Characterisation of genetic complexity in chronic lymphocytic leukaemia

Adele T Timbs

Oxford Brookes University

A thesis submitted in partial fulfilment of the requirements of the award of Doctor of Philosophy

This research was carried out in collaboration with both the University of Oxford and the Oxford University Hospitals NHS Foundation Trust

January 2017

Abstract

Chronic lymphocytic leukaemia (CLL) is the most common adult leukaemia in the Western world, characterised by the accumulation of mature B-cells. The disease has a heterogeneous course whereby some patients do not require treatment for several years while others follow a more aggressive disease progression and require early therapeutic action. Although several prognostic markers have been described, it is still difficult to predict the prognosis of the disease at diagnosis or to explain the diversity associated with the disease. The goal of this thesis was therefore, to characterise the genomic complexity and molecular drivers of clonal expansion and maintenance of CLL using targeted next generation sequencing (NGS) and high resolution single nucleotide polymorphism (SNP) array.

NGS analysis of the immunoglobulin variable heavy-chain (IgHV) in 497 pre-treatment CLL patients found IgHV subclones in 18.5%, far higher than previously reported. SNP array analysis in 411 CLL patients refined the minimally overlapping regions (MOR) on del(6q) to the *ATG5* and *PRDM1* genes. Further autophagy related genes were also found to be present in other MORs. Analysis of the autophagy pathway at the protein level showed higher autophagy activity in primary CLL cells compared with age-matched controls.

NGS-IgHV enabled the existing Sanger-sequencing based prognostic system to be refined. Correlation to the clinical outcome data showed that this new classification was prognostically significant. Furthermore, the identification of IgHV subclones imply that the leukaemia initiating event in CLL takes place prior to recombination. Concurrently, the detection of altered autophagy levels suggests a role for autophagy in CLL. Collectively these data provide new insights into the pathogenesis of CLL.

For my children,

Never be afraid to try, as you cannot succeed if you do not try.

Acknowledgements

Firstly, I would like to thank my supervisors. To Anna Schuh, thank you for this opportunity. It has been a great privilege to work with you over the years and I hope I will continue to do so. My appreciation also to Susan Brooks for her expert advice and encouragement.

Thanks to all my colleagues in the Molecular Haematology Department. Special thanks to Shirley Henderson whose support and enthusiasm has been tireless throughout this project. Also to Sarah Darko and the other members of the White Cell Team, past and present, for their patience and understanding.

I wish to give special thanks to my colleague and good friend, Adam Burns. Thank you for all the help, encouragement and coffee! Thanks also to Ruth Clifford, Pauline Robbe and David Bruce, who have all become my good friends over the course of this thesis. To Basile Stamatopoulos, thank you for the opportunity to work with you. It was a very insightful and rewarding experience. Thank you to Helene Dreau and the other members of the NGS core, as well as to Dimitris Vavoulis for patiently answering all my statistics questions. Also thanks to Prof Anna Katharina Simon, Dr Kanchan Phadwal and their group for their assistance with the Amnis ImageStream.

Thanks to my parents Trudy and Neil for their unwavering support even when managing their own problems. To my in-laws, Georgina and Robert for watching the children over many Saturdays. Also, I would like to thank my friends, Kathyrn Lewis and Mary Palides, for providing much needed childcare to enable me to write in peace!

Finally, I wish to thank my husband, Michael. Your understanding, support and encouragement was the energy I needed to complete this work. I couldn't have done it without you.

Individual Contributions

This thesis is the result of my work with the additional contributions listed below.

The work described in Chapters 3 and 4 using the LymphoTrack® IGH Somatic Hypermutation was a collaborative project performed with Dr Basile Stamatopoulos. We both contributed equally to the experimental work and data analysis. The Circos plot was kindly provided by Dr Adam Burns.

The work described in Chapters 4 and 5 using SNP arrays was part of a larger project whereby Dr Ruth Clifford, Pauline Robbe, Matthieu Lm and I all performed the experimental work. Dr Ruth Clifford, Dr Samantha Knight, Dr Anna Schuh, Pauline Robbe and I all undertook the primary data analysis in Nexus. Additionally, a large proportion of the DNA used for this work was extracted by Maite Cabes.

Finally, the work described in Chapter 6 on the Amnis ImageStream used a protocol previously devised by Dr Kanchan Phadwal and Prof Anna Katharina Simon.

Publications arising from this work

Stamatopoulos B*, **Timbs A***, Bruce D, Smith T, Clifford R, Robbe P, Burns A, Vavoulis DV, Lopez L, Antoniou P, Mason J, Dreau H, Schuh A. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. Leukemia. 2016. doi: 10.1038/leu.2016.307. [Epub ahead of print] *joint co-authors

Guièze R, Robbe P, Clifford R, de Guibert S, Pereira B, **Timbs A**, Dilhuydy MS, Cabes M, Ysebaert L, Burns A, Nguyen-Khac F, Davi F, Véronèse L, Combes P, Le Garff-Tavernier M, Leblond V, Merle-Béral H, Alsolami R, Hamblin A, Mason J, Pettitt A, Hillmen P, Taylor J, Knight SJ, Tournilhac O, Schuh A. Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. Blood. 2015; 126(18):2110-7

Clifford R, Louis T, Robbe P, Ackroyd S, Burns A, **Timbs AT**, Wright Colopy G, Dreau H, Sigaux F, Judde JG, Rotger M, Telenti A, Lin YL, Pasero P, Maelfait J, Titsias M, Cohen DR, Henderson SJ, Ross MT, Bentley D, Hillmen P, Pettitt A, Rehwinkel J, Knight SJ, Taylor JC, Crow YJ, Benkirane M, Schuh A. SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. Blood. 2014;123(7):1021-31

Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, Feller SM, Grocock R, Henderson S, Khrebtukova I, Kingsbury Z, Luo S, McBride D, Murray L, Menju T, **Timbs A**, Ross M, Taylor J, Bentley D. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. Blood. 2012;120(20):4191-6

Knight SJ, Yau C, Clifford R, **Timbs AT**, Sadighi Akha E, Dréau HM, Burns A, Ciria C, Oscier DG, Pettitt AR, Dutton S, Holmes CC, Taylor J, Cazier JB, Schuh A. Quantification of subclonal distributions of recurrent genomic aberrations in paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia. Leukemia. 2012;26(7):1564-75.

Contents

ABSTRACT I

ACKNOWLEDGEMENTS V

INDIVIDUAL CONTRIBUTIONS VII

PUBLICATIONS ARISING FROM THIS WORK IX

CONTENTS XI

LIST OF FIGURES XIX

LIST OF TABLES XXI

LIST OF ABBREVIATIONS XXIII

CHAPTER 1 INTRODUCTION 1

1.1 Chronic Lymphocytic Leukaemia 1

1.2 B Lymphocytes 3

1.2.1 VDJ Recombination 4

1.2.2 Somatic Hypermutation 15

1.3 The IgH in CLL 17

1.3.1 IgHV Mutational Status 18

1.3.2 Stereotyping 21

1.4 Prognostic Markers in CLL 27

1.4.1 Abnormalities affecting the TP53 gene 27

1.4.2 Chromosomal Aberrations 29

1.4.3 Other Genetic Markers 33

1.5 Risk stratification 36

1.6 Cancer as an Evolutionary Process 39

1.6.1 Clonal Evolution and Tumour Heterogeneity in CLL 42

1.6.2 The Leukaemic Stem Cell in CLL 44

1.7 Aims of thesis 47

CHAPTER 2 MATERIALS AND METHODS 49

2.1 Cell Preparation 49

2.1.1 Ficoll Preparation of Peripheral Blood Mononuclear Cells 49

2.1.2 Thawing Frozen Cells 50

2.1.3 Isolating B-Cells 50

2.2 Extraction of Genomic DNA 51

2.3 Extraction of RNA and cDNA preparation 52

2.3.1 Preparation of cells for RNA extraction 52

2.3.2 RNA extraction using the QIAamp RNA Blood Mini Kit 53

2.3.3 cDNA preparation 54

2.4 Nucleic Acid Assessment 54

2.4.1 Nanodrop 54

2.4.2 Qubit 55

2.4.3 Bioanalyser 55

2.5 Polymerase Chain Reaction (PCR) 56

2.6 Sanger Sequencing 56

2.6.1 PCR amplification of the IgHV locus using the InvivoScribe Somatic 56

2.6.2 PCR clean-up using MicroClean 57

2.6.3 Cycle sequencing 57

2.6.4 Ethanol precipitation 58

2.6.5 Capillary electrophoresis and Data Analysis 58

2.7 Next Generation Sequencing 59

2.7.1 AMPure clean-up 59

2.7.2 Nextera 59

2.7.3 KapaQuant 60

2.7.4 MiSeq Preparation and Loading 61

2.7.5 Analysis of the IgHV locus using the LymphoTrack® IGH Somatic Hypermutation Assay Panel 61

2.7.6 Targeted Next Generation Sequencing using the Illumina TruSeq Custom Amplicon assay 62

2.8 Single Nucleotide Polymorphism Array 64

2.8.1 Illumina Genome-wide SNP Platform hybridisations 64

2.8.2 Data Analysis with Biodiscovery Nexus 65

2.9 Data Analysis 65

CHAPTER 3 USING NEXT GENERATION SEQUENCING FOR ANALYSIS OF THE IMMUNOGLOBULIN HEAVY CHAIN LOCUS 68

3.1 Introduction 68

3.2 Materials and Methods 70

3.2.1 Patient Samples 70

3.2.2 Sanger Sequencing 70

3.2.3 454 Analysis 70

3.2.4 Illumina MiSeq 72

3.3 Results 74

3.3.1 454 Analysis 74

3.3.2 MiSeq Analysis using an in-house assay 79

3.3.3 MiSeq Analysis using the Invivoscribe Lymphotrack SHM Kit 88

3.4 Discussion 91

3.4.1 Use of the two NGS Platforms 91

3.4.2 Comparison of methodologies 94

3.4.3 Subclonal Analysis 95

3.4.4 Conclusions 99

CHAPTER 4 SUBCLONAL IGHV REARRANGEMENTS ARE CLINICALLY RELEVANT IN CHRONIC LYMPHOCYTIC LEUKAEMIA 100

4.1 Introduction 100

4.2 Materials and Methods 101

4.2.1 Patient Samples 101

4.2.2 Invivoscribe LymphoTrack IGH Somatic Hypermutation Assay 102

4.2.3 Targeted NGS and SNP Array Analysis 102

4.2.4 Statistical Analysis 102

4.3 Results 103

4.3.1 Factors effecting the detection of subclones 103

4.3.2 NGS identifies 5 prognostically significant subgroups in an unselected patient cohort (Population A) 109

4.3.3 IgHV by NGS defines a very poor prognostic subgroup in a treatment-selected cohort (Population B) 115

4.3.4 NGS-IgHV is an independent prognostic marker 117

4.3.5 IgHV subclones are phylogenetically unrelated 123

4.3.6 Subclonal analysis using NGS-IgHV shows an evolving picture 125

4.4 Discussion 127

4.4.1 NGS-IgHV classification refines the SSeq-IgHV prognostic subgroups 127

4.4.2 IgHV subclones detected by NGS provide evidence for a precursor B-cell as the leukaemic stem in CLL 128

4.4.3 Stereotypy analysis suggests that the IgHV subclones are phylogenetically unrelated 130

4.4.4 Conclusions 131

CHAPTER 5 WHOLE GENOME ANALYSIS USING SINGLE NUCLEOTIDE POLYMORPHISM ARRAYS 132

5.1 Introduction 132

5.2 Materials and Methods 135

5.2.2 Genome-wide SNP Platform hybridizations and analysis 135

5.2.3 Sanger sequencing..... 135

5.2.4 Targeted NGS gene panel..... 135

5.3 Results..... 136

5.3.1 Common cytogenetic aberrations are recurrent within this cohort... 136

5.3.2 Large copy number alterations are recurrent in CLL..... 142

5.3.3 Clinical significance of recurrent large copy number alterations is uncertain..... 154

5.3.4 Many pre-treatment abnormalities persist at relapse and other recurrent CNAs emerge..... 156

5.3.5 Further investigation of the del (6q) 157

5.3.6 Genes involved in autophagy are recurrently found in MORs in CLL.. 158

5.3.7 Screening of relapse samples identifies mutations not present at diagnosis..... 161

5.4 Discussion..... 163

5.4.1 SNP arrays can detect CNAs and cnLOH in CLL..... 163

5.4.2 Genes required for leukaemogenesis are present in relapse samples 165

5.4.3 Screening of relapse samples identifies mutations not present at diagnosis..... 165

5.4.4 Genes in the autophagy pathway are recurrently mutated using SNP array analysis..... 166

5.4.5 Conclusions..... 167

CHAPTER 6 ANALYSIS OF AUTOPHAGY IN B-CLL SAMPLES REVEALS INCREASED AUTOPHAGY ACTIVITY WHEN COMPARED TO NORMAL AGE-MATCHED B-CELLS 169

6.1 Introduction 169

6.1.1 The Autophagy Pathway 170

6.1.2 Role of Autophagy in Cancer 175

6.1.3 Autophagy in CLL 177

6.2 Materials and Methods 179

6.2.1 Quantitative real-time PCR of autophagy genes 179

6.2.2 Preparation of Cells for Autophagy Analysis 180

6.2.3 Imagestream 181

6.2.4 Statistical Analysis 183

6.3 Results 184

6.3.1 Quantitative PCR 184

6.3.2 Analysis of the autophagy levels using the Amnis Imagestream 186

6.3.3 Basal autophagy levels are higher in CLL cells than age-matched controls 189

6.3.4 CLL cells have increased autophagy levels upon induction 189

6.3.5 Increased autophagy levels in CLL cells is not a feature of $CD19^{+}CD5^{+}$ cells 190

6.3.6 Higher autophagy is not linked to cell death 191

6.4 Discussion 193

6.4.1 CLL cells have a higher level of autophagy activity than B-cells from normal age-matched controls 193

6.4.2 Increased autophagy is not a feature of B1-a cells 194

6.4.3 Cells with high autophagy levels are not undergoing apoptosis..... 195

6.4.4 Conclusions..... 196

CHAPTER 7 DISCUSSION 197

7.1 Introduction..... 197

7.2 Summary of Findings..... 197

7.3 The relationship between aging and development of CLL 201

7.4 CLL as a subclonal disease 202

7.5 The role of autophagy in CLL pathogenesis..... 205

7.6 Future work and directions..... 207

7.7 Conclusion 210

REFERENCES..... 211

APPENDICES..... 258

List of Figures

Figure 1.1 Immunoglobulin structure.. 5

Figure 1.2. The IgH locus undergoes VDJ recombination 6

Figure 1.3. VDJ recombination uses deletional or inversional rearrangement.. 8

Figure 1.4. Process of V(D)J recombination. 10

Figure 1.5. Mechanisms of somatic hypermutation.. 16

Figure 3.1 The subclones detected in CLL044, CLL063 and CLL056 77

Figure 3.2 The different subclones detected in CLL003 78

Figure 3.3. The dominant clone represents an average of 90% of the reads in samples with one major clone 85

Figure 3.4. Subclonality in L76. 86

Figure 3.5. Repertoire of the VDJ rearrangement in samples of 2 healthy age-matched probands. 87

Figure 3.6. NGS can detect multiple specific VH family rearrangements. 90

Figure 4.1. NGS IgHV profiles of B cell from healthy donors. 104

Figure 4.2. There is good concordance between cDNA and gDNA 105

Figure 4.3. Polyclonal background B-cells are undetectable in CLL samples. 107

Figure 4.4. There is concordance between singleplex and multiplex PCR 108

Figure 4.5. NGS-IgHV refines SSeq classification. 110

Figure 4.6. NGS-IgHV defines 5 different prognostic subgroups. 111

Figure 4.7. NGS-IgHV defines 5 different prognostic subgroups in Binet Stage A patients.. 113

Figure 4.8. NGS-IgHV defines a poor prognostic subgroup in a treatment selected clinical trial cohort. 116

Figure 4.9. NGS-IgHV is independent from other prognostic markers 122

Figure 4.10. Circos plot visualization of V to D and V to J associations. 124

Figure 5.1. TP53 deletions are recurrent in CLL. 137

Figure 5.2. ATM deletions are recurrent in CLL.. 138

Figure 5.3.The 11q deletion in CLL073 removes the ATM gene. 139

Figure 5.4. Partial trisomy of chromosome 12 in CLL108 140

Figure 5.5. Deletions at 2q. 145

Figure 5.6. The MOR at 6q21 contains two genes; PRDM1 and ATG5. 146

Figure 5.7. Deletions of 8p are recurrent in CLL.. 149

Figure 5.8. Deletions at 9p at recurrent in CLL. 150

Figure 5.9. Two MORs can be defined at 18p. 152

Figure 5.10. Recurrent gain at 19p.. 153

Figure 6.1 The autophagy pathway. 171

Figure 6.2. ATG5 and Becn1 expression ratio in normal B-cells and CLL cells. 185

Figure 6.3. ImageStream analysis of autophagy. 187

Figure 6.4. CLL cells have a higher basal autophagy level than age-matched control B-cells. 188

Figure 6.5 Induced autophagy levels are higher in CLL cells than in age-matched controls. 190

Figure 6.6 Basal autophagy level in the B-cell subsets of the age-matched controls. 191

Figure 6.7 The increase in autophagy activity in the CLL cells is not associated with cell death..... 192

List of Tables

Table 1.1. Staging systems in CLL. 2

Table 1.2. Characteristics of the major stereotyped subsets. 24

Table 1.3 Prognostic risk stratification as devised by Rossi, Rasi *et al* (2013) 37

Table 1.4. Prognostic risk stratification as devised by the Internal CLL-IPI Working Group 38

Table 1.5. Summary of the DNA sequence motifs recognised by some of the common causes of somatic mutations. 41

Table 3.1. Comparison summary of SSeq and the dominant clone identified using NGS from the 454. 75

Table 3.2 Technical result summary of sequential samples on the 454 platform… 76

Table 3.3. Concordance between MiSeq and SSeq for 10 patients. 80

Table 3.4. Summary of Subclone Analysis from the Miseq. 82

Table 4.1. Summary of healthy age-matched controls 101

Table 4.2. Number of deaths in Binet Stage A patients. 112

Table 4.3. Predictive value of NGS-IgHV. 114

Table 4.4. Univariate Cox regression for Population A for TFS 118

Table 4.5. Multivariate Cox regression for Population A for TFS. 119

Table 4.6. Multivariate Cox regression for Population B. 120

Table 4.7. ADM304 with 100% intra-clonal CDR3-AA sequence homology.. 125

Table 4.8. NGS-IgHV analysis of diagnosis-relapse pairs. 126

Table 5.1. Summary of recurring copy number alterations in a pre-treatment CLL cohort. 143

Table 5.2. Summary of the recurrent CNA in a larger pre-treatment cohort. 155

Table 5.3. Presence or loss of copy number alteration at relapse. 157

Table 5.4. Summary of autophagy genes identified in recurrent MORs. 159

Table 5.5. Summary of differences at diagnosis and relapse exclusively of patients without noticeable difference by array 162

List of Abbreviations

AID	Activation induced deaminase
APE1	Apurinic/apyrimidinic endonuclease 1
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
ATM	Ataxia telangiectasia mutated
BCR	B-cell receptor
BDS	Bright detail similarity
BER	Base-excision repair
BIRC3	Baculoviral IAP repeat containing 3
cDNA	Complementary DNA
CDR	Complementary determining regions
CGH	Comparative genomic hybridization
CHD2	Chromodomain helicase DNA binding protein 2
CHIP	Clonal haematopoiesis of indeterminant potential
CLL	Chronic lymphocytic leukaemia
CNA	Copy number alterations
cnLOH	Copy neutral loss of heterozygosity
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DSB	Double-stranded breaks
EDTA	Ethylenediaminetetraacetic acid
EGR2	Early growth response 2
ERGIC	ER-Golgi-intermediate complex
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FBXW7	F-box and WD repeat containing protein 7
FCR	Fludarabine, cyclophosphamide and rituximab
FISH	Fluorescent in-situ hybridization
FOXO	Forkhead box protein O
FR	Framework regions
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC	Germinal centres
GTC	Guanidine thiocyanate
HOPS	Homotypic fusion and vacuolar protein sorting
HR	Hazard ratios
Ig	Immunoglobulin
IgH	immunoglobulin heavy-chain
IgHD	Immunoglobulin diverse heavy-chain
IgHJ	Immunoglobulin joining heavy-chain
IgHV	Immunoglobulin variable heavy-chain
IgL	Immunoglobulin light-chain
IgLV	Immunoglobulin variable light-chain
Igλ	Immunoglobulin lambda
Igκ	Immunoglobulin kappa
IM	Isolation membrane
KDE	Kappa deleting element
KEGG	Kyoto Encyclopedia of Genes and Genomes
LPL	Lipoprotein lipase
M	Mutated
MARS	Most abundant reads

MBL	Monoclonal B-cell lymphocytosis
MMR	Mismatch repair
MOR	Minimally overlapping region
MPDZ	Multiple PDZ Domain Protein
MTOC	Microtubule organising centre
MYD88	Myeloid differentiation primary response 88
MZB	Marginal zone B-cells
NBD	Nonmer binding domain
NFIB	Nuclear factor IB
NFκB	Nuclear factor kappa beta
NGS	Next generation sequencing
NHEJ	Nonhomologous end joining
NICD	Notch intercellular domain
NPV	Negative predictive value
NTC	No template control
OS	Overall survival
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PCH	Pericentrometric heterochromatin
PCR	Polymerase chain reaction
PFS	Progression free survival
POT1	Protection of telomeres 1
PPV	Positive predicative value
PRAME	Preferentially expressed antigen in melanoma
RAG	Recombination activating gene
RBC	Red blood cell
RILP	Rab-interacting lysosomal protein
ROMA	Representational oligonucleotide microarray analysis
ROS	Reactive oxygen species
RS	Richter syndrome
RSS	Recombination signal sequence
SAMHD1	SAM domain and HD domain 1
SEM	Standard error of the mean
SF3B1	Splicing factor 3B subunit 1
SHM	Somatic hypermutation
SNARE	Soluble NSF attachment protein receptor
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variations
TdT	Terminal deoxynucleotidyl transferase
TFS	Treatment free survival
TP53	Tumour protein 53
TSCA	TruSeq custom amplicon
TTFT	Time to first treatment
UM	Unmutated
UNG	Uracil-DNA glycosylase
UVRAG	UV Radiation Resistance Associated
VAF	Variant allele frequency
VCF	Variant calling file
WGS	Whole genome sequencing
WES	Whole exome sequencing
WIPI	WD repeat domain phosphoinositide-interacting protein

Chapter 1 Introduction

1.1 Chronic Lymphocytic Leukaemia

Chronic lymphocytic leukaemia (CLL) is the most common leukaemia in the UK. There are over 3400 new diagnoses every year representing around 38% of all leukaemias and 1% of total cancer. CLL incidence is strongly associated with age since over 70% of new diagnoses occur in the 65 years and over age-group. Males are more likely to develop the disease, having an incidence rate of 4.9 per 100,000 versus 2.6 in females (Cancer Research UK 2016a). Of interest, CLL is more common in those of a white racial background, particularly in comparison to Asian populations where CLL accounts for less than 10% of all leukaemias (Dores *et al* 2007), although this may also be due to a decrease in early diagnoses in these populations (Gross *et al* 2008, Yang *et al* 2015).

CLL is a haematological malignancy affecting B lymphocytes, a type of white blood cell. The CLL cells have increased proliferation as well as a decreased apoptosis rate, leading to their accumulation in the blood and bone marrow. Patients are typically asymptomatic at diagnosis, the disease having been detected through a routine full blood count showing a high white cell count, although some patients may present with anaemia, abnormal bleeding or suffer from infection due to the abnormal CLL cells

overwhelming the normal blood cells, preventing them from functioning normally.

Table 1.1. Staging systems in CLL. Adapted from Rai *et al*. (1975) and Binet *et al*. (1981).

Stage	Features	Median Survival (yrs)
Binet et al (1981)		
A	Involvement with <3 lymphoid areas	>10
B	Involvement with ≥3 lymphoid areas	5
C	Haemoglobin < 100g/l and/or platelet count < 100 x 10^9/l	2
Rai et al (1975)		
0	Lymphocyte count > 15 x 10^9/l	12
I	Lymphadenopathy	9
II	Lymphocyte count > 15 x 10^9/l with splenomegaly or hepatomegaly	7
III	Haemoglobin < 110 g/l	1-2
IV	Platelet count < 100 x 10^9/l	1-2

Currently there is no cure for CLL aside from a haematopoietic stem cell transplantation which is not usually suitable for those over 65 years of age. The disease has a very heterogeneous course whereby some patients have a life expectancy comparable to the general population (Parikh *et al* 2014, Rossi, Rasi, *et al* 2013) whereas others follow a more aggressive disease progression requiring early therapeutic action. Consequently, there is great diversity in the survival time with advanced stage (Binet stage C) patients having an overall survival (OS) of 1-3 years whilst Binet stage A patients have an OS of over 10 years (Table 1.1) (Binet *et al* 1981, Rai *et al* 1975,

Zenz *et al* 2010). For some patients their disease will transform to an aggressive lymphoma know as Richter syndrome (RS) which has a median overall survival (OS) of only 8-19 months (Eyre *et al* 2016, Rossi *et al* 2011).

Patients are not treated until they display symptoms (CLL Trialists' Collaborative Group 1999), therefore most patients remain untreated for several years with 30% never requiring therapeutic intervention (Bachow and Lamanna 2016). Once symptoms develop, patients are assessed for fitness based on their age, existing co-morbidities and susceptibility to infection. Patients deemed fit are treated with the chemo-immunotherapy regime of fludarabine, cyclophosphamide and rituximab (FCR) (Oscier *et al* 2012). Patients for whom this treatment is unsuitable are treated instead with chlorambucil with either ofatumumab or obinotuzumab (Goede *et al* 2014, Hillmen *et al* 2015), or alternatively with bendamustine and rituximab (Gentile, Zirlik, *et al* 2016). However, most patients will eventually relapse and the disease follows a chronic relapsing pattern.

1.2 B Lymphocytes

B cell lymphocytes are a subset of white blood cells that form part of the adaptive immune system. Adaptive immunity is highly specific towards a certain pathogen in comparison to the innate immune response which provides a more generic response. In addition, the adaptive system can remember a pathogen and improve its response on each subsequent

encounter with the same pathogen, thereby providing life-long immunity. The adaptive immune system's diverse specificity comes from genetic manipulation during the cells' development through V(D)J recombination and somatic hypermutation (SHM) (Murphy and Weaver 2016a).

The principle function of the B-lymphocyte is to produce immunoglobulins (Ig). Ig are Y-shaped proteins (Figure 1.1) that recognise and bind to a distinct target called an antigen which can be a protein expressed on the cellular surface of the pathogen or a toxin that it produces. A site in the antigen, known as the epitope, binds to the tip of the Ig in a region called the paratope. Each B-lymphocyte produces a unique paratope which will only have specificity for a single epitope. There are five different types or classes of Ig, known as isotypes. The isotypes have different functions but express the same paratope as their parent B-cell. Membrane-bound Ig form part of the B-cell receptor (BCR) which triggers cell activation upon antigen binding (Murphy and Weaver 2016).

1.2.1 VDJ Recombination

The Ig is composed of two heavy chains (IgH) and two light chains (IgL). Each chain has a variable region at its N terminus and a constant region at its C terminus. The constant region controls for the class of the Ig whereas the variable region contains the paratope. The paratope is composed of three hypervariable regions known as the complementary determining

regions (CDR) held together by four relatively conserved framework regions (FR) (Figure 1.1). Although all three CDRs interact directly with the attacking antigen, the CDR3 has been shown to be the crucial factor for antigen specificity recognition (Xu and Davis 2000) .

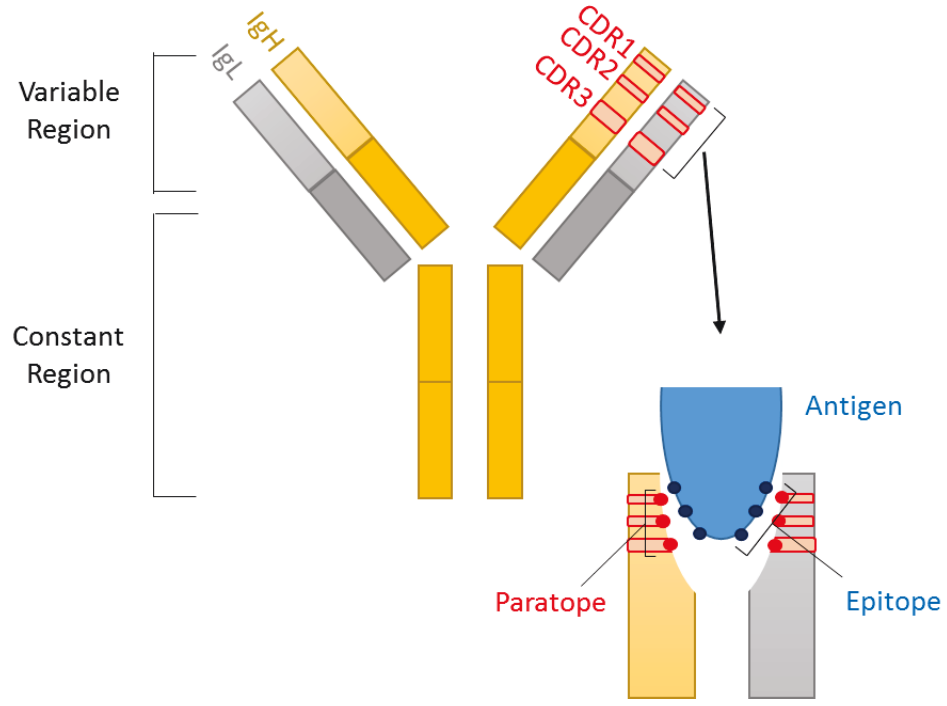


Figure 1.1 Immunoglobulin structure. Adapted from Hwang *et al*. (2014).

The gene for the heavy chain is located on chromosome 14 and diversity of the variable region is coded for through the composition of three types of genes; the diverse (D), joining (J) and variable (V) regions. The light chain can be either an Ig kappa (Igκ) or Ig lambda (Igλ), the genes for which are located on chromosomes 2 and 22, respectively, and are composed of a V and J gene. Together these genes can be rearranged to produce 2 x 10^6

different combinations (Figure 1.2). Additionally, modifications are introduced into the rearranged genes during the recombination process through the addition and removal of nucleotides. Diversity is further increased by SHM, a process of extremely high mutation rate that is at least 10^5 times greater than the normal rate of mutation.

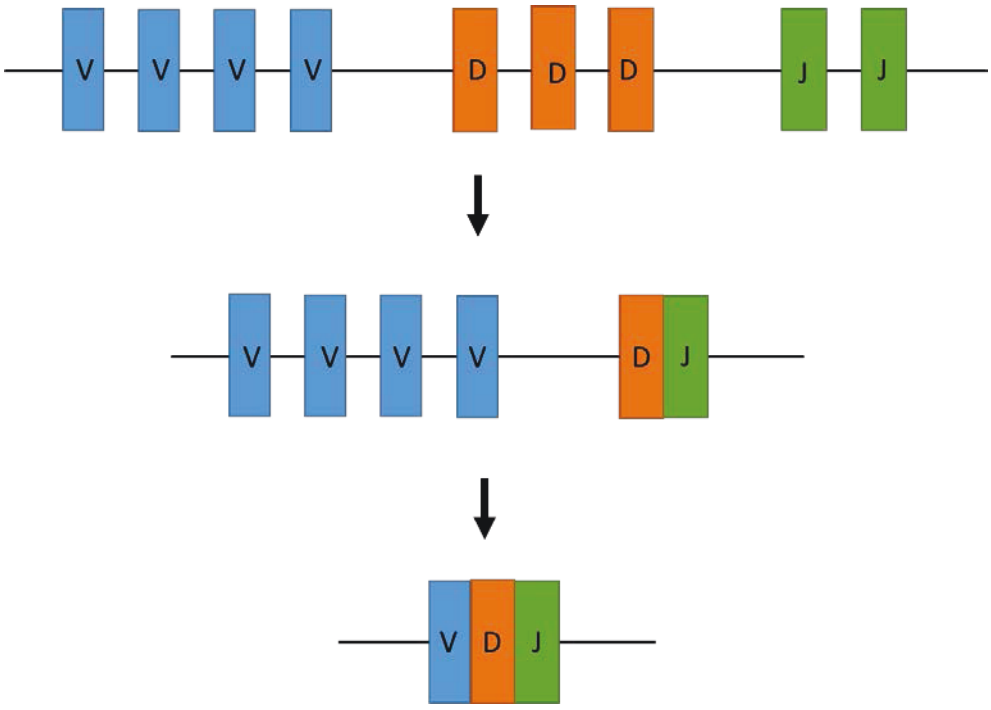


Figure 1.2. The IgH locus undergoes VDJ recombination during B-cell development. The D and J genes are rearranged first followed by the V gene. Adapted from Roth (2014).

The site of recombination is demarcated by the recombination signal sequence (RSS) which is composed of a conserved heptamer and nonamer region separated by either a 12bp or 23bp spacer sequence. Recombination typically occurs between RSSs of different spacer lengths known as the 12/23 rule. Consequently, within the IgL, the Vλ and Jκ both

have 23bp spacers whilst the Vκ and Jλ have 12bp spacers. Similarly, in the IgH the VH-gene and JH-gene segments both have spacers of 23bp whereas the DH-gene segment has a 12bp spacer on both sides (Ramsden *et al* 1994). Therefore, following the 12/23 rule, the amalgamation of Vλ-Jλ, Vκ-Jκ, DH-JH and VH-DH are permitted whilst the direct combination of the VH- and JH-genes are prevented.

Recombination can occur using both deletional and inversional rearrangement (Figure 1.3). In the more common deletional recombination, both gene segments are in the same transcriptional orientation. The DNA forms a loop enabling the 12bp spacer and 23bp spacer to come into alignment. During this type of recombination, the loop containing the signalling ends is excised when the coding ends are joined together. Inversional rearrangement takes place less frequently and occurs when the segments are in opposing orientations. To get the spacers into the correct position, the DNA forms a coil. In this instance the signalling ends are also incorporated into the DNA strand.

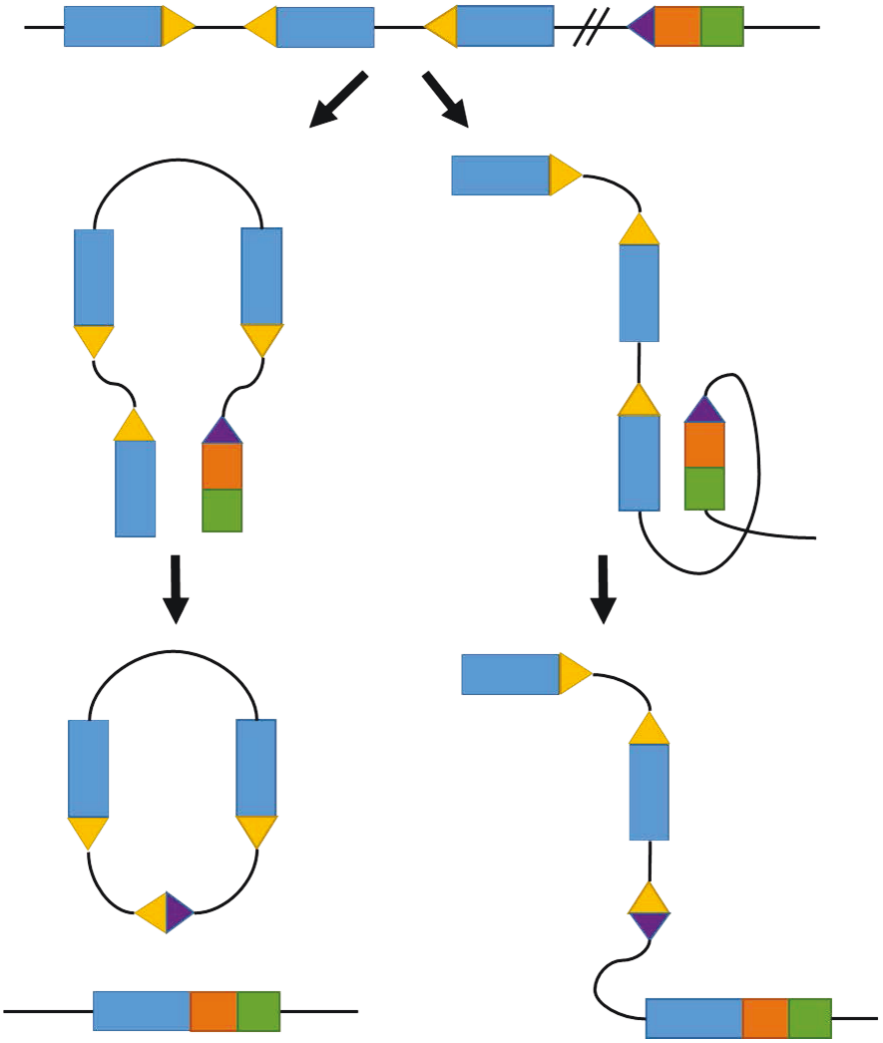


Figure 1.3. VDJ recombination uses deletional or inversional rearrangement. The left side of the figure shows deletional recombination used when the gene segments are in the same transcriptional orientation. The DNA forms a loop which is excised. The right side shows inversional recombination used when the gene segments are in opposing transcriptional orientations. In this instance, the DNA forms a coil which is retained following recombination. Blue box – V-gene, orange box – D-gene, green box – J-gene, yellow triangle – 23bp RSS, purple triangle – 12bp RSS. Adapted from Murphy and Weaver (2016).

The recombination activating gene (RAG) complex guides alignment of the opposing RSSs. This complex is comprised of a heterotetramer of the lymphocyte-specific proteins RAG1 and RAG2 (Lapkouski *et al* 2015), with a DNA-bending protein which can be either high mobility group box 1 or 2

(HMGB1 or 2) (Grundy *et al* 2009, Kim *et al* 2015). The RAG complex binds to the RSS sequence through the nonmer binding domain (NBD) on RAG1 then recruits the opposing RSS into the complex (Jones and Gellert 2002) bringing the gene pairs together. The RAG complex then cuts a single strand of the DNA between the RSS and the coding region of the gene segment. A 3'OH group is generated by this cut which then undergoes transesterification with the opposing strand, thereby creating a hairpin loop on the coding ends. The ends are then repaired using the nonhomologous end joining (NHEJ) pathway. The repair proteins, Ku70 and Ku80, bind to both the signalling and coding ends and recruit the DNA-dependant protein kinase catalytic subunit (DNA-PKcs) to the coding ends. The DNA-PKcs initiates the enzyme Artemis to randomly open the hairpin coding ends through hydrolysis of their phosphodiester bonds. The haphazard nature of this action has the potential to create overhanging ends which introduce further modifications through the addition of palindromic nucleotides (P) that are complementary to the overhanging end. Further alterations in the sequence can be made by the removal of nucleotides through exonuclease activity and through the insertion of non-template nucleotides (N) by the template-independent DNA polymerase terminal deoxynucleotidyltransferase (TdT). Finally, DNA ligase IV in combination with XRCC4 and the Cernunnos protein ligates the ends together (Figure 1.4) (Helmink and Sleckman 2012, Roth 2014, Schatz and Swanson 2011, Teng and Schatz 2015).

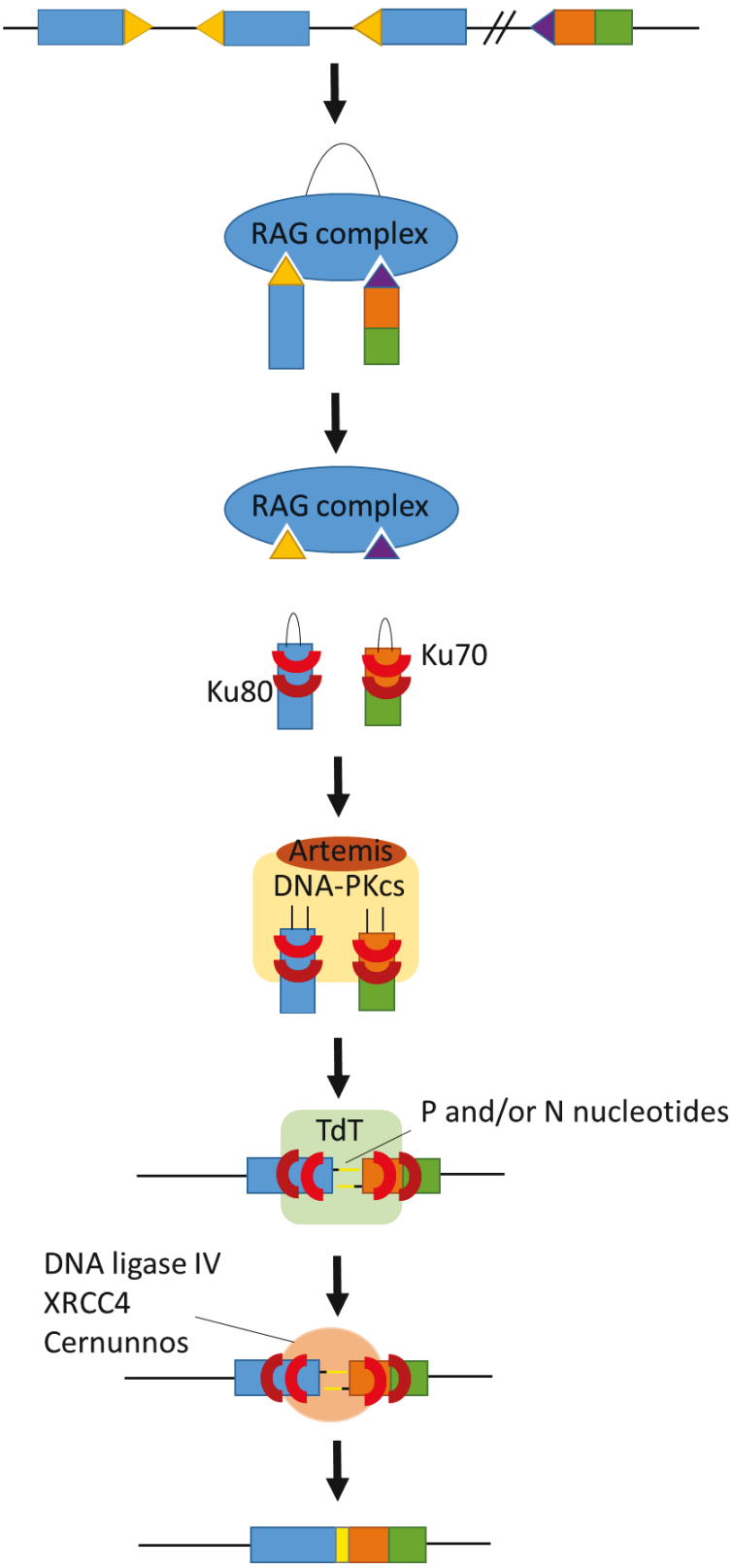


Figure 1.4. Process of V(D)J recombination. The RAG complex brings the gene pairs into alignment then cleaves the RSS leaving a hairpin loop. Ku70 and Ku80 recruit Artemis and DNA-PKcs to initiate random opening of the hairpin loops. P and/or N nucelotides are inserted, before the ends are ligated by DNA ligase IV. Blue box – V-gene, orange box – D-gene, green box – J-gene, yellow triangle – 23bp RSS, purple triangle – 12bp RSS, yellow box- newly inserted nucleotides, P- palindromic. N-non-template, TdT- terminal deoxynucleotidyltransferase. Adapted from Murphy and Weaver (2016).

The rearranged DNA must code for a functioning protein, otherwise the cell will undergo apoptosis (Lam *et al* 1997). To be functional, the recombined DNA must be (1) in-frame; that is, in multiples of three, so as to be compatible with protein synthesis; (2) have no stop codons which would result in a truncated protein; (3) contain only functional genes - that is, they must not have base alterations in their splice sites, recombination signalling sites, regulatory elements or in conserved amino acids important for correct protein folding; and (4) the protein produced through recombination must be able to pair with the protein of the second chain to produce a complete Ig (Belessi 2009). Due to the imprecise nature of the junctional joining, it is estimated that only a third of recombination produces a functional rearrangement. Therefore, there is a high level of wastage in this process. It is possible for some cells to undergo receptor editing which enables a second V-gene to replace the one in the non-functioning rearrangement. This is more common in the Igκ, but has also been described in the IgH and the Igλ (Luning Prak *et al* 2011).

V(D)J recombination takes place during the development of the B-cell in an ordered fashion and allows for multiple attempts at producing a functioning protein. The IgH gene rearranges first during the pro-B cell phase. DJ recombination occurs on both alleles in the early pro-B cell followed by the joining of the V gene to one of the DJ in the late pro-B cell (Alt *et al* 1984). If the resulting rearrangement is productive, it feeds back

to prevent the appendage of the V-gene to the DJ on the second allele in a process called allelic exclusion. If the rearrangement is out-of-frame, then the second allele undergoes V recombination. Once a productive IgH rearrangement has been generated, the heavy chain is associated with a surrogate light chain to form a pre-BCR which is expressed on the cell surface. Expression of the pre-BCR triggers proliferation and facilitates differentiation to the pre-B cell. In the pre-B cell, the light chain undergoes VJ rearrangement (Hesslein and Schatz 2001, Melchers 2005), with the Igκ, typically undergoing recombination before the Igλ. This was first implicated by the κ:λ ratio of 3:1, that is commonly found in the human B-cell repertoire (Gorman *et al* 1996, Takeda *et al* 1996). Secondly, it has been shown that cells expressing Igλ often have had rearranged Igκ, whereas the Igλ has not undergone recombination in Igκ expressing cells. This is not always the case however as there are some instances whereby Igλ expressing cells have been found to have unrearranged Igκ genes (Berg *et al* 1990).

Traditionally in genetics, allelic exclusion is monoallelic, in that only one allele is expressed whilst the other is silenced. In the Ig genes, in contrast, both Ig allele transcripts are expressed; therefore, allelic exclusion occurs through the generation of a non-functioning, unproductive or dysfunctional transcript on one allele. The stochastic model of allelic exclusion proposes that since the method of producing a functioning,

pairing-compatible rearrangement is so inefficient, it is highly improbable for the second allele to be able to replicate the process to produce a second functioning transcript. Nonetheless, the likelihood of dual expressing B-cell has been estimated to be as high as 20% based on this model (Wabl and Steinberg 1992). Studies using flow cytometry have demonstrated that in actuality this phenomenon occurs in less than 1 in 10000 of B-cells (Barreto and Cumano 2000), suggesting that there must be some feedback regulating allelic exclusion.

In the feedback inhibition model, the process of allelic exclusion occurs in response to the proliferation that is triggered when the pre-BCR is expressed on the cell surface. Proliferation is associated with increased signalling through the interleukin-7 receptor (Il-7R) pathway. The Il-7R pathway activates the signal transducer and activator of transcription (STAT)5 proteins whilst repressing the forkhead box protein O (FOXO) transcription factors. FOXO induces transcription of *Rag1* and *Rag2* (Amin and Schlissel 2008), whereas STAT5 repositions the *Rag* locus to the repressive nuclear environment of the pericentrometric heterochromatin (PCH). Therefore, the increase in activated STAT5 as well as a decrease in FOXO leads to decreased levels of RAG1 and RAG2 (Bertolino *et al* 2005, Johnson *et al* 2012). In addition, RAG2 is further reduced during the cell cycle since it is rapidly degraded during the G1 to S transition (Jiang *et al* 2005, Lee and Desiderio 1999).

Allelic exclusion has also been shown to occur during recombination. The ataxia telangiectasia mutated (ATM) kinase is activated in response to DNA double-stranded breaks (DSB), such as those generated by the Rag complex. This has been shown to prevent recombination from taking place on the second allele following Rag-mediated DNA cleavage on the first allele. ATM repositions the second un-cleaved allele to the PCH where it is inaccessible to the recombinase (Hewitt *et al* 2009). Indeed mice deficient for ATM have been shown to have an increased frequency of bi-allelic B-cells in their repertoire (Steinel *et al* 2013, 2014).

In terms of antibody diversity, the capacity of a B-cell to produce multiple functioning antibodies may at first appear beneficial. However, this can lead to problems with tolerance, the system that prevents the production of autoreactive Ig. Indeed, in the autoimmune disease systemic lupus erythematous (SLE), a disorder whereby antibodies are targeted towards the patient's own DNA and nucleoproteins, deficiencies have been reported in the kappa deleting element (KDE), whose function it is to remove unwanted Igκ rearrangements, leading to the dual expression of a functional Igκ and Igλ (Fraser *et al* 2015).

1.2.2 Somatic Hypermutation

Following successful V(D)J recombination, the immature B-lymphocyte expresses the IgM on its cell surface as its BCR, and migrates to the germinal centres (GC) so that it may encounter a compatible antigen. The compatible antigen activates the B-cell through binding of the BCR. Once activated, the B-cell undergoes a process called somatic hypermutation (SHM) whereby multiple single point mutations are introduced to the V(D)J and its flanking DNA. SHM occurs most commonly in the region coding for the CDRs of the Ig thereby affecting the strength of the interaction between the B-cell and the antigen (Wilson *et al* 1998).

SHM is initiated by the activation induced deaminase (AID) enzyme in the first stage of a two-step process (Figure 1.5). The AID enzyme only acts on single stranded DNA; therefore, it is only functional on DNA that is actively being transcribed. AID works by deaminating a cytidine to a uridine which creates a mismatch with the pairing guanidine as well as introducing a foreign nucleoside to the DNA strand. Replication of the strand at this stage leads to a transition mutation at this site. Alternatively, the uridine stimulates the second step of SHM through the initiation of the DNA repair pathways including the mismatch repair (MMR) and the base-excision repair (BER) pathways. The BER pathway utilises uracil-DNA-glycosylase (UNG) to excise the uracil thereby creating an abasic site in the DNA. Provided there are no further modifications, the DNA polymerase Rev1 will

insert a deoxycytidine opposite this abasic site resulting in a C:G transversion (Nelson *et al* 1996, Prakash *et al* 2005). Another option is for the mismatch to be detected by the MSH2-MHS6 heterodimer from the MMR pathway. The mismatch and its bordering DNA are then removed by EXO1. The resulting patch is then filled in by the low fidelity polymerase Polη, which is prone to A:T errors during synthesis creating mismatches in the surrounding sequence (Rada *et al* 2004).

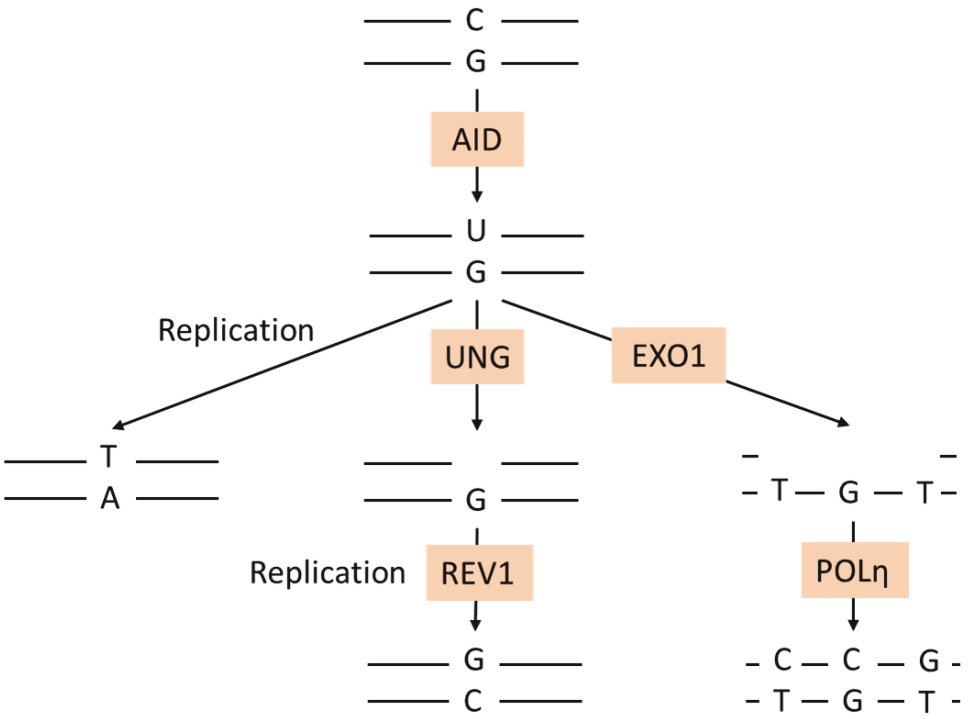


Figure 1.5. Mechanisms of somatic hypermutation. AID deaminases the cytosine to a uracil. a. the uracil is replicated to a thymidine. b. The uracil is excised by the UNG enzyme, then a cytosine incorporated in the opposite strand during replication. c. The surrounding DNA is excised by the EXO1 enzyme, then repaired with the low fidelity Polη polymerase. Adapted from Odegard and Schatz (2006).

Another method of repairing the error is through a process called gene conversion. In this instance, following excision of the uracil by Uracil-DNA glycosylase (UNG), the surrounding DNA is also removed by the enzyme apurinic/apyrimidinic endonuclease 1 (APE1). The resulting gap is then repaired using the neighbouring genes as templates. While gene conversion is a common mechanism of diversification in birds, it is not thought to be used in humans. However, there is some evidence that it may have role in receptor editing (Darlow and Stott 2006), a process that initiates a secondary rearrangement in a functioning rearranged V(D)J in order to remove an autoreactive BCR (Luning Prak *et al* 2011).

1.3 The IgH in CLL

CLL is a neoplasm affecting B-lymphocytes, resulting in the increase of mature B-cells. Since these are cells that have left the bone marrow, their IgH locus would have undergone differentiation. Additionally, as a cancer is the uncontrolled growth of clonal cell, and as each B cell expresses a single IgH rearrangement, the IgH locus could act as a clonal marker of CLL. Initial analysis of the IgH locus in CLL led to the assumption that CLL cells were antigen naïve as early studies only demonstrated germline IgHV sequences (Kipps *et al*. 1989). However, a literature review in the early 90's of VDJ transcripts in CLL found that 39 of 75 VH sequences differed significantly from the germline indicating that these cells had undergone SHM (Schroeder & Dighiero 1994).

1.3.1 IgHV Mutational Status

The finding that some patients with CLL had a mutated IgHV (Fais *et al* 1998) fuelled further analysis of the region. This lead to the breakthrough discovery that IgHV mutation status was an important prognostic marker. Two studies independently identified that patients could be divided into two groups based on those whom had undergone SHM and those that had not. In both studies, an unmutated (UM) IgHV region, defined as greater than 98% homology to the germline sequence, was found to be associated with a worse prognosis than those found to be hypermutated (M), with the UM patients having a median survival of 9 years compared to over 17 years for the M patients (Damle *et al* 1999, Hamblin *et al* 1999).

The 98% cut-off that is used to distinguish the UM and M groups is an arbitrary value chosen to take into account germline polymorphisms and possible sequencing error (Ghia *et al* 2007), thereby alleviating the need for germline sequencing of the IgHV. Therefore, caution should be used when interpreting borderline case as there are studies which suggest that the greater the level of mutation the better the prognosis (Tobin *et al* 2005). Furthermore it has been shown that, in terms of prognosis, 97% is equivalent to 98% (Davis *et al* 2016), and a cut-off of 96% homology has been proposed as a better discriminating factor (Kröber *et al* 2002).

The distinction between the M and UM groups is reflected on a protein level with the two groups showing biological differences. Patients with an UM IgHV have a higher expression of ZAP-70 (Crespo *et al* 2003, Rassenti *et al* 2004, Rosenwald *et al* 2001, Wiestner *et al* 2003), a protein tyrosine kinase involved in BCR signalling (Chen *et al* 2002). The UM group have also been found to have higher levels of cells expressing CD38 (Damle *et al* 1999, 2002, Kröber *et al* 2002, Matrai *et al* 2001), a cell surface marker that is detectable on immature B-cells, but lost once the cells have left the GC (Hamblin 2003). They also express high levels of lipoprotein lipase (LPL) (Oppezzo *et al* 2008), beta-2-microglobulin (Oscier *et al* 2010) and serum thymidine kinase (Magnac *et al* 2003).

Genetic differences have also been reported; telomeres are shorter in UM CLL compared with M CLL (Damle *et al*. 2003) and the two groups have been shown to have different gene expression profiles (Klein *et al* 2001, Rosenwald *et al* 2001). Furthermore, the groups are known to correlate with certain genomic abnormalities since genotypes associated with a poorer prognosis are more recurrently described in the UM subgroup, such as the poor prognosis cytogenetic abnormalities the 17q and 11p deletions (Section 1.4.2) (Kröber *et al* 2002). Additionally *KRAS* mutations have been found exclusively in UM-CLL (Herling *et al* 2016).

1.3.1.1 The IGHV3-21 Gene

Even though the IgHV mutational status has been repeatedly demonstrated to correlate with prognosis (Kröber *et al* 2002, Oscier *et al* 2010), there are exceptions, the most reported of which is found in patients expressing IGHV3-21. This has been reported as M and UM yet these patients are known to have an inferior survival (Lin *et al* 2003, Tobin *et al* 2002) as well as a shorter time to first treatment (TTFT) then M-CLL (Ghia *et al* 2008) regardless of their mutational status. IGHV3-21-CLL share characteristics with UM-CLL, such as elevated expression of ZAP-70 and CD38 as well an increased incidence of poor prognostic cytogenetics (Kienle *et al* 2013, Thorselius *et al* 2006). It was speculated that these cases may in fact be UM-CLL that appeared mutated due to polymorphisms in the IGHV3-21 gene (Nollet *et al* 2002). However, this was not confirmed by germline analysis of M-IGHV3-21 cases, and further analysis of these mutations found that around half occurred in the hypermutation hotspots, suggesting they were caused by SHM (Tobin *et al* 2003). Furthermore, gene expression studies have shown that patients with IGHV3-21 rearrangements have a distinct gene expression profile that differs from both M- and UM-CLL (Falt *et al* 2005, Kienle *et al* 2006). Interestingly, IGHV3-21 has an higher incidence in Northern Europe - of around 11% (Lin *et al* 2003, Tobin *et al* 2002, 2003) - compared to Southern Europe where it is found in around 3% of CLL cases (Ghia *et al* 2005).

1.3.2 Stereotyping

Although B-cells can produce an extraordinary diverse catalogue of immunoglobulins, the repertoire identified in CLL cells has been found to be skewed with overrepresentation of selected heavy chain genes that are inconsistent with the repertoire in age-matched controls (Potter *et al*. 2003; Fais *et al*. 1998). Over 35% of CLL cases express either IGHV1-69, IGHV4-34, IGHV3-7 or IGHV3-23 (Fais *et al* 1998, Murray *et al* 2008, Stamatopoulos *et al* 2005). Additionally, certain V-genes correlate with mutational status, since around a quarter of all UM-CLL express IGHV1-69, whereas the IGHV3-family are found in nearly 60% of M-CLL (Murray *et al* 2008). This selection bias is also seen with the IgHJ, whereby IGHJ4 is overrepresented in M-CLL and IGHJ6 is more frequently reported in UM-CLL (Stamatopoulos *et al* 2007). The IgL has also been shown to be biased in its usage in CLL and is frequently found to have the same mutation status as its heavy chain counterpart (Stamatopoulos *et al* 2005).

Bias has also been reported in the V(D)J gene combinations. A large-scale study of 1220 patients with CLL noted that there was restricted use of the IgHJ and IgHD genes rearranged with some IgHV genes, leading to the creation of VH CDR3s with conserved amino acid sequences. In particular, 1% of the cases analysed had an IGHV1-69/IGHD3-16/IGHJ3 rearrangement. Furthermore, these cases also used the same IgLV gene, IgκV3-20 (also known as A27) (Widhopf *et al* 2004). Another study reported

a recurrent IGHV4-39/IGHD6-13/IGHJ5 rearrangement in a fifth of its cohort. As with the previous study, the cases all utilised the same IgLV gene - in this instance the IgκV1-39 (Ghiotto *et al* 2004). Restricted IgL usage has also been described in the IGHV3-21 rearrangement whereby the IGλV3-21 (Vλ2-14) is reported in around 75% of cases (Ghia *et al* 2008, Thorselius *et al* 2006, Tobin *et al* 2003). These recurrent features all implied that there were similarities in the BCR in many unrelated cases of CLL.

Analysis of the BCR sequence in 30 patients with a functional IGHV3-21 rearrangement showed that 40% of them had homologous heavy chain variable region complementary determining region 3 (VH-CDR3) (Figure 1.1). For seven of these patients, the amino acid sequence in the VH-CDR3 was identical, and a further five differed by a single amino acid (Tobin *et al* 2003). Another study of 255 patients identified five subsets of VH-CDR3 sequences that had over 60% amino acid identity. The authors termed this phenomenon as stereotypy and were able to classify 8% of their cohort into one of these five subsets (Messmer *et al* 2004). Larger scale studies found that stereotyped VH-CDR3s were present in up to a third of CLL cases. Restricted Ig gene usage was a feature of stereotyped BCRs as non-stereotyped had a more diverse Ig gene repertoire. This implies that these two classifications of CLL may have different ontogeny and may therefore have different features and potentially require different therapies

(Agathangelidis *et al* 2012, Darzentas *et al* 2009, Stamatopoulos *et al* 2007).

In the largest stereotyping study to date of over 7000 CLL cases, the authors revised the criteria for determining stereotypy in the BCR to make it more stringent. Additionally, since the different IgHV genes are known to originate from three ancestral clans (Kirkham *et al* 1992), stereotyped BCRs had to contain IgHV genes from the same clan since these have phylogenetically similar sequences. Therefore, BCRs were classified as stereotyped if they had:

- ≥50% amino acid identity i.e. at least 50% of the amino acid sequence matched precisely
- ≥70% amino acid similarity i.e. at least 70% of the amino acid sequence had similar amino acids based on their physiochemical properties
- IgHV genes belonged to the same clan
- Identical VH-CDR3 length
- Identical sequence pattern offsets i.e. sequence patterns were at the exact same location in the VH-CDR3

Table 1.2. Characteristics of the major stereotyped subsets. AA- amino acid. Adapted from Agathangelidis *et al* (2012).

Subset No	Freq in CLL (%)	Freq in stereotyped CLL (%)	IGHV gene(s)	IGHJ gene(s)	VH-CDR3 AA length	VH-CDR3 AA sequence pattern
#2	2.8	9.2	V3-21	J6	9	[AVLI]x[DE]xxxM[DE]x
#1	2.4	8.0	Clan I genes	J4	13	ARx[NQ]W[AVLI]xxxxFDx
#4	1.0	3.2	V4-34	J6	20	[AVLI]RGxxxxxxx[KRH]RYYYYGx[DE]x
#6	0.9	2.9	V1-69	J3	21	ARGGxYDY[AVLI]WGSYRxx[DE][AVLI]FDx
#5	0.7	2.3	V1-69	J6	20	ARxxxxxx[AVLI]xxxYYYYxMDx
#3	0.6	1.8	V1-69	J6	22	Axxxxx[AVLI][AVLI]VxxAxxxxYYGMDx
#8	0.5	1.5	V4-39	J5	19	AxxxxYSSSWxxxxNWFDP
#31	0.4	1.3	Clan III genes	J6	21	ARxxxxxxxxxxxxYYYxMDx
#16	0.3	1.1	V4-34	J6	24	AxRFYCxGxxCxxxxxxYYxGxDx
#77	0.3	1.1	V4 (V4-4, V4-59)	Various	14	[AVLI]RGxxx[ST]GWxxxxx
#7C	0.3	1.0	V1-69	J6	24	AxxxxxxDFW[ST]GYxxxxYYYxxDx
#28A	0.3	1.0	V1 (V1-2)	J6	17	ARxxxGxxYYYYYGMDx
#201	0.3	1.0	V4-34	J3	17	ARRxxxWxxxxxDxxDx
#59	0.3	1.0	Clan I genes	Various	12	AxxxDFWSGxxx
#12	0.3	1.0	V1-2, V1-46	J4	19	ARDxxYYDSSGYY[ST]xxxDx
#14	0.3	0.9	V4 (V4-4)	J4	10	x[KRH]GGxWxFDx
#64B	0.3	0.9	V3 (V3-48)	J6	21	A[KRH][DE]xx[AVLI]VVxxx[AVLI]xYYYYGMDx
#99	0.3	0.9	Clan I genes	J4	14	ARxQWLxxxxxFDx
#202	0.3	0.9	Clan III genes	J3, J4	14	A[KRH]xxxGxx[AVLI]xxxDx

Using these criteria, 952 different BCR sequences were detected that were quasi-identical in at least two cases. Of these, the authors identified 19 major subsets (Table 1.2) that were found in 12% of the overall cohort and 41% of stereotyped cases, with subset #1, #2 and #4 comprising half of those (20% overall) (Agathangelidis *et al* 2012). These subsets have been found to share clinical and molecular features in addition to BCR similarities.

Subset #2 is the most commonly occurring subset, found in 2.8% of all CLL cases. It is associated with poor prognosis similar to that of UM-CLL despite the majority having undergone some level of SHM (Bomben *et al* 2009, Ghia *et al* 2005, Stamatopoulos *et al* 2007) The inferior survival reported in patients expressing IGHV3-21 has been confirmed to be a feature of subset #2, as non-subset #2 IGHV3-21 behave in a similar manner to their respective counterparts depending on their mutation status (Baliakas, Agathangelidis, *et al* 2015, Bomben *et al* 2009). Subset #2 patients have a higher incidence of *SF3B1* mutations, especially in comparison to the other major subsets #1 and #8 (Strefford *et al* 2013), as well as being enriched for 11q deletions (Thorselius *et al* 2006) and *ATM* mutations (Navrkalova *et al* 2016).

Subset #1 is the next most common, found in 2.4% of CLL cases (Agathangelidis *et al* 2012). The IgHV in these patients are found to be

exclusively UM. They are reported to have an inferior median survival of 84 months compared to 234 months in the non-subset # 1 cohort, a group of patients expressing the same UM-IgHV genes as the subset #1 patients but without the stereotyped BCR (Bomben *et al* 2009, Stamatopoulos *et al* 2007). Subset #1 patients have been found to have an increased frequency of *NOTCH1* mutations which is found in around 40% of the cohort (Rossi, Spina, *et al* 2013).

Patients in subset #4 are typically younger with a median age of 43 years at diagnosis. These patients are reported to have an indolent disease and the median survival is yet to be determined (Stamatopoulos *et al* 2007). This is supported by an absence of unfavourable molecular markers such as poor prognostic cytogenetics, UM-IgHV and mutations in *TP53*, *SF3B1* and *NOTCH1* (Maura *et al* 2015, Rossi, Spina, *et al* 2013).

The other 16 major subsets combined make up 20% of stereotyped cases of CLL. Within these subsets #6, #5, #3, #8, #31, #7C, #28A, #59, #12, #64B, #99 and #202 are usually UM whereas #16, #77, #201, and #14 are generally M (Agathangelidis *et al* 2012). The UM subsets have a substantially shorter TTFT than the M subsets (Baliakas *et al* 2014). Subset #8 is associated with a particularly poor prognosis with patients having a high risk of transformation to RS (Rossi *et al* 2009).

The findings that some patients with CLL had undergone SHM as well as the identification of stereotyping in the BCR of CLL cells has given rise to the theory that antigens/superantigens may drive pathogenesis in CLL (Stamatopoulos *et al* 2007). The hypothesis proposes that these antigens stimulate cell proliferation through activation of the BCR, although whether this is limited to the stage at which malignant transformation takes place or if this is ongoing throughout the disease has yet to be established.

1.4 Prognostic Markers in CLL

In addition to the IgHV mutation status there have been other genetic aberrations that have been identified in CLL that have been found to be predictive of prognosis. These include large copy number alterations (CNA) that can be detected in the chromosomes as well as single nucleotide variations (SNV) and small insertion/deletion (indel) mutations in recurrent genes.

1.4.1 Abnormalities affecting the TP53 gene

TP53 is a vital tumour suppressor gene which codes for the protein p53. The p53 protein is a key regulator of cell division that acts by triggering various pathways involved in preventing the proliferation of aberrant DNA.

It does this through arresting the cell cycle and initiating DNA repair or apoptosis in response to cellular stress (Lane 1992). Abnormalities in the *TP53* gene have been described in many cancers including CLL (Olivier *et al* 2010, Rivlin *et al* 2011). In all cancers, defective *TP53* is associated with poorer prognosis and treatment resistance (Petitjean *et al* 2007). *TP53* defects occur in ~10% of pre-treated CLL and are associated with significantly shorter TFS and OS (Dohner *et al* 1995, Gonzalez *et al* 2011, Landau *et al* 2013, Zenz *et al* 2010).

TP53 is located on the p-arm of chromosome 17. In CLL, both deletions of the 17p locus as well as mutations in the gene itself have been described. Additionally, the identification of one type of *TP53* mutation is strongly associated with presence of another i.e. patients with a 17p deletion often have a mutation in the remaining *TP53* gene (Zenz *et al* 2008). *TP53* mutations are known to cause refractoriness to traditional CLL chemotherapy as most conventional chemotherapy agents operate through the *TP53* pathway. However *TP53* mutated patients have been shown to respond to treatment with novel therapies, such as Idelalisib and ibrutinib, that by-pass the *TP53* pathway and instead target the B-cell receptor pathway to prevent the proliferation of B-cells (Burger *et al* 2014, Byrd *et al* 2014, Farooqui *et al* 2015, Furman *et al* 2014).

1.4.2 Chromosomal Aberrations

Around 80% of CLL patients have chromosomal aberrations (Dohner *et al* 2000). Of these, there are four key chromosomal abnormalities that have been identified as commonly occurring in CLL. These are the 13q14 deletion which is associated with good prognosis, trisomy 12 which is thought to be an intermediary marker and the 11q23 and17p deletions which confer a poorer prognosis. Other abnormalities have also been detected, such as the 6q deletion, but these occur at a lower frequency (Dohner *et al* 1999).

The 13q14 deletion is the most common cytogenetic abnormality found in up to 61% of CLL cases (Dohner *et al* 2000, Edelmann *et al* 2013). It is thought to be a favourable prognostic marker when found in isolation with patients having a median TTFT of 6 years and OS of over 12 years (Van Dyke *et al* 2016). However, it has also been associated with a reduced progression free survival (PFS) time when found in greater than 65% of nuclei and also when associated with the more severe 17p deletion (Van Dyke *et al* 2009). Furthermore, it has also been found that patients with larger deletions of the 13q have a shorter TTFT of under 3 years (Parker *et al* 2011).

Extensive characterisation over several large cohort studies have defined the predominant minimally overlapping region (MOR) in the 13q deletion as containing *DLEU1*, *DLEU2* and the microRNAs *miR15a* and *miR16-1* (Knight *et al* 2012, Ouillette, Collins, Shakhan, Li, Li, *et al* 2011, Parker *et al* 2011). This is consistent with studies that have shown that patients with 13q deletions have reduced expression of *miR15a* and *miR16-1* (Calin *et al* 2002, Ouillette, Collins, Shakhan, Li, Li, *et al* 2011). However more recent data have suggested that only *DLEU1* and the first 2 exons of *DLEU2* are involved in the MOR (Edelmann *et al* 2013). There is evidence that *DLEU2* acts as a promotor for the microRNAs (Kasar *et al* 2014, Lerner *et al* 2009), therefore defects in *DLEU2* alone may still cause their downregulation.

The prognostic significance of trisomy 12 is uncertain. In isolation it is considered an intermediate marker with patients having a similar TTFT and OS as those with normal karyotype (Dohner *et al* 2000, Van Dyke *et al* 2016). However, its role in disease progression is still undetermined as although it is not an indicator of aggressive disease (Catovsky *et al* 2007), nor does its frequency increase over time (Gunnarsson *et al* 2011), it does have associations with transformation to RS (Chigrinova *et al* 2013). In keeping with this, trisomy 12 is reported to be in 87% of subset #8 patients who also have a high incidence of Richter transformation (Rossi, Spina, *et al* 2013). Furthermore, *NOTCH1* mutated trisomy 12 CLL has a more

aggressive disease course reducing OS by over a third (Balatti *et al* 2012, Del Giudice *et al* 2012).

11q23 deletions are found in up to 27% of CLL cases and are more common in patients with an early disease onset as well as those with a marked lymphadenopathy (Byrd *et al* 2004, Edelmann *et al* 2013, Hernández *et al* 2015). These patients have a shorter remission period when treated with chemotherapy alone (Catovsky *et al* 2007). However, since the introduction of the immunotherapy drug rituximab as a standard therapy, the number of patients having a PFS of over 3 years has doubled (Fischer *et al* 2016, Hallek *et al* 2010). The MOR has been refined to the *ATM* gene (Edelmann *et al* 2013) which was previously found to be downregulated in patients with 11q deletions (Kienle *et al* 2005). The *ATM* gene is important in cell cycle regulation and can induce apoptosis in response to DNA damage (Bredemeyer *et al* 2006) as well as having a role in V(D)J recombination (Section 1.2.1 above). In addition to deletions of the 11q, mutations in *ATM* have been reported in 12% of CLL cases and are associated with a reduced OS (Austen *et al* 2005). Furthermore *ATM* mutations on the residual allele have been reported in 36% of patients with 11q deletions (Austen *et al* 2007). Patients with biallelic inactivation of *ATM* have been shown to have shorter TTFT, an inferior response to treatment as well as an increased risk of disease progression and shorter OS comparable with *TP53* mutated CLL

(Austen *et al* 2007, Hurtado *et al* 2015, Lozano-Santos *et al* 2016, Skowronska *et al* 2012).

Array technologies have been shown to be a powerful tool which can be used to identify genetic aberrations in cancers. Comparative genomic hybridization (CGH) arrays operate by comparing the copy numbers of differentially labelled test and reference DNA samples (Barrett *et al*. 2004). Single nucleotide polymorphism (SNP) arrays are composed of oligonucleotide probes which correspond to the allelic variants of selected SNPs. SNP arrays have the added advantage of being able to detect copy neutral loss of heterozygosity (cnLOH) which is when a cell contains two copies of the same allele (Gunderson *et al*. 2005).

CGH arrays have been found to have a high concordance with fluorescent in-situ hybridisation (FISH) analysis in the detection of chromosomal abnormalities in CLL, although are unable to detect abnormalities at low levels (Schwaenen *et al*. 2004; Tyybakinoja *et al*. 2007) despite detection sensitivity having improved from 53% of the total cell population down to around 25% (Patel *et al*. 2007). Studies using CGH arrays identified novel recurring regions of interest such as small deletions on areas of chromosomes 8 and 2 (Grubor *et al*. 2009). The long arm of chromosome 22 has also been reported to have recurrent deletions particularly in a region containing the *PRAME* gene (Gunn *et al*. 2008). However, recent

evidence suggests that this is an artefact of Igλ rearrangement and not pathogenic in nature (Mraz *et al* 2013).

Genome analysis of CLL patients using SNP arrays have identified a previously undescribed gain of the 2p16 locus which contains the *REL* and *BCL11A* oncogenes. These gains occurred in patients with UM-IGHV and are more frequent in patients with a 17p deletion (Pfeifer *et al*. 2007; Forconi *et al*. 2008). Later studies identified 78 SNPs that were implicated in PFS (Sellick *et al*. 2008), six novel risk loci (Di Bernardo *et al*. 2008) and a novel MOR on chromosome 15 (Edelmann *et al* 2013). Furthermore, it has been demonstrated that elevated numbers of CNAs are associated with progressive disease and shorter survival (Knight *et al* 2012, Kujawski *et al* 2008, Ouillette, Collins, Shakhan, Li, Peres, *et al* 2011).

1.4.3 Other Genetic Markers

In addition to chromosomal abnormalities, genome-wide screening technologies such as whole genome (WGS) and whole exome sequencing (WES) have identified further genetic aberrations that are recurrent in CLL, including mutations in the *NOTCH1*, *SF3B1*, *BIRC3* and *MYD88* genes (Baliakas *et al* 2014, Landau *et al* 2013, Schuh *et al* 2012). Although these genes are mutated at a low frequency in cohorts of untreated patients with CLL (<10%) (Foà *et al* 2013), there have been many reported associations

with refractory disease and inferior survival rates (Jethwa *et al* 2013, Nadeu *et al* 2016, Wang *et al* 2011).

Notch1 is a transmembrane receptor that upon activation undergoes cleavage of the Notch intercellular domain (NICD). NICD translocates to the nucleus where it associates with the CSL DNA-binding complex to initiate transcription (Kopan and Ilagan 2009). Mutations in the *NOTCH1* gene are reported in up to 12% of pre-treated CLL (Chiaretti *et al* 2014, Jethwa *et al* 2013, Pozzo *et al* 2015, Puente *et al* 2011, Rossi, Rasi, *et al* 2012) with their frequency increasing in patients with advanced disease (Jethwa *et al* 2013, Villamor *et al* 2013). These mutations, predominately a 2-bp frameshift deletion (c.7544_7545delCT) in exon 34, are all predicted to disrupt the PEST domain (Chiaretti *et al* 2014, Puente *et al* 2011, 2015, Rossi, Rasi, *et al* 2012), preventing the degradation of NICD by F-box and WD repeat containing protein 7 (FBXW7) leading to the accumulation of NICD in the nucleus. *NOTCH1* mutations are enriched in trisomy 12 patients (Balatti *et al* 2012, Del Giudice *et al* 2012) as well as in the subset #8 cohort (Rossi, Spina, *et al* 2013). Mutations in *FBXW7* have also been reported in CLL, often in association with trisomy 12 (Jeromin *et al* 2014, Wang *et al* 2011). Interestingly, recent data have demonstrated that patients with a *NOTCH1* mutation do not benefit from treatment with rituximab (Bo *et al* 2014, Stilgenbauer *et al* 2014).

Mutations in splicing factor 3B subunit 1 (*SF3B1*) are reported in around 10% of pre-treated CLL cases and are considered to be a poor prognostic marker as they are associated with a short TTFT and OS (Jeromin *et al* 2014, Nadeu *et al* 2016, Wang *et al* 2011). Similar to *NOTCH1,* their mutation frequency increases over time and in relapse cohorts (Baliakas, Hadzidimitriou, *et al* 2015, Guièze *et al* 2015, Jethwa *et al* 2013). *SF3B1* mutations are strongly correlated with subset #2 (Rossi, Rasi, *et al* 2013, Strefford *et al* 2013) as well as deletions in 11q (Jeromin *et al* 2014). The SF3B1 protein is a component of the small nuclear ribonucleoprotein complex that functions in the initiating phases of RNA splicing (Shin and Manley 2004). Patients with mutations in *SF3B1* in CLL have been shown to have aberrant splicing (Quesada *et al* 2012, Wang *et al* 2011). As with *NOTCH1*, mutations have also been reported in CLL in genes that operate downstream to *SF3B1*, including *XPO1* (Quesada *et al* 2012) and *DDX3X* (Landau *et al* 2015, Wang *et al* 2011).

Mutations have also been recurrently described at frequencies of less than 5% in pre-treatment CLL in baculoviral IAP repeat containing 3 (*BIRC3*) and myeloid differentiation primary response 88 (*MYD88*) (Puente *et al* 2011, Quesada *et al* 2012, Wang *et al* 2011). These genes both function upstream in the NFκB pathway, therefore their disruption leads to constitutive activation of NFκB (Puente *et al* 2011, Rossi, Fangazio, *et al* 2012). Despite this, *BIRC3* is associated with a poor TTFT (Baliakas, Hadzidimitriou, *et al*

2015), shorter OS than that seen in wild-type patients (Nadeu *et al* 2016). It is highly correlated with UM-IgHV whereas *MYD88* mutations are exclusively found in M-IgHV (Baliakas, Hadzidimitriou, *et al* 2015, Landau *et al* 2015, Puente *et al* 2011, Wang *et al* 2011) and have not been shown to have a negative impact on OS (Puente *et al* 2011).

Several other genes have been reported to be recurrently mutated in CLL, albeit at very low incidences. These include the DNA repair genes SAM domain and HD domain 1 (*SAMHD1*), protection of telomeres 1 (*POT1*) and chromodomain helicase DNA binding protein 2 (*CHD2*), all found at frequencies of around 3% (Clifford *et al* 2014, Landau *et al* 2015, Quesada *et al* 2012, Wang *et al* 2011). Also the oncogenes *KRAS* and *BRAF* have reported in 2% of CLL cases (Landau *et al* 2013, Zhang *et al* 2011). More recently, mutations in the ribosomal protein S15 (*RPS15*) gene were found at a high frequency in a cohort of patients who relapsed within 2 years of FCR treatment (Ljungström *et al* 2016).

1.5 Risk stratification

Despite the identification of a plethora of reportedly clinically significant markers, there is still great uncertainty when predicting the TTFT for many patients. Mutations affecting the *TP53* gene have a well described correlation with short TTFT and chemo-refractoriness in the patients

harbouring them. However, there are a cohort of patients without these poor prognostic factors that will do badly.

Table 1.3 Prognostic risk stratification as devised by Rossi, Rasi et al (2013)

	Cytogenetics	**Molecular mutations**
High Risk	del(17p)	*TP53* and/or *BIRC3*
Intermediate Risk	del(11q)	*NOTCH1* and/or *SF3B1*
Low Risk	tri(12)	-
Very Low Risk	del(13q)	-

To address this issue, recently devised prognostic algorithms have incorporated genetic markers into their scoring (Rossi, Rasi, *et al* 2013, The International CLL-IPI Working Group 2016). The model proposed by Rossi, Rasi, *et al* (2013) stratifies patients into four risk categories based on the presence or absence of certain molecular characteristics (Table 1.3). Application of this model to 1274 patients determined that the OS over 10 years was 29.1% for the high-risk group, 37.1% for the intermediate group, 57.3% for the low risk group and 69.3% for very low risk group. However, neither Jeromin *et al* (2014) or Baliakas, Hadzidimitriou, *et al* (2015) were able to discriminate between the high and intermediate risk categories for TTFT in their cohorts of 930 and 3490 patients, respectively.

Table 1.4. Prognostic risk stratification as devised by the Internal CLL-IPI Working Group (2016). Very high risk = 7–10, high risk = 4–6, intermediate risk = 2–3, low risk = 0–1.

Factor	Adverse Feature	Assigned Score
TP53	Mutated or deleted	4
IgHV	Unmutated	2
β2-microglobulin	>3.5mg/L	2
Rai Stage **Binet Stage**	I-VI B-C	1
Age	>65	1

The new prognostic index proposed by the International CLL-IPI Working Group (2016) also classifies patients into four groups. However, *TP53* and IgHV mutation status are the only molecular markers used. The index works by applying a prognostic scoring system (The International CLL-IPI Working Group 2016) to separate patients into very high risk (score 7–10), high risk (score 4–6), intermediate risk (score 2–3) or low risk (score 0–1). On a cohort of 1214 patients, the proportion of patients surviving over 5 years was 23.2% for the very high-risk group, 63.3% for the low risk group, 79.3% for the intermediate group, and 93.2% for low risk group. Three independent validation cohorts supported these results bringing the total number of patients tested to 1839. The index has been criticised for two reasons. Firstly, since the data are based on clinical trial patients they do

not consider 'watch and wait' patients. Therefore, the results have a limited application in determining TTFT, which can be valuable in identifying patients at risk of rapid disease progression, particularly as many newly diagnosed patients are classified as Binet Stage A (Baliakas, Hadzidimitriou, *et al* 2015). Secondly, the clinical trials were performed prior to the introduction of chemo-immunotherapy and therefore may not be suited in determining OS in patients treated with these therapies. However, subsequent studies have validated the index for determining TTFT (Gentile, Shanafelt, *et al* 2016, Molica *et al* 2016) as well as OS in patients treated with chemo-immunotherapy (da Cunha-Bang *et al* 2016).

1.6 Cancer as an Evolutionary Process

The theory that all cancers arise from a single clone that has undergone progressive acquisition of somatic mutations that confer a survival advantage, was first postulated forty years ago (Nowell 1976). If this is the case, it would therefore follow that cancer is governed by the same tenants underlying evolutionary biology; that is, the generation of genetic variation, the influence of genetic drift, i.e. random life and death events and Darwinian selection, i.e. the survival of those with a fitness advantage. Genetic variation occurs within the tumour through the acquisition of a somatic mutation. The survival of this mutation relies firstly on genetic drift, as even beneficial mutations have been shown to die out before they have a chance to expand (Levy *et al* 2015). Should the mutation escape

extinction and provide a survival advantage to the tumour, it will then gradually increase in abundance and, since the acquisition of somatic mutations is a random event, a heterogeneous population of cancer cells is created. Indeed, there are multiple studies which demonstrate intra-clonal heterogeneity across a variety of cancer types as well as providing evidence supporting cancer progression as an evolutionary process (Anderson *et al* 2011, Awad *et al* 2013, Diaz Jr *et al* 2012, Gerlinger *et al* 2012, 2014, Landau *et al* 2015, Schuh *et al* 2012).

Whilst the model of clonal evolution in cancer is now well accepted, the acquisition of recurrent mutations across inter-clonal samples is, at first, difficult to explain in the clonal evolution theory of cancer. If it is assumed that genetic variation is a random event, how then can the recurrence of specific mutations be explained? It is known that most cancers have an elevated mutation rate (Jones *et al* 2008, Loeb *et al* 2003). In addition, the processes that have been found to be responsible for generating somatic mutations (Alexandrov *et al* 2013) are biased to targeting certain DNA sequence patterns (Table 1.5). For instance, the AID enzyme, that is essential in SHM, preferentially deaminases cytosine found in a 5'-WRCN-3' sequence. Furthermore, it has been hypothesised that even with a slow mutation rate of 1 mutation per cell division, a tumour containing 10^{10} cells would have developed every point mutation at least once (Wang, Walters *et al* 2014).

Table 1.5. Summary of the DNA sequence motifs recognised by some of the common causes of somatic mutations. Adapted from Roberts and Gordenin (2014). *AID, activation-induced cytidine deaminase; APOBEC, apolipoprotein B mRNA editing enzyme catalytic polypeptide-like; POLη, DNA polymerase η, ROS, reactive oxygen species, POLε, DNA polymerase ε.*

Cause of Mutation	Sequence Motif	Base Change
APOBEC	T**C**W	C>K
AID	WR**C**N	C>K
POLη	N**T**W	T>N
5-meCpG deamination	N**C**G	C>T
ROS	N**C**N	C>A
Tobacco	N**C**N	C>A
POLε	T**C**G	C>T
	T**C**T	C>A

In contrast to the evolution theory of cancer development, some recent studies have suggested that mutations are acquired rapidly in a relatively few, or even a single, catastrophic event. This 'Big Bang' model of genomic derangement suggests that tumour subclones are all formed during the initiation of the cancer as opposed to from later clonal expansions. Furthermore, the pervasiveness of these subclones is more dependent on when they occur during tumour development as opposed to being selected

for due to a survival advantage (Baca *et al* 2013, Gao *et al* 2016, Sottoriva *et al* 2015, Stephens *et al* 2011).

1.6.1 Clonal Evolution and Tumour Heterogeneity in CLL

CLL is a good model of disease evolution since most, if not all, patients have the pre-malignant disorder of monoclonal B-cell lymphocytosis (MBL) (Landgren *et al* 2009). MBL is an asymptotic condition characterised by the clonal expansion of B-cells at a frequency of less than 5000 cells per cubic millilitre (Strati and Shanafelt 2015). MBL has been shown to have cytogenetic and molecular abnormalities that are thought to be early drivers of CLL pathogenesis (Landgren *et al* 2009, Ojha, Secreto *et al* 2014, Rasi *et al* 2012). However, since the level of MBL transformation to CLL is particularly low at a rate of 1.1% a year (Rawstron *et al* 2008b), yet the frequency of some CLL-like mutations such as the del13q are reported at around 50% (Henriques *et al* 2014, Morabito *et al* 2013), there must be addition factors driving the pathogenesis of CLL from a background of MBL.

Tumour progression is further demonstrated in CLL through the development of Richter syndrome. Around 5% of CLL cases will transform into RS which has an incidence rate of 0.5% per year (Parikh *et al* 2013). In most cases of RS, the tumour is clonally related to the preceding CLL and transformation has taken place in a linear fashion. This is demonstrated through possession of the same IgH rearrangement and mutational status

(Rossi, Spina, *et al* 2012) as well as through maintenance of mutations present in the CLL clone (Chigrinova *et al* 2013, Fabbri *et al* 2013).

Work performed by Campbell *et al* (2008) in the early days of NGS showed that ultra-deep sequencing analysis of the IgHV could be used to demonstrate intraclonal diversification. Furthermore, analysis of the phylogenetic relationship between the subclones as well as their evolution from the germline, indicated a complex model of tumour development whereby some subclones denoted intermediaries to the dominant clone, others signified divergent clones, whilst a proportion represented ongoing evolution. The presence of divergent clones detectable through NGS analysis of the IgHV was confirmed in a later study (Bashford-Rogers *et al* 2013).

Nowadays, intra-tumour heterogeneity is well reported in CLL through the use of NGS technologies, or the analysis of consecutive samples, or a combination of both (Knight *et al* 2012, Landau *et al* 2013, Schuh *et al* 2012). Furthermore, these studies have demonstrated how these subclones can expand over time to become the major clone in response to external pressures, particularly since patients who have undergone treatment have a greater degree of clonal evolution (Landau *et al* 2013). This is supported by studies that have shown that the presence of subclonal mutations are associated with the same prognosis as clonal

mutations in certain genes such as *TP53* (Rossi *et al* 2014) and *NOTCH1* (Nadeu *et al* 2016).

1.6.2 The Leukaemic Stem Cell in CLL

The cell of origin in CLL remains controversial. Since some CLL cells are known to have undergone SHM (Damle *et al* 1999, Hamblin *et al* 1999) and all have clonal IgHV rearrangements (Fais *et al* 1998), it was proposed that CLL cells derive from mature B-cells (Chiorazzi and Ferrarini 2003). Later it was found that CLL cells shared many features with marginal zone B- (MZB) cells (Chiorazzi and Ferrarini 2011). However, a more recent study found that CLL and MZB cells have distinct expression profiles and instead proposed a previously undefined mature $CD5^+$ B-cell as the cell of origin (Seifert *et al* 2012).

Relapsed CLL is known to have a distinct profile genetically to its diagnostic predecessor since mutations that enable the cell to survive treatment were able to expand whereas those with mutations susceptible to therapy died out (Schuh *et al* 2012, Stilgenbauer *et al* 2007). In spite of these transient changes, it is well accepted that the IgHV remains stable over the course of the disease, in that the V(D)J rearrangement does not change from diagnosis to relapse (Provan *et al* 1996). This suggests that the CLL and the subsequent relapse clone must have arisen from a B-cell population that has undergone V(D)J recombination, supporting the notion that the cell of

origin is a mature B-cell. Further corroboration is shown in cases that have transformed to RS as these have the same IgHV rearrangement as their CLL predecessor (Rossi, Spina, *et al* 2012). Moreover, it is now widely accepted that CLL develops from pre-malignant condition MBL. MBL has been shown to be oligo-clonal in terms of IgHV rearrangements in a high proportion of cases (Klinger *et al* 2015, Lanasa *et al* 2010), yet bi-clonal IgHV rearrangements are only reported in <5% of CLL cases (Plevova *et al* 2014, Rassenti and Kipps 1997, Visco *et al* 2013). This may suggest that a subclone of the MBL population acquires further mutations, potentially as an indirect result of antigen activation of the BCR, which instigates the transformation to CLL.

Whilst these provide evidence for a mature B-cell as the normal counterpart for CLL, the origin of the leukaemic stem cell in CLL is yet to be elucidated. Certainly, the presence of oligo-clonal MBL suggest that the leukaemia initiating event must occur in an early progenitor B-cell, at least, and could be related to the normal ageing process that leads to B-cell oligoclonality in the age-matched healthy population (Gibson *et al* 2009). Recent data show that leukaemia inducing mutations, such as those in *SF3B1,* can be detected in the progenitor B-cells of patients with CLL (Damm *et al* 2014). Furthermore, studies in mice have shown that engrafting haemopoetic stem cells from patients with CLL into immunodeficient mice gives rise to clonal B-cells with V(D)J

rearrangements that are independent of the original clone (Kikushige *et al* 2011).

Nonetheless, an early B-cell as the CLL leukaemic stem cell would be expected to give rise to a proportion of CLL expressing multiple V(D)J rearrangements. Yet, using conventional methods, only 5% of CLL are reported to have bi-clonal IgHV rearrangements (Plevova *et al* 2014, Rassenti and Kipps 1997, Visco *et al* 2013). At such low frequencies, it is possible that these actually represent two lymphoproliferative diseases that have arisen independently, as opposed to one disorder possessing two IgHV rearrangements. Indeed, clonally unrelated RS was considered as a secondary disease that had arisen independently to the CLL due to the presence of genetic lesions different to their CLL counterparts (Rossi *et al* 2011). However, further studies found that they possess the same V(D)J rearrangement suggesting that they arose as a result of branching evolution in the pre-CLL phase (Fabbri *et al* 2013). Identification of a sub-clone within CLL possessing the same DJ rearrangement as the dominant clone but with a different V-gene would provide further evidence that the CLL derives from a haematopoietic stem cell.

1.7 Aims of thesis

The primary aim of this thesis was to characterise the genomic complexity and molecular drivers of clonal expansion and maintenance of chronic lymphocytic leukaemia (CLL) in the context of clinical intervention. More specifically, the objectives were to develop, validate and apply methods to identify and begin to functionally characterise the various genetic abnormalities within paired pre-treatment and relapse samples using (1) ultra-deep NGS sequencing of the IgHV and (2) genome wide array technology. As a final step, the approach was applied to samples from large cohorts of patients with robust clinical outcome to determine the prognostic/predictive significance of these findings.

Chapter 2 Materials and Methods

This chapter describes the protocols used in this work. Additional protocols, including primer sequences used for specific projects, are detailed in the relevant chapters and/or the appendices.

2.1 Cell Preparation

2.1.1 Ficoll Preparation of Peripheral Blood Mononuclear Cells

Whole blood or bone marrow from EDTA tubes was transferred into 50ml centrifuge tubes. Phosphate buffered saline [PBS] (10010031, Life Technologies, Carlsbad, California, USA) with 1% fetal bovine serum [FBS] (10270-106, Life Technologies) was added at 2.5x volumes. 4ml of Ficoll (17-5442-02, GE Healthcare, Little Chalfont, Buckinghamshire, UK) was added to 15ml centrifuge tubes. 10ml of blood/PBS mix was carefully layered on top of the Ficoll. The tubes were centrifuged for 20mins at 400g at 4°C. The upper layer was removed and discarded, whilst the white cell fraction was transferred to a new 15ml tube. PBS was added to make a total volume of 14ml, then the tubes centrifuged for 5mins at 280g. The supernatant was removed and discarded and 500ul of RBC lysis solution (158904, Qiagen, Hilden, Germany) added. The tubes were incubated at room temperature for 10mins then made up to a volume of 10ml with Dulbecco's Modified Eagle Medium [DMEM] (11995-065, Life

Technologies). The tubes were centrifuged for 5mins at 280g and the supernatant removed and discarded. PBS was added to give a final concentration of $1x10^7$ cells/ml. Cells were stored in 2ml cryogenic vials containing 1ml of cell preparation.

2.1.2 Thawing Frozen Cells

Thawing of frozen cells must be carried out in a timely manner to prevent cell death. Vials containing cells are agitated in a 37°C waterbath till thawed. Cells are then transferred dropwise to 15ml centrifuge tubes containing 10ml DMEM. The tubes are then centrifuged for 5mins at 180g and the supernatant removed. 1ml of PBS was then added.

2.1.3 Isolating B-Cells

B-cells were isolated using the EasySep Human CD19 selection kit (18054, Stem Cell Technologies, Vancouver, Canada). Equal amounts of peripheral blood and 1x EasySep RBC lysis buffer were mixed in a 14ml falcon tube. CD19 EasySep Positive Selection Cocktail was added to a final concentration of 25ul/ml and the tubes incubated at room temperature for 15mins. EasySep Magnetic Nanoparticles were then added to a final concentration of 25ul/ml and the tubes incubated at room temperature for 10mins. The samples were made up to 10ml with PBS containing 1% FBS and incubated at room temperature in the Big Easy EasySep Magnet (18001, Stem Cell Technologies) for 10mins. The supernatant was then

discarded by inverting the magnet. The remaining cells were pipette mixed with 10ml PBS with 1%FBS, then incubated for 5mins at room temperature in the magnet, before the supernatant was again discarded. This was repeat twice more and the remaining cells resuspended in 400ul of PBS with 1% FBS.

2.2 Extraction of Genomic DNA

The QIAamp DNA mini kit (51306, Qiagen) uses spin columns to isolate DNA. Unless otherwise stated, the reagents used for this protocol were provided in this kit. The sample is first lysed with buffers and enzymes that also stabilise the DNA and enhance its binding to the silica-gel membrane within the column. Wash buffers are used to remove impurities, and then the DNA is eluted into low-salt buffer (Buffer AE).

Once thawed, cells were made up to 200ul with PBS in a 1.5ml microcentrifuge tube. 200ul of buffer AL and 20ul Qiagen protease was added and the tubes pulse-vortexed. The tubes were then incubated at 56°C for 10mins. 200ul ethanol was added and the tubes pulse-vortexed. The mix was transferred to a QIAamp spin column and centrifuged at 6000g for 1min. The spin column was then transferred to a clean 2ml collection tube and 500ul Buffer AW1 added. The column was centrifuged at 6000g for 1min and then transferred to a clean collection tube. 500ul

Buffer AW2 was added and the column centrifuged at 20,000g for 3min. The column was transferred to a new collection tube and centrifuged at 20,000g for a further 1min. The column was then transferred to a 1.5ml microcentrifuge tube and 200ul Buffer AE added to the membrane. The column was incubated at room temperature for 1min then centrifuged at 6000g for 1min.

2.3 Extraction of RNA and cDNA preparation

2.3.1 Preparation of cells for RNA extraction

2.3.1.1 From Whole Blood

Blood was transferred to 50ml tubes and made up to 50ml with RBC lysis solution. Tubes were rotary mixed at room temperature for 10mins then centrifuged at 1100g for 10mins. The supernatant was discarded then 20ml of RBC lysis solution added. Tubes were rotary mixed for 5mins then centrifuged as before and the supernatant was discarded. RBC lysis solution was added to give a maximum concentration of $3.5x10^7$ cells/ml. The cell preparation was transferred to 1.5ml microcentrifuge tubes in 1ml volumes. The tubes were spun at 18,000g for 2mins then the supernatant discarded. 600ul of Buffer RLT (79216, Qiagen) containing 1% β-mercaptoethanol (M3148, Sigma Aldrich, St. Louis, Missouri, USA) was then added and the cells homogenised with a 20-gauge needle. Samples were then either stored at -70°C or processed to the next step.

2.3.1.2 From Ficoll Preparations or Sorted Cells

Both types of cell preparations have already undergone red cell lysis. Therefore, these cells start at the microcentrifuge tube stage and progress as described in section 2.3.1.1.

2.3.2 RNA extraction using the QIAamp RNA Blood Mini Kit

RNA extraction from the GTC lysate was performed using the QIAamp RNA Blood Mini Kit (52304, Qiagen). First, 600ul 70% ice-cold ethanol was added to the microcentrifuge tubes and the tubes pipette-mixed. The mix was transferred to a QIAamp spin column and centrifuged at 18,000g for 15s. The spin column was then transferred to a clean 2ml collection tube and 700ul Buffer RW1 added. The column was centrifuged at 18,000g for 15s and then transferred to a clean collection tube. 500ul Buffer RPE was added and the column centrifuged at 18,000g for 15secs. A further 500ul Buffer RPE was added and the tubes centrifuged for 2mins. The column was transferred to a new collection tube and centrifuged at 18,000g for a further 1min. The column was air-dried at room temperature for 5mins, then transferred to a 1.5ml microcentrifuge tube and 30ul molecular-grade water (W4502, Sigma Aldrich) added to the membrane. The column was incubated at room temperature for 2min then centrifuged at 18,000g for 2min.

2.3.3 cDNA preparation

6ul of RNA was transferred to a 0.5ml microcentrifuge tube and 2ul 5x Vilo, 1ul 10X Superscript (11754250, Life Technologies) and 1ul molecular-grade water added. The tubes were then incubated at 25°C for 10mins, then 42°C for 60mins and finally the reaction terminated at 85°C for 5mins. The cDNA was then made up to a final volume of 50ul with molecular grade water.

2.4 Nucleic Acid Assessment

2.4.1 Nanodrop

A Nanodrop 2000c UV-Vis spectrophotometer (ThermoFisher Scientific, Waltham, Massachusetts, USA) was used to determine nucleic acid concentration and purity. With the measurement arm raised, 1.5ul of test sample was pipetted onto the pedestal. The measurement arm was then closed so that the sample was held between the light source and the detector. The Nanodrop 2000 software package (v1.6.198) determines the measurement and record the results for each sample. Prior to analysis of the first sample, the baseline measurement is established by analysing 1ul of the solution used to re-suspend the DNA/cDNA/RNA sample. The sample pedestal is wiped clean with distilled water between each test sample. A 260/280 ratio of ~1.8 was considered as ‘pure’ DNA and of ~2.0 as ‘pure’ RNA.

2.4.2 Qubit

For some samples, a Qubit 2.0 Fluorometer (Thermo Fisher Scientific) was also used to quantify DNA specimens using the ds BR assay kit (Q32850, Thermo Fisher Scientific). Since this method can differentiate between DNA and other contaminates it is a more accurate methodology than spectrophotometric based methods. A working dilution of Qubit reagent was prepared by diluting the Qubit dsDNA reagent by a factor of 1:200 into Qubit dsDNA Buffer. To prepare the two standards, 10ul of each standard was diluted into 190ul of the Qubit working solution. The test samples were prepared by diluting 1ul of sample into 199ul of Qubit working solution. The tubes were then mixed by vortexing and left at room temperature for 2mins. The standards were measured first to generate a standard curve, against which each sample is measured.

2.4.3 Bioanalyser

Following amplification and cleaning, PCR product was assessed for purity using the Agilent High Sensitivity DNA Assay (5067-4626, Agilent Technologies, Santa Clara, California, USA) on the Agilent 2100 Bioanalyser. For each sample, 1ul of product was loaded into a single well of the chip and processed using a 2100 Bioanalyzer instrument (Agilent Technologies) as per the manufacturer's instructions.

2.5 Polymerase Chain Reaction (PCR)

Conventional PCR was employed to amplify regions of interest within the DNA. Unless otherwise stated, a standard 25ul reaction mix comprised of 12.5ul Multiplex PCR MasterMix (206145, Qiagen), 1ul of forward and reverse primer (10uM of each primer) and 1ul of template DNA (25-250ng/ul). Primers were designed using the Primer3 software (http://bioinfo.ut.ee/primer3-0.4.0/) using all the default settings except for adjusting the GC clamp to 1. The amplification conditions consisted of an initial denaturation step of 95°C for 10mins followed by 35 cycles of denaturation at 95°C for 30secs, primer annealing at 60°C for 1min and primer extension at 72°C for 2mins, followed by a final 10min extension at 72°C.

2.6 Sanger Sequencing

2.6.1 PCR amplification of the IgHV locus using the InvivoScribe Somatic

Samples were amplified using the IGH Somatic Hypermutation Assay v2.0 (5-101-0041, Invivoscribe, San Diego, California, USA). 1ul of DNA/cDNA was added to 25ul of Hypermutation Mix 1 containing 0.1ul AmpliTaq Gold (N808-0241, Applied Biosystems, Foster City, California, USA). The amplification conditions consisted of an initial denaturation step of 95°C for 7mins followed by 35 cycles of denaturation at 95°C for 45secs, primer

annealing at 60°C for 45secs and primer extension at 72°C for 90secs, followed by a final 10min extension at 72°C.

2.6.2 PCR clean-up using MicroClean

Following amplification, equal volumes of PCR product and MicroClean reagent (2MCL50, Microzone Ltd, Heywards Heath, UK) were mixed by vortexing in a PCR plate and left at room temperature for 5mins. The plate was then centrifuged at 4700rpm for 40mins, and then spun inverted for 60secs at 38rpm on a folded paper towel to remove the supernatant. The cleaned product was re-suspended in DNase free water up to half the volume of starting PCR product and allowed to stand for 10mins to re-hydrate.

2.6.3 Cycle sequencing

A 10ul Sanger PCR mix comprised of 1x ready reaction premix, 1x BigDye sequencing buffer (4337456, Thermo Fisher Scientific), 30pmol of sequencing primer and 1ul of purified PCR product. The amplification conditions consisted of an initial denaturation step of 96°C for 1min followed by 25 cycles of 96°C for 10secs, 50°C for 5secs and 60°C for 4mins.

2.6.4 Ethanol precipitation

5µl of freshly prepared 125mM EDTA (93302, Sigma Aldrich) and 60µl of cold 100% ethanol was added to each cycle sequencing reaction. The sample was vortex mixed and incubated at room temperature for 15mins. The plate was centrifuged at 3700rpm for 30mins at 4°C. The plate was then inverted to remove the supernatant and washed again in 60µl of ice cold 70% ethanol. The samples were centrifuged at 3500rpm for 15mins, the supernatant removed and the samples allowed to air-dry. The dried samples were re-suspended in 8.5µl of Hi-Di formamide (4311320, Thermo Fisher Scientific) prior to sequencing.

2.6.5 Capillary electrophoresis and Data Analysis

Samples underwent capillary electrophoresis on a 3130 Genetic Analyzer (Applied Biosystems). Samples were run on a 96-well plate using the Sequencing Analysis Application.

Following capillary electrophoresis, raw sequencing was first analysed using the sequencing analysis software (v5.4, Applied Biosystems), to generate quality values and basecalling data files. These output files can then be used for downstream analysis. IgHV SHM analysis was performed by importing a FASTA file into IMGT/VQUEST (Lefranc *et al* 2009). Variant identification was performed using Mutation Surveyor (SoftGenetics, State College, Pennsylvania, USA). This software takes analysed ABI files and

performs a direct base-to-base comparison against a FASTA file of the wild-type sequence.

2.7 Next Generation Sequencing

2.7.1 AMPure clean-up

Following amplification, the libraries were purified to select amplicons for re-sequencing, and remove excess reaction components that could interfere with the downstream processes. Purification is performed using magnetic beads which will bind and elute DNA using specific buffers.

PCR products were made up to 25ul using elution buffer (19086, Qiagen) in a 96-well microplate. 25ul of AMPure XP Beads (A63882, Beckman Coulter, High Wycombe, UK) were added to each sample and vortex mixed at 1800RPM for 2mins. The samples were incubated at room temperature for 10mins then the plate was placed onto a magnetic stand for 2mins. The supernatant was removed using a multi-channel pipette and the samples washed twice in freshly-prepared 80% ethanol. The plate was air-dried for 10mins then the dried samples were resuspended in 25ul of elution buffer.

2.7.2 Nextera

Amplified cleaned PCR products were tagged with the Index 1 (i7) and Index 2 (i5) adapters and sequences required for cluster formation for

MiSeq sequencing using the Nextera XT DNA library kit (FC-131-1096, Illumina, San Diego, California, USA). The i7 and i5 tags uniquely barcode the sequences from individual samples so that they can be identified post sequencing. A 50ul Nextera mix contained 2ul of cleaned PCR product, 15ul of Nextera PCR Mastermix (NPM), 5ul i5 tag and 5ul i7 tag. The amplification conditions consisted of an initial step of 92°C for 3min followed by 95°C for 30sec then 5 cycles of 95°C for 10secs, 55°C for 30secs and 72°C for 30secs and a final extension of 72°C for 5mins. The samples were then cleaned again using the AmpPure technique.

2.7.3 KapaQuant

Once cleaned, the labelled PCR products were quantified using the Kapa Library Quantification kit (KK4835, KapaBiosystems, Woburn, Massachusetts, USA). A 20ul reaction mix consisted of 12ul KAPASYBR Fast qPCR mix, 0.4ul low Rox and 4ul of standard or 1:1000 diluted PCR product. PCR was performed in the 7500 Real-Time PCR System (Applied Biosystems). The amplification conditions consisted of an initial denaturation step of 95°C for 5mins followed by 35 cycles of denaturation at 95°C for 30secs and primer annealing at 60°C for 45secs. Absolute quantification for each sample was determined against the standard curve using the KAPA Library Quantification Data Analysis Template.

2.7.4 MiSeq Preparation and Loading

Samples underwent next generation sequencing using the either the MiSeq Reagent Kit v2 300-cycle kit (MS-102-2002, Illumina) or the MiSeq Reagent v3 600-cycle kit (MS-102-3003, Illumina) on the MiSeq Instrument (Illumina). Cleaned PCR products were quantified using the Qubit BR assay kit (Section 2.4.2), then normalised to 4nM. 2ml of each normalised product was then pooled, and then requantified using the Qubit BR assay kit. 5ul of pooled product was added to 5ul of freshly prepared 0.2M NaOH then incubated at room temperature for 5mins, and then 990ul of HT1 added. In a new microtube, this library was diluted to the required loading concentration with HT1 and 11.5pmol PhiX. 600ul of the final library was then loaded onto the cartridge.

2.7.5 Analysis of the IgHV locus using the LymphoTrack® IGH Somatic Hypermutation Assay Panel

1.25ul of template cDNA/DNA (25-250ng/ul) was amplified in a 12ul reaction mix comprised of 11.25ul LymphoTrack IGH SHM Leader PCR MasterMix (71210069, Invivoscribe) and 0.1ul AmpliTaq Gold (Thermo Fisher Scientific). The amplification conditions consisted of an initial denaturation step of 95°C for 7mins followed by 33 cycles of denaturation at 95°C for 45secs, primer annealing at 60°C for 45secs and primer extension at 72°C for 90secs, followed by a final 10min extension at 72°C. Next the PCR products were purified and quantified as previously described

(Sections 2.7.1 and 2.4.2 respectively) and then analysed using the MiSeq Reagent v3 600-cycle kit (Section 2.7.4) to generate FASTQ data.

The FASTQ data generated by the MiSeq was analysed using the LymphoTrack IGHV SHM Software (Invivoscribe). First, data was analysed using the LymphoTrack_IGH_SHM_MiSeq_1.0 software to generate a fastq_read_summary file. This file was then visualised using the LymphoTrack_IGH_SHM_Visualization.xlsm macro file on Excel. Reads that were out of frame or that contained stop codons were discounted from the analysis. Sequences were also analysed using IMGT HighV-Quest (Alamyar *et al* 2012).

2.7.6 Targeted Next Generation Sequencing using the Illumina TruSeq Custom Amplicon assay

A TruSeq Custom Amplicon assay (TSCA) (Illumina) was used to analyse genes which had been reported as recurrently mutated in CLL. The libraries were prepared using the TSCA kit (FC-130-1001, Illumina). Unless otherwise stated, the reagents used for this protocol were provided in this kit. First the oligos were hybridised to the sample template by adding 5ul of TSCA mastermix to 10ul of DNA (150-250ng) and 35ul of OHS2. This was then heated to 95°C then left to cool to 40°C for ~90mins. Unbound oligos were then removed by washing twice with SW1 and once with UB1 in a filter plate. The products were then extended and ligated by adding 45ul of

ELM4 to each sample and incubating at 37°C for 45mins. Next 25ul of freshly prepared 50nM NaOH was added and the samples incubated at room temperature for 5mins. 20ul of this product was added to a PCR mastermix consisting of 4ul i5 tag, 4ul i7 tag, 21.6ul PMM2 and 0.4ul TDP1. The amplification conditions consisted of an initial denaturation step of 95°C for 3mins followed by 25 cycles of denaturation at 95°C for 30secs, primer annealing at 66°C for 30secs and primer extension at 72°C for 60secs, followed by a final 5min extension at 72°C. PCR products were purified and quantified as previously described (Sections 1.7.1 and 1.4.2 respectively) and then analysed using the MiSeq Reagent v2 300-cycle kit (Section 1.7.4).

The variant calling file (VCF) generated by the MiSeq was analysed in VariantStudio v2.2 (Illumina). Variants were filtered to remove those with less than 10 reads, those with a variant allele frequency (VAF) of less than 5%, as well as those with no consequence. Variants which had been validated in dbSNP (v138) were assumed to be germline and were also excluded from the analysis.

2.8 Single Nucleotide Polymorphism Array

2.8.1 Illumina Genome-wide SNP Platform hybridisations

Single Nucleotide Polymorphism (SNP) arrays were used to detect copy number alterations (CNAs). Initial arrays were performed using 1M-Duo SNP chips (WG-311-1102, Illumina); however, due to product discontinuation, subsequent samples were analysed using HumanOmni1-Quad SNP chips (WG-311-1110, Illumina). Unless otherwise stated, the reagents used for this protocol were provided in these kits.

250ng of DNA was denatured by adding 0.1M NaOH and incubating at room temperature for 10mins. Next the denatured DNA underwent whole-genome amplified by adding 68ul MA2 and 75ul MSM to each sample, and then incubating for 20-24hrs at 37°C. The amplified samples were then enzymatically fragmented by adding 50ul FMS and incubating at 37°C for 1hr. Next, the fragmented samples were precipitated by adding 100ul PMS and incubating at 37°C for 5mins, then adding 300ul isopropanol (I9516, Sigma Aldrich), mixing well, incubating at 4°C for 30mins and then centrifuging at 4700rpm for 20mins, before removing the supernatant. The pellets were left to air-dry for 1hr, then resuspended in 46ul of RA1 by incubating at 48°C for 1hr. Samples were hybridized onto the BeadChips by rocking in a hybridisation oven at 48˚C for 16-24hrs and then the BeadChips were washed in PB1. The hybridised primers were extended with labeled nucleotides and then stained. Finally, the BeadChips were

coated with XC4 and then analysed on the iScan (Illumina). The data was processed using GenomeStudioV2009.2 (Illumina) to generate a txt file.

2.8.2 Data Analysis with Biodiscovery Nexus

The txt file generated by GenomeStudio was analysed using Nexus 8 Discovery Edition (BioDiscovery, El Segundo, California, USA) with the following settings (SNPRank Segmentation): Significance Threshold 1x 10-5; Max Contiguous Probe Spacing (Kbp) 1000.0; Min number of probes per segment 5; High Gain 0.6; Gain 0.2; Loss -0.2; Big Loss -1.0; 3:1 sex chromosome gain 1.2; Homozygous Frequency threshold 0.95; Homozygous Value Threshold 0.8; Heterozygous Imbalance Threshold 0.4; Minimum LOH Length (Kb) 20; percentage outliers to remove 3%. All Nexus processed plots were also inspected visually to scan for changes not identified using these analysis settings.

2.9 Data Analysis

TFS and OS distributions were plotted using Kaplan–Meier estimates and were compared using the log-rank test. TFS and OS were calculated from the time of diagnosis until the date of first treatment and the date of death, respectively. All deaths included in this analysis were CLL-related. When the death was CLL-unrelated, the death event was not included in the OS analysis. All tests were two-sided. An affect was considered

statistically significant at $P<0.05$. All analyses were performed with Graphpad Prism 5.0 software (La Jolla, California, USA), IBM SPSS 13.0 software (Chicago, Illinois, USA) or Microsoft Excel (Richmond, Washington, USA).

Chapter 3 Using Next Generation Sequencing for Analysis of the Immunoglobulin Heavy Chain Locus

3.1 Introduction

It is widely accepted that the mutational status of the IgHV gene is independently prognostically significant in CLL (Damle *et al* 1999, Hamblin *et al* 1999). Current practice is to determine the mutational status using SSeq (Ghia *et al* 2007). SSeq presents challenges when analysing the IgHV region, however, as it only enables analysis of the dominant clone unless further manipulation is performed such as gel extraction or cloning. In addition, SSeq analysis of the IgHV region has a reported failure rate of up to 20% (Burger *et al* 2014).

NGS has the capacity to analyse multiple sequences simultaneously so it is highly suited to analysing samples with sub-clonal populations. Campbell *et al* (2008) and Bashford-Rogers *et al* (2013) have both previously demonstrated that it is possible to detect sub-clones in CLL through the analysis of the IgHV region. However, neither study investigated the significance of these subclones in terms of correlation with disease progression or pathogenesis. As discussed previously, CLL is widely thought

of as a monoclonal disorder in terms of IgHV, that solely affects mature B-cells. However, if it were possible to identify IgHV subclones from different Vh families this would suggest that the pre-leukaemic stem cell would be pre-B in origin, if not earlier. Therefore, the aim of the study in this chapter was to develop and validate an NGS method to analyse the IgHV region. Three methods were used; an in-house assay on the Roche 454 platform, an in-house assay on the Illumina MiSeq platform and the Invivoscribe Lymphotrack IGH Somatic Hypermutation Assay.

3.2 Materials and Methods

3.2.1 Patient Samples

3.2.1.1 In-house Assays

Unsorted cells from 15 patients with CLL were used in these studies (From population C, Appendix 2). For the 454 analysis, DNA was extracted as previously described (Section 2.2) and for the MiSeq analysis cDNA was extracted as previously described (Section 2.3). In addition, cDNA from 2 non-CLL patients was used to illustrate the normal B-cell repertoire.

3.2.1.2 Invivoscribe IGH Somatic Hypermutation Assay

For this study cDNA was extracted from CD19+ purified cells from a cohort of 270 patients (Population A, Appendix 2)

3.2.2 Sanger Sequencing

Samples were amplified and analysed using the InvivoScribe Somatic Hypermutation Kit V2.0 as described previously (Section 2.6.1).

3.2.3 454 Analysis

3.2.3.1 Sample Preparation and Running

The clone-specific VDJ rearrangement in the IgHV was amplified using the specific Vh family primers (GS FLX primers, Appendix 1) in a 50ul mix

containing 25ul Phusion PCR Buffer (M0532S, NEB), 1ul of forward and reverse primer (10uM of each primer) and 1ul DNA extracted from sequential samples taken at diagnosis, post first treatment and at subsequent relapse on 3 CLL patients with an unmutated IgHV and 1 patient with a hypermutated IgHV. The amplification conditions consisted of an initial denaturation step of 94°C for 3min followed by 30 cycles of 94°C for 30secs, 60°C for 45sec and 72°C for 1mins, followed by a final 7min extension at 72°C. PCR product was then cleaned using the AmpPure method (Section 2.7.1) and quantified using the Bioanalyser (Section 2.4.3). These products were sequenced on a 454-FLX (Roche Diagnostics, Basel, Switzerland).

3.2.3.2 Data Analysis

Resulting sequences were analysed using a Perl script. The primers were removed and reads of less than 200bp were eliminated from the analysis. The reads were then grouped in Excel to identify recurrent sequences. Reads with less than 100 copies were discarded from the analysis. The remaining reads were then visualised in Jalview (Waterhouse *et al* 2009) to identify differences between the sequences.

3.2.4 Illumina MiSeq

3.2.4.1 In House Assay

3.2.4.1.1 Sample Preparation

1ul of cDNA from 10 patients with CLL and 2 non-CLL patients was amplified for each Vh family individually using the MiSeq Leader and JH primers (Appendix 1) in a 25ul mix containing 5ul Hotstart HiFidelity PCR Buffer and 0.1ul HiFidelity polymerase (202602, Qiagen), 1ul of forward and reverse primer (10uM of each primer). The amplification conditions consisted of an initial denaturation step of 95°C for 5mins followed by 35 cycles of 94°C for 15secs, 57°C for 1min and 72°C for 2mins, followed by a final 10min extension at 72°C. PCR product was then cleaned using AmpPure (Section 2.7.1).The product was then quantified using the Qubit (Section 2.4.2) and 2ul of each Vh product pooled for each sample giving 12 pooled samples. These were tagged using the Nextera protocol (Section 2.7.2), then cleaned again using the AmpPure method. The cleaned products were then quantified using the KapaQuant kit (Section 2.7.3). From this a 10pMol library with 10% PhiX was prepared and loaded onto the Illumina MiSeq (Section 2.7.4)

3.2.4.1.2 Data Analysis

The FASTA files generated from the run were paired together using FLASH (Magoč and Salzberg 2011) and the primer sequences removed using Cutadapt (Martin 2011). Files were then uploaded to Galaxy

(https://usegalaxy.org/) and collapsed to count recurring reads. The collapsed reads were then exported as a FASTA file which was uploaded to IMGT HighV-Quest (Alamyar *et al* 2012) online analysis tools to enable identification of the IGH rearrangement. This analysis produces a compressed folder comprising of several txt files. The 1_Summary file was then analysed using Excel to generate a csv file. Bubble plots were generated with the csv file using the ggplot2 package (Wickham 2009) in R-Studio (RStudio Team 2015).

3.2.4.2 Invivoscribe LymphoTrack IGH Somatic Hypermutation Assay

Samples were amplified using the LymphoTrack IGH Somatic Hypermutation Assay, loaded onto the MiSeq and analysed using the LymphoTrack IGHV SHM Software (Section 2.7.5).

3.3 Results

3.3.1 454 Analysis

3.3.1.1 Concordance with Sanger sequencing

To determine whether NGS would be appropriate for investigating the IgHV region, DNA from five patients was analysed using NGS on the 454 platform and compared to the gold-standard method of SSeq. Comparison of the dominant clone identified by NGS in the diagnostic sample showed that the sequence was identical to the one determined by SSeq (Table 3.1).

3.3.1.2 Sub-clonal Analysis

Sub-clonal analysis was performed on four of the five patients. Since this analysis only used the VH-family specific primer the subclones all contained the same V-, J- and D-genes (known as clonotype) as the major clone. Subclones with reads of over 100 copies in at least one sample (diagnosis, post-treatment or relapse) were visually assessed in Jalview (Waterhouse *et al* 2009) for differences. For CLL044 and CLL063, six subclones were identified, for CLL056 seven subclones were detected and for CLL003 21 subclones were found (Figure 3.1 Figure 3.2, Table 3.2). Of the 36 sub-clones identified over the four samples, 33 were caused by indels (92%). Two of the remaining three sub-clones were the result of transition mutations, a C to T mutation in CLL056 and a G to A mutation in CLL003. The final sub-clone, in CLL003, had both a transition and an indel mutation.

Table 3.1. Comparison summary of SSeq and the dominant clone identified using NGS from the 454. Sequences obtained from both techniques were analysed using the IMGT V-Quest tool to determine V-gene, V-gene identity to germline, J-gene, J-gene identity to germline and D-gene.

Sample	Method	V-gene	V-gene Identity to Germline	J-gene	J-gene Identity to Germline	D-gene
CLL003	SSeq	IGHV1-69*01	219/219 nt	IGHJ6*02	43/47 nt	IGHD3-3*01
	NGS	IGHV1-69*01	219/219 nt	IGHJ6*02	43/47 nt	IGHD3-3*01
CLL044	SSeq	IGHV3-30*03	206/223 nt	IGHJ4*02	23/31 nt	IGHD6-13*01
	NGS	IGHV3-30*03	206/223 nt	IGHJ4*02	23/31 nt	IGHD6-13*01
CLL056	SSeq	IGHV3-48*03	221/223 nt	IGHJ4*02	26/31 nt	IGHD6-6*01
	NGS	IGHV3-48*03	221/223 nt	IGHJ4*02	26/31 nt	IGHD6-6*01
L13	SSeq	IGHV3-72	220/224 nt	IGHJ6*02	40/46 nt	IGHD2-2*02
	NGS	IGHV3-72	220/224 nt	IGHJ6*02	40/46 nt	IGHD2-2*02
CLL063	SSeq	IGHV3-21*01	218/223 nt	IGHJ6*02	31/36 nt	IGHD2-15*01
	NGS	IGHV3-21*01	218/223 nt	IGHJ6*02	31/36 nt	IGHD2-15*01

3.3.1.3 Clones remain stable over time

Sequential samples taken at post-treatment and subsequent relapse was available on four patients. To determine the clone stability over time DNA from these samples was analysed on the 454 platform. The technical results of this analysis are summarised in Table 3.2. Analysis of the sequential data showed that all sub-clones were present at all the time-points. For three of the four samples, CLL044, CLL063 and CLL056, the

clones remained stable over time, with the major clone making up at least 90% of the reads in each sample (Figure 3.1). In CLL056, the most dominant subclone occupied a maximum of 1.3% of the reads with the remaining subclones each having less than 1% individually and collectively they represented 1.7%, 1.4% and 1.7% of the analysed reads at diagnosis, post-treatment and relapse, respectively. The subclones in CLL063 were a greater proportion of the analysed reads with a combined average of 5.8%. As with CLL056, there was a more dominant subclone which in this case, appeared to be expanding at relapse. CLL044 was the only M sample in this cohort. For this patient, the most dominant subclone occupied an average of 5.9% of the reads, although this clone appeared to be marginally growing out at relapse. The other subclones detected in this patient are consistent at diagnosis and relapse, collectively representing 1.3% of the reads in both cases, although there was a slight expansion in these clones post-treatment to around 4%.

Table 3.2 Technical result summary of sequential samples on the 454 platform

		CLL003	CLL063	CLL056	CLL044
No. of reads over 250bp / total number of reads % reads used in analysis	Diagnosis	38659 / 43669 88%	28624 / 41262 69%	34571 / 41098 84%	16089 / 27556 58%
	Post-treatment	34702 / 44030 79%	27532 / 35188 78%	34851 / 38018 92%	28329 / 33076 84%
	Relapse	46815 / 55045 85%	25867 / 34941 74%	31915 / 39370 81%	13156 / 27556 48%
No. of recurrent clones in over 100 reads		21	6	7	6

a. CLL044

Clone (position and variant)	Diagnosis	Post Treatment	Relapse
Dominant	10275	22070	7735
253 +G	15	324	1
258 +A, 279 +G and 285 +T	14	303	1
279 +G and 285 +T	95	280	85
285 +T	19	108	25
296 +G	576	842	765

b. CLL063

Clone (position and variant)	Diagnosis	Post-treatment	Relapse
Dominant	20093	19351	17903
49 +G	186	72	179
49 +G and 70 +G	281	129	164
70 +G	181	213	134
101 -GTA	264	212	191
235 +G	368	311	624

c. CLL056

Clone (position and variant)	Diagnosis	Post Treatment	Relapse
Dominant	25353	30941	24027
69 +G	174	75	200
254 +T	128	153	31
265 +G	292	187	330
270 - G	37	101	35
277 +T	8	120	3
284 C-T	104	6	96

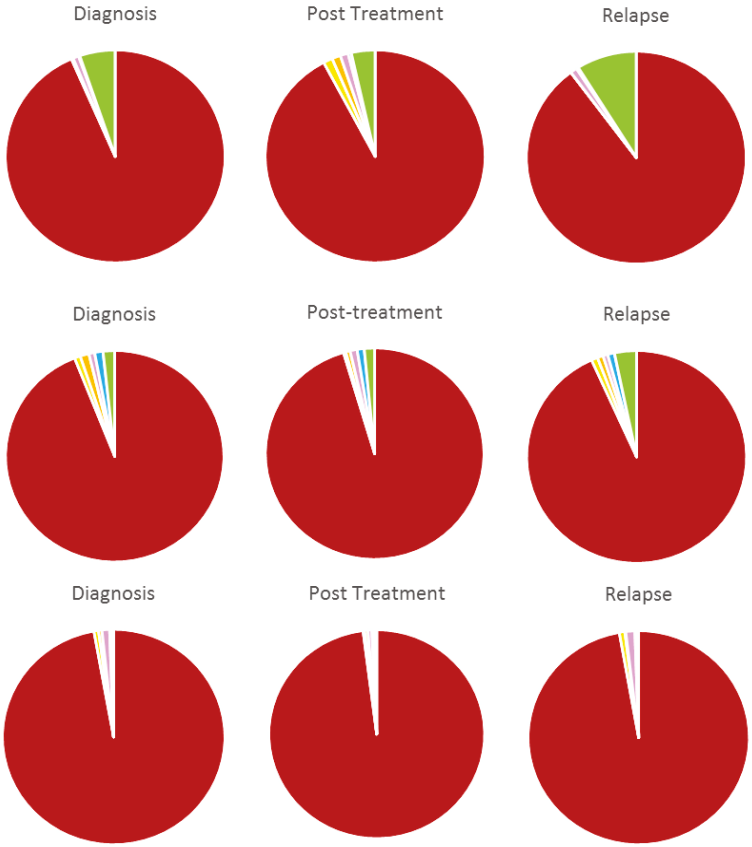


Figure 3.1 The different subclones detected in a.CLL044, b.CLL063 and c.CLL056 over the 3 time-points

Clone (position and variant)	Diagnosis	Post Treatment	Relapse
Dominant	15043	17178	24602
38 +A	223	109	359
38 + A and 51 + A	1	31	120
56 +C	6	11	375
56 +C and 256 -T	1	1	214
100 +T	31	7	158
165 G-A	6	3943	33
165 G-A and 257 +T	0	123	0
231 +T and 250 +T	10	145	22
231 +T, 250 +T and 256 -T	196	186	97
231 +T, 236 -T, 250 +T and 256 -T	23	26	107
231 +T, 236 -T, 246 -T, 250 +T and 256 -T	112	148	201
231 +T, 236 -T, 246 -T, 250 +T, 256 -T and 322 +G	7	4	285
236 -T and 256 -T	1	4	127
250 +T	23	396	60
250 +T and 256 -T	662	105	863
256 -T	10853	113	4158
257 +T	0	264	33
293 -TAC	152	107	125
308 +G	156	315	362
314 +G	1	68	105

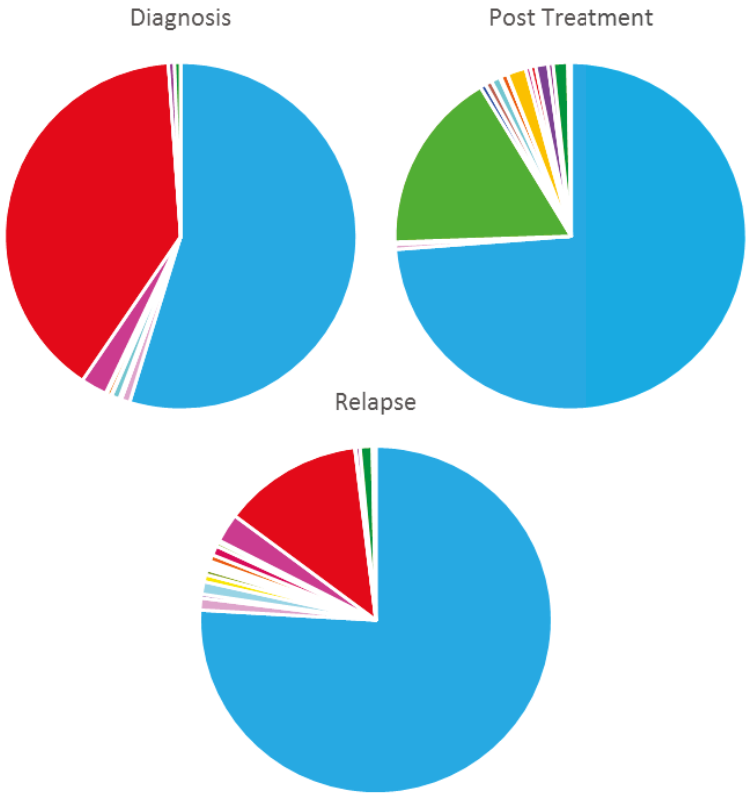


Figure 3.2 The different subclones detected in CLL003 over the 3 time-point

One patient, CLL003, however, had fluctuating levels of sub-clones as illustrated by the IgHV region (Figure 3.2). At diagnosis, the major clone occupied just 55% of the total reads. This clone expanded following treatment and at relapse to about 75% of the reads. The most dominant subclone at diagnosis represented 39% of the total reads. However, this dramatically decreased post-treatment with an 80-fold reduction to just 0.49%. At relapse, this clone had begun to expand again having increased to 13% of the reads.

3.3.2 MiSeq Analysis using an in-house assay

3.3.2.1 Concordance with Sanger sequencing

To determine whether the MiSeq NGS platform would also be appropriate for investigating the IgHV locus, MiSeq analysis of the IgHV region was performed on cDNA from 10 patients and compared to SSeq. As with the 454 platform, all rearrangements found by SSeq were also detectable by MiSeq NGS (Table 3.3). The average number of reads obtained for the dominant clone was 170,761 (range 3,342-661,730).

Table 3.3. Concordance between MiSeq and SSeq for 10 patients. In each case the sequences obtained from the SSeq and the MiSeq conveyed identical VH gene and the % homology to the germline (as determined using the IMGT/ Vquest tool). UM unmutated, M mutated

	MiSeq		Sanger		
Pt No	**VH Gene**	**% homology**	**VH Gene**	**% homology**	**IgHV status**
CLL129-Abs	V1-69	100	V1-69	100	UM
CLL130_Abs	V1-69	100	V1-69	100	UM
L43_Abs	V1-69	99.65	V1-69	99	UM
L49_abs	V3-23	98.26	V3-23	98.26	UM
L74_Abs	V3-23	97.22	V3-23	97.22	M
L76_abs	V3-23	96.18	V3-23	96.2	M
RA_Abs	V3-48	96.53	V3-48	96	UM
L48_Abs	V3-13	96.14	V3-13	95.5	M
TM_Abs	V3-9	93.06	V3-9	93.4	UM

3.3.2.2 Sub-clonal Analysis

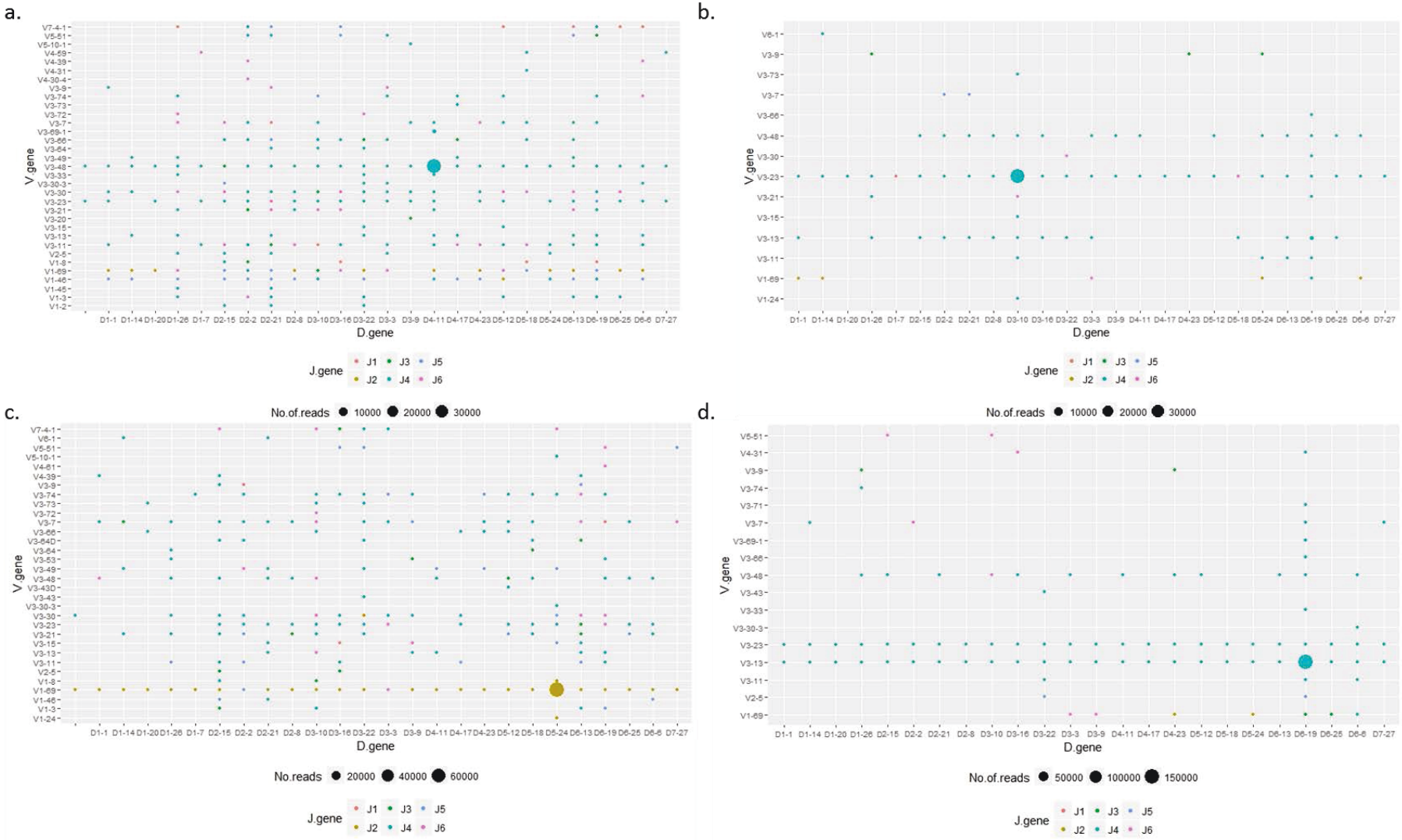
The work on the 454 had shown that frequency of subclones of the same clonotype as the major clone remained stable over time suggesting that VDJ recombination was either an early event in the leukogenesis of CLL or preceded the development of CLL. To determine whether VDJ recombination was an early event, different clonotype subclones other than the one detected in the major clone were analysed. Since the 454 study showed that a VDJ rearrangement could have several subclones which differed by a few bases, recurrent reads were further grouped together, following the IGMT/HighV-Quest analysis, if they were deemed identical through possession of the same clonotype (VDJ rearrangememt), regardless of SHM or productivity. Subsequently, the major clone (as defined by SSeq) was found to represent a median of 90.0% (range 60.7-98.8) of the total reads (Figure 3.3) and in five of the nine samples this

accounted for over 90% of the total reads compared to the minor clones which made up less than 2% (Table 3.4).

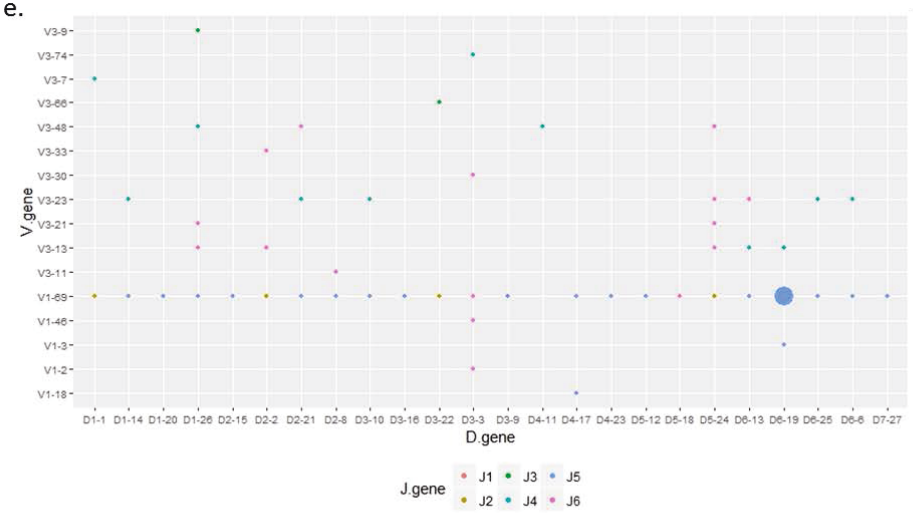
In four samples, the major clone comprised less than 90% of the reads. For three of these samples, the minor clones represented an average 18.8% (range 10.1-39.3) of the total reads (Figure 3.3f-h). Further analysis of the subclones identified in TM found the 2nd most frequent read, a V3-13, J4, D6-19 clonotype, was detected in 1% of the total reads. A V3-13, J4, D6-19 clonotype was also found in CLL130 composed of 1.4% of the total reads and L74 was found to have a V3-21, J6, D5-24 clonotype that comprised 1.2% of the total reads.

Table 3.4. Summary of Subclone Analysis from the Miseq. Fasta files were analysed using the IMGT HighVquest online tool to identify the V,D, and J genes as well as the productivity of the rearrangement. To determine the number and % of reads representing the dominant clones reads that had the same VDJ rearrangement as the dominant clone regardless of SHM or productivity were grouped together. Minor clones were defined as any read that differed by at least 1 gene (V, D or J). The number of copies of the minor clone was defined as the number of unique reads that differed from the VDJ rearrangement found in the dominant clone

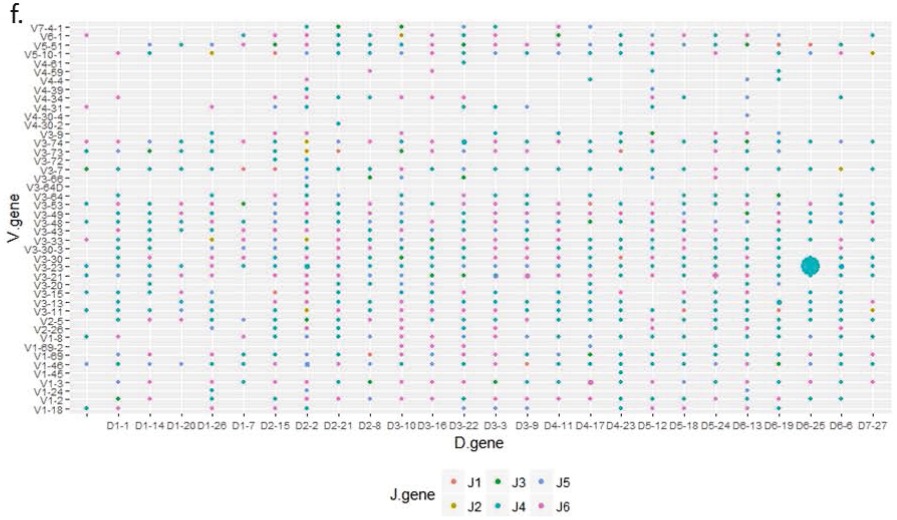
Sample	Total No. of Reads	No. of Unique Reads	No. of Reads Representing Dominant Clone	% Reads Representing Dominant Clone	No. of Reads Representing Minor Clones	% Reads Representing Minor Clones	No. of Reads Representing Productive Minor Clones	Range in the No. of Copies of the Minor Clones (min to max)
CLL129	33704	16137	32604	96.7	294	0.9	211	1-8
CLL130	3805	254	3342	87.8	401	10.5	305	1-24
L43	191914	65976	181142	94.4	2283	1.2	1519	1-87
L48	669631	192511	661730	98.8	746	0.1	346	1-11
L49	154352	64813	138856	90.0	1695	1.1	1283	1-99
L74	104359	55939	77070	73.9	15921	15.3	13711	1-399
L76	261387	143129	158635	60.7	102752	39.3	65928	1-9005
RA	190607	77256	184565	96.8	3083	1.6	2395	1-142
TM	9830	7088	8091	82.3	996	10.1	773	1-11


a.
b.
c.
d.
V.gene
D.gene
J.gene
J1
J2
J3
J4
J5
J6
No.of.reads
10000
20000
30000
No.reads
20000
40000
60000
50000
100000
150000

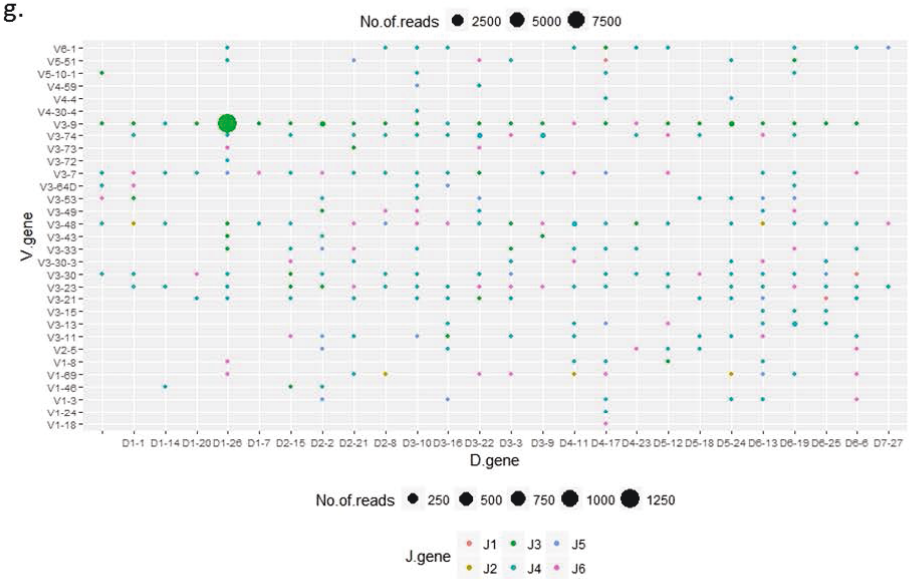
e.


V.gene
V3-9
V3-74
V3-7
V3-66
V3-48
V3-33
V3-30
V3-23
V3-21
V3-13
V3-11
V1-69
V1-46
V1-3
V1-2
V1-18
D1-1 D1-14 D1-20 D1-26 D2-15 D2-2 D2-21 D2-8 D3-10 D3-16 D3-22 D3-3 D3-9 D4-11 D4-17 D4-23 D5-12 D5-18 D5-24 D6-13 D6-19 D6-25 D6-6 D7-27
D.gene
J.gene
J1 J3 J5
J2 J4 J6

f.


V.gene
D.gene
J.gene
J1 J3 J5
J2 J4 J6

g.


No.of.reads
2500 5000 7500
V.gene
D.gene
No.of.reads
250 500 750 1000 1250
J.gene
J1 J3 J5
J2 J4 J6

h.

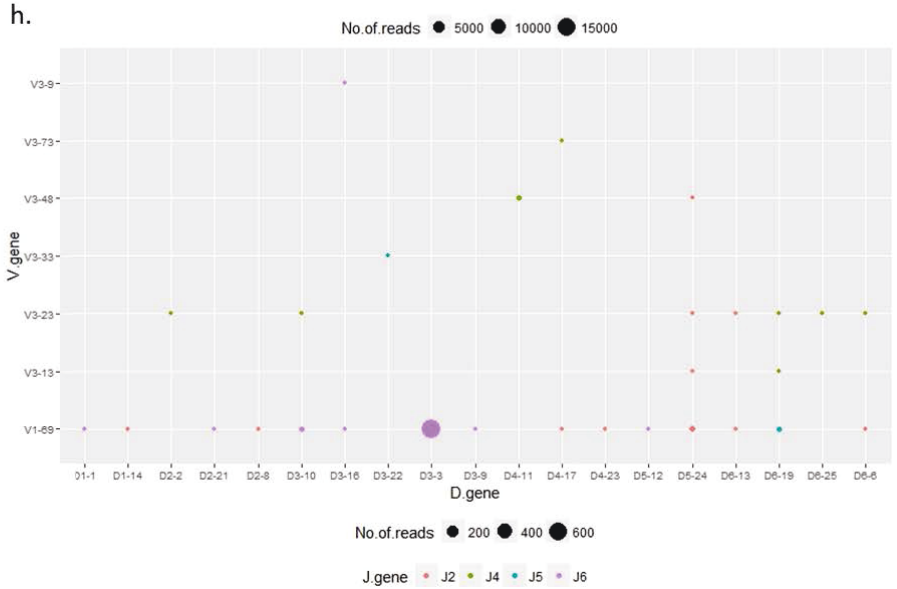

No.of.reads
5000 10000 15000
V.gene
V3-9
V3-73
V3-48
V3-33
V3-23
V3-13
V1-69
D.gene
No.of.reads
200 400 600
J.gene
J2 J4 J5 J6

Figure 3.3. The dominant clone represents an average of 90% of the reads in samples with one major clone. The bubbleplots show the frequency of the different clonotypes found in 8 of the CLL samples. a.RA, b.L49, c. L43, d.L48, e.CLL129, f.L74, g.TM, h. CLL130. The V-gene is plotted along the X-axis, the D-gene along the Y-axis and the J-gene is represented by the colour of the bubble. The size of the bubble reflects the frequency of the clonotype. These 8 samples were all found to be IgHV-monoclonal as the major clone represents an average of 90% of the reads (range 74-99).

In one sample, L76, it was possible to detect a second clonotype comprised of a V1-69, a J6 and a D3-3 with 100% homology to the germline V1-69 gene, indicating that this was an UM subclone. This subclone represented 28% of the total reads for this patient (Figure 3.4a). The major clone identified by SSeq was a M V3-23 gene alongside a J4 and D6-6 rearrangement. The major clone would be correlated with good prognosis, whereas the minor clone would be associated with a poorer outcome. Both clones can be visualised on agarose gel electrophoresis of the IgHV, amplified using the individual VH leader primers (Fais *et al* 1998) (Figure 3.4b).

3.3.2.3 The B-cell repertoire

In each CLL sample, it was possible to observe VDJ clonotypes other than the major clone that had been identified using SSeq. To determine whether these rearrangements represented the normal background B-cell repertoire, cDNA from two aged-matched healthy controls underwent IgHV analysis on the MiSeq using the in-house assay. C1 generated a total of 401,090 reads and C2 produced 17,569 reads. Of these reads 78% and 87%

respectively were unique. In comparison, an average of 42% of the total reads were unique for the nine CLL samples.

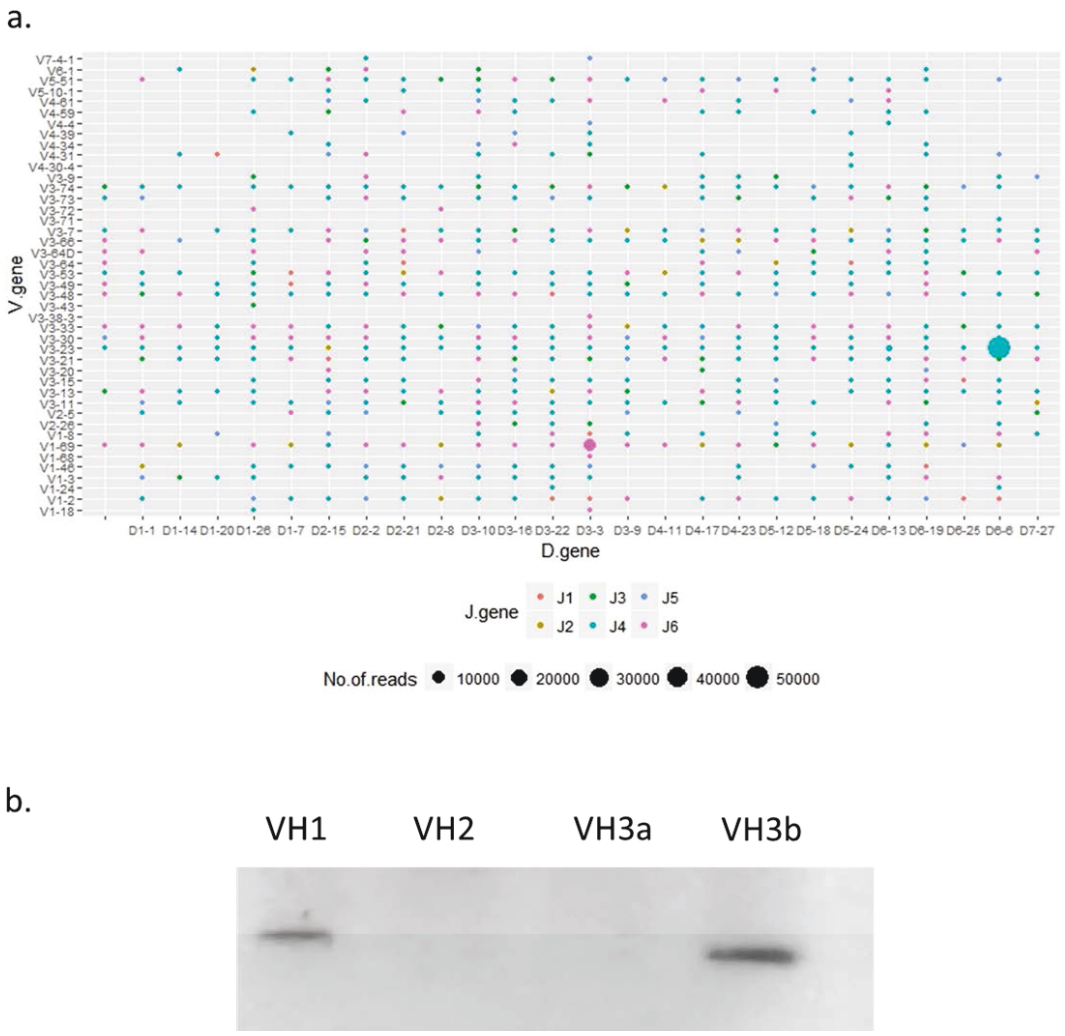


Figure 3.4. Subclonality in L76. a. The bubble-plot shows both the VH3-23, J4, D6-6 clonotype (blue) as well as the V1-69, J6, D3-3 clonotype (pink). b. Electrophoresis on 3% agarose gel of IgH amplified using the specific VH leader primer for VH1, VH2 and VH3. There are 2 leader primers for VH3; VH3a and VH3b.

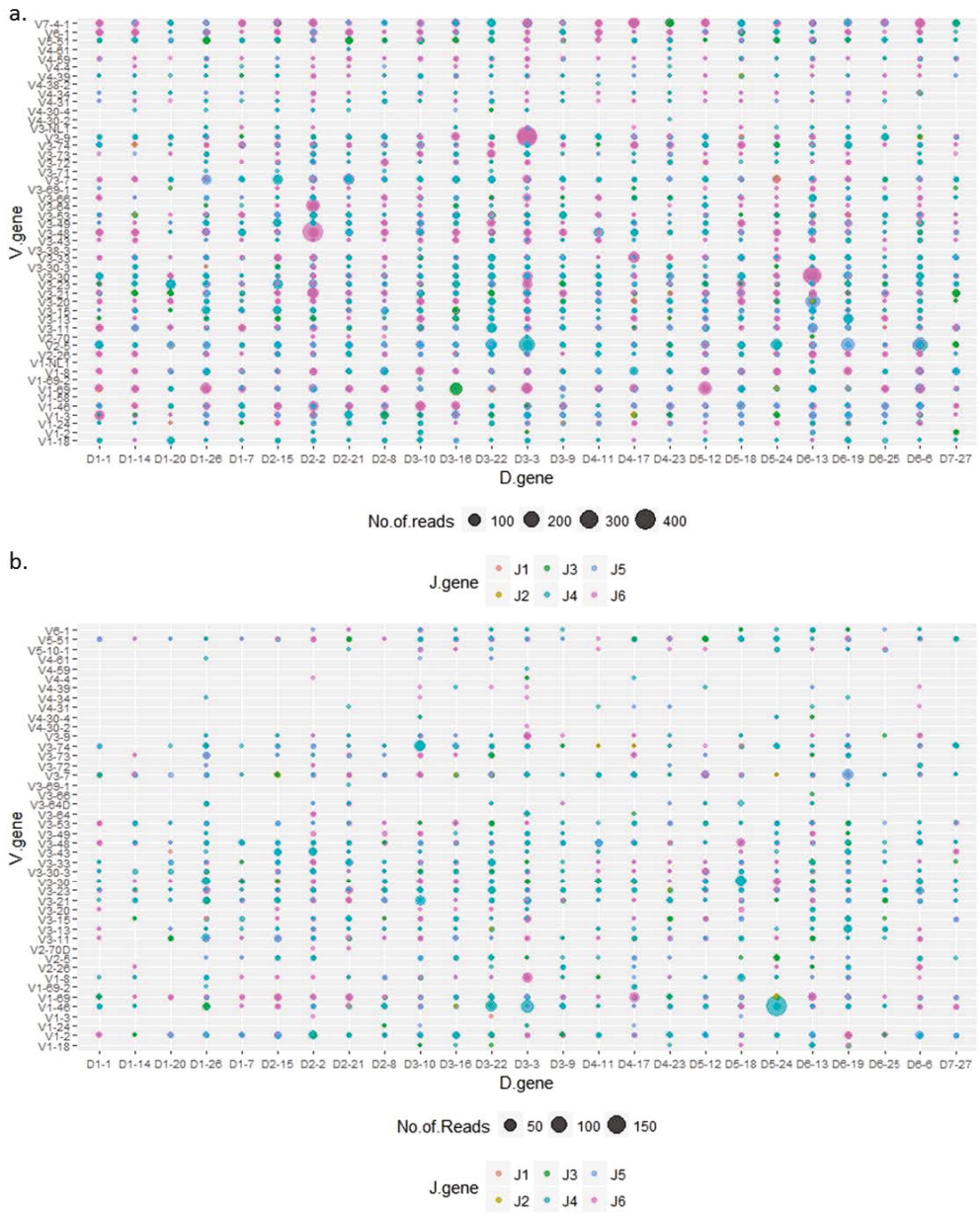


Figure 3.5. Repertoire of the VDJ rearrangement in samples of 2 healthy age-matched controls. Both C1 (a) and C2 (b) show a polyclonal picture.

The frequency of the different VDJ clonotypes in these samples was variable. In C1, the most abundant rearrangements were a V3-9, J6, D3-3 rearrangement, a V3-21, J6, D6-13 rearrangement and a V3-48, J6, D2-2 rearrangement (Figure 3.5a). These represented 1.7%, 1.0% and 0.99% of the total reads, respectively. There appeared to be only one dominantly

recurring clonotype in C2 which was composed of a V1-46, J4, D5-24 rearrangement that comprised 3.1% of the total reads (Figure 3.5b).

3.3.3 MiSeq Analysis using the Invivoscribe Lymphotrack SHM Kit

The Invivoscribe Lymphotrack SHM kit is a commercially available kit that uses a single multiplex PCR mastermix. It therefore has a shorter processing pipeline than the in-house methodology as it does not require the Nextera tagging step or the Kapa quantification step. Furthermore, the kit is supplied with a built-in analysis software which enables rapid analysis of the NGS data obtained from the MiSeq.

3.3.3.1 Concordance with Sanger sequencing

To determine whether the Invivoscribe Lymphotrack SHM kit was suitable for NGS IgHV analysis, it was used on cDNA from 270 patients with CLL and compared to SSeq. As with the both the 454 platform and the in-house MiSeq assay, all rearrangements found by SSeq were also detectable using the Invivoscribe kit and bioinformatics pipeline. Previous analysis using SSeq had identified that 50.0% (135/270) of patients were IgHV M, 38.5% (104/270) were UM and 11.5% (31/270) were believed to be polyclonal as SSeq was inconclusive due to the appearance of more than one sequencing read. In comparison, NGS could identify productive clonotypes in all patients. For samples with conclusive SSeq results, the concordance

between SSeq and NGS for the major clone was 98.3% (235/239); 87.9% (210/239) had identical results and 10.5% (25/239) had the same clonotype however the percentage homology to the germ-line differed slightly with a median error of 0.6% (range, 0.2 -1.6). For the remaining 1.7% (4/239), SSeq detected another clone which was less abundant by NGS, but still detectable.

3.3.3.2 Sub-clonal Analysis

The subclonal analysis using the Invivoscribe kit focussed solely on the first 200 most abundant reads (MARs). This was primarily because these were the sequences presented by the Lymphotrack software but also because the analysis of the healthy age-matched controls in section 1.3.2.3 showed that normal B-cells could comprise up to 3.1% of the total reads and the 200 MARs accounted for an average of 24% of the total reads (range 5-50.4), therefore suggesting that any low level clonotypes represented CLL subclones as opposed to the normal B-cell background.

Analysis of the 200 MARs found that 76.3% (206/270) had only one clonotype. Interestingly, in addition to the 11.5% (31/270) patients found to be polyclonal by SSeq analysis, a further 33 patients were found to have more than one productive VDJ rearrangement. This meant that 23.7% (64/270) of this patient cohort had multiple productive clonotypes. To verify the presence of these subclones, five patients with subclones of

different Vh families underwent IgHV PCR amplification using individual and specific VH family primer. For each patient, the NGS-predicted VH family amplification was always observed even when the subclone was present at a low frequency (<5% of the MARs) (Figure 3.6). This suggested that the subclones identified in this patient cohort were not a bioinformatic artefact.

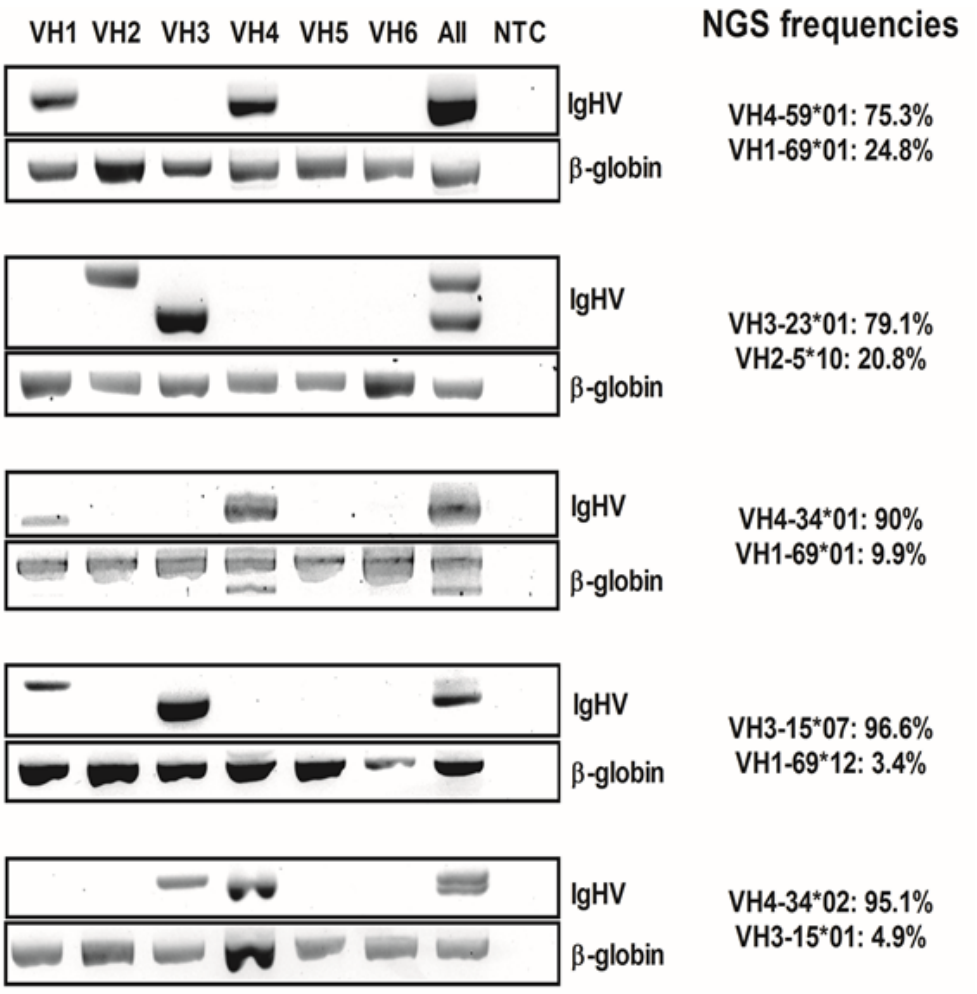


Figure 3.6. NGS can detect multiple specific VH family rearrangements. Electrophoresis on 2% agarose gel of IgHV amplified using specific VH leader primer for 5 subclonal representative patients. NGS clone frequencies are also provided. Even low frequency clones (<5%) are confirmed by electrophoresis. Beta globin was used as endogenous control. All: multiplex including all VH primers as positive control. NTC: no template control.

3.4 Discussion

The IGH locus undergoes VDJ rearrangement during B-cell development to produce a rearrangement that is unique to the cell. In mature B-cell neoplasms, such as CLL, the locus therefore provides a marker of clonality. Previous studies have found the level of SHM that the rearrangement has undergone to be an independent marker of prognosis (Damle *et al* 1999, T J Hamblin *et al* 1999). Additionally, it is well known that there is a bias amongst the VDJ rearrangements found in patients with CLL as stereotyping can be demonstrated in over a third of patients in their CDR3 (Section 1.2) (Agathangelidis *et al* 2012, 2014). Correlation of this stereotyping to clinical data has enabled recurrent subsets which share clinical and biological features to be identified (Bomben *et al* 2009, Rossi and Gaidano 2010). This strongly implies that CLL may be antigen driven and, therefore, further analysis of the IgHV locus may elucidate the pathogenesis of CLL and potentially be beneficial in determining patient outcome. The aim of the study described in this chapter was to determine if NGS could be utilised to interrogate the IgHV locus, particularly for detecting subclones. It further sought to identify the best platform and method for NGS interrogation of the IgHV locus.

3.4.1 Use of the two NGS Platforms

The amplified size of the rearranged VDJ can vary depending on the components of the rearrangement and the primers used to amplify the

product. Both DNA and cDNA amplified using the FR1 primers will produce a fragment of 310-360bp (van Dongen *et al* 2003), whereas use of the leader primers will generate a 410-490 fragment for cDNA and a 500-570bp product for DNA (Szankasi and Bahler 2010). Leader primers tend to be the first option for IgHV analysis as the whole of the gene is amplified, enabling total interrogation of the locus (Ghia *et al* 2007). These primers, however, can be problematic as the larger fragment can cause reduced PCR efficiency, particularly when using DNA. This may also present a challenge for certain NGS platforms. The FR1 primers, in contrast, are more robust but as they are located within the IgHV gene it is not possible to obtain the full sequence; therefore, percent homology to the germline is estimated (Langerak 2009). Both the FR1 and the leader primers were used in this study.

Two NGS platforms were utilised; the Roche 454-FLX and the Illumina MiSeq. The 454-FLX was the original instrument of choice as it was the only platform at the time that was capable of analysing reads of up 400bp. The results obtained from the 454 were consistent with the SSeq demonstrating that it would have been suitable for IgHV analysis. However, whilst the 454 was capable of both forward and reverse sequencing it did not, at that time, perform paired-end sequencing and therefore, there was no confirmation of the individual sequences. This can be beneficial particularly at the end of the reads when the reagents are nearing

exhaustion and the quality of the data decreases, especially for this analysis whereby single base differences may represent a true subclone or a sequencing artefact.

Developments in the Illumina Miseq chemistry meant that it became possible to perform 300-cycle pair-end sequencing enabling 600bp sequencing. The added advantage of this system was that as the sequencing was paired-end, the forward reaction would sequence one-way then the reverse would start at the other end of the fragment and sequence back towards the start. This meant that a fragment of 500bp would have an overlap of 100bp in the middle, enabling confirmation of base changes in this region of the sequencing. Since the overlapping region was the end of the fragment to be sequenced in both directions, this was the area most prone to sequencing errors caused by the exhaustion of the reagents. Double confirmation of base-changes in the region provided assurance as to its reliability.

Irrespective of the advantages associated with either platform, results obtained from both the Roche 454 and the Illumina MiSeq were consistent with the traditional gold-standard method of SSeq, showing that either was suitable for this type of analysis.

3.4.2 Comparison of methodologies

The methods used in this study included two in-house designed assays and one commercial kit. The kit was released after the development of the two in-house assays and offered a more rapid sequencing procedure as well as having the additional benefit of an analysis software. This was a significant advantage since the 454 methodology relied heavily on bioinformatic support which was not readily available at the time of the study, whilst the bioinformatics analysis of the in-house MiSeq assay used multiple command-line software tools that were not designed specifically for that type of investigation, resulting in a time-consuming bioinformatic analysis. In comparison, the software provided with the Invivoscribe kit was easy and relatively quick to use. Furthermore, the software produced multiple graphics to aid in the interpretation of the data.

The Invivoscribe kit also offered a more rapid PCR processing pipeline in comparison to the in-house MiSeq assay. Whilst the method used for the in-house MiSeq assay controlled for PCR bias through the amplification of each VH-family independently, it also created a more timely and complicated process as each sample had eight PCR mixes (one for each VH-family and a NTC). This meant that it was difficult to process large numbers of samples, limiting the potential usefulness of this assay.

Therefore, the Invivoscribe Lymphotrack Kit was chosen as the most suitable assay due to its simple processing and user friendly bioinformatics.

3.4.3 Subclonal Analysis

3.4.3.1 454

Using the VH family specific primers, it was possible to identify subclones in all the samples using the 454 platform. Reads of less than 250bp and recurrent reads under 100 copies, unless present in sample from a different time-point, were excluded from this analysis. Reads of less than 250bp were discounted as they would not enable complete examination of the IgHV region as well as the possibility that they were the result of sequenced primer dimerisation. Reads with less than 100 copies were also omitted on the assumption that they may be the result of polymerase errors.

NGS sequencing analysis of sub-clonal populations can present some problems in regards to what represents a true sequencing variant and what is artefact generated during the initial amplification stage or during sequencing. The Phusion Taq used for the 454 sequencing has a reported error rate of $5.5x10^{-7}$ mutations/bp amplified (Li *et al* 2006), although McInerney *et al* (2014) found the error rate to be $2.6x10^{-6}$. In their study, McInerney *et al* found both transition and indel mutation polymerase

errors when using the Phusion Taq but no transversions. This is comparable with the subclones identified in this study, so it is possible that the differences are the result of polymerase errors in the initial amplification. However, using a 30 cycle PCR to generate a 300bp product would mean 0.4% of reads would have an error. This study produced an average of 38400 reads per sample equating to just 154 reads with polymerase mutations.

Most of the subclones detected were caused by indel mutations (92%). The pyrosequencing chemistry utilised by the 454 is prone to indel errors in homopolymer tracts (Wang *et al* 2007). Boyd *et al* (2009), Bashford-Rogers *et al* (2013) and Campbell *et al* (2008) all report errors to be at a higher rate than other types of sequencing error and that they are inserted at either the beginning or the end of short (>4bp) homopolymer runs (Campbell *et al*). A further consequence of these particular artefacts is the erroneous insertion of the homopolymer nucleotide at its next genuine occurrence following the tract. To overcome this problem, they discounted indels from their analysis, suggesting that the resulting frameshift would produce a non-functioning protein. It is possible that the indels detected in the subclones in this chapter are the result of sequencing artefacts. However, these subclones were reproducible as they were recurrent in all three samples from the same patient, albeit at differing proportions in some instances.

Of the four patients who had sequential samples, three demonstrated relatively constant subclonal frequencies over the three time-points. CLL003 however had fluctuating sub-clonal fractions. This patient had a rapidly progressing disease and died within three years of diagnosis. Therefore, this may account for the change in the clones over time. Furthermore, this patient subsequently went on to have whole genome sequencing on sequential samples which showed a very similar picture (Schuh *et al* 2012). Therefore, it is possible that the IgHV analysis may represent a surrogate measurement for the subclonal dynamics identified in this patient.

3.4.3.2 MiSeq

Both the in-house methodology and the Invivoscribe Lymphotrack kit used the leader primers to amplify cDNA to enable detection of the whole of the IgHV region that was expressed by the CLL cells. Focussing the analysis solely on expressed IgHV rearrangements prevented non-functioning unproductive rearrangements from inappropriately skewing the data.

The in-house methodology used all the VH family primers together, as opposed to the family-specific primer used in the 454 study. This approach enabled detection of a second VDJ rearrangement in one patient. In this

instance the major clone defined by SSeq identified a M rearrangement which would conventionally be associated with good prognosis. However, NGS further identified an UM minor clone. Bi-clonal productive rearrangements of the IgHV locus have previously been reported (Kern *et al* 2014, Plevova *et al* 2014) with one study reporting that patients carrying rearrangements of discordant mutational status had shorter TFS and OS then patients with two hypermutated rearrangements (Visco *et al* 2013). Current guidelines however recommend that it is not possible to offer a conclusive clinical interpretation of the result (Langerak *et al* 2011).

The release of the Invivoscribe Lymphotrack kit meant that it was possible to perform IgHV-NGS on a larger cohort pf patients. Analysis of 270 patients with CLL found that the number of patients with multiple productive rearrangements was 24%. This is higher than has been found in previous studies where the percentage of CLL patients with more than one IgHV clone has been reported in <5% of CLL cases (Plevova *et al* 2014, Rassenti and Kipps 1997, Visco *et al* 2013). Since these other studies used SSeq, the increased sensitivity of NGS and its ability to amplify and sequence mixtures of DNA templates more reliably compared to SSeq, could provide an explanation for the detection of a larger number of subclonal IgHV patients. However, it is also possible that these subclones could represent the normal polyclonal B-cell background.

3.4.4 Conclusions

The work in this chapter has demonstrated that NGS is, in principle, highly suited to analysis of the IgH, particularly for detection of subclones. Whilst all three methodologies were comparable with the current gold standard of SSeq, the Invivoscribe Lymphotrack Kit was shown to be superior in terms of sample processing and bioinformatic analysis. The discovery that around a quarter of CLL cases have multiple productive VDJ clonotypes provides evidence for a leukaemia initiating event that takes place prior to recombination. However, further work needs to be done to establish the sensitivity and false positive rate for sub-clonal detection using NGS as well as determining the clinical implications of these IgHV subclones within a CLL patient cohort.

Chapter 4 Subclonal IgHV rearrangements are clinically relevant in chronic lymphocytic leukaemia

4.1 Introduction

The work described in the previous chapter found that patients with subclonal IgHV rearrangements occurred at a higher frequency then has been previously described (Plevova *et al* 2014, Rassenti and Kipps 1997, Visco *et al* 2013), yet the clinical relevance of this is unclear. Whilst the identification of subclones that have different productive VDJ clonotypes would support a precursor B-cell as the leukaemic stem cell in CLL, it is also possible that these subclones are the part of the normal poly-clonal B-cell background.

The work described in this chapter aimed to further investigate these IgHV subclones. This was done firstly by analysing the normal B-cell repertoire to establish if the subclones are background normal B-cells. Secondly, it was determined if these clones had any effect on prognosis.

4.2 Materials and Methods

4.2.1 Patient Samples

Population A used cDNA from a cohort of 270 unselected patients (Appendix 2). Population B used gDNA from a cohort of 227 pre-treatment CLL patients enrolled into two UK based FCR clinical trials (appendix 2). As a control, cDNA was extracted from purified B-cells taken from 10 age-matched controls (Table 4.1).

Table 4.1. Summary of healthy age-matched controls

Sample	Age (years)	$CD19^+$ cells (%)
AMCONTROL1	77	92.2
AMCONTROL2	86	92.3
AMCONTROL3	80	94.6
AMCONTROL4	90	97.4
AMCONTROL5	75	96.9
AMCONTROL6	71	92.3
AMCONTROL7	75	98.3
AMCONTROL8	59	97.2
AMCONTROL9	62	97.7
AMCONTROL10	61	97.3

4.2.2 Invivoscribe LymphoTrack IGH Somatic Hypermutation Assay

Samples were amplified using the LymphoTrack IGH Somatic Hypermutation Assay, loaded onto the MiSeq and analysed using the LymphoTrack IGHV SHM Software (Section 2.7.6)

4.2.3 Targeted NGS and SNP Array Analysis

DNA from population B was analysed using SNP arrays (Section 2.8) and a targeted TSCA panel (Section 2.7.6 and Appendix 3).

4.2.4 Statistical Analysis

All analyses were performed with Graphpad Prism 5.0 software or IBM SPSS 13.0 software (Section 2.9)

4.3 Results

4.3.1 Factors effecting the detection of subclones

4.3.1.1 The normal polyclonal background B-cell repertoire

To assess the diversity of the normal B-cell repertoire, cDNA from 10 healthy age-matched volunteers was analysed using the Invivoscribe NGS IGH kit. The analysis was limited to the first 200 most abundant reads (MARs) to mimic the analysis of the CLL samples. Polyclonality was demonstrated in all cases with the healthy controls expressing a median of 46.5 (range 28-53) different V-genes in the 200 MARs. These V-genes were found in combination with different J-genes as is demonstrated through the detection of multiple VJ clonotypes (Figure 4.1). The most common VJ clonotype comprised an average of 25% of the 200 MARs (range 7.6-49.9). Since, typically, most pre-treated CLL peripheral blood samples contain over 90% CLL cells (Ternynck *et al* 1974), polyclonal B-cell contamination would be found in less than 0.4% of the total reads. However, to ensure that subclones were real, CLL samples were only considered as subclonal if the minor clone represented over 2.5% of the 200 MARs.

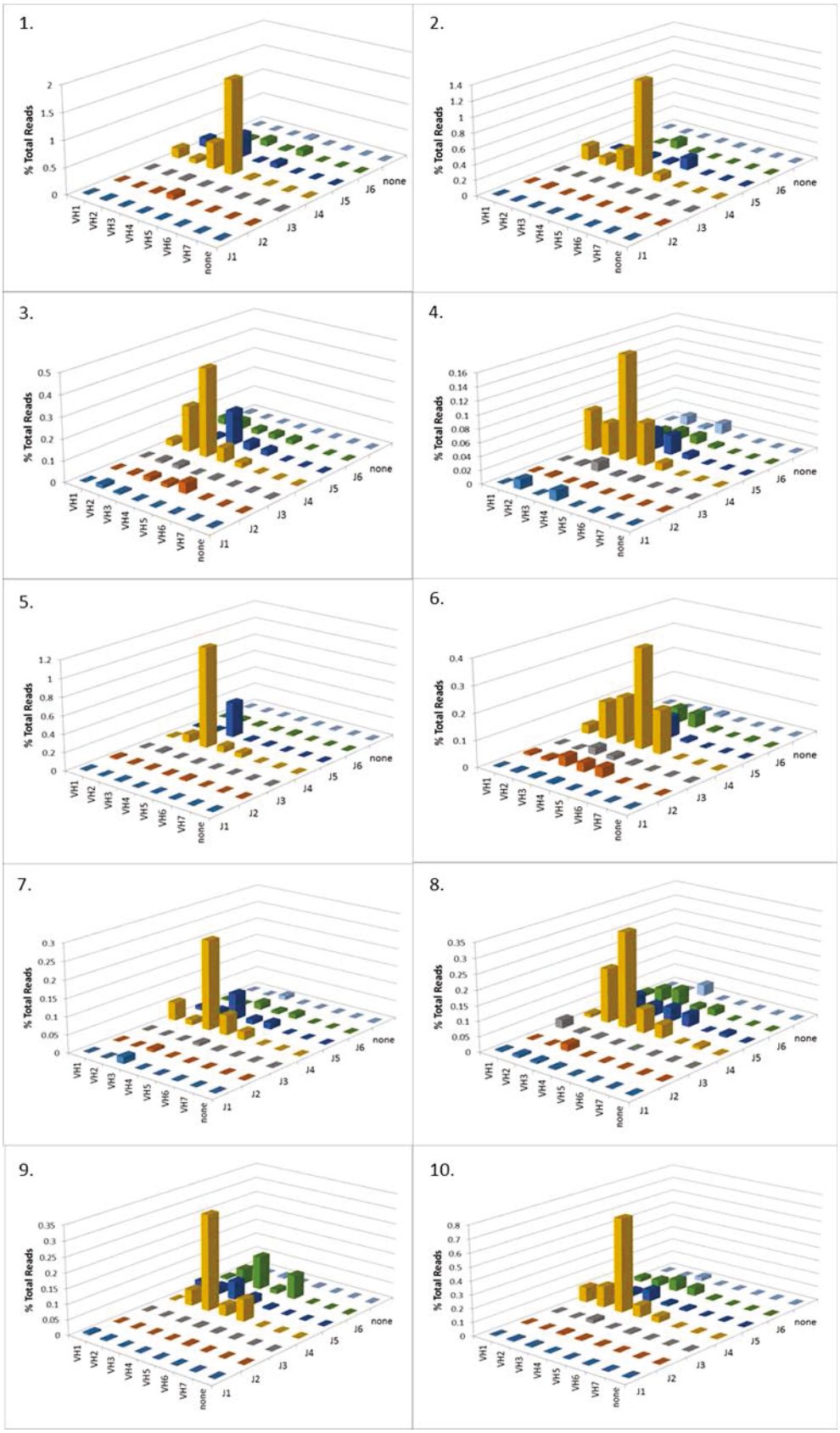


Figure 4.1. NGS IgHV profiles of B cell from healthy donors. Panels 1 to 10 represent the NGS IgHV profile of 10 purified B cells samples from healthy age-matched donors. The graphs are calculated by the lymphtrack software (Invivoscribe).

4.3.1.2 Effect of using cDNA or DNA on IgHV subclonal detection

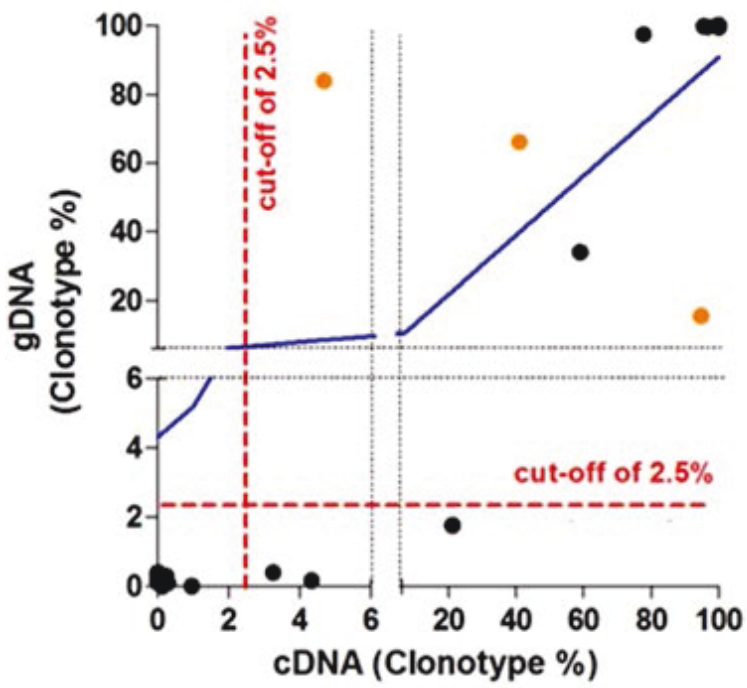


Figure 4.2. There is good concordance between cDNA and gDNA in the detection of subclones although cDNA is superior in detecting low level clones. NGS IgHV analysis of cDNA and gDNA from 8 patients. Black dots are productive rearrangements and orange dots are unproductive rearrangements. R^2=0.70, p<0.0001

To determine the effect of using cDNA versus gDNA, cDNA and gDNA samples from eight patients were analysed in parallel (Figure 4.2). Whilst the data showed a significant correlation (R^2=0.70, p<0.0001) and overall, using the 2.5% cutoff, the concordance between cDNA and gDNA was 87% (20/23), there were some discrepancies. Firstly, three subclones that were detected using cDNA were found below the 2.5% cutoff in gDNA and would therefore been discounted from the analysis. Also, in one instance, two unproductive rearrangements were detected at a lower frequency using cDNA compared to gDNA.

4.3.1.3 Normal B-cell contamination into CLL samples does not introduce false subclones

Since it was demonstrated that cDNA can detect more subclones than gDNA, and it is widely accepted that CLL cells have a decreased expression of Ig in comparison to normal B-cells (Batata and Shen 1992, Ludescher *et al* 1992, Merson and Brochier 1988, Ternynck *et al* 1974), it is possible that by using cDNA the impact of the background B-cells may be artificially elevated and cause the detection of false subclones. To investigate this further, cDNA from purified B-cells taken from three healthy controls was serially diluted into cDNA from one clonal and from one subclonal CLL sample (Figure 4.3). In both incidences, the normal B-cell contamination polyclonal was undetectable in the first 200 MARs even at a 50:50 dilution ratio and with all normal B-cell clonotypes considered. This suggests that the subclones detected in the CLL samples are not the background normal B-cells and provides additional support for the 2.5% cutoff for the detection of subclones.

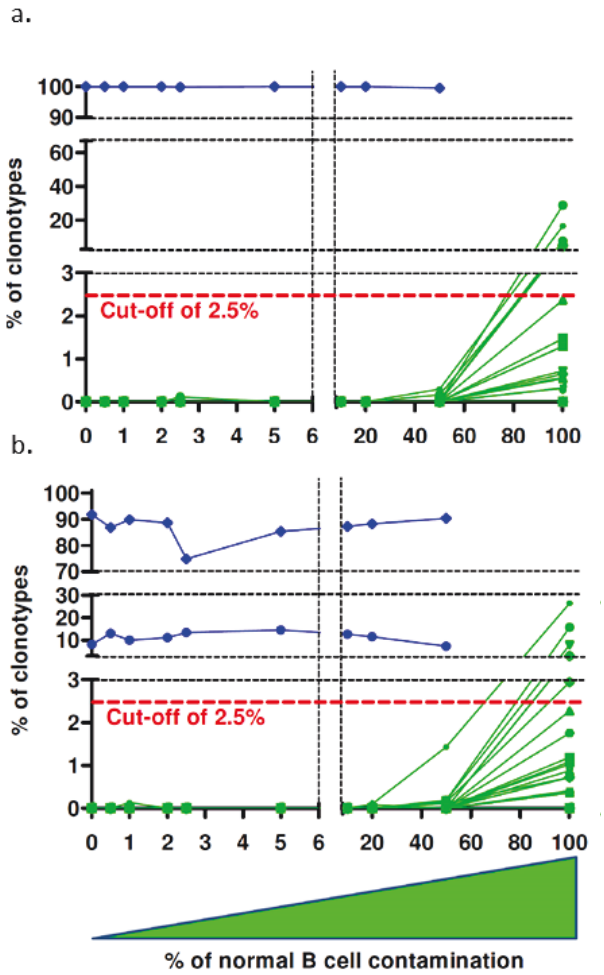


Figure 4.3. Polyclonal background B-cells are undetectable in CLL samples. Monoclonal (a) and biclonal (b) CLL cDNA (blue) was serially contaminated with cDNA from purified normal B cells (green). The frequency of each clonotype was calculated for the 200 most abundant clones. Polyclonal B-cells were not detectable in CLL samples with up to 50% normal B-cell contamination.

4.3.1.4 Effect of multiplex PCR versus singleplex PCR on IgHV subclonal detection

To determine if using a multiplex PCR mix, as provided in the Invivoscribe kit, could influence the detection of small subclones that would possibly not be detected by a normalized PCR approach through amplification bias, each of the seven Vh-leader forward primers were amplified with the

consensus JH reverse primer in individual and separate PCR reactions for six patients with CLL. Following amplification, the seven reactions were normalised, pooled and sequenced by NGS. In parallel, the samples were also analysed using the Invivoscribe kit. 23 different clonotypes were detected in total amongst the six patients and, whilst some clonotypes were detected in the individual reaction but not in the multiplexed reaction and vice versa, these were all below the 2.5% background cut-off. Importantly, the frequency of major clonotypes between individual reaction and multiplex PCR were significantly correlated with a R^2 of 0.9992 (Figure 4.4).

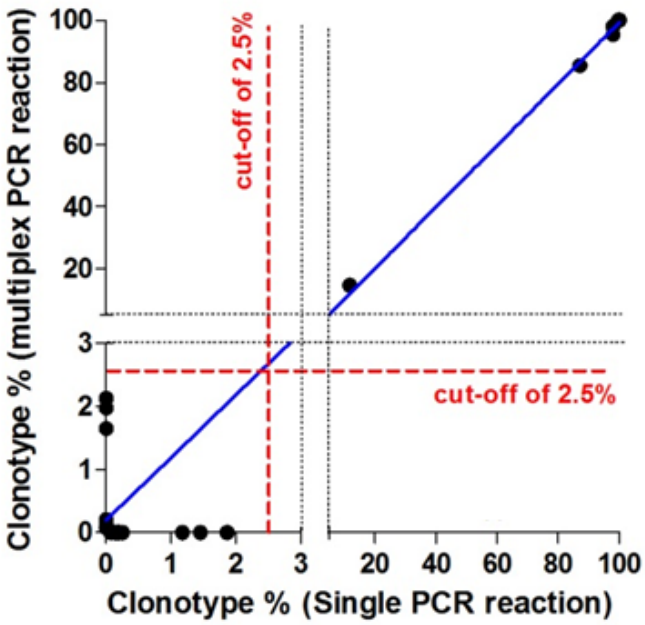


Figure 4.4. There is significant concordance between singleplex and multiplex PCR amplification of the IgHV locus. cDNA from 6 patients was amplified in 7 individual reactions for each of the Vh-family leader primers prior to NGS and compared to results obtained using the Invivoscribe kit. R2= 0.9992, p<0.0001.

4.3.2 NGS identifies five prognostically significant subgroups in an unselected patient cohort (Population A)

NGS analysis of the IgHV locus in a cohort of 270 unselected patients (population A) found that 23.7% (64/270) had multiple productive rearrangements. Of these 64, 65.6% (42/64) had two subclones, 18.8% (12/64) had three subclones and 15.6% (10/64) had four or more subclones. The median frequency of the second, third and fourth most abundant clones was 10.3% (range 2.5-50%), 3.7% (range 2.5-25.2%) and 6.5% (range 2.5-15.5%) respectively.

Since NGS could identify different subclones in some patients, five different categories were created: patients with (i) multiple M clones (MultiM), (ii) 1 M clone, (iii) a mix of M-UM clones, (iv) 1 UM clone (+subset #2 VH3-21), (v) multiple UM (Multi-UM) clones. SSeq had previously identified 135 M patients with a TFS of 131 months, 102 UM patients with a median TFS of 33 months and 31 polyclonal patients with a median TFS of over 280 months. Using the new NGS classification the SSeq-M patients could be further divided into three subgroups with a median TFS of >280 (MultiM), 131 (M), 56 (Mix) months (P=0.0037), the SSeq-UM patients into three subgroups with a median TFS of 94 (Mix), 29 (UM), and 40 (Multi-UM) month (P=0.0049) and the polyclonal patients, previously inconclusive by SSeq, into three subgroups with a median TFS of >280 (Multi-M), 111 (M), and 4 months (Multi-UM) (P=0.0167) (Figure 4.5).

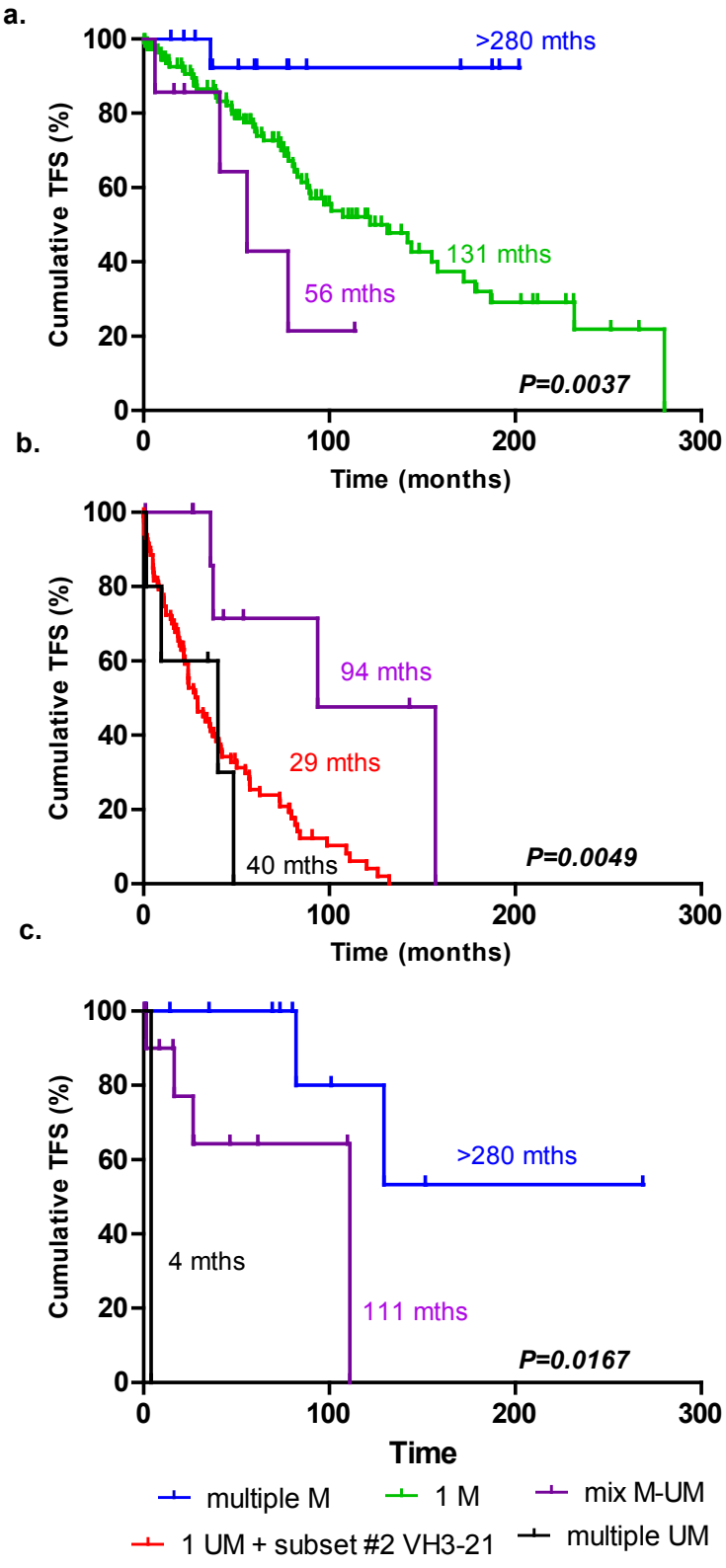


Figure 4.5. NGS-IgHV refines SSeq classification. TFS according to IgHV mutational status determined by SSeq, was plotted with Kaplan-Meier methods for all patients. NGS-IgHV classification was after applied to M patients by SSeq (a), to UM patients by SSeq (b) and to polyclonal patients inconclusive by SSeq (c) in terms of TFS. Difference between curves was assessed with the log rank test and median survival for each subgroup of patients was provided.

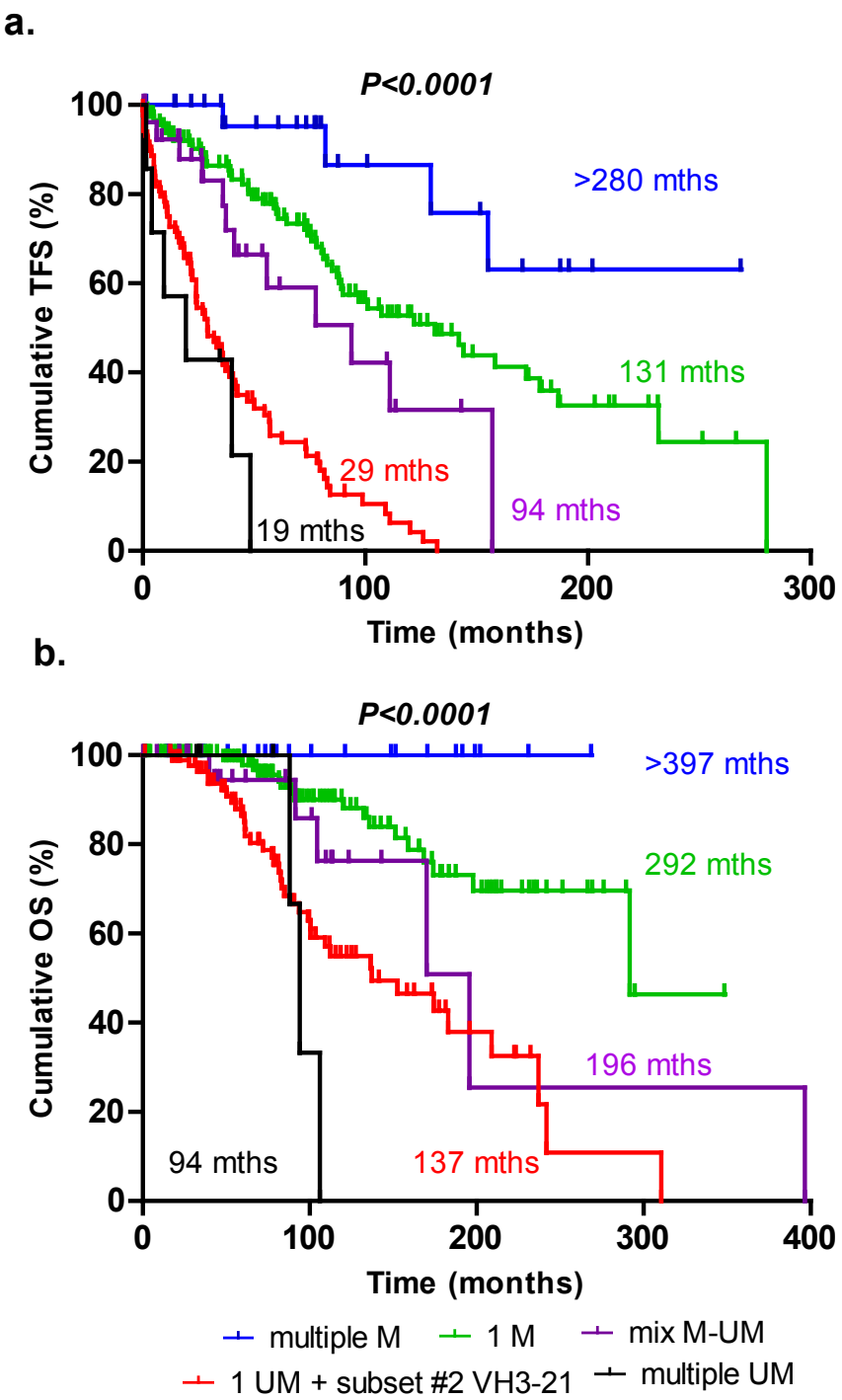


Figure 4.6. NGS-IgHV defines 5 different prognostic subgroups. TFS (a) and OS (b), according to NGS-IgHV classification, were plotted with Kaplan-Meier methods. Difference between curves was assessed with the log rank test and median survival for each subgroup of patients was provided.

Applying the NGS-IgHV classification to the whole cohort identified five subgroups with median TFS of >280 (Multi-M), 131 (M), 94 (Mix), 29 (UM),

19 (Multi-UM) months (P<0.0001) and a median OS of >397, 292, 196, 137, 94 months (P<0.0001) (Figure 4.6).

Table 4.2. Number of deaths in Binet Stage A patients.

	No of Patients	**No of deaths**	**%**
Multi-M	25	0	0
M	98	9	9.2
Mix	27	5	18.5
UM	41	11	26.8
Multi-UM	4	1	25

Since Binet Stage A patients account for over 80% of new diagnoses (Baliakas, Hadzidimitriou, *et al* 2015), the model was applied to Binet Stage A patients only. In Population A, Stage A patients represented 72.6% (196/270) of the total patients. The model was found to be a significant prognostic marker of TFS and OS for these patients (p<0.001) (Figure 4.7) although the median OS for the UM group was higher at 311 months than for both the Mix (196 months) and M (292 months) groups despite the UM group having a higher death rate (Table 4.2). This is most likely due to the low number of deaths in the Binet Stage A cohort; 13.3% (26/196) versus 23.0% across the cohort (62/270).

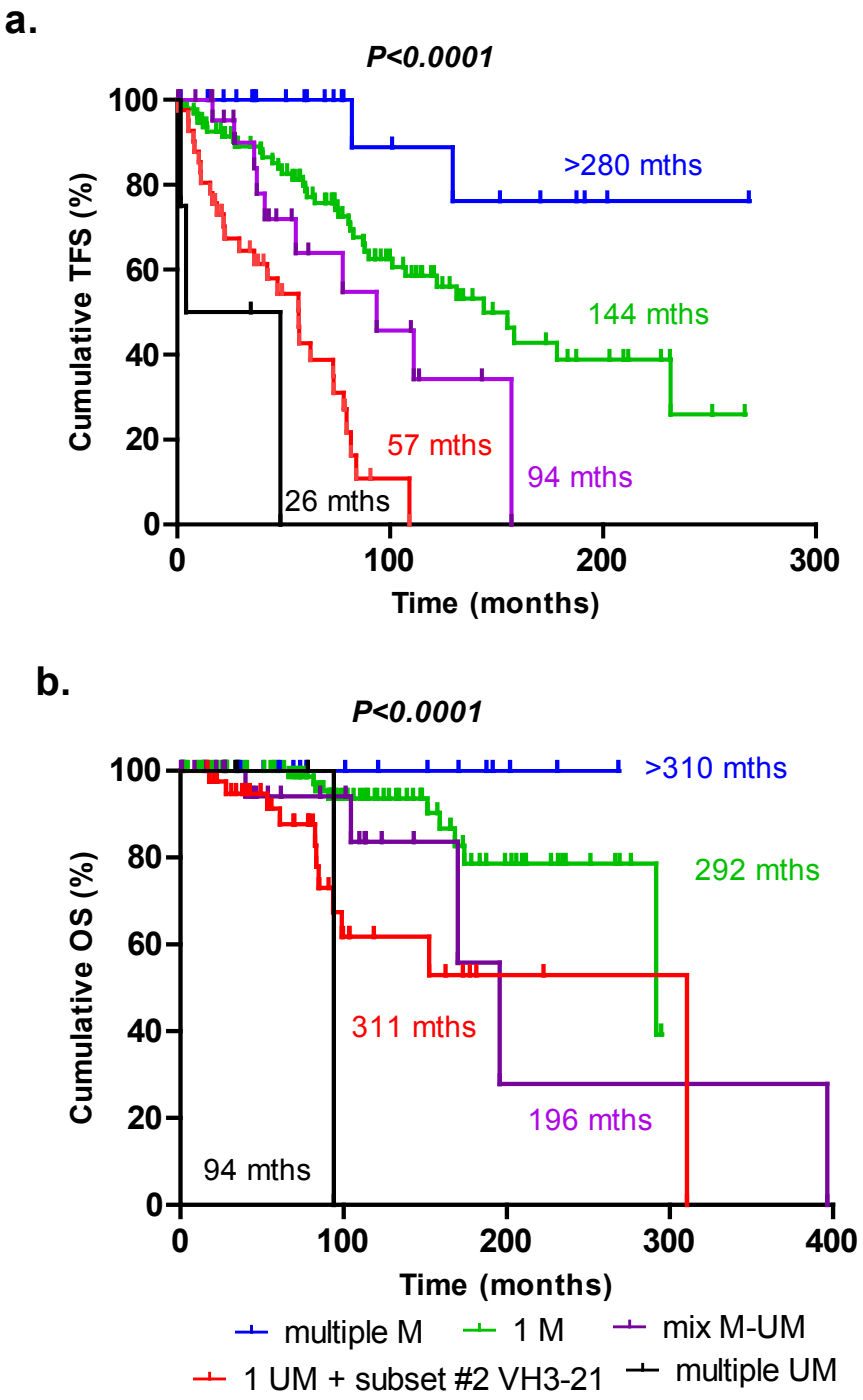


Figure 4.7. NGS-IgHV defines 5 different prognostic subgroups in Binet Stage A patients. TFS (a) and OS (b), according to NGS-IgHV classification, were plotted with Kaplan-Meier methods. Difference between curves was assessed with the log rank test and median survival for each subgroup.

Table 4.3. Predictive value of NGS-IgHV. PPV positive predictive value, NPV negative predictive value.
** patient with at least a follow up of 2 years or receiving a treatment within 2 years after diagnosis.*
*** patients with at least a follow up of 10 years or died within 10 years after diagnosis*

Treatment Prediction

		No of Patients*	%	*Treated within 4 years of diagnosis	Untreated 4 years after diagnosis	PPV %	NPV %
Sseq	**M**	107	49	23	84		79
	UM	90	41	62	28	69	
	Polyclonal	20	9	5	15	undetermined	
NGS	**Multi-M**	19	9	1	18		95
	1M	97	45	21	76		78
	Mix M-UM	17	8	7	10	undetermined	
	1UM	79	36	56	23	71	
	Multi-UM	5	2	4	1	80	

Survival Prediction

		No of Patients**	%	Deceased within 10 years of diagnosis	Alive 10 years after diagnosis	PPV %	NPV %
Sseq	**M**	56	44	8	48		86
	UM	58	46	31	27	53	
	Polyclonal	12	10	3	9	undetermined	
NGS	**Multi- M**	9	7	0	9		100
	1M	55	44	8	47		85
	Mix M-UM	8	6	3	5	undetermined	
	1UM	52	41	29	23	56	
	Multi-UM	2	2	2	0	75	

To determine the predictive power of the NGS-IgHV classification the positive predicative value (PPV) and negative predictive value (NPV) were calculated for both the NGS-IgHV and SSeq-IgHV subgroups. NGS-IgHV had a PPV of treatment initiation within 4 years of diagnosis, of 80% in patients with multiple UM clones and a NPV of 95% in patient with multiple M clones, compared to a PPV of 69% for SSeq-UM and a NPV of 79% for SSeq-M. NGS-IgHV could also predict an OS of less than 10 years from diagnosis with a PPV of 75% for multiple UM and a NPV of 100% for multiple M subgroups compared with 53% for SSeq-UM and 86% for SSeq-M. For patients with only one clone (M or UM), similar PPV and NPV were observed between SSeq and NGS (Table 4.3).

4.3.3 IgHV by NGS defines a very poor prognostic subgroup in a treatment-selected cohort (Population B)

To validate the NGS-IgHV model, it was applied to a homogeneous clinical trial population, composed of 227 patients treated with FCR-based regimen (population B). In this cohort, NGS identifies productive clonal rearrangements in 99.1% of patients (225/227). The number of multiple rearrangements detected was less than population A at 11.5% (26/227). Furthermore, since population B is solely composed of patients that had received treatment, this introduced a bias in the patient population and therefore the five NGS-IgHV categories were not confirmed in this clinical trial population. This is unsurprising given that the frequency of multi-M

patients decreased from 10% in population A to just 3% in population B, whilst the frequency of Multi-UM patients increased from 2% to 9%. Indeed, the great majority of Multi-M patients in population A (90%) never required treatment and were therefore, by definition, would have been excluded from this clinical trial population.

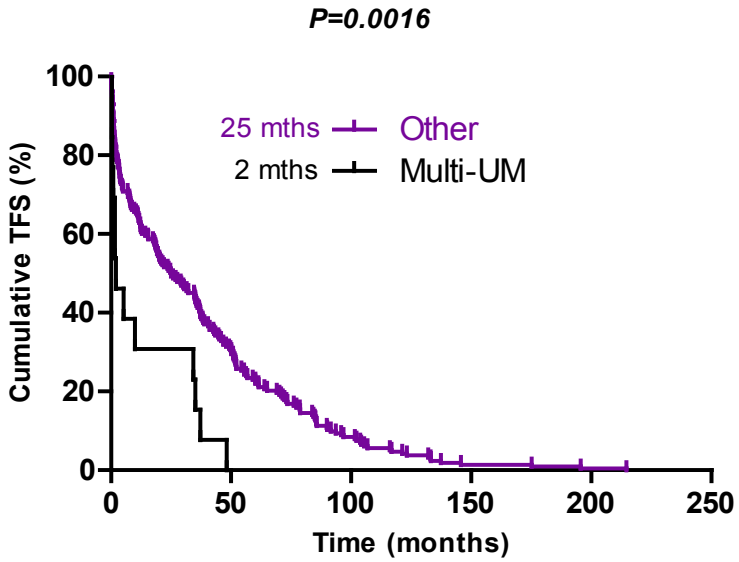


Figure 4.8. NGS-IgHV defines a poor prognostic subgroup in a treatment selected clinical trial cohort. Difference between curves was assessed using an univariate analysis using with the log rank test and median survival for each subgroup of patients is provided.

Despite this patient bias, it was possible to observe that Multi-UM patients had a far worse prognosis in terms of TFS compared to the rest of population B as these patients had a median TFS of just 2 months

(p=0.0016) (Figure 4.8). Additionally, the Multi-UM patients had a higher rate of progression compared to the rest of population B at 30.8% (4/13) compared to 21.9% (46/210) within 5 years of treatment, although neither group had reached its median progression free survival (PFS) at the time of writing.

4.3.4 NGS-IgHV is an independent prognostic marker

To compare how the NGS-IgHV model compared with other prognostic models described in Sections 1.5, a Cox regression analysis was used in Population A to determine the hazard ratios (HR). The presence of a 17p deletion was binarised (1); Binet stage A, B and C were named 0, 0.5 and 1, respectively; Döhner cytogenetic classification was named as 0, 0.25, 0.5, 0.75, and 1 for 13q del, normal karyotype, trisomy 12, 11q del and 17p del, respectively, the CLL-IPI groups low risk, moderate risk, high risk and very high risk were called 0, 0.33, 0.66 and 1, respectively; and for NGS-IgHV, the five categories – (i) Multi-M; (ii) M; (iii) Mix M-UM; (iv) UM (+VH3-21), (V) Multi-UM - were named as 0, 0.25, 0.5, 0.75 and 1, respectively. Using this type of naming meant that the HR represents the hazard ratio between the very favourable prognosis group and the very unfavourable prognosis group whilst considering the intermediate groups in an ordinal manner. A univariate Cox regression analysis (n=167) found that NGS-IgHV mutational status was the only significant HR compared to all tested factors for TFS (7.845, 95% CI 2.73 – 22.539, P<0.0001) (Table 4.4).

Table 4.4. Univariate Cox regression for Population A for TFS. HR hazard ratio, CI confidence interval. Binet stage calculated according to Binet et al. (1981), Döhner classified according to Döhner *et al* (2000) *and CLL-IPI calculated according to* The International CLL-IPI Working Group (2016).

	P Value	**HR**	**95% CI**	
			Lower	**Higher**
Döhner	0.836	0.921	0.424	2.003
NGS IgHV	<0.0001	7.845	2.731	22.539
Binet	0.124	1.639	0.874	3.074
CLL-IPI	0.146	2.659	0.712	9.928
del17p	0.229	1.898	0.668	5.393

In a multivariate stepwise analysis (n=167) of these models, both IgHV mutational status by NGS (P<0.0001) and the CLL-IPI model (P<0.0001) were selected as independent and significant predictors of TFS (Table 4.5). None of these models were found to be significant for OS but this is most likely due to the small population size.

Table 4.5. Multivariate Cox regression for Population A for TFS. HR hazard ratio, CI confidence interval. Binet stage calculated according to Binet et al. (1981), Döhner classified according to Döhner et al (2000) and CLL-IPI calculated according to The International CLL-IPI Working Group (2016).

		P Value	HR	95% CI	
				Lower	Higher
Step 1.	**Döhner**	0.836	0.921	0.424	2.003
	NGS IgHV	<0.0001	7.845	2.731	22.539
	Binet	0.124	1.639	0.874	3.074
	CLL-IPI	0.146	2.659	0.712	9.928
	del17p	0.229	1.898	0.668	5.393
Step 2.	**NGS IgHV**	<0.0001	7.693	2.716	21.790
	Binet	0.110	1.658	0.892	3.082
	CLL-IPI	0.149	2.634	0.706	9.821
	del17p	0.205	1.798	0.725	4.455
Step 3.	**NGS IgHV**	<0.0001	5.755	2.285	14.494
	Binet	0.229	1.428	0.799	2.551
	CLL-IPI	0.004	4.488	1.615	12.473
Step 4.	**NGS IgHV**	<0.0001	5.642	2.286	13.925
	CLL-IPI	<0.0001	6.215	2.681	14.409

For 204 patients of Population B, a Cox regression multivariate analysis was performed. For this analysis, factors were binarised as either 'having' (1) or 'not having' (0), except for SSeq-IgHV which was named as 0 for M and 1 for UM. The factors included in the regression were mutation analysis by targeted NGS of commonly the mutated genes TP53, ATM, NOTCH1, SF3B1, and BIRC3; SNP array detection of the common cytogenetic abnormalities deletions 17p, 11q, 13q and trisomy 12; SSeq- IgHV and NGS-IgHV Multi-

UM (1) or not (0). For this analysis, TP53 mutation and 17p deletion were merged, as were ATM mutation and 11q deletion. Only the presence of a multi-UM was found to be an independent and significant predictor of TFS (P<0.0001) (Table 4.6).

Table 4.6. Multivariate Cox regression for Population B. HR hazard ratio, CI confidence interval.

		P Value	HR	95% CI	
				Lower	Higher
Step 1	**TP53 Mut/Del**	0.499	1.197	0.710	2.019
	ATM Mut/Del	0.155	1.275	0.912	1.781
	BIRC3 Mut	0.664	0.882	0.500	1.555
	Tri12	0.409	1.209	0.771	1.896
	NOTCH1 Mut	0.882	1.032	0.684	1.556
	SF3B1 Mut	0.655	1.084	0.762	1.543
	Del13q	0.336	0.857	0.626	1.173
	MultiUM	<0.0001	3.086	1.671	5.696
Step 2	**TP53 Mut/Del**	0.501	1.197	0.710	2.018
	ATM Mut/Del	0.157	1.273	0.911	1.778
	BIRC3 Mut	0.679	0.889	0.510	1.550
	Tri12	0.411	1.207	0.770	1.893
	SF3B1 Mut	0.660	1.082	0.761	1.540
	Del13q	0.309	0.853	0.627	1.159
	MultiUM	<0.0001	3.101	1.685	5.704
Step 3	**TP53 Mut/Del**	0.491	1.202	0.713	2.026
	ATM Mut/Del	0.171	1.261	0.905	1.756
	Tri12	0.461	1.176	0.764	1.812
	SF3B1 Mut	0.643	1.087	0.764	1.545
	Del13q	0.287	0.847	0.624	1.149

	MultiUM	<0.0001	3.102	1.686	5.706
Step 4	**TP53 Mut/Del**	0.533	1.178	0.703	1.974
	ATM Mut/Del	0.175	1.258	0.903	1.751
	Tri12	0.522	1.146	0.755	1.739
	Del13q	0.256	0.839	0.620	1.136
	MultiUM	<0.0001	3.104	1.688	5.709
Step 5	**ATM Mut/Del**	0.200	1.238	0.893	1.717
	Tri12	0.578	1.124	0.745	1.697
	Del13q	0.266	0.843	0.623	1.140
	MultiUM	<0.0001	3.118	1.695	5.734
Step 6	**ATM Mut/Del**	0.240	1.206	0.882	1.649
	Del13q	0.163	0.818	0.616	1.085
	MultiUM	<0.0001	3.042	1.665	5.559
Step 7	**Del13q**	0.129	0.804	0.607	1.065
	MultiUM	<0.0001	3.051	1.670	5.571
Step 8	**MultiUM**	<0.0001	3.185	1.748	5.803

To determine if there was any correlation between the five NGS-IgHV categories and other recurrent genetic abnormalities in CLL, a heatmap association chart was plotted for 204 patients in population B (Figure 4.9). Visual inspection of this heatmap shows no obvious associations between the five NGS-IgHV subgroups and other common genetic abnormalities.

Figure 4.9. NGS-IgHV is independent from other prognostic markers. Heatmap showing the distribution of recurrent genetic abnormalities in the different NGS-IgHV subgroups.

4.3.5 IgHV subclones are phylogenetically unrelated

Since multiple studies have demonstrated that clones from independent CLL patients displayed similar, or even identical, CDR3 amino acid regions, the frequency of stereotyped receptors in the 358 productive sequences obtained from population A were investigated. Using previously described methods (Bomben *et al* 2009), 36.9% (132/358) of sequences were found to have over 70% of identity when all the sequences were considered, which is in line with previous reports (Agathangelidis *et al* 2012). However, since more than one sequence was analysed per patient due to the presence of subclones, 47.8% (129/270) of patients had at least one stereotyped clone. This value decreases to 37.8% (102/270) of patients when the analysis is performed using only the major clone. In population B, 33.5% (90/269) of sequences were identified as having greater than 70% identity and 37.8% (85/225) of patients had at least one stereotyped receptor.

When both populations were combined subclonal patients constituted 18.1% (90/497) of the total cohort. Of these patients, only 13.3% (12/89) expressed two subclones containing the same VH family gene. In these cases, the clonotypes had different D and J genes (Figure 4.10) suggesting that the clones had arisen independently, during recombination.

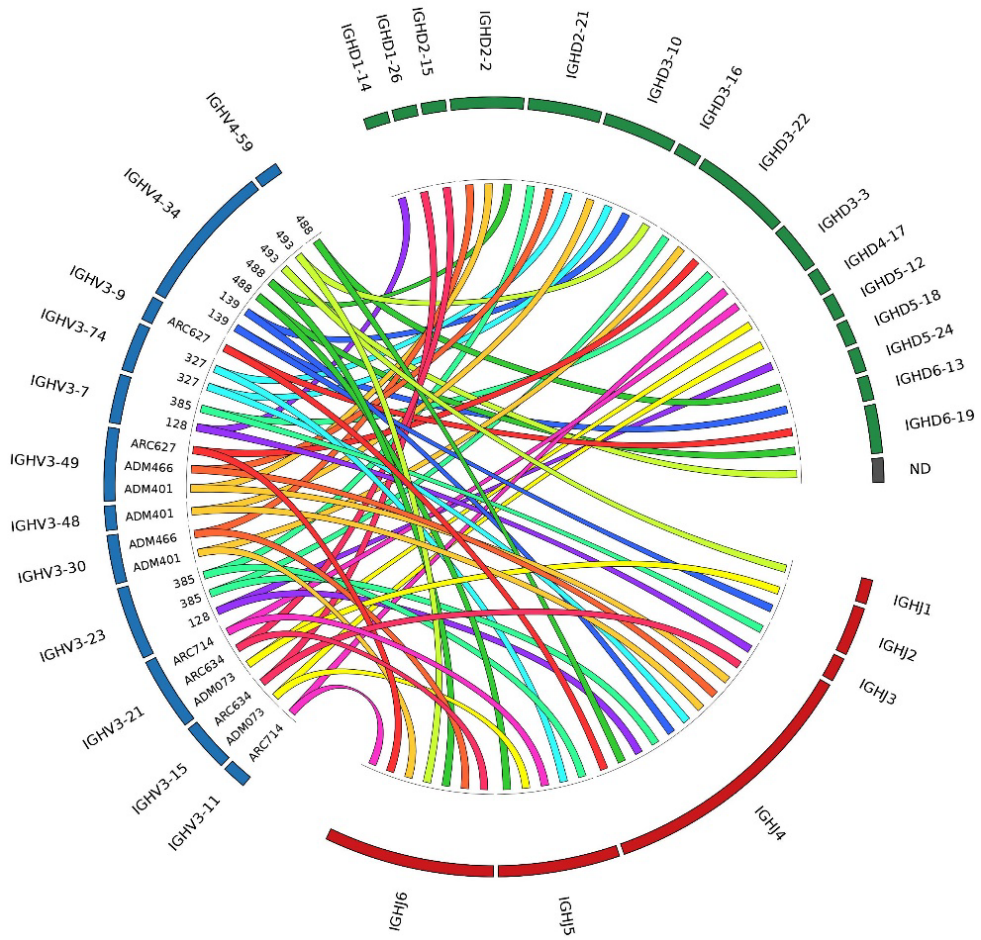


Figure 4.10. Circos plot visualization of V to D and V to J associations in the 12 patients expressing subclonal IgHV rearrangements from the same VH family.

Further stereotyping analysis of the intra-clonal sequences identified that only 6.7% (6/90) of subclonal patients expressed related CDR3s, as defined as greater than 70% identity. Interestingly, one of these patients with intra-clonal similarity, had 100% CDR3 AA sequence homology despite both subclones expressing completely different VH-family genes (Table 4.7).

Table 4.7. ADM304 with 100% intra-clonal CDR3-AA sequence homology. Freq – frequency, CDR3 – complementary determining region 3, AA – amino acid

	Freq (%)	V-gene	J-gene	D-gene	CDR3 AA seq
ADM304	95.8	IGHV3-33	IGHJ6	IGHD3-16	RDLGSDVRLYYYYGMDVW
	3.0	IGHV1-3	IGHJ6	IGHD3-16	RDLGSDVRLYYYYGMDVW

4.3.6 Subclonal analysis using NGS-IgHV shows an evolving picture

Work described in the Chapter 3 suggested that the IgHV remained stable between diagnosis and relapse. To investigate this further, 31 samples were screened at diagnosis and relapse (Table 4.8). Of the 31 patients, four had subclones at diagnosis (13%), which is in keeping with population B, since all the patients in this cohort have also undergone treatment. Of these four, two were multi-UM and two were mixed. At relapse, three had lost their subclones and one had detectable subclones although the clonal burden had reduced from 17.8% to 2.6%. Two patients had acquired subclones at relapse (CLL082 and CLL156) which were not detectable in the 200 MARs. Interestingly analysis of all the unique reads in the diagnosis sample showed that for CLL156 (reads=140,168) the subclones were present at diagnosis but were below the 2.5% threshold, whilst for CLL082 (reads=109,804) the subclone was undetectable in the diagnostic sample.

Table 4.8. NGS-IgHV analysis of diagnosis-relapse pairs. UM – unmutated, M -mutated, Freq - frequency

	DIAGNOSIS			RELAPSE		
Sample	**V-gene**	**IgHV Status**	**Freq (%)**	**V-gene**	**IgHV Status**	**Freq (%)**
ARR002	IGHV3-21	UM	100	IGHV3-21	UM	100
CLL003	IGHV1-69	UM	100	IGHV1-69	UM	100
CLL023	IGHV3-9	UM	100	IGHV3-9	UM	100
CLL044	IGHV3-30	M	100	IGHV3-30	M	100
CLL050	IGHV3-30	UM	100	IGHV3-30	UM	100
CLL056	IGHV3-48	UM	100	IGHV3-48	UM	100
CLL063	IGHV3-21	UM	100	IGHV3-21	UM	100
CLL079	IGHV5-51	UM	84.2	IGHV5-51	UM	97.3
	IGHV3-48	UM	17.8	IGHV3-48	UM	2.6
CLL080	IGHV3-49	M	100	IGHV3-49	M	100
CLL081	IGHV1-24	UM	100	IGHV1-24	M	100
CLL082	IGHV4-34	M	100	IGHV4-34	M	87.3
	-	-	-	IGHV1-18	UM	12.7
CLL083	IGHV1-18	UM	100	IGHV1-18	UM	100
CLL084	IGHV1-46	UM	100	IGHV1-46	UM	100
CLL085	IGHV3-15	UM	94.0	IGHV3-15	UM	100
	IGHV5-51	UM	6.0	-	-	-
CLL086	IGHV3-23	UM	100	IGHV3-23	UM	100
CLL087	IGHV4-31	UM	100	IGHV4-31	UM	100
CLL088	IGHV4-34	M	92.2	IGHV4-34	M	100
	IGHV3-15	UM	5.6	-	-	-
CLL089	IGHV3-48	UM	94.6	IGHV3-48	UM	97.6
	IGHV3-30	M	2.9	-	-	-
CLL090	IGHV3-30	M	100	IGHV3-30	M	100
CLL091	IGHV1-6	M	100	IGHV1-6	M	100
CLL092	IGHV1-69	UM	100	IGHV1-69	UM	100
CLL094	IGHV1-2	UM	100	IGHV1-2	UM	100
CLL096	IGHV4-34	M	100	IGHV4-34	M	100
CLL098	IGHV4-4	UM	100	IGHV4-4	UM	100
CLL099	IGHV3-33	UM	1001	IGHV3-33	UM	100
CLL103	IGHV3-11	UM	100	IGHV3-11	UM	100
CLL106	IGHV1-2	UM	100	IGHV1-2	UM	100
CLL107	IGHV1-2	UM	100	IGHV1-2	UM	100
CLL108	IGHV1-69	UM	100	IGHV1-69	UM	100
CLL145	IGHV3-23	M	100	IGHV3-23	M	100
CLL156	IGHV1-24	UM	100	IGHV1-24	UM	59.5
	-	-	-	IGHV3-23	UM	24.6
	-	-	-	IGHV3-30	M	9.1

4.4 Discussion

The identification of subclones containing different productive VDJ clonotypes provides strong evidence that the leukaemia initiating event in CLL occurs early in B-cell development before the cell undergoes VDJ recombination. This suggests that a precursor B-cell is the leukaemic stem cell in CLL. Prior to the work described in this chapter, the percentage of CLL cases with more than one IgHV clone had been reported as relatively rare at a frequency of less than 5% (Plevova *et al* 2014, Rassenti and Kipps 1997, Visco *et al* 2013). In contrast, by using NGS, this study reports a frequency five times higher, at 24%, in an unselected patient cohort.

4.4.1 NGS-IgHV classification refines the SSeq-IgHV prognostic subgroups

In this study, it was shown that IgHV-NGS could stratify patients into five different categories, with respect to TFS and OS, in an unselected CLL population. Patients with multi-M rearrangements had the best prognosis, even compared to patients with only one M clone. This is in line with Visco *et al*. (2013) who observed that biclonally M patients had a significantly lower CD38, more favourable cytogenetic lesions and a more indolent clinical course with the longest TTFT. Similarly, they found multi-UM to have a short TTFT. In contrast to the work described in this chapter, they reported that patients with mixed IgHV rearrangements had the worst

prognosis, although this might be due to the decreased sensitivity of the techniques used.

The new NGS-IgHV classification can precisely classify patients with multiple IgHV rearrangements for which SSeq is inconclusive and improved prognostication for 92 out of 270 patients. Importantly, the poor prognosis of multiple UM patients detected by NGS-IgHV could be confirmed in an independent clinical trial population and showed a very short median TFS of 2 months. The presence of multiple IgHV re-arrangements was the strongest independent prognostic indicator in regression analysis of 11 different variables including important driver mutations.

4.4.2 IgHV subclones detected by NGS provide evidence for a precursor B-cell as the leukaemic stem in CLL

The ability of NGS to detect a higher frequency of subclones in comparison to SSeq is unsurprising. NGS has a far greater sensitivity than SSeq which is often reported at around 20% (Tsiatis *et al* 2010). Furthermore, since NGS sequences each amplicon individually it is possible for the separate subclones to be detected as well as preventing larger subclones from masking the presence of smaller ones.

Since NGS has been shown to detect low level IgHV subclones at frequencies of down to 0.02% (Campbell *et al* 2008), it is necessary to exclude the normal B-cell background as the origin of these minor subclones. This was done by limiting the analysis to the first 200 MARs and further restricted by using a 2.5% cutoff for the classification of a subclone. Confirmation of the rationality of these restrictions was demonstrated through a series of validation experiments. Of interest, the comparison of gDNA and cDNA showed some differences. Firstly, that gDNA detected more unproductive subclones. This observation is not unexpected, however, as it is compatible with the instability of unproductive rearrangement at the RNA level (Ghia *et al* 2007, Langerak *et al* 2011). Secondly that cDNA had a far greater sensitivity for productive subclonal detection. This may in part explain the difference in the frequency of subclonal detection between Population A and B, since it suggests that gDNA subclones may be discarded as the normal B-cell background.

Several reports have previously described the presence of multiple CLL clones (Blachly *et al* 2015, Kriangkum *et al* 2015, Plevova *et al* 2014) as well as multiple clones with a discordant IgHV mutational status (Langerak *et al* 2011, Visco *et al* 2013). The presence of multiple productive rearrangements in one quarter of CLL patients could be explained by a lack of allelic exclusion (Rassenti and Kipps 1997), which would suggest that both rearrangements are actually present in the same cell. However, a

recent study using single cell sequencing demonstrated that there was only one productive IgHV clonotype per cell (Kriangkum *et al* 2015). Furthermore, the study in this chapter has demonstrated incidences of three or more clonotypes. Taken together, this suggests that multiple rearrangements correspond to multiple leukemic subclones as was previously assumed (Plevova *et al* 2014, Visco *et al* 2013), implying that the leukemia-initiating events occur prior to IgHV re-arrangement in an unrearranged early B cell progenitor.

4.4.3 Stereotypy analysis suggests that the IgHV subclones are phylogenetically unrelated

Interclonal stereotyping of the CDR3 has been reported in a third of CLL cases (Agathangelidis *et al* 2012, Darzentas *et al* 2009, Stamatopoulos *et al* 2007), leading to the theory that CLL pathogenesis is driven through the activation of the BCR by a common antigen or superantigen. Conversely, stereotyping analysis of the intraclonal CDR3s of subclonal IgHV patients in this study showed that, in most cases, there was no phylogenetic relationship. This contradicts this theory, since it would be expected that different clones in a given patient would have similar CDR3 region, if the leukemic clones emerged in response to a single common extracellular antigen. However recent work shows that Ig-BCRs of unrelated clonotypes may recognize common epitopes that were unpredictable from the molecular features of the Ig-BCR, including the CDR3 composition and

length (Mimmi *et al* 2016). Further in-depth studies of the hematopoietic stem cell and early B cell progenitor compartments in these patients will be required to confirm these observations.

4.4.4 Conclusions

In conclusion, the incidence of cases with multiple productive IgHV subclonal rearrangements is higher than previously reported at 24.4%. These subclones are likely to significantly contribute to the biological heterogeneity of this disease and point towards leukemia-initiating events that pre-date IgHV rearrangement. From a practical perspective, NGS-IgHV resolves the diagnostic uncertainties in a significant number of patients, enables the detection of small subclones and refines the previous SSeq-based IgHV prognostic classification system. Finally, a new NGS-IgHV classification with five different clinically relevant prognostic categories has been proposed.

Chapter 5 Whole genome analysis using Single Nucleotide Polymorphism arrays

5.1 Introduction

The work described in the previous chapter demonstrated that the incidence of patients with multiple productive IgHV rearrangements is much higher than previously reported. Firstly, this finding implies potential biological heterogeneity of CLL, but it is not known whether the different IgHV subclones are also genetically and functionally different. Secondly, it suggests that leukaemia-initiating events are likely to occur prior to VDJ recombination in an early B-cell leukaemic progenitor. However, it does not provide a mechanism for leukaemogenesis *per se*. Thirdly, similar to the IgHV mutation status (Damle *et al* 1999, Hamblin *et al* 1999), the new NGS-IgHV classification of CLL patients was also of independent prognostic significance, and there was no association between the five NGS-IgHV categories and the common chromosomal aberrations detected in CLL, which have also been linked to prognosis (Dohner *et al* 1995, Gonzalez *et al* 2011, Landau *et al* 2013, Zenz *et al* 2010, Austen *et al* 2007, Hurtado *et al* 2015, Lozano-Santos *et al* 2016, Skowronska *et al* 2012, Jethwa *et al* 2013, Nadeu *et al* 2016, Wang *et al* 2011).

Using conventional cytogenetics, the incidence of a chromosomal abnormality in CLL is around 80% (Dohner *et al* 2000). However, through utilising both CGH and SNP array technology it has been possible to detect copy number alterations (CNA) in over 90% of cases (Gunn *et al* 2008, Gunnarsson *et al* 2011, Kay *et al* 2010, Ouillette, Collins, Shakhan, Li, Peres, *et al* 2011). Whilst CGH arrays have been reported to have a better CNA detection capacity than SNP arrays, particularly for small regions, this was done using relatively low resolution SNP arrays (<300K coverage) (Gunnarsson *et al* 2008). Furthermore, an additional benefit of SNP arrays over CGH arrays is the ability to detect copy neutral loss of heterozyogisty (cnLOH), which has been shown to be useful in identifying mutations in TP53 and ATM, which would have otherwise appeared normal using FISH or CGH array analysis (Edelmann *et al* 2013, Ouillette, Collins, Shakhan, Li, Peres, *et al* 2011). Therefore, a high resolution array with over one million SNP probes, would have a greater sensitivity than previous SNP arrays used to study CLL to date (Gunnarsson *et al* 2008, Pfeifer *et al* 2006) as well as the ability to detect cnLOH, making it an ideal tool for a systematic genome-wide analysis of recurrent chromosomal aberrations in CLL. The hypothesis was that these abnormalities may shed light on the molecular drivers of leukaemogenesis and relapse and identify additional novel predictors of response.

The work presented in the this chapter therefore aimed to:

1. Describe the different CNAs in a cohort of patients before therapy and subsequent relapse and identify potential driver genes by MOR analysis. It was assumed that subclones containing driver mutations would be present at relapse whilst subclones with passenger mutations would decrease in number, or completely die out.
2. Validate a selection of these in a larger cohort of clinically well annotated patients

3. Correlate findings to clinical outcome

5.2 Materials and Methods

5.2.1 Patient Samples

The test project used gDNA from a pre-treatment cohort of 95 patients that were typical of that found in a tertiary referral centre with a high proportion of unmutated and treatment refractory cases (population C, Appendix 2). For 42 patients gDNA from subsequent relapse was also used. The validation project used gDNA from a cohort of 242 CLL patients enrolled into four UK based FCR clinical trials (population B and population D, appendix 2). All gDNA was extracted using the method described in section 2.2.

5.2.2 Genome-wide SNP Platform hybridizations and analysis

DNA was analysed using SNP arrays as previously described (Section 2.8)

5.2.3 Sanger sequencing

Sanger sequencing (Section 2.6) was performed on gDNA from 40 CLL patients using the *ATG5* and *BLIMP1* primers (Appendix 1).

5.2.4 Targeted NGS gene panel

DNA from 10 CLL patients was analysed using a TSCA (Section 2.7.6) CLL gene targeted NGS panel (Appendix 3) and the LymphoTrack IGH Somatic Hypermutation Assay (Section 2.7.5).

5.3 Results

5.3.1 Common cytogenetic aberrations are recurrent within this cohort

5.3.1.1 17p Deletions

In the test cohort of 95 patients, 8.4% (8/95) were found to have a 17p deletion that removed the *TP53* gene (Figure 5.1). One of these patients (CLL094) had a subclonal deletion that was found in approximately 40% of the sample. All the deletions removed the p-terminus of chromosome 17. The longest deletion was 22.7Mb in length and the shortest was 16.5Mb. In addition, a further four patients (4.2%) were found to have cnLOH over the *TP53* gene, including one patient (CLL145) with a subclonal cnLOH in ~50% of the sample. As per the deletions, all the cnLOH affected the p-terminus of chromosome 17 with the length ranging from 8.1Mb to 21.1Mb.

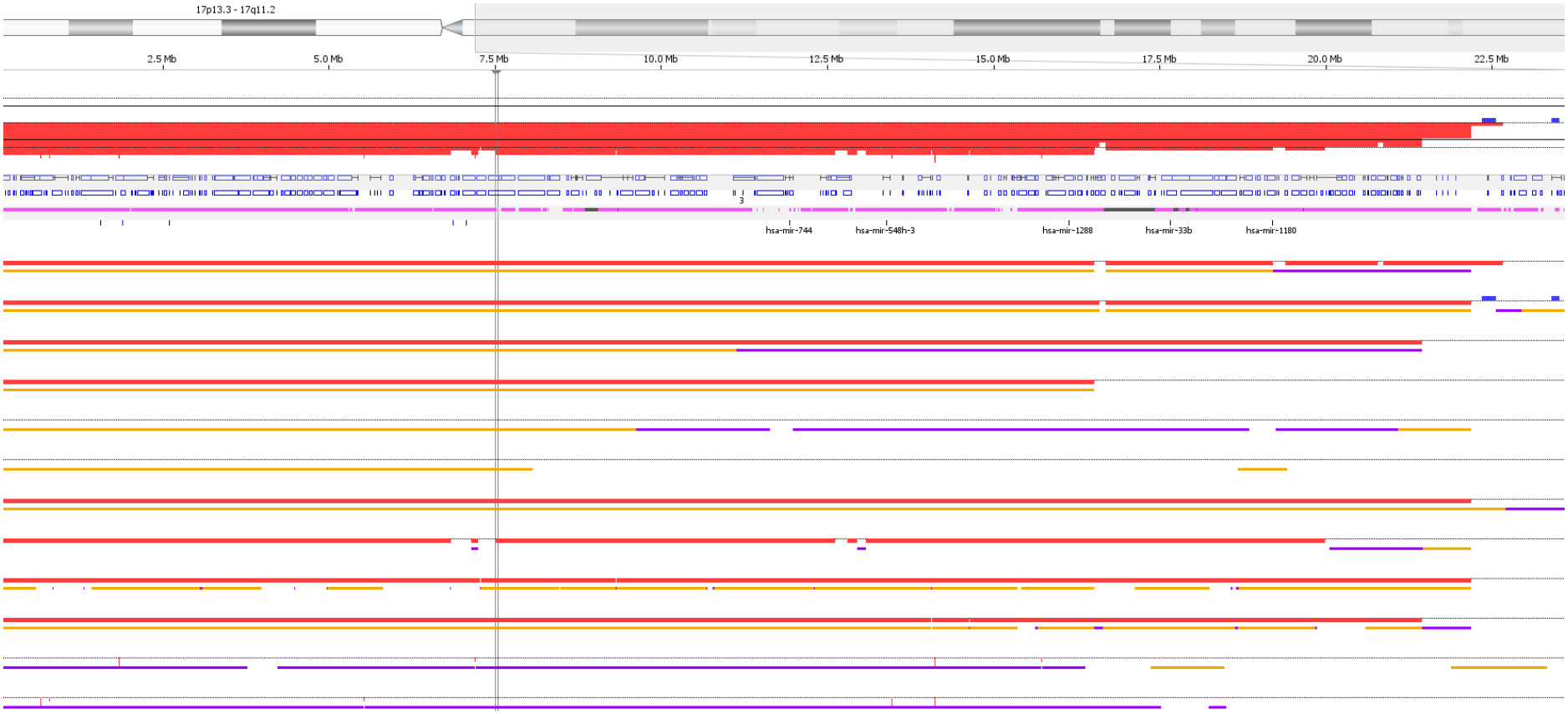


Figure 5.1. TP53 deletions are recurrent in CLL. Nexus (BioDiscovery) image showing Illumina 1M- Duo data calls from 12 patients with deletions/cnLOH of Chr17p. Red - CNA loss, orange - LOH; purple - allelic imbalance. The grey line marks the location of the TP53 gene.

Figure 5.2. ATM deletions are recurrent in CLL. Nexus (BioDiscovery) image showing Illumina 1M- Duo data calls from 18 patients with deletions/cnLOH of Chr17p. Red - CNA loss, orange - LOH; purple - allelic imbalance. The grey line marks the location of the ATM gene.

5.3.1.2 11q Deletions

In total, 18.9% (18/95) patients had an 11q abnormality over the *ATM* gene (Figure 5.2). Of these 18 patients, 13 had a deletion and five had cnLOH. The size of the abnormalities ranged from 0.2Mb to 72.3Mb. Interestingly the smallest deletion (CLL073), although affecting three genes; *C11orf65*, *NPAT* and *ATM*, only completely removed *ATM* (Figure 5.3).

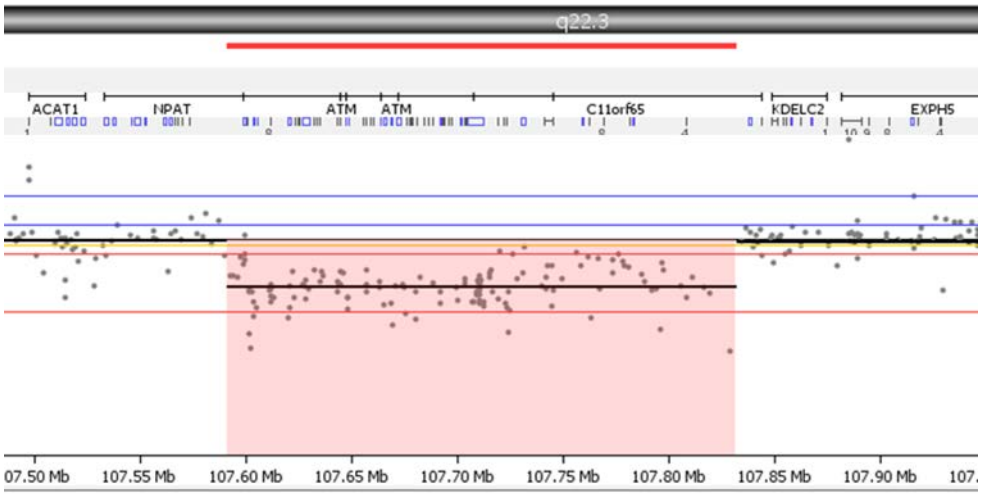


Figure 5.3.The 11q deletion in CLL073 only completely removes the ATM gene. Nexus (BioDiscovery) image showing Illumina 1M- Duo data call for CLL073 at 11q22. Red - CNA loss.

5.3.1.3 Trisomy 12

13.7% (13/95) of patients had a trisomy 12. One additional patient (CLL 108) had a partial trisomy 12 from 12p.12.2 to 12q21.31 that was 63.9Mb in size. Interestingly, the rest of chromosome 12 for this patient had cnLOH (Figure 5.4). Another patient (CLL189) also had a large region of cnLOH over chromosome 12 which covered from 12q13.11 to 12qter.

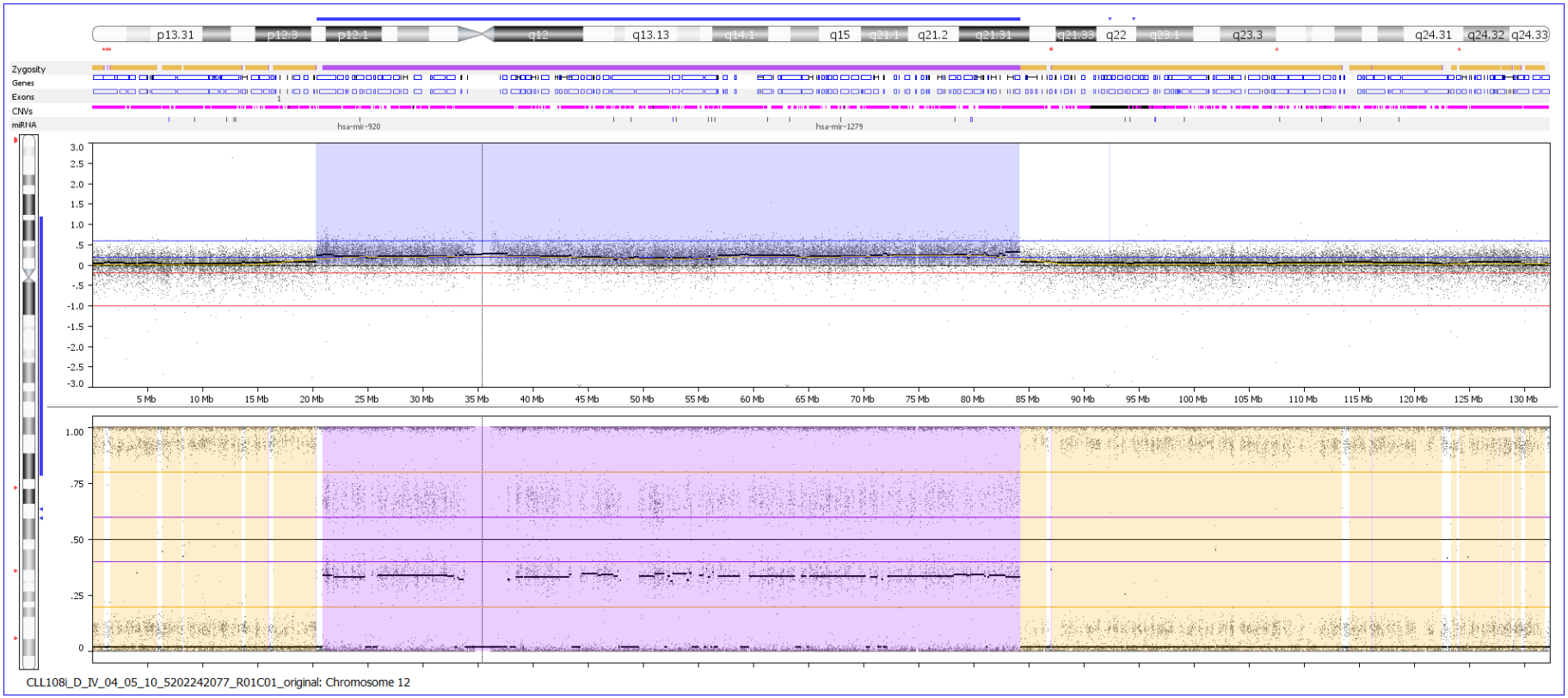


Figure 5.4. Partial trisomy of chromosome 12 in CLL108. Nexus (BioDiscovery) image showing Illumina 1M- Duo data calls for CLL108 at chromosome 12. Blue - CNA gain, orange - LOH; purple - allelic imbalance.

5.3.1.4 13q14 Deletions

52.6% of patients (50/95) were found to have either cnLOH, a heterozygous or homozygous deletion at the 13q14 locus. The heterozygous deletion was the most common abnormality detected in 66.0% (33/50) of this cohort. The size of the heterozygous deletion ranged from 0.6Mb to 78.8Mb in size, whereas the homozygous deletion which was identified in 28% (14/50) went from 0.15Mb to 2.6Mb. The remaining 6% (3/50) had cnLOH ranging from 0.7Mb to 66.5Mb. Mosaicism was common, with 16% (8/50) of cases having a 13q14 deletion subclonal population, averaging 67.5% (range 30%-90%) of the total clone.

The MOR for the 13q14 deletion was defined by two samples, CLL088 and MCLL100, and was 7364bp (0.007Mb) in length. This region contained just exon 2 of *DLEU2*. Mapping the MOR of the biallelic deletion was more complicated as two samples, CLL075 and MCLL148 had intermittent homozygous deletions. Therefore, two MOR regions were identified for the biallelic deletions. The first was a 0.11Mb region that was demarcated by CLL096 and CLL148. This area encompassed the first three exons of *DLEU2*. The second biallelic MOR was characterised by CLL075 and CLL105. This was 0.06Mb in length and contained exon 5 of *DLEU1*. Interestingly, miR15a and miR16-1 did not map to any of these MORs, although they were located about 9Kb (0.009Mb) downstream from the monoallelic MOR.

5.3.2 Large copy number alterations are recurrent in CLL

Since other chromosomal abnormalities have been reported in CLL, additional recurrent genetic aberrations were investigated. Abnormalities were defined as recurrent if they were present in more than three samples (3% of the cohort) and further investigated if at least two of these samples had deletions or gains i.e. cnLOH was not examined in isolation. Using these criteria, abnormalities were identified on 2q, 6q, 8p, 9p, 10q, 18p, 19p and 20q (Table 5.1). Whilst a recurrent deletion was identified on 22q, these were discounted as artefacts of Igλ rearrangement (Mraz *et al* 2013).

5.3.2.1 Del(2q)

There were two regions of interest on 2q, both of which were defined by deletions or cnLOH. The larger affected area was located at 2q22-2q24 whereby two patients had a deletion (CLL068, CLL081) and two had cnLOH (CLL095, CLL103). The deletions were both smaller than the cnLOH at 13.7Mb and 20.8Mb compared with 113.4Mb and 30.7Mb for the cnLOH. The two deletions defined the MOR as 11.3Mb in length (Figure 5.5a). There were many genes described within this area.

The smaller region of interest on chromosome 2 was at 2q36 and affected four patients. There were two patients with deletions (CLL043, CLL112) and

two patients with cnLOH (CLL076, CLL095). The MOR was smaller at 0.43Mb and contained the genes *SP140L*, *SP100* and *CAB39* (Figure 5.5b).

Table 5.1. Summary of recurring copy number alterations in a pre-treatment CLL cohort (n=95). CNA – copy number alteration, MOR – minimally overlapping region, cnLOH – copy neutral loss of heterozygosity

Chromosomal Region	Type of CNA	No of samples with CNA	Size of MOR (Mb)	No of genes in MOR	Genes
17p	Deletion/ cnLOH	12	8.1	337	Many inc *TP53*
11q	Deletion/ cnLOH	18	0.2	3	*C11orf65*, *NPAT, ATM*
12	Gain	15	63.9	803	Many
13q14	Deletion/ cnLOH	50	0.007	1	*DLEU2*
2q22-2q24	Deletion/ cnLOH	4	11.3	56	Many
2q36	Deletion/ cnLOH	4	0.4	4	*SP140*L, *SP100*, *CAB39*, *LOC151475*
6q21	Deletion/ cnLOH	4	0.7	2	*PRDM1*, *ATG5*
8p22p21.2	Deletion/ cnLOH	3	3.8	31	Many
8p12	Deletion/ cnLOH	6	1.0	0	-
9p23p22.3	Deletion/ cnLOH	5	2.3	9	*MPDZ*, *NFIB*, *TYRP1* + 6 non-protein coding
10q24.32	Deletion/ cnLOH	4	1.2	42	Many
18p11.21	Deletion/ cnLOH	4	6.6	69	Many
18p11.31	Deletion/ cnLOH	4	4.0	83	Many
19p13.2	Gain	3	0.8	49	Many

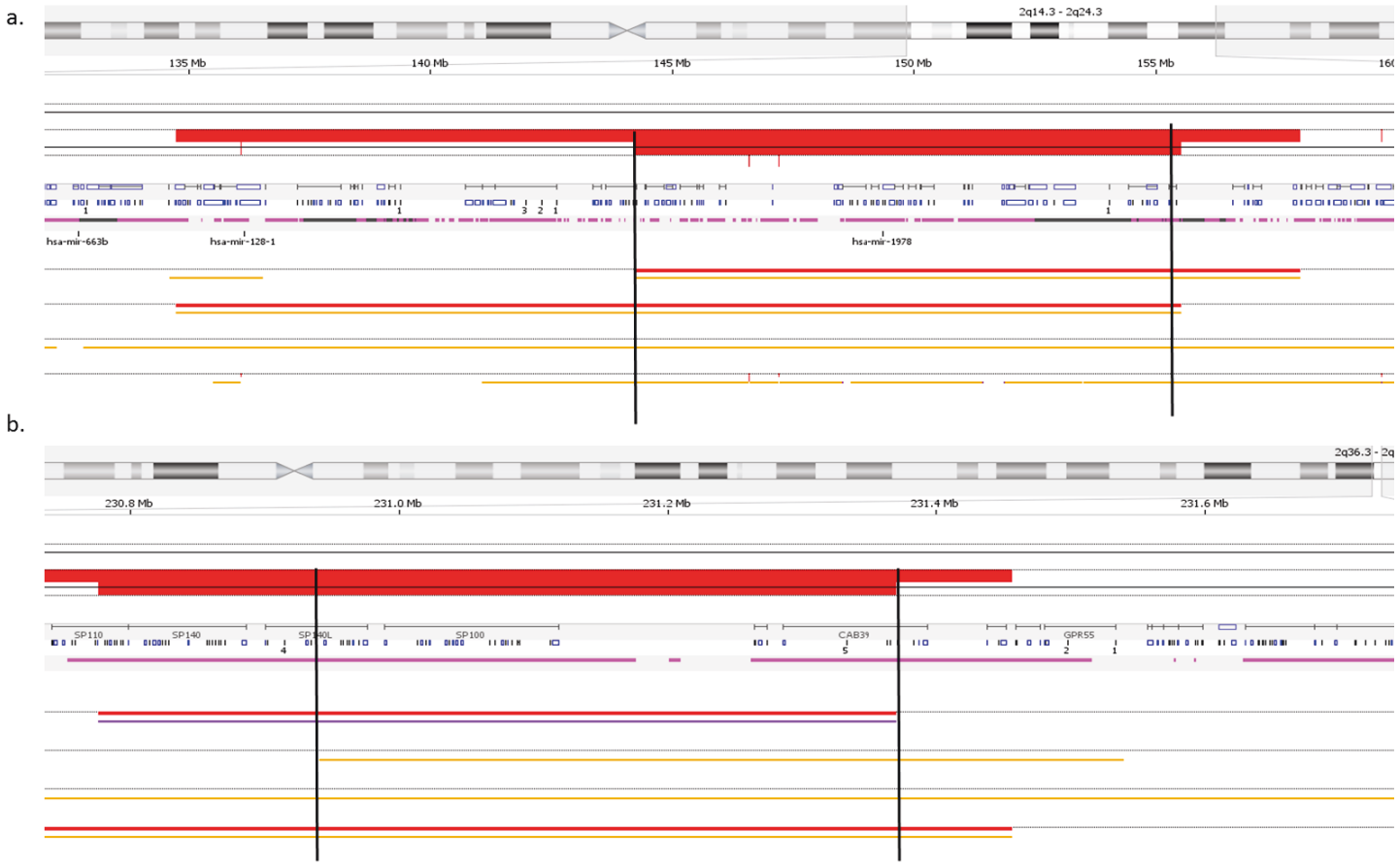


Figure 5.5. Deletions at 2q. Nexus (BioDiscovery) image showing Illumina 1M- Duo data calls for 4 patients at a. 2q22-2q24 and b. 2q36.The MOR boundary is indicated by the black lines. Red - CNA loss, orange – LOH

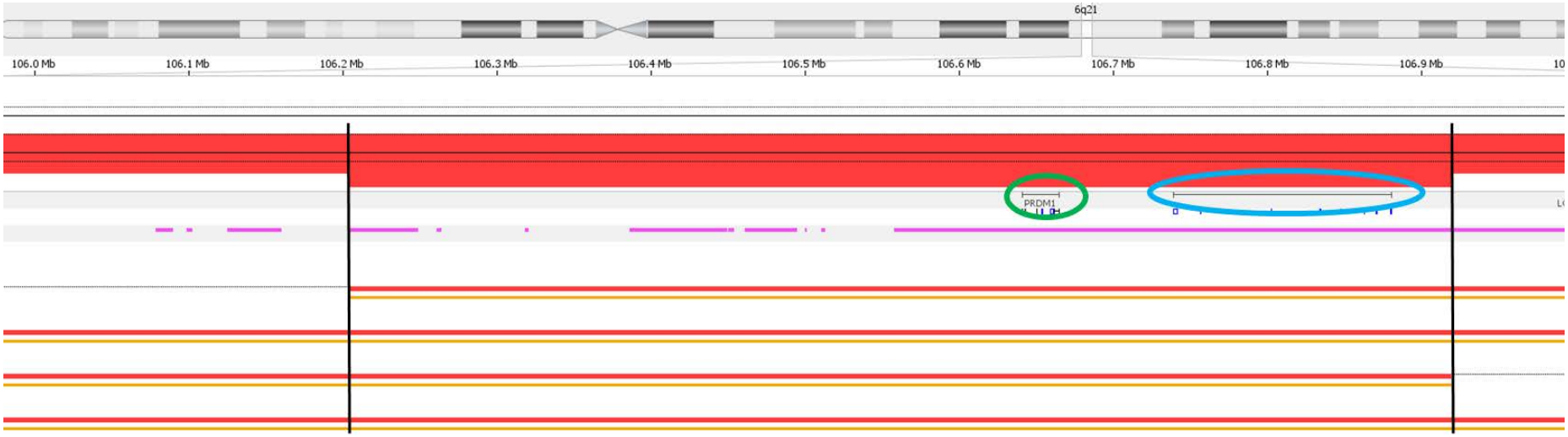


Figure 5.6. The MOR at 6q21 contains two genes; PRDM1 and ATG5. Nexus (BioDiscovery) image showing Illumina 1M- Duo data calls for 4 patients at 6q21.The MOR boundary is indicated by the two black lines. Red - CNA loss, orange – LOH, green circle – PRMD1, blue circle – ATG5

5.3.2.2 Del(6q)

A recurrent deletion at 6q21 was identified in four patients in this cohort (4.2%). The deletions ranged in size from 5.9Mb to 87Mb with an average length of 41.9Mb. The MOR was 0.71Mb and defined by CLL061 and CLL095 (Figure 5.6). There are only two genes within this region; PR Domain 1 (*PRDM1*) and Autophagy related 5 (*ATG5*). *PRMD1* is a regulator of plasma cell differentiation (Shapiro-Shelef *et al* 2003), whereas *ATG5* is an important component of the autophagy pathway. Interestingly an additional patient, CLL083, had a small cnLOH over this site which was just under 0.5Mb (498,480bp).

5.3.2.3 Del(8p)

There are two regions of interest on 8p. The first at 8p22p21.2 was shown by deletions in 3 samples. The 3.8Mb MOR was defined solely by one sample, CLL078 (Figure 5.7a). 31 genes resided at this location including CCR4-NOT Transcription Complex Subunit 7 (*CNOT7*) which negatively regulates cell proliferation (Doidge *et al* 2012) and the tumour suppressor Platelet Derived Growth Factor Receptor Like (*PDGFRL*). The other location on chromosome 8 was at 8p12 and was found in two patients with deletions and four patients with cnLOH. The MOR was 1.0Mb in length and contained no genes although the gene Dual Specificity Phosphatase 26 (*DUPS26*) was located just upstream (Figure 5.7b). This gene has been

shown to have a tumour suppressor role in certain cancers (Patterson *et al* 2010).

a.

b.

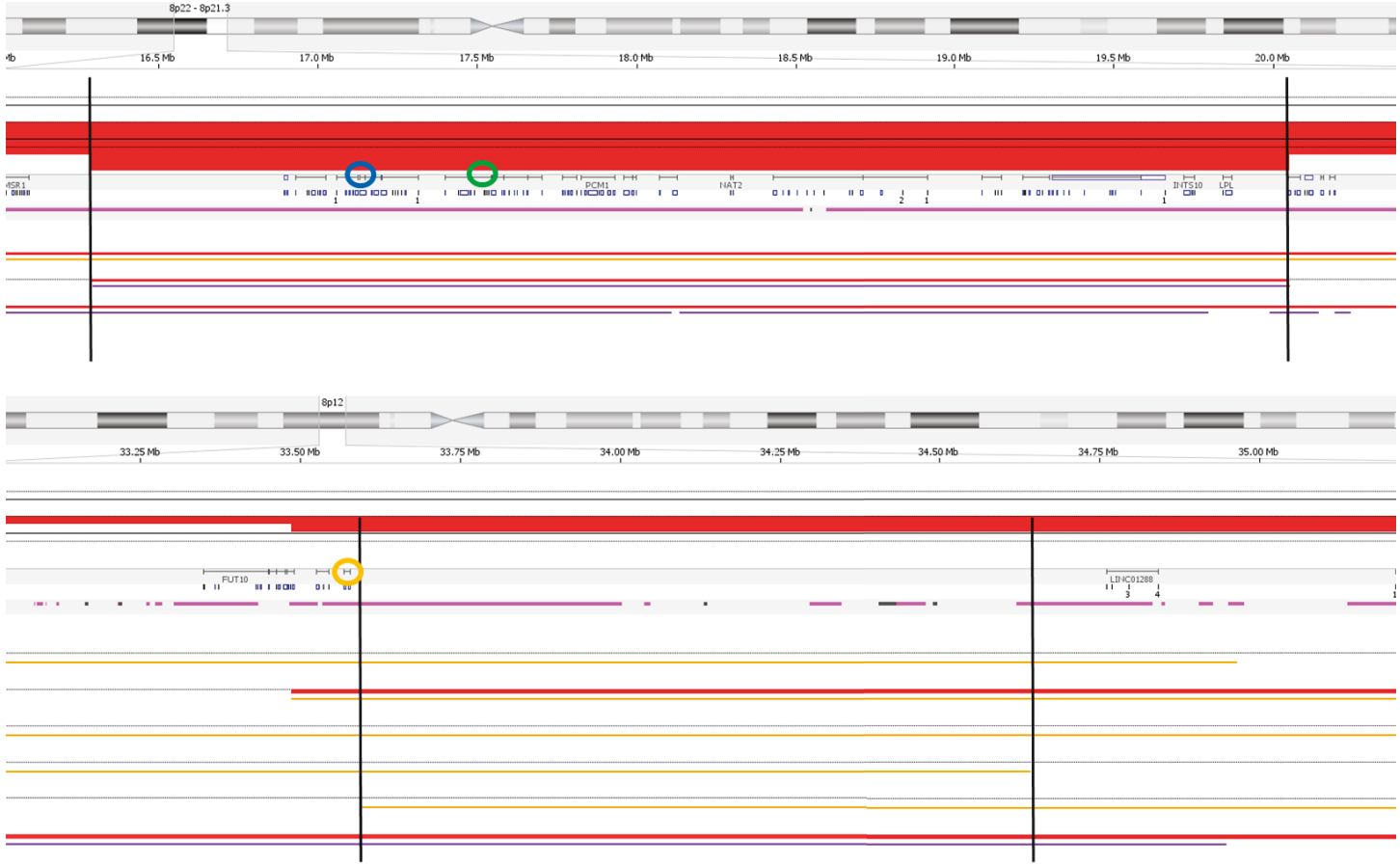


Figure 5.7. Deletions of 8p are recurrent in CLL.. Nexus (BioDiscovery) image showing Illumina 1M- Duo data calls for a. 3 patients at 8p22p21.2 and b. 6 patients at 8p12.The MOR boundary is indicated by the black lines. Red - CNA loss, orange – LOH, green circle – PDGFRL, blue circle – CNOT7, orange circle – DUSP26

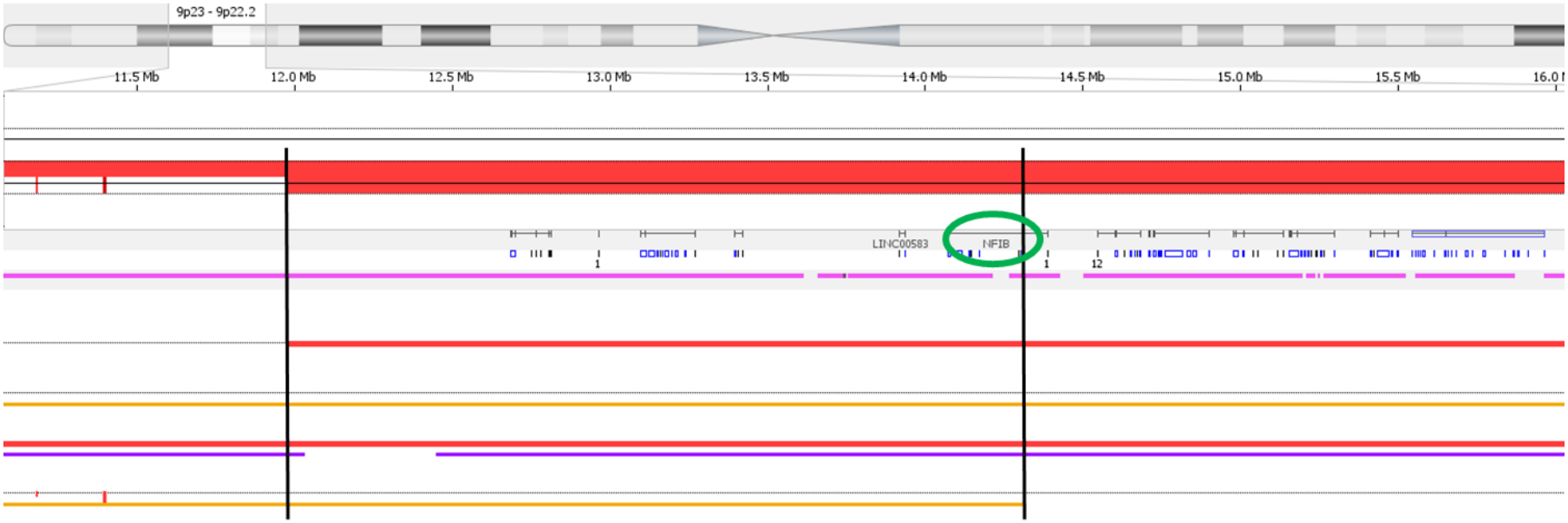


Figure 5.8. Deletions at 9p at recurrent in CLL. Nexus (BioDiscovery) image showing Illumina 1M- Duo data calls for 4 patients at 9p23p22.3. The MOR boundary is indicated by the black lines. Red - CNA loss, orange – LOH, green circle – NFIK

5.3.2.4 Del(9p)

A recurrent deletion was identified at 9p23p22.3. This deletion affected three patients and cnLOH was found in a further two patients (Figure 5.8). The MOR for this region was defined by two samples CLL043 and CLL105 and was 2.3Mb in length. There were three protein coding genes in this region; Tyrosinase Related Protein 1 (*TYRP1*), Multiple PDZ Domain Protein (*MPDZ*) and Nuclear Factor IB (*NFIB*).

5.3.2.5 Del(10q)

Four patients were found to have a deletion on the 10q locus with a further patient having cnLOH. The size of the deletions ranged from 9.1Mb to 1.2Mb, whilst the cnLOH was 2.2Mb in length. The MOR was defined by one patient, CLL041. The 1.2Mb region contained many genes including the transcription factor Nuclear Factor Kappa B Subunit 2 (*NFKB2*).

5.3.2.6 Del(18p)

The MOR at 18p was difficult to define since one sample (CLL106) had a split deletion, therefore two MORs were identified from the same three samples. The length of the first MOR was 6.5Mb and the second was 4.3Mb. Both regions contain several genes (Figure 5.9).

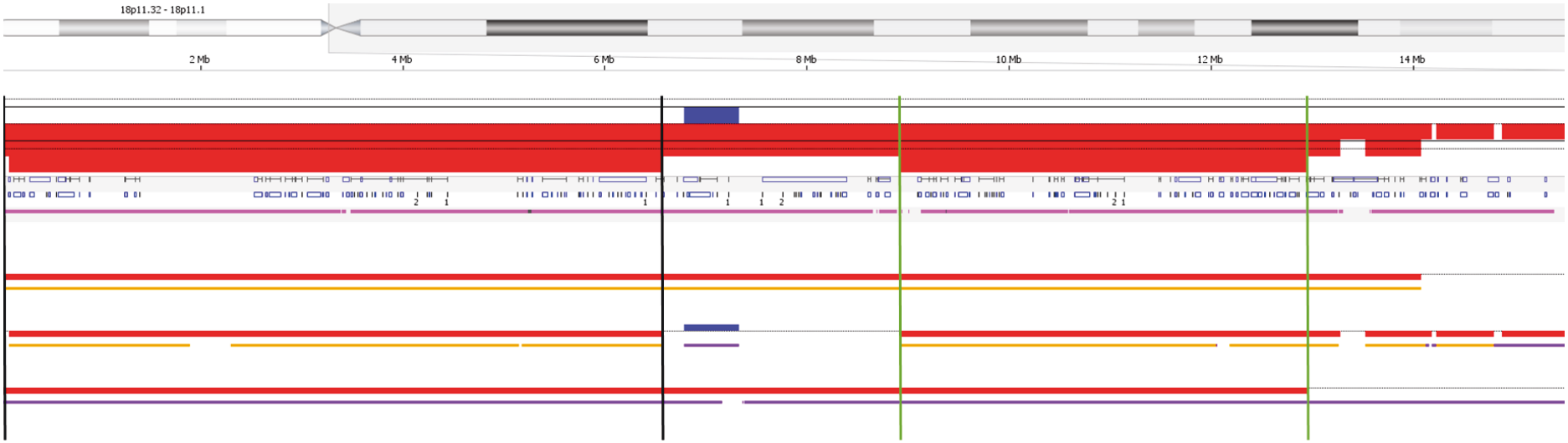


Figure 5.9. Two MORs can be defined at 18p. Nexus (BioDiscovery) image showing Illumina 1M- Duo data calls for 3 patients at 18p. The first MOR boundary is indicated by the black lines and the second by the green lines. Red - CNA loss, orange – LOH.

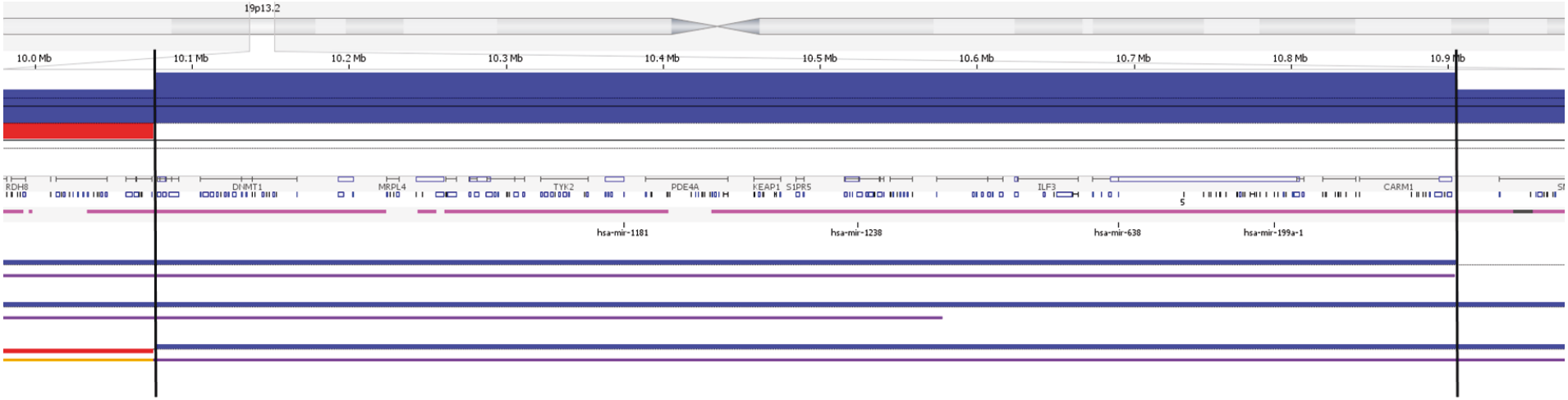


Figure 5.10. Recurrent gain at 19p. Nexus (BioDiscovery) image showing Illumina 1M- Duo data calls for 3 patients at 19p. The first MOR boundary is indicated by the black lines. Blue - CNA gain, purple – allelic imbalance.

5.3.2.7 Gain(19p)

A recurrent gain was identified in three patients on chromosome 19. One patient (CLL048) had trisomy 19, whereas two other patients had smaller gains that were 1.8Mb and 3.6Mb in length. The MOR for this region was 0.8Mb and contained over 40 genes, including DNA methyltransferase 1 (*DNMT1*) and autophagy related 4D cysteine peptidase (*ATG4D*) (Figure 5.10).

5.3.2.8 Other abnormalities

Whilst other previously reported abnormalities were identifiable across the cohort, these did not fulfil the above stipulated criteria (Section 5.3.2) and were not therefore investigated in detail. Gain of 2p (Chapiro *et al* 2010) were found in two patients, for example, whilst the gain(8q) and trisomy 18 (Edelmann *et al* 2013) were only detected in single patients. Additionally, large regions of cnLOH were observed at 20q (Clifford *et al* 2014) in two patients. However as there was only one incidence of a deletion at this site, these were not evaluated further.

5.3.3 Clinical significance of recurrent large copy number alterations is uncertain

A comprehensive validation of these findings using a well-annotated pre-treatment cohort of clinical trial patients (n=250) was performed as part of a separate project in the laboratory. All the 14 CNAs identified in the test

cohort were found in this larger cohort except for the deletion at 8p12. Out of 250 patients, there were 49 incidences of these recurrent abnormalities found in 36 patients (14.4%). For five of the abnormalities, it was possible to further refine the MORs, in some cases substantially reducing the number of genes in the MOR (Table 5.2).

Table 5.2. Summary of the recurrent CNA in a larger pre-treatment cohort. CNA – copy number alteration, MOR – minimally overlapping region.

CNA	No. of samples	Refined MOR	No. of genes in new MOR	Genes
Del(2q22-2q24)	1	Yes	8	*ARHGAP15*, *GTDC1*, *TEX41*, *ZEB2* + others
Del(2q36)	3	No	4	*SP140*L, *SP100*, *CAB39*, *LOC151475*
Del(6q21)	10	No	2	*PRDM1*, *ATG5*
Del(8p22p21.2)	4	No	31	Many
Del(8p12)	0	No	0	-
Del(9p23p22.3)	2	Yes	1	*NFIB*
Del(10q24.32)	5	Yes	4	*AS3MT*, *BORCS7*, *CNNM2*, *NT5C2*
Del(18p11.21)	9	Yes	12	Many
Del(18p11.31)	8	Yes	25	Many
Gain(19p13.2)	7	No	49	Many

Analysis of the TFS in patients with recurrent CNAs found that patients with the del(6q21) (n=10) had the shortest median TFS of 10.6 months whilst patients with the gain(19p) (n=8) had the longest median TFS of 54.0 months compared to 30.5 months for TP53-wildtype patients (n=213) and 12.3 for TP53-mutated patients (n=21). However, this finding was not significant most likely due to the low frequency of the CNAs.

5.3.4 Many pre-treatment abnormalities persist at relapse and other recurrent CNAs emerge

To determine how these abnormalities had evolved at relapse, 42 relapse samples of the test cohort were analysed using SNP arrays. The results are summarised in Table 5.3. Of the 14 abnormalities identified in the pre-treatment cohort, only the 10q deletion and the partial trisomy 12 were no longer present in the paired relapse sample. Validation of these findings in a well-annotated relapse/refractory cohort (n=67) of clinical trial patients confirmed all except for gain(19p) which was not detected in this cohort. As with the pre-treatment cohort the del(6q21) remained the most frequent CNA, after the common abnormalities, found in 9% (6/67) of these samples.

Table 5.3. Presence or loss of copy number alteration at relapse.

Region	Lost at relapse	Retained at relapse	Gained at relapse
Del(17p)	0	5	1
Del(11q)	0	8	1
Tri(12)	1 (loss of partial trisomy)	7	0
Del(13q14)	0	21	0
Del(2q22-2q24)	0	3	0
Del(2q36)	0	2	1
Del(6q21)	0	4	0
Del(8p22p21.2)	0	n/a	1
Del(8p12)	0	2	1
Del(9p23p22.3)	0	2	0
Del(10q24.32)	2	0	2
Del(18p11.21)	0	1	0
Del(18p11.31)	0	1	0
Gain(19p13.2)	0	1	1

5.3.5 Further investigation of the del (6q)

SNP array analysis of diagnosis/relapse pairs identified several CNAs that are recurrent in CLL and which persist at relapse, suggesting that they may be from the original leukaemic clone. Although multiple potentially interesting regions had been identified, the del(6q21) abnormality was

investigated further since, after the common abnormalities, it was the most frequent CNA found in 4.9% (20/411) of samples, it persisted at relapse and patients had a shorter TFS.

The deletion at 6q in this cohort was found to contain just two genes; *PRDM1* and *ATG5*. To determine if additional genetic abnormalities could be found in either gene, exon SSeq was performed on DNA from 40 patients within this cohort including the four patients found to have the 6q deletion. However, no mutations were detected in either gene.

5.3.6 Genes involved in autophagy are recurrently found in MORs in CLL

Since SSeq did not identify any mutations in either *ATG5* or *PRDM1* in a cohort of 40 patients, nor was it possible to refine the MOR further by combining the data from all the studies (n=411), all MORs identified were interrogated by pathway analysis. Using a combination of the KEGG pathway database (Kanehisa *et al* 2016, Ogata *et al* 1999) and literature annotations, a number of other genes were classified as related to the autophagy processes (Table 5.4).

*Table 5.4. Summary of autophagy genes identified in recurrent MORs. Chr -chromosome, CNA – copy number alteration, MOR – minimally overlapping region. * Additional samples had trisomy 12 which affected this location. Others – uncharacterised genes*

Gene	Chr	Function in Autophagy	Type of CNA	No. of samples affected	No. of genes in MOR	Other genes in MOR
ATG4D	19	Protein targeting / autophagosome formation	Gain/ cnLOH	9	14	*AP1M2, CDKN2D, DNM2, ILF3, KRI1, MIR1238, MIR638, QTRT1, S1PR5, SLC44A2 + others*
ATG5	6	Autophagosome formation	Del/ cnLOH	13	2	*PRDM1*
CAMKK2	12	Regulator of autophagy	Del/ cnLOH*	6	3	*ANAPC5, P2RX4*
CDKN1B	12	Regulator of autophagy	Del/ cnLOH*	3	25	*Many*
CDKN2A	9	Regulator of autophagy	Del/ cnLOH	10	7	*CDKN2B, MTAP, Mir_384 + others*
KBTBD6	13	Regulator of autophagy	Del/ cnLOH	42	6	*KBTBD7, Mir_324, MTRF1 + others*

KBTBD7	13	Regulator of autophagy	Del/ cnLOH	42	6	*KBTBD6, Mir_324, MTRF1 + others*
NEK9	14	Autophagosome formation	Del/ cnLOH	29	3	*TMED10, ZC2HC1C*
PRKAG1	12	Regulator of autophagy	Del/ cnLOH*	10	13	*Many*
RB1	13	Regulator of autophagy	Del/ cnLOH	121	3	*LPAR6, RCBTB2*
SESN1	6	Regulator of autophagy	Del/ cnLOH	21	3	*ARMC2, CEP57L1*
STK4	20	Activation of LC3	Del/ cnLOH	10	21	*Many*
STK7 (PKN2)	1	Regulator of autophagy	cnLOH	7	2	*GTF2B*

5.3.7 Screening of relapse samples identifies mutations not present at diagnosis

Analysis of the 42 paired samples showed that 40% (n=17) had no change in the subsequent relapse sample using SNP array analysis. Of these samples, there was sufficient DNA on nine to perform both a targeted NGS gene panel and the NGS-IgHV.

Analysis of these nine samples showed that changes at relapse were detectable in seven (78%) (Table 5.5). For two samples (CLL87 and CLL103) there were no differences between the diagnostic and subsequent relapse sample. Both patients were UM and had trisomy 12. Of the seven patients with changes at relapse, three (CLL098, CLL99 and CLL102) had altered VAFs, implying a shift in the subclonal levels. The other four samples had either lost mutations and/or acquired novel variants suggesting that a subclone had either died out completely or a new subclone had been created. At relapse, CLL088 had lost the IgHV3-15 subclone and gained an *EGR2* variant, although further analysis showed that the *EGR2* deletion was detectable in 1.9% of the reads in the diagnostic sample (depth = 529) and that the IGHV3-15 subclone was present in 0.0019% (5/267,440) of the relapse reads. CLL050 was similar in that the *KRAS* variant was also detectable in the diagnosis sample, at a VAF of 3.6%. In contrast, the IGHV3-30 subclone in CLL089 was completely absent at relapse (read depth = 24,770) and the emerging IGHV1-18 subclone in CLL082 was undetectable in the diagnostic sample even with a depth of 109,804 reads.

Table 5.5. Summary of differences at diagnosis and relapse exclusively of patients without noticeable difference by array

	Diagnosis			**Relapse**	**Comments**
Sample	**Array**	**IgHV**	**NGS Panel**		
CLL050	Del6q14, Del11q12, Del13q14 Tri7p21	UM IGHV3-30	None detected	8.4% KRAS p.L19F	KRAS detectable in D at 3.6%
CLL082	Del13q13 HomDel13q14	M IGHV4-34	None detected	12% UM IGHV1-18	-
CLL087	Tri12, HomDelXp11	UM IGHV4-31	None detected	No changes	-
CLL088	Del13q14	92.2% M IGHV4-34 5.6% UM IGHV3-15	None detected	Loss of IGHV3-15 Gain 5% EGR2 p.A304del	EGR2 detectable in D at 1.9%
CLL089	Tri1q21, Tri1q25	94.6% UM IGHV3-48 2.9% M IGHV3-30	18.3% XPO p.E571Q	Loss of IGHV3-30 XPO increased to 42.7%	-
CLL098	None detected	UM IGHV4-4	7.9% NOTCH1 p.P2514Rfs*4 10% NOTCH1 p.Q2406*	NOTCH1 p.P2514Rfs*4 decreased to 2.6% NOTCH1 p.Q2406* increased to 24.8%	-
CLL099	Tri1q21	UM IGHV3-33	18.1% RSP15 p.P131T	RSP15 decreased to 2.2%	
CLL102	Tri12		40.3% IRF4 p.N102K	IRF4 increased to 47.6%	
CLL103	Tri1q21, cnLOH2q22, cnLOH9p24, Tri10q26, cnLOH13q33, Tri12, Tri22q11	UM IGHV3-11	29.7% NOTCH1 p.H2106Y	No changes	

5.4 Discussion

The data in this chapter demonstrates the validity of using high-resolution SNP arrays to characterise and quantify copy-number aberrations in CLL. The analysis reveals several recurrent, and therefore non-random, CNAs within the cohort, some of which persist into relapse. Furthermore, the refining and characterisation of the MORs of these CNAs has enabled candidate genes and pathways involved in leukaemogenesis and disease progression to be identified.

5.4.1 SNP arrays can detect CNAs and cnLOH in CLL

The frequency of the common mutations detected in the test cohort was in keeping with previous studies, with the 13q14 deletion being the most common aberration detected in the test cohort, found in 52.6% of samples (Gunnarsson *et al* 2010, Ouillette, Collins, Shakhan, Li, Peres, *et al* 2011, Pfeifer *et al* 2006). MOR analysis of these regions were also consistent with previous studies with both the heterozygous and homozygous 13q14 deletions containing part of the DLEU2 gene, as well as the 11q and 17p deletions MOR mapping to the ATM and TP53 genes, respectively (Edelmann *et al* 2013).

Analysis of other recurrent abnormalities was limited to those present in at least three samples, with a CNA in at least two of the samples. cnLOH in

isolation was excluded since studies using paired tumour and germline DNA samples have found that cnLOH in the absence of an associated CNA are often germline in origin (Edelmann *et al* 2013, Pfeifer *et al* 2006). The CNAs detected in the test cohort using this criteria had all been previously reported (Dohner *et al* 2000, Edelmann *et al* 2013, Gunnarsson *et al* 2011, Ouillette, Collins, Shakhan, Li, Peres, *et al* 2011). However other recurrently reported abnormalities such as the gain(2p) and trisomy 18, were excluded as they were present at low frequencies in this cohort (<3%).

Analysis of the MORs in an extended cohort of 411 samples identified several interesting genes that may have a role in leukaemogenesis. *PRMD1* and *ATG5* were the only genes in the MOR at del(6q) whilst the MOR at 2p36 contained the genes *SP100* and *SP140L,* and *NFIB* was found to be the sole gene in the MOR of the del(9p) abnormality. *NFIB* is a transcription factor that has been reported to be an oncogene (Dooley *et al* 2011, Zhang *et al* 2015) and has been found to have a role in megakaryocyte differentiation (Lu Chen *et al* 2014, Rice *et al* 2011). *SP100* and *SP140L* have previously been implicated in familial CLL (Di Bernardo *et al* 2008). Whilst these are all interesting genes for further investigation, the del(6q) was by far the most frequent CNA found in 4.9% (20/411) of samples, compared to 2.2% (9/411) for the del(2p36) and 1.9% (8/411) for the del(9p).

5.4.2 Genes required for leukaemogenesis are present in relapse samples

To identify possible genetic drivers of tumour development and relapse, paired pre-treatment and relapse samples were analysed. Previous studies using paired samples have focussed on clonal evolution in the context of other clinical markers such as IgHV mutation status, as opposed to the implications of disease drivers (Gunnarsson *et al* 2010) or solely on the common abnormalities (Ouillette *et al* 2013). Therefore, this study concentrated on identifying genetic abnormalities that were present at relapse since, although it is possible that mutations that initiate leukaemogenesis may not be necessary for the subsequent relapse, if they were present in the pre-leukaemic clone it would be assumed that they would be present in all subclones and in the relapse sample.

5.4.3 Screening of relapse samples identifies mutations not present at diagnosis

Three of the nine samples that underwent a full screen of SNP array, targeted NGS and NGS-IgHV, were found to have acquired novel subclones as determined by the detection of a new variant. In two of these samples, deep sequencing analysis of the novel gene variants found that these were present at diagnosis suggesting that treatment had selected for them. Interestingly the third sample had gained a subclone with an UM IgHV

which was not detected at diagnosis. Attainment of an IgHV-UM subclone in a IgHV-M patient has previously been reported (Rose-Zerilli *et al* 2016); however in this case it was possible to detect other genetic lesions that were associated with the UM subclone in the diagnostic sample, suggesting that the UM subclone was present at diagnosis but undetectable using conventional SSeq methodology. In contrast, it was not possible to detect the emerging clone in the diagnostic sample in the study described in this chapter despite obtaining over 100,000 reads. Whilst it is well recognised that CLL samples can acquire novel variants through random mutagenesis (Amin *et al* 2016, Landau *et al* 2015, Ouillette *et al* 2013, Schuh *et al* 2012), the acquisition of previously undetectable IgHV subclone would support the theory that the susceptibility for a cell to progress to CLL is present early in B-cell development (Kikushige *et al* 2011).

5.4.4 Genes in the autophagy pathway are recurrently mutated using SNP array analysis

The del(6q) has been recurrently reported in CLL (Dohner *et al* 2000, Gunnarsson *et al* 2011, Ouillette, Collins, Shakhan, Li, Peres, *et al* 2011, Stilgenbauer *et al* 1999) and was persistent in the relapse samples. The work described in this chapter refines, for the first time, the del(6q) MOR to the *ATG5* and *PRMD1* genes. *PRDM1*, as a B-cell regulator of differentiation, seemed the most likely target, especially since it has previously been reported to act as a tumour suppressor in other lymphoid

malignancies (Hangaishi and Kurokawa 2010). However, downregulation of the autophagy pathway has also been implicated in the myeloid leukaemias (Watson *et al* 2011). SSeq mutation screening of both genes did not reveal any variants in a cohort of 40 CLL patients, however genetic mutations in CLL are often described at a low frequency so it is possible that the sample size was too small. Nevertheless, WES of a cohort of 278 patients only found one variant in *PRDM1* and none in *ATG5* (Landau *et al* 2015).

Extended analysis in a cohort of 411 patients was unable to further refine the del(6q) MOR. However, several key genes involved in the autophagy pathway were found to be located within recurrent CNAs, suggesting that the autophagic pathway may be altered in CLL cells. Furthermore, *MYD88*, which is mutated in around 5% of CLL cases (Puente *et al* 2011, Lili Wang *et al* 2011) has been shown to trigger autophagy as part of antigen presentation (Shi and Kehrl 2008). The expression of *PRDM1*, the other potential target of the del(6q), has been shown to be regulated in plasma cells through autophagy (Pengo *et al* 2013). Together, this suggests that autophagy may be a significant role in the pathogenesis of CLL.

5.4.5 Conclusions

In conclusion, the work described in this chapter provides further support for an early B-cell as the CLL leukaemic stem cell. Furthermore, the refining

of the del(6q) MOR and the identification of multiple autophagy genes in other MORs has elucidated a potential new mechanism for CLL pathogenesis through the autophagy pathway. Further work is necessary to determine the role of autophagy in CLL.

Chapter 6 Analysis of autophagy in B-CLL samples reveals increased autophagy activity when compared to normal age-matched B-cells

6.1 Introduction

The work described in the previous chapter showed that in patients with CLL there are recurring abnormalities in genes involved in the autophagy process. Autophagy is a mechanism of degradation whereby redundant or defective cellular components are disassembled and recycled. Autophagy has been shown to have a dual role in cancer in that it can act both as a tumour suppressor and a tumour promoter. Impaired DNA repair because of the downregulation of autophagy can promote tumorigenesis. Conversely, an increase in autophagy activity can support tumour cells through the provision of energy and nutrients (Eskelinen 2011).

Since the autophagy pathway interacts with many genes that have implications in CLL, it is therefore possible that autophagy may be involved in the pathogenesis of CLL. The aim of the study in this chapter was to further investigate the role of the autophagy process in CLL.

6.1.1 The Autophagy Pathway

Macroautophagy (hereafter referred to as autophagy) is the highly-conserved process of degradation of cytoplasmic components in a double membraned structure called the autophagosome. The autophagosome encapsulates superfluous proteins targeted for destruction, then fuses with the lysosome and degradation takes place. Basal autophagy plays an important role in maintaining normal cellular haemostasis through the disposal of misfolded or damaged proteins and organelles, and protein aggregates. Autophagy also acts as a survival mechanism and is up-regulated in response to cellular stresses such as starvation, to provide additional nutrients through the recycling of cytosolic components. Additionally, autophagy has an immunological function through the elimination of microbes as well as a role in antigen presentation.

6.1.1.1 Formation of the Autolysosome

Autophagy is a multistep process. It starts with nucleation whereby the necessary factors are recruited to form the phagophore. Once formed, the phagophore is elongated to encapsulate the factors targeted for degradation. When the autophagosome has matured and sealed, it is then fused to the lysosome. The lysosomal components are sequestered to the now autolysosome and the elements within degraded (Figure 6.1) (Choi *et al* 2013).

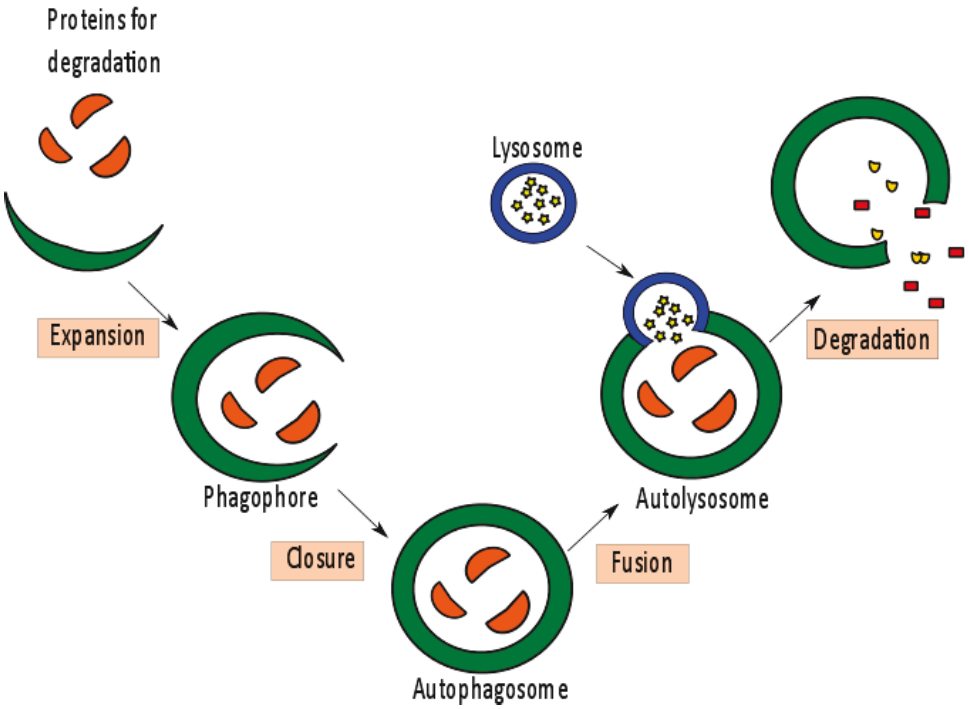


Figure 6.1 The autophagy pathway. Proteins targeted for degradation are encapsulated by the autophagosome, which then fuses to the lysosome. The lysosomal factors are released into the autolysosome and the proteins degraded. Adapted from Choi *et al.* (2013).

6.1.1.1.1 Nucleation

Classical induction of autophagy is signalled through the serine/ threonine kinase mTOR. m-TOR resides inside the mTORC1 complex, alongside raptor, mLST8, PRAS40 and deptor. When activated, the MTORC1 complex suppresses the ULK1 complex which is composed of ULK1/2, FIP200, ATG13 and ATG101 (Hara *et al* 2008, Hosokawa, Hara, *et al* 2009, Hosokawa, Sasaki, *et al* 2009) as well as the beclin1 regulator, Ambra1 (Nazio *et al* 2013). Upon detachment from MTORC1, ULK1 is activated and is able to autophosphorylate itself and phosphorylate FIP200 and ATG13 (Ganley *et al* 2009). Stimulation of the ULK complex initiates the nucleation process of autophagy by activating the PI3K complex.

In mammalian cells the PI3k complex is formed from the class III phosphoinositide 3-kinase Vps34 alongside p150 (Volinia *et al* 1995), Barkor (Sun *et al* 2008), and beclin1. Barkor and p150 act as chaperones directing the complex to the site of autophagosome production while Vsp34 produces phosphatidylinositol-3-phosphate (PI-3-P) which acts as a marker of the site. The complex recruits WIPI1 (Proikas-Cezanne *et al* 2004), WIPI2 (Polson *et al* 2010) and DFCP1. DFCP1 is a key component of the omegasome (Axe *et al* 2008), a PI-3-P-enriched ER subdomain which acts as a conduit for the formation of the phagophore or isolation membrane (IM) as it is also known (Hayashi-Nishino *et al* 2009, Uemura *et al* 2014, Ylä-Anttila *et al* 2009). Although the site of autophagosome formation is yet to be elucidated, the most favoured model is that development takes place at the ER-mitochondria contact site (Anao *et al* 2015, Hamasaki *et al* 2013).

6.1.1.1.2 Expansion and Closure

The formation and elongation of the phagophore requires two ubiquitin-like systems. In the first system, the ATG16L complex is formed. The initial step in the production of the ATG16L complex is the activation of ATG12 by the E1 enzyme ATG7. This is then conjugated to ATG5 by the E2 enzyme ATG10 (Mizushima *et al* 1998). ATG16L then binds to ATG5, through its N-terminal domain, and to other ATG12-ATG5-conjugated ATG16L through its coil-coiled region, to form a homo-oligomer (Mizushima *et al* 2003). The

complex is then recruited to the autophagosome formation site by WIPI2 (Dooley *et al* 2014) where it acts as an E3 enzyme by recruiting and activating the LC3-conjugated E2-enzyme ATG3 (Fujita *et al* 2008, Sakoh-Nakatogawa *et al* 2013).

In the second system, the ATG8 homologues; the microtubule-associated protein light chain 3 (LC3) subfamilies, are conjugated to phosphatidylethanolamide (PE). ATG3, alongside ATG7, converts LC3-I to LC3-II (Kabeya *et al* 2004). The ATG16L complex also lipidates the other ATG8 homologues; the Golgi-associated ATPase enhancer of 16 kDa (GATE-16)/γ-aminobutyric acid (GABA)-receptor- associated protein (GABARAP) subfamilies. The lipidated LC3s and GATE16/GABARAPs are then recruited to the phagophore membrane. Both subfamilies have been shown to be essential for the formation of the autophagosome with LC3s acting at the early stages of development and GATE16/GABARAPs having a role in phagophore maturation (Weidberg *et al* 2010).

The phagophore membrane expands through the fusion of lipids which is mediated by LC3 through the facilitation of hemifusion events. This is when only the outer layer of two membranes fuse together. The lipids can be obtained from either the ER (Hayashi-Nishino *et al* 2009, Ylä-Anttila *et al* 2009), the mitochondria (Hailey *et al* 2010), the cell membrane (Ravikumar *et al* 2010) and the Golgi via the ER-Golgi-intermediate complex (ERGIC)

(Ge *et al* 2013), or from a combination of these depending on the autophagy stimulus and availability of the organelles (Tooze and Yoshimori 2010).

The mechanisms surrounding autophagosome closure are still unknown however it is thought that the GABARAPs act to dissociate the ATG16L complex from the autophagosome (Weidberg *et al* 2010) as it is not found on mature autophagosomes (Mizushima *et al* 2003). Additionally it is known that both ATG2 and ATG9 are essential for sealing of the membrane and defects in either prevent the autophagosome from closing (Corcelle-termeau *et al* 2016, Velikkakath *et al* 2012).

6.1.1.1.3 Fusion

Once sealed, the autophagosome needs to be transported to the microtubule organising centre (MTOC) where it is fused to the lysosome. To achieve this the autophagosome moves along microtubule tracks assisted by the small GTPase Rab7 and its effectors; Rab-interacting lysosomal protein (RILP) (Jordens *et al* 2001) and FYCO1 (Pankiv *et al* 2010).

Fusion to the lysosome is co-ordinated by Rab GTPases, particularly Rab7, facilitated by PLEKHM1 which simultaneously binds Rab7, LC3 and the homotypic fusion and vacuolar protein sorting (HOPS) complex (McEwan *et*

al 2015). The HOPS complex, a multimer of six vacuolar protein sorting proteins (Balderhaar and Ungermann 2013), has two functions; firstly it acts as a tethering molecule to position the necessary compartments intended for fusion into close proximity and secondly it primes the SNARE proteins for fusion (Jiang *et al* 2014, Wartosch *et al* 2015). The autophagosome-bound SNARE protein STX17, in complex with the SNAP29 SNARE and stabilised by Barkor (Diao *et al* 2015), attaches to the lysosomal-bound SNARE VAMP8 enabling fusion between the autophagosome and the lysosome (Itakura *et al* 2012) to produce the autolysosome.

6.1.2 Role of Autophagy in Cancer

6.1.2.1 Autophagy as a Tumour Suppressor

Autophagy can act as a tumour suppressor by removing damaged proteins which can increase levels of oxidative stress leading to DNA damage. There is also cross-talk between autophagy with the apoptotic pathway. Autophagy has been shown to prime cells for programmed cell death (Young *et al* 2012) and may be used as an alternative death pathway for cells with impaired apoptosis (Y Zhao *et al* 2010).

Deficiencies of key autophagy genes has been described in several tumours further implying a tumour suppressor role for autophagy. *BECN1* which codes for beclin1, is recurrently mono-allelically deleted in several

carcinomas (Aita *et al* 1999) and loss of the up-regulators of *BECN1*, *Ambra1* and *UVRAG*, have also been shown to increase proliferation in tumour cells (Fimia *et al* 2007, Liang *et al* 2006). Absence of the ATG5 protein is associated with shorter PFS in melanoma (He Liu *et al* 2013) and mouse knockouts for *ATG5* and *ATG7* have been shown to develop diseases symptomatic of myelodysplastic syndrome and acute myeloid leukaemia (Mortensen *et al* 2011).

Deficiency in the autophagy process leads to an increase in defective proteins and organelles which can stimulate oxidative stress through the production of reactive oxygen species (ROS). Additionally a decrease in autophagy activity results in the build-up of p62/SQSTM1, an adaptor protein used to deliver damaged cellular components to the autophagosome for degradation (Pankiv *et al* 2007). Accumulation of p62 has been shown to lead to the production of ROS which in turn causes DNA damage. Furthermore, increased p62 has also been linked to dysregulation of the NF-κB pathway (Mathew *et al* 2009).

6.1.2.2 Autophagy as a Tumour Promoter

Upregulation of autophagy can provide a survival mechanism for tumour cells. Indeed inhibition of autophagy has been demonstrated to increase chemosensitivity and cell death in many carcinomas (Koukourakis *et al* 2015, Levy *et al* 2014, Ma *et al* 2011, Ojha, Singh *et al* 2014, Zhao *et al*

2014). Increased autophagy can also provide nutrients to sustain tumour survival particularly during periods of cellular stress (Guo *et al* 2011) and tumour cells have been shown to be dependent on autophagy for the removal of ROS (Yang *et al* 2011).

Apoptosis and autophagy are activated by similar triggers and can be stimulated simultaneously, sequentially or independently. Autophagy has been shown to prevent tumour cells from undergoing apoptosis by degrading key apoptotic factors (Hou *et al* 2010, Wang *et al* 2015). Moreover the inhibition of autophagy has been shown to initiate apoptosis suggesting a tumour promoter role (Boya *et al* 2005).

6.1.3 Autophagy in CLL

The role of autophagy in CLL is largely unknown. It has been suggested that the basal autophagy level in primary CLL cells provides a survival mechanism against cell death (El-Khoury *et al* 2014). Up-regulation of autophagy in CLL in response to cellular stress has also previously been described and demonstrated to promote cell survival (Kovaleva *et al* 2012, Mahoney *et al* 2012). This is consistent with studies that demonstrate that the inhibition of autophagy decreases CLL cell viability (Amrein *et al* 2011, Mahoney *et al* 2013). Furthermore two studies have established that one mechanism of action of the histone deacetylase inhibitor drugs, which have been shown to cause cytotoxicity in CLL cells, is to block the autophagy

pathway (El-Khoury *et al* 2014, MacCallum *et al* 2013). In contrast, a recent study has shown that CLL cells with impaired autophagy have a marked resistance to fludarabine (Bologna *et al* 2016), although this contradicts a previous study which suggests that autophagy provides little protection from fludarabine (Mahoney *et al* 2012).

Genes involved in the P13K complex such as *BECN1* have been showed to have reduced expression levels in CLL when compared with normal B-cells (Kristensen *et al* 2015). Additionally, the whole genome screening approaches described in the previous chapter have identified recurrent mutations in key autophagy genes such as *ATG5* and *ATG4D*. Together this data suggests that there may be a role for autophagy in CLL pathogenesis.

6.2 Materials and Methods

6.2.1 Quantitative real-time PCR of autophagy genes

cDNA from 11 CLL samples (from population C, Appendix 2) and from sorted B cells (Section 2.1.3) from healthy donors was used in this analysis. Real time PCR was performed using Applied Biosystems Taqman Gene Expression assay kits for ATG5 (Hs00169468_m1), Beclin1 (Hs00186838_m1) and GAPDH (Hs99999905_m1). Samples of cDNA (2ul) were made up into 10ul reaction mixes using Taqman MasterMix (4304437, Thermo Fisher Scienitific) and the corresponding Taqman Gene Expression probe. Detection was performed using the Rotorgene 6000 (Qiagen). The PCR cycling conditions started with a 2min incubation at 50°C, then an initial denaturation of 95°C for 10mins followed by 50 cycles of 95°C for 15sec and 60°C for 1min. Quantification of the of the samples' transcripts was performed using the Comparative Quantification method of the Rotor Gene 6.0 software. In this method, the Ct is defined as the point at which the exponential phase of amplification begins and the reaction efficiency is calculated between this point and the maximum rate of exponential amplification. From this the level of individual transcripts relative to the reference sample can be calculated using the formula:

$$\text{Relative Quantity} = \text{Efficiency}^{\text{ReferenceCt-SampleCt}}$$

From this the relative expression ratios of *BECN1* and *ATG5* to *GAPDH* for each sample can be determined (Mccurdy *et al* 2008).

6.2.2 Preparation of Cells for Autophagy Analysis

Cells from 8 CLL patients (from population C, appendix 2) and 4 age-matched healthy donors were collected and the PBMC isolated and stored (Section 2.1.1). FACS buffer contained PBS with 1% FBS (12003C, Sigma Aldrich) and 0.05% sodium azide (71289, Sigma Aldrich). Cells were thawed (Section 2.1.2), then incubated either in RPMI (11875-085, Thermo Fisher Scientific) with 10% FBS with or without 10ug/ml of each of the lysosomal inhibitors E64d (E8640, Sigma) and Pepstatin A (P5318, Sigma) or in HBSS (14060, Invitrogen) with or without E64D and Pepstatin A for 2hrs at 37°C. The cells were then harvested and immunolabelled with human anti-CD5 (1:100 FACS buffer, PE, 561897, BD Biosciences, San Jose, California, USA) and anti-CD19 (1:100 FACS buffer, Pacific blue, MHCD1928, Invitrogen, Carlsbad, California, USA). To prepare the cells for autophagy analysis using the ImageStream (Amnis, Seattle, Washington, USA), the cells were stained with the lysosomal marker Lyso-ID Red (1:1000 assay buffer, ENZ-51005-500, Enzo Life Sciences, Farmingdale, New York, USA) then incubated with fixative (00-8222-49, eBiosciences, San Diego, California, USA) and washed with permeabilisation buffer (008333-56, eBiosciences). A mouse anti-LC3 (1:1000 permeabilisation buffer, 0231-100/LC3-5F10, Nanotools, Teningen, Germany) was then added followed by immunolabelling with either alexa fluor 488 goat anti-mouse IgG (H+L) (1:500 permeabilisation buffer, Invitrogen) or alexa fluor 546 goat anti-mouse IgG (H+L) (1:500 permeabilisation buffer, Invitrogen). For the caspase-3 analysis, cells were immunolabelled with FITC-conjugated rabbit anti-active caspase-3 (1:20

FACS buffer, BD Biosciences), post permeablisation. A small fraction of each sample was separately stained with LIVE/DEAD (1:2000 PBS, Invitrogen). Cells were then resuspended in 50ul FACS buffer.

6.2.3 Imagestream

Samples were processed on the Amnis ImageStream using the integrated INSPIRE software. First, the ImageStream was calibrated using speed beads (400040, Amnis) according to the manufacturer's instructions. Next unlabelled and single colour fluorescent controls for each of the antibodies were run and finally the experimental samples were processed last. Acquisition started once the core stream had stabilised. At least 20000 cells per sample were collected and the cell classifier field had the area lower limit set to 25. Each sample generated a raw data file.

6.2.3.1 Imagestream Analysis

Following image acquisition, cells were analysed using the IDEAS 6.1 software package (Amnis). A compensation file was generated from the single colour control files. Doublet cells and out-of-focus cells were discarded by gating out bright field area vs aspect ratio and bright field gradient root mean square > 300 respectively. Cells were gated on CD19 and CD5 double positives in the CLL samples and on the CD19 positives for the normal controls.

These subsets were then gated on $LC3^{+}Lyso^{+}$. Autophagy activity in this study was based on the co-localisation of LC3 and lysosome, which occurs during autolysosome formation. The autolysosome formation levels were calculated based on the percentage of co-localisation bright detail similarity (BDS) of LC3 and lysosome in the $LC3^{+}Lyso^{+}$ subset. BDS is derived from the log-transformed, non-mean normalised Pearson's correlation coefficient (r) calculated from the pixel intensity of the fluorescence of the two markers of interest. The r value determines the similarity in the fluorescence of the two markers on a scale of 0 (no correlation) to 1 (complete correlation). Since r is a linear value it can lead to compression at the high end of the range. This is overcome by log-transforming the r value to give the BDS value (Beum *et al*. 2006; Phadwal *et al*. 2012). The BDS of LC3 and lysosome was ascertained for each sample analysed on the Imagestream and both the mean BDS and the percentage of cells gated as BDS^{hi} were used to establish the autophagy activity. A previous autophagy study using the Imagestream described bimodal histograms when analysing the BDS in the cells of interest, enabling the second peak to be termed BDS^{hi} (Phadwal *et al*. 2012). In contrast, the histograms obtained in this study had a more normal distribution pattern, therefore BDS^{hi} was defined as the region to the right of the curve. To enable comparison between samples the average of the peak for all the CLL samples was calculated as 1.66. Cells with a BDS greater than this value were defined as BDS^{hi} and this was applied to all the samples analysed. Both the mean BDS and the

proportion of cells achieving BDS^{hi} were determined for all samples to determine the autophagy activity.

6.2.4 Statistical Analysis

Statistical analysis was performed using Graphpad Prism 5.0. p-values were calculated using 2-tailed independent sample t-test and 2-tailed Mann-Whitney U test.

6.3 Results

6.3.1 Quantitative PCR

To determine if the copy number loss identified by the array over the *ATG5* gene had an effect, qPCR was used to quantify gene expression of the *ATG5* transcript in 11 CLL samples and B-cells isolated from four healthy donors. One of the B-cell controls was used as a calibrator. Overall comparison showed there was no significant change in *ATG5* expression levels when compared to *GAPDH* (Figure 6.2a). qPCR analysis of the *BECN1* transcript however, showed almost a 50% decrease in the level of expression when normalised to *GAPDH* (Figure 6.2b).

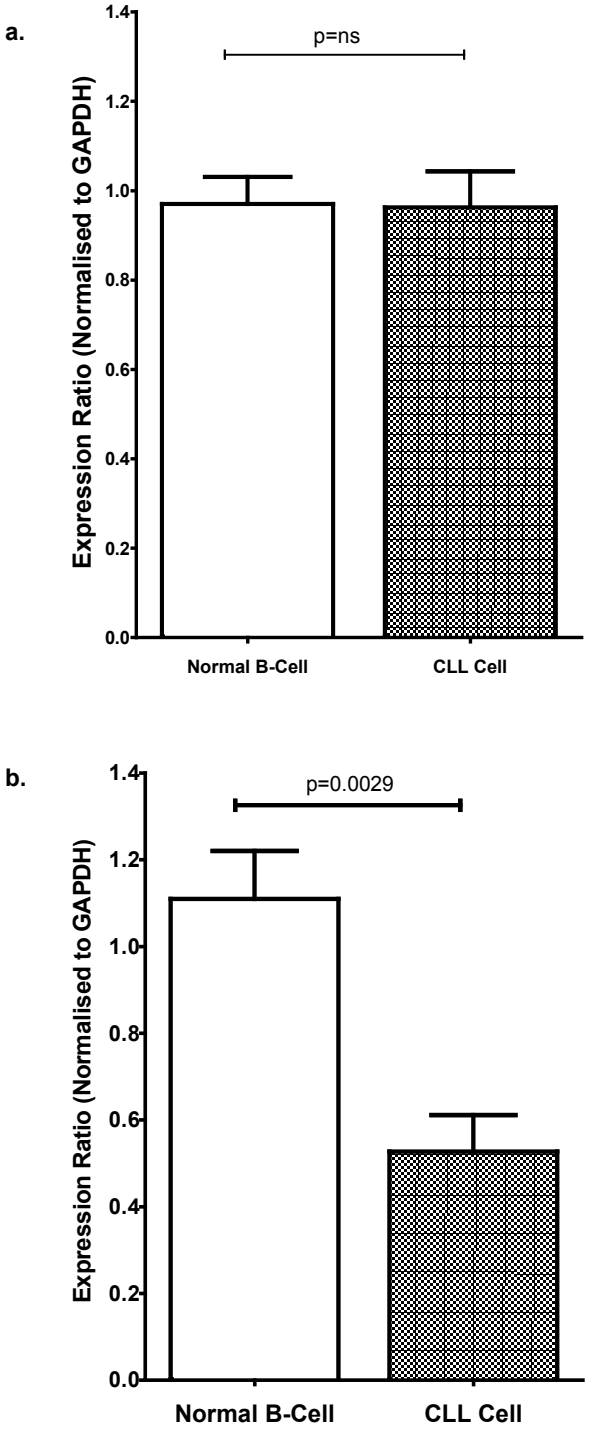


Figure 6.2. ATG5 and BECN1 expression ratio in normal B-cells and CLL cells. qPCR of cDNA from 11 CLL samples and B-cells isolated from 4 healthy donors for a. ATG5 and b.BECN1 (mean ± SEM, [t-Test]), ns – not significant

6.3.2 Analysis of the autophagy levels using the Amnis Imagestream

The Amnis Imagestream is an imaging flow cytometer that enables the simultaneous assessment of morphological characteristics alongside fluorescence signals in large numbers of cells. Furthermore, the location and intensity of the fluorescence signals can be determined and quantified. It is highly suited for assessing autophagy levels in primary cells as it is able to calculate co-localisation of autophagosomal markers whilst identifying a subpopulation of cells by their surface markers (Figure 6.3a). Assessment of autophagy in this study is based on the co-localisation of LC3 and lysosome which characterises autolysosome formation (Figure 6.3b). In addition to analysing the basal autophagy levels, analysis of co-localisation in the presence of the lysosomal inhibitors E64D and Pepsatin A were also calculated. The inhibitors prevent the fusion of lysosome to the autophagosome and halts the autophagic process. This therefore implies the degree of autolysosomal formation at the period the inhibitors were added.

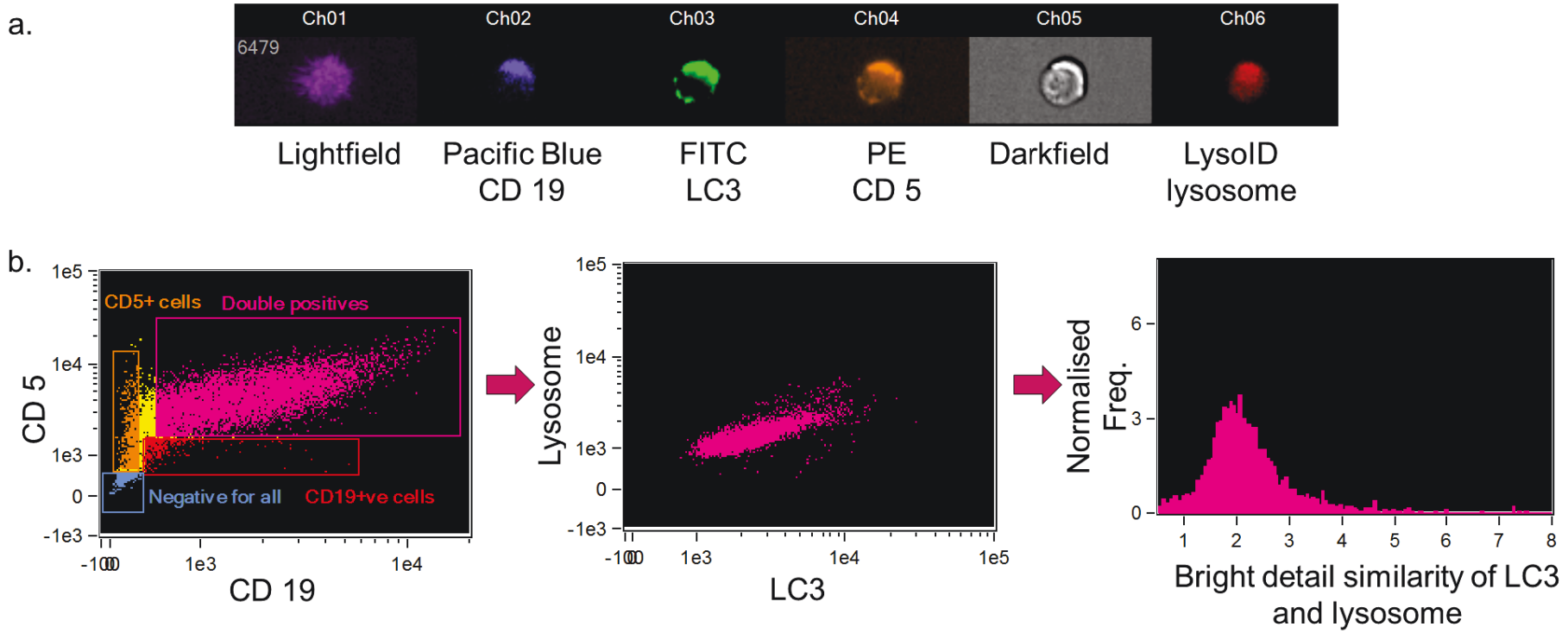


Figure 6.3. ImageStream analysis of autophagy. a. ImageStream images of one cell for each channel. b. Analysis pathway for determining autophagy levels in CLL cells. Cells are first gated on $CD5^{+}CD19^{+}$ double positives. The co-localisation of lysosome and LC3 is then determined in this gated population. Finally, the distribution of co-localisation is demonstrated graphically.

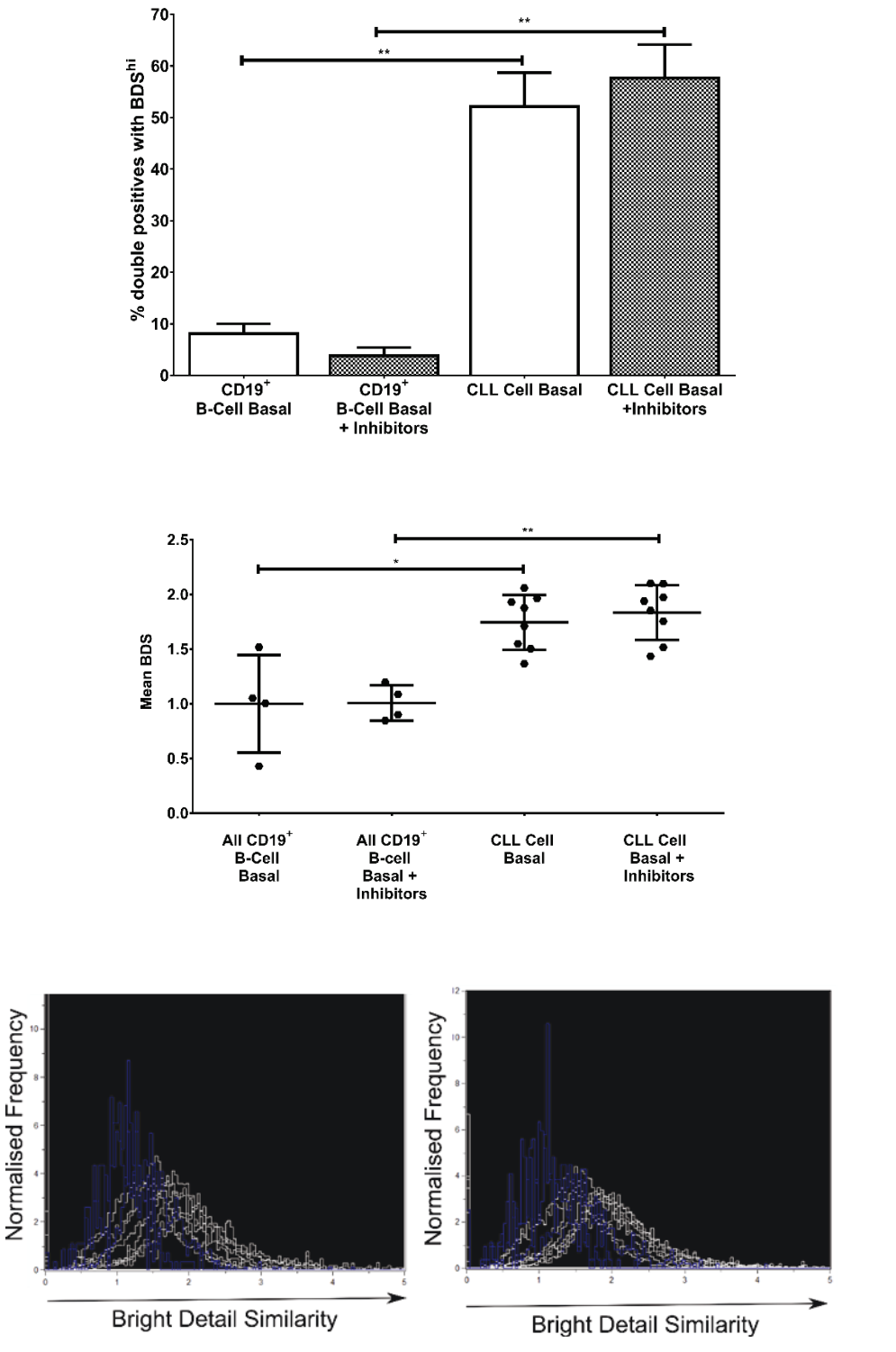


*Figure 6.4. CLL cells have a higher basal autophagy level than age-matched control B-cells. Autophagy levels in basal and basal + I from 8 CLL patients and 4 age-matched controls comparing a. Percentage of double positives with BDS^{hi} in (mean ± SEM, p=0.004 [Mann-Whitney Test]) b. Mean BDS (Mean ± SEM, *p=0.016, **p=0.004 [Mann-Whitney Test]). Overlay of BDS histograms from CLL (white) and normal B-cells (blue) under c. basal and d. basal + I conditions.*

6.3.3 Basal autophagy levels are higher in CLL cells than age-matched controls

To assess the basal level of autophagy in CLL, primary cells from patients and healthy age-matched controls were immunostained with anti-CD5 and anti-CD19 and the level of autophagy activity established. The autophagy level of the normal B-cells was similar to what has been previously described (Phadwal *et al* 2012). Analysis of the %BDS^{hi} demonstrated that the CLL cells have nearly a six-fold increase in autophagy activity compared to the normal B-cells (Figure 6.4a). The increase in autophagy is further supported by analysis of the mean BDS which shows a 70% increase in the CLL cells. (Figure 6.4b) An overlay of the BDS histograms (Figure 6.4c and d) for the CLL cells (blue) and the normal B-cells (white) provided further evidence of the increase.

6.3.4 CLL cells have increased autophagy levels upon induction

Autophagy can be stimulated during cellular stress to provide nutrients and energy. Therefore, cells were incubated in starvation buffer for 2hours prior to immunostaining to ascertain how the induction of autophagy activity in CLL cells compares to normal B-cells. Analysis of %BDS^{hi} showed that the CLL cells have nearly a 50% increase in induced autophagy activity compared to normal B-cells (Figure 6.5). Interestingly the data showed no difference between the induced and basal autophagy level in the CLL cells

(p=0.63) which differs from what has been previously reported (Kovaleva *et al* 2012).

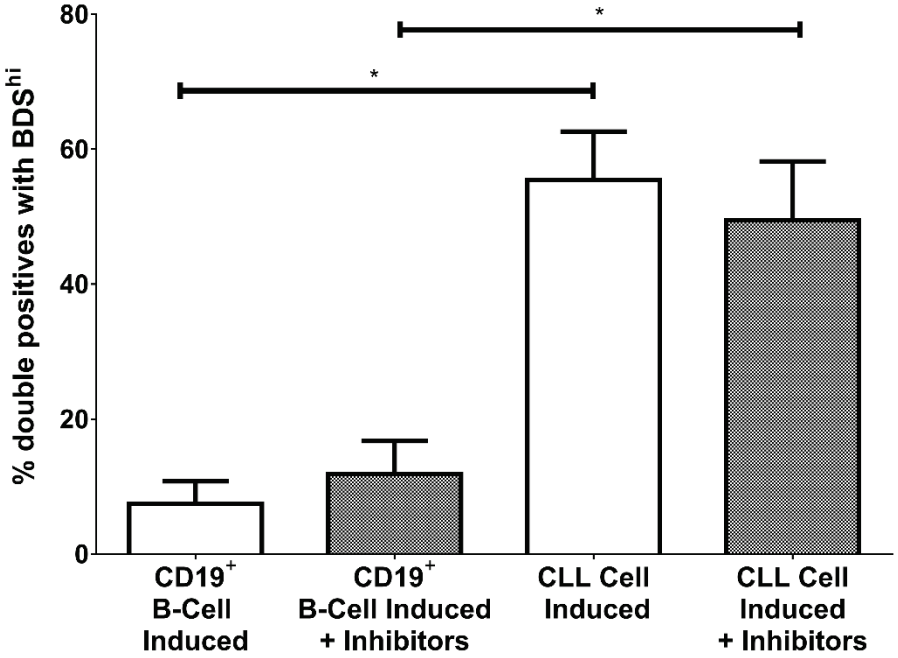


*Figure 6.5. Induced autophagy levels are higher in CLL cells than in age-matched controls. Percentage of double positives with BDS^{hi} in induced and induced + I from 8 CLL patients and 4 age-matched controls Mean ± SEM, *p=0.012, *p=0.032 [Mann-Whitney Test]).*

6.3.5 Increased autophagy levels in CLL cells is not a feature of $CD19^+CD5^+$ cells

CLL cells express CD19 and CD5 on their surface. Although CD19 is found on all mature B-cells (Nadler *et al* 1983, Wang *et al* 2012), CD5 is only expressed on B-1a cells which represent <20% of the B-cells compartment (Cabezudo *et al* 1997, Vuillier *et al* 1991). These cells have a different function to the $CD19^+CD5^-$ cells, so it is possible that the increase in autophagy is a feature of the $CD19^+CD5^+$ phenotype. Therefore, the $CD19^+$ B-cell cohort isolated from the age-matched controls was further delineated to the $CD5^+$ fraction. Analysis of the mean BDS in this subset

showed no significant difference in autophagy activity compared to the $CD19^{+}CD5^{-}$ cohort (Figure 6.6).

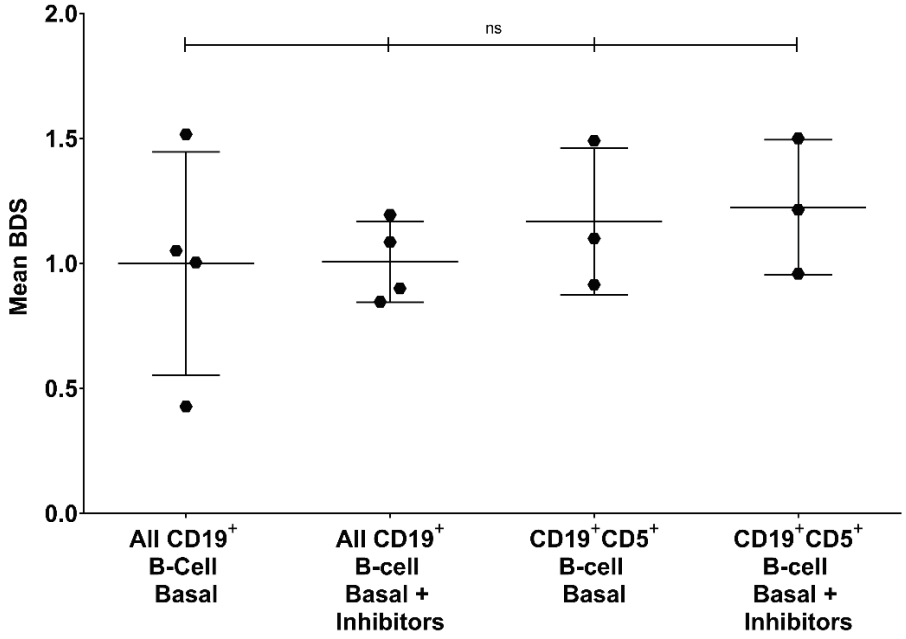
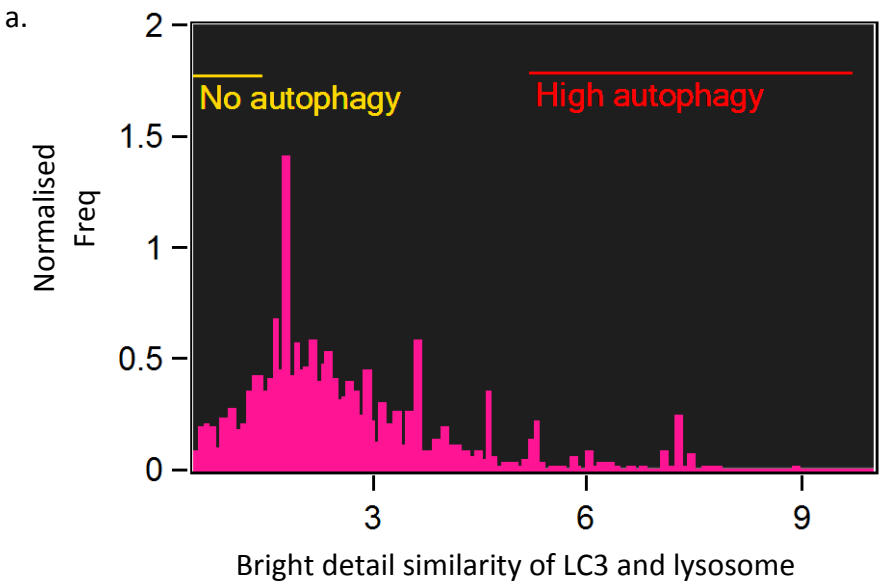


Figure 6.6. Basal autophagy level in the B-cell subsets of the age-matched controls. Mean BDS in basal and basal + inhibitors for all $CD19^{+}$ cells and in the $CD19^{+}Cd5^{+}$ cohort (mean ± SEM, p=0.86 and 0.23 respectively [Mann-Whitney U Test]) ns – not significant

6.3.6 Higher autophagy is not linked to cell death

Since an increase in autophagy could be a response to increased cellular death, the level of caspase-3 was determined in the cells. For this, the CLL cells were identified solely with an anti-CD19 marker, and the autophagy activity and caspase-3 level determined. Analysis of the caspase-3 level in CLL cells with high and no autophagy were compared. No significant change in the caspase-3 level was detected in any of the three patient samples (p-value=0.7) (Figure 6.7).



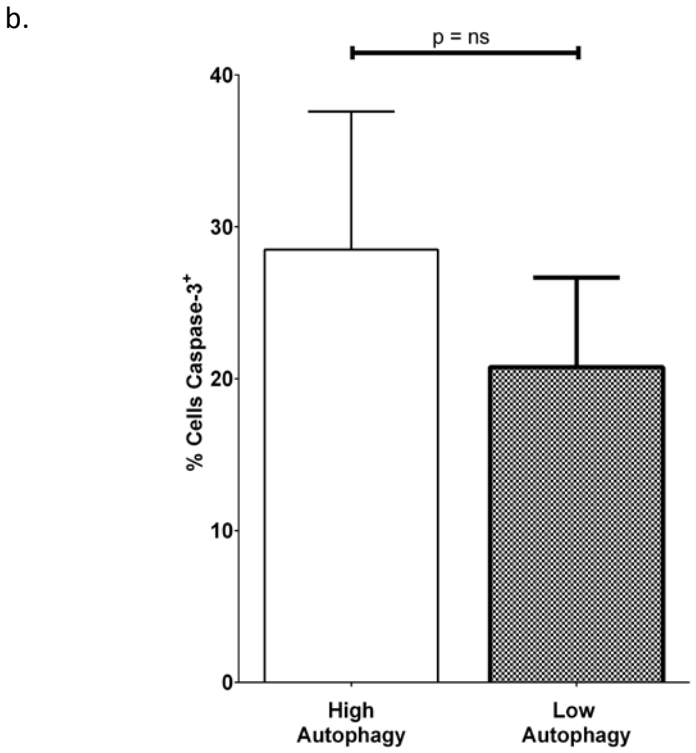


Figure 6.7 The increase in autophagy activity in the CLL cells is not associated with cell death. a. BDS overlay demonstrating the cells define as high and low autophagy. b. Percentage of cells with caspase-3 activity in the high and low autophagy subsets (mean ± SEM, p=0.7 [Mann Whitney U Test]).

6.4 Discussion

Since genes involved in the autophagy pathway were found to be recurrently mutated in patients with CLL it raises the question of a role for autophagy in the pathogenesis of CLL. This study has demonstrated that CLL cells have higher autophagy activity than normal age-matched B-cells and shown that this increase is neither a feature of B1-a cells or a response to an increase in apoptosis.

6.4.1 CLL cells have a higher level of autophagy activity than B-cells from normal age-matched controls

In this study, it has been demonstrated that CLL cells have a higher level of autophagy than B-cells from healthy age-matched controls. Normal $CD19^{+}$ B-cells are known to have low or undetectable levels of basal autophagy (Lee & Sugden 2008; Phadwal *et al*. 2012). Furthermore, mature B-cell can exist in the absence of functioning autophagy as has been demonstrated by ATG5 knockout mice suggesting that autophagy is not essential for B-cell survival (Miller *et al*. 2008; Arnold *et al*. 2016). The increased autophagy seen in the CLL cells supports previous studies that have shown that CLL cells use autophagy as a survival mechanism (El-Khoury *et al* 2014, Kovaleva *et al* 2012, Mahoney *et al* 2012).

Accumulation of autolysosomes, as demonstrated by the co-localisation of LC3 and lysosome, can potentially be cause by either an overproduction of autophagosomes or by a deficiency in the downstream machinery of the autophagy pathway, and could therefore represent defective autophagy rather than a true increase in activity (Mizushima *et al* 2010). To overcome this potential problem, the study in this chapter also analysed co-localisation in cells that have been treated with the lysosomal protease inhibitors Pepstain A and E64D. These inhibitors block the degradation of LC3-II causing an accumulation of autolysosomes (Tanida *et al* 2005). The increase in co-localisation in the CLL cells versus normal age-matched B-cells was still detectable under these conditions suggesting that rise in the number of autolysosomes in the CLL cells was not caused by defective autophagy.

6.4.2 Increased autophagy is not a feature of B1-a cells

B-1a cells are found naturally in the B-cell component. It is estimated that $CD5^+$ B-cells account for up 25% of the B-cell population (Cabezudo *et al* 1997, Vuillier *et al* 1991). B1-a cells, in contrast to B2-cells, are involved in the innate immune response. Since B1-a cells differ in their functionality to B2 cells it may be that they also differ in their levels of autophagic activity. Indeed a key function of B-1a cells is to remove autoantigens (Botto and Walport 2002, Ehrenstein and Notley 2010, Kaveri *et al* 2012, Tsiantoulas *et al* 2012) and it is possible that autophagy is the mechanism by which this

is achieved. Furthermore ATG5 has been shown to be important in the maintaining B1-a cells in the peripheral blood although it is not essential for B2 cells (Miller *et al* 2008) implying that these subsets may have distinct autophagy activity.

To determine if B1-a cells had an autophagy level that was different compared to the rest of the B-cell population, the B-cell compartment was further delineated into $CD5^+$ and CD5- cells in the normal age-matched controls. Comparison of these subgroups yielded no significant difference suggesting that the increase in autophagy activity was not a feature of $CD5^+CD19^+$ cell types.

6.4.3 Cells with high autophagy levels are not undergoing apoptosis

There is a large amount of cross-talk between the regulators and components of the autophagy and apoptosis pathways as well as an overlap in the activation of both pathways (Maiuri *et al* 2007). Consequently both processes are induced in response to the same stimuli (Marino *et al* 2014) and autophagy has been shown to be induced prior to apoptosis (Bhutia *et al* 2010). It is possible that the up-regulation of autophagy detected in the CLL cells in this study may be in response to apoptosis. Therefore, analysis of caspase-3 was done to exclude cell death as a reason for the increase in autophagy. Caspase-3 is activated early in

apoptosis and is therefore a marker of apoptosis. There was no increase in activated caspase-3 in the cells that had high levels of autophagy when compared to the cells with low levels of autophagy. This suggests that the increase in autophagy seen in the CLL cells was not a by-product of apoptosis.

There is some evidence that autophagy may act as a second death pathway, ergo it is possible that the increase in autophagy in the CLL cells is caused by autophagic cell death. As the mechanism of autophagic cell death is not yet understood, it is not possible to exclude this as an option through this assay, as autophagic cell death is independent of caspase-3 (Yang Liu *et al* 2013). However *ATG5* and *BECN1* expression has been found to be upregulated during autophagic cell death but not under autophagy survival conditions (Shimizu *et al* 2004) and since the work described in this chapter did not detect an increase in *ATG5* expression in the CLL samples, it is likely that the CLL cells are not undergoing autophagic cell death.

6.4.4 Conclusions

In conclusion, the work in this chapter has shown that CLL cells have higher levels of autophagy activity when compared to normal age-matched B-cells. The increase is not a feature of normal $CD5^+$ B-cells and is also not in response to programmed cell death. This further supports the idea that CLL

cells use autophagy as a survival mechanism and that inhibition of autophagy may sensitise CLL to treatment.

Chapter 7 Discussion

7.1 Introduction

Over the course of my thesis, the understanding of the molecular drivers of CLL pathogenesis has significantly increased due, mostly, to the increasing availability of NGS technologies. At the beginning of this work, CLL was thought to be a relatively stable disease originating in the secondary lymphoid organs. However, our work and that of others (Damm *et al* 2014, Kikushige *et al* 2011, Knight *et al* 2012, Landau *et al* 2015, Ouillette *et al* 2013, Puente *et al* 2011, Schuh *et al* 2012, Zenz *et al* 2009), provides evidence that supports the presence of ongoing clonal evolution and of a pre-leukaemic stem cell that originates from an un-rearranged progenitor cell in the bone marrow.

7.2 Summary of Findings

My thesis aimed to characterise the genomic complexity of CLL. It sought to identify candidate molecular drivers responsible for clonal expansion and relapse following clinical intervention and to define predictors of poor outcome. A range of techniques were utilised, including genome-wide SNP array as well as NGS of both the IgHV region and recurrently mutated

genes. Paired pre-treatment and relapse samples were used to allow the identification of emerging subclones. Finally, where possible, findings were correlated with clinical outcomes to determine prognostic significance or validated at protein level using flow cytometry.

At the start of this thesis, in 2009, NGS technology was in its infancy. Three platforms were available, the Roche 454 GS Flx, the Illumina Genome Analyser and the Applied Biosystems SOLiD sequencer. Each were designed primarily for high-throughput large-scale sequencing (Mardis 2008). Only the release of benchtop sequencers made NGS accessible to a wider user group, including routine diagnostic laboratories (Loman *et al* 2012). The work in Chapter 3 describes efforts to validate NGS for investigation of the IgHV region using these technologies. At first the Roche 454 was the only platform with the capacity to perform 400bp sequencing, required to cover the whole of the IgHV. It was only replaced by the Illumina MiSeq when this platform could perform 600bp sequencing.

Chapter 3 introduces the concept of IgHV subclones. The work by Campbell *et al* (2008) was the first to demonstrate ongoing clonal evolution in CLL through NGS analysis of the IgHV region using a VH-family specific primer. The work in Chapter 3 expands on this by utilising all the VH-primers to search for subclones with different VDJ rearrangements, the presence of which would indicate a pre-B cell leukaemic stem cell. The detection of two

productive VDJ rearrangements has previously been reported in CLL at low frequencies (Plevova *et al* 2014, Rassenti and Kipps 1997, Visco *et al* 2013). The work in this chapter found that by using NGS it was possible to detect multiple productive IgHV subclones in 24% of CLL patients. This finding provides evidence for leukaemia-initiating events that pre-date IgHV rearrangement.

Chapter 4 sought to determine whether the IgHV subclones formed part of the CLL or whether these were contaminating B-cells from the normal polyclonal background. Through a series of validation experiments it was possible to show that these subclones were indeed part of the CLL. Furthermore, it was demonstrated that these subclones refined the previous SSeq-based IgHV prognostic classification system (Damle *et al* 1999, Hamblin *et al* 1999).

Chapter 5 used a cohort of 411 pre-treatment and relapse samples to detect recurrent genetic abnormalities. This work primarily used high resolution SNP array to screen for genome-wide copy number aberrations. At the time, NGS was not available and CGH- and SNP-array were the techniques of choice to screen for global changes in the genome. Prior to the start of this study there had been limited work in CLL performed using SNP-arrays (Gunnarsson *et al* 2008, Pfeifer *et al* 2006). This chapter expands on these studies by using high-resolution SNP arrays with over one

million probes, on a larger cohort of patients (411 versus 10 and 70, respectively). Furthermore, this work included paired diagnosis and relapse samples enabling, for the first time, clonal evolution to be demonstrated in CLL at a global level through the identification of emerging subclones (Knight *et al* 2012).

Finally, the work in Chapter 6 showed that CLL cells had higher levels of autophagy when compared with normal B-cells from healthy age-matched controls. This work followed on from the observation in Chapter 5, that *ATG5* was one of two genes in the MOR of the del(6q) and that autophagy related genes could be identified in other MORs that occurred recurrently in CLL. There have been limited studies investigating the role of autophagy in the pathogenesis of CLL. Whilst a few studies have suggested that CLL cells up-regulate autophagy in response to stress (Kovaleva *et al* 2012, Mahoney *et al* 2012) and that inhibition of autophagy decreases the viability of CLL cells (El-Khoury *et al* 2014, Mahoney *et al* 2013), to date there has been no work investigating the autophagy levels in CLL cells, compared with age-matched normal controls.

7.3 The relationship between aging and development of CLL

Both Chapters 4 and 6 use $CD19^+$ B-cells from age-matched healthy volunteers as controls. CD19 is a reliable hallmark of B-cells since it is expressed on the B-cell surface from the pro-B phase (Wang *et al* 2012).

The control samples are acquired from age-matched volunteers since CLL is found predominately in the over 65 age-group (Cancer Research UK, 2016b). Age-matched controls ensure, as far as possible, that any abnormalities are a characteristic of the CLL rather than of aging. Both autophagy and B-cells are known to be affected by the aging process. Autophagy levels are thought to decrease with age due to the altered expression and availability of its components (Carroll *et al* 2013, Phadwal *et al* 2012). Therefore, aged-matched controls are vital in differentiating between the effects of the CLL and the effects of aging.

Likewise, age is associated with a higher frequency of haematopoietic disorders. The incidence of clonal haematopoiesis of indeterminate potential (CHIP), an entity defined by the presence of somatically acquired mutations in blood cells, is around 10% in the over 60s (Genovese *et al* 2014, Jaiswal *et al* 2014, Steensma *et al* 2015, Xie *et al* 2014). Monoclonal B cell lymphopoesis (MBL) has been reported in just under 20% of the healthy population aged over 70 (Almeida *et al* 2011). Similar to CHIP, MBL

is a relatively benign condition, although mutations found in some haematological malignancies have been reported in MBL (Ojha, Secreto *et al* 2014, Rawstron *et al* 2008b). The validation work in Chapter 4 provides further support for the presence of oligoclonal B cell populations in the over 70s since it was possible to identify low frequency clonal IgHV subclones in every normal control. It is important to stress that whilst these cells were $CD19^{+}$, they were also CD5-. Therefore, these cells do not represent MBL, but are typical of age-related B-cell expansion (Gibson *et al* 2009, Frasca and Blomberg 2009). However, whilst both CHIP and MBL are not uncommon, the incidence of transformation is relatively low at around 1% per year (Genovese *et al* 2014, Jaiswal *et al* 2014, Rawstron *et al* 2008b). Therefore, whilst these diseases represent a pre-leukaemic state, a second hit is required for transformation to CLL and this is likely to involve antigenic stimulation in the germinal centre.

7.4 CLL as a subclonal disease

CLL has traditionally been thought of as a monoclonal disease characterised by the presence of a single productive VDJ rearrangement. Although the detection of two productive rearrangements had been reported it was argued that only one would be transcribed, translated, and expressed on the cell surface (Rosenquist 2009). This was further supported by early work suggesting that the presence of two productive clones was, in most instances, caused by rearrangement of both alleles in

one cell due to a lack of allelic exclusion otherwise known as allelic inclusion (Langerak *et al* 2011, Rassenti and Kipps 1997).

Whilst it was accepted that two rearrangements may represent a biclonal disease, this was thought to be rare since the clones were required to be both morphologically and phenotypically distinct in order to be classified as such (Sánchez *et al* 2002). Later work by Plevova *et al* (2014) found several incidences of CLL with a homogeneous phenotype alongside multiple productive rearrangements. The group excluded allelic inclusion, through the detection of partial VJ rearrangements since this would suggest that full VDJ recombination had not taken place on both alleles. Single cell sequencing has been used more recently to show that there is only one productive rearrangement per cell (Kriangkum *et al* 2015). Therefore, it is highly likely that each rearrangement corresponds to an individual leukaemia subclone, that is, that the different rearrangements are present in different cells.

The data presented in Chapters 3 and 4 demonstrate that the incidence of multiple productive rearrangements in CLL is far higher than previously reported (Plevova *et al* 2014, Rassenti and Kipps 1997, Visco *et al* 2013), suggesting that CLL, like other cancers and leukaemias, is a subclonal disease. Furthermore, the work in Chapter 5 provides further evidence of subclones through the detection of expanding mutations at relapse.

Indeed, the notion of subclones in CLL, as defined as by subclonal mutations and CNAs, is well supported by current literature (Amin *et al* 2016, Grubor *et al* 2008, Jethwa *et al* 2013, Knight *et al* 2012, Landau *et al* 2013, 2015, Ljungström *et al* 2016, Rasi *et al* 2016, Rossi *et al* 2014, Schuh *et al* 2012).

It has been proposed that the presence of multiple productive rearrangements represents CLL co-existing alongside ongoing MBL with a CLL-phenotype (Plevova *et al* 2014). Both clonal and subclonal IgHV CLL-like MBL has been reported (Klinger *et al* 2015, Rawstron *et al* 2008a), so it is possible that there is underlying MBL in multiply productive CLL. Through ultra-deep targeted NGS, it was possible to quantify the IgHV subclones detected in Chapters 4 and 5. A biclonal IgHV was detected in 65% of multiple productive CLL with an average frequency of 10% for the second clone. It is possible that this indicates CLL with a monoclonal MBL background. Likewise, three IgHV subclones were detected in 19%, and over four subclones were detected in 16% of multiply productive CLL patients, which may suggest CLL with an oligoclonal MBL background. It is difficult to distinguish subclonal CLL from monoclonal CLL with background MBL, since the discriminating factor between CLL and MBL diagnostically is their white cell count in the peripheral blood (Strati and Shanafelt 2015). However, there is evidence that MBL does not have stereotyped CDR3 receptors (Dagklis *et al* 2008). Chapter 4 shows that some IgHV subclones

did demonstrate stereotyping which may favour a subclonal CLL model. Regardless of the origin of the subclones, their very presence is supportive of an IgHV subclonal disease, which contrasts with the traditional concept of a monoclonal disorder.

7.5 The role of autophagy in CLL pathogenesis

The work in Chapter 6 shows that CLL cells have altered autophagy when compared with normal B-cells. The reason behind the investigation into autophagy was the finding that *ATG5* is recurrently mono-allelically deleted in CLL. It would be assumed that the deletion of one copy of the *ATG5* gene would result in some loss of expression, yet this was not seen using quantitative PCR. Furthermore, a functional loss of *ATG5* would be predicted to reduce the level of autophagy, however, increased autophagy was reported even in the patient with the *ATG5* deletion. SNPs in the *ATG5* gene and in its preceding intergenic region have been shown to increase *ATG5* expression (Zhou *et al* 2011) therefore it is possible that the loss of the balancing SNP may cause upregulation of the *ATG5* gene. However, this study did not find an increase in *ATG5* expression in the CLL samples, although *BECN1* expression was downregulated, a finding that has been previously reported (Kristensen *et al* 2015). Therefore, it is worth noting that mRNA levels do not always correspond to protein activity (Vogel and Marcotte 2012) and that the expression levels of *ATG5* and *BECN1* may not affect autophagy activity in CLL.

On the other hand, it has been shown that autophagy can take place in mammalian cells in the absence of the ATG5 protein (Nishida *et al* 2009). Furthermore, inactivation of *ATG5* was found to accelerate oncogenesis in mice, although it also decreased the level of progression (Rao *et al* 2014). This is in keeping with the dual role of autophagy in cancer, such that dysregulated autophagy can lead to tumourigenesis, yet functioning autophagy provides a survival advantage to cancers once they have developed. Whilst the absence of *ATG5* and *BECN1* may have implications for the development of CLL, it is not possible to explain the relationship, if any, between *BECN1* expression levels and the increase in autophagy reported in Chapter 6.

Normal B-cells have a low level of autophagy activity (Phadwal *et al* 2012). Although not essential for B-cell development, it has been shown that autophagy is required at a basal level to sustain B-cells long-term in the periphery (Arnold *et al* 2016, McLeod and He 2010, Miller *et al* 2008). B-cells are known to use autophagy in the processing of antigens prior to their presentation to T-cells (Clark *et al* 2004, Schmid *et al* 2007, Watanabe and Tsubata 2009). Furthermore, antigen stimulation of the BCR has been shown to induce autophagy in B-cells (Watanabe *et al* 2008).

Chapter 6 found that CLL cells have a higher level of autophagy activity than their counterpart B-cells. Autophagy is known to provide a survival advantage to tumour cells through the provision of nutrients (Guo *et al* 2011) and the prevention of apoptosis (Hou *et al* 2010, Wang *et al* 2015). Recent work suggests that autophagy may also play a role in disabling the anti-tumour immune response (Lévy *et al* 2015, Rao *et al* 2014, Wei and Guan 2012). Stimulation of the BCR has also been shown to induce autophagy in CLL cells (Smith *et al* 2015) and since it has been proposed that activation of the BCR may be the trigger in transitioning MBL to CLL (Sutton and Rosenquist 2014), it is possible that autophagy may be instrumental in this evolution. Indeed, UM-CLL has been reported to have higher autophagy activity than M-CLL (Smith *et al* 2015), providing further evidence that increased autophagy confers a survival advantage in CLL.

7.6 Future work and directions

Although my work has provided some insight into the genomic complexity of CLL, it also raises further questions for investigation.

Firstly, it would be beneficial to confirm the IgHV prognostic classification system in an independent cohort. Whilst validation of this model in another clinical trial cohort may be subjected to the same issues of sample bias that was reported in population B in Chapter 4, examination of a well

annotated non-trial cohort would provide further support for this model. Additionally, since the ARCTIC and AdMIRe trial samples used for population B had limited follow-up, it would be worth re-evaluating the data in a further five years to ascertain if the model works for this cohort in terms of OS. Long-term follow-up of the CLL8 trial has shown that around 50% of patients with M-CLL obtained a long-term PFS following FCR treatment (Stilgenbauer *et al* 2014). It would be very interesting to test this cohort for their NGS-IgHV status to see if they would fit the prognostic model, as well as potentially differentiating the long-term PFS M-CLL patients from the other M-CLL patients.

The work performed in Chapters 3 and 4 used a 2.5% cut-off to distinguish between subclones and B-cell background. Early NGS work by Campbell *et al* (2008) reported subclones as low as 0.02% although they excluded B-cell background by solely interrogating the dominant clonal IgHV. The validation work in Chapter 4 shows that less than 0.4% of the reads would be normal B-cells, suggesting that a cutoff of 2.5% is very conservative. It also shows that using gDNA reduces the occurrence of IgHV subclones since they fall below the 2.5% cut-off. Therefore, it would be interesting to see if using a lower cut-off increases the frequency of IgHV subclonal CLL and the effect this would have on the NGS-IgHV prognostic model.

It would be interesting to correlate the IgHV subclones with other mutations and to analyse sequential samples to see how subclones evolve over time. Currently, it is difficult to use NGS to determine both IgHV mutation status and the presence of gene mutations simultaneously due to the different processing techniques involved. It may be possible to add IgHV to a TSCA panel, however I think that it will prove difficult to achieve acceptable sequencing coverage without compromising the integrity of the rest of the panel. Whilst it may be possible to use WGS for this, again I predict that there will be issues at the bioinformatics stage when piecing together the 100bp reads to obtain a full and accurate coverage of the IgHV. Additionally, WGS has an average depth of 75x, which would be insufficient for the detection of some IgHV subclones. It is possible that third generation sequencing technologies such as Nanopore technology may solve this problem since it is able to process large lengths of unfragmented DNA and may offer a real alternative to single cell sequencing (Jain *et al* 2016, Lu *et al* 2016).

Finally, further work on the role of autophagy in CLL and MBL may provide valuable insight into the pathogenesis of CLL, as well as offering a target for treatment. Investigation of the link between BCR signalling and increased autophagy in CLL (Smith *et al* 2015) may elucidate whether autophagy is one survival mechanism by which CLL cells evade anticancer immune-surveillance and persevere following treatment. Indeed, inhibition of

autophagy in CLL has been recurrently shown to decrease the viability of CLL cells (Amrein *et al* 2011, Mahoney *et al* 2013). Furthermore, work on the activation of autophagy through BCR signalling may be instrumental in the determining how CLL may develop through antigen stimulation.

7.7 Conclusion

Overall this thesis has contributed to our knowledge of the genomic architecture of CLL through the utilisation of NGS and SNP array. CLL was found to harbour IgHV subclones that are likely to contribute to the biological heterogeneity of this disease and point towards leukaemia-initiating events pre-dating IgHV rearrangement. Finally, CLL cells were found to carry chromosomal copy number changes that affected a number of autophagy-related genes which may provide an alternative mechanism whereby CLL cells become resistant to apoptosis.

References

Agathangelidis A, Darzentas N, Hadzidimitriou A, Brochet X, Murray F, Yan X-J, Davis Z, van Gastel-Mol EJ, Tresoldi C, Chu CC, Cahill N, Giudicelli V, Tichy B, Pedersen LB, Foroni L, Bonello L, Janus A, Smedby K, Anagnostopoulos A, Merle-Beral H, Laoutaris N, Juliusson G, di Celle PF, Pospisilova S, Jurlander J, Geisler C, Tsaftaris A, Lefranc M-P, Langerak AW, Oscier DG, Chiorazzi N, Belessi C, Davi F, Rosenquist R, Ghia P and Stamatopoulos K (2012) Stereotyped B-cell receptors in one third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapeutic interventions. *Blood* 119(19): 4467–4476.

Agathangelidis A, Vardi A, Baliakas P and Stamatopoulos K (2014) Stereotyped B-cell receptors in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 55(10): 2252–2261.

Aita VM, Liang XH, Murty V V, Pincus DL, Yu W, Cayanis E, Kalachikov S, Gilliam TC and Levine B (1999) Cloning and genomic organization of beclin 1, a candidate tumor suppressor gene on chromosome 17q21. *Genomics* 59(1): 59–65.

Alamyar E, Giudicelli V, Li S, Duroux P and Lefranc MP (2012) IMGT/Highv-quest: The IMGT web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Research* 8(1): 1–15.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio S a JR, Behjati S, Biankin A V, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens J a, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Imielinsk M, Jäger N, Jones DTW, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott P a, Pajic M, Papaemmanuil E, Paradiso A, Pearson J V, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt ANJ, Valdés-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ and Stratton MR (2013) Signatures of mutational processes in human cancer. *Nature* 500: 415–21.

Almeida J, Nieto WG, Teodosio C, Pedreira CE, López A, Fernández-Navarro P, Nieto A, Rodríguez-Caballero A, Muñoz-Criado S, Jara-Acevedo M, Romero A and

Orfao A (2011) CLL-like B-lymphocytes are systematically present at very low numbers in peripheral blood of healthy adults. *Leukemia* 25(4): 718–722.

Alt FW, Yancopoulos GD, Blackwell TK, Wood C, Thomas E, Boss M, Coffman R, Rosenberg N, Tonegawa S and Baltimore D (1984) Ordered rearrangement of immunoglobulin heavy chain variable region segments. *Embo Journal* 3(6): 1209–19.

Amin NA, Seymour E, Saiya-Cork K, Parkin B, Shedden K and Malek SN (2016) A quantitative analysis of subclonal and clonal gene mutations before and after therapy in chronic lymphocytic leukemia. *Clinical Cancer Research* 22(17): 4525–4535.

Amin RH and Schlissel MS (2008) Foxo1 directly regulates the transcription of recombination-activating genes during B cell development. *Nature Immunology* 9(6): 613–22.

Amrein L, Soulières D, Johnston JB and Aloyz R (2011) p53 and autophagy contribute to dasatinib resistance in primary CLL lymphocytes. *Leukemia Research* 35(1): 99–102.

Anao TN, Oike MK, Amaguchi JY, Asaki MS and Chiyama YU (2015) Cellular localization and tissue distribution of endogenous DFCP1 protein. *Biomedical Research* 36(2): 121–133.

Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, Kempski H, Moorman A V., Titley I, Swansbury J, Kearney L, Enver T and Greaves M (2011) Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* 469(7330): 356–61.

Arnold J, Murera D, Arbogast F, Fauny J-D, Muller S and Gros F (2016) Autophagy is dispensable for B-cell development but essential for humoral autoimmune responses. *Cell Death and Differentiation* 23(5): 853–864.

Austen B, Powell JE, Alvi A, Edwards I, Hooper L, Starczynski J, Taylor AMR, Fegan C, Moss P and Stankovic T (2005) Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL. *Blood* 106(9): 3175–3182.

Austen B, Skowronska A, Baker C, Powell JE, Gardiner A, Oscier DG, Majid A, Dyer M, Siebert R, Taylor AM, Moss PA and Stankovic T (2007) Mutation status of the residual ATM allele is an important determinant of the cellular response to

chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion. *Journal of Clinical Oncology* 25(34): 5448–5457.

Awad MM, Katayama R, McTigue M, Liu W, Deng Y-L, Brooun A, Friboulet L, Huang D, Falk MD, Timofeevski S, Wilner KD, Lockerman EL, Khan TM, Mahmood S, Gainor JF, Digumarthy SR, Stone JR, Mino-Kenudson M, Christensen JG, Iafrate a J, Engelman J a and Shaw AT (2013) Acquired resistance to crizotinib from a mutation in CD74-ROS1. *The New England Journal of Medicine* 368(25):

Axe EL, Walker SA, Manifava M, Chandra P, Roderick HL, Habermann A, Griffiths G and Ktistakis NT (2008) Autophagosome formation from membrane compartments enriched in phosphatidylinositol 3-phosphate and dynamically connected to the endoplasmic reticulum. *The Journal of Cell Biology* 182(4): 685–701.

Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, Van Allen E, Kryukov G V., Sboner A, Theurillat JP, Soong TD, Nickerson E, Auclair D, Tewari A, Beltran H, Onofrio RC, Boysen G, Guiducci C, Barbieri CE, Cibulskis K, Sivachenko A, Carter SL, Saksena G, Voet D, Ramos AH, Winckler W, Cipicchio M, Ardlie K, Kantoff PW, Berger MF, Gabriel SB, Golub TR, Meyerson M, Lander ES, Elemento O, Getz G, Demichelis F, Rubin MA and Garraway LA (2013) Punctuated evolution of prostate cancer genomes. *Cell* 153(3): 666–677.

Bachow SH and Lamanna N (2016) Evolving Strategies for the Treatment of Chronic Lymphocytic Leukemia in the Upfront Setting. *Current Hematologic Malignancy Reports* 11(1): 61–70.

Balatti V, Bottoni A, Palamarchuk A, Alder H, Rassenti LZ, Kipps TJ, Pekarsky Y and Croce CM (2012) NOTCH1 mutations in CLL associated with trisomy 12. *Blood* 119(2): 329–331.

Balderhaar HJ kleine and Ungermann C (2013) CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion. *Journal of Cell Science* 126(Pt 6): 1307-16

Baliakas P, Agathangelidis A, Hadzidimitriou A, Sutton L-A, Minga E, Tsanousa A, Scarfò L, Davis Z, Yan XJ, Shanafelt TD, Plevova K, Sandberg Y, Vojdeman FJ, Boudjogra M, Tzenou T, Chatzouli M, Chu CC, Veronese S, Gardiner A, Mansouri L, Smedby KE, Pedersen LB, Moreno D, Van Lom K, Giudicelli V, Francova HS, Nguyen-Khac F, Panagiotidis P, Juliusson G, Angelis L, Anagnostopoulos A, Lefranc

MP, Facco M, Trentin L, Catherwood M, Montillo M, Geisler CH, Langerak AW, Pospisilova S, Chiorazzi N, Oscier DG, Jelinek DF, Darzentas N, Belessi C, Davi F, Ghia P, Rosenquist R and Stamatopoulos K (2015) Not all IGHV3-21 chronic lymphocytic leukemias are equal: Prognostic considerations. *Blood* 125(5): 856–859.

Baliakas P, Hadzidimitriou A and Sutton L-A (2014) Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study. *The Lancet Haemtology* 1: e74-84.

Baliakas P, Hadzidimitriou A, Sutton L-A, Rossi D, Minga E, Villamor N, Larrayoz M, Kminkova J, Agathangelidis A, Davis Z, Tausch E, Stalika E, Kantorova B, Mansouri L, Scarfò L, Cortese D, Navrkalova V, Rose-Zerilli M, Smedby KE, Juliusson G, Anagnostopoulos A, Makris a M, Navarro A, Delgado J, Oscier DG, Belessi C, Stilgenbauer S, Ghia P, Pospisilova S, Gaidano G, Campo E, Strefford JC, Stamatopoulos K and Rosenquist R (2015) Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia* (April): 1–8.

Barreto V and Cumano A (2000) Frequency and characterization of phenotypic Ig heavy chain allelically included IgM-expressing B cells in mice. *Journal of Immunology* 164(2): 893–9.

Bashford-Rogers R, Palser AL, Huntly BJ, Rance R, Vassiliou GS, Follows G a and Kellam P (2013b) Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. *Genome Research* 23: 1874–1884.

Berg J, Mcdowell M, Jäck H-M and Wabl M (1990) Immunoglobulin λ Gene Rearrangement Can Precede κ Gene Rearrangement. *Developmental Immunology* 1(1): 53–57.

Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick GS, Wild R, Sullivan K, Vijayakrishnan J, Wang Y, Pittman AM, Sunter NJ, Hall AG, Dyer MJ, Matutes E, Dearden C, Mainou-Fowler T, Jackson GH, Summerfield G, Harris RJ, Pettitt AR, Hillmen P, Allsup DJ, Bailey JR, Pratt G, Pepper C, Fegan C, Allan JM, Catovsky D and Houlston RS (2008) A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nature Genetics* 40(10): 1204–1210.

Bertolino E, Reddy K, Medina KL, Parganas E, Ihle J and Singh H (2005) Regulation of interleukin 7-dependent immunoglobulin heavy-chain variable gene rearrangements by transcription factor STAT5. *Nature immunology* 6(8): 836–43.

Bhutia SK, Dash R, Das SK, Azab B, Su Z-ZZ, Lee S-GG, Grant S, Yacoub A, Dent P, Curiel DT, Sarkar D and Fisher PB (2010) Mechanism of autophagy to apoptosis switch triggered in prostate cancer cells by antitumor cytokine melanoma differentiation-associated gene 7/interleukin-24. *Cancer Research* 70(9): 3667–76.

Binet JL, Auquier A, Dighiero G, Chastang C, Piguet H, Goasguen J, Vaugier G, Potron G, Colona P, Oberling F, Thomas M, Tchernia G, Jacquillat C, Boivin P, Lesty C, Duault MT, Monconduit M, Belabbes S and Gremy F (1981) A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer* 48(1): 198–206.

Blachly JS, Ruppert AS, Zhao W, Long S, Flynn J, Flinn I, Jones J, Maddocks K, Andritsos L, Ghia EM, Rassenti LZ, Kipps TJ, de la Chapelle A and Byrd JC (2015) Immunoglobulin transcript sequence and somatic hypermutation computation from unselected RNA-seq reads in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 112(14): 4322–7.

Bo MD, Del Principe MI, Pozzo F, Ragusa D, Bulian P, Rossi D, Capelli G, Rossi FM, Niscola P, Buccisano F, Bomben R, Zucchetto A, Maurillo L, de Fabritiis P, Amadori S, Gaidano G, Gattei V and Del Poeta G (2014) NOTCH1 mutations identify a chronic lymphocytic leukemia patient subset with worse prognosis in the setting of a rituximab-based induction and consolidation treatment. *Annals of Hematology* 93(10): 1765–1774.

Bologna C, Buonincontri R, Serra S, Vaisitti T, Audrito V, Brusa D, Pagnani A, Coscia M, Arena GD, Mereu E, Piva R, Furman RR, Rossi D, Gaidano G, Terhorst C and Deaglio S (2016) SLAMF1 regulation of chemotaxis and autophagy determines CLL patient response. *The Journal of Clinical Investigation* 126(16): 1–14.

Bomben R, Dal Bo M, Capello D, Forconi F, Maffei R, Laurenti L, Rossi D, Del Principe MI, Zucchetto A, Bertoni F, Rossi FM, Bulian P, Cattarossi I, Ilariucci F, Sozzi E, Spina V, Zucca E, Degan M, Lauria F, Del Poeta G, Efremov DG, Marasca R, Gaidano G and Gattei V (2009) Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: Results from an Italian multicentre study. *British Journal of Haematology* 144(4): 492–506.

Botto M and Walport MJ (2002) C1q, autoimmunity and apoptosis. *Immunobiology* 205(4–5): 395–406.

Boya P, Gonzalez-Polo, Rosa-Ana Casares N, Perfettini J-L, Dessen P, Larochette N, Métivier D, Meley D, Souquere S, Yoshimori T, Pierron G, Codogno P and Kroemer G (2005) Inhibition of Macroautophagy Triggers Apoptosis Inhibition of Macroautophagy Triggers Apoptosis. *Molecular and Cellular Biology* 25(3): 1025–1040.

Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, Jones CD, Simen BB, Hanczaruk B, Nguyen KD, Nadeau KC, Egholm M, Miklos DB, Zehnder JL and Fire AZ (2009) Measurement and clinical monitoring of human lymphocyte clonality by massively parallel {VDJ} pyrosequencing. *Science Translational Medicine* 1(12): 12ra23.

Bredemeyer AL, Sharma GG, Huang CY, Helmink BA, Walker LM, Khor KC, Nuskey B, Sullivan KE, Pandita TK, Bassing CH and Sleckman BP (2006) ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* 442(7101): 466–470.

Burger JA, Keating MJ, Wierda WG, Hartmann E, Hoellenriegel J, Rosin NY, de Weerdt I, Jeyakumar G, Ferrajoli A, Cardenas-Turanzas M, Lerner S, Jorgensen JL, Nogueras-Gonz??lez GM, Zacharian G, Huang X, Kantarjian H, Garg N, Rosenwald A and O'Brien S (2014) Safety and activity of ibrutinib plus rituximab for patients with high-risk chronic lymphocytic leukaemia: A single-arm, phase 2 study. *The Lancet Oncology* 15(10): 1090–1099.

Byrd JC, Brown JR, O'Brien S, Barrientos JC, Kay NE, Reddy NM, Coutre SE, Tam CS, Mulligan SP, Jaeger U, Devereux S, Barr PM, Furman RR, Kipps TJ, Cymbalista F, Pocock C, Thornton P, Caligaris-Cappio F, Robak T, Delgado J, Schuster SJ, Montillo M, Schuh A, de Vos S, Gill D, Bloor A, Dearden C, Moreno C, Jones JJ, Chu AD, Fardis M, McGreivy J, Clow F, James DF and Hillmen P (2014) Ibrutinib versus Ofatumumab in Previously Treated Chronic Lymphoid Leukemia. *New England Journal of Medicine* 371(3): 213–223.

Byrd JC, Stilgenbauer S and Flinn IW (2004) Chronic lymphocytic leukemia. *American Society of Hematology Education Program* 163–183. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15561682

Cabezudo E, Matutes E, Ramratten R, Morilla R and Catovsky D (1997) Analysis of residual disease in chronic lymphocytic leukemia by flow cytometry. *Leukemia* 11: 1909–1914.

Calin G, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti LZ, Kipps T, Negrini M, Bullrich F and Croce CM (2002) Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 99(24): 15524–15529.

Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, Follows GA, Green AR, Futreal PA and Stratton MR (2008) Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 105(35): 13081–13086.

Cancer Research UK (2016a) *Chronic lymphocytic leukaemia (CLL) Key Stats*. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-cll/incidence#heading-One

Cancer Research UK (2016b) *Cancer mortality statistics*. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality.

Carroll B, Hewitt G and Korolchuk VI (2013) Autophagy and ageing: implications for age-related neurodegenerative diseases. *Essays in Biochemistry* 55: 119–31.

Catovsky D, Richards S, Matutes E, Oscier DG, Dyer MJ, Bezares RF, Pettitt AR, Hamblin TJ, Milligan DW, Child JA, Hamilton MS, Dearden CE, Smith AG, Bosanquet AG, Davis Z, Brito-Babapulle V, Else M, Wade R and Hillmen P (2007) Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet* 370(9583): 230–239.

Chapiro E, Leporrier N, Radford-Weiss I, Bastard C, Mossafa H, Leroux D, Tigaud I, De Braekeleer M, Terré C, Brizard F, Callet-Bauchu E, Struski S, Veronese L, Fert-Ferrer S, Taviaux S, Lesty C, Davi F, Merle-Béral H, Bernard OA, Sutton L, Raynaud SD and Nguyen-Khac F (2010) Gain of the short arm of chromosome 2 (2p) is a frequent recurring chromosome aberration in untreated chronic lymphocytic leukemia (CLL) at advanced stages. *Leukemia Research* 34(1): 63–68.

Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E, Downes K, Macaulay IC, Bielczyk-Maczynska E, Coe S, Farrow S, Poudel P, Burden F, Jansen

SBG, Astle WJ, Attwood A, Bariana T, de Bono B, Breschi A, Chambers JC, Consortium B, Choudry FA, Clarke L, Coupland P, van der Ent M, Erber WN, Jansen JH, Favier R, Fenech ME, Foad N, Freson K, van Geet C, Gomez K, Guigo R, Hampshire D, Kelly AM, Kerstens HHD, Kooner JS, Laffan M, Lentaigne C, Labalette C, Martin T, Meacham S, Mumford A, Nürnberg S, Palumbo E, van der Reijden BA, Richardson D, Sammut SJ, Slodkowicz G, Tamuri AU, Vasquez L, Voss K, Watt S, Westbury S, Flicek P, Loos R, Goldman N, Bertone P, Read RJ, Richardson S, Cvejic A, Soranzo N, Ouwehand WH, Stunnenberg HG, Frontini M and Rendon A (2014) Transcriptional diversity during lineage commitment of human blood progenitors. *Science* 345(6204): 1251033.

Chen L, Widhopf G, Huynh L, Rassenti LZ, Rai KR, Weiss A and Kipps TJ (2002) Expression of ZAP-70 is associated with increased B-cell receptor signaling in chronic lymphocytic leukemia. *Receptor* 100(13): 4609–4614.

Chiaretti S, Marinelli M, Del Giudice I, Bonina S, Piciocchi A, Messina M, Vignetti M, Rossi D, Di Maio V, Mauro FR, Guarini A, Gaidano G and Foà R (2014) NOTCH1, SF3B1, BIRC3 and TP53 mutations in patients with chronic lymphocytic leukemia undergoing first-line treatment: correlation with biological parameters and response to treatment. *Leukemia & Lymphoma* 55(12): 2785–92.

Chigrinova E, Rinaldi A, Kwee I, Rossi D, Rancoita PM V, Strefford JC, Oscier DG, Stamatopoulos K, Papadaki T, Berger F, Young KH, Murray F, Rosenquist R, Greiner TC, Chan WC, Orlandi EM, Lucioni M, Marasca R, Inghirami G, Ladetto M, Forconi F, Cogliatti S, Votavova H, Swerdlow SH, Stilgenbauer S, Piris MA, Matolcsy A, Spagnolo D, Nikitin E, Zamò A, Gattei V, Bhagat G, Ott G, Zucca E, Gaidano G and Bertoni F (2013) Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. *Blood* 122(15): 2673–82.

Chiorazzi N and Ferrarini M (2003) B Cell Chronic Lymphocytic Leukemia : Lessons Learned from Studies of the B Cell Antigen Receptor Biases in V Gene Use. *Annual Review of Immunology* 21: 841–94.

Chiorazzi N and Ferrarini M (2011) Cellular origins of chronic lymphocytic leukemia : cautionary notes and additional considerations and possibilities. *Blood* 117(6): 1781–1791.

Choi AMK, Ryter SW and Levine B (2013) Autophagy in human health and disease. *The New England Journal of Medicine* 368(7): 651–62.

Clark MR, Massenburg D, Siemasko K, Hou P and Zhang M (2004) B-cell antigen receptor signaling requirements for targeting antigen to the MHC class II presentation pathway. *Current Opinion in Immunology* 16(3): 382–387.

Clifford R, Louis T, Robbe P, Ackroyd S, Burns A, Timbs AT, Wright Colopy G, Dreau H, Sigaux F, Judde JG, Rotger M, Telenti A, Lin YL, Pasero P, Maelfait J, Titsias M, Cohen DR, Henderson SJ, Ross MT, Bentley D, Hillmen P, Pettitt AR, Rehwinkel J, Knight SJL, Taylor JC, Crow YJ, Benkirane M and Schuh A (2014) SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood* 123(7): 1021–1031.

CLL Trialists' Collaborative Group (1999) Chemotherapeutic options in chronic lymphocytic leukemia: a meta-analysis of the randomized trials. CLL Trialists' Collaborative Group. *Journal of the National Cancer Institute* 91(10): 861–8.

Corcelle-termeau E, Vindeløv SD and Hämälistö S (2016) Excess sphingomyelin disturbs ATG9A trafficking and autophagosome closure. *Autophagy* 12(5): 1–17.

Crespo M, Bosch F, Villamor N, Bellosillo B, Colomer D, Rozman M, Marcé S, López-Guillermo A, Campo E and Montserrat E (2003) ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *The New England Journal of Medicine* 348(18): 1764–1775.

Dagklis A, Fazi C, Sala C, Cantarelli V, Scielzo C, Massacane R, Toniolo D, Caligaris-Cappio F, Stamatopoulos K and Ghia P (2008) The immunoglobulin gene repertoire of low-count chronic lymphocytic leukemia (CLL)-like monoclonal B lymphocytosis is different from CLL: diagnostic implications for clinical monitoring. *Blood* 114: 26–32.

Damle RN, Ghiotto F, Valetto A, Albesiano E, Fais F, Yan XJ, Sison CP, Allen SL, Kolitz J, Schulman P, Vinciguerra VP, Budde P, Frey J, Rai KR, Ferrarini M and Chiorazzi N (2002) B-cell chronic lymphocytic leukemia cells express a surface membrane phenotype of activated, antigen-experienced B lymphocytes. *Blood* 99(11): 4087–4093.

Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, Buchbinder A, Budman D, Dittmar K, Kolitz J, Lichtman SM, Schulman P, Vinciguerra VP, Rai KR, Ferrarini M and Chiorazzi N (1999) Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 94(6): 1840–1847.

Damm F, Mylonas E, Cosson A, Yoshida K, Della Valle V, Mouly E, Diop M, Scourzic L, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Kikushige Y, Davi F, Lambert J,

Gautheret D, Merle-Béral H, Sutton L, Dessen P, Solary E, Akashi K, Vainchenker W, Mercher T, Droin N, Ogawa S, Nguyen-Khac F and Bernard OA (2014) Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discovery* 4(9): 1088–1101.

Darlow JM and Stott DI (2006) Gene conversion in human rearranged immunoglobulin genes. *Immunogenetics* 58(7): 511–522.

Darzentas N, Hadzidimitriou A, Murray F, Hatzi K, Josefsson P, Laoutaris N, Moreno C, Anagnostopoulos A, Jurlander J, Tsaftaris a, Chiorazzi N, Belessi C, Ghia P, Rosenquist R, Davi F and Stamatopoulos K (2009) A different ontogenesis for chronic lymphocytic leukemia cases carrying stereotyped antigen receptors: molecular and computational evidence. *Leukemia* 24(1): 125–132.

Davis Z, Forconi F, Parker A, Gardiner A, Thomas P, Catovsky D, Rose-Zerilli M, Strefford JC and Oscier DG (2016) The outcome of Chronic lymphocytic leukaemia patients with 97% IGHV gene identity to germline is distinct from cases with <97% identity and similar to those with 98% identity. *British Journal of Haematology* 173(1): 127–136.

Diao J, Liu R, Rong Y, Zhao M, Zhang J, Lai Y, Zhou Q, Wilz LM, Li J, Vivona S, Pfuetzner RA, Brunger AT and Zhong Q (2015) ATG14 promotes membrane tethering and fusion of autophagosomes to endolysosomes. *Nature* 520(7548): 563–566.

Diaz Jr L a., Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, Allen B, Bozic I, Reiter JG, Nowak M a., Kinzler KW, Oliner KS and Vogelstein B (2012) The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 486(7404): 4–7.

Dohner H, Fischer K, Bentz M, Hansen K, Benner A, Cabot G, Diehl D, Schlenk R, Coy J, Stilgenbauer S and *et al*. (1995) p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. *Blood* 85(6): 1580–1589.

Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, Dohner K, Bentz M and Lichter P (2000) Genomic aberrations and survival in chronic lymphocytic leukemia. *New England Journal of Medicine* 343(26): 1910–1916.

Dohner H, Stilgenbauer S, Dohner K, Bentz M, Lichter P, Döhner H and Döhner K (1999) Chromosome aberrations in B-cell chronic lymphocytic leukemia:

reassessment based on molecular cytogenetic analysis. *Journal of Molecular Medicine* 77(2): 266–81.

Doidge R, Mittal S, Aslam A and Winkler GS (2012) The Anti-Proliferative Activity of BTG/TOB Proteins Is Mediated via the Caf1a (CNOT7) and Caf1b (CNOT8) Deadenylase Subunits of the Ccr4-Not Complex. *PLoS ONE* 7(12):e51331.

van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, Garcia-Sanz R, van Krieken JH, Droese J, Gonzalez D, Bastard C, White HE, Spaargaren M, Gonzalez M, Parreira A, Smith JL, Morgan GJ, Kneba M and Macintyre EA (2003) Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 17: 2257–2317.

Dooley AL, Winslow MM, Chiang DY, Banerji S, Stransky N, Dayton TL, Snyder EL, Senna S, Whittaker C a, Bronson RT, Crowley D, Barretina J, Garraway L, Meyerson M and Jacks T (2011) Nuclear factor I / B is an oncogene in small cell lung cancer. *Genes and Development* 25: 1470–1475.

Dooley HC, Razi M, Polson HEJ, Girardin SE, Wilson MI and Tooze S (2014) WIPI2 Links LC3 Conjugation with PI3P, Autophagosome Formation, and Pathogen Clearance by Recruiting Atg12-5-16L1. *Molecular Cell*. 55(2): 238–252.

Dores GM, Anderson WF, Curtis RE, Landgren O, Ostroumova E, Bluhm EC, Rabkin CS, Devesa SS and Linet MS (2007) Chronic lymphocytic leukaemia and small lymphocytic lymphoma: overview of the descriptive epidemiology. *British Journal of Haematology* 139(5): 809–19.

Van Dyke DL, Shanafelt TD, Call TG, Zent CS, Smoley S a, Rabe KG, Schwager SM, Sonbert JC, Slager SL and Kay NE (2009) A comprehensive evaluation of the prognostic significance of 13q deletions in patients with B-chronic lymphocytic leukaemia. *British Journal of Haematology* 148(4): 544–550.

Van Dyke DL, Werner L, Rassenti LZ, Neuberg D, Ghia EM, Heerema NA, Dal Cin P, Dell Aquila M, Sreekantaiah C, Greaves AW, Kipps TJ and Kay NE (2016) The Dohner fluorescence in situ hybridization prognostic classification of chronic lymphocytic leukaemia (CLL): The CLL Research Consortium experience. *British Journal of Haematology* 173(1): 105–113.

Edelmann J, Holzmann K, Miller F, Winkler D, Bühler A, Zenz T, Bullinger L, Kühn MWM, Gerhardinger A, Bloehdorn J, Radtke I, Su X, Ma J, Pounds S, Hallek M,

Lichter P, Korbel J, Busch R, Mertens D, Downing JR, Stilgenbauer S, Döhner H, Dc W, Edelmann J, Holzmann K, Miller F, Winkler D, Bu A, Zenz T, Bullinger L, Stilgenbauer S and Do H (2013) High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood* 120(24): 4783–4794.

Ehrenstein MR and Notley C a (2010) The importance of natural IgM: scavenger, protector and regulator. *Nature Reviews of Immunology* 10(11): 778–86.

El-Khoury V, Pierson S, Szwarcbart E, Brons NHC, Roland O, Cherrier-De Wilde S, Plawny L, Van Dyck E and Berchem G (2014) Disruption of autophagy by the histone deacetylase inhibitor MGCD0103 and its therapeutic implication in B-cell chronic lymphocytic leukemia. *Leukemia* 28(8): 1636–46.

Eskelinen E-L (2011) The dual role of autophagy in cancer. *Current Opinion in Pharmacology* 11(4): 294–300.

Eyre TA, Clifford R, Bloor A, Boyle L, Roberts C, Cabes M, Collins GP, Devereux S, Follows G, Fox CP, Gribben J, Hillmen P, Hatton CS, Littlewood TJ, Mccarthy H, Murray J, Pettitt AR, Soilleux E, Stamatopoulos B, Love SB, Wotherspoon A and Schuh A (2016) NCRI phase II study of CHOP in combination with ofatumumab in induction and maintenance in newly diagnosed Richter syndrome. *British Journal of Haematology* 175(1):43-54

Fabbri G, Khiabanian H, Holmes AB, Wang J, Messina M, Mullighan CG, Pasqualucci L, Rabadan R and Dalla-Favera R (2013) Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome. *The Journal of Experimental Medicine* 210(11): 2273–88.

Fais F, Ghiotto F, Hashimoto S, Sellars B, Valetto A, Allen SL, Schulman P, Vinciguerra VP, Rai K, Rassenti LZ, Kipps TJ, Dighiero G, Schroeder HW, Ferrarini M and Chiorazzi N (1998) Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *Journal of Clinical Investigation* 102(8): 1515–1525.

Falt S, Merup M, Tobin G, Thunberg U, Rosenquist R and Wennborg A (2005) Distinctive gene expression pattern in V H 3-21 utilizing B-cell chronic lymphocytic leukemia V H gene analysis. *Blood* 106(2): 681–689.

Farooqui MZH, Valdez J, Martyr S, Aue G, Saba N, Niemann CU, Herman SEM, Tian X, Marti G, Soto S, Hughes TE, Jones J, Lipsky A, Pittaluga S, Stetler-Stevenson M, Yuan C, Lee YS, Pedersen LB, Geisler CH, Calvo KR, Arthur DC, Maric I, Childs R,

Young NS and Wiestner A (2015) Ibrutinib for previously untreated and relapsed or refractory chronic lymphocytic leukaemia with TP53 aberrations: A phase 2, single-arm trial. *The Lancet Oncology* 16(2): 169–176.

Fimia GM, Stoykova A, Romagnoli A, Giunta L, Di Bartolomeo S, Nardacci R, Corazzari M, Fuoco C, Ucar A, Schwartz P, Gruss P, Piacentini M, Chowdhury K and Cecconi F (2007) Ambra1 regulates autophagy and development of the nervous system. *Nature* 447(7148): 1121–1125.

Fischer K, Bahlo J, Fink AM, Goede V, Herling CD, Cramer P, Langerbeins P, Von Tresckow J, Engelke A, Maurer C, Kovacs G, Herling M, Tausch E, Kreuzer KA, Eichhorst B, Böttcher S, Seymour JF, Ghia P, Marlton P, Kneba M, Wendtner CM, Döhner H, Stilgenbauer S and Hallek M (2016) Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: Updated results of the CLL8 trial. *Blood* 127(2): 208–215.

Foà R, Del Giudice I, Guarini A, Rossi D and Gaidano G (2013) Clinical implications of the molecular genetics of chronic lymphocytic leukemia. *Haematologica* 98(5): 675–685.

Frasca D and Blomberg BB (2009) Effects of aging on B cell function. *Current Opinion in Immunology* 21(4): 425–430.

Fraser LD, Zhao Y, Lutalo PMK, D'Cruz DP, Cason J, Silva JS, Dunn-Walters DK, Nayar S, Cope AP and Spencer J (2015) Immunoglobulin light chain allelic inclusion in systemic lupus erythematosus. *European Journal of Immunology* 45(8): 2409–2419.

Fujita N, Itoh T, Omori H, Fukuda M, Noda T and Yoshimori T (2008) The Atg16L complex specifies the site of LC3 lipidation for membrane biogenesis in autophagy. *Molecular Biology of the Cell* 19(5): 2092–100.

Furman RR, Sharman JP, Coutre SE, Cheson BD, Pagel JM, Hillmen P, Barrientos JC, Zelenetz AD, Kipps TJ, Flinn I, Ghia P, Eradat H, Ervin T, Lamanna N, Coiffier B, Pettitt AR, Ma S, Stilgenbauer S, Cramer P, Aiello M, Johnson DM, Miller LL, Li D, Jahn TM, Dansey RD, Hallek M and O'Brien SM (2014) Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *The New England Journal of Medicine* 370(11): 997–1007.

Ganley IG, Lam DH, Wang J, Ding X, Chen S and Jiang X (2009) ULK1.ATG13.FIP200 complex mediates mTOR signaling and is essential for autophagy. *Journal of Biological Chemistry* 284(18): 12297–305.

Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, Tsai P-C, Casasent A, Waters J, Zhang H, Meric-Bernstam F, Michor F and Navin NE (2016) Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nature Genetics* 48(10): 1–15.

Ge L, Melville D, Zhang M and Schekman R (2013) The ER-Golgi intermediate compartment is a key membrane source for the LC3 lipidation step of autophagosome biogenesis. *eLife* 2: e00947.

Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose S a., Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, Purcell SM, Svantesson O, Landén M, Höglund M, Lehmann S, Gabriel SB, Moran JL, Lander ES, Sullivan PF, Sklar P, Grönberg H, Hultman CM, McCarroll S a., Ph D, Rose S a., Bakhoum SF, Ph D, Chambert K, Mick E, Neale BM, Ph D, Fromer M, Ph D, Purcell SM, Ph D, Svantesson O, Sullivan PF, Sklar P, Ph D, Grönberg H, Ph D, Hultman CM, Ph D, McCarroll S a. and Ph D (2014) Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine* 371(26): 2477–2487.

Gentile M, Shanafelt TD, Rossi D, Laurenti L, Mauro FR, Molica S, Cutrona G, Uccello G, Campanelli M, Vigna E, Tripepi G, Chaffee KG, Parikh SA, Bossio S, Recchia AG, Innocenti I, Pasquale R and Neri A (2016) Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood* 128(16): 2093–2096.

Gentile M, Zirlik K, Ciolli S, Mauro FR, Di Renzo N, Mastrullo L, Angrilli F, Molica S, Tripepi G, Giordano A, Di Raimondo F, Selleri C, Coscia M, Musso M, Orsucci L, Mannina D, Rago A, Giannotta A, Ferrara F, Herishanu Y, Shvidel L, Tadmor T, Scortechini I, Ilariucci F, Murru R, Guarini A, Musuraca G, Mineo G, Vincelli I, Arcari A, Tarantini G, Caparrotti G, Chiarenza A, Levato L, Villa MR, De Paolis MR, Zinzani PL, Polliack A and Morabito F (2016) Combination of bendamustine and rituximab as front-line therapy for patients with chronic lymphocytic leukaemia: Multicenter, retrospective clinical practice experience with 279 cases outside of controlled clinical trials. *European Journal of Cancer* 60: 154–165.

Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, Fisher R, McGranahan N, Matthews N, Santos CR, Martinez P, Phillimore B, Begum S, Rabinowitz A, Spencer-Dene B, Gulati S, Bates PA, Stamp G, Pickering L, Gore M, Nicol DL, Hazell S, Futreal PA, Stewart A and Swanton C (2014) Genomic

architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature Genetics* 46(3): 225–33.

Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler AP, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA and Swanton C (2012) Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine* 366(10): 883–892.

Ghia EM, Jain S, Widhopf G. F. 2nd, Rassenti LZ, Keating MJ, Wierda WG, Gribben JG, Brown JR, Rai KR, Byrd JC, Kay NE, Greaves AW and Kipps TJ (2008) Use of IGHV3-21 in chronic lymphocytic leukemia is associated with high-risk disease and reflects antigen-driven, post-germinal center leukemogenic selection. *Blood* 111(10): 5101–5108.

Ghia P, Stamatopoulos K, Belessi C, Moreno C, Stella S, Guida G, Michel A, Crespo M, Laoutaris N, Montserrat E, Anagnostopoulos A, Dighiero G, Fassas A, Caligaris-Cappio F and Davi F (2005) Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: The lesson of the IGHV3-21 gene. *Blood* 105(4): 1678–1685.

Ghia P, Stamatopoulos K, Belessi C, Moreno C, Stilgenbauer S, Stevenson FK, Davi F and Rosenquist R (2007) ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia* 21(1): 1–3.

Ghiotto F, Fais F and Valetto A (2004) Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. *Journal of Clinical Investigation* 113(7):1008-1016

Gibson KL, Wu YC, Barnett Y, Duggan O, Vaughan R, Kondeatis E, Nilsson BO, Wikby A, Kipling D and Dunn-Walters DK (2009) B-cell diversity decreases in old age and is correlated with poor health status. *Aging Cell* 8(1): 18–25.

Del Giudice I, Rossi D, Chiaretti S, Marinelli M, Tavolaro S, Gabrielli S, Laurenti L, Marasca R, Rasi S, Fangazio M, Guarini A, Gaidano G and Foà R (2012) NOTCH1 mutations in +12 chronic lymphocytic leukemia (CLL) confer an unfavorable prognosis, induce a distinctive transcriptional profiling and refine the intermediate prognosis of +12 CLL. *Haematologica* 97(3): 437–441.

Goede V, Fischer K, Busch R, Engelke A, Eichhorst B, Wendtner CM, Chagorova T, de la Serna J, Dilhuydy M-S, Illmer T, Opat S, Owen CJ, Samoylova O, Kreuzer K-A, Stilgenbauer S, Döhner H, Langerak AW, Ritgen M, Kneba M, Asikanius E, Humphrey K, Wenger M and Hallek M (2014) Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *The New England Journal of Medicine* 370(12): 1101–10.

Gonzalez D, Martinez P, Wade R, Hockley S, Oscier DG, Matutes E, Dearden CE, Richards SM, Catovsky D and Morgan GJ (2011) Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: Results from the LRF CLL4 trial. *Journal of Clinical Oncology* 29(16): 2223–2229.

Gorman JR, van der Stoep N, Monroe R, Cogne M, Davidson L and Alt FW (1996) The Ig(kappa) enhancer influences the ratio of Ig(kappa) versus Ig(lambda) B lymphocytes. *Immunity* 5(3): 241–252.

Gross SA, Zhu X, Bao L, Ryder J, Le A, Chen Y, Wang XQ and Irons RD (2008) A prospective study of 728 cases of non-Hodgkin lymphoma from a single laboratory in Shanghai, China. *International Journal of Hematology* 88(2): 165–173.

Grubor V, Krasnitz A, Troge JE, Meth JL, Lakshmi B, Kendall JT, Yamrom B, Alex G, Pai D, Navin N, Hufnagel LA, Lee Y-HH, Cook K, Allen SL, Rai KR, Damle RN, Calissano C, Chiorazzi N, Wigler M and Esposito D (2008) Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). *Blood* 113(6): 1294–1303.

Grundy GJ, Ramon-Maiques S, Dimitriadis EK, Kotova S, Biertumpfel C, Heymann JB, Steven AC, Gellert M and Yang W (2009) Initial Stages of V(D)J Recombination: The Organization of RAG1/2 and RSS DNA in the Postcleavage Complex. *Molecular Cell* 35(2): 217–227.

Guièze R, Robbe P, Clifford R, De Guibert S, Pereira B, Timbs A, Dilhuydy MS, Cabes M, Ysebaert L, Burns A, Nguyen-Khac F, Davi F, Véronèse L, Combes P, Garff-Tavernier M Le, Leblond V, Merle-Béral H, Alsolami R, Hamblin A, Mason J, Pettitt AR, Hillmen P, Taylor J, Knight SJL, Tournilhac O and Schuh A (2015) Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. *Blood* 126(18): 2110–2117.

Gunn SR, Bolla AR, Barron LL, Gorre ME, Mohammed MS, Bahler DW, Mellink CH, van Oers MH, Keating MJ, Ferrajoli A, Coombes KR, Abruzzo L V and Robetorye RS (2008) Array CGH analysis of chronic lymphocytic leukemia reveals frequent cryptic monoallelic and biallelic deletions of chromosome 22q11 that include the PRAME gene. *Leukemia Research* 33: 1276–1281.

Gunnarsson R, Isaksson a, Mansouri M, Göransson H, Jansson M, Cahill N, Rasmussen M, Staaf J, Lundin J, Norin S, Buhl AM, Smedby KE, Hjalgrim H, Karlsson K, Jurlander J, Juliusson G and Rosenquist R (2010) Large but not small copy-number alterations correlate to high-risk genomic aberrations and survival in chronic lymphocytic leukemia: a high-resolution genomic screening of newly diagnosed patients. *Leukemia* 24(1): 211–5.

Gunnarsson R, Mansouri L, Isaksson A, Göransson H, Cahill N, Jansson M, Rasmussen M, Lundin J, Norin S, Buhl AM, Smedby KE, Hjalgrim H, Karlsson K, Jurlander J, Geisler C, Juliusson G and Rosenquist R (2011) Array-based genomic screening at diagnosis and during follow-up in chronic lymphocytic leukemia. *Haematologica* 96(8): 1161–9.

Gunnarsson R, Staaf J, Jansson M, Ottesen AM, Göransson H, Liljedahl U, Ralfkiaer U, Mansouri M, Buhl AM, Smedby KE, Hjalgrim H, Syvänen A-C, Borg A, Isaksson A, Jurlander J, Juliusson G and Rosenquist R (2008) Screening for copy-number alterations and loss of heterozygosity in chronic lymphocytic leukemia--a comparative study of four differently designed, high resolution microarray platforms. *Genes, Chromosomes & Cancer* 47(8): 697–711.

Guo JY, Chen HY, Mathew R, Fan J, Strohecker AM, Karsli-Uzunbas G, Kamphorst JJ, Chen G, Lemons JMS, Karantza V, Coller HA, DiPaola RS, Gelinas C, Rabinowitz JD and White E (2011) Activated Ras requires autophagy to maintain oxidative metabolism and tumorigenesis. *Genes and Development* 25(5): 460–470.

Hailey DW, Rambold AS, Satpute-Krishnan P, Mitra K, Sougrat R, Kim PK and Lippincott-Schwartz J (2010) Mitochondria Supply Membranes for Autophagosome Biogenesis during Starvation. *Cell* 141(4): 656–667.

Hallek M, Fischer K, Fingerle-Rowson G, Fink AM, Busch R, Mayer J, Hensel M, Hopfinger G, Hess G, Von Grunhagen U, Bergmann MA, Catalano J, Zinzani PL, Caligaris-Cappio F, Seymour JF, Berrebi A, Jager U, Cazin B, Trneny M, Westermann A, Wendtner CM, Eichhorst BF, Staib P, Buhler A, Winkler D, Zenz T, Bottcher S, Ritgen M, Mendila M, Kneba M, Dohner H and Stilgenbauer S (2010)

Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: A randomised, open-label, phase 3 trial. *The Lancet* 376(9747): 1164–1174.

Hamasaki M, Furuta N, Matsuda A, Nezu A, Yamamoto A, Fujita N, Oomori H, Noda T, Haraguchi T, Hiraoka Y, Amano A and Yoshimori T (2013) Autophagosomes form at ER-mitochondria contact sites. *Nature* 495(7441): 389–93.

Hamblin TJ (2003) CD38: what is it there for? *Blood* 102(6): 1939–1940.

Hamblin TJ, Davis Z, Gardiner A, Oscier DG and Stevenson FK (1999) Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* 94(6): 1848–1854.

Hamblin TJ, Davis Z, Gardiner A, Oscier DG and Stevenson FK (1999) Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* 94(6): 1848–1854.

Hangaishi A and Kurokawa M (2010) Blimp-1 is a tumor suppressor gene in lymphoid malignancies. *International Journal of Hematology* 91(1): 46–53.

Hara T, Takamura A, Kishi C, Iemura SI, Natsume T, Guan JL and Mizushima N (2008) FIP200, a ULK-interacting protein, is required for autophagosome formation in mammalian cells. *Journal of Cell Biology* 181(3): 497–510.

Hayashi-Nishino M, Fujita N, Noda T, Yamaguchi A, Yoshimori T and Yamamoto A (2009) A subdomain of the endoplasmic reticulum forms a cradle for autophagosome formation. *Nature Cell Biology* 11(12): 1433–7.

Helmink BA and Sleckman BP (2012) The Response to and Repair of RAG-Mediated DNA Double-Strand Breaks. *Annual Review of Immunology* 30(1): 175–202.

Henriques A, Rodríguez-Caballero A, Criado I, Langerak AW, Nieto WG, Lécrevisse Q, González M, Cortesão E, Paiva A, Almeida J and Orfao A (2014) Molecular and cytogenetic characterization of expanded B-cell clones from multiclonal versus monoclonal B-cell chronic lymphoproliferative disorders. *Haematologica* 99(5): 897–907.

Herling CD, Klaumunzer M, Krings Rocha C, Altmuller J, Thiele H, Bahlo J, Kluth S, Crispatzu G, Herling M, Schiller J, Engelke A, Tausch E, Dohner H, Fischer K, Goede V, Nurnberg P, Reinhardt HC, Stilgenbauer S, Hallek M and Kreuzer KA (2016)

Complex karyotypes, KRAS and POT1 mutations impact outcome in CLL after chlorambucil based chemo- or chemoimmunotherapy. *Blood* 128:395-404.

Hernández JÁ, Hernández-Sánchez M, Rodríguez-Vicente AE, Grossmann V, Collado R, Heras C, Puiggros A, Martín AÁ, Puig N, Benito R, Robledo C, Delgado J, González T, Queizán JA, Galende J, De La Fuente I, Martín-Núñez G, Alonso JM, Abrisqueta P, Luño E, Marugán I, González-Gascón I, Bosch F, Kohlmann A, González M, Espinet B and Hernández-Rivas JM (2015) A low frequency of losses in 11q chromosome is associated with better outcome and lower rate of genomic mutations in patients with chronic lymphocytic leukemia. *PLoS ONE* 10(12): 1–16.

Hesslein DGT and Schatz DG (2001) *Factors and forces controlling V(D)J recombination*. *Advances in Immunology* 78:169-232

Hewitt SL, Yin B, Ji Y, Chaumeil J, Marszalek K, Tenthorey J, Salvagiotto G, Steinel N, Ramsey LB, Ghysdael J, Farrar M a, Sleckman BP, Schatz DG, Busslinger M, Bassing CH and Skok J a (2009) RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nature Immunology* 10(6): 655–664.

Hillmen P, Robak T, Janssens A, Babu KG, Kloczko J, Grosicki S, Doubek M, Panagiotidis P, Kimby E, Schuh A, Pettitt AR, Boyd T, Montillo M, Gupta I V., Wright O, Dixon I, Carey JL, Chang CN, Lisby S, McKeown A and Offner F (2015) Chlorambucil plus ofatumumab versus chlorambucil alone in previously untreated patients with chronic lymphocytic leukaemia (COMPLEMENT 1): A randomised, multicentre, open-label phase 3 trial. *The Lancet* 385(9980): 1873–1883.

Hosokawa N, Hara T, Kaizuka T, Kishi C, Takamura A, Miura Y, Iemura S, Natsume T, Takehana K, Yamada N, Guan J, Oshiro N and Mizushima N (2009) Nutrient-dependent mTORC1 Association with the ULK1 – Atg13 – FIP200 Complex Required for Autophagy. *Molecular Biology of the Cell* 20: 1981–1991.

Hosokawa N, Sasaki T, Iemura SI, Natsume T, Hara T and Mizushima N (2009) Atg101, a novel mammalian autophagy protein interacting with Atg13. *Autophagy* 5(7): 973–979.

Hou W, Han J, Lu C, Goldstein LA and Rabinowich H (2010) Autophagic degradation of active caspase-8: A crosstalk mechanism between autophagy and apoptosis. *Autophagy* 6(7): 891–900.

Hurtado AM, Chen-Liang T-H, Przychodzen B, Hamedi C, Muñoz-Ballester J, Dienes B, García-Malo MD, Antón AI, de Arriba F, Teruel-Montoya R, Ortuño FJ, Vicente V,

Maciejewski JP and Jerez A (2015) Prognostic signature and clonality pattern of recurrently mutated genes in inactive chronic lymphocytic leukemia. *Blood Cancer Journal* 5: e342.

Hwang J, Alt FW and Yeap L-S (2014) Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination. *Microbiology Spectrum* 3(2): 1–26.

Itakura E, Kishi-Itakura C and Mizushima N (2012) The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. *Cell* 151(6): 1256–69.

Jain M *et al* (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* 17(1): 239. 1103-0.

Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman P V, Mar BG, Lindsley RC, Mermel CH, Burtt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen H a, Ladenvall C, Getz G, Correa A, Banahan BF, Gabriel S, Kathiresan S, Stringham HM, McCarthy MI, Boehnke M, Tuomilehto J, Haiman C, Groop L, Atzmon G, Wilson JG, Neuberg D, Altshuler D and Ebert BL (2014) Age-related clonal hematopoiesis associated with adverse outcomes. *The New England Journal of Medicine* 371(26): 2488–98.

Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, Alpermann T, Roller A, Kohlmann A, Haferlach T, Kern W and Schnittger S (2014) SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia* 28(1): 108–17.

Jethwa A, Hullein J, Stolz T, Blume C, Sellner L, Jauch A, Sill M, Kater AP, te Raa GD, Geisler C, van Oers M, Dietrich S, Dreger P, Ho AD, Paruzynski A, Schmidt M, von Kalle C, Glimm H and Zenz T (2013) Targeted resequencing for analysis of clonal composition of recurrent gene mutations in chronic lymphocytic leukaemia. *British Journal of Haematology* 163(4): 496–500.

Jiang H, Chang FC, Ross AE, Lee J, Nakayama K, Nakayama K and Desiderio S (2005) Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle. *Molecular Cell* 18(6): 699–709.

Jiang P, Nishimura T, Sakamaki Y, Itakura E, Hatta T, Natsume T and Mizushima N (2014) The HOPS complex mediates autophagosome-lysosome fusion through interaction with syntaxin 17. *Molecular Biology of the Cell* 25(8): 1327–1337.

Johnson K, Chaumeil J, Micsinai M, Wang JMH, Ramsey LB, Baracho G V., Rickert RC, Strino F, Kluger Y, Farrar MA and Skok JA (2012) IL-7 Functionally Segregates the Pro-B Cell Stage by Regulating Transcription of Recombination Mediators across Cell Cycle. *The Journal of Immunology* 188(12): 6084–6092.

Jones J and Gellert M (2002) Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. *EMBO Journal* 21(15): 4162–4171.

Jones S, Chen W-D, Parmigiani G, Diehl F, Beerenwinkel N, Antal T, Traulsen A, Nowak MA, Siegel C, Velculescu VE, Kinzler KW, Vogelstein B, Willis J and Markowitz SD (2008) Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences of the United States of America* 105(11): 4283–8.

Jordens I, Fernandez-Borja M, Marsman M, Dusseljee S, Janssen L, Calafat J, Janssen H, Wubbolts R and Neefjes J (2001) The Rab7 effector protein RILP controls lysosomal transport by inducing the recruitment of dynein-dynactin motors. *Current Biology* 11(21): 1680–1685.

Kabeya Y, Mizushima N, Yamamoto A, Oshitani-Okamoto S, Ohsumi Y and Yoshimori T (2004) LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *Journal of Cell Science* 117(13): 2805–12.

Kanehisa M, Sato Y, Kawashima M, Furumichi M and Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* 44(D1): D457–D462.

Kasar S, Underbayev C, Yuan Y, Hanlon M, Aly S, Khan H, Chang V, Batish M, Gavrilova T, Badiane F, Degheidy H, Marti G and Raveche E (2014) Therapeutic Implications of Activation of the Host Gene (Dleu2) Promoter for miR-15a/16-1 in Chronic Lymphocytic Leukemia (CLL). *Oncogene* 33(25): 3307–3315.

Kaveri S V, Silverman GJ and Bayry J (2012) Natural IgM in immune equilibrium and harnessing their therapeutic potential. *Journal of Immunology* 188(3): 939–45.

Kay NE, Eckel-Passow JE, Braggio E, Vanwier S, Shanafelt TD, Van Dyke DL, Jelinek DF, Tschumper RC, Kipps T, Byrd JC and Fonseca R (2010) Progressive but previously untreated CLL patients with greater array CGH complexity exhibit a less durable response to chemoimmunotherapy. *Cancer Genetics and Cytogenetics* 203(2): 161–8.

Kern W, Bacher U, Schnittger S, Dicker F, Alpermann T, Haferlach T and Haferlach C (2014) Flow cytometric identification of 76 patients with biclonal disease among 5523 patients with chronic lymphocytic leukaemia (B-CLL) and its genetic characterization. *British Journal of Haematology* 164(4): 565–569.

Kienle D, Benner A, Krober A, Winkler D, Mertens D, Buhler A, Seiler T, Jager U, Lichter P, Dohner H and Stilgenbauer S (2005) Distinct gene expression patterns in chronic lymphocytic leukemia defined by usage of specific VH genes. *Blood* 107: 2090–2093.

Kienle D, Benner A, Kröber A, Winkler D, Mertens D, Bühler A, Seiler T, Jäger U, Lichter P, Döhner H, Stilgenbauer S, Dc W, Kienle D, Benner A, Kro A, Winkler D, Mertens D, Bu A, Seiler T and Ja U (2006) Distinct gene expression patterns in chronic lymphocytic leukemia defined by usage of specific VH genes. *Blood* 107(5): 2090–2093.

Kikushige Y, Ishikawa F, Miyamoto T, Shima T, Urata S, Yoshimoto G, Mori Y, Iino T, Yamauchi T, Eto T, Niiro H, Iwasaki H, Takenaka K and Akashi K (2011) Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. *Cancer Cell* 20(2): 246–259.

Kim M-S, Lapkouski M, Yang W and Gellert M (2015) Crystal structure of the V(D)J recombinase RAG1–RAG2. *Nature* 518(7540): 507–511.

Kirkham PM, Mortari F, Newton J a and Schroeder HW (1992) Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. *EMBO* 11(2): 603–9.

Klein U, Tu Y, Stolovitzky GA, Mattioli M, Cattoretti G, Husson H, Freedman A, Inghirami G, Cro L, Baldini L, Neri A, Califano A and Dalla-Favera R (2001) Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *Journal of Experimental Medicine* 194(11): 1625–1638.

Klinger M, Zheng J, Elenitoba-Johnson KSJ, Perkins SL, Faham M and Bahler DW (2015) Next-generation IgVH sequencing CLL-like monoclonal B-cell lymphocytosis reveals frequent oligoclonality and ongoing hypermutation. *Leukemia*. 30: 1–28.

Knight SJL, Yau C, Clifford R, Timbs AT, Sadighi Akha E, Dréau HM, Burns A, Ciria C, Oscier DG, Pettitt AR, Dutton S, Holmes CC, Taylor J, Cazier J-B and Schuh A (2012) Quantification of subclonal distributions of recurrent genomic aberrations in

paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia. *Leukemia* 26, 1564 - 1575.

Kopan R and Ilagan MXG (2009) The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism. *Cell* 137(2): 216–233.

Koukourakis MI, Kalamida D, Mitrakas A, Pouliliou S, Kalamida S, Sivridis E and Giatromanolaki A (2015) Intensified autophagy compromises the efficacy of radiotherapy against prostate cancer. *Biochemical and Biophysical Research Communications* 461: 268–274.

Kovaleva V, Mora R, Park YJ, Plass C, Chiramel AI, Bartenschlager R, Döhner H, Stilgenbauer S, Pscherer A, Lichter P and Seiffert M (2012) miRNA-130a targets ATG2B and DICER1 to inhibit autophagy and trigger killing of chronic lymphocytic leukemia cells. *Cancer Research* 72(7): 1763–72.

Kriangkum J, Motz SN, Mack T, Beiggi S, Baigorri E, Kuppusamy H, Belch AR, Johnston JB and Pilarski LM (2015) Single-Cell Analysis and Next-Generation Immuno-Sequencing Show That Multiple Clones Persist in Patients with Chronic Lymphocytic Leukemia. *Plos One* 10(9): e0137232.

Kristensen L, Kristensen T, Abildgaard N, Thomassen M, Frederiksen M, Mourits-Andersen T and Møller MB (2015) High expression of PI3K core complex genes is associated with poor prognosis in chronic lymphocytic leukemia. *Leukemia Research* 39(6): 555–560.

Kröber A, Seiler T, Benner A, Bullinger L, Brückle E, Lichter P, Döhner H, Stilgenbauer S, Krober A, Seiler T, Benner A, Bullinger L, Bruckle E, Lichter P, Dohner H and Stilgenbauer S (2002) V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood* 100(4): 1410–1416.

Kujawski L, Ouillette P, Erba H, Saddler C, Jakubowiak A, Kaminski M, Shedden K and Malek SN (2008) Genomic complexity identifies patients with aggressive chronic lymphocytic leukemia. *Blood* 112(5): 1993–2003.

Lanasa MC, Allgood SD, Volkheimer a D, Gockerman JP, Whitesides JF, Goodman BK, Moore JO, Weinberg JB and Levesque MC (2010) Single-cell analysis reveals oligoclonality among 'low-count' monoclonal B-cell lymphocytosis. *Leukemia* 24(1): 133–140.

Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Sivachenko A, Wang L, Wan Y, Zhang W, Shukla SA,

Vartanov A, Fernandes SM, Saksena G, Cibulskis K, Tesar B, Gabriel S, Hacohen N, Meyerson M, Lander ES, Neuberg D, Brown JR, Getz G and Wu CJ (2013) Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell* 152(4): 714–726.

Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, Kluth S, Bozic I, Lawrence M, Böttcher S, Carter SL, Cibulskis K, Mertens D, Sougnez CL, Rosenberg M, Hess JM, Edelmann J, Kless S, Kneba M, Ritgen M, Fink A, Fischer K, Gabriel S, Lander ES, Nowak MA, Döhner H, Hallek M, Neuberg D, Getz G, Stilgenbauer S and Wu CJ (2015) Mutations driving CLL and their evolution in progression and relapse. *Nature* 526(7574): 525–30.

Landgren O, Albitar M, Ma W, Abbasi F, Hayes RB, Ghia P, Marti GE and Caporaso NE (2009) B-Cell Clones as Early Markers for Chronic Lymphocytic Leukemia. *New England Journal of Medicine* 360(7): 659–667.

Lane DP (1992) Cancer. p53, guardian of the genome. *Nature* 358(6381): 15–16.

Langerak AW (2009) From the patient to the sequence. In: Ghia P, Rosenquist R and Davi F (eds) *Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia*. Milan: Wolters Kluwer Health, 23–31.

Langerak AW, Davi F, Ghia P, Hadzidimitriou A, Murray F, Potter KN, Rosenquist R, Stamatopoulos K and Belessi C (2011) Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. *Leukemia* 25(6): 979–84.

Lapkouski M, Chuenchor W, Kim MS, Gellert M and Yang W (2015) Assembly pathway and characterization of the RAG1/2-DNA paired and signal-end complexes. *Journal of Biological Chemistry* 290(23): 14618–14625.

Lee J and Desiderio S (1999) Cyclin A/CDK2 Regulates V(D)J Recombination by Coordinating RAG-2 Accumulation and DNA Repair. *Immunity* 11(6): 771–781.

Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, Hadi-Saljoqi S, Sasorith S, Lefranc G and Kossida S (2009) IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Research* 37: D1006–D1012

Lerner M, Harada M, Lovén J, Castro J, Davis Z, Oscier DG, Henriksson M, Sangfelt O, Grandér D and Corcoran MM (2009) DLEU2, frequently deleted in malignancy, functions as a critical host gene of the cell cycle inhibitory microRNAs miR-15a and miR-16-1. *Experimental Cell Research* 315(17): 2941–2952.

Lévy J, Cacheux W, Bara MA, L'Hermitte A, Lepage P, Fraudeau M, Trentesaux C, Lemarchand J, Durand A, Crain A-M, Marchiol C, Renault G, Dumont F, Letourneur F, Delacre M, Schmitt A, Terris B, Perret C, Chamaillard M, Couty J-P and Romagnolo B (2015) Intestinal inhibition of Atg7 prevents tumour initiation through a microbiome-influenced immune response and suppresses tumour growth. *Nature Cell Biology* 17(8): 1062–1073.

Levy JMM, Thompson JC, Griesinger AM, Amani V, Donson AM, Birks DK, Morgan MJ, Mirsky DM, Handler MH, Foreman NK and Thorburn A (2014) Autophagy inhibition improves chemosensitivity in BRAFV600E brain tumors. *Cancer Discovery* 4(7): 773–780.

Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS and Sherlock G (2015) Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519(7542): 181–6.

Li M, Diehl F, Dressman D, Vogelstein B and Kinzler KW (2006) BEAMing up for detection and quantification of rare sequence variants. *Nature Methods* 3(2): 95–97.

Liang C, Feng P, Ku B, Dotan I, Canaani D, Oh B-H and Jung JU (2006) Autophagic and tumour suppressor activity of a novel Beclin1-binding protein UVRAG. *Nature Cell Biology* 8(7): 688–99.

Lin K, Manocha S, Harris RJ, Matrai Z, Sherrington P and Pettitt AR (2003) High frequency of p53 dysfunction and low level of VH mutation in chronic lymphocytic leukemia patients using the Vh3-21 gene segment. *Blood* 102(3): 1145–1146.

Liu H *et al* (2013) Down-regulation of autophagy-related protein 5 (ATG5) contributes to the pathogenesis of early-stage cutaneous melanoma. *Science Translational Medicine* 5(202): 202ra123.

Liu Y, Shoji-Kawata S, Sumpter RM, Wei Y, Ginet V, Zhang L, Posner B, Tran K a, Green DR, Xavier RJ, Shaw SY, Clarke PGH, Puyal J and Levine B (2013) Autosis is a Na+,K+-ATPase-regulated form of cell death triggered by autophagy-inducing peptides, starvation, and hypoxia-ischemia. *Proceedings of the National Academy of Sciences of the United States of America* 110(51): 20364–71.

Ljungström V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K, Ntoufa S, Baliakas P, Clifford R, Sutton L-A, Blakemore SJ, Stavroyianni N, Agathangelidis A, Rossi D, Höglund M, Kotaskova J, Juliusson G, Belessi C, Chiorazzi N, Panagiotidis P, Langerak AW, Smedby KE, Oscier DG, Gaidano G, Schuh A, Davi F, Pott C,

Strefford JC, Trentin L, Pospisilova S, Ghia P, Stamatopoulos K, Sjöblom T and Rosenquist R (2016) Whole-exome sequencing in relapsing chronic lymphocytic leukemia: Clinical impact of recurrent RPS15 mutations. *Blood* 127(8): 1007–1016.

Loeb LA, Loeb KR and Anderson JP (2003) Multiple mutations and cancer. *Proceedings of the National Academy of Sciences of the United States of America* 100(3): 776–81.

Loman NJ, Misra R V, Dallman TJ, Constantinidou C, Gharbia SE, Wain J and Pallen MJ (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology* 30(5): 434–9.

Lozano-Santos C, García-Vela JA, Pérez-Sanz N, Nova-Gurumeta S, Fernandez-Cuevas B, Gomez-Lozano N, Sánchez-Beato M, Sanchez-Godoy P, Bueno JL and Garcia-Marco JA (2016) Biallelic *ATM* alterations detected at diagnosis identify a subset of treatment-naïve chronic lymphocytic leukemia patients with reduced overall survival similar to patients with p53 deletion. *Leukemia & Lymphoma* 8194: 1–7.

Lu H, Giordano F and Ning Z (2016) Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics* 14(5): 265–279.

Luning Prak ET, Monestier M and Eisenberg RA (2011) B cell receptor editing in tolerance and autoimmunity. *Annals of the New York Academy of Sciences* 1217(1): 96–121.

Ma XH, Piao S, Wang D, Mcafee QW, Nathanson KL, Lum JJ, Li LZ and Amaravadi RK (2011) Measurements of tumor cell autophagy predict invasiveness, resistance to chemotherapy, and survival in melanoma. *Clinical Cancer Research* 17(10): 3478–3489.

MacCallum SF, Groves MJ, James J, Murray K, Appleyard V, Prescott AR, Drbal A a, Nicolaou A, Cunningham J, Haydock S, Ganley IG, Westwood NJ, Coates PJ, Lain S and Tauro S (2013) Dysregulation of autophagy in chronic lymphocytic leukemia with the small-molecule Sirtuin inhibitor Tenovin-6. *Scientific Reports* 3: 1275.

Magnac C, Porcher R, Davi F, Nataf J, Payelle-Brogard B, Tang RP, Oppezzo P, Levy V, Dighiero G and Ajchenbaum-Cymbalista F (2003) Predictive value of serum thymidine kinase level for Ig-V mutational status in B-CLL. *Leukemia* 17(1): 133–137.

Magoč T and Salzberg SL (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21): 2957–2963.

Mahoney E, Lucas DM, Gupta S V, Wagner AJ, Herman SEM, Smith LL, Yeh Y-Y, Andritsos L, Jones J a, Flynn JM, Blum K a, Zhang X, Lehman A, Kong H, Gurcan M, Grever MR, Johnson AJ and Byrd JC (2012) ER stress and autophagy: new discoveries in the mechanism of action and drug resistance of the cyclin-dependent kinase inhibitor flavopiridol. *Blood* 120(6): 1262–73.

Mahoney E, Maddocks K, Flynn J, Jones J, Cole SL, Zhang X, Byrd JC and Johnson AJ (2013) Identification of endoplasmic reticulum stress-inducing agents by antagonizing autophagy: a new potential strategy for identification of anti-cancer therapeutics in B-cell malignancies. *Leukemia & lymphoma* 54(12): 2685–92.

Maiuri MC, Zalckvar E, Kimchi A and Kroemer G (2007) Self-eating and self-killing: crosstalk between autophagy and apoptosis. *Nature Reviews in Molecular Cell Biology* 8(9): 741–752.

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24(3): 133–141.

Marino G, Niso-Santano M, Baehrecke EH and Kroemer G (2014) Self-consumption: the interplay of autophagy and apoptosis. *Nature Reviews in Molecular Cell Biology* 15(2): 81–94.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1): 10–12.

Mathew R, Karp CM, Beaudoin B, Vuong N, Chen G, Chen HY, Bray K, Reddy A, Bhanot G, Gelinas C, DiPaola RS, Karantza-Wadsworth V and White E (2009) Autophagy Suppresses Tumorigenesis through Elimination of p62. *Cell* 137(6): 1062–1075.

Matrai Z, Lin K, Dennis M, Sherrington P, Zuzel M, Pettitt AR and Cawley JC (2001) CD38 expression and Ig V H gene mutation in B-cell chronic lymphocytic leukemia. *Blood* 97(6): 1901–1902.

Maura F, Cutrona G, Mosca L, Matis S, Lionetti M, Fabris S, Agnelli L, Colombo M, Massucco C, Ferracin M, Zagatti B, Reverberi D, Gentile M, Recchia AG, Bossio S, Rossi D, Gaidano G, Molica S, Cortelezzi A, Di Raimondo F, Negrini M, Tassone P, Morabito F, Ferrarini M and Neri A (2015) Association between gene and miRNA expression profiles and stereotyped subset #4 B-cell receptor in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 56(11): 3150–3158.

Mccurdy RD, Mcgrath JJ and Mackay-sim A (2008) Validation of the comparative quantification method of real-time PCR analysis and a cautionary tale of housekeeping gene selection. *Gene Therapy and Molecular Biology* 12: 15–24.

McEwan DG, Popovic D, Gubas A, Terawaki S, Suzuki H, Stadel D, Coxon FP, MirandadeStegmann D, Bhogaraju S, Maddi K, Kirchof A, Gatti E, Helfrich MH, Wakatsuki S, Behrends C, Pierre P and Dikic I (2015) PLEKHM1 regulates autophagosome-lysosome fusion through HOPS complex and LC3/GABARAP proteins. *Molecular Cell* 57(1): 39–54.

McInerney P, Adams P and Hadi MZ (2014) Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Molecular Biology International* 2014: 287430.

McLeod IX and He Y (2010) Roles of autophagy in lymphocytes: reflections and directions. *Cellular & Molecular Immunology* 7(2): 104–7.

Melchers F (2005) The pre-B-cell receptor: selector of fitting immunoglobulin heavy chains for the B-cell repertoire. *Nature Reviews. Immunology* 5(7): 578–84.

Messmer BT, Albesiano E, Efremov DG, Ghiotto F, Allen SL, Kolitz J, Foa R, Damle RN, Fais F, Messmer D, Rai KR, Ferrarini M and Chiorazzi N (2004) Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *Journal of Experimental Medicine* 200(4): 519–525.

Miller BC, Zhao Z, Stephenson LM, Cadwell K, Pua HH, Lee HK, Iwasaki A, He Y, Swat W and Iv HWV (2008) The autophagy gene ATG5 plays an essential role in B lymphocyte development. *Autophagy* 4: 309–314.

Mimmi S, Vecchio E, Iaccino E, Rossi M, Lupia A, Albano F, Chiurazzi F, Fiume G, Pisano A, Ceglia S, Pontoriero M, Golino G, Tassone P, Quinto I, Scala G and Palmieri C (2016) Evidence of shared epitopic reactivity among independent B-cell clones in Chronic Lymphocytic Leukemia patients. *Leukemia* 30(12):2419-2422

Mizushima N, Kuma A, Kobayashi Y, Yamamoto A, Matsubae M, Takai T, Natsume T, Ohsumi Y and Yoshimori T (2003) Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. *Journal of Cell Science* 116(9): 1679–1688.

Mizushima N, Noda T, Yoshimori T, Tanaka Y, Ishii T, George MD, Klionsky DJ, Ohsumi M and Ohsumi Y (1998) A protein conjugation system essential for autophagy. *Nature* 395: 395–398.

Mizushima N, Yoshimori T and Levine B (2010) Methods in mammalian autophagy research. *Cell* 140(3): 313–26.

Molica S, Shanafelt TD, Giannarelli D, Gentile M, Mirabelli R, Cutrona G, Levato L, Di Renzo N, Di Raimondo F, Musolino C, Angrilli F, Famà A, Recchia AG, Chaffee KG, Neri A, Kay NE, Ferrarini M and Morabito F (2016) The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: Independent validation in a prospective cohort of early stage patients. *American Journal of Hematology* 91(11):1090-1095

Morabito F, Mosca L, Cutrona G, Agnelli L, Tuana G, Ferracin M, Zagatti B, Lionetti M, Fabris S, Maura F, Matis S, Gentile M, Vigna E, Colombo M, Massucco C, Recchia AG, Bossio S, De Stefano L, Ilariucci F, Musolino C, Molica S, Di Raimondo F, Cortelezzi A, Tassone P, Negrini M, Monti S, Rossi D, Gaidano G, Ferrarini M and Neri A (2013) Clinical monoclonal B lymphocytosis versus rai 0 chronic lymphocytic leukemia: A comparison of cellular, cytogenetic, molecular, and clinical features. *Clinical Cancer Research* 19(21): 5890–5900.

Mortensen M, Soilleux EJ, Djordjevic G, Tripp R, Lutteropp M, Sadighi-Akha E, Stranks AJ, Glanville J, Knight S, Jacobsen S-EW, Kranc KR and Simon AK (2011) The autophagy protein Atg7 is essential for hematopoietic stem cell maintenance. *The Journal of Experimental Medicine* 208(3): 455–67.

Mraz M, Stano Kozubik K, Plevova K, Musilova K, Tichy B, Borsky M, Kuglik P, Doubek M, Brychtova Y, Mayer J and Pospisilova S (2013) The origin of deletion 22q11 in chronic lymphocytic leukemia is related to the rearrangement of immunoglobulin lambda light chain locus. *Leukemia Research* 37(7): 802–808.

Murphy K and Weaver C (2016) *Janeway's Immunobiology*. Taylor & Francis Group.

Murray F, Darzentas N, Hadzidimitriou A, Tobin G, Boudjogra M, Scielzo C, Laoutaris N, Karlsson K, Baran-Marzsak F, Tsaftaris A, Moreno C, Anagnostopoulos A, Caligaris-cappio F, Vaur D, Ouzounis C, Belessi C, Ghia P, Davi F, Rosenquist R, Stamatopoulos K, Dc W, Scielzo C and Moreno C (2008) Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia : implications for the role of antigen selection in leukemogenesis. *Blood* 111(3): 1524–1533.

Nadeu F, Delgado J, Royo C, Baumann T, Stankovic T, Pinyol M, Jares P, Navarro A, Martin-Garcia D, Bea S, Salaverria I, Oldreive C, Aymerich M, Suarez-Cisneros H,

Rozman M, Villamor N, Colomer D, Lopez-Guillermo A, Gonzalez M, Alcoceba M, Terol MJ, Colado E, Puente XS, Lopez-Otin C, Enjuanes A and Camp (2016) Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood* 127(17): 2122–2131.

Nadler LM, Anderson KC, Marti G, Bates M, Park E and Schlossman SF (1983) B4 a human B lymphocyte-associated antigen expressed on normal, mitogen-activated and malignant B lymphocytes. *Journal of Immunology* 131(1):244-50.

Navrkalova V, Young E, Baliakas P, Radova L, Sutton L-A, Plevova K, Mansouri L, Ljungstrom V, Ntoufa S, Davies Z, Juliusson G, Smedby KE, Belessi C, Panagiotidis P, Touloumenidou T, Davi F, Langerak AW, Ghia P, Strefford JC, Oscier DG, Mayer J, Stamatopoulos K, Pospisilova S, Rosenquist R and Trbusek M (2016) ATM mutations in major stereotyped subsets of chronic lymphocytic leukemia: enrichment in subset #2 is associated with markedly short telomeres. *Haematologica*. 101(9):e369-73

Nazio F, Strappazzon F, Antonioli M, Bielli P, Cianfanelli V, Bordi M, Gretzmeier C, Dengjel J, Piacentini M, Fimia GM and Cecconi F (2013) mTOR inhibits autophagy by controlling ULK1 ubiquitylation, self-association and function through AMBRA1 and TRAF6. *Nature Cell Biology* 15(4): 406–16.

Nelson JR, Lawrence CW and Hinkle DC (1996) Deoxycytidyl transferase activity of yeast REV1 protein. *Nature*, 729–731.

Nishida Y, Arakawa S, Fujitani K, Yamaguchi H, Mizuta T, Kanaseki T, Komatsu M, Otsu K, Tsujimoto Y and Shimizu S (2009) Discovery of Atg5/Atg7-independent alternative macroautophagy. *Nature* 461(7264): 654–8.

Nollet F, Cauwelier B, Billiet J, Selleslag D, Van Hoof A, Louwagie A and Criel A (2002) Do B-cell chronic lymphocytic leukemia patients with Ig VH3-21 genes constitute a new subset of chronic lymphocytic leukemia? *Blood* 100(3): 1097–1099.

Nowell P (1976) The clonal evolution of tumor cell populations. *Science* 194: 23–28.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H and Kanehisa M (1999) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 27(1): 29–34.

Ojha J, Secreto C, Rabe K, Ayres-Silva J, Tschumper R, Dyke D V, Slager S, Fonseca R, Shanafelt T, Kay N and Braggio E (2014) Monoclonal B-cell lymphocytosis is

characterized by mutations in CLL putative driver genes and clonal heterogeneity many years before disease progression. *Leukemia* 28(12): 2395–8.

Ojha R, Singh SK, Bhattacharyya S, Dhanda RS, Rakha A, Mandal AK and Jha V (2014) Inhibition of grade dependent autophagy in urothelial carcinoma increases cell death under nutritional limiting condition and potentiates the cytotoxicity of chemotherapeutic agent. *Journal of Urology* 191(6): 1889–1898. A

Olivier M, Hollstein M and Hainaut P (2010) TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harbor Perspectives in Biology* 2(1): a001008.

Oppezzo P, Vasconcelos Y, Settegrana C, Jeannel D, Vuillier F, Legarff-tavernier M, Kimura EY, Bechet S, Dumas G, Brissard M, Merle-béral H, Yamamoto M, Dighiero G and Davi F (2008) The LPL / ADAM29 expression ratio is a novel prognosis indicator in chronic lymphocytic leukemia The LPL / ADAM29 expression ratio is a novel prognosis indicator in chronic lymphocytic leukemia. *Blood* 106(2): 650–657.

Oscier DG, Dearden C, Erem E, Fegan C, Follows G, Hillmen P, Illidge T, Matutes E, Milligan DW, Pettitt AR, Schuh A and Wimperis J (2012) Guidelines on the diagnosis, investigation and management of chronic lymphocytic leukaemia. *British Journal of Haematology* 159(5): 541–564.

Oscier DG, Wade R, Davis Z, Morilla A, Best G, Richards S, Else M, Matutes E and Catovsky D (2010) Prognostic factors identified three risk groups in the LRF CLL4 trial, independent of treatment allocation. *Haematologica* 95(10): 1705–1712.

Ouillette P, Collins R, Shakhan S, Li J, Li C, Shedden K and Malek SN (2011) The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia. *Clinical Cancer Research* 17(21): 6778–6790.

Ouillette P, Collins R, Shakhan S, Li J, Peres E, Kujawski L, Kaminski M, Li C, Shedden K, Malek SN, Dc W and Talpaz M (2011) Acquired genomic copy number aberrations and survival in chronic lymphocytic leukemia. *Blood* 118(11): 3051–3061.

Ouillette P, Saiya-Cork K, Seymour E, Li C, Shedden K and Malek SN (2013) Clonal evolution, genomic drivers, and effects of therapy in chronic lymphocytic leukemia. *Clinical Cancer Research* 19(11): 2893–2904.

Pankiv S, Alemu EA, Brech A, Bruun JA, Lamark T, Overvatn A, Bjorkoy G and Johansen T (2010) FYCO1 is a Rab7 effector that binds to LC3 and PI3P to mediate

microtubule plus end - Directed vesicle transport. *Journal of Cell Biology* 188(2): 253–269.

Pankiv S, Clausen TH, Lamark T, Brech A, Bruun JA, Outzen H, Overvatn A, Bjorkoy G and Johansen T (2007) p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *Journal of Biological Chemistry* 282(33): 24131–24145.

Parikh SA, Rabe KG, Call TG, Zent CS, Habermann TM, Ding W, Leis JF, Schwager SM, Hanson CA, Macon WR, Kay NE, Slager SL and Shanafelt TD (2013) Diffuse large B-cell lymphoma (Richter syndrome) in patients with chronic lymphocytic leukaemia (CLL): A cohort study of newly diagnosed patients. *British Journal of Haematology* 162(6): 774–782.

Parikh SA, Rabe KG, Kay NE, Call TG, Ding W, Schwager SM, Bowen DA, Conte M, Jelinek DF, Slager SL and Shanafelt TD (2014) Chronic lymphocytic leukemia in young (< 55 years) patients: A comprehensive analysis of prognostic factors and outcomes. *Haematologica* 99(1): 140–147.

Parker H, Rose-Zerilli M, Parker a, Chaplin T, Wade R, Gardiner a, Griffiths M, Collins a, Young BD, Oscier DG and Strefford JC (2011) 13Q Deletion Anatomy and Disease Progression in Patients With Chronic Lymphocytic Leukemia. *Leukemia* 25(3): 489–97.

Patterson KI, Brummer T, Daly RJ and O'Brien PM (2010) DUSP26 negatively affects the proliferation of epithelial cells, an effect not mediated by dephosphorylation of MAPKs. *Biochimica et Biophysica Acta - Molecular Cell Research* 1803(9): 1003–1012.

Pengo N, Scolari M, Oliva L, Milan E, Mainoldi F, Raimondi A, Fagioli C, Merlini A, Mariani E, Pasqualetto E, Orfanelli U, Ponzoni M, Sitia R, Casola S and Cenci S (2013) Plasma cells require autophagy for sustainable immunoglobulin production. *Nature Immunology* 14(3): 298–305.

Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P and Olivier M (2007) TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene* 26(15): 2157–2165.

Pfeifer D, Pantic M, Skatulla I, Rawluk J, Kreutz C, Martens UM, Fisch P, Timmer J and Veelken H (2006) Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. *Blood* 109(3): 1202–1210.

Phadwal K, Alegre-abarrategui J, Watson AS, Pike L, Anbalagan S, Hammond EM, Wade-martins R, Mcmichael A, Klenerman P and Simon AK (2012) A novel method for autophagy detection in primary cells: Impaired levels of macroautophagy in immunosenescent T cells. *Autophagy* 8(April): 677–689.

Plevova K, Francova HS, Burckova K, Brychtova Y, Doubek M, Pavlova S, Malcikova J, Mayer J, Tichy B and Pospisilova S (2014) Multiple productive immunoglobulin heavy chain gene rearrangements in chronic lymphocytic leukemia are mostly derived from independent clones. *Haematologica* 99(2): 329–38.

Polson HEJ, de Lartigue J, Rigden DJ, Reedijk M, Urbé S, Clague MJ and Tooze S a (2010) Mammalian Atg18 (WIPI2) localizes to omegasome-anchored phagophores and positively regulates LC3 lipidation. *Autophagy* 6(4): 506–22.

Pozzo F, Bittolo T, Arruga F, Bulian P, Macor P, Tissino E, Gizdic B, Rossi FM, Bomben R, Zucchetto A, Benedetti D, Degan M, D'Arena G, Chiarenza A, Zaja F, Pozzato G, Rossi D, Gaidano G, Del Poeta G, Deaglio S, Gattei V and Dal Bo M (2015) NOTCH1 mutations associate with low CD20 level in chronic lymphocytic leukemia: evidence for a NOTCH1 mutation-driven epigenetic dysregulation. *Leukemia* 30: 182–189.

Prakash S, Johnson R and Prakash L (2005) Eukayotic translesion synthesis DNA polymerases: Specificity of Structure and Function. *Annual Review of Biochemistry* 74(1): 317–353.

Proikas-Cezanne T, Waddell S, Gaugel A, Frickey T, Lupas A and Nordheim A (2004) WIPI-1alpha (WIPI49), a member of the novel 7-bladed WIPI protein family, is aberrantly expressed in human cancer and is linked to starvation-induced autophagy. *Oncogene* 23(58): 9314–25.

Provan D, Bartlett-pandite L, Zwicky C, Neuberg D, Maddocks A, Corradini P, Soiffer R, Ritz J, Nadler LM and Gribben JG (1996) Eradication of polymerase chain reaction-detectable chronic lymphocytic leukemia cells is associated with improved outcome after bone marrow transplantation. *Blood* 88(6): 2228–2235.

Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, Munar M, Rubio-Pérez C, Jares P, Aymerich M, Baumann T, Beekman R, Belver L, Carrió A, Castellano G, Clot G, Colado E, Colomer D, Costa D, Delgado J, Enjuanes A, Estivill X, Ferrando AA, Gelpí JL, González B, González S, González M, Gut M, Hernández-Rivas JM, López-Guerra M, Martín-García D, Navarro A, Nicolás P, Orozco M, Payer ÁR, Pinyol M, Pisano DG, Puente DA, Queirós AC, Quesada V,

Romeo-Casabona CM, Royo C, Royo R, Rozman M, Russiñol N, Salaverría I, Stamatopoulos K, Stunnenberg HG, Tamborero D, Terol MJ, Valencia A, López-Bigas N, Torrents D, Gut I, López-Guillermo A, López-Otín C and Campo E (2015) Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* 526(7574): 519–524.

Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, Escaramis G, Jares P, Beà S, González-Díaz M, Bassaganyas L, Baumann T, Juan M, López-Guerra M, Colomer D, Tubío JMC, López C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernández JM, Puente D a, Freije JMP, Velasco G, Gutiérrez-Fernández A, Costa D, Carrió A, Guijarro S, Enjuanes A, Hernández L, Yagüe J, Nicolás P, Romeo-Casabona CM, Himmelbauer H, Castillo E, Dohm JC, de Sanjosé S, Piris M a, de Alava E, San Miguel J, Royo R, Gelpí JL, Torrents D, Orozco M, Pisano DG, Valencia A, Guigó R, Bayés M, Heath S, Gut M, Klatt P, Marshall J, Raine K, Stebbings L a, Futreal PA, Stratton MR, Campbell PJ, Gut I, López-Guillermo A, Estivill X, Montserrat E, López-Otín C and Campo E (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 475(7354): 101–5.

Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Bea S, Pinyol M, Martinez-Trillos A, Lopez-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Gine E, Hernandez JM, Gonzalez-Diaz M, Puente DA, Velasco G, Freije JM, Tubio JM, Royo R, Gelpi JL, Orozco M, Pisano DG, Zamora J, Vazquez M, Valencia A, Himmelbauer H, Bayes M, Heath S, Gut M, Gut I, Estivill X, Lopez-Guillermo A, Puente XS, Campo E and Lopez-Otin C (2012) Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genetics* 44(1): 47–52.

Rada C, Di Noia JM and Neuberger MS (2004) Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Molecular Cell* 16(2): 163–171.

Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN and Pasternack BS (1975) Clinical staging of chronic lymphocytic leukemia. *Blood* 46(2): 219–234.

Ramsden D a, Baetz K and Wu GE (1994) Conservation of sequence in recombination signal sequence spacers. *Nucleic acids research* 22(10): 1785–96.

Rao S, Tortola L, Perlot T, Wirnsberger G, Novatchkova M, Nitsch R, Sykacek P, Frank L, Schramek D, Komnenovic V, Sigl V, Aumayr K, Schmauss G, Fellner N,

Handschuh S, Glösmann M, Pasierbek P, Schlederer M, Resch GP, Ma Y, Yang H, Popper H, Kenner L, Kroemer G and Penninger JM (2014) A dual role for autophagy in a murine model of lung cancer. *Nature communications* 5: 3056.

Rasi S, Khiabanian H, Ciardullo C, Terzi-di-Bergamo L, Monti S, Spina V, Bruscaggin A, Cerri M, Deambrogi C, Martuscelli L, Biasi A, Spaccarotella E, De Paoli L, Gattei V, Foa' R, Rabadan R, Gaidano G and Rossi D (2016) Clinical Impact of Small Subclones Harboring Notch1, Sf3B1 or Birc3 Mutations in Chronic Lymphocytic Leukemia. *Haematologica* 123(14): 2139–2148.

Rasi S, Monti S, Spina V, Foa R, Gaidano G and Rossi D (2012) Analysis of NOTCH1 mutations in monoclonal B-cell lymphocytosis. *Haematologica* 97(1): 153–154.

Rassenti LZ, Huynh L, Toy TL, Chen L, Keating MJ, Gribben JG, Neuberg DS, Flinn IW, Rai KR, Byrd JC, Kay NE, Greaves A, Weiss A and Kipps TJ (2004) ZAP-70 compared with immunoglobulin heavy-chain gene mutation status as a predictor of disease progression in chronic lymphocytic leukemia. *The New England Journal of Medicine* 351(9): 893–901.

Rassenti LZ and Kipps TJ (1997) Lack of allelic exclusion in B cell chronic lymphocytic leukemia. *The Journal of experimental medicine* 185(8): 1435–1445.

Ravikumar B, Moreau K, Jahreiss L, Puri C and Rubinsztein DC (2010) Plasma membrane contributes to the formation of pre-autophagosomal structures. *Nature Cell Biology* 12(8): 747–757.

Rawstron AC, Bennett F, O'Connor S, Kwok M, Fenton JAL, Plummer M, de Tute R, Owen RG, Richards SJ, Jack AS and Hillmen P (2008a) Monoclonal B-Cell Lymphocytosis and Chronic Lymphocytic Leukemia. *New England Journal of Medicine* 359(6): 575–583.

Rawstron AC, Bennett FL, O'Connor SJ, Kwok M, Fenton JAL, Plummer M, de Tute R, Owen RG, Richards SJ, Jack AS and Hillmen P (2008b) Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med* 359(6): 575–583.

Rice KL, Lin X, Wolniak K, Ebert BL, Berkofsky-Fessler W, Buzzai M, Sun Y, Xi C, Elkin P, Levine R, Golub T, Gilliland DG, Crispino JD, Licht JD and Zhang W (2011) Analysis of genomic aberrations and gene expression profiling identifies novel lesions and pathways in myeloproliferative neoplasms. *Blood Cancer Journal* 1(11): e40.

Rivlin N, Brosh R, Oren M and Rotter V (2011) Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes Cancer* 2(4): 466–474.

Roberts SA and Gordenin DA (2014) Hypermutation in human cancer genomes: footprints and mechanisms. *Nature Reviews in Cancer* 14(12): 786–800.

Rose-Zerilli M, Gibson J, Wang J, Tapper W, Davis Z, Parker H, Larrayoz M, McCarthy H, Walewska R, Forster J, Gardiner A, Steele AJ, Chelala C, Ennis S, Collins A, Oakes C, Oscier DG and Strefford JC (2016) Longitudinal copy number, whole exome and targeted deep sequencing of 'good risk' IGHV-mutated CLL patients with progressive disease. *Leukemia* 30: 1–10.

Rosenquist R (2009) How to report IG sequence data in clinical routine: difficult cases to categorize. In: Ghia P, Rosenquist R and Davi F (eds) *Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia*. Milan: Wolters Kluwer Health, 133–124.

Rosenwald A, Alizadeh AA, Widhopf G, Simon R, Davis RE, Yu X, Yang L, Pickeral OK, Rassenti LZ, Powell J, Botstein D, Byrd JC, Grever MR, Cheson BD, Chiorazzi N, Wilson WH, Kipps TJ, Brown PO and Staudt LM (2001) Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *Journal of Experimental Medicine* 194(11): 1639–1647.

Rossi D, Fangazio M, Rasi S, Vaisitti T, Monti S, Cresta S, Chiaretti S, Del Giudice I, Fabbri G, Bruscaggin A, Spina V, Deambrogi C, Marinelli M, Fam?? R, Greco M, Daniele G, Forconi F, Gattei V, Bertoni F, Deaglio S, Pasqualucci L, Guarini A, Dalla-Favera R, Fo?? R and Gaidano G (2012) Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. *Blood* 119(12): 2854–2862.

Rossi D and Gaidano G (2010) Biological and clinical significance of stereotyped B-cell receptors in chronic lymphocytic leukemia. *Haematologica* 95(12): 1992–1995.

Rossi D, Khiabanian H, Spina V, Ciardullo C, Bruscaggin A, Famà R, Monti S, Deambrogi C, Paoli L De, Wang J, Gattei V, Guarini A, Foà R, Rabadan R and Gaidano G (2014) Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood* 123(14): 2139–2148.

Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F, Marasca R, Laurenti L, Bruscaggin A, Cerri M, Monti S, Cresta S, Fama R, De Paoli L, Bulian P, Gattei V,

Guarini A, Deaglio S, Capello D, Rabadan R, Pasqualucci L, Dalla-Favera R, Foa R and Gaidano G (2012) Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood* 119(2): 521–529.

Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, Deambrogi C, Khiabanian H, Serra R, Bertoni F, Forconi F, Laurenti L, Marasca R, Dal-bo M, Rossi FM, Bulian P, Nomdedeu J, Poeta G Del, Gattei V, Pasqualucci L, Rabadan R and Foa R (2013) Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* 121(8): 1403–1412.

Rossi D, Spina V, Bomben R, Rasi S, Dal-Bo M, Bruscaggin A, Rossi FM, Monti S, Degan M, Ciardullo C, Serra R, Zucchetto A, Nomdedeu J, Bulian P, Grossi A, Zaja F, Pozzato G, Laurenti L, Efremov DG, Di-Raimondo F, Marasca R, Forconi F, Del Poeta G, Gaidano G and Gattei V (2013) Association between molecular lesions and specific B-cell receptor subsets in chronic lymphocytic leukemia. *Blood* 121(24): 4902–4905.

Rossi D, Spina V, Cerri M, Rasi S, Deambrogi C, De Paoli L, Laurenti L, Maffei R, Forconi F, Bertoni F, Zucca E, Agostinelli C, Cabras A, Lucioni M, Martini M, Magni M, Deaglio S, Ladetto M, Nomdedeu JF, Besson C, Ramponi A, Canzonieri V, Paulli M, Marasca R, Larocca LM, Carbone A, Pileri S, Gattei V and Gaidano G (2009) Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to richter syndrome. *Clinical Cancer Research* 15(13): 4415–4422.

Rossi D, Spina V, Deambrogi C, Rasi S, Laurenti L, Stamatopoulos K, Arcaini L, Lucioni M, Rocque GB, Xu-Monette ZY, Visco C, Chang J, Chigrinova E, Forconi F, Marasca R, Besson C, Papadaki T, Paulli M, Larocca LM, Pileri SA, Gattei V, Bertoni F, Foa R, Young KH and Gaidano G (2011) The genetics of Richter syndrome reveals disease heterogeneity and predicts survival after transformation. *Blood* 117(12): 3391–3401.

Rossi D, Spina V, Forconi F, Capello D, Fangazio M, Rasi S, Martini M, Gattei V, Ramponi A, Larocca LM, Bertoni F and Gaidano G (2012) Molecular history of Richter syndrome: Origin from a cell already present at the time of chronic lymphocytic leukemia diagnosis. *International Journal of Cancer* 130(12): 3006–3010.

Roth DB (2014) V(D)J Recombination : Mechanism , Errors , and Fidelity. *Microbiology Spectrum* 2(6): 1–11.

RStudio Team (2015) RStudio: Integrated Development Environment for R. Boston, MA. Available at: http://www.rstudio.com/.

Sakoh-Nakatogawa M, Matoba K, Asai E, Kirisako H, Ishii J, Noda NN, Inagaki F, Nakatogawa H and Ohsumi Y (2013) Atg12-Atg5 conjugate enhances E2 activity of Atg3 by rearranging its catalytic site. *Nature Structural & Molecular Biology* 20(4): 433–9.

Sánchez ML, Almeida J, Vidriales B, López-Berges MC, García-Marcos M a, Moro MJ, Corrales a, Calmuntia MJ, San Miguel JF and Orfao A (2002) Incidence of phenotypic aberrations in a series of 467 patients with B chronic lymphoproliferative disorders: basis for the design of specific four-color stainings to be used for minimal residual disease investigation. *Leukemia* 16: 1460–1469.

Schatz DG and Swanson PC (2011) V(D)J Recombination: Mechanisms of Initiation. *Annual Review of Genetics* 45(1): 167–202.

Schmid D, Pypaert M and Münz C (2007) Antigen-Loading Compartments for Major Histocompatibility Complex Class II Molecules Continuously Receive Input from Autophagosomes. *Immunity* 26(1): 79–92.

Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, Feller SM, Grocock R, Henderson S, Khrebtukova I, Kingsbury Z, Luo S, McBride D, Murray L, Menju T, Timbs A, Ross M, Taylor J and Bentley D (2012) Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* 120(20): 4191–4196.

Seifert M, Sellmann L, Bloehdorn J, Wein F, Stilgenbauer S, Dürig J and Küppers R (2012) Cellular origin and pathophysiology of chronic lymphocytic leukemia. *The Journal of Experimental Medicine* 209(12): 2183–98.

Shapiro-Shelef M, Lin KI, McHeyzer-Williams LJ, Liao J, McHeyzer-Williams MG and Calame K (2003) Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* 19(4): 607–620.

Shi CS and Kehrl JH (2008) MyD88 and Trif target Beclin 1 to trigger autophagy in macrophages. *Journal of Biological Chemistry* 283(48): 33175–33182.

Shimizu S, Kanaseki T, Mizushima N, Mizuta T, Arakawa-Kobayashi S, Thompson CB and Tsujimoto Y (2004) Role of Bcl-2 family proteins in a non-apoptotic programmed cell death dependent on autophagy genes. *Nature Cell Biology* 6(12): 1221–1228.

Shin C and Manley JL (2004) Cell signalling and the control of pre-mRNA splicing. *Nature Reviews Molecular Cell Biology* 5(9): 727–738.

Skowronska A, Parker A, Ahmed G, Oldreive C, Davis Z, Richards S, Dyer M, Matutes E, Gonzalez D, Taylor AMR, Moss P, Thomas P, Oscier DG and Stankovic T (2012) Biallelic ATM inactivation significantly reduces survival in patients treated on the United Kingdom leukemia research fund chronic lymphocytic leukemia 4 trial. *Journal of Clinical Oncology* 30(36): 4524–4532.

Smith L, Hogg E, Haynes A, Strefford JC, Forconi F, Stevenson FK, Packham G, Cragg MS, Tumbarello D, Blunt MD and Steele AJ (2015) Biological Significance of B Cell Receptor Mediated Regulation of Autophagy in Chronic Lymphocytic Leukemia [Abstract]. *Blood* 126(23): 4130 LP-4130.

Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, Marjoram P, Siegmund K, Press MF, Shibata D and Curtis C (2015) A Big Bang model of human colorectal tumor growth. *Nature Genetics* 47(3): 209–216.

Stamatopoulos K, Belessi C, Hadzidimitriou A, Smilevska T, Kalagiakou E, Hatzi K, Stavroyianni N, Athanasiadou A, Tsompanakou A, Papadaki T, Kokkini G, Paterakis G, Saloum R, Laoutaris N, Anagnostopoulos A and Fassas A (2005) Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood* 106(10): 3575–3583.

Stamatopoulos K, Belessi C, Moreno C, Boudjogra M, Guida G, Smilevska T, Belhoul L, Stella S, Stavroyianni N, Crespo M, Hadzidimitriou A, Sutton L, Bosch F, Laoutaris N, Anagnostopoulos A, Montserrat E, Fassas A, Dighiero G, Caligaris-Cappio F, Merle-Beral H, Ghia P and Davi F (2007) Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood* 109(1): 259–270.

Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP and Ebert BL (2015) Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* 126(1): 9–16.

Steinel NC, Fisher MR, Yang-Iott KS and Bassing CH (2014) The Ataxia Telangiectasia Mutated and Cyclin D3 Proteins Cooperate To Help Enforce TCRβ and IgH Allelic Exclusion. *The Journal of Immunology* 193(6): 2881–2890.

Steinel NC, Lee B-S, Tubbs AT, Bednarski JJ, Schulte E, Yang-Iott KS, Schatz DG, Sleckman BP and Bassing CH (2013) The ataxia telangiectasia mutated kinase controls Igκ allelic exclusion by inhibiting secondary Vκ-to-Jκ rearrangements. *The Journal of Experimental Medicine* 210(2): 233–9.

Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal S, Leroy C, Jia M, Menzies A, Butler AP, Teague JW, Quail MA, Burton J, Swerdlow H, Carter NP, Morsberger LA, Iacobuzio-Donahue C, Follows GA, Green AR, Flanagan AM, Stratton MR, Futreal PA and Campbell PJ (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144(1): 27–40.

Stilgenbauer S, Bullinger L, Benner A, Wildenberger K, Bentz M, Dohner K, Ho AD, Lichter P and Dohner H (1999) Incidence and clinical significance of 6q deletions in B cell chronic lymphocytic leukemia. *Leukemia* 13: 1331–1334.

Stilgenbauer S, Sander S, Bullinger L, Benner A, Leupolt E, Winkler D, Krober A, Kienle D, Lichter P and Dohner H (2007) Clonal evolution in chronic lymphocytic leukemia: acquisition of high-risk genomic aberrations associated with unmutated VH, resistance to therapy, and short survival. *Haematologica* 92(9): 1242–1245.

Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K, Bühler A, Böttcher S, Ritgen M, Kneba M, Winkler D, Tausch E, Hoth P, Edelmann J, Mertens D, Bullinger L, Bergmann MA, Kless S, Mack S, Jager U, Patten N, Wu L, Wenger M, Fingerle-Rowson G, Lichter P, Cazzola M, Wendtner C-M, Fink A, Fischer K, Busch R, Hallek M and Dohner H (2014) Gene Mutations and Treatment Outcome in Chronic Lymphocytic Leukemia: results from the CLL8 trial. *Blood* 123(21): 3247–3255.

Strati P and Shanafelt TD (2015) Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. *Blood* 126(4): 454–62. Available at:

Strefford JC, Sutton L-A, Baliakas P, Agathangelidis A, Malčíková J, Plevova K, Scarfò L, Davis Z, Stalika E, Cortese D, Cahill N, Pedersen LB, di Celle PF, Tzenou T, Geisler C, Panagiotidis P, Langerak a W, Chiorazzi N, Pospisilova S, Oscier DG, Davi F, Belessi C, Mansouri L, Ghia P, Stamatopoulos K and Rosenquist R (2013) Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. *Leukemia* 27(April): 2196–9.

Sun Q, Fan W, Chen K, Ding X, Chen S and Zhong Q (2008) Identification of Barkor as a mammalian autophagy-specific factor for Beclin 1 and class III

phosphatidylinositol 3-kinase. *Proceedings of the National Academy of Sciences of the United States of America* 105(49): 19211–6.

Sutton L-A and Rosenquist R (2014) Deciphering the molecular landscape in chronic lymphocytic leukemia: Time frame of disease evolution. *Haematologica* 100(1): 7–16.

Szankasi P and Bahler DW (2010) Clinical laboratory analysis of immunoglobulin heavy chain variable region genes for chronic lymphocytic leukemia prognosis. *The Journal of Molecular Diagnostics* 12(2): 244–9.

Takeda S, Sonoda E and Arakawa H (1996) The K:L ratio of immature B cells. *Immunology Today* 17: 200–201.

Tanida I, Minematsu-Ikeguchi N, Ueno T and Kominami E (2005) Lysosomal turnover, but not a cellular level, of endogenous LC3 is a marker for autophagy. *Autophagy* 1(2): 84–91.

Teng G and Schatz DG (2015) Regulation and Evolution of the RAG Recombinase. *Advances in Immunology* 128:1-39

Ternynck T, Dighiero G, Follezou J and Binet J-L (1974) Comparison of Normal and CLL Lymphocyte Surface Ig Determinants Using Peroxidase-labeled Antibodies .I. Detection and Quantitation of Light Chain Determinants. *Blood* 43(6):789-95.

The International CLL-IPI Working Group (2016) An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *The Lancet Oncology* 2045(16): 1–12.

Thorselius M, Krober A, Murray F, Thunberg U, Tobin G, Bu A, Kienle D, Albesiano E, Maffei R, Wiley J, Vilpo J, Laurell A, Merup M, Karlsson K, Chiorazzi N, Marasca R, Do H, Stilgenbauer S, Rosenquist R, Thorselius M, Krober A, Buhler A, Dao-Ung LP, Roos G and Dohner H (2006) Strikingly homologous immunoglobulin gene rearrangements and poor outcome in V H 3-21 – using chronic lymphocytic leukemia patients independent of geographic origin and mutational status. *Blood* 107(7): 2889–2894.

Tobin G, Thunberg U, Johnson A, Eriksson I, Soderberg O, Karlsson K, Merup M, Juliusson G, Vilpo J, Enblad G, Sundstrom C, Roos G and Rosenquist R (2003) Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. *Blood* 101(12): 4952–4957.

Tobin G, Thunberg U, Johnson A, Thörn I, Söderberg O, Hultdin M, Botling J, Enblad G, Sällström J, Sundström C, Roos G, Rosenquist R and Tho I (2002) Somatically mutated Ig VH3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood* 99(6): 2262–2264.

Tobin G, Thunberg U, Laurell A, Karlsson K, Aleskog A, Willander K, Soderberg O, Merup M, Vilpo J, Hultdin M, Sundstrom C, Roos G and Rosenquist R (2005) Patients with chronic lymphocytic leukemia with mutated VH genes presenting with Binet stage B or C form a subgroup with a poor outcome. *Haematologica* 90(4): 465–469.

Tooze S a and Yoshimori T (2010) The origin of the autophagosomal membrane. *Nature cell biology* 12(9): 831–835.

Tsiantoulas D, Gruber S and Binder CJ (2012) B-1 cell immunoglobulin directed against oxidation-specific epitopes. *Frontiers in Immunology* 3: 1–6.

Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR and Murphy KM (2010) Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *The Journal of Molecular Diagnostics* 12(4): 425–32.

Uemura T, Yamamoto M, Kametaka A, Sou Y-S, Yabashi A, Yamada A, Annoh H, Kametaka S, Komatsu M and Waguri S (2014) A cluster of thin tubular structures mediates transformation of the endoplasmic reticulum to autophagic isolation membrane. *Molecular and cellular biology* 34(9): 1695–706.

Velikkakath AKG, Nishimura T, Oita E, Ishihara N and Mizushima N (2012) Mammalian Atg2 proteins are essential for autophagosome formation and important for regulation of size and distribution of lipid droplets. *Molecular Biology of the Cell* 23(5): 896–909.

Villamor N, Conde L, Martínez-Trillos a, Cazzola M, Navarro a, Beà S, López C, Colomer D, Pinyol M, Aymerich M, Rozman M, Abrisqueta P, Baumann T, Delgado J, Giné E, González-Díaz M, Hernández JM, Colado E, Payer a R, Rayon C, Navarro B, José Terol M, Bosch F, Quesada V, Puente XS, López-Otín C, Jares P, Pereira a, Campo E and López-Guillermo a (2013) NOTCH1 mutations identify a genetic subgroup of chronic lymphocytic leukemia patients with high risk of transformation and poor outcome. *Leukemia* 27: 1100–6.

Visco C, Moretta F, Falisi E, Facco M, Maura F, Novella E, Nichele I, Finotto S, Giaretta I, Ave E, Perbellini O, Guercini N, Scupoli MT, Trentin L, Trimarco V, Neri

A, Semenzato G, Rodeghiero F, Pizzolo G and Ambrosetti A (2013) Double productive immunoglobulin sequence rearrangements in patients with chronic lymphocytic leukemia. *American Journal of Hematology* 88(4): 277–282.

Vogel C and Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* 13(4): 227–232.

Volinia S, Dhand R, Vanhaesebroeck B, MacDougall LK, Stein R, Zvelebil MJ, Domin J, Panaretou C and Waterfield MD (1995) A human phosphatidylinositol 3-kinase complex related to the yeast Vps34p-Vps15p protein sorting system. *EMBO Journal* 14(14): 3339–3348.

Vuillier F, Scott-Algara D and Dighiero G (1991) Extensive analysis of lymphocyte subsets in normal subjects by three-color immunofluorescence. *Nouvelle Revue Française d'Hématologie* 33(1): 31–8.

Wabl M and Steinberg C (1992) Allelic exclusion model questioned. *Nature* 359(6394): 370–371.

Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M and Shafer RW (2007) Characterization of mutation spectra with ultra-deep pyrosequencing: Application to HIV-1 drug resistance. *Genome Research* 17(8): 1195–1201.

Wang K, Wei G and Liu D (2012) CD19: a biomarker for B cell development, lymphoma diagnosis and therapy. *Experimental Hematology & Oncology* 1(1): 36.

Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, Deluca DS, Zhang L, Zhang W, Vartanov AR, Fernandes SM, Goldstein NR, Folco EG, Cibulskis K, Tesar B, Sievers QL, Shefler E, Gabriel S, Hacohen N, Reed R, Meyerson M, Golub TR, Lander ES, Neuberg D, Brown JR, Getz G and Wu CJ (2011) SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. *New England Journal of Medicine* 365(26): 2497–2506.

Wang S, Li X, Wang Q and Xiu Z (2015) Autophagy inhibitor sensitizes MCF-7 breast cancer cells to novel cyclic tetrapeptide CTS203-induced caspase-9-dependent apoptotic cell death. *Neoplasma* 62(2): 220–229.

Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, Multani A, Zhang H, Zhao R, Michor F, Meric-Bernstam F and Navin NE (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512(7513): 1–15.

Wartosch L, Gunesdogan U, Graham SC and Luzio JP (2015) Recruitment of VPS33A to HOPS by VPS16 Is Required for Lysosome Fusion with Endosomes and Autophagosomes. *Traffic* 727–742.

Watanabe K, Ichinose S, Hayashizaki K and Tsubata T (2008) Induction of autophagy by B cell antigen receptor stimulation and its inhibition by costimulation. *Biochemical and Biophysical Research Communications* 374(2): 274–281.

Watanabe K and Tsubata T (2009) Autophagy connects antigen receptor signaling to costimulatory signaling in B lymphocytes. *Autophagy* 5(1):108-10.

Waterhouse AM, Procter JB, Martin DMA, Clamp M and Barton GJ (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9): 1189–1191.

Watson AS, Mortensen M and Simon AK (2011) Autophagy in the pathogenesis of myelodysplastic syndrome and acute myeloid leukemia. *Cell Cycle* 10(11): 1719–1725.

Wei H and Guan JL (2012) Pro-tumorigenic function of autophagy in mammary oncogenesis. *Autophagy* 8(1): 129–131.

Weidberg H, Shvets E, Shpilka T, Shimron F, Shinder V and Elazar Z (2010) LC3 and GATE-16/GABARAP subfamilies are both essential yet act differently in autophagosome biogenesis. *EMBO Journal* 29(11): 1792–1802.

Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Available at: http://ggplot2.org.

Widhopf GF, Rassenti LZ, Toy TL, Gribben JG, Wierda WG, Kipps TJ and Dc W (2004) Chronic lymphocytic leukemia B cells of more than 1 % of patients express virtually identical immunoglobulins. *Blood* 104(8): 2499–2504.

Wiestner A, Rosenwald A, Barry TS, Wright G, Davis RE, Henrickson SE, Zhao H, Ibbotson RE, Orchard J a, Davis Z, Stetler-Stevenson M, Raffeld M, Arthur DC, Marti GE, Wilson WH, Hamblin TJ, Oscier DG and Staudt LM (2003) ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. *Blood* 101(12): 4944–4951.

Wilson PC, de Bouteiller O, Liu YJ, Potter K, Banchereau J, Capra JD and Pascual V (1998) Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. *Journal of Experimental Medicine* 187(1): 59–70.

Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, McMichael JF, Schmidt H, Yellapantula V, Miller CA, Ozenberger BA, Welch JS, Link DC, Walter MJ, Mardis ER, Dipersio JF, Chen F, Wilson RK, Ley TJ and Ding L (2014) Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature medicine* 20(12): 1472–8.

Xu JL and Davis MM (2000) Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 13(1): 37–45.

Yang S, Wang X, Contino G, Liesa M, Sahin E, Ying H, Bause A, Li Y, Stomme JM, Dell'Antonio G, Mautner J, Tonon G, Haigis M, Shirihai OS, Doglioni C, Bardeesy N and Kimmelman AC (2011) Pancreatic cancers require autophagy for tumor growth. *Genes and Development* 25(7): 717–729.

Yang SM, Li JY, Gale RP and Huang XJ (2015) The mystery of chronic lymphocytic leukemia (CLL): Why is it absent in Asians and what does this tell us about etiology, pathogenesis and biology? *Blood Reviews* 29(3): 205–213.

Ylä-Anttila P, Vihinen H, Jokitalo E and Eskelinen EL (2009) 3D tomography reveals connections between the phagophore and endoplasmic reticulum. *Autophagy* 5(8): 1180–1185.

Young MM, Takahashi Y, Khan O, Park S, Hori T, Yun J, Sharma AK, Amin S, Hu CD, Zhang J, Kester M and Wang HG (2012) Autophagosomal membrane serves as platform for intracellular death-inducing signaling complex (iDISC)-mediated caspase-8 activation and apoptosis. *Journal of Biological Chemistry* 287(15): 12455–12468.

Zenz T, Eichhorst B, Busch R, Denzel T, Häbe S, Winkler D, Bühler A, Edelmann J, Bergmann MA, Hopfinger G, Hensel M, Hallek M, Döhner H and Stilgenbauer S (2010) TP53 mutation and survival in chronic lymphocytic leukemia. *Journal of Clinical Oncology* 28(29): 4473–4479.

Zenz T, Habe S, Denzel T, Mohr J, Winkler D, Buhler A, Sarno A, Groner S, Mertens D, Busch R, Hallek M, Dohner H and Stilgenbauer S (2009) Detailed analysis of p53 pathway defects in fludarabine-refractory chronic lymphocytic leukemia (CLL): dissecting the contribution of 17p deletion, TP53 mutation, p53-p21 dysfunction, and miR34a in a prospective clinical trial. *Blood* 114(13): 2589–2597.

Zenz T, Krober A, Scherer K, Habe S, Buhler A, Benner A, Denzel T, Winkler D, Edelmann J, Schwanen C, Dohner H and Stilgenbauer S (2008) Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia:

results from a detailed genetic characterization with long-term follow-up. *Blood* 112(8): 3322–3329.

Zhang Q, Cao L-Y, Cheng S-J, Zhang A-M, Jin X-S and Li Y (2015) p53-induced microRNA-1246 inhibits the cell growth of human hepatocellular carcinoma cells by targeting NFIB. *Oncology Reports* 33(12): 1335–1341.

Zhang X, Reis M, Khoriaty R, Li Y, Ouillette P, Samayoa J, Carter H, Karchin R, Li M, Diaz L a, Velculescu VE, Papadopoulos N, Kinzler KW, Vogelstein B and Malek SN (2011) Sequence analysis of 515 kinase genes in chronic lymphocytic leukemia. *Leukemia* 25(12): 1908–10.

Zhao D, Yuan H, Yi F, Meng C and Zhu Q (2014) Autophagy prevents doxorubicin-induced apoptosis in osteosarcoma. *Molecular Medicine Reports* 9(5): 1975–1981.

Zhao Y, Yang J, Liao W, Liu X, Zhang H, Wang S, Wang D, Feng J, Yu L and Zhu WG (2010) Cytosolic FoxO1 is essential for the induction of autophagy and tumour suppressor activity. *Nat Cell Biol* 12(7): 665–675.

Zhou X, Lu X, Lv J, Yang H, Qin L, Zhao M, Su Y, Li Z and Zhang H (2011) Genetic association of PRDM1-ATG5 intergenic region and autophagy with systemic lupus erythematosus in a Chinese population. *Annals of the Rheumatic Diseases* 70(7): 1330–1337.

Appendices

APPENDIX 1 - TABLE OF PRIMERS 260

APPENDIX 2 – PATIENT CHARACTERISTICS 264

Population A 264

Population B 265

Population C 266

Population D 267

APPENDIX 3 – NGS CLL PANEL 268

Appendix 1 - Table of primers

Primer	Sequence (5'-3')
ATG5 Exon 1 (F)	CTGGGTTAGGCAGAACACG
ATG5 Exon 1 (R)	ACAAGGTGGACACACACACG
ATG5 Exon 2 (F)	TGCAAGGATCTGACTAATGCTC
ATG5 Exon 2 (R)	CCCATTTGCCACAATCAATG
ATG5 Exon 3 (F)	AAGAACACGGCTGTTTTTCC
ATG5 Exon 3 (R)	TGCTTAATAATGCAGAAAAATTCAC
ATG5 Exon 4 (F)	TTAAAGCCCCTGACATTTGG
ATG5 Exon 4 (R)	AATGGGACGAAGGAGAAATG
ATG5 Exon 5 (F)	TTGAAAAACTGGGGGATATAGTTC
ATG5 Exon 5 (R)	AATCTGGGCACAGAGGCTAC
ATG5 Exon 6 (F)	TTTCATCTCTTCATGTGAGGTATTC
ATG5 Exon 6 (R)	TGTCTGAGGCTTTCATAAATGG
ATG5 Exon 7 (F)	AAAAGGCACCTAATGCCAAC
ATG5 Exon 7 (R)	GAAATGTTTTAATGTTGCTGATTG
ATG5 Exon 8a (F)	TTGTTGGGTTTCTTTCTTGG
ATG5 Exon 8a (R)	TTCGTTAAGGAAAGATGGGTTTAC
ATG5 Exon 8b (F)	TTGAACTTTAGCTCATGAAAGTGG
ATG5 Exon 8b (R)	TCAGTGAAAATCGCAAAAGG
ATG5 Exon 8c (F)	CGATCATGGTTTTAGATCCCATA
ATG5 Exon 8c (R)	TTTTAAATAAAGACGGACACAACA
BLIMP1 exon 1 (F)	GCCGAGTGGCTAAGGAAATC
BLIMP1 exon 1 (R)	GCAGAAGCATGTTTCTACTGC
BLIMP1 exon 2 (F)	CATAGCCTCTCAGAAGGAGC
BLIMP1 exon 2 (R)	GTAGGGAGATTTGGCACCAG
BLIMP1 exon 3 (F)	GGCATCATTAATGTCTGTTTACTTATC
BLIMP1 exon 3 (R)	CGCACCTTAGTCCCAGCTAC
BLIMP1 exon 4 (F)	CTTACCTGTTTCCGCCCTG
BLIMP1 exon 4 (R)	CAGCATGTCTGGACCAATCC

BLIMP1 exon 5a (F)	CTAGCCCTCTGTGTAATCGC
BLIMP1 exon 5a (R)	GTTGTTGATGCCATTCATGC
BLIMP1 exon 5b (F)	AGTACGCTCACTACCCCAAGTTC
BLIMP1 exon 5b (R)	CTCACAGCCCCTTGGACTG
BLIMP1 exon 6 (F)	GAGCCAGCTTGAGAGCAGAG
BLIMP1 exon 6 (R)	TGGGAGGGTGACTCACAGAC
BLIMP1 exon 7a (F)	CCCGTTGGCAACTCTTAATC
BLIMP1 exon 7a (R)	GCCTTTGCCTTGTTCATGC
BLIMP1 exon 7b (F)	AAATGGTTTCCCCTCACCTC
BLIMP1 exon 7b (R)	TGGGGAAATTTTCGCAGTG
BLIMP1 exon 7c (F)	AGTCCTGTGGCCATTCAGAG
BLIMP1 exon 7c (R)	ACCGACGTGATTGTGAGGTC
BLIMP1 exon 7d (F)	TGAGCCAAGCCATGTAAAAG
BLIMP1 exon 7d (R)	TGGAGATGTCGCTCACTGAC
Miseq Leader VH1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGGACTGGACCTGGAGG
Miseq Leader VH2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACRCTCCTGCTGCTGACCA
Miseq Leader VH3a	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTGGGTTTTCCTTGTTGC
Miseq Leader VH3b	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGGAGTTKGGRCTGAGCTG
Miseq Leader VH4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTCCCAGATGGGGTCCTG
Miseq Leader VH5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCCTCCTGGCTGTTCTCC
Miseq Leader VH6	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGTCTCCTTCCTCATCTTCC
Miseq JH consensus	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTACCTGAGGAGACGGTGACC
GS FLX VH1-FR1 (1)	GCCTCCCTCGCGCCATCAGTCGTATGGCCTCAGTGAAGGTCTCCTGCAAG
GS FLX VH2-FR1 (1)	GCCTCCCTCGCGCCATCAGTCGTATGTCTGGTCCTACGCTGGTGAAACCC
GS FLX VH3-FR1 (1)	GCCTCCCTCGCGCCATCAGTCGTATCTGGGGGGTCCCTGAGACTCTCCTG

GS FLX VH4-FR1 (1)	GCCTCCCTCGCGCCATCAGTCGTATCTTCGGAGACCCTGT CCCTCACCTG
GS FLX VH5-FR1 (1)	GCCTCCCTCGCGCCATCAGTCGTATCGGGGAGTCTCTGAA GATCTCCTGT
GS FLX VH6-FR1 (1)	GCCTCCCTCGCGCCATCAGTCGTATTCGCAGACCCTCTCA CTCACCTGTG
GS FLX JH Consensus (1)	GCCTTGCCAGCCCGCTCAGTCGTATCTTACCTGAGGAGAC GGTGACC
GS FLX VH1-FR1 (2)	GCCTCCCTCGCGCCATCAGTCGCGAGGCCTCAGTGAAGGT CTCCTGCAAG
GS FLX VH2-FR1 (2)	GCCTCCCTCGCGCCATCAGTCGCGAGTCTGGTCCTACGCT GGTGAAACCC
GS FLX VH3-FR1 (2)	GCCTCCCTCGCGCCATCAGTCGCGACTGGGGGGTCCCTGA GACTCTCCTG
GS FLX VH4-FR1 (2)	GCCTCCCTCGCGCCATCAGTCGCGACTTCGGAGACCCTGT CCCTCACCTG
GS FLX VH5-FR1 (2)	GCCTCCCTCGCGCCATCAGTCGCGACGGGGAGTCTCTGAA GATCTCCTGT
GS FLX VH6-FR1 (2)	GCCTCCCTCGCGCCATCAGTCGCGATCGCAGACCCTCTCA CTCACCTGTG
GS FLX JH Consensus (2)	GCCTTGCCAGCCCGCTCAGTCGCGACTTACCTGAGGAGAC GGTGACC
GS FLX VH1-FR1 (3)	GCCTCCCTCGCGCCATCAGACGTGCGGCCTCAGTGAAGGT CTCCTGCAAG
GS FLX VH2-FR1 (3)	GCCTCCCTCGCGCCATCAGACGTGCGTCTGGTCCTACGCT GGTGAAACCC
GS FLX VH3-FR1 (3)	GCCTCCCTCGCGCCATCAGACGTGCCTGGGGGGTCCCTGAGACTC TCCTG
GS FLX VH4-FR1 (3)	GCCTCCCTCGCGCCATCAGACGTGCCTTCGGAGACCCTGT CCCTCACCTG
GS FLX VH5-FR1 (3)	GCCTCCCTCGCGCCATCAGACGTGCCGGGGAGTCTCTGAA GATCTCCTGT
GS FLX VH6-FR1 (3)	GCCTCCCTCGCGCCATCAGACGTGCTCGCAGACCCTCTCA CTCACCTGTG
GS FLX JH Consensus (3)	GCCTTGCCAGCCCGCTCAGACGTGCCTTACCTGAGGAGAC GGTGACC

Appendix 2 – Patient Characteristics

Population A

A cohort of 270 unselected pre-treatment CLL patients. FCR - Fludarabine, cyclophosphamide and rituximab

Characteristics	Number of patients 270
Mean Age (years)	63 (range 34-90)
Gender	
Female	123
Male	147
Binet	
Stage A	195
Stage B-C	75
SSeq IgHV	
Unmutated	104
Mutated	135
Not known	31
Treatment Status	
Required Treatment	122
Did not require treatment	148
Type of Treatment	
Chlorambucil alone/mini-CHOP/cyclophosphamid	50
Fludarabine alone/cladribine/bendamustine/ fludarabine-cyclophosphamide	25
Rituximab + bendamustine/chlorambucil/GA101/ CHOP/cyclophosphamide	12
FCR	49
Ibrutinib/lenalidomide/lumiliximab/alentuzumab	16

Population B

A cohort of 250 pre-treatment CLL patients enrolled onto either the ARCTIC (Attenuated dose Rituximab with ChemoTherapy In CLL) or AdMIRe (Does the ADdition of Mitoxantrone Improve Response) trials. FCR - Fludarabine, cyclophosphamide and rituximab, FCMR - Fludarabine, cyclophosphamide, mitoxantrone and rituximab.

Characteristics	Number of patients 250
Mean Age (years)	62 (range 33-80)
Gender	
Female	68
Male	182
Binet	
Stage A	34
Stage B-C	216
SSeq IgHV	
Unmutated	120
Mutated	92
Not known	38
TP53 status	
Mutated	20
Not mutated	230
Type of Treatment	
FCR	122
FCMR	61
FCMminiR	53
FCminiR/FCR	14

Population C

A cohort of 94 unselected CLL patients. FC – Fludarabine and cyclophosphamide FCR - Fludarabine, cyclophosphamide and rituximab

Characteristics	Number of patients 94
Mean Age (years)	67 (range 36-87)
Gender	
Female	28
Male	66
Binet	
Stage A	19
Stage B-C	36
Not known	39
SSeq IgHV	
Unmutated	33
Mutated	22
Not known	39
TP53 status	
Mutated	17
Not mutated	77
Type of Treatment	
Chlorambucil	16
Chlorambucil + rituximab/ofatumumab	3
FC	12
FCR	11
Ofatumumab	1
Alemtuzumab	1
Chemotherapy refractory	18
Unknown	15

Population D

A cohort of 67 previously treated CLL patients enrolled in either the CLL201 trial (a randomized phase 2 trial of fludarabine, cyclophosphamide, and mitoxantrone with or without rituximab in previously treated CLL) or the CLL202 trial (a phase 2 study of subcutaneous alemtuzumab plus fludarabine in patients with fludarabine refractory CLL). FCMR - Fludarabine, cyclophosphamide, mitoxantrone and rituximab, CamFlud – Combination alemtuzumab and fludarabine

Characteristics	Number of patients N=67
Mean Age (years)	64 (range 32-79)
Gender	
Female	14
Male	53
Unknown	2
Binet	
Stage A	9
Stage B-C	28
Not known	30
SSeq IgHV	
Unmutated	20
Mutated	14
Not known	33
TP53 status	
Mutated	19
Not mutated	48
No of Previous Treatments	
1	12
>1	24
Unknown	31
Type of Treatment	
FCMR	19
FCMminiR	20
CamFlud	29

Appendix 3 – NGS CLL panel

The targeted TSCA NGS panel was designed by Pauline Robbe and Ruth Clifford and was designed to cover hotspots in the following genes:

ATM
BIRC3
BCOR
BRAF
CDKN2A
CHD2
DDX3X
EGR2
FBXW7
HIST1H1E
IRF4
ITPKB
KLHL6
KRAS
MED12
MYD88
NOTCH1
NRAS
POT1
PTEN
RIPK1
SAMHD1
SF3B1
TGIF1
TP53
XPO1
ZFPM2
ZMYM3