>T;8913A>G;9989T>G;11563T>C	2DL2*00101_unconfirmed_var24	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	-137C>A;5966delT;6377A>G;6423delATATA;9989T>G;11563T>C	2DL2*00101_unconfirmed_var25	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6423delATATA;9989T>G	2DL2*00101_unconfirmed_var26	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	4146insAT;4149A>G;4160G>C;5113delGA;5966delT;6423delATATA;9989T>G;11563T>C;13518G>A	2DL2*00101_unconfirmed_var28	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6423delATATA;9362C>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var30	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5113insGA;5966delT;6311delAT;6419delATATA;6425A>T;6427A>T;7071C>A;9989T>G	2DL2*00101_unconfirmed_var31	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6376delTATATA;6425A>T;6427A>T;9989T>G	2DL2*00101_unconfirmed_var32	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6376delTATATA;6427A>T	2DL2*00101_unconfirmed_var34	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6326delTATG;6331A>G;6333G>A;6335A>G;6376delTA;9989T>G	2DL2*00101_unconfirmed_var35	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6334delTA;6376delTATA;6427A>T;9989T>G	2DL2*00101_unconfirmed_var36	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6376delTATATA;6427A>T;9989T>G;11563T>C;11903G>A	2DL2*00101_unconfirmed_var37	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6118C>T;6421delATATATA;9989T>G;11563T>C	2DL2*00101_unconfirmed_var39	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	1216G>A;6311delAT;6424delTATAT;9989T>G;11563T>C	2DL2*00101_unconfirmed_var40	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	-137C>A;1302G>C;5966delT;6376delTATATA;6383A>G;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var41	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6419delATATATA;6427A>T;9989T>G	2DL2*00101_unconfirmed_var43	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6417delATATATA;6425A>T;6427A>T;9989T>G	2DL2*00101_unconfirmed_var44	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	1216G>A;6421delATATATA;8214C>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var47	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6421delATATATA	2DL2*00101_unconfirmed_var48	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5697C>T;5966delT;6421delATATATA;9989T>G;11563T>C	2DL2*00101_unconfirmed_var50	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	299G>T;1039C>G;5966delT;6377A>G;6421delATATATA;9989T>G	2DL2*00101_unconfirmed_var51	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6311delAT;6376delTATATA;6427A>T;9989T>G	2DL2*00101_unconfirmed_var55	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6334delTA;6423delATATA;9989T>G;12647G>A	2DL2*00101_unconfirmed_var56	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6421delATATATA;9989T>G;11563T>C;11903G>A;12436G>C	2DL2*00101_unconfirmed_var57	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5113delGA;6376delTATATA;6427A>T;9989T>G	2DL2*00101_unconfirmed_var59	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6421delATATATA;9989T>G;11563T>C	2DL2*00101_unconfirmed_var60	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	4744delAG;5966delT;6376delTATATA;6423A>T;6425A>T;6427A>T;9989T>G	2DL2*00101_unconfirmed_var61	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	4146insAT;4149A>G;4160G>C;5966delT;6419delATATATATA;9989T>G;10076delT;11563T>C;12259G>C;13518G>A	2DL2*00101_unconfirmed_var63	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6311insAT;6417delATATATATATA;8259C>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var64	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	1705A>T;6419delATATATATA;9989T>G;11563T>C;14483insC;14511insATGGTTTCGGTAATGGGACTCTTTTCTTGCCTAAGGCTTCCGGTGT;14511C>T	2DL2*00101_unconfirmed_var65	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	-137C>A;5966delT;6377A>G;6417delATATATATA;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var66	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6417delATATATATA;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var67	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6419delATATATATA;9989T>G;11563T>C	2DL2*00101_unconfirmed_var68	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	-137C>A;5966delT;6377A>G;6419delATATATATA;9989T>G;11563T>C	2DL2*00101_unconfirmed_var69	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6419delATATATATA;9989T>G	2DL2*00101_unconfirmed_var70	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6376delTATATATATA;6427A>T;9989T>G	2DL2*00101_unconfirmed_var71	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6376delTATATATATA;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var72	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	1216G>A;6376delTATATATATA;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var73	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6376delTATATATATA;6427A>T;9989T>G	2DL2*00101_unconfirmed_var74	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL2*00101	2DL2*0010103	770A>G;6118C>T;6417delATATATATATA;9989T>G;11563T>C	2DL2*00101_unconfirmed_var75	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6376delTATATATATA;6427A>T;9989T>G;11563T>C;11903G>A	2DL2*00101_unconfirmed_var76	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	2662C>T;6417delATATATATATA;9989T>G;11563T>C;11903G>A;12436G>C	2DL2*00101_unconfirmed_var77	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6417delATATATATATA;9989T>G	2DL2*00101_unconfirmed_var79	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	6415delATATATATATA;6427A>T;9989T>G	2DL2*00101_unconfirmed_var81	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6417delATATATATATA;9989T>G;11563T>C;11903G>A;12436G>C	2DL2*00101_unconfirmed_var82	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	1216G>A;5966delT;6415delATATATATATA;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var83	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6417delATATATATATA;9989T>G;11563T>C	2DL2*00101_unconfirmed_var84	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6417delATATATATATA;9989T>G;12647G>A	2DL2*00101_unconfirmed_var85	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6415delATATATATATA;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var87	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	1216G>A;5966delT;6417delATATATATATA;8214C>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var88	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6311delAT;6419delATATATATA;9989T>G	2DL2*00101_unconfirmed_var89	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	1216G>A;6376delTATATATATATA;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var90	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6311delAT;6371A>G;6375G>A;6417delATATATATA;6427A>T;9989T>G;11618delA	2DL2*00101_unconfirmed_var91	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6417delATATATATATA;9989T>G;11563T>C;11903G>A;11967delT	2DL2*00101_unconfirmed_var92	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	1216G>A;6334delTA;6415delATATATATATA;6427A>T;9989T>G;11563T>C	2DL2*00101_unconfirmed_var94	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6415delATATATATATA;9989T>G;11563T>C	2DL2*00101_unconfirmed_var95	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	5966delT;6334delTA;6417delATATATATATA;9989T>G;11563T>C;11903G>A	2DL2*00101_unconfirmed_var96	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	634delG;1092delG;2102delG;6311delAT;6417delATATATATATA;9989T>G;11563T>C;11903G>A	2DL2*00101_unconfirmed_var104	Unconfirmed	N/A
2DL2*00101	2DL2*0010103	2204delG;5966delT;6324delTATATG;6337A>G;6346C>T;6354T>C;6362C>T;6364T>C;6421delATATA;6427A>T;9989T>G;11411delA;11443delT;11618delA;11967delT	2DL2*00101_unconfirmed_var105	Unconfirmed	N/A
2DL2*00101	2DL2*0010104	5966delT;6423delT;8913A>G;9659T>A;9660G>C;11458G>A;11563T>C;12987C>T;14049T>C	2DL2*00101_unconfirmed_var9	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	6421delATAT;6424delT	2DL2*00101_var1	Confirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6376delT;6376delA;6376delTA	2DL2*00101_var3	Confirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6421delATAT;6424delT	2DL2*00101_var4	Confirmed	N/A
2DL2*00101	2DL2*0010106	6419delATATAT;6424delT	2DL2*00101_var7	Confirmed	N/A
2DL2*00101	2DL2*0010106	2807G>C;5966delT;6376delT;6376delA;6376delTATATATA;6992T>C	2DL2*00101_var15	Confirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6417delATATATATT;6425delT;6425delT;12634G>A	2DL2*00101_var18	Confirmed	N/A
2DL2*00101	2DL2*0010106	6413delATATATATATATT;6425delT;6425delT	2DL2*00101_var21	Confirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6413delATATATATATATT;6425delT;6425delT	2DL2*00101_var22	Confirmed	N/A
2DL2*00101	2DL2*0010106	6423A>T;6424delT;6424delT;8272C>T	2DL2*00101_unconfirmed_var3	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6424delT;6424delT	2DL2*00101_unconfirmed_var4	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	6421delA;6421delT;6421delA;6574G>A	2DL2*00101_unconfirmed_var6	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	157insGGA;158A>G;5966delT;6326delTATG;6331A>G;6333G>A;6335A>G;6423A>T;6424delT;6424delT;6992T>C	2DL2*00101_unconfirmed_var8	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	4361C>T;6207G>C;6424delT;6424delT;6424delTT;7835G>A	2DL2*00101_unconfirmed_var10	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6311delAT;6423A>T;6424delT;6424delT	2DL2*00101_unconfirmed_var12	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6419delA;6419delT;6419delATA;7167A>T;7168G>C	2DL2*00101_unconfirmed_var13	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	2807G>C;5966delT;6419delA;6419delT;6419delATA;6992T>C	2DL2*00101_unconfirmed_var15	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6376delT;6376delA;6376delTATA	2DL2*00101_unconfirmed_var18	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	6421delATATT;6425delT;6425delT	2DL2*00101_unconfirmed_var20	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6419delATATAT;6424delT	2DL2*00101_unconfirmed_var27	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6417delA;6417delT;6417delATATA;8188A>G	2DL2*00101_unconfirmed_var29	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	4361C>T;5966delT;6207G>C;6423delATTTT;6427delT;6427delT;7835G>A	2DL2*00101_unconfirmed_var33	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL2*00101	2DL2*0010106	6417delATATATAT;6424delT	2DL2*00101_unconfirmed_var38	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	2807G>C;6415delA;6415delT;6415delATATATA;6992T>C	2DL2*00101_unconfirmed_var42	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	2807G>C;5966delT;6376delT;6376delA;6376delTATATA;6992T>C	2DL2*00101_unconfirmed_var49	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	2807G>C;5966delT;6415delA;6415delT;6415delATATATA;6992T>C	2DL2*00101_unconfirmed_var52	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6417delATATATAT;6424delT	2DL2*00101_unconfirmed_var53	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6421delATATTTT;6427delT;6427delT;7473C>T;10675G>C;10711G>A;10727G>A;10729C>T	2DL2*00101_unconfirmed_var54	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	2807G>C;6376delT;6376delA;6376delTATATATA;6421A>T;6992T>C	2DL2*00101_unconfirmed_var58	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6420delTATATTTTT;6428delT;7434C>T	2DL2*00101_unconfirmed_var62	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6414delTATATATATAT;6424delT;12634G>A	2DL2*00101_unconfirmed_var78	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	6415delATATATATATT;6425delT;6425delT	2DL2*00101_unconfirmed_var80	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	3087C>G;5966delT;6415delATATATATATT;6425delT;6425delT;12523T>C;12634G>A	2DL2*00101_unconfirmed_var86	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	6413delATATATATATATT;6425delT;6425delT;12634G>A	2DL2*00101_unconfirmed_var93	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6415delATATATATATTTT;6427delT;6427delT	2DL2*00101_unconfirmed_var97	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	6411delATATATATATATATT;6425delT;6425delT	2DL2*00101_unconfirmed_var98	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	6411delATATATATATATATT;6425delT;6425delT;12634G>A	2DL2*00101_unconfirmed_var100	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	4576T>G;5966delT;6411delATATATATATATATT;6425delT;6425delT	2DL2*00101_unconfirmed_var102	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	5966delT;6409delATATATATATATATAT;6424delT;6992T>C	2DL2*00101_unconfirmed_var103	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	4576T>G;6311delAT;6411delATATATATATATATT;6425delT;6425delT	2DL2*00101_unconfirmed_var107	Unconfirmed	N/A
2DL2*00101	2DL2*0010106	3048G>A;6326delTATG;6331A>G;6333G>A;6335A>G;6422delTAT;6424delT;6424delTTTTTT;6839delT;7473C>T;8576delT;10675G>C;10711G>A;10727G>A;10729C>T	2DL2*00101_unconfirmed_var116	Unconfirmed	N/A
2DL2*00101	2DL2*0010107	6377A>G;6394insT;6395insA;6396insT;6397insA;6398insT;6399insA;6400insT;6401insA;6402insT;6403insA;6404insT;6405insA;6406insT;6407insA;6408insT;6409insA;6410insT	2DL2*00101_unconfirmed_var101	Unconfirmed	N/A
2DL2*00101	2DL2*0010107	6394insT;6395insA;6396insT;6397insA;6398insT;6399insA;6400insT;6401insA;6402insT;6403insA;6404insT;6405insA;6406insT;6407insA;6408insT;12634G>A	2DL2*00101_unconfirmed_var106	Unconfirmed	N/A
2DL2*00101	2DL2*0010107	5966delT;6395T>A;6396insT;6397insA;6398insT;6399insA;6400insT;6401insA;6402insT;6403insA;6404insT;6405insA;6406insT;6407insA;6408insT;6409insA;6410insT;8272C>T;11563T>C	2DL2*00101_unconfirmed_var108	Unconfirmed	N/A
2DL2*00101	2DL2*0010107	6394insT;6395insA;6396insT;6397insA;6398insT;6399insA;6400insT;6401insA;6402insT;6403insA;6404insT;6405insA;6406insT;8272C>T	2DL2*00101_unconfirmed_var109	Unconfirmed	N/A
2DL2*00101	2DL2*0010107	5966delT;6395T>A;6396insT;6397insA;6398insT;6399insA;6400insT;6401insA;6402insT;6403insA;6404insT;6405insA;6406insT;6407insA;6408insT;8272C>T;9736T>G	2DL2*00101_unconfirmed_var110	Unconfirmed	N/A
2DL2*00101	2DL2*0010107	5966delT;6396insT;6397insA;6398insT;6399insA;6400insT;6401insA;6402insT;6403insA;6404insT;6405insA;6406insT;6407insA;8272C>T	2DL2*00101_unconfirmed_var111	Unconfirmed	N/A
2DL2*00101	2DL2*0010107	6394insT;6395insA;6396insT;6397insA;6398insT;6424insT;6425insT	2DL2*00101_unconfirmed_var112	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	5966delT;6421delTT	2DL2*00301_var6	Confirmed	N/A
2DL2*00301	2DL2*0030101	6420delT;6426delTT	2DL2*00301_var8	Confirmed	N/A
2DL2*00301	2DL2*0030101	5966delT;6420delT;6426delTT	2DL2*00301_var13	Confirmed	N/A
2DL2*00301	2DL2*0030101	6377A>G	2DL2*00301_unconfirmed_var6	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	4370A>G;5966delT;6419A>T	2DL2*00301_unconfirmed_var11	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	6377A>G;6420delT	2DL2*00301_unconfirmed_var12	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	5767C>A;6419A>T;6420delT	2DL2*00301_unconfirmed_var13	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	6419delAT	2DL2*00301_unconfirmed_var16	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	6421delTT	2DL2*00301_unconfirmed_var19	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	196insG;258T>C;287insG;319insG;503G>A;624T>A;898insC;904C>T;1816C>A;5966insT;6328delT;6328delA;6334delTA;6426delTTT	2DL2*00301_unconfirmed_var20	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL2*00301	2DL2*0030101	6311delA;6311delT;6419A>T;6420delT	2DL2*00301_unconfirmed_var25	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	5966delT;6311delA;6311delT;6377A>G;6420delT	2DL2*00301_unconfirmed_var28	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	4636T>A;6334delTA;6421delTT	2DL2*00301_unconfirmed_var29	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	1158C>G;6334delTA;6421delTT	2DL2*00301_unconfirmed_var30	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	5966delT;6311delA;6311delT;6421delTT	2DL2*00301_unconfirmed_var32	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	6334delTA;6377A>G;6420delT;6426delTT	2DL2*00301_unconfirmed_var34	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	6334delTA;6420delT;6426delTT	2DL2*00301_unconfirmed_var39	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	5113insGA;6326delT;6326delA;6334delTA;6371A>G;6375G>A;6419delATTT	2DL2*00301_unconfirmed_var41	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	5966delT;6311delA;6311delT;6420delT;6426delTT;10530C>T	2DL2*00301_unconfirmed_var46	Unconfirmed	N/A
2DL2*00301	2DL2*0030101	5966delT;6328delT;6328delA;6334delTA;6426delTTT;8448delA	2DL2*00301_unconfirmed_var58	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	6380delTA	2DL2*00301_var1	Confirmed	N/A
2DL2*00301	2DL2*0030102	5966delT;6311delA;6311delT	2DL2*00301_var2	Confirmed	N/A
2DL2*00301	2DL2*0030102	5966delT;6380delTA	2DL2*00301_var3	Confirmed	N/A
2DL2*00301	2DL2*0030102	5966delT;6426delTAT	2DL2*00301_var5	Confirmed	N/A
2DL2*00301	2DL2*0030102	5966delT;6381A>G;6426delTAT	2DL2*00301_var9	Confirmed	N/A
2DL2*00301	2DL2*0030102	5966delT;6311delA;6311delT;6380delTA	2DL2*00301_var12	Confirmed	N/A
2DL2*00301	2DL2*0030102	6311insATAT;6428delT	2DL2*00301_unconfirmed_var1	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	4361C>T;6424insT;6425insA;6426insT	2DL2*00301_unconfirmed_var2	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	1158C>G;6424insT	2DL2*00301_unconfirmed_var3	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	1158C>G;6377A>G;6424insT	2DL2*00301_unconfirmed_var4	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	12859G>C	2DL2*00301_unconfirmed_var5	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	1158C>G;6334delTA;6424insT	2DL2*00301_unconfirmed_var7	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	1158C>G;5966delT;6311delA;6311delT;6329G>A;6331A>G;6333G>A;6335A>G;6378insT;6379insA;6427A>T	2DL2*00301_unconfirmed_var8	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	6427delA	2DL2*00301_unconfirmed_var9	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	1158C>G;6311delA;6311delT;6427delA	2DL2*00301_unconfirmed_var14	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	4510C>G;5966delT;6380delTA	2DL2*00301_unconfirmed_var15	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	1158C>G;5966delT;6334delTA;6427A>T	2DL2*00301_unconfirmed_var17	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	6426delTAT	2DL2*00301_unconfirmed_var18	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	1158C>G;6311delA;6311delT;6427delA;8455delA	2DL2*00301_unconfirmed_var21	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	1158C>G;5966delT;6311delA;6311delT;6427delA	2DL2*00301_unconfirmed_var22	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	5966delT;6380delTA;11618delA	2DL2*00301_unconfirmed_var24	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	637G>T;1158C>G;5966delT;6311delA;6311delT;6427delA;8448delA;14205delA	2DL2*00301_unconfirmed_var35	Unconfirmed	N/A
2DL2*00301	2DL2*0030102	5966delT;6311delA;6311delT;6426delTAT	2DL2*00301_unconfirmed_var38	Unconfirmed	N/A
2DL2*00301	2DL2*0030103	6380delTATA;6423T>A;6425T>A;6427T>A	2DL2*00301_unconfirmed_var26	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	6384insT;6385insA;6425T>A;7358T>C	2DL2*00301_var7	Confirmed	N/A
2DL2*00301	2DL2*0030106	6415T>A;6416insT;7358T>C	2DL2*00301_var10	Confirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6384insT;6385insA;6425T>A;7358T>C	2DL2*00301_var11	Confirmed	N/A
2DL2*00301	2DL2*0030106	6415T>A;6417T>A;6419insA;7358T>C	2DL2*00301_var14	Confirmed	N/A
2DL2*00301	2DL2*0030106	6425T>A;7358T>C	2DL2*00301_var16	Confirmed	2DL2*0030108 (LR593893)
2DL2*00301	2DL2*0030106	5966delT;6425T>A;7358T>C	2DL2*00301_var17	Confirmed	N/A
2DL2*00301	2DL2*0030106	6425delT;7358T>C	2DL2*00301_var18	Confirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6425T>A;6427T>A;7358T>C	2DL2*00301_var19	Confirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6425delT;7358T>C	2DL2*00301_var20	Confirmed	N/A
2DL2*00301	2DL2*0030106	6423delATT	2DL2*00301_var22	Confirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6423delATT	2DL2*00301_var23	Confirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6311insATATAT;6425delT;7358T>C	2DL2*00301_unconfirmed_var10	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	4636T>A;6384insT;6385insA;6425T>A;7358T>C	2DL2*00301_unconfirmed_var27	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	6415T>A;6417T>A;6418insT;7358T>C	2DL2*00301_unconfirmed_var31	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6415T>A;6417T>A;6419insA;7358T>C	2DL2*00301_unconfirmed_var33	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	6371A>G;6375G>A;6415T>A;6416insT;7358T>C;12744delG	2DL2*00301_unconfirmed_var36	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	6311delA;6311delT;6384insT;6385insA;6425T>A;7358T>C	2DL2*00301_unconfirmed_var37	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6425T>A;6427T>A;7358T>C;12516C>G	2DL2*00301_unconfirmed_var40	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	6425delT;12987T>C	2DL2*00301_unconfirmed_var42	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6311delA;6311delT;6384insT;6385insA;6425T>A;7358T>C	2DL2*00301_unconfirmed_var44	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	6371A>G;6375G>A;6425delT;7358T>C	2DL2*00301_unconfirmed_var45	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6392T>C;6425T>A;7358T>C;12744G>A	2DL2*00301_unconfirmed_var47	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	4636T>A;6371A>G;6375G>A;6425delT;7358T>C;7569C>T	2DL2*00301_unconfirmed_var48	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6425T>A;6427T>A;7358T>C;9104delA	2DL2*00301_unconfirmed_var49	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	6311delA;6311delT;6425T>A;6427T>A;7358T>C	2DL2*00301_unconfirmed_var50	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6425T>A;6427T>A;7358T>C;8455delA	2DL2*00301_unconfirmed_var51	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL2*00301	2DL2*0030106	6311delA;6311delT;6415T>A;6417T>A;6428delT;7358T>C	2DL2*00301_unconfirmed_var52	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6386delTA;6425T>A;7358T>C	2DL2*00301_unconfirmed_var53	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6311delA;6311delT;6425T>A;7358T>C	2DL2*00301_unconfirmed_var54	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	6311delA;6311delT;6425delT;7358T>C	2DL2*00301_unconfirmed_var55	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6311delA;6311delT;6371A>G;6375G>A;6415T>A;6427delT;7358T>C	2DL2*00301_unconfirmed_var56	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	6334delTA;6415T>A;6427delT;6998delT;7358T>C	2DL2*00301_unconfirmed_var60	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6328delT;6328delA;6334delTA;6425T>A;7358T>C	2DL2*00301_unconfirmed_var61	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	5966delT;6311delA;6311delT;6371A>G;6374delTG;6425T>A;6427T>A;7358T>C	2DL2*00301_unconfirmed_var62	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	1138delC;1965delC;4744delAG;6425delT;7358T>C;13593delC	2DL2*00301_unconfirmed_var64	Unconfirmed	N/A
2DL2*00301	2DL2*0030106	4636T>A;5966delT;6345delACAC;6386delTA;6425T>A;7358T>C	2DL2*00301_unconfirmed_var65	Unconfirmed	N/A
2DL2*00301	2DL2*0030107	295A>G;6411T>A;6412insT;12055G>A	2DL2*00301_var21	Confirmed	N/A
2DL2*00301	2DL2*0030107	295A>G;5966delT;6411T>A;6412insT;12055G>A	2DL2*00301_var24	Confirmed	N/A
2DL2*00301	2DL2*0030107	295A>G;5966delT;6311delA;6311delT;6411T>A;6412insT;12055G>A	2DL2*00301_var25	Confirmed	N/A
2DL2*00301	2DL2*0030107	295A>G;5966delT;6311insATATAT;6411T>A;6412insT;12055G>A	2DL2*00301_unconfirmed_var23	Unconfirmed	N/A
2DL2*00301	2DL2*0030107	295A>G;5601T>G;5966delT;6411T>A;6412insT;12055G>A	2DL2*00301_unconfirmed_var57	Unconfirmed	N/A
2DL2*00301	2DL2*0030107	295A>G;6425T>A;6427T>A;12055G>A	2DL2*00301_unconfirmed_var59	Unconfirmed	N/A
2DL2*00301	2DL2*0030107	295A>G;5966delT;6425T>A;6427T>A;8455delA;12055G>A	2DL2*00301_unconfirmed_var63	Unconfirmed	N/A
2DL2*00301	2DL2*0030107	295A>G;5966delT;6311delA;6311delT;6334delTA;6411T>A;6412insT;12055G>A	2DL2*00301_unconfirmed_var66	Unconfirmed	N/A
2DL2*00301	2DL2*0030107	295A>G;6390delTATA;6425T>A;6427T>A;12055G>A	2DL2*00301_unconfirmed_var67	Unconfirmed	N/A
2DL3*00101	2DL3*0010101	2774C>A;5968delT;6308insTA;6398delT;6400T>A;6402T>A;6404T>A	2DL3*00101_unconfirmed_var96	Unconfirmed	N/A
2DL3*00101	2DL3*0010101	2216T>C;6334C>T;6348T>C;6354A>T;6356C>A;6358T>C;6399T>A	2DL3*00101_unconfirmed_var97	Unconfirmed	N/A
2DL3*00101	2DL3*0010101	2774C>A;6308insTA;6398delTTT;6402T>A;6404T>A	2DL3*00101_unconfirmed_var98	Unconfirmed	N/A
2DL3*00101	2DL3*0010101	2774C>A;6308insTA;6398delTTT;6402T>A	2DL3*00101_unconfirmed_var100	Unconfirmed	N/A
2DL3*00101	2DL3*0010101	503A>G;1816A>C;2048T>G;2376T>G;2779T>A;3187C>A;4551A>G;4860G>A;5114delGA;5769C>G;6078T>C;6334C>T;6348T>C;6354A>T;6356C>A;6358T>C;14398C>G;14401G>T	2DL3*00101_unconfirmed_var103	Unconfirmed	N/A
2DL3*00101	2DL3*0010101	363insG;1282delG;2774C>A;5968delT;6334C>T;6348T>C;6354A>T;6356C>A;6358T>C;6399T>A;6401T>A;6403T>A;12676delG	2DL3*00101_unconfirmed_var104	Unconfirmed	N/A
2DL3*00101	2DL3*0010101	2216T>C;5114delGA;5968delT;6398delTTT;7422delT	2DL3*00101_unconfirmed_var164	Unconfirmed	N/A
2DL3*00101	2DL3*0010101	2774C>A;5968delT;6368delTATATATATATATATATATATATATATATATTTT	2DL3*00101_unconfirmed_var320	Unconfirmed	N/A
2DL3*00101	2DL3*0010102	2216T>C;6333T>C;6346C>T;6352T>A;6354A>C;6356C>T;10899G>C;11751T>C	2DL3*00101_unconfirmed_var75	Unconfirmed	N/A
2DL3*00101	2DL3*0010102	5114delGA;5968delT;10899G>C;11751T>C	2DL3*00101_unconfirmed_var84	Unconfirmed	N/A
2DL3*00101	2DL3*0010102	5968delT;6308delT;6308delA;6398delTA;10899G>C;11751T>C	2DL3*00101_unconfirmed_var89	Unconfirmed	N/A
2DL3*00101	2DL3*0010102	4511C>G;5968delT;6308delT;6308delA;6399delA;8105G>C;8435delA;10056delT;10899G>C;11751T>C	2DL3*00101_unconfirmed_var92	Unconfirmed	N/A
2DL3*00101	2DL3*0010102	2216T>C;5968delT;6396delTATAT;10899G>C;11751T>C	2DL3*00101_unconfirmed_var94	Unconfirmed	N/A
2DL3*00101	2DL3*0010102	52C>T;5968delT;6308delT;6308delA;6397delATA;10899G>C;11751T>C	2DL3*00101_unconfirmed_var95	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;4942insTTTTAGAT;4942C>T;5968delT;6391T>A;6393T>A;6395insA	2DL3*00101_var18	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;6308insTATA	2DL3*00101_var21	Confirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;6393T>A;6395insA;6396insT;6397insA;6402T>A;6404T>A	2DL3*00101_var22	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6308insTA;6391insA;6392insT;6393insA;6399delT	2DL3*00101_var23	Confirmed	N/A
2DL3*00101	2DL3*0010103	2216T>C;4511G>C;6391T>A;6393insA;6394insT;6395insA	2DL3*00101_var25	Confirmed	N/A
2DL3*00101	2DL3*0010103	2216T>C;4511G>C;5968delT;6391T>A;6393insA;6394insT;6395insA	2DL3*00101_var26	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;6391insA;6392insT;6393insA;6399delT	2DL3*00101_var27	Confirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6391insA;6392insT;6393insA;6399delT	2DL3*00101_var29	Confirmed	N/A
2DL3*00101	2DL3*0010103	4014C>G;4100A>C;4147insAT;4150A>G;6398delT	2DL3*00101_var31	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6308insTA;6398delT	2DL3*00101_var35	Confirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5968delT;6391insA;6392insT;6393insA;6394insT;6395insA;6399delTTTT	2DL3*00101_var36	Confirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5968delT;6391T>A;6393T>A;6395T>A;6397insA	2DL3*00101_var37	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6391T>A;6393insA	2DL3*00101_var38	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5927A>G;6399T>A	2DL3*00101_var39	Confirmed	2DL3*0010113 (LR593975)
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5968delT;6336C>T;6356T>C;6362A>T;6364C>A;6366T>C;6399T>A	2DL3*00101_var40	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6336C>T;6356T>C;6362A>T;6364C>A;6366T>C	2DL3*00101_var41	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT	2DL3*00101_var42	Confirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;6398delT;6400T>A;6402T>A	2DL3*00101_var43	Confirmed	N/A
2DL3*00101	2DL3*0010103	6398delT	2DL3*00101_var44	Confirmed	N/A
2DL3*00101	2DL3*0010103	6397delA	2DL3*00101_var45	Confirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;6391insA;6392insT;6393insA;6394insT;6395insA;6399delTTTTTT	2DL3*00101_var46	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6398delT	2DL3*00101_var48	Confirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5968delT;6398delT;6400T>A;6402T>A	2DL3*00101_var50	Confirmed	N/A
2DL3*00101	2DL3*0010103	3211G>A;5968delT;6397delA;11643A>G	2DL3*00101_var51	Confirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6398delT;6400T>A	2DL3*00101_var52	Confirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6398delT	2DL3*00101_var53	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6308delT;6308delA;6391insA	2DL3*00101_var54	Confirmed	N/A
2DL3*00101	2DL3*0010103	6349delAT;6362A>T;6364C>A;6366T>C	2DL3*00101_var55	Confirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6365delAT	2DL3*00101_var56	Confirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6398delTT	2DL3*00101_var58	Confirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;4942insTTTTAGAT;4942C>T;6393T>A;6395insA;6396insT;6397insA;6402T>A	2DL3*00101_unconfirmed_var78	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;4942insTTTTAGAT;4942C>T;6391insA;6392insT;6393insA;6394insT;6395insA;6399delTTTT	2DL3*00101_unconfirmed_var83	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;4942insTTTTAGAT;4942C>T;5968delT;6391T>A;6393T>A;6395insA;7916T>C	2DL3*00101_unconfirmed_var85	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;4942insTTTTAGAT;4942C>T;5968delT;6391T>A;6393T>A;6395T>A	2DL3*00101_unconfirmed_var87	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;4942insTTTTAGAT;4942C>T;6398delT;6400T>A;6402T>A	2DL3*00101_unconfirmed_var90	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;4942insTTTTAGAT;4942C>T;5114delGA;5968delT;6391T>A;6393T>A;6395insA	2DL3*00101_unconfirmed_var93	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6336C>T;6350T>C;6353insA;6354insT;6355insA;6356insT;6357insA;6358insA;6360T>C;6362A>T;6364C>T;6399T>A	2DL3*00101_unconfirmed_var99	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6391insA;6392insT;6393insA;6394insT;6395insA;14030C>T	2DL3*00101_unconfirmed_var102	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;6308insTA;6391insA;6392insT;6393insA;6399delT	2DL3*00101_unconfirmed_var105	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2216T>C;4511G>C;5968delT;6336C>T;6350T>C;6351insA;6352insT;6399insATA;6400T>A	2DL3*00101_unconfirmed_var107	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;6391insA;6392insT;6393insA;6394insT;6395insA;6399delT;6401T>A	2DL3*00101_unconfirmed_var108	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6308insTA;6363insA;6364insT	2DL3*00101_unconfirmed_var112	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5658G>A;6393T>A;6395insA;6396insT;6397insA;6402T>A	2DL3*00101_unconfirmed_var113	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6391insA;6392insT;6393insA;6394insT;6395insA;6399delT	2DL3*00101_unconfirmed_var114	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6308insTA;6391T>A;6393insA	2DL3*00101_unconfirmed_var117	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5968delT;6336C>T;6350T>C;6351insA;6352insT;6399insA;6400T>A;6402T>A	2DL3*00101_unconfirmed_var118	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5114delGA;5968delT;6391insA;6392insT;6393insA;6394insT;6395insA;14030C>T	2DL3*00101_unconfirmed_var119	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;6363insA;6364insT;6399T>A;12832C>T	2DL3*00101_unconfirmed_var120	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00101	2DL3*0010103	173G>A;211delG;212G>T;214T>G;842C>T;3269C>T;4511G>C;5873G>A;6331insATAT;6391T>A;6400delT	2DL3*00101_unconfirmed_var121	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6391insA;6392insT;6393insA;14030C>T	2DL3*00101_unconfirmed_var123	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;6363insA;6364insT;6399T>A;6401T>A;6403T>A	2DL3*00101_unconfirmed_var125	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2193A>C;4511G>C;5968delT;6391T>A;6393insA;6394insT;6395insA;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var127	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5828A>C;5968delT;6363insA;6364insT;6399T>A;11875G>A	2DL3*00101_unconfirmed_var132	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6010C>A;6336C>T;6350T>C;6351insA;6352insT;6399T>A;11590G>A	2DL3*00101_unconfirmed_var135	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;6391insA;6392insT;6393insA;6394insT;6395insA;6399delTTTT	2DL3*00101_unconfirmed_var136	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;6350T>C;6351insA;6352insT;6398delT;7688G>A;9428C>G;10468T>A	2DL3*00101_unconfirmed_var137	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5968delT;6336C>T;6350T>C;6351insA;6352insT;6399T>A	2DL3*00101_unconfirmed_var138	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6308insTA;6399T>A	2DL3*00101_unconfirmed_var139	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6363insA;6364insT;6399T>A;14030C>T	2DL3*00101_unconfirmed_var142	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5927A>G;5968delT;6363insA;6364insT;6399T>A	2DL3*00101_unconfirmed_var143	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	503A>G;4511G>C;6399T>A;8887G>T	2DL3*00101_unconfirmed_var145	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6391insA;12313G>A;14030C>T	2DL3*00101_unconfirmed_var147	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C	2DL3*00101_unconfirmed_var149	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;6399T>A;6401T>A;6403T>A	2DL3*00101_unconfirmed_var151	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6391insA;14030C>T	2DL3*00101_unconfirmed_var152	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6391insA	2DL3*00101_unconfirmed_var154	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5114delGA;5968delT;6308insTA;6391T>A;6393insA	2DL3*00101_unconfirmed_var155	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;6336C>T;6356T>C;6362A>T;6364C>A;6366T>C;6399T>A	2DL3*00101_unconfirmed_var158	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C	2DL3*00101_unconfirmed_var159	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2193A>C;4511G>C;6399T>A;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var160	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5114delGA;5828A>C;5968delT;6308insTA;6390insT	2DL3*00101_unconfirmed_var162	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5114insGA;6398delTT;6401T>A	2DL3*00101_unconfirmed_var163	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6399T>A;14030C>T	2DL3*00101_unconfirmed_var167	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	349insG;4511G>C;5968delT;6398delT;6400T>A	2DL3*00101_unconfirmed_var168	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4183T>C;5968delT;6399T>A	2DL3*00101_unconfirmed_var169	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5968delT;6399T>A;6401T>A	2DL3*00101_unconfirmed_var170	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6399T>A	2DL3*00101_unconfirmed_var175	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6336C>T;6356T>C;6362A>T;6364C>A;6366T>C;6399T>A	2DL3*00101_unconfirmed_var176	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;6398delT	2DL3*00101_unconfirmed_var177	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5658G>A;6398delT;6400T>A;6402T>A	2DL3*00101_unconfirmed_var178	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;6308delT;6308delA;6391insA	2DL3*00101_unconfirmed_var180	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6399T>A;6401T>A	2DL3*00101_unconfirmed_var182	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6399T>A;12832C>T	2DL3*00101_unconfirmed_var183	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6399T>A	2DL3*00101_unconfirmed_var184	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2193A>C;4511G>C;5968delT;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var185	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	6398delT;14398C>G;14401G>T	2DL3*00101_unconfirmed_var188	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	6398delT;11643A>G	2DL3*00101_unconfirmed_var189	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5927A>G;5968delT;6399T>A	2DL3*00101_unconfirmed_var190	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	6308delT;6308delA;6391insA	2DL3*00101_unconfirmed_var191	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5114delGA;6391insA;6392insT;6393insA;6394insT;6395insA;6399delTTTT	2DL3*00101_unconfirmed_var192	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5114insGA;5968delT;6398delTTT;8542A>G	2DL3*00101_unconfirmed_var194	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2901C>A;4511G>C;6398delTT	2DL3*00101_unconfirmed_var195	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2342C>T;5968delT;6398delT;6400T>A	2DL3*00101_unconfirmed_var196	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5927A>G;5968delT;6398delT;6400T>A	2DL3*00101_unconfirmed_var197	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6398delT;13590A>G	2DL3*00101_unconfirmed_var198	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	6349delAT;6362A>T;6364C>A;6366T>C;14398C>G;14401G>T	2DL3*00101_unconfirmed_var200	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5114delGA;5968delT;6308insTA;6398delT;6400T>A	2DL3*00101_unconfirmed_var202	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6398delT;6400T>A	2DL3*00101_unconfirmed_var203	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5968delT;6391insA;6392insT;6393insA;6394insT;6395insA;6399delTTTTTT	2DL3*00101_unconfirmed_var204	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00101	2DL3*0010103	2193A>C;4511G>C;5968delT;6398delT;6400T>A;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var205	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6398delT;6400T>A;7688G>A;9428C>G	2DL3*00101_unconfirmed_var208	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;6398delTT	2DL3*00101_unconfirmed_var209	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2193A>C;4511G>C;5968delT;6398delT;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var210	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;6365delAT	2DL3*00101_unconfirmed_var214	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2193A>C;4511G>C;6398delTT;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var215	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6349delAT;6362A>T;6364C>A;6366T>C;12549G>A	2DL3*00101_unconfirmed_var220	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5114delGA;5968delT;6336C>T;6356T>C;6362A>T;6364C>A;6366T>C;6399T>A	2DL3*00101_unconfirmed_var222	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	363insG;2774C>A;4511G>C;5114delGA;5968delT;6391insA;6392insT;6393insA;6394insT;6395insA;6399delTTTTTT	2DL3*00101_unconfirmed_var226	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5114delGA;6308delT;6308delA;6391insA	2DL3*00101_unconfirmed_var229	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2774C>A;4511G>C;5114delGA;6391insA;6392insT;6393insA;6394insT;6395insA;6399delTTTTTT	2DL3*00101_unconfirmed_var230	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	3211G>A;5968delT;6349delAT;6362A>T;6364C>A;6366T>C;7943C>G;11643A>G	2DL3*00101_unconfirmed_var231	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6398delTT;14398C>G;14401G>T	2DL3*00101_unconfirmed_var232	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2193A>C;4511G>C;5114delGA;5968delT;7688G>A;9428C>G;10535G>A;11928G>A;12905A>G	2DL3*00101_unconfirmed_var234	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5968delT;6349delAT;6362A>T;6364C>A;6366T>C;13020G>T	2DL3*00101_unconfirmed_var236	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2193A>C;4511G>C;5968delT;6398delTT;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var238	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4014C>G;4100A>C;4147insAT;4150A>G;5968delT;6396delTAT;6977delT	2DL3*00101_unconfirmed_var239	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6365delAT	2DL3*00101_unconfirmed_var240	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	6308delT;6308delA;6398delT;8329C>A;14398C>G;14401G>T	2DL3*00101_unconfirmed_var243	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	2444G>A;6396delTAT	2DL3*00101_unconfirmed_var244	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	3211G>A;5968delT;6395delATA;11643A>G;14286G>A	2DL3*00101_unconfirmed_var254	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	3211G>A;5968delT;6395delATA;11643A>G	2DL3*00101_unconfirmed_var257	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;5114delGA;5968delT;6335delA;6335delC;6391T>A;6393insA;6977delT	2DL3*00101_unconfirmed_var278	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5114delGA;5968delT;6365delAT	2DL3*00101_unconfirmed_var279	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	3211G>A;5968delT;6365delATAT;11643A>G	2DL3*00101_unconfirmed_var280	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	1028delT;4511G>C;5114delGA;6335delA;6335delC;6391T>A;6393insA;6787delT;11417delT;11942delT	2DL3*00101_unconfirmed_var304	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5968delT;6363insA;6364insT;14030C>T;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00101_unconfirmed_var325	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	5114insGA;6398delT;6400T>A;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00101_unconfirmed_var326	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	4511G>C;6336C>T;6350T>C;6351insA;6352insT;6397delA;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00101_unconfirmed_var327	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00101	2DL3*0010103	3211G>A;6334T>C;6347delACAT;6362A>T;6364C>A;6366T>C;6397A>T;10899C>G;11643A>G;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00101_unconfirmed_var329	Unconfirmed	N/A
2DL3*00101	2DL3*0010103	3211G>A;5114delGA;6349delAT;6362A>T;6364C>A;6366T>C;6397A>T;11643A>G;12337C>A;14233delC;14242delACCTCTCCAACCTAACTGGCTTACTTCCTAGTCTACTTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCA;14367C>T;14370C>A;14371C>T;14378C>A	2DL3*00101_unconfirmed_var330	Unconfirmed	N/A
2DL3*00101	2DL3*0010104	4511C>G;6395T>A;13561C>G;13563A>T;14030C>T	2DL3*00101_unconfirmed_var126	Unconfirmed	N/A
2DL3*00101	2DL3*0010104	5828A>C;6397delA;13561C>G;13563A>T	2DL3*00101_unconfirmed_var141	Unconfirmed	N/A
2DL3*00101	2DL3*0010104	659G>A;5968delT;6398delT;13561C>G;13563A>T	2DL3*00101_unconfirmed_var157	Unconfirmed	N/A
2DL3*00101	2DL3*0010104	5968delT;6395insA;6399delTTT;10218G>A;13561C>G;13563A>T	2DL3*00101_unconfirmed_var165	Unconfirmed	N/A
2DL3*00101	2DL3*0010104	5114delGA;5927A>G;5968delT;6333delAC;6395insA;6399delTT;13561C>G;13563A>T	2DL3*00101_unconfirmed_var253	Unconfirmed	N/A
2DL3*00101	2DL3*0010106	2377C>G;6398delT;12761G>T	2DL3*00101_unconfirmed_var172	Unconfirmed	N/A
2DL3*00101	2DL3*0010106	5968delT;6398delT;12761G>T	2DL3*00101_unconfirmed_var213	Unconfirmed	N/A
2DL3*00101	2DL3*0010106	5968delT;6396delTATT;12761G>T	2DL3*00101_unconfirmed_var281	Unconfirmed	N/A
2DL3*00101	2DL3*0010107	5968delT;6308insTATA;8598T>C;12839delG	2DL3*00101_unconfirmed_var110	Unconfirmed	N/A
2DL3*00101	2DL3*0010107	5968delT;6395A>T;8598T>C;12839delG	2DL3*00101_unconfirmed_var181	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	5968delT;6400insT;10606T>C	2DL3*00101_var67	Confirmed	N/A
2DL3*00101	2DL3*0010108	327A>G;328insGGAGGGT;329T>G;2774C>A;5114delGA;6389insA;6390insT;6391insA;6392insT;6393insA;6402delTTT;10606T>C	2DL3*00101_unconfirmed_var122	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	4511C>G;5114insGA;5948A>T;5968delT;6308insTA;6400insT;10606T>C;14249C>T	2DL3*00101_unconfirmed_var166	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2774C>A;5968delT;6308insTA;6389T>A;6391insA;6392insT;6393insA;10606T>C	2DL3*00101_unconfirmed_var171	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2774C>A;3838C>G;5968delT;6308insTA;6388insT;6389insA;6402T>A;10606T>C	2DL3*00101_unconfirmed_var201	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2193A>C;6308insTA;6389insA;7688G>A;9428C>G;10606T>C;11928G>A	2DL3*00101_unconfirmed_var206	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	5968delT;6336C>T;6345insA;6346insC;6400insT;10606T>C	2DL3*00101_unconfirmed_var218	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2774C>A;5968delT;6389T>A;6391insA;6392insT;6393insA;10606T>C	2DL3*00101_unconfirmed_var221	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	5968delT;6336C>T;6345insA;6346insC;6400insT;10606T>C;12087C>G	2DL3*00101_unconfirmed_var223	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2774C>A;5114delGA;5968delT;6308insTA;6389T>A;6391insA;6392insT;6393insA;10606T>C	2DL3*00101_unconfirmed_var227	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	5968delT;6345insA;6346insC;6400insT;7688G>A;9428C>G;10468T>A;10606T>C	2DL3*00101_unconfirmed_var233	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	6388insT;6389insA;6402T>A	2DL3*00101_unconfirmed_var241	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2774C>A;5658G>A;5968delT;6388insT;6389insA;6402T>A;10606T>C	2DL3*00101_unconfirmed_var247	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	5968delT;6399insT;6400insT;10606T>C	2DL3*00101_unconfirmed_var249	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2774C>A;5968delT;6388insT;6389insA;6402T>A;10606T>C	2DL3*00101_unconfirmed_var251	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2193A>C;5968delT;6388insT;6389insA;6402T>A;7688G>A;9428C>G;10606T>C;11928G>A	2DL3*00101_unconfirmed_var258	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	363insG;4511C>G;5968delT;6400insT;10606T>C	2DL3*00101_unconfirmed_var266	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	6389T>A;6391insA;10606T>C	2DL3*00101_unconfirmed_var267	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	503A>G;5968delT;6400insT;10606T>C;12211C>A	2DL3*00101_unconfirmed_var274	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	4511C>G;5968delT;6400insT;10606T>C	2DL3*00101_unconfirmed_var275	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	4511C>G;5549C>G;5968delT;6400insT;10606T>C	2DL3*00101_unconfirmed_var282	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2774C>A;5114delGA;5968delT;6388insT;6389insA;6402T>A;10606T>C	2DL3*00101_unconfirmed_var290	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	4511C>G;5968delT;6400A>T;10606T>C	2DL3*00101_unconfirmed_var292	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	6401delT;7805T>C;10606T>C;13366A>T	2DL3*00101_unconfirmed_var296	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00101	2DL3*0010108	5114delGA;5968delT;6400insT;10606T>C	2DL3*00101_unconfirmed_var298	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2774C>A;5968delT;6389T>A;6977delT;10606T>C	2DL3*00101_unconfirmed_var303	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	2774C>A;5968delT;6387delA;10606T>C	2DL3*00101_unconfirmed_var306	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	5968delT;6401delT;6532T>C;7805T>C;10606T>C	2DL3*00101_unconfirmed_var308	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	5968delT;6387delA;7688G>A;8702C>T;9428C>G;10606T>C	2DL3*00101_unconfirmed_var309	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	1690G>A;5968delT;6387delA;10606T>C	2DL3*00101_unconfirmed_var310	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	5114delGA;6308delT;6308delA;6401delT	2DL3*00101_unconfirmed_var318	Unconfirmed	N/A
2DL3*00101	2DL3*0010108	5927A>G;6382delTATATA;10606T>C	2DL3*00101_unconfirmed_var319	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	6390insT;6391insA;6400insT	2DL3*00101_var49	Confirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6390insT;6391insA;6400insT	2DL3*00101_var57	Confirmed	N/A
2DL3*00101	2DL3*0010109	2193A>C;5968delT;6308insTA;6391insA;7688G>A;9428C>G;11928G>A	2DL3*00101_var59	Confirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5968delT;6308insTA;6391insA	2DL3*00101_var60	Confirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;6390insT;6391insA	2DL3*00101_var61	Confirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5968delT;6308insTA	2DL3*00101_var62	Confirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6390insT;6391insA	2DL3*00101_var63	Confirmed	N/A
2DL3*00101	2DL3*0010109	4511C>G;6388insT;6400A>T	2DL3*00101_var64	Confirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6308delTA;6390insT;6391insA;6400insT	2DL3*00101_var65	Confirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;6391T>A	2DL3*00101_var66	Confirmed	2DL3*0010115 (LR593924)
2DL3*00101	2DL3*0010109	3211G>A;4511C>G;5968delT;6398A>T;6400A>T;11643A>G;14286G>A	2DL3*00101_var69	Confirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5968delT;6391T>A	2DL3*00101_var70	Confirmed	N/A
2DL3*00101	2DL3*0010109	4511C>G;5968delT;6389delA	2DL3*00101_var71	Confirmed	N/A
2DL3*00101	2DL3*0010109	390insGAGATATGGGCCTGGAGTGGAGATATGGGTCTGGAGGTGGAGATACGGGCCTGCAGTA;5968delT;6390insT;6391insA;6400insT	2DL3*00101_unconfirmed_var1	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;6391T>A;6393insA;6394insT;6395insA;6396insT;6397insA	2DL3*00101_unconfirmed_var161	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;6391T>A;6393T>A;6394insT;6395insA;6396insT;6397insA	2DL3*00101_unconfirmed_var179	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	6390insT;6391insA;6392insT;6393insA;14357G>C	2DL3*00101_unconfirmed_var186	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;6390insT;6391insA;6392insT;6393insA;6402T>A	2DL3*00101_unconfirmed_var187	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5968delT;6333insAC;6390insT;6391insA;6402T>A	2DL3*00101_unconfirmed_var193	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	6390insT;6391insA;6392insT;12832C>T	2DL3*00101_unconfirmed_var199	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6308insTATA;6391insA;8262delA	2DL3*00101_unconfirmed_var207	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6308insTATA;6746T>C	2DL3*00101_unconfirmed_var211	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;6333insAC;6391T>A;6393insA	2DL3*00101_unconfirmed_var212	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;6308insTA;6391insA	2DL3*00101_unconfirmed_var216	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6308insTA;6391insA	2DL3*00101_unconfirmed_var219	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6391insA;6392insT;6393insA;8883C>T	2DL3*00101_unconfirmed_var224	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5968delT;6333insAC;6400insT	2DL3*00101_unconfirmed_var225	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6214T>A;6391insA;6392insT;6393insA	2DL3*00101_unconfirmed_var228	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2723G>C;3947G>T;6399insT;6400insT;6589T>C	2DL3*00101_unconfirmed_var235	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6333insAC;6400insT;7688G>A;9428C>G;10468T>A	2DL3*00101_unconfirmed_var237	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	6308delTA;6390insT;6391insA;6392insT;6393insA	2DL3*00101_unconfirmed_var242	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	1905G>A;6390insT;6391insA	2DL3*00101_unconfirmed_var246	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6399insT;6400insT	2DL3*00101_unconfirmed_var248	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5968delT;6334C>T;6390insT;6391insA	2DL3*00101_unconfirmed_var250	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2980C>G;5968delT;6308insTA;7688G>A;9428C>G	2DL3*00101_unconfirmed_var252	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2193A>C;5968delT;6390insT;6391insA;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var255	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6390insT;6391insA;6402T>A	2DL3*00101_unconfirmed_var256	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	749G>A;2774C>A;5968delT;6308insTA;6334C>T	2DL3*00101_unconfirmed_var259	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5114delGA;5968delT;6391T>A;6393T>A;6394insT;6395insA;6396insT;6397insA	2DL3*00101_unconfirmed_var260	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5968delT;6390insT;6391insA;6402T>A	2DL3*00101_unconfirmed_var262	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5114delGA;6391insA;6392insT;6393insA;12832C>T	2DL3*00101_unconfirmed_var263	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6390insT;6391insA;14357G>C	2DL3*00101_unconfirmed_var264	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	6400insT	2DL3*00101_unconfirmed_var265	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	4014C>G;4100A>C;4147insAT;4150A>G;4511C>G;5968delT;6400A>T;11406T>A	2DL3*00101_unconfirmed_var268	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	4511C>G;6400insT	2DL3*00101_unconfirmed_var270	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	1905G>A;6400insT	2DL3*00101_unconfirmed_var271	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6391insA;14357G>C	2DL3*00101_unconfirmed_var272	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	1847T>C;4147insAT;4217G>C;4511C>G;5968delT;6389delA;7322C>T;9568T>A	2DL3*00101_unconfirmed_var273	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00101	2DL3*0010109	2774C>A;6391T>A;6393T>A;11335T>C;11336A>G	2DL3*00101_unconfirmed_var276	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6308delTA;6391insA;6392insT;6393insA	2DL3*00101_unconfirmed_var277	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6400insT	2DL3*00101_unconfirmed_var283	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	6308delTA;6399insT;6400insT	2DL3*00101_unconfirmed_var284	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5968delT;6391T>A;11810T>C	2DL3*00101_unconfirmed_var286	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2193A>C;5122insG;5968delT;6334C>T;6401delT;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var287	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5114delGA;6308delTA;6391insA;6392insT;6393insA	2DL3*00101_unconfirmed_var289	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2723G>C;3947G>T;5968delT;6308delTA;6390insT;6391insA;6589T>C	2DL3*00101_unconfirmed_var291	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	4014C>G;4100A>C;4147insAT;4150A>G;4511C>G;5114delGA;5968delT;6400A>T	2DL3*00101_unconfirmed_var294	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5114delGA;6400insT	2DL3*00101_unconfirmed_var295	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5114delGA;6391T>A;6393T>A;6395insA	2DL3*00101_unconfirmed_var297	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2342C>T;4511C>G;5968delT;6401delT	2DL3*00101_unconfirmed_var299	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	503A>G;5968delT;6389delA;8887G>T	2DL3*00101_unconfirmed_var300	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5968delT;6308delTA;6390insT;6391insA;6402T>A;6589T>C;6977delT;12341G>A	2DL3*00101_unconfirmed_var302	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	3211G>A;4511C>G;5968delT;6387A>T;6389delA;11643A>G	2DL3*00101_unconfirmed_var305	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	3211G>A;4511C>G;5968delT;6385A>T;6387A>T;6389delA;10899C>G;11643A>G;14286G>A	2DL3*00101_unconfirmed_var307	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	349insG;2342C>T;4511C>G;5114delGA;5968delT	2DL3*00101_unconfirmed_var311	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5114delGA;5968delT;6391insA;12940G>C	2DL3*00101_unconfirmed_var312	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5114delGA;6308delTATA;6390insT;6391insA;6392insT;6393insA	2DL3*00101_unconfirmed_var313	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2193A>C;5968delT;6333delAC;6398A>T;6400A>T;7688G>A;9428C>G;11928G>A	2DL3*00101_unconfirmed_var314	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;5114delGA;5968delT;6109G>A;6333delAC;6391T>A;6393T>A;6394insT;6395insA	2DL3*00101_unconfirmed_var315	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	5114delGA;5968delT;6308delTA;6391insA;14357G>C	2DL3*00101_unconfirmed_var316	Unconfirmed	N/A
2DL3*00101	2DL3*0010109	2774C>A;6390insT;6391insA;6392insT;6393insA;6402T>A;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00101_unconfirmed_var328	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6357insA;6358insT;6359insA;6360insT;6361insA;6362insA;6364T>C;6368A>T;6370C>T;6399T>A;8887G>T	2DL3*00101_var2	Confirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;6388insT;6389insA;8887G>T	2DL3*00101_var3	Confirmed	N/A
2DL3*00101	2DL3*0010110	6367insA;6368insT;6369insA;6370insT;6399T>A	2DL3*00101_var4	Confirmed	N/A
2DL3*00101	2DL3*0010110	6367insA;6368insT;6369insA;6370insT;6399T>A;8887G>T	2DL3*00101_var5	Confirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6367insA;6368insT;6369insA;6370insT;6399T>A;8887G>T	2DL3*00101_var7	Confirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6108C>T;6367insA;6368insT;6369insA;6370insT	2DL3*00101_var8	Confirmed	N/A
2DL3*00101	2DL3*0010110	5114delGA;6385insA;6386insT;6387insA;6388insT;6389insA;8887G>T	2DL3*00101_var9	Confirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;8887G>T	2DL3*00101_var11	Confirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6385insA;6386insT;6387insA	2DL3*00101_var12	Confirmed	N/A
2DL3*00101	2DL3*0010110	6334T>C;6357insA;6358insT;6360C>T;6364T>A;6366T>C;6368A>T;6370C>T;6399T>A;8887G>T	2DL3*00101_var13	Confirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6369insA;6370insT;6399T>A;8887G>T	2DL3*00101_var14	Confirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6369insA;6370insT	2DL3*00101_var15	Confirmed	N/A
2DL3*00101	2DL3*0010110	6399T>A;6401T>A;6403T>A;9169G>C	2DL3*00101_var16	Confirmed	2DL3*0010114 (LR593994)
2DL3*00101	2DL3*0010110	5114delGA;5968delT;6308delT;6308delA;6367insA;6368insT;6369insA;6370insT;6399T>A	2DL3*00101_var17	Confirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00101	2DL3*0010110	5968delT;6347insA;6348insC;6349insA;6350insT;6351insA;6352insT;6353insA;6354insT;6355insA;6356insA;6357insA;6358insC;6360C>T;6368A>T;6370C>T;6399insA;7559T>G;9521A>C;12088T>A	2DL3*00101_unconfirmed_var2	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	1789T>G;6336C>T;6338C>T;6340C>T;6349insA;6350insC;6351insA;6352insC;6353insA;6354insC;6393insA;6394insT;6395insA;6396insT;6397insA;6400T>A	2DL3*00101_unconfirmed_var3	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5114delGA;5968delT;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA;6392insT;6393insA;6394insT;6395insA;6396insT;6397insA;7559T>G;9521A>C;12088T>A	2DL3*00101_unconfirmed_var4	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308insTA;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA	2DL3*00101_unconfirmed_var5	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6384insT;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT	2DL3*00101_unconfirmed_var6	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6357insA;6358insT;6359insA;6360insA;6361insA;6362insC;6363insA;6364insT;6368A>T;6370C>T;6399T>A	2DL3*00101_unconfirmed_var7	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	503G>A;4511C>G;5114delGA;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA;6392insT;6393insA	2DL3*00101_unconfirmed_var8	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA;6400T>A	2DL3*00101_unconfirmed_var9	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6333T>C;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA	2DL3*00101_unconfirmed_var10	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6334T>C;6357insA;6358insT;6359insA;6360insA;6361insA;6362insC;6363insA;6364insT;6368A>T;6370C>T	2DL3*00101_unconfirmed_var11	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA;8887G>T	2DL3*00101_unconfirmed_var14	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA;6400T>A	2DL3*00101_unconfirmed_var15	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;6388insT;6389insA	2DL3*00101_unconfirmed_var16	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;6388insT;6389insA;8887G>T;9178G>A	2DL3*00101_unconfirmed_var17	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6308delT;6308delA;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA;8887G>T	2DL3*00101_unconfirmed_var18	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308delT;6308delA;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA;6392insT;6393insA;6399delT;8887G>T	2DL3*00101_unconfirmed_var19	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6357insA;6358insT;6359insA;6360insT;6361insA;6362insA;6364T>C;6368A>T;6370C>T	2DL3*00101_unconfirmed_var20	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;6388insT;6389insA;8692G>T;8887G>T	2DL3*00101_unconfirmed_var21	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6357insA;6358insT;6359insA;6360insT;6361insA;6362insA;6364T>C;6368A>T;6370C>T;10899C>G	2DL3*00101_unconfirmed_var22	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;6388insT;6389insA;6400T>A	2DL3*00101_unconfirmed_var23	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6108C>T;6385insA;6386insT;6387insA;6388insT;6389insA	2DL3*00101_unconfirmed_var24	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	1789T>G;5968delT;6308insTA;6385insA;6386insT;6387insA	2DL3*00101_unconfirmed_var28	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6385insA;6386insT;6387insA;6388insT;6389insA	2DL3*00101_unconfirmed_var29	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5114delGA;5968delT;6308delT;6308delA;6332T>C;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA;6392insT;6393insA	2DL3*00101_unconfirmed_var30	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308insTA;6385insA;6386insT;6387insA	2DL3*00101_unconfirmed_var31	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6385insA;6386insT;6387insA;6388insT;6389insA;8887G>T	2DL3*00101_unconfirmed_var33	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308delT;6308delA;6385insA;6386insT;6387insA;6388insT;6389insA;6390insT;6391insA;8887G>T	2DL3*00101_unconfirmed_var34	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6367insA;6368insT;6369insA;6370insT;6399T>A;12211C>A	2DL3*00101_unconfirmed_var35	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	739T>A;5114delGA;5968delT;6385T>A;6387insA;6388insT;6389insA;6390insT;6391insA;6392insT;6393insA	2DL3*00101_unconfirmed_var36	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00101	2DL3*0010110	5968delT;6336C>T;6355insA;6356insC;6357insA;6358insC;6360C>T;6366T>A;6368A>C;6370C>T;6399T>A;8887G>T	2DL3*00101_unconfirmed_var37	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6367insA;6368insT;6369insA;6370insT;6399T>A;12211C>A	2DL3*00101_unconfirmed_var38	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;12211C>A	2DL3*00101_unconfirmed_var39	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6308delT;6308delA;6385T>A;6387insA;6388insT;6389insA;6390insT;6391insA;11451T>G	2DL3*00101_unconfirmed_var40	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6334T>C;6348C>T;6353insA;6354insA;6355insA;6356insC;6368A>T;6370C>T	2DL3*00101_unconfirmed_var41	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA	2DL3*00101_unconfirmed_var42	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6367insA;6368insT;6369insA;6370insT;6399T>A;7075G>A;8887G>T	2DL3*00101_unconfirmed_var43	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6108C>T;6367insA;6368insT;6369insA;6370insT;6399T>A	2DL3*00101_unconfirmed_var44	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6384insT;6385insA;6386insT;8887G>T	2DL3*00101_unconfirmed_var45	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	1789T>G;6369insA;6370insT;6399T>A	2DL3*00101_unconfirmed_var47	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5114delGA;5968delT;6385insA;6386insT;6387insA;6388insT;6389insA;8887G>T	2DL3*00101_unconfirmed_var48	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6348C>T;6354T>A;6356A>C;6358C>T;6386insT;6387insA;6388insT	2DL3*00101_unconfirmed_var49	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6385insA;6386insT;6387insA;8887G>T	2DL3*00101_unconfirmed_var50	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6385insA;6386insT;6387insA;6400T>A;8887G>T	2DL3*00101_unconfirmed_var51	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	503G>A;2774C>A;5968delT;6385insA;6386insT;6387insA;6400T>A	2DL3*00101_unconfirmed_var53	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6385insA;6386insT;6387insA;6400T>A	2DL3*00101_unconfirmed_var54	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308delT;6308delA;6385insA;6386insT;6387insA;6388insT;6389insA;8887G>T	2DL3*00101_unconfirmed_var55	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308delT;6308delA;6385insA;6386insT;6387insA;6388insT;6389insA	2DL3*00101_unconfirmed_var56	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5114delGA;5968delT;6385insA;6386insT;6387insA;6388insT;6389insA	2DL3*00101_unconfirmed_var57	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	503G>A;2774C>A;5114delGA;6334T>C;6346C>T;6349insA;6350insT;6351insA;6352insT;6353insA;6354insA;6355insA;6356insC;6357insA;6358insT;6360C>T;6368A>T;6370C>T;6398delTTTTTT	2DL3*00101_unconfirmed_var58	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6369insA;6370insT;6399T>A;8023T>C	2DL3*00101_unconfirmed_var60	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6334T>C;6357insA;6358insT;6360C>T;6364T>A;6366T>C;6368A>T;6370C>T;6399T>A;8887G>T	2DL3*00101_unconfirmed_var61	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6108C>T;6308delT;6308delA;6367insA;6368insT;6369insA;6370insT	2DL3*00101_unconfirmed_var62	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6385insA;6386insT;6387insA;6977delT	2DL3*00101_unconfirmed_var63	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6369insA;6370insT;6399T>A;12211C>A	2DL3*00101_unconfirmed_var64	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6308delT;6308delA;6385insA;6386insT;6387insA;8887G>T	2DL3*00101_unconfirmed_var66	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308delT;6308delA;6385insA;6386insT;6387insA	2DL3*00101_unconfirmed_var67	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308delT;6308delA;6385insA;6386insT;6387insA;8887G>T	2DL3*00101_unconfirmed_var68	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308delT;6308delA;6385T>A;6387insA;6388insT;6389insA;8887G>T	2DL3*00101_unconfirmed_var69	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6308delT;6308delA;6384insT;6385insA;6386insT	2DL3*00101_unconfirmed_var70	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6108C>T;6308delT;6308delA;6369insA;6370insT	2DL3*00101_unconfirmed_var71	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	74A>G;4922A>G;5968delT;6334T>C;6348C>T;6354T>A;6356A>C;6358C>T;6387T>A;11328G>T;13743G>C	2DL3*00101_unconfirmed_var73	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6399T>A;6401T>A;6403T>A;9169G>C	2DL3*00101_unconfirmed_var74	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6399T>A	2DL3*00101_unconfirmed_var76	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5114delGA;6348C>T;6354T>A;6356A>C;6358C>T;6387T>A;6389insA;8887G>T	2DL3*00101_unconfirmed_var77	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	74A>G;4922A>G;5968delT;6332T>C;6334T>C;6345delAC;6360C>T;6366T>A;6368A>C;6370C>T;6399T>A;11328G>T;13743G>C	2DL3*00101_unconfirmed_var80	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6333delAT;6348C>T;6354T>A;6356A>C;6358C>T;6387T>A	2DL3*00101_unconfirmed_var81	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5114delGA;5968delT;6308delT;6308delA;6385insA;6386insT;6387insA;6388insT;6389insA;6399delTTT	2DL3*00101_unconfirmed_var82	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00101	2DL3*0010110	-175A>G;-20C>A;52C>A;345delG;503G>A;2048delT;3227delG;4475insT;5114delGA;6308delT;6308delA;6334C>T;6385T>A;6387insA;11620delC	2DL3*00101_unconfirmed_var88	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;6388insT;6389insA;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00101_unconfirmed_var321	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	6385insA;6386insT;6387insA;6388insT;6389insA;8887G>T;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00101_unconfirmed_var322	Unconfirmed	N/A
2DL3*00101	2DL3*0010110	5968delT;6367insA;6368insT;6369insA;6370insT;6399T>A;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00101_unconfirmed_var323	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	6396delTAT;14398C>G;14401G>T	2DL3*00101_var19	Confirmed	N/A
2DL3*00101	2DL3*0010111	6396delTATT;14398C>G;14401G>T	2DL3*00101_var20	Confirmed	N/A
2DL3*00101	2DL3*0010111	6394delTATAT	2DL3*00101_var24	Confirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6394delTATATT;14398C>G;14401G>T	2DL3*00101_var30	Confirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6399insA	2DL3*00101_unconfirmed_var86	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6399T>A	2DL3*00101_unconfirmed_var91	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	6359delATAT;14398C>G;14401G>T	2DL3*00101_unconfirmed_var106	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5114delGA;6308delT;6308delA;6398delT	2DL3*00101_unconfirmed_var111	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6394delTATAT;14398C>G;14401G>T	2DL3*00101_unconfirmed_var116	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6394delTATATT;10148C>T;14398C>G;14401G>T	2DL3*00101_unconfirmed_var133	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6396delTATTTT;14398C>G;14401G>T	2DL3*00101_unconfirmed_var134	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6394delTATATT	2DL3*00101_unconfirmed_var140	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	6334T>C;6348C>T;6354T>A;6356A>C;6358C>T;6392delTATATATT;10148C>T;14398C>G;14401G>T	2DL3*00101_unconfirmed_var146	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6308delT;6308delA;6396delTATTT;14398C>G;14401G>T	2DL3*00101_unconfirmed_var148	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5114delGA;5968delT;6394delTATAT	2DL3*00101_unconfirmed_var153	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6392delTATATATT	2DL3*00101_unconfirmed_var173	Unconfirmed	N/A
2DL3*00101	2DL3*0010111	5968delT;6334T>C;6348C>T;6354T>A;6356A>C;6358C>T;6392delTATATATT;14398C>G;14401G>T	2DL3*00101_unconfirmed_var174	Unconfirmed	N/A
2DL3*00101_c.142C>A	2DL3*0010108	2774C>A;3488C>A;5968delT;6308insTA;6388insT;6389insA;6402T>A;10606T>C	2DL3*00101_c.142C>A_var1	Confirmed	N/A
2DL3*00101_c.274C>G	2DL3*0010109	3620C>G;5968delT;6399insT;6400insT	2DL3*00101_c.274C>G_var2	Confirmed	N/A
2DL3*00101_c.549C>T	2DL3*0010103	2216T>C;4511G>C;5410C>T;5968delT;6391T>A;6393insA;6394insT;6395insA	2DL3*00101_c.549C>T_var3	Confirmed	N/A
2DL3*00101_c.598G>A	2DL3*0010103	5459G>A;5968delT;6363insA;6364insT;14030C>T	2DL3*00101_c.598G>A_var4	Confirmed	N/A
2DL3*00110	None available		2DL3*00110_var1	Confirmed	2DL3*00110 (LR593955)
2DL3*00110	None available		2DL3*00110_var2	Confirmed	N/A
2DL3*00110	None available		2DL3*00110_var3	Confirmed	N/A
2DL3*00110	None available		2DL3*00110_var4	Confirmed	N/A
2DL3*00110	None available		2DL3*00110_unconfirmed_var1	Unconfirmed	N/A
2DL3*00110	None available		2DL3*00110_unconfirmed_var2	Unconfirmed	N/A
2DL3*00110	None available		2DL3*00110_unconfirmed_var3	Unconfirmed	N/A
2DL3*00110	None available		2DL3*00110_unconfirmed_var4	Unconfirmed	N/A
2DL3*00110	None available		2DL3*00110_unconfirmed_var5	Unconfirmed	N/A
2DL3*00110	None available		2DL3*00110_unconfirmed_var6	Unconfirmed	N/A
2DL3*00110	None available		2DL3*00110_unconfirmed_var7	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00110	None available		2DL3*00110_unconfirmed_var8	Unconfirmed	N/A
2DL3*00110	None available		2DL3*00110_unconfirmed_var11	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATA;6389insATATA;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_var1	Confirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATATA;6336C>T;6391T>A;6393T>A;6395delTTTT;14277C>T	2DL3*00201_var2	Confirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATATA;6398delT;6398delT;6398delT;6398delT;6398T>A;14277C>T	2DL3*00201_var3	Confirmed	N/A
2DL3*00201	2DL3*0020101	6308insTA;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_var4	Confirmed	N/A
2DL3*00201	2DL3*0020101	6398delT;6398delT;6398T>A;6571A>T;14330C>G;14342G>A	2DL3*00201_var5	Confirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6390delT	2DL3*00201_var6	Confirmed	N/A
2DL3*00201	2DL3*0020101	6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_var7	Confirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6398delT;6398delT	2DL3*00201_var8	Confirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_var9	Confirmed	N/A
2DL3*00201	2DL3*0020101	4141C>A;6344C>T;6346C>T;6348C>T;6350C>T;6358T>C;6360T>C;6362T>C;6372T>C;6398delT;6398delT;6398delT;6398delT;6398delT;6398T>A;14277C>T	2DL3*00201_var10	Confirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6398delT;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_var11	Confirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6334T>C;6348C>T;6354T>C;6356C>T;6390delT;6390delTTTT;6571A>T;14330C>G;14342G>A	2DL3*00201_var12	Confirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_var13	Confirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6396delT;6396delA;6396delT;6396delT	2DL3*00201_var14	Confirmed	N/A
2DL3*00201	2DL3*0020101	6308insTATATATA;6389insATATA;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var3	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATATA;6348T>C;6389insATATA;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var8	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6333insATATAC;6389insATATA;6391T>A;6393T>A;6395T>A;6419A>G;9178G>A;14277C>T	2DL3*00201_unconfirmed_var9	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	-152C>A;6333insATATATAC;6389insA;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var10	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6335insATATATAC;6348C>T;6354T>C;6356C>T;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var13	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308insTA;6389T>A;6391T>A;6393T>A;6395T>A;6397insA;6398insTTTT;14277C>T	2DL3*00201_unconfirmed_var15	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATA;6389T>A;6391T>A;6393T>A;6395T>A;6397insA;6398insTAT;12459T>G;12670delG;14277C>T	2DL3*00201_unconfirmed_var16	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATA;6389insATA;6391T>A;6393T>A;6395T>A;14135C>T;14277C>T	2DL3*00201_unconfirmed_var18	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6334C>T;6348T>C;6354C>T;6356T>C;6389T>A;6391T>A;6393T>A;6395T>A;6397insA;6398insTATAT;14277C>T	2DL3*00201_unconfirmed_var19	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;6336C>T;6390insTATT;9416insT;11831G>A	2DL3*00201_unconfirmed_var20	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968delT;6335insATATAC;6348C>T;6354T>C;6356C>T;11831G>A	2DL3*00201_unconfirmed_var22	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATATA;6390delT;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var23	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATATATA;6390delT;6390delTT;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var24	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5114insG;6308insTA;6334C>T;6389insA;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var25	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATA;6389insA;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var26	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6334C>T;6336C>T;6348T>C;6354C>T;6356T>C;6389T>A;6391T>A;6393T>A;6395T>A;6397insA;6398insTTTT;6787delT;14277C>T	2DL3*00201_unconfirmed_var27	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	-152C>A;5968delT;6308insTATA;6389insA;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var28	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6334T>C;6348C>T;6354T>C;6356C>T;6391T>A;6393T>A;6395T>A;6397T>A;6398insTATT;12816C>G;14277C>T	2DL3*00201_unconfirmed_var29	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00201	2DL3*0020101	4850T>C;6308insTATATA;6388delTAT;11831G>A	2DL3*00201_unconfirmed_var30	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308insTATATATA;6390delT;6390delTTTT;6395T>A;14277C>T	2DL3*00201_unconfirmed_var31	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6348C>T;6354T>C;6356C>T;6391T>A;6393T>A;6395T>A;6397T>A;6398insTT;14277C>T	2DL3*00201_unconfirmed_var33	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATATA;6390delT;6390delTT;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var34	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308insTATATA;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var35	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968delT;6308insTA;6334C>T;6389insT;6961G>A;11831G>A	2DL3*00201_unconfirmed_var36	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308insTATATA;6398delT;6398delT;6398delT;6398delT;6398T>A;14277C>T	2DL3*00201_unconfirmed_var37	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATA;6391T>A;6393T>A;6395T>A;6397delTT;14277C>T	2DL3*00201_unconfirmed_var38	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968delT;6308insTA;11831G>A	2DL3*00201_unconfirmed_var39	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968delT;6308insTATA;6398delT;6398delT;6398T>A;11831G>A	2DL3*00201_unconfirmed_var40	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6348T>C;6354C>T;6356T>C;6389insA;6391T>A;6393T>A;6395T>A;6419A>G;9178G>A;14277C>T	2DL3*00201_unconfirmed_var41	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5132insG;5968delT;6334C>T;6348T>C;6354C>T;6356T>C;6389insA;6391T>A;6393T>A;14277C>T	2DL3*00201_unconfirmed_var42	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308insTATATA;6390delT;6390delTTTT;6395T>A;14277C>T	2DL3*00201_unconfirmed_var44	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATA;6336C>T;6390delT;6390delTT;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var45	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATATA;6390delT;6390delTTTT;6395T>A;14277C>T	2DL3*00201_unconfirmed_var46	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6333T>C;6333T>C;6344C>T;6346C>T;6350T>C;6354C>T;6389insA;6391T>A;6393T>A;6395T>A;12816C>G;14277C>T	2DL3*00201_unconfirmed_var47	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;6348C>T;6350C>T;6362T>C;6364T>C;6368C>T;6372T>C;6398T>A;11831G>A	2DL3*00201_unconfirmed_var48	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	1569G>C;6334C>T;6348T>C;6354C>T;6356T>C;6389T>A	2DL3*00201_unconfirmed_var49	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968insT;6308insTA;6398delT;6398delT;6398delT;11831G>A	2DL3*00201_unconfirmed_var50	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;6348C>T;6350C>T;6352C>T;6362T>C;6364T>C;6366T>C;6368C>T;6374T>C;6398T>A;6398T>A;11831G>A	2DL3*00201_unconfirmed_var51	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5717G>A;5968delT;6334C>T;6336C>T;6338C>T;6348T>C;6350T>C;6354C>T;6358T>C;6389T>A;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var52	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	-152C>A;320delG;5114insG;5115insA;5968delT;6336C>T;6390delT;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var53	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6336C>T;6350T>C;6356C>T;6358T>C;6390delT;6391T>A;6393T>A;6395T>A;6419A>G;9178G>A;14277C>T	2DL3*00201_unconfirmed_var54	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5630C>T;5968delT;6334C>T;6348T>C;6354C>T;6356T>C;6389T>A;6391T>A;6393T>A;14277C>T	2DL3*00201_unconfirmed_var55	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6336C>T;6350T>C;6356C>T;6358T>C;6390delT;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var57	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6336C>T;6338C>T;6340C>T;6350T>C;6352T>C;6354T>C;6356C>T;6362T>C;6391T>A;6393T>A;6395T>A;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var58	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5630C>T;5968delT;6334C>T;6336C>T;6348T>C;6350T>C;6354C>T;6358T>C;6389T>A;6391T>A;6393T>A;14277C>T	2DL3*00201_unconfirmed_var59	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6334C>T;6348T>C;6354C>T;6356T>C;6389T>A;6391T>A;6393T>A;7916T>C;14277C>T	2DL3*00201_unconfirmed_var60	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6336C>T;6350T>C;6356C>T;6358T>C;6391T>A;6393T>A;6395T>A;6397delTT;14277C>T	2DL3*00201_unconfirmed_var61	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5630C>T;5968delT;6336C>T;6338C>T;6350T>C;6352T>C;6356C>T;6360T>C;6390delT;6391T>A;14277C>T	2DL3*00201_unconfirmed_var62	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968delT;6390delT;6391T>A;11831G>A	2DL3*00201_unconfirmed_var63	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00201	2DL3*0020101	6389delAT	2DL3*00201_unconfirmed_var64	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308insTA;6398delT;6398delT;6398delT;6398delT;6398T>A;14277C>T	2DL3*00201_unconfirmed_var65	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968insT;6344C>T;6358T>C;6364C>T;6366T>C;6398delT;6398delT;6398delT;11831G>A	2DL3*00201_unconfirmed_var66	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTATA;6398delT;6398delT;6398delT;6398delT;6398delT;6571A>T;8535T>C;14330C>G;14342G>A	2DL3*00201_unconfirmed_var67	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308insTA;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var68	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6390delT;6390delTT;6393T>A;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var69	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	1569G>C;6398delT;6398delT;6398delT;6398delT	2DL3*00201_unconfirmed_var70	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	2854A>G;5968delT;6336C>T;6390delT;6390delTT;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var71	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6342T>C;6356C>T;6362T>C;6364C>T;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var72	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5122G>C;6344C>T;6358T>C;6364C>T;6366T>C;6398delT;6398delT;6398delT;6398delT;11831G>A	2DL3*00201_unconfirmed_var74	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	1569G>C;5968delT;6344C>T;6358T>C;6364C>T;6366T>C;6398delT;6398delT;6398delT;6398delT	2DL3*00201_unconfirmed_var75	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968delT;6348C>T;6350C>T;6362T>C;6368C>T;6370T>C;6398delT;6398delA;6398delT;6398delT;11831G>A	2DL3*00201_unconfirmed_var76	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	1569G>C;6388delTATTT	2DL3*00201_unconfirmed_var77	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;6398delT;6398delA;6398delT;6398delA;6398delT;11831G>A	2DL3*00201_unconfirmed_var78	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6388delTAT;7155C>T;8372G>A	2DL3*00201_unconfirmed_var79	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6389insA	2DL3*00201_unconfirmed_var80	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6388delTAT	2DL3*00201_unconfirmed_var81	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	3782C>G;6308delT;6308delA;6398delT;6398delT;6398delT;6571A>T	2DL3*00201_unconfirmed_var82	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6336T>C;6338T>C;6350C>T;6352C>T;6356T>C;6360C>T;6394delT;6394delT;6394delT;6394delT;6394delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var84	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	1133C>T;1137A>T;1138C>T;5968delT;6308A>T;6398delT;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var85	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6398delT;6398delT;6398delT;6398delT;6398delT;6398delT;6398T>A;14277C>T	2DL3*00201_unconfirmed_var86	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6308delT;6308delA;6398delT;6398delT	2DL3*00201_unconfirmed_var87	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4247T>C;5968delT;6388delTATTT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var88	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	1569G>C;5968delT;6398delT;6398delA;6398delT;6398delA;6398delT	2DL3*00201_unconfirmed_var89	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968delT;6308delT;6308delA;6388delTAT;11831G>A	2DL3*00201_unconfirmed_var90	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var91	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6398delT;6398delA;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var92	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6344C>T;6398delT;6398delT;6398delT;6398delT;6398delT;6398delT;6398T>A;14277C>T	2DL3*00201_unconfirmed_var93	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	1569G>C;5968delT;6398delT;6398delA;6398delT;6398delT;6398delT;6398delT	2DL3*00201_unconfirmed_var94	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var95	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6398delT;6398delT;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var96	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6398delT;6398delA;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var97	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6398delT;6398delT;6398delT;6398delT;6571A>T	2DL3*00201_unconfirmed_var98	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00201	2DL3*0020101	4850T>C;5968delT;6348C>T;6362T>C;6368C>T;6370T>C;6398delT;6398delA;6398delT;6398delT;6398delT;6398delT;11831G>A	2DL3*00201_unconfirmed_var99	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5968delT;6398delT;6398delA;6398delT;6398delT;6398delT;6398delT;11831G>A	2DL3*00201_unconfirmed_var100	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	4850T>C;5122G>C;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6398delT;11831G>A	2DL3*00201_unconfirmed_var101	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	162C>T;163delTGGGCCTG;172A>G;173G>C;179A>G;180T>G;181C>A;183G>T;1938delC;5968delT;6308insTATA;6390delT;6391T>A;6393T>A;6395T>A;6397T>A;14277C>T	2DL3*00201_unconfirmed_var102	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6346T>C;6360C>T;6366T>C;6368C>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var103	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6386delTATAT	2DL3*00201_unconfirmed_var104	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	155delTGGA;162delCTG;167C>G;173delGTG;177A>G;4212delC;5968delT;6308insTATATATA;6336C>T;6390delT;6390delTT;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var105	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6398delT;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var106	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6398delT;6398delA;6398delT;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var107	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6337delA;6337delT;6337delA;6337delT;6337delA;6337delC;6390delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var108	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6398delT;6398delT;6398delT	2DL3*00201_unconfirmed_var109	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	1569G>C;5968delT;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6398delT	2DL3*00201_unconfirmed_var110	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6346T>C;6360C>T;6366T>C;6368C>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var111	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6308delT;6308delA;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var112	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6346T>C;6360C>T;6366T>C;6368C>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6398delT;6398delT;6571A>T	2DL3*00201_unconfirmed_var113	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6388delTAT	2DL3*00201_unconfirmed_var114	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6308delT;6308delA;6398delT;6398delT;6398delT;6398delT;6571A>T;12313G>A	2DL3*00201_unconfirmed_var115	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6343delA;6343delT;6343delA;6343delT;6343delAC;6398delT;6398delT	2DL3*00201_unconfirmed_var117	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6348C>T;6350C>T;6352C>T;6354C>T;6356C>T;6362T>C;6364T>C;6366T>C;6370T>C;6378T>C;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6398delT;6398delT;6571A>T	2DL3*00201_unconfirmed_var118	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	992C>A;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6388delTAT	2DL3*00201_unconfirmed_var120	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6308delT;6308delA;6344C>T;6396delT;6396delA;6396delT;6396delT;6396delT;6571A>T	2DL3*00201_unconfirmed_var121	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6398delT;6398delT;6398delT;6398delT;6398delT;6571A>T	2DL3*00201_unconfirmed_var123	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6396delT;6396delA;6396delT;6396delT;6396delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var124	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6386delTATATTT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var125	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6386delTATATTT	2DL3*00201_unconfirmed_var126	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6388delTAT	2DL3*00201_unconfirmed_var127	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	1938delC;6308delT;6308delA;6386delTATATTT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var128	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6308insTATA;6389T>A;6391T>A;6393T>A;6395T>A;6397insA;6398insTATAT;14277C>T	2DL3*00201_unconfirmed_var129	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6308delT;6308delA;6396delT;6396delA;6396delT;6396delT;6396delT;6396delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var131	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	6308delT;6308delA;6308delT;6308delA;6396delT;6396delA;6396delT;6396delT;6396delT;6396delT;6571A>T;11879C>G;14330C>G;14342G>A	2DL3*00201_unconfirmed_var132	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6389A>T;6391delT;6391delTTT	2DL3*00201_unconfirmed_var134	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6395delATAT;6398delT;6398delT	2DL3*00201_unconfirmed_var135	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6396delT;6396delA;6396delT;6396delT;6396delT;6396delT;11475G>T	2DL3*00201_unconfirmed_var138	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;4850T>C;5968delT;6308insTATA;6391T>A;6393insA;6394insT;6395insA;6396insT;11831G>A	2DL3*00201_unconfirmed_var140	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6396delT;6396delA;6396delT;6396delT;6396delT;6396delT;6396delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var142	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6386delTATAT	2DL3*00201_unconfirmed_var144	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6388delTATTT	2DL3*00201_unconfirmed_var147	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6389A>T;6393delT;6393delTTTTT;6571A>T	2DL3*00201_unconfirmed_var150	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	880delG;1890delC;1966delC;3122delC;5116delGA;5968delT;6398delT;6398delT;6398delT;6398delT;6398delT;6571A>T;11805delG;12676delG;14330C>G;14342G>A	2DL3*00201_unconfirmed_var159	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;4850T>C;6308insTA;6389insT;9416insT;11831G>A	2DL3*00201_unconfirmed_var166	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6337delA;6337delT;6337delA;6337delT;6337delA;6337delT;6337delA;6337delC;6386delTATATTT;6571A>T	2DL3*00201_unconfirmed_var168	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6308insTATA;6348C>T;6354T>C;6356C>T;6391T>A;6393T>A;6395T>A;6397T>A;14277C>T	2DL3*00201_unconfirmed_var172	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	5968delT;6337delA;6337delT;6337delA;6337delT;6337delA;6337delT;6337delA;6337delC;6388delTATTTT;13609G>A	2DL3*00201_unconfirmed_var175	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;4745delAGAG;5968delT;6308insTATATA;6389insA;6391T>A;6393T>A;6395T>A;14277C>T	2DL3*00201_unconfirmed_var177	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6308insTATA;6390delT;6390delTT;6393T>A;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var178	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;4929A>T;5968delT;6308insTATA;6390delT;6390delTT;6393T>A;14277C>T	2DL3*00201_unconfirmed_var179	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6348T>C;6354C>T;6356T>C;6389T>A;6391T>A;6393T>A;6419A>G;9178G>A;14277C>T	2DL3*00201_unconfirmed_var180	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;4100A>C;4147insAT;5968delT;6356C>T;6362T>C;6364C>T;6398delT;6398delT;6398delT;6398delT;6398delT;6398delT;8714T>G;14277C>T	2DL3*00201_unconfirmed_var183	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6308insTA;6398delT;6398delT;6398delT;6398delT;6398delT;6398delT;6398T>A;14277C>T	2DL3*00201_unconfirmed_var184	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6334T>C;6348C>T;6354T>C;6356C>T;6390delT;6390delTTTT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var185	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6308insTA;6398delT;6398delT;6398delT;6398delT;6398delT;6398delT;6398delT;14277C>T	2DL3*00201_unconfirmed_var186	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6342T>C;6356C>T;6362T>C;6364C>T;6398delT;6398delT;6398delT;6398delT;6398delT;6398delT;6398delT;14277C>T	2DL3*00201_unconfirmed_var187	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;4850T>C;5968delT;6348C>T;6398delT;6398delA;6398delT;6398delT;6398delT;6398delT;11831G>A	2DL3*00201_unconfirmed_var189	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6308delT;6308delA;6308delT;6308delA;6398delT;6398delT;6398delT;6398delT;6571A>T	2DL3*00201_unconfirmed_var190	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6308delT;6308delA;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var191	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5114insG;5115insA;5968delT;6308delT;6308delA;6308delT;6308delA;6398delT;6398delT;6398delT;6398T>A;14277C>T	2DL3*00201_unconfirmed_var192	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;1569G>C;5968delT;6360C>T;6366T>C;6368C>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6398delT;8909G>C	2DL3*00201_unconfirmed_var193	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6308delT;6308delA;6308delT;6308delA;6398delT;6398delT;6398delT;6398delT;6398delT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var194	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6390delT;6390delTTT;11475G>T	2DL3*00201_unconfirmed_var195	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6308delT;6308delA;6308delT;6308delA;6389A>T;6391delT;6391delTTT	2DL3*00201_unconfirmed_var197	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6308delT;6308delA;6308delT;6308delA;6389A>T;6393delT;6393delTTTTT;6571A>T	2DL3*00201_unconfirmed_var199	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6308delT;6308delA;6308delT;6308delA;6390delT;6390delTTTT;6571A>T;12313G>A	2DL3*00201_unconfirmed_var200	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6337delA;6337delT;6337delA;6337delT;6337delA;6337delC;6388delTATTTTT;6571A>T;14330C>G;14342G>A	2DL3*00201_unconfirmed_var202	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6336C>T;6338C>T;6350T>C;6352T>C;6356C>T;6360T>C;6390delT;6391T>A;6393T>A;6395T>A;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00201_unconfirmed_var210	Unconfirmed	N/A
2DL3*00201	2DL3*0020101	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;4850T>C;5968delT;6308insTA;6398delT;6398delT;6398delT;6398delT;11831G>A;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCCTCATGCTGTTCCACCTCCCCTCAGACTAGCTTTCAGCCTTCTGTCAGCAGTAAAACTTATATATTTTTTA;14443A>G;14444A>C;14447A>G;14450T>G;14453A>G	2DL3*00201_unconfirmed_var211	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	3334A>G;5078G>A;6308delT;6308delA;6308delT;6308delA;6395delT	2DL3*00201_unconfirmed_var1	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	4100C>A;4147delA;4147delT;5078G>A;5968delT;6333delA;6333delT;6333delA;6333delT;6333delA;6333delT;6333delA;6333delT;6395T>A;6397T>A;6398insTATTT;8714G>T	2DL3*00201_unconfirmed_var2	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	-152C>A;4100C>A;4147delA;4147delT;5078G>A;5968delT;6308delT;6308delA;6308delT;6308delA;6395T>A;6397delT;8714G>T	2DL3*00201_unconfirmed_var4	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00201	2DL3*0020102	4100C>A;4147delA;4147delT;5078G>A;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6334C>T;6336C>T;6348T>C;6350T>C;6354C>T;6358T>C;6395T>A;8714G>T	2DL3*00201_unconfirmed_var5	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	4100C>A;4147delA;4147delT;5078G>A;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6395T>A;6754G>A;8714G>T	2DL3*00201_unconfirmed_var6	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	3334A>G;5078G>A;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6389delATATA	2DL3*00201_unconfirmed_var11	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	4100C>A;4147delA;4147delT;5078G>A;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6393delATT;8714G>T	2DL3*00201_unconfirmed_var12	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	4100C>A;4147delA;4147delT;4521T>C;5078G>A;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6334C>T;6336C>T;6348T>C;6350T>C;6354C>T;6358T>C;8714G>T	2DL3*00201_unconfirmed_var14	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	4100C>A;4147delA;4147delT;5078G>A;5968delT;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6334C>T;6336C>T;6348T>C;6350T>C;6354C>T;6358T>C;6395T>A;8714G>T	2DL3*00201_unconfirmed_var17	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6392delTATTTT;8428delA;11341delA	2DL3*00201_unconfirmed_var21	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	3334A>G;5078G>A;5968delT;6335delA;6335delT;6335delA;6335delT;6335delA;6335delT;6335delA;6335delT;6335delA;6335delC;6351delATACAT;6360T>C;6395T>A	2DL3*00201_unconfirmed_var32	Unconfirmed	N/A
2DL3*00201	2DL3*0020102	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5078G>A;5968delT;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6395delTTT	2DL3*00201_unconfirmed_var141	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6334C>T;6348T>C;6349delA;6349delT;6349delA;6349delC;6354T>C;9943T>G	2DL3*00201_var16	Confirmed	N/A
2DL3*00201	2DL3*0020103	6398delT;6398delA;6398delT;6398delA;6398delT;8142G>A	2DL3*00201_var17	Confirmed	N/A
2DL3*00201	2DL3*0020103	6334C>T;6348T>C;6354C>T;6356T>C;6398delT;6398delA;6398delT;6398delA;6398delT;9943T>G	2DL3*00201_var18	Confirmed	N/A
2DL3*00201	2DL3*0020103	6398delT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_var19	Confirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_var20	Confirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;14409C>T	2DL3*00201_var21	Confirmed	N/A
2DL3*00201	2DL3*0020103	6395delATAT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_var22	Confirmed	N/A
2DL3*00201	2DL3*0020103	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6334C>T;6348T>C;6354C>T;6356T>C;6397A>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;9943T>G	2DL3*00201_var23	Confirmed	N/A
2DL3*00201	2DL3*0020103	6334C>T;6348T>C;6354C>T;6356T>C;6398delT;6398delA;6722T>C;9943T>G	2DL3*00201_unconfirmed_var116	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6398delT;6398delA	2DL3*00201_unconfirmed_var119	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6398delT;6398delT;13361A>G	2DL3*00201_unconfirmed_var130	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6398delT;6398delA;6398delT;6398delT;8142G>A	2DL3*00201_unconfirmed_var136	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6334C>T;6348T>C;6349delA;6349delT;6349delA;6349delC;6354T>C;9943T>G;12446G>A	2DL3*00201_unconfirmed_var137	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6348T>C;6354C>T;6356T>C;6399delA;6399delT;6399delA;6419A>G;9178G>A	2DL3*00201_unconfirmed_var139	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6355delA;6355delT;6355delA;6355delT;8142G>A	2DL3*00201_unconfirmed_var143	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6334C>T;6348T>C;6354C>T;6356T>C;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var145	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6355delA;6355delT;6355delA;6355delT	2DL3*00201_unconfirmed_var146	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6398delT;6398delA;6398delT;6398delA;6398delT;7379G>A	2DL3*00201_unconfirmed_var148	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6355delA;6355delT;6355delA;6355delT;9943T>G	2DL3*00201_unconfirmed_var149	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00201	2DL3*0020103	6334C>T;6336C>T;6338C>T;6340C>T;6342C>T;6344C>T;6346C>T;6354C>T;6356T>C;6358T>C;6360T>C;6362T>C;6364T>C;6366T>C;6368T>C;6370T>C;6378T>C;6397A>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6419A>G	2DL3*00201_unconfirmed_var153	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6334C>T;6348T>C;6354C>T;6356T>C;6397A>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;9943T>G	2DL3*00201_unconfirmed_var154	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6334C>T;6348T>C;6354C>T;6356T>C;6397A>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;9943T>G	2DL3*00201_unconfirmed_var155	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6397delAT;6398delA;6398delT;6398delA;6398delT;6398delT;14409C>T	2DL3*00201_unconfirmed_var156	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6397delAT;6398delA;6398delT;6398delA;6398delT;6398delT;8271A>C	2DL3*00201_unconfirmed_var157	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6397A>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6796A>T;13214G>A	2DL3*00201_unconfirmed_var158	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6334C>T;6348T>C;6354C>T;6356T>C;6397delAT;6398delA;6398delT;6398delA;6398delT;6398delT;9943T>G	2DL3*00201_unconfirmed_var160	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6332T>C;6346C>T;6352T>C;6354C>T;6397delAT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var161	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6306insT;6319delA;6332delT;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var162	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6396delTAT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var163	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6353delACA;6355delT;6355delA;6355delT;6355delA;6355delT;6356T>C;9943T>G	2DL3*00201_unconfirmed_var164	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6334C>T;6348T>C;6354C>T;6356T>C;6393A>T;6395A>T;6396delTAT;6398delA;6398delT;6398delA;6398delT;6398delT;9943T>G	2DL3*00201_unconfirmed_var165	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5132insG;6309delA;6325delA;6334C>T;6348T>C;6354C>T;6356T>C;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6977delT	2DL3*00201_unconfirmed_var167	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6308delTA;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var169	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6334C>T;6348T>C;6354C>T;6356T>C;6396delTAT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var170	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	2213G>A;6395delATAT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var171	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	2980C>G;5968delT;6396delTAT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var173	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6396delTAT;6398delA;6398delT;6398delA;6398delT;6398delT;14409C>T	2DL3*00201_unconfirmed_var174	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	5968delT;6395delATAT;6398delA;6398delT;6398delA;6398delT;6398delT;14409C>T	2DL3*00201_unconfirmed_var176	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6308delTA;6392delTATATATA;6399delT;6399delA;6399delT;6399delT;6399delT;14277T>C	2DL3*00201_unconfirmed_var182	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	180delTC;319delG;416delG;6331delATACAC;6348T>C;6358T>C;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT;6398delTTT	2DL3*00201_unconfirmed_var188	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6398delT;6398delT;11942delT	2DL3*00201_unconfirmed_var196	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var198	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var201	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00201	2DL3*0020103	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6334C>T;6348T>C;6354C>T;6356T>C;6398delT;6398delA;6398delT;6398delA;6398delT;9943T>G;11942delT	2DL3*00201_unconfirmed_var203	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6332T>C;6346C>T;6352T>C;6354C>T;6397A>T;6398delT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var204	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;6353delACA;6355delT;6355delA;6355delT;6355delA;6355delT;6356T>C;9943T>G	2DL3*00201_unconfirmed_var205	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;5968delT;6353delACA;6355delT;6355delA;6355delT;6355delA;6355delT;6356T>C;9943T>G	2DL3*00201_unconfirmed_var206	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	319delG;319delG;319delG;319delC;319delC;319delT;319delG;319delG;319delA;319delG;319delT;319delG;319delG;319delA;319delG;319delA;319delT;319delA;319delT;3183G>C;5968delT;6394delTATAT;6398delA;6398delT;6398delA;6398delT;6398delT	2DL3*00201_unconfirmed_var207	Unconfirmed	N/A
2DL3*00201	2DL3*0020103	6334C>T;6347delA;6347delT;6347delA;6347delT;6347delA;6347delT;6356T>C;9943T>G;14246delCTCCAACCTAACTGGCTTACTTCCTAG;14277T>C;14279delTGAGGCTGCAATCACACTGAGGAACTCACAATTCCAAACATACAAGAGGCTCCCTCTTAACACGGCACTTAGACACGTGCTGTTCCACCTTCCC;14373T>G;14378C>A;14382T>C	2DL3*00201_unconfirmed_var209	Unconfirmed	N/A
2DL3*00201_c.171C>A	2DL3*0020101	3517C>A;6334C>T;6336C>T;6348T>C;6350T>C;6354C>T;6358T>C;6389T>A;6391T>A;6393T>A;14277C>T	2DL3*00201_c.171C>A_var5	Confirmed	N/A
2DL3*00201_c.618A>C	2DL3*0020101	5479A>C;6308delT;6308delA;6308delT;6308delA;6308delT;6308delA;6390delT;6390delTT	2DL3*00201_c.618A>C_var6	Confirmed	N/A
2DL3*003	None available		2DL3*003_var1	Confirmed	None available (LR594007)
2DL3*003	None available		2DL3*003_unconfirmed_var2	Unconfirmed	N/A
2DL3*003	None available		2DL3*003_unconfirmed_var3	Unconfirmed	N/A
2DL3*003	None available		2DL3*003_unconfirmed_var4	Unconfirmed	N/A
2DL3*003	None available		2DL3*003_unconfirmed_var5	Unconfirmed	N/A
2DL3*003	None available		2DL3*003_unconfirmed_var6	Unconfirmed	N/A
2DL3*003	None available		2DL3*003_unconfirmed_var7	Unconfirmed	N/A
2DL3*003	None available		2DL3*003_unconfirmed_var8	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_var1	Confirmed	2DL3*00501 (LR593920)
2DL3*00501	None available		2DL3*00501_var2	Confirmed	N/A
2DL3*00501	None available		2DL3*00501_var3	Confirmed	N/A
2DL3*00501	None available		2DL3*00501_var4	Confirmed	N/A
2DL3*00501	None available		2DL3*00501_var5	Confirmed	N/A
2DL3*00501	None available		2DL3*00501_var6	Confirmed	N/A
2DL3*00501	None available		2DL3*00501_var7	Confirmed	N/A
2DL3*00501	None available		2DL3*00501_var8	Confirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var1	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var2	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var3	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var4	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var5	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var6	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var7	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var8	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var9	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var11	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var12	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var13	Unconfirmed	N/A
2DL3*00501	None available		2DL3*00501_unconfirmed_var14	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
2DL3*00501_c.709G>A	None available		2DL3*00501_c.709G>A_var7	Confirmed	2DL3*036 (LR593968)
2DL3*010	2DL3*010	346delA;346delG;346delT;346delG;346delG;346delA;346delG;346delA;346delT;346delA;346delT;346delG;346delG;346delG;346delC;346delC;346delT;346delG;346delA;5970delT;6396delTATA;6399delT;6399delA;6399delT;6399delA;6399delTATAT;10407delT	2DL3*010_unconfirmed_var1	Unconfirmed	N/A
2DL3*01202	None available		2DL3*01202_unconfirmed_var1	Unconfirmed	N/A
2DL3*030	None available		2DL3*030_unconfirmed_var1	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	11915T>A	3DL1*00101_var1	Confirmed	3DL1*0010109 (LR593945)
3DL1*00101	3DL1*0010101	5388C>T;7914T>C;11915T>A	3DL1*00101_var2	Confirmed	3DL1*0010107 (LR593922)
3DL1*00101	3DL1*0010101	6109G>A;11915T>A	3DL1*00101_var3	Confirmed	3DL1*0010106 (LR593990)
3DL1*00101	3DL1*0010101	5388C>T;6108A>C;11915T>A	3DL1*00101_var4	Confirmed	3DL1*0010108 (LR593932)
3DL1*00101	3DL1*0010101	5388C>T;11915T>A	3DL1*00101_var5	Confirmed	3DL1*0010111 (LR593993)
3DL1*00101	3DL1*0010101	95C>T;11915T>A	3DL1*00101_var6	Confirmed	3DL1*0010110 (LR593985)
3DL1*00101	3DL1*0010101	5388C>T;6043delT;11915T>A	3DL1*00101_var8	Confirmed	N/A
3DL1*00101	3DL1*0010101	6043delT;6109G>A;11915T>A	3DL1*00101_var9	Confirmed	N/A
3DL1*00101	3DL1*0010101	5388C>T;6043delT;7914T>C;11915T>A	3DL1*00101_var10	Confirmed	N/A
3DL1*00101	3DL1*0010101	6043delT;11915T>A	3DL1*00101_var11	Confirmed	N/A
3DL1*00101	3DL1*0010101	5388C>T;6043delT;6108A>C;11915T>A	3DL1*00101_var12	Confirmed	N/A
3DL1*00101	3DL1*0010101	-303insG;-298insG;4096T>C;8836C>G;11915T>A	3DL1*00101_unconfirmed_var2	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	-317insGG;6109G>A;11915T>A	3DL1*00101_unconfirmed_var3	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	-303insG;5388C>T;7914T>C;11915T>A	3DL1*00101_unconfirmed_var6	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	7650G>A;11915T>A;12938G>C	3DL1*00101_unconfirmed_var7	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	5710A>G;11915T>A	3DL1*00101_unconfirmed_var8	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	2412C>A;5388C>T;7914T>C;11915T>A	3DL1*00101_unconfirmed_var9	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	346C>T;3905G>A;5388C>T;11915T>A	3DL1*00101_unconfirmed_var10	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	2580delG;6043insT;8678delT;11915T>A;12559T>C;12560G>A	3DL1*00101_unconfirmed_var11	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	6043delT;7252T>A;11915T>A	3DL1*00101_unconfirmed_var12	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	5388C>T;6057delT;7914T>C;11915T>A	3DL1*00101_unconfirmed_var16	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	5664delT;6109G>A;11915T>A	3DL1*00101_unconfirmed_var18	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	11915T>A;11952delG	3DL1*00101_unconfirmed_var19	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	4096T>C;6043delT;8836C>G;11915T>A	3DL1*00101_unconfirmed_var20	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	95C>T;6043delT;11915T>A	3DL1*00101_unconfirmed_var22	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	5388C>T;11226delT;11915T>A	3DL1*00101_unconfirmed_var23	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	2337C>T;6043delT;11915T>A	3DL1*00101_unconfirmed_var24	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	5388C>T;5664delT;7914T>C;11226delT;11915T>A	3DL1*00101_unconfirmed_var26	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	999delG;5388C>T;6043delT;7914T>C;11915T>A	3DL1*00101_unconfirmed_var27	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	3723delA;6109G>A;11915T>A;11958delG	3DL1*00101_unconfirmed_var28	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	5388C>T;6043delT;6108A>C;10877delA;11915T>A	3DL1*00101_unconfirmed_var29	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	6043delT;11226delT;11915T>A	3DL1*00101_unconfirmed_var30	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	5664delT;6043delT;11915T>A	3DL1*00101_unconfirmed_var31	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	5664delT;6043delT;6109G>A;11915T>A	3DL1*00101_unconfirmed_var32	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	5388C>T;5664delT;6055delTC;6057T>C;11915T>A	3DL1*00101_unconfirmed_var34	Unconfirmed	N/A
3DL1*00101	3DL1*0010101	3723delAAAA;6109G>A;11915T>A	3DL1*00101_unconfirmed_var41	Unconfirmed	N/A
3DL1*00101	3DL1*0010102		3DL1*0010102	Confirmed	N/A
3DL1*00101	3DL1*0010102	6043delT	3DL1*00101_var7	Confirmed	N/A
3DL1*00101	3DL1*0010102	-301A>G;-300insAGT;-299T>G;6043delT	3DL1*00101_unconfirmed_var4	Unconfirmed	N/A
3DL1*00101	3DL1*0010102	8359delT	3DL1*00101_unconfirmed_var13	Unconfirmed	N/A
3DL1*00101	3DL1*0010102	2009C>T;6043delT	3DL1*00101_unconfirmed_var14	Unconfirmed	N/A
3DL1*00101	3DL1*0010102	5664delT	3DL1*00101_unconfirmed_var17	Unconfirmed	N/A
3DL1*00101	3DL1*0010102	5664delT;6043delT;8471G>A	3DL1*00101_unconfirmed_var25	Unconfirmed	N/A
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;10939G>A	3DL1*00101_var13	Confirmed	N/A
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;6043delT;10939G>A	3DL1*00101_var14	Confirmed	N/A
3DL1*00101	3DL1*0010103	-301A>G;-300insAGT;-299T>G;473C>T;828delAGAG;10939G>A	3DL1*00101_unconfirmed_var15	Unconfirmed	N/A
3DL1*00101	3DL1*0010103	828delAGAG;3203G>A;10939G>A	3DL1*00101_unconfirmed_var36	Unconfirmed	N/A
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;9138A>C;10939G>A	3DL1*00101_unconfirmed_var37	Unconfirmed	N/A
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;3899C>T;10939G>A	3DL1*00101_unconfirmed_var38	Unconfirmed	N/A
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;8274A>T;10939G>A	3DL1*00101_unconfirmed_var39	Unconfirmed	N/A
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;8105C>T;10939G>A	3DL1*00101_unconfirmed_var40	Unconfirmed	N/A
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;5514C>G;10939G>A	3DL1*00101_unconfirmed_var42	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;5664delT;10939G>A	3DL1*00101_unconfirmed_var43	Unconfirmed	N/A
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;6043delT;10939G>A;11504A>G	3DL1*00101_unconfirmed_var44	Unconfirmed	N/A
3DL1*00101	3DL1*0010103	473C>T;828delAGAG;10939G>A;11952delG;11984delT	3DL1*00101_unconfirmed_var45	Unconfirmed	N/A
3DL1*00101_c.1149C>G;1152C>T	3DL1*0010103	473C>T;828delAGAG;10939G>A;12982C>G;12985C>T	3DL1*00101_c.1149C>G;1152C>T_var1	Confirmed	N/A
3DL1*00101_c.1303A>G	3DL1*0010103	-390delTATGGGCCTAGAGGTGGAGT;-283G>T;2027G>A;3130T>C;4463G>A;7705C>G;8350C>A;8387G>C;9132C>T;11710A>G;11814G>A;12723C>G;13254A>G	3DL1*00101_c.1303A>G_var2	Confirmed	N/A
3DL1*002	3DL1*002		3DL1*002	Confirmed	N/A
3DL1*002	3DL1*002	6215C>T	3DL1*002_var1	Confirmed	3DL1*0020103 (LR593970)
3DL1*002	3DL1*002	6215C>T;8532C>T	3DL1*002_var2	Confirmed	3DL1*0020104 (LR593935)
3DL1*002	3DL1*002	6215C>T;10953A>C	3DL1*002_var3	Confirmed	3DL1*0020102 (LR593973)
3DL1*002	3DL1*002	6043delT;6215C>T	3DL1*002_var5	Confirmed	N/A
3DL1*002	3DL1*002	6043delT	3DL1*002_var6	Confirmed	N/A
3DL1*002	3DL1*002	6043delT;6215C>T;10953A>C	3DL1*002_var7	Confirmed	N/A
3DL1*002	3DL1*002	6215C>T;6657delT;6661A>C;8236G>A	3DL1*002_var8	Confirmed	N/A
3DL1*002	3DL1*002	5664delT;6043delT;6215C>T	3DL1*002_var9	Confirmed	N/A
3DL1*002	3DL1*002	2240A>G;2243insAACC;2244C>A;2250A>C;2251G>A;2253A>G;2255C>A;2259insGGG;2265G>A;2267A>G;2268insGGGGAC;2272insT;2286insCA;2290insGCCAAATT;2291T>A;2296insA;2323insCA;2330insGGGG;2377insACG;2380G>A;2381A>T;6215C>T;7697insA	3DL1*002_unconfirmed_var1	Unconfirmed	N/A
3DL1*002	3DL1*002	-336insG	3DL1*002_unconfirmed_var3	Unconfirmed	N/A
3DL1*002	3DL1*002	6043insT;6215C>T	3DL1*002_unconfirmed_var4	Unconfirmed	N/A
3DL1*002	3DL1*002	5805G>C;6215C>T;10953A>C	3DL1*002_unconfirmed_var5	Unconfirmed	N/A
3DL1*002	3DL1*002	6215C>T;11927G>T	3DL1*002_unconfirmed_var6	Unconfirmed	N/A
3DL1*002	3DL1*002	3978T>A;6215C>T	3DL1*002_unconfirmed_var7	Unconfirmed	N/A
3DL1*002	3DL1*002	657C>T;6215C>T;6838T>C	3DL1*002_unconfirmed_var8	Unconfirmed	N/A
3DL1*002	3DL1*002	5664delT;6215C>T	3DL1*002_unconfirmed_var9	Unconfirmed	N/A
3DL1*002	3DL1*002	6043delT;6215C>T;11226delT	3DL1*002_unconfirmed_var10	Unconfirmed	N/A
3DL1*002	3DL1*002	3660delA;6043delT;6215C>T;8706insT;10626delA;10953A>C	3DL1*002_unconfirmed_var11	Unconfirmed	N/A
3DL1*002	3DL1*002	6043delT;6215C>T;10461delA	3DL1*002_unconfirmed_var12	Unconfirmed	N/A
3DL1*002	3DL1*002	6043delT;6215C>T;10877delA	3DL1*002_unconfirmed_var13	Unconfirmed	N/A
3DL1*002	3DL1*002	6043delT;6215C>T;6221delT	3DL1*002_unconfirmed_var14	Unconfirmed	N/A
3DL1*002	3DL1*002	6215C>T;9103delA;11226delT	3DL1*002_unconfirmed_var15	Unconfirmed	N/A
3DL1*002	3DL1*002	5664delT;6043delT;6215C>T;10953A>C	3DL1*002_unconfirmed_var16	Unconfirmed	N/A
3DL1*002	3DL1*002	1902delC;6215C>T;11958delG	3DL1*002_unconfirmed_var17	Unconfirmed	N/A
3DL1*002	3DL1*002	3660delA;6043delT;6215C>T;7697delA	3DL1*002_unconfirmed_var18	Unconfirmed	N/A
3DL1*002	3DL1*002	5664delT;6043delT;11226delT	3DL1*002_unconfirmed_var19	Unconfirmed	N/A
3DL1*002	3DL1*002	3723delA;5664delT;6043delT;6215C>T;10877delA	3DL1*002_unconfirmed_var20	Unconfirmed	N/A
3DL1*002_c.1195G>C	3DL1*002	10581G>C;13146G>C	3DL1*002_c.1195G>C_var3	Confirmed	3DL1*1190101 (LR593948)
3DL1*002_c.1195G>C	3DL1*002	13146G>C	3DL1*002_c.1195G>C_var4	Confirmed	3DL1*1190102 (LR594008)
3DL1*00401	3DL1*0040101		3DL1*0040101	Confirmed	N/A
3DL1*00401	3DL1*0040101	1728A>G	3DL1*00401_var2	Confirmed	3DL1*0040106 (LR593901)
3DL1*00401	3DL1*0040101	9475G>A	3DL1*00401_var3	Confirmed	3DL1*0040105 (LR593915)
3DL1*00401	3DL1*0040101	5093A>T	3DL1*00401_var4	Confirmed	3DL1*0040108 (LR593957)
3DL1*00401	3DL1*0040101	1728A>G;9668A>G	3DL1*00401_var6	Confirmed	3DL1*0040104 (LR593913)
3DL1*00401	3DL1*0040101	6055T>C;6056delC	3DL1*00401_var8	Confirmed	N/A
3DL1*00401	3DL1*0040101	5093A>T;6055T>C;6056delC	3DL1*00401_var9	Confirmed	N/A
3DL1*00401	3DL1*0040101	6055T>C;6056delC;7696delA	3DL1*00401_var12	Confirmed	N/A
3DL1*00401	3DL1*0040101	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG;6955insT	3DL1*00401_var14	Confirmed	N/A
3DL1*00401	3DL1*0040101	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG;9475G>A	3DL1*00401_var15	Confirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
3DL1*00401	3DL1*0040101	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG	3DL1*00401_var16	Confirmed	N/A
3DL1*00401	3DL1*0040101	5856T>C	3DL1*00401_unconfirmed_var2	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	2665G>C	3DL1*00401_unconfirmed_var3	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	11224G>A	3DL1*00401_unconfirmed_var4	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	2167T>C;9475G>A	3DL1*00401_unconfirmed_var5	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	1728A>G;5245C>T	3DL1*00401_unconfirmed_var7	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	1728A>G;12689delC	3DL1*00401_unconfirmed_var8	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	5664delT;10580C>T	3DL1*00401_unconfirmed_var9	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	1728A>G;6055T>C;6056delC	3DL1*00401_unconfirmed_var10	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	-182insA;5664delT;6055T>C;6056delC;9475G>A	3DL1*00401_unconfirmed_var11	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	7696delA	3DL1*00401_unconfirmed_var12	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	5664delT;9475G>A	3DL1*00401_unconfirmed_var13	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	-328delG	3DL1*00401_unconfirmed_var14	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	1728A>G;6055T>C;6056delC;9668A>G	3DL1*00401_unconfirmed_var17	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	1728A>G;5245C>T;6055T>C;6056delC	3DL1*00401_unconfirmed_var18	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	2262C>A;6055T>C;6056delC	3DL1*00401_unconfirmed_var19	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	11226delT	3DL1*00401_unconfirmed_var20	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	6055T>C;6056delC;8359delT	3DL1*00401_unconfirmed_var21	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	6055T>C;6056delC;7696delA;9475G>A	3DL1*00401_unconfirmed_var22	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	6055T>C;6056delC;11226delT;12689C>T	3DL1*00401_unconfirmed_var23	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	5093A>T;6055T>C;6056delC;11226delT	3DL1*00401_unconfirmed_var24	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	1728A>G;5664delT;6055T>C;6056delC;7696delA	3DL1*00401_unconfirmed_var25	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	-308delG;-308delG;-308delC;-308delC;-308delT;-308delG;-308delG;-308A>G;-308G>C;-308delG;6043delT;7696delA	3DL1*00401_unconfirmed_var29	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG;6055T>C;6056delC;6955insT	3DL1*00401_unconfirmed_var32	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG;6055T>C;6056delC;6955insT;7696delA	3DL1*00401_unconfirmed_var35	Unconfirmed	N/A
3DL1*00401	3DL1*0040101	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG;5093A>T;6055T>C;6056delC	3DL1*00401_unconfirmed_var36	Unconfirmed	N/A
3DL1*00401	3DL1*0040102		3DL1*0040102	Confirmed	N/A
3DL1*00401	3DL1*0040102	12548G>C	3DL1*00401_var1	Confirmed	3DL1*0040107 (LR593908)
3DL1*00401	3DL1*0040102	5588G>A;7650G>A	3DL1*00401_var5	Confirmed	3DL1*0040109 (LR593976)
3DL1*00401	3DL1*0040102	6055T>C;6056delC	3DL1*00401_var7	Confirmed	N/A
3DL1*00401	3DL1*0040102	6055T>C;6056delC;12548G>C	3DL1*00401_var10	Confirmed	N/A
3DL1*00401	3DL1*0040102	6055T>C;6056delC;7696delA	3DL1*00401_var11	Confirmed	N/A
3DL1*00401	3DL1*0040102	6055T>C;6056delC;7696delA;12548G>C	3DL1*00401_var13	Confirmed	N/A
3DL1*00401	3DL1*0040102	12191G>A	3DL1*00401_unconfirmed_var6	Unconfirmed	N/A
3DL1*00401	3DL1*0040102	-328delG;12548G>C	3DL1*00401_unconfirmed_var15	Unconfirmed	N/A
3DL1*00401	3DL1*0040102	3046C>T;6055T>C;6056delC	3DL1*00401_unconfirmed_var16	Unconfirmed	N/A
3DL1*00401	3DL1*0040102	5664delT;6055T>C;6056delC;7696delA	3DL1*00401_unconfirmed_var26	Unconfirmed	N/A
3DL1*00401	3DL1*0040102	-335delG;5664delT;6055T>C;6056delC;9538C>T	3DL1*00401_unconfirmed_var27	Unconfirmed	N/A
3DL1*00401	3DL1*0040102	6055T>C;6056delC;6954T>G;6955delG;7696delA;11226delT	3DL1*00401_unconfirmed_var28	Unconfirmed	N/A
3DL1*00401	3DL1*0040102	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG;8443G>A	3DL1*00401_unconfirmed_var30	Unconfirmed	N/A
3DL1*00401	3DL1*0040102	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG	3DL1*00401_unconfirmed_var31	Unconfirmed	N/A
3DL1*00401	3DL1*0040102	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG;12548G>C	3DL1*00401_unconfirmed_var33	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
3DL1*00401	3DL1*0040102	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG;7696delA;12548G>C	3DL1*00401_unconfirmed_var38	Unconfirmed	N/A
3DL1*00401	3DL1*0040102	-339delA;-339delG;-339delT;-339delG;-339delG;-339delA;-339delG;-339delA;-339delT;-339delC;-339delT;-339delG;-339delG;-339delG;-339delC;-339delC;-339delT;-339delG;-339delG;7696delA	3DL1*00401_unconfirmed_var39	Unconfirmed	N/A
3DL1*00401_c.605delCCT	3DL1*0040101	2973delCCT	3DL1*00401_c.605delCCT_var5	Confirmed	N/A
3DL1*00402	None available		3DL1*00402_var1	Confirmed	3DL1*00402 (LR593984)
3DL1*00402	None available		3DL1*00402_var2	Confirmed	N/A
3DL1*00402	None available		3DL1*00402_unconfirmed_var1	Unconfirmed	N/A
3DL1*00402	None available		3DL1*00402_unconfirmed_var2	Unconfirmed	N/A
3DL1*00402	None available		3DL1*00402_unconfirmed_var3	Unconfirmed	N/A
3DL1*00402	None available		3DL1*00402_unconfirmed_var4	Unconfirmed	N/A
3DL1*00402	None available		3DL1*00402_unconfirmed_var5	Unconfirmed	N/A
3DL1*00402	None available		3DL1*00402_unconfirmed_var6	Unconfirmed	N/A
3DL1*00402	None available		3DL1*00402_unconfirmed_var7	Unconfirmed	N/A
3DL1*00501	3DL1*0050101		3DL1*0050101	Confirmed	N/A
3DL1*00501	3DL1*0050101	2561G>A;2569C>G;5305C>A	3DL1*00501_var3	Confirmed	3DL1*0050111 (LR593978)
3DL1*00501	3DL1*0050101	3065A>T	3DL1*00501_var4	Confirmed	3DL1*0050106 (LR593900)
3DL1*00501	3DL1*0050101	2610T>C;2611G>A	3DL1*00501_var5	Confirmed	3DL1*0050110 (LR593926)
3DL1*00501	3DL1*0050101	5664delT	3DL1*00501_var7	Confirmed	N/A
3DL1*00501	3DL1*0050101	6043delT	3DL1*00501_var8	Confirmed	N/A
3DL1*00501	3DL1*0050101	6954T>G;6955delG	3DL1*00501_var10	Confirmed	N/A
3DL1*00501	3DL1*0050101	5664delT;6043delT	3DL1*00501_var11	Confirmed	N/A
3DL1*00501	3DL1*0050101	-339insAGTGGAGATCTGGGCCTGG	3DL1*00501_unconfirmed_var2	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	2240A>G;2243insAC;2258insCCAGG;2258A>G;2267A>G;2268insGGAC;2290A>G;2291T>C;2292insCAATA;2367insT;2373insCCC;11226delT	3DL1*00501_unconfirmed_var3	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	2258insCCCAGG;2258A>G;2297G>A;2298A>G;2299T>A;2300G>T;2303G>T;2305insAAGGTG;2318insA;2341insG	3DL1*00501_unconfirmed_var4	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	11730T>G	3DL1*00501_unconfirmed_var6	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	10427C>A	3DL1*00501_unconfirmed_var8	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	5305C>A;9430A>T	3DL1*00501_unconfirmed_var9	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	-176insG;211delG	3DL1*00501_unconfirmed_var11	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	6783G>A	3DL1*00501_unconfirmed_var13	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	6043delT;6059T>C	3DL1*00501_unconfirmed_var19	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	2561G>A;2569C>G;5305C>A;6043delT	3DL1*00501_unconfirmed_var20	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	1319delC	3DL1*00501_unconfirmed_var21	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	687G>T;5664delT	3DL1*00501_unconfirmed_var22	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	4354delA;5305C>A;6043delT	3DL1*00501_unconfirmed_var27	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	6043delT;7697delA	3DL1*00501_unconfirmed_var28	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	6043delT;10877delA;11226delT	3DL1*00501_unconfirmed_var30	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	6057delT;7755delT;10877delA	3DL1*00501_unconfirmed_var31	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	5664delT;6043delT;6221delC	3DL1*00501_unconfirmed_var33	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	6043delT;7697delA;11226delT	3DL1*00501_unconfirmed_var34	Unconfirmed	N/A
3DL1*00501	3DL1*0050101	-339delAGTGGAGATCTGGGCCTGG;2091G>A	3DL1*00501_unconfirmed_var35	Unconfirmed	N/A
3DL1*00501	3DL1*0050103		3DL1*0050103	Confirmed	N/A
3DL1*00501	3DL1*0050103	2610T>C;2611G>A;4051A>T	3DL1*00501_var1	Confirmed	3DL1*0050108 (LR593961)
3DL1*00501	3DL1*0050103	1130A>C	3DL1*00501_var2	Confirmed	3DL1*0050109 (LR593899)
3DL1*00501	3DL1*0050103	9869C>A	3DL1*00501_var6	Confirmed	3DL1*0050107 (LR593929)
3DL1*00501	3DL1*0050103	2610T>C;2611G>A;4051A>T;6043delT	3DL1*00501_var9	Confirmed	N/A
3DL1*00501	3DL1*0050103	-554G>A;-422A>C;-339insAGTGGAGATCTGGGCCTGG;186C>G;187T>A;190C>T;191C>G;201A>T;205G>T;211G>A;1881C>G;2610T>C;2611G>A;3181G>C;3201T>G;4234insT;4235T>A;5139C>G;5304insA;6357insT	3DL1*00501_unconfirmed_var1	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	-298insG	3DL1*00501_unconfirmed_var5	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	1130A>C;12859C>G	3DL1*00501_unconfirmed_var7	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
3DL1*00501	3DL1*0050103	2027A>G;3746G>A;4235T>A;4385T>G;5139C>G;6948T>G;7277A>T;7752C>T;7914T>C;8030G>A;8537T>C;8590A>G;8674G>A;8678C>T;8714T>G;8726C>T;8740insT;8767delT;8767delG;8767delC;8807T>G;8811A>C;8871G>A;9110C>G;9248G>A;9307T>C;9443C>A;9444G>C;9533A>G;9764insC;9765insA;9781T>C;9791C>G;9796G>C;9814T>G;9831A>C;9834C>G;9850A>G;9950C>A;9970A>G;10049C>A;10065C>G;10135C>T;10422T>A;10556A>G;10868C>G;10939G>A;11013C>T;11015A>C;11114C>G;11710G>A;11814A>G;12241G>A;12489C>G;13016A>G;13309G>A	3DL1*00501_unconfirmed_var10	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	5127T>C	3DL1*00501_unconfirmed_var12	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	6043delT	3DL1*00501_unconfirmed_var15	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	6954T>G;6955delG	3DL1*00501_unconfirmed_var16	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	4928T>C;6043delT	3DL1*00501_unconfirmed_var17	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	6043delT;9869C>A	3DL1*00501_unconfirmed_var18	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	1130A>C;6043delT	3DL1*00501_unconfirmed_var24	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	3181G>C;3201T>G;4157C>T;4235T>A;5664delT;6043delT;10029G>A	3DL1*00501_unconfirmed_var25	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	7681T>C;7696delC;7700A>G;7755delT	3DL1*00501_unconfirmed_var26	Unconfirmed	N/A
3DL1*00501	3DL1*0050103	482delG;1130A>C;10626delA	3DL1*00501_unconfirmed_var29	Unconfirmed	N/A
3DL1*00701	3DL1*0070101		3DL1*0070101	Confirmed	N/A
3DL1*00701	3DL1*0070101	6299C>T	3DL1*00701_var1	Confirmed	3DL1*0070106 (LR593944)
3DL1*00701	3DL1*0070101	11909C>T	3DL1*00701_var2	Confirmed	3DL1*0070105 (LR593928)
3DL1*00701	3DL1*0070101	6043delT;6299C>T	3DL1*00701_var4	Confirmed	N/A
3DL1*00701	3DL1*0070101	-193A>G;-192insA;6043delT	3DL1*00701_unconfirmed_var1	Unconfirmed	N/A
3DL1*00701	3DL1*0070101	1625G>A;6043delT	3DL1*00701_unconfirmed_var4	Unconfirmed	N/A
3DL1*00701	3DL1*0070101	6043delT;12559T>C	3DL1*00701_unconfirmed_var5	Unconfirmed	N/A
3DL1*00701	3DL1*0070101	9494delT	3DL1*00701_unconfirmed_var6	Unconfirmed	N/A
3DL1*00701	3DL1*0070101	6043delT;7128delT;11909C>T	3DL1*00701_unconfirmed_var7	Unconfirmed	N/A
3DL1*00701	3DL1*0070101	3639delA;5664delT;6055delTC;6057T>C;10877delA	3DL1*00701_unconfirmed_var8	Unconfirmed	N/A
3DL1*00701	3DL1*0070104	12928G>A	3DL1*00701_var3	Confirmed	3DL1*0070104 (LR593938)
3DL1*008	3DL1*008		3DL1*008	Confirmed	N/A
3DL1*008	3DL1*008	4108T>C	3DL1*008_var1	Confirmed	3DL1*0080102 (LR593905)
3DL1*008	3DL1*008	6043delT	3DL1*008_var2	Confirmed	N/A
3DL1*008	3DL1*008	5664delT	3DL1*008_var3	Confirmed	N/A
3DL1*008	3DL1*008	2240A>G;2243insAC;2255C>A;2258insCCCAGGGG;2258A>G;2277A>G;2280insAC;2281C>A;2290A>G;2291T>G;2293insTAG;2300insGATTCCG;2301T>G;2308insG;2337insGGC;2343insGGC;2371insT;2381insA;6043delT	3DL1*008_unconfirmed_var1	Unconfirmed	N/A
3DL1*008	3DL1*008	-301A>G;-300insAGT;-299T>G	3DL1*008_unconfirmed_var2	Unconfirmed	N/A
3DL1*008	3DL1*008	6043insT	3DL1*008_unconfirmed_var3	Unconfirmed	N/A
3DL1*008	3DL1*008	10515G>A	3DL1*008_unconfirmed_var4	Unconfirmed	N/A
3DL1*008	3DL1*008	9305T>C	3DL1*008_unconfirmed_var5	Unconfirmed	N/A
3DL1*008	3DL1*008	6G>A;6043delT	3DL1*008_unconfirmed_var7	Unconfirmed	N/A
3DL1*008	3DL1*008	4108T>C;6043delT	3DL1*008_unconfirmed_var8	Unconfirmed	N/A
3DL1*008	3DL1*008	6043delT;11814G>A;12241A>G;12723C>G	3DL1*008_unconfirmed_var9	Unconfirmed	N/A
3DL1*008	3DL1*008	6043delT;11226delT	3DL1*008_unconfirmed_var10	Unconfirmed	N/A
3DL1*008	3DL1*008	4354delA;6043delT	3DL1*008_unconfirmed_var11	Unconfirmed	N/A
3DL1*008	3DL1*008	6043delT;11226delT;12856delC	3DL1*008_unconfirmed_var12	Unconfirmed	N/A
3DL1*008	3DL1*008	5664delT;6043delT;11226delT	3DL1*008_unconfirmed_var13	Unconfirmed	N/A
3DL1*008_c.1316A>T	3DL1*008	13267A>T	3DL1*008_c.1316A>T_var6	Confirmed	N/A
3DL1*009	None available		3DL1*009_var1	Confirmed	3DL1*0090102 (LR593982)
3DL1*009	None available		3DL1*009_var2	Confirmed	3DL1*0090104 (LR593991)
3DL1*009	None available		3DL1*009_var3	Confirmed	3DL1*0090103 (LR593951)
3DL1*009	None available		3DL1*009_var4	Confirmed	3DL1*0090101 (LR593977)
3DL1*009	None available		3DL1*009_unconfirmed_var1	Unconfirmed	N/A
3DL1*009	None available		3DL1*009_unconfirmed_var2	Unconfirmed	N/A
3DL1*009	None available		3DL1*009_unconfirmed_var3	Unconfirmed	N/A
3DL1*009	None available		3DL1*009_unconfirmed_var4	Unconfirmed	N/A
3DL1*009	None available		3DL1*009_unconfirmed_var5	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
3DL1*009	None available		3DL1*009_unconfirmed_var6	Unconfirmed	N/A
3DL1*009	None available		3DL1*009_unconfirmed_var7	Unconfirmed	N/A
3DL1*009	None available		3DL1*009_unconfirmed_var8	Unconfirmed	N/A
3DL1*009	None available		3DL1*009_unconfirmed_var10	Unconfirmed	N/A
3DL1*009	None available		3DL1*009_unconfirmed_var11	Unconfirmed	N/A
3DL1*01501	3DL1*0150101		3DL1*0150101	Confirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G;7542A>G	3DL1*01502_var2	Confirmed	3DL1*0150214 (LR593896)
3DL1*01502	3DL1*0150201	3623A>G;6148T>C;7542A>G	3DL1*01502_var3	Confirmed	3DL1*0150215 (LR593919)
3DL1*01502	3DL1*0150201	3623A>G;6148T>C;7542A>G;10581G>A	3DL1*01502_var4	Confirmed	3DL1*0150216 (LR593959)
3DL1*01502	3DL1*0150201	3623A>G;6043delT;7542A>G	3DL1*01502_var7	Confirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G;5664delT;6043delT;7542A>G	3DL1*01502_var9	Confirmed	N/A
3DL1*01502	3DL1*0150201	-298insG;3623A>G;7542A>G	3DL1*01502_unconfirmed_var2	Unconfirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G;7542A>G;8095T>A	3DL1*01502_unconfirmed_var5	Unconfirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G;6548T>G;7542A>G;11013T>C;11015C>A	3DL1*01502_unconfirmed_var6	Unconfirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G	3DL1*01502_unconfirmed_var7	Unconfirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G;6043delT;6148T>C;7542A>G;10581G>A	3DL1*01502_unconfirmed_var8	Unconfirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G;6043delT;7542A>G;8095T>A	3DL1*01502_unconfirmed_var10	Unconfirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G;6221delT;7542A>G	3DL1*01502_unconfirmed_var11	Unconfirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G;5664delT;6043delT;6148T>C;7542A>G	3DL1*01502_unconfirmed_var15	Unconfirmed	N/A
3DL1*01502	3DL1*0150201	3623A>G;6043delT;7542A>G;11226delT	3DL1*01502_unconfirmed_var17	Unconfirmed	N/A
3DL1*01502	3DL1*0150203		3DL1*0150203	Confirmed	N/A
3DL1*01502	3DL1*0150204		3DL1*0150204	Confirmed	N/A
3DL1*01502	3DL1*0150204	9139G>A	3DL1*01502_var5	Confirmed	3DL1*0150218 (LR593979)
3DL1*01502	3DL1*0150204	6043delT	3DL1*01502_var6	Confirmed	N/A
3DL1*01502	3DL1*0150204	5664delT;6043delT	3DL1*01502_var8	Confirmed	N/A
3DL1*01502	3DL1*0150204	-275insA	3DL1*01502_unconfirmed_var1	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	-336insG	3DL1*01502_unconfirmed_var3	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	11622C>T;11623C>G	3DL1*01502_unconfirmed_var4	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	6043delT;12655G>A	3DL1*01502_unconfirmed_var9	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	469C>T;6043delT	3DL1*01502_unconfirmed_var12	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	7697delA	3DL1*01502_unconfirmed_var13	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	6043delT;10877delA	3DL1*01502_unconfirmed_var14	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	6043delT;13466delC	3DL1*01502_unconfirmed_var16	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	9496delT;11952delG	3DL1*01502_unconfirmed_var18	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	5664delT;6055delTC;6057T>C	3DL1*01502_unconfirmed_var20	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	6043delT;7697delA;10877delA;11226delT	3DL1*01502_unconfirmed_var21	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	4150delAGAC;9139G>A	3DL1*01502_unconfirmed_var22	Unconfirmed	N/A
3DL1*01502	3DL1*0150204	-218insA;407delG;3660delA;5664delT;6055delTC;6057T>C;10626delA	3DL1*01502_unconfirmed_var23	Unconfirmed	N/A
3DL1*01701	3DL1*01701	6548T>G	3DL1*01701_unconfirmed_var1	Unconfirmed	N/A
3DL1*01701	3DL1*01701	6043delT;6548T>G	3DL1*01701_unconfirmed_var2	Unconfirmed	N/A
3DL1*019	None available		3DL1*019_var1	Confirmed	3DL1*019 (LR593962)
3DL1*019	None available		3DL1*019_var2	Confirmed	N/A
3DL1*019	None available		3DL1*019_unconfirmed_var1	Unconfirmed	N/A
3DL1*019	None available		3DL1*019_unconfirmed_var2	Unconfirmed	N/A
3DL1*02001	3DL1*0200102	6289C>G	3DL1*02001_var1	Confirmed	3DL1*0200103 (LR593894)
3DL1*02001	3DL1*0200102	6043delT;6289C>G	3DL1*02001_var2	Confirmed	N/A
3DL1*02001	3DL1*0200102	6043insT;6289C>G	3DL1*02001_unconfirmed_var1	Unconfirmed	N/A
3DL1*02001	3DL1*0200102	6289C>G;8566C>T	3DL1*02001_unconfirmed_var2	Unconfirmed	N/A
3DL1*02001	3DL1*0200102	4265A>C;6289C>G	3DL1*02001_unconfirmed_var3	Unconfirmed	N/A
3DL1*02001	3DL1*0200102	6289C>G;12488delG	3DL1*02001_unconfirmed_var4	Unconfirmed	N/A
3DL1*02001	3DL1*0200102	4265A>C;5664delT;6043delT;6289C>G	3DL1*02001_unconfirmed_var5	Unconfirmed	N/A
3DL1*02001	3DL1*0200102	4265A>C;5664delT;6043delT;6221delT;6289C>G;8359delT;11226delT	3DL1*02001_unconfirmed_var6	Unconfirmed	N/A
3DL1*03101	3DL1*0310101		3DL1*0310101	Confirmed	N/A
3DL1*033	None available		3DL1*033_unconfirmed_var1	Unconfirmed	N/A
3DL1*053	None available		3DL1*053_var1	Confirmed	3DL1*053 (LR593903)
3DL1*053	None available		3DL1*053_unconfirmed_var1	Unconfirmed	N/A
3DL1*053	None available		3DL1*053_unconfirmed_var2	Unconfirmed	N/A
3DL1*099	None available		3DL1*099_unconfirmed_var1	Unconfirmed	N/A
3DL1*099	None available		3DL1*099_unconfirmed_var2	Unconfirmed	N/A
3DL1*099	None available		3DL1*099_unconfirmed_var3	Unconfirmed	N/A
3DL1*109	None available		3DL1*109_unconfirmed_var1	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
3DS1*01301	3DS1*0130101		3DS1*0130101	Confirmed	3DS1*0130101 (LR593910)
3DS1*01301	3DS1*0130101	4762G>A	3DS1*01301_var1	Confirmed	N/A
3DS1*01301	3DS1*0130101	5405G>A	3DS1*01301_var2	Confirmed	N/A
3DS1*01301	3DS1*0130101	9684A>G	3DS1*01301_var4	Confirmed	N/A
3DS1*01301	3DS1*0130101	1698G>T	3DS1*01301_var5	Confirmed	3DS1*0130101 (LR593925)
3DS1*01301	3DS1*0130101	1174T>A;13736C>T	3DS1*01301_var6	Confirmed	N/A
3DS1*01301	3DS1*0130101	8110A>G	3DS1*01301_var7	Confirmed	N/A
3DS1*01301	3DS1*0130101	7940A>C	3DS1*01301_var8	Confirmed	N/A
3DS1*01301	3DS1*0130101	7010T>C	3DS1*01301_var9	Confirmed	3DS1*0130101 (LR594000)
3DS1*01301	3DS1*0130101	6445delT	3DS1*01301_var10	Confirmed	N/A
3DS1*01301	3DS1*0130101	6445delT;8110A>G	3DS1*01301_var11	Confirmed	N/A
3DS1*01301	3DS1*0130101	6445delT;7010T>C	3DS1*01301_var12	Confirmed	N/A
3DS1*01301	3DS1*0130101	8756G>A	3DS1*01301_unconfirmed_var1	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2056C>G	3DS1*01301_unconfirmed_var2	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	12189A>G;12191C>G;;13736C>T	3DS1*01301_unconfirmed_var3	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	5500G>A	3DS1*01301_unconfirmed_var4	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	13155A>G	3DS1*01301_unconfirmed_var5	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2457delA	3DS1*01301_unconfirmed_var6	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	4931G>A;6445delT	3DS1*01301_unconfirmed_var7	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	1209G>T;6066delT	3DS1*01301_unconfirmed_var8	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	1174T>A;6445delT;;13736C>T	3DS1*01301_unconfirmed_var9	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	4097delA	3DS1*01301_unconfirmed_var10	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6445delT;8835C>G	3DS1*01301_unconfirmed_var11	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	4111G>C;6445delT	3DS1*01301_unconfirmed_var13	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6066delT	3DS1*01301_unconfirmed_var14	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	8099delA	3DS1*01301_unconfirmed_var15	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	1209G>T;6445delT;10270G>A	3DS1*01301_unconfirmed_var16	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	12189A>G;12191C>G;;13592delC;13736C>T	3DS1*01301_unconfirmed_var18	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	12522delG	3DS1*01301_unconfirmed_var19	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6445delT;11206T>C	3DS1*01301_unconfirmed_var20	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6445delT;10045C>G	3DS1*01301_unconfirmed_var21	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2385delG	3DS1*01301_unconfirmed_var22	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6066delT;6445delT	3DS1*01301_unconfirmed_var23	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6445delT;8099delA	3DS1*01301_unconfirmed_var24	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	4762G>A;6066delT;6445delT	3DS1*01301_unconfirmed_var25	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6445delT;6624delT	3DS1*01301_unconfirmed_var26	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6445delT;7010T>C;12522delG	3DS1*01301_unconfirmed_var27	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2488delA;6445delT	3DS1*01301_unconfirmed_var28	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2441delA;6445delT	3DS1*01301_unconfirmed_var29	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2457delA;6445delT;12189A>G;12191C>G;;13736C>T	3DS1*01301_unconfirmed_var30	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6445delT;6521T>C;13868delA	3DS1*01301_unconfirmed_var31	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2457delA;6445delT;8110A>G	3DS1*01301_unconfirmed_var32	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2424delAC;6445delT;8110A>G	3DS1*01301_unconfirmed_var33	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	6066delT;6445delT;11069G>A;12522delG	3DS1*01301_unconfirmed_var34	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	5665delA;6445delT;8099delA;8110A>G;11276delA	3DS1*01301_unconfirmed_var36	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	1174T>A;4404delGAT;6445delT;;13736C>T	3DS1*01301_unconfirmed_var37	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2621delG;2659delC;4647delA;6066delT	3DS1*01301_unconfirmed_var38	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2385delG;2457delA;4647delA;;13850delC	3DS1*01301_unconfirmed_var39	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	-629delG;-482delG;-14insC;-14delG;-14delG;3895delA;4048G>A;9788delG	3DS1*01301_unconfirmed_var40	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2353delG;2457delA;2573delG;2803delC;6445delT	3DS1*01301_unconfirmed_var41	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	266T>C;2457delA;2493delA;2573delG;2679delG;4609delA;6445delT	3DS1*01301_unconfirmed_var42	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2372delA;2441delA;2573delG;2601delG;2681delGT	3DS1*01301_unconfirmed_var43	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2441delA;2457delA;2573delG;2741delT;2748delA;4609delA;6445delT	3DS1*01301_unconfirmed_var44	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2441delA;2457delA;2484delA;2493delA;2573delG;2601delG;2621delG;2659delC;2748delA;4609delA;6445delT	3DS1*01301_unconfirmed_var45	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2599delCA;2601G>A;2608delC;2614delG;2621delG;2635delG;2665delC;2673delG;2679delG;2701delGTC;2704T>C	3DS1*01301_unconfirmed_var46	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2385delG;2441delA;2461delAT;2463A>T;2601delG;2665delC;2681delGT;4404delGAT;6445delT	3DS1*01301_unconfirmed_var47	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2455T>A;2456C>A;2459A>T;2465delATTAAT;2478delT;2486delAG;2488A>G;2529delG;2573delG;2586C>T;2587delTG;6445delT	3DS1*01301_unconfirmed_var48	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
3DS1*01301	3DS1*0130101	2441delA;2460delAAT;2497delA;2507delT;2516C>T;2517delTGA;2529delG;2573delG;2601delG;3753delT;12522delG	3DS1*01301_unconfirmed_var49	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2460delAATAAAT;2471T>A;2473A>T;2474T>A;2475C>T;2476A>C;2490delACC;2493A>C;2498delAG;2506delGT;6445delT	3DS1*01301_unconfirmed_var50	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2353delG;2385delG;2398delT;2438delGT;2440G>T;2441A>G;2461delAT;2463A>T;2486A>G;2487G>A;2489delAAC;2493A>C;2514delC;2545delT;2601delG;2661C>A;2663delAT;2665C>T;2741delT;2748delA;4609delA	3DS1*01301_unconfirmed_var51	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	-752A>G;205T>G;722delC;1129insA;2302delAC;2305A>C;2353delG;2398delT;2441delA;2461delAT;2463A>T;2486A>G;2487G>A;2489delAAC;2493A>C;2503delA;2573delG;2601delG;2651delGA;2665delC;2748delA;2812delA;4609delA;6066delT;6445delT;6624delT;7010T>C	3DS1*01301_unconfirmed_var52	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2408delCCAAGA;2420delCACT;2428T>C;2429C>T;2430delCAGCCTG;2441delAA;2457delA;2465delAT;;13868delA	3DS1*01301_unconfirmed_var53	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2262delA;2300delA;2321delGT;2333delGC;2353delG;2385delG;2398delT;2441delA;2461delAT;2463A>T;2490delAC;2493A>C;2529delG;2573delG;2601delG;2608delC;2621delG;2661C>A;2663delAT;2665C>T;2681delGT;2721delC;6445delT	3DS1*01301_unconfirmed_var54	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2390delG;2397delGT;2415delA;2461A>T;2462T>A;2464A>T;2466T>A;2472delAATCAATTAATTAAAGAAACC;2499G>A;6445delT	3DS1*01301_unconfirmed_var55	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2457delA;2465delAT;2468A>T;2480delA;2661C>A;2663delATC;2666C>T;2674G>A;2676delACAGG;2687G>C;2688G>A;2689delTCACTGATGT;2711delTGAG;2715T>G	3DS1*01301_unconfirmed_var56	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	1781delC;1781delA;1781delG;1781delT;1781delG;1781delA;1781delC;1781delT;1781C>G;1781T>C;1781C>T;1781A>C;1781A>T;1781delT;1781delT;1781delA;1781delA;1781delT;1781delT;1781delA;1781delA;1781delA;1781delT;1781delC;1781delA;1781delA;1781delT;1781delT;1781delA;1781delA;1781delA;1781delT;1781delT;1781delA;1781delA;1781delA;1781delG;1781delA;1783C>T;1784C>A;5756delT	3DS1*01301_unconfirmed_var57	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2460A>T;2464A>T;2467delTAATTAATCAATTAATTAAAGAAACC;2500G>A;2501A>G;2502delGAAGGT;2509G>T	3DS1*01301_unconfirmed_var58	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	1174T>A;2400T>C;2401C>A;2402A>G;2404delTGAGCCAAGAC;2416A>G;2419C>A;2423T>A;2424A>C;2425C>T;2444A>G;2445delAGTGAC;2453T>A;2456C>T;2457A>C;2465delATTAAT;2481delATTAAAG;2489A>G;2491C>A;2518delGA;2531delGC;2533A>C;2534G>A;2546delTA;2548C>A;2554delG;;13736C>T	3DS1*01301_unconfirmed_var59	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2443A>G;2445delAGTGACTGT;2460A>T;2464A>T;2467delTAATTAATCAATTAATTAAAGAAACCAA;2496C>T;2500G>A;2502G>C;2506G>A;2507T>G;2509delG	3DS1*01301_unconfirmed_var60	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2415delA;2445delAGTGACT;2454C>G;2460A>T;2465A>T;2466T>A;2468delAATTAATCAATTAATTAAGAAACC;2500delGAGA;2506G>A;2507T>G;2509G>T;2548delC;2573delG;6445delT	3DS1*01301_unconfirmed_var61	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	2410delA;2436delG;2444delAAGT;2455delTCAAAAAT;2466T>A;2478delT;2486delAG;2488A>G;2495delAC;2497A>C;2499G>A;2507delT;2516C>T;2517delTGA;2528A>G;2529G>A;2530delGGCAGGAT;2542delATGT;2551C>G;2553delAGGCT;2558C>T;2586C>T;2587delTG;2599delCA;2601G>A;6445delT	3DS1*01301_unconfirmed_var62	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	-176insG;2441delA;2457delA;2488delA;2499delG;2512T>C;2513delACCCTGAGATCAGCA;2528A>C;2529G>A;2539C>T;2541G>A;2542delATGTTAC;2559delCAT;2583delCT;2595G>C;2598delCCAG;2633delA;2651delGA;2661C>A;2662delAAT;2665C>T;2674delGA;2681delGT;2748delA	3DS1*01301_unconfirmed_var63	Unconfirmed	N/A

Allele CDS Description	Most homologous full length allele	gDNA discrepancy description	Sequence ID	Confirmed?	New allele name (Accession number)
3DS1*01301	3DS1*0130101	2461delAT;2463A>T;2484delA;2554delG;2562delC;2568delAGGA;2574G>A;2577T>G;2586C>T;2587T>G;2588G>A;2589G>C;2591A>G;2595insG;2600A>T;2601G>A;2602delGGGCCACCCTATGGAAGC;2622G>A;2625C>G;2634delAG;2640delA;2646delGGCA;2651G>C;2661C>A;2662delAAT;2665C>T;2673delG;2681delGT;2697delGTCTG;2715delT;2721delC;6445delT;8110A>G	3DS1*01301_unconfirmed_var65	Unconfirmed	N/A
3DS1*01301	3DS1*0130101	1780delA;1780delA;1780delC;1780delA;1780delA;1780delG;1780delG;1780delC;1780delA;1780delG;1780delA;1780delG;1780delC;1780delC;1780delA;1780C>G;1780A>C;1780C>A;1780G>T;1780delT;1780delG;1780delA;1780A>G;1780A>T;1780A>T;1780A>T;1780A>T;1780A>T;1780A>C;1780delA;1780delT;1780delT;1780delA;1780delA;1780delT;1780delC;1780delA;1780delA;1780delT;1780delT;1780delA;1780delA;1780delT;1780delT;1780delA;1780delA;1780delA;1780delG;1781A>G;1785A>C;1787delACAAGG;1893delG;3905delA;3905delG;3905delA;3905delT;3905delA;5375delT;5790delTC;5792T>C	3DS1*01301_unconfirmed_var66	Unconfirmed	N/A
3DS1*01301	3DS1*0130105	2409delCA;2416delAC;2424delAC;2434C>T;2435delTG;2444A>G;2445delAGT;2455delTCAA;2465delATTAAT;2475delCAATTA;2483T>C;2485A>T;2487G>T;2493delA;2500G>A;2501A>G;2502G>A;2504A>G;2519delAGATCAGCAAGGGCAGGATGCTGATGTTACC;2557insG;2561insG;2566insA;6250T>G;6445delT	3DS1*01301_unconfirmed_var67	Unconfirmed	N/A
3DS1*01301	3DS1*0130105	2404T>A;2406delAGCCAAGA;2420C>T;2424delACACTC;2434delCTGGG;2441A>T;2442A>G;2449delACTGT;2457delA;2464A>T;2467T>C;2470delTTAATCAATTAAT;2486A>T;2487G>A;2489A>G;2491C>A;2542delATGTTAC;2549C>T;2552C>A;2553A>C;2554G>A;2559C>A;2560delAT;2565C>T;2567T>G;2571delAAGGGGT;2584T>C;2585C>T;2586delCTG;2698delT;6250T>G;6445delT;	3DS1*01301_unconfirmed_var68	Unconfirmed	N/A
3DS1*01301_c.1144insAC;1155G>C	3DS1*0130101	6934delT;12678A>G;12680C>G;13867insAC;13877C>G;14225C>T	3DS1*01301_c.1144insAC;1155G>C_unconfirmed_var6	Unconfirmed	N/A
3DS1*049N	None available		3DS1*049N_var1	Confirmed	None available (LR593921)
3DS1*049N	None available		3DS1*049N_unconfirmed_var1	Unconfirmed	N/A
3DS1*049N	None available		3DS1*049N_unconfirmed_var2	Unconfirmed	N/A
3DS1*049N	None available		3DS1*049N_unconfirmed_var3	Unconfirmed	N/A
3DS1*049N	None available		3DS1*049N_unconfirmed_var4	Unconfirmed	N/A

Supplementary Table J. Estimated KIR2DL2/3~KIR2DL1 haplotype frequency and linkage disequilibrium

KIR2DL2/3 allele	KIR2DL1 allele	Haplotype frequency	Coefficient of LD (D)	Normalised coefficient of LD (D´)
2DL3*00101	2DL1*00302	0.365	0.205	0.920
2DL3*00201	2DL1*00201	0.225	0.159	0.912
2DL2*00301	2DL1*absent	0.115	0.098	0.912
2DL2*00101	2DL1*00401	0.105	0.088	0.866
2DL3*00501	2DL1*00101	0.039	0.037	0.978
2DL3*00101	2DL1*00201	0.033	-0.082	0.712
2DL2*00101	2DL1*absent	0.024	0.004	0.036
2DL3*00110	2DL1*00201	0.009	0.005	0.378
2DL3*00201	2DL1*008	0.008	0.005	0.761
2DL2*00301	2DL1*00401	0.008	-0.007	0.479
2DL3*00101	2DL1*032N	0.007	0.004	1.000
2DL3*00110	2DL1*00302	0.007	0.000	0.028
2DL3*003	2DL1*00302	0.006	0.003	0.640
2DL2*00101	2DL1*007	0.004	0.004	1.000
2DL3*00501	2DL1*00201	0.004	-0.008	0.652
2DL3*00101	2DL1*037	0.003	0.001	1.000
2DL3*00101	2DL1*00401	0.003	-0.047	0.948
2DL3*00201	2DL1*00401	0.003	-0.026	0.910
2DL2*00101	2DL1*00402	0.003	0.003	0.768
2DL2*00101	2DL1*004_808delA	0.002	0.001	1.000
2DL3*00101	2DL1*008	0.002	-0.002	0.564
2DL3*003	2DL1*00201	0.002	0.000	0.195
2DL3*005_709G>A	2DL1*00101	0.001	0.001	1.000
2DL3*002_171C>A	2DL1*00201	0.001	0.001	1.000
2DL3*002_618A>C	2DL1*00201	0.001	0.001	1.000
2DL3*00201	2DL1*002_110C>G	0.001	0.001	1.000
2DL3*00201	2DL1*002_169C>G	0.001	0.001	1.000
2DL3*00201	2DL1*002_568G>A	0.001	0.001	1.000
2DL3*00201	2DL1*002_950G>T	0.001	0.001	1.000
2DL3*001_142C>A	2DL1*00302	0.001	0.001	1.000
2DL3*001_274C>G	2DL1*00302	0.001	0.001	1.000
2DL3*001_549C>T	2DL1*00302	0.001	0.001	1.000
2DL3*001_598G>A	2DL1*00302	0.001	0.001	1.000
2DL3*01202	2DL1*004_13G>T	0.001	0.001	1.000
2DL2*00301	2DL1*007_1027A>G	0.001	0.001	1.000
2DL3*010	2DL1*010	0.001	0.001	1.000
2DL3*00201	2DL1*021	0.001	0.001	1.000
2DL2*00101	2DL1*024	0.001	0.001	1.000
2DL3*00201	2DL1*034	0.001	0.001	1.000
2DL3*00101	2DL1*002_22A>G	0.001	0.000	1.000
2DL3*00101	2DL1*003_842C>G	0.001	0.000	1.000
2DL3*00101	2DL1*003_853A>C	0.001	0.000	1.000
2DL3*00101	2DL1*003_910C>G	0.001	0.000	1.000
2DL3*00101	2DL1*003_963G>A	0.001	0.000	1.000
2DL3*00101	2DL1*020	0.001	0.000	1.000
2DL3*00101	2DL1*absent	0.001	-0.057	0.985
2DL3*00501	2DL1*00302	0.001	-0.016	0.950
2DL3*00110	2DL1*00401	0.001	-0.001	0.578
2DL2*00101	2DL1*010	0.001	0.001	0.419
2DL3*030	2DL1*00201	0.001	0.000	0.309
2DL3*030	2DL1*00302	0.001	0.000	0.190
2DL2*00301	2DL1*00402	0.001	0.000	0.086

Supplementary Table K. Subgroup population sizes for donor KIR2DL1 allelic polymorphism analysis on the overall, RIC and MAC cohorts

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
Donor KIR expression phenotype									
Expressed	388	NT	NT	378	NT	NT	369	NT	NT
Not expressed	11	NT	NT	11	NT	NT	10	NT	NT
Donor KIR2DL1 R^{245}									
KIR2DL1 R^{245} pos	356	119	106	349	116	105	337	111	102
KIR2DL1 R^{245} neg	31	10	10	28	7	10	31	10	10
Donor KIR2DL1 C^{245}									
KIR2DL1 C^{245} pos	113	37	34	109	33	34	110	36	33
KIR2DL1 C^{245} neg	274	92	82	268	90	81	258	85	79
Donor KIR2DL1 R/C^{245} heterotypic									
R/C^{245} heterotypic	82	27	24	81	26	24	79	26	23
R^{245} monotypic	274	92	82	268	90	81	258	85	79
C^{245} monotypic	31	10	10	28	7	10	31	10	10
Donor KIR2DL1 R^{245} CNV*									
0 copies	26	11	15	24	9	15	26	11	15
1 copy	113	52	60	110	51	58	110	52	57
2 copies	171	82	87	168	80	86	158	74	82
Donor KIR2DL1 C^{245} CNV*									
0 copies	232	107	122	226	104	119	218	99	116
1 copy	68	35	33	66	33	33	66	35	31
2 copies	10	3	7	10	3	7	10	3	7
Donor high avidity KIR2DL1									
High avidity KIR2DL1 present	204	67	57	202	66	57	195	63	57
High avidity KIR2DL1 absent	184	63	59	176	58	58	174	59	55
Donor only high avidity KIR2DL1									
Only high avidity KIR2DL1 present	86	27	28	86	27	28	85	27	28
Low/intermediate avidity KIR2DL1 present	302	103	88	292	97	87	284	95	84
Donor intermediate avidity KIR2DL1									
Intermediate avidity KIR2DL1 present	238	82	68	232	80	67	222	75	64
Intermediate avidity KIR2DL1 absent	150	48	48	146	44	48	147	47	48

Supplementary Table K (continued)

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
Donor only intermediate avidity KIR2DL1									
Only intermediate avidity KIR2DL1 present	103	36	35	98	34	34	94	32	32
Low/high avidity KIR2DL1 present	285	94	81	280	90	81	275	90	80
Donor low avidity KIR2DL1									
Low avidity KIR2DL1 present	114	38	34	110	34	34	111	37	33
Low avidity KIR2DL1 absent	274	92	82	268	90	81	258	85	79
Donor only low avidity KIR2DL1									
Only low avidity KIR2DL1 present	31	10	10	28	7	10	31	10	10
High/intermediate avidity KIR2DL1 present	357	120	106	350	117	105	338	112	102
Donor total avidity grouping*									
Low (total avidity≤3)	119	54	64	114	51	62	117	54	62
High (total avidity≥4)	191	91	98	188	89	97	177	83	92
Donor KIR2DL1 V^{-17}									
KIR2DL1 V^{-17} pos	317	108	94	307	102	93	298	100	90
KIR2DL1 V^{-17} neg	71	22	22	71	22	22	71	22	22
Donor KIR2DL1 F^{-17}									
KIR2DL1 F^{-17} pos	183	60	50	181	59	50	174	55	50
KIR2DL1 F^{-17} neg	205	70	66	197	65	65	195	67	62
Donor KIR2DL1 V/F^{-17} heterotypic									
V/F^{-17} heterotypic	112	38	28	110	37	28	103	33	28
V^{-17} monotypic	205	70	66	197	65	65	195	67	62
F^{-17} monotypic	71	22	22	71	22	22	71	22	22
Donor KIR2DL1 V^{-17} CNV*									
0 copies	63	27	36	63	27	36	63	27	36
1 copy	139	66	71	133	62	69	131	61	68
2 copies	109	53	55	107	52	54	101	50	50
Donor KIR2DL1 F^{-17} CNV*									
0 copies	155	70	83	148	66	80	146	67	77
1 copy	124	65	58	123	64	58	117	60	56
2 copies	32	11	21	32	11	21	32	11	21

* RIC and MAC cohorts include paediatric and HLA-mismatched transplants. NT= Not tested.

Supplementary Table L. Subgroup population sizes for donor KIR2DL2/3 allelic polymorphism analysis on the overall, RIC and MAC cohorts

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
Donor KIR2DL2/3 P^{16}~R^{148}									
KIR2DL2/3 P^{16}~R^{148} pos	292	138	151	285	134	148	276	130	143
KIR2DL2/3 P^{16}~R^{148} neg	30	14	16	28	12	16	29	13	16
Donor KIR2DL2/3 R^{16}~C^{148}									
KIR2DL2/3 R^{16}~C^{148} pos	145	67	77	139	63	75	141	66	74
KIR2DL2/3 R^{16}~C^{148} neg	177	85	90	174	83	89	164	77	85
Donor KIR2DL2/3 R^{16}~C^{148} CNV									
0 copies	176	84	90	173	82	89	163	76	85
1 copy	114	52	61	111	51	59	111	52	58
2 copies	30	14	16	28	12	16	29	13	16
Donor KIR2DL2/3 Q^{35}									
KIR2DL2/3 Q^{35} pos	286	135	148	279	131	145	270	127	140
KIR2DL2/3 Q^{35} neg	36	17	19	34	15	19	35	16	19
Donor KIR2DL2/3 E^{35}									
KIR2DL2/3 E^{35} pos	164	79	84	158	75	82	158	77	80
KIR2DL2/3 E^{35} neg	158	73	83	155	71	82	147	66	79
Donor KIR2DL2/3 E^{35}CNV									
0 copies	157	72	83	154	70	82	146	65	79
1 copy	127	61	65	124	60	63	122	60	61
2 copies	36	17	19	34	15	19	35	16	19
Donor KIR2DL2/3 R^{50}									
KIR2DL2/3 R^{50} pos	23	14	9	23	14	9	21	13	8
KIR2DL2/3 R^{50} neg	299	138	158	290	132	155	284	130	151
Donor KIR2DL2/3 T^{200}									
Donor KIR2DL2/3 T^{200} pos	313	148	162	304	142	159	297	140	154
Donor KIR2DL2/3 T^{200} neg	9	4	5	9	4	5	8	3	5
Donor KIR2DL2/3 I^{200}									
Donor KIR2DL2/3 I^{200} pos	86	48	38	83	45	38	82	47	35
Donor KIR2DL2/3 I^{200} neg	236	104	129	230	101	126	223	96	124

Supplementary Table L (continued)

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
Donor KIR2DL2/3 I^{200} CNV									
0 copies	235	103	129	229	100	126	222	95	124
1 copy	76	43	33	74	41	33	73	43	30
2 copies	9	4	5	9	4	5	8	3	5
Donor KIR2DL2/3 P^{208}									
Donor KIR2DL2/3 P^{208} pos	296	143	150	287	137	147	279	134	142
Donor KIR2DL2/3 P^{208} neg	26	9	17	26	9	17	26	9	17
Donor KIR2DL2/3 L^{208}									
Donor KIR2DL2/3 L^{208} pos	127	64	62	125	62	62	122	60	61
Donor KIR2DL2/3 L^{208} neg	195	88	105	188	84	102	183	83	98
Donor KIR2DL2/3 L^{208} CNV									
0 copies	26	9	17	26	9	17	26	9	17
1 copy	100	54	45	99	53	45	95	50	44
2 copies	194	87	105	187	83	102	182	82	98
Donor KIR2DL2/3 R^{297}									
Donor KIR2DL2/3 R^{297} pos	290	138	149	281	132	146	274	130	141
Donor KIR2DL2/3 R^{297} neg	32	14	18	32	14	18	31	13	18
Donor KIR2DL2/3 H^{297}									
Donor KIR2DL2/3 H^{297} pos	147	74	72	145	72	72	141	70	70
Donor KIR2DL2/3 H^{297} neg	175	78	95	168	74	92	164	73	89
Donor KIR2DL2/3 H^{297} CNV									
0 copies	174	77	95	167	73	92	163	72	89
1 copy	114	59	54	113	58	54	109	56	52
2 copies	32	14	18	32	14	18	31	13	18

Supplementary Table M. Subgroup population sizes for donor KIR3DL1/S1 allelic polymorphism analysis on the overall, RIC and MAC cohorts

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
KIR3DL1 expression									
KIR3DL1 negative phenotype	48	NT	NT	45	NT	NT	44	NT	NT
KIR3DL1 positive phenotype	284	NT	NT	281	NT	NT	272	NT	NT
Donor high expression KIR3DL1 alleles									
High expression KIR3DL1 allele positive	225	108	114	222	106	113	216	102	111
High expression KIR3DL1 allele negative	107	49	57	104	46	57	100	45	54
Donor low expression KIR3DL1 alleles									
Low expression KIR3DL1 allele positive	112	53	58	112	53	58	108	52	55
Low expression KIR3DL1 allele negative	220	104	113	214	99	112	208	95	110
Donor null expression KIR3DL1 alleles									
Null expression KIR3DL1 allele positive	101	43	58	99	41	58	95	39	56
Null expression KIR3DL1 allele negative	231	114	113	227	111	112	221	108	109
Rudimentary donor KIR3DL1 expression score									
Null (score=0)	48	23	24	45	20	24	44	20	23
Low (score≤2)	164	77	86	162	75	86	155	72	82
High (score≥3)	120	57	61	119	57	60	117	55	60
MSK model – 10/10 HLA matched									
Strongly inhibitory	63	27	34	62	26	34	61	26	33
Weakly inhibitory	70	32	38	69	32	37	67	29	38
Non-inhibitory	113	58	54	109	54	54	106	54	51
MSK model – HLA mismatched, donor ligand									
Strongly inhibitory	94	42	50	93	41	50	92	41	49
Weakly inhibitory	99	48	51	98	48	50	92	43	49
Non-inhibitory	139	67	70	135	63	70	132	63	67
MSK model – HLA mismatched, recipient ligand									
Strongly inhibitory	92	39	51	91	38	51	90	38	50
Weakly inhibitory	101	50	51	100	50	50	94	45	49
Non-inhibitory	139	68	69	135	64	69	132	64	66

Supplementary Table M (continued)

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
MSK model – 10/10 HLA matched									
Strongly inhibitory	63	27	34	62	26	34	61	26	33
Weak/non-inhibitory	183	90	92	178	86	91	173	83	89
MSK model – HLA mismatched, donor ligand									
Strongly inhibitory	94	42	50	93	41	50	92	41	49
Weak/non-inhibitory	238	115	121	233	111	120	224	106	116
MSK model – HLA mismatched, recipient ligand									
Strongly inhibitory	92	39	51	91	38	51	90	38	50
Weak/non-inhibitory	240	118	120	235	114	119	226	109	115
Donor KIR3DL1 G^{238} alleles									
KIR3DL1 G^{238} positive	246	119	125	244	118	124	235	113	120
KIR3DL1 G^{238} negative	86	38	46	82	34	46	81	34	45
Donor KIR3DL1 R^{238} alleles									
KIR3DL1 R^{238} positive	83	32	49	81	31	48	81	31	48
KIR3DL1 R^{238} negative	249	125	122	245	121	122	235	116	117
Donor KIR3DL1 I^{320} alleles									
KIR3DL1 I^{320} positive	275	133	139	272	131	138	263	126	134
KIR3DL1 I^{320} negative	57	24	32	54	21	32	53	21	31
Donor KIR3DL1 V^{320} alleles									
KIR3DL1 V^{320} positive	17	5	12	17	5	12	17	5	12
KIR3DL1 V^{320} negative	315	152	159	309	147	158	299	142	153
Donor KIR3DL1 intermediate functionality alleles									
Intermediate functionality KIR3DL1 allele positive	236	117	117	234	116	54	225	111	112
Intermediate functionality KIR3DL1 allele negative	96	40	54	92	36	116	91	36	53
Donor high avidity KIR3DL1 alleles									
High avidity KIR3DL1 allele positive	96	49	46	96	49	46	92	48	43
High avidity KIR3DL1 allele negative	236	108	125	320	103	124	224	99	122
Donor low avidity KIR3DL1 alleles									
Low avidity KIR3DL1 allele positive	235	110	122	232	108	121	226	104	119
Low avidity KIR3DL1 allele negative	97	47	49	94	44	49	90	43	46

Supplementary Table M (continued)

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
Donor high avidity KIR3DL1 allele CNV									
0 copies	236	108	125	230	103	124	224	99	122
1 copy	82	43	38	82	43	38	78	42	35
2 copies	14	6	8	14	6	8	14	6	8
Donor low avidity KIR3DL1 allele CNV									
0 copies	97	47	49	94	44	49	90	43	46
1 copy	161	77	82	159	75	82	154	73	79
2 copies	74	33	40	73	33	39	72	31	40
Donor-recipient KIR3DL1 interaction score									
Interaction score=0	97	NT	NT	93	NT	NT	90	NT	NT
Interaction score<50	29	NT	NT	29	NT	NT	26	NT	NT
Interaction score ≥50	55	NT	NT	55	NT	NT	51	NT	NT

Supplementary Table N. Normalised KIR3DL1-HLA-Bw4 interaction scores

HLA-Bw4 ligand	KIR3DL1*001	KIR3DL1*005	KIR3DL1*015
HLA-B*57:01	99.7	100	96.8
HLA-B*58:01	91.1	80.4	80.4
HLA-B*57:03	89.1	76	64.8
HLA-B*49:01	81.8	70.8	70.7
HLA-A*32:01	80.9	59.9	87.6
HLA-B*53:01	78.9	68.1	71.2
HLA-B*38:01	67.3	61.5	48.8
HLA-B*15:13	53.1	47.4	32.6
HLA-B*59:01	48.9	44.5	29.1
HLA-B*47:01	46.6	51.4	24.5
HLA-B*44:03	46.3	81.5	26.1
HLA-B*15:16	45.9	40.7	32.5
HLA-B*51:02	45.0	42.4	32.2
HLA-B*51:01	42.1	40.2	30.8
HLA-A*24:03	34.8	51.2	17.6
HLA-B*44:02	31.8	46.2	22.9
HLA-A*24:02	31.2	47.7	8.91
HLA-B*52:01	28.0	24.1	15.5
HLA-B*37:01	23.4	35.2	7.24
HLA-B*27:05	6.92	21.3	0.00
HLA-A*23:01	4.95	23.9	0.00
HLA-B*13:01	0.00	9.12	0.00

These values were obtained from ref [298].

Supplementary Table O. Subgroup population sizes for donor multi-locus allelic polymorphism analysis on the overall, RIC and MAC cohorts

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
KIR2DL3*00101~KIR2DL1*00302									
KIR2DL3*00101~KIR2DL1*00302 pos	182	88	91	176	85	88	168	81	84
KIR2DL3*00101~KIR2DL1*00302 neg	138	62	76	136	60	76	135	60	75
KIR2DL3*00201~KIR2DL1*00201									
KIR2DL3*00201~KIR2DL1*00201 pos	118	58	59	117	57	59	114	55	58
KIR2DL3*00201~KIR2DL1*00201 neg	202	92	108	195	88	105	189	86	101
KIR2DL2*00101~KIR2DL1*absent									
KIR2DL2*00101~KIR2DL1*absent pos	66	26	39	61	23	37	66	26	39
KIR2DL2*00101~KIR2DL1*absent neg	254	124	128	251	122	127	237	115	120
KIR2DL2*00301~KIR2DL1*00401									
KIR2DL2*00301~KIR2DL1*00401 pos	60	30	30	59	29	30	58	30	28
KIR2DL2*00301~KIR2DL1*00401 neg	260	120	137	253	116	134	245	111	131

Supplementary Table P. Subgroup population sizes for donor-recipient KIR allele GVH matching in the overall, RIC and MAC cohorts

Test	5 year OS			5 year DFS, 5 year relapse and 1 year NRM			aGVHD		
	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort	Entire cohort	RIC cohort	MAC cohort
Donor-recipient KIR2DL1 allele GVH matched									
Matched	97	46	49	96	45	49	93	44	47
Mismatched	255	115	137	247	110	134	241	109	129
Donor-recipient KIR2DL2/3 allele GVH matched									
Matched	52	25	27	51	24	27	50	23	27
Mismatched	177	85	90	170	80	88	167	81	84
Donor-recipient KIR3DL1/S1 allele GVH matched									
Matched	15	10	5	14	9	5	15	10	5
Mismatched	230	105	121	226	102	120	216	97	115
Total donor-recipient GVH mismatches at tested loci									
≤3	90	48	41	88	46	41	83	44	38
≥4	74	31	42	72	30	41	70	29	40

Supplementary Figure A. Primer-binding site alignments within the 5′ UTR

A

```
KIR2DL1     AAACTACAAAACTCCAGAA----------------------------TTTACAGG-TGGGGTTTTTACT
KIR2DL2     AACTTATAGAACCCCTTAAAATTTGGTAACCTGAGTCCTCTGATTTGTTATTATAGGTTATTTAGTTTGCT
KIR2DL3     AACTTATAGAACCMCTTAAAATTTGGTAACCTGAGTCCTCTGATTTGTTATTATAGGTTATTTAGTTTGCT
KIR2DL4     ACACGACCAGACCTTCAA----------------------------TTGACATATTGTG-TTTTTGCT
KIR2DL5A    AAACTACAAAACTCCAGAA----------------------------TTTACAGG-TGTGGTTTTTGCT
KIR2DL5B    AAACTACAAAACTCCAGAA----------------------------TTTACAGG-TGTGGTTTTTGCT
KIR2DS1     AATCTACAAAAATCCAGAG----------------------------TTTAAATG-TGTGGTTTTTGCT
KIR2DS2     AACTTATAGAACCCCTTAAAATTTGGTAACCTGAGTCCTCTGATTTGTTATTATAGGTTATTTAGTTTGCT
KIR2DS3     AAACTATAAAAATCCAGAA----------------------------TTTACATG-TGTGGTTTTTGCT
KIR2DS4     AAACTACAAAAATCCAGAA----------------------------TTTACATG-TGTGGTTTTTGCT
KIR2DS5     AAACTATAAAAATCCAGAA----------------------------TTTACATG-TGTGGTTTTTGCT
KIR2DP1     AAWCTACAAAAMTCCAGAR----------------------------TTTAMAKG-TGTGGTTTTTGCT
KIR3DL1     AATCTACAAAAATCCAGAA----------------------------TTTACATGTTGTGATTTTYGCT
KIR3DL2     AACTTATAGAACCCCTTAAAATTTGGTAACCTGAGTTCTCTGATTTGTTATTATAGGTTATTTAGTTTGCT
KIR3DL3     TCACCACTGCATTCCAAAC----TGG----------------------GTGACGAAGTGAGATTGCRTCT
KIR3DS1     AATCTACAAAAATCCAGAA----------------------------TTTACATGTTGTGATTTTTGCT
KIR3DP1     AAACTACAAAACTCCAGAA----------------------------TTTACAGG-TGTGGTTTTTGCT
```

B

```
KIR2DL1     CTTCCC-TGGGAATTTAAATCATTTTAACTGGTTCTGCTGTAA
KIR2DL2     TCCGCC-TCCCAGGTTCAAGCTATTCTGATGCCTCTGGTTTAG
KIR2DL3     TCCGTC-TCCCAGGTTCAAGCTATTCTGATGCCTCTGGTTTAG
KIR2DL4     TTTTCTACTGGGATTTAAATCATTTTATCTGTTTCTGGCTTAA
KIR2DL5A    CTTCCC-TGGGAATTTAAATCATTTTAGCTGGTTCTGCTGTAA
KIR2DL5B    CTTCCC-TGGGAATTTAAATCATTTTAGCTGGTTCTGCTGTAA
KIR2DS1     CTTCCC-TGGGAATTTAAATCATTTTAACTGGTTCTGCTGTAA
KIR2DS2     TCCGCC-TCCCAGGTTCAAGCTATTCTGATGCCTCTGGTTTAG
KIR2DS3     CTTCCC-TGGGAATTTAAATCATTTTAACTGGTTCTGCTGTAA
KIR2DS4     CTTCCC-TGGGAATTTAAATCATTTGAACTGGTTCTGCTGTAA
KIR2DS5     CTTYCC-TGGGARTTTAAATCATTTTAACTGGTTTTGCTGTAA
KIR2DP1     CTTCCC-TGGGAATTTAAATCATTTTAACTGGTTCTGCTGYAA
KIR3DL1     CTTCCC-TGGGAGTTTAAATCATTTGAACTGGKTCTGCTGTAA
KIR3DL2     TCCGCC-TCCCAGGTTCAAGCAATTCTCGTGCCTCAGGTTTAG
KIR3DL3     -------TGGCAGATTGTGACATACTGCCAGCTGC--------
KIR3DS1     CTTCCC-TGGGAGTTTAAATCATTTGAACTGGTTCTGCTGTAA
KIR3DP1     CTTCCC-TGGGAATTTAAATCATTTTAGCTGGTTCTGCTGTAA
```

C

```
KIR2DL1     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR2DL2     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR2DL3     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR2DL4     CACATGTTGTGGTCAATGTGTCAACTGCACGATCCGGGCC
KIR2DL5A    CCCGTGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR2DL5B    CCCATGATGTAGTCAACATGTAAGCTGCATGGGCAGGGCG
KIR2DS1     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR2DS2     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR2DS3     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGAG
KIR2DS4     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR2DS5     CCCRTGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR2DP1     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR3DL1     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR3DL2     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR3DL3     CCCATGATGTGGTCAGCATGTAAACTGCATGA-----GCC
KIR3DS1     CCCATGATGTGGTCAACATGTAAACTGCATGGGCAGGGCG
KIR3DP1     CCCATGATGTRGTCAACATGTAARCTGCATGGGCAGGGCG
```

Supplementary Figure A continued

D

```
KIR2DL1     TTTCATGTTAGCACAGATTTTAGGCATCTCGTGTTCGGGAGGTTGGATCTCAGACGTGTTTTGAGTTGGGT
KIR2DL2     TTTAATGTAAGCACAGAATTCAATCATCTCGTGTATGAGAGGTTGGATCTGAGACGTCTTTTGAGTCTGGT
KIR2DL3     TTTAATGTAAGCACAGAATTCAATCATCTCGTGTATGAGAGGTTGGATCTGAGACGTCTTTTGAGTCTGGT
KIR2DL4     CCTCATGTGAGTGCAGAATTCAATCGTCCCGTGCAGGGGTAAGTGAGTCTGAGATGTGTTTTGAGCCTGGC
KIR2DL5A    TTTCACGTTAGCACAGATTTTAGGCATCTTGTGTTCGGGAGGTTGGATCTGAGACGTGTTGTGAGTTGGT
KIR2DL5B    TTTCACGTTAGCACAGATTTTAGGCATCTTGTGTTCRGGAGGTTGGATCTGAGACGTGTTKTGAGTTGGT
KIR2DS1     TTTAATGTAAGCACAGAATTCAATCATCTCGTGTATGAGAGGTTGGATCTGAGACGTSTTTTGAGTCTGGT
KIR2DS2     TTTCACGTTAGCACAGATTTTAGGCATCTCGTGTTCAGGAGGTTGGATCTGAGACGTGTTTTGAGTTGGT
KIR2DS3     TTTCACGTTAGCACAGATTTTAGGCATCCTGTGTTCMGGAGGTTGGATCTGAGACGTGTTTTGAGTTGGT
KIR2DS4     TTTCATGTTAGCACAGATTTTAGGCATCTCATGTTTGGGAGGTTGGATCTAAGACRTGTTTTGAGTTGGT
KIR2DS5     TTTCACGTTAGCASAGATTTTAGGCATCYTGTGTTCVGGAGGTTGGATCTGAGACGTGTTKTGAGTTGGT
KIR2DP1     TTTCACGTTAGCACAGATTTTAGGCATCTCGTGTTCAGGAGGTTGGATCTGAGACGTGTTTTGAGTTGGT
KIR3DL1     TTTCATGTTAGCACAGATTTTAGGCATCTCRTGTTCGGGAGGTTGGATCTRAGACGTGTTTTGAGTTGGT
KIR3DL2     TTTAATGTAAGCACAGAATTCAATCACCTCATGTGTGAGAGGTTGGATCTGAGACRTCTTTTGAGTCTGGT
KIR3DL3     TTTAATGTAAGCCAAGCATTCAGTCATCTCCTGTATGAGAGATTGGATCTGAGACGTGTTTTGAGTTGGT
KIR3DS1     TTTCATGTTAGCACAGATTTTAGGCATCTCGTGTTCRGGAGGTTGGATCTGAGACGTGTTTTGAGTTGGT
KIR3DP1     TTTCACGTTAGCACAGATTTTAGGCATCTTGTGTTCGGGAGGTTGGATCTGAGACGTGTTGTGAGTTGGT
```

E

```
KIR2DL1     TGATCC--GCCCACCTCGGCTTCCCAACGTGCT-GGGGA
KIR2DL2     ACATGGCAGCCTTTGTCAACTA-TAAAGGGACTGTGTAT
KIR2DL3     ACATGGCAGCCTTTGTCAACTA-TAAAGGGACTGTGTAT
KIR2DL4     ATATT----CCAGTATCAAGTT-GGAGGATGTTATCAGT
KIR2DL5A    TGATCC--GCCCACCTCGGCTTCCCAACGTGCTGGGGAA
KIR2DL5B    TGATCC--GCCCACCTCGGCTTCCCAACGTGCTGGGGAA
KIR2DS1     ACATGGCAGCCTTTGTCAACTA-TAAAGGGACTGTGTAT
KIR2DS2     TGATCC--GCCCACCTCGGCTTCCCGACGTGCTGGGGGA
KIR2DS3     TGATTC--GCCCACCGCGGCTTCCCAACATGCTGGGGGA
KIR2DS4     TGATCC--GCCCACCTCGGCTTCCCRACCTGCTGGGGGA
KIR2DS5     TGATTC--GCCCACCTCGGCTTCCAAACATGCTGGGGGA
KIR2DP1     TGATCC--GCCCACCTCGGCTTCCCGACGTGCTGGGGGA
KIR3DL1     TGAT-C--ACTCACYTCGGCTT-CCAAAGTGCTGGGGAA
KIR3DL2     RCATGGCAGCCTTTGTCAACTA-TRAAGGGACTGTGTAT
KIR3DL3     TCTCAG--CACTTTGGGAGGCT-GAAGTAGACAGATCAC
KIR3DS1     TGAT-C--ACTCACCTCGGCTT-CCAAAGTGCTGGGGAA
KIR3DP1     TGATCY--GCCCACCTCGGCTTCCCAACGTGCTGGGGAA
```

F

```
KIR2DL1     GGCCAGGCTGCTCTCAAACTCCTTATCTCAGTTGATCC--GCCCACCTCGGCT
KIR2DL2     TGTTGAGAAAATATTCTATTCCACCTTAAACTACATGGCAGCCTTTGTCAACT
KIR2DL3     TGTTGAGAAAATATTCTATTCCACCTTAAACTACATGGCAGCCTTTGTCAACT
KIR2DL4     AAATGTGTAGATTTCAACCTTCAGTTATTGCAATATT----CCAGTATCAAGT
KIR2DL5A    GACCAGGCTGCTCTCAAACTCCTTATCTCAGTTGATCC--GCCCACCTCGGCT
KIR2DL5B    GGCCAAGCTGCTCTCAAACTCCTTATCTCAGTTGATCC--GCCCACCTCGGCT
KIR2DS1     TGTTGAGAAAATATTCTATTCCACCTTAAACTACATGGCAGCCTTTGTCAACT
KIR2DS2     -GCCAGGCTGCTCTCAAACTCCTTATATCAGTTGATCC--GCCCACCTCGGCT
KIR2DS3     GGCCAGGCTGCTCTCAAACTCCTCATCTCAGTTGATTC--GCCCACCGCGGCT
KIR2DS4     GGCCAGGCTGCTCTCAATCTCCTCATCTCAGTTGATCC--GCCCACCTCGGCT
KIR2DS5     GGCCAGGCTGCTCTCAAACTCCTCATCTCAGTTGATTC--GCCCACCTCGGCT
KIR2DP1     GGCCAGGCTGCTCTCAAWCYCCTYATMTCAGTTGATCC--GCCCACCTCGGCT
KIR3DL1     GGCCAGGCTGGTCTCAAACTCCTGACCTCGGTTGAT-C--ACTCACYTCGGCT
KIR3DL2     TRTTGAGAAAATATTCTATTCCACCTTAAACTRCATGGCAGCCTTTGTCAACT
KIR3DL3     TCCCGGCCAGGCATGGTGGCTCACACCTGTAATCTCAG--CACTTTGGGAGGC
KIR3DS1     GGCCAGGCTGGTCTCAAACTCCTGACCTCGGTTGAT-C--ACTCACCTCGGCT
KIR3DP1     GGCCARGCTGCTCTCAAACTCCTTATCTCAGTTGATCY--GCCCACCTCGGCT
```

Supplementary Figure A continued

G

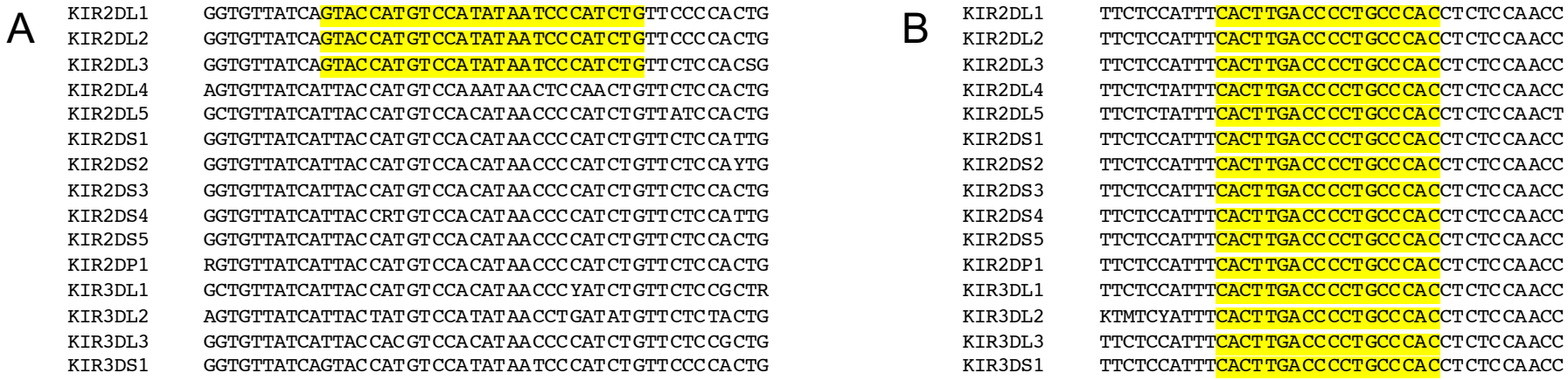
```
KIR2DL1     TTTCATGTTAGCACAGATTTTAGGCATCTCGTGTTCGGGAGGTTG
KIR2DL2     TTTAATGTAAGCACAGAATTCAATCATCTCGTGTATGAGAGGTTG
KIR2DL3     TTTAATGTAAGCACAGAATTCAATCATCTCGTGTATGAGAGGTTG
KIR2DL4     CCTCATGTGAGTGCAGAATTCAATCGTCCCGTGCAGGGGTAAGTG
KIR2DL5A    TTTCACGTTAGCACAGATTTTAGGCATCTTGTGTTCGGGAGGTTG
KIR2DL5B    TTTCACGTTAGCACAGATTTTAGGCATCTTGTGTTCRGGAGGTTG
KIR2DS1     TTTAATGTAAGCACAGAATTCAATCATCTCGTGTATGAGAGGTTG
KIR2DS2     TTTCACGTTAGCACAGATTTTAGGCATCTCGTGTTCAGGAGGTTG
KIR2DS3     TTTCACGTTAGCACAGATTTTAGGCATCCTGTGTTCMGGAGGTTG
KIR2DS4     TTTCATGTTAGCACAGATTTTAGGCATCTCATGTTTGGGAGGTTG
KIR2DS5     TTTCACGTTAGCASAGATTTTAGGCATCYTGTGTTCVGGAGGTTG
KIR2DP1     TTTCACGTTAGCACAGATTTTAGGCATCTCGTGTTCAGGAGGTTG
KIR3DL1     TTTCATGTTAGCACAGATTTTAGGCATCTCRTGTTCGGGAGGTTG
KIR3DL2     TTTAATGTAAGCACAGAATTCAATCACCTCATGTGTGAGAGGTTG
KIR3DL3     TTTAATGTAAGCCAAGCATTCAGTCATCTCCTGTATGAGAGATTG
KIR3DS1     TTTCATGTTAGCACAGATTTTAGGCATCTCGTGTTCRGGAGGTTG
KIR3DP1     TTTCACGTTAGCACAGATTTTAGGCATCTTGTGTTCGGGAGGTTG
```

H

```
KIR2DL1     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR2DL2     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR2DL3     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR2DL4     TGTTGTGGTCAATGTGTCAACTGCACGATCCGGGCCCCTCACCACA
KIR2DL5A    TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR2DL5B    TGATGTRGTCAACATGTAARCTGCATGGGCAGGGCGCCAAATAACA
KIR2DS1     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR2DS2     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR2DS3     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGAGCCAAATAACA
KIR2DS4     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR2DS5     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR2DP1     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR3DL1     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCRAATAACA
KIR3DL2     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR3DL3     TGATGTGGTCAGCATGTAAACTGCAT----GAGCCCMTCACAACA
KIR3DS1     TGATGTGGTCAACATGTAAACTGCATGGGCAGGGCGCCAAATAACA
KIR3DP1     TGATGTRGTCAACATGTAARCTGCATGGGCAGGGCGCCAAATAACA
```

Supplementary Figure A. Primer-binding site alignments within the 5′ UTR for the final PCR strategy

Small regions of the 5′ UTR consensus alignment are shown. Yellow highlighting represents the primer-binding sites for KIR2DL1 (A), KIR2DL2/3/S2 (B), KIR2DL4 (C), KIR2DL5/3DP1 (D), KIR2DS5 (E), KIR3DL1/S1 (F), KIR3DL2 (G) and KIR3DL3 (H). Blue highlighting in D represents the alternative 5′ primer for KIR2DL5*0020202, 003 and 00602 alleles. Dashes (-) represent gaps in an aligned consensus.



Supplementary Figure B. Primer-binding site alignments within the 3′ UTR for final PCR strategy

Small regions of the 3′ UTR consensus alignment, excluding KIR3DP1, are shown. Yellow highlighting represents the primer-binding sites for KIR2DL1/2/3 (A) and the generic 3′ KIR primer (B).