THE UNIVERSITY of ADELAIDE

Investigating rare pathogenic mutations of the extended Fanconi Anaemia DNA repair pathway in Acute Myeloid Leukaemia (AML)

Kyaw Ze Ya Maung

B.Sc. Genetics (Hons)

A thesis submitted for the degree of Doctor of Philosophy in the School of Medicine at the University of Adelaide

July 2017

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Kyaw Ze Ya Maung

July 2017

Acknowledgements

A PhD is often considered as a giant step for just one man and a tiny step for mankind. However, for those who have gone through and completed this trial, it was about self-discovery. This journey was made possible and easier by the many interactions that I have had with people in my life.

I would firstly like to thank my supervisors Professor Richard D'Andrea and Dr James Gray for their faith in me, the opportunity to undertake this 4 year study and for sharing their wealth scientific knowledge and research experiences. I would also like to thank my co-supervisors, Dr Debora Casolari and Dr Sarah Bray, who have constantly been there to guide and help me through my candidature regardless of time, situations or triviality of my needs as a student. I would next like to thank Dr Saumya Samaraweera for her guidance and support throughout my candidature as well.

I would like to thank the members of the Acute Leukaemia Laboratory, past and present, for making my time in this laboratory a very pleasant one. Also, Mr Mahmoud Bassal, Mr Nick Li and Mr Jesse Cheah, my fellow candidates, I thank you for walking this road alongside me.

I would also like to thank my collaborators in Professor Thomas Gonda for providing great scientific knowledge and assistance to the Acute Leukaemia Laboratory, as well as Associate Professor Paul Leo and Mr Stephen Pederson for their assistance in handling and processing the bioinformatics aspects of my project.

Last but definitely not least, I would like to thank my dearest siblings for their endless support and motivation during these 4 years. Most importantly, I would like to thank my farther for always showing me that a man always sees his tasks through to the end and my mother for always been supportive of my decisions to pursue a scientific career. Though I have been away on a different continent for the past seven years, my family has been the main driving force behind me throughout this journey.

Abbreviations

1000 Genome	1000 Genome project (database)
A-EJ	Alternate end joining repair
AKT	AKT serine/threonine kinase
ALDH1A1	aldehyde dehydrogenase 1 family member A1
ALLG	Australian Leukaemia & Lymphoma Group
AML	Acute myeloid leukaemia
APL	Acute promyelocytic leukaemia
ASC	Adult stem cell
A-T	ataxia telangiectasia
ATM	ATM Serine/Threonine Kinase
ATRA	All-trans retinoic acid
ATRA	ATR Serine/Threonine Kinase
ATRIP	ATR interacting protein
BARD1	BRCA1 Associated RING Domain 1
BER	Base excision repair
BIC	Breast cancer Information Core
BLM	Bloom's syndrome RecQ Like Helicase
BM	Bone marrow
BMMNC	Bone marrow mononuclear cells
BRCA	Breast cancer gene
BrCa	Breast cancer
BrdU	Bromodeoxyuridine
BRIP1	BRCA1 Interacting Protein C-Terminal Helicase 1
BS	Bloom's syndrome
CADD	Combined Annotation Dependent Depletion
Cas9	CRISPR associated protein 9
CBF	Core binding factor
CBFA2	Core binding factor subunit-alpha 2
CD	Cluster of differentiation
CEBPA	CCAAT/enhancer-binding protein alpha
CHEK	Checkpoint kinases
c-Kit	Mouse tyrosine-protein kinase Kit
CLL	Chronic lymphoid leukaemia
CLP	Common lymphoid progenitor
CML	Chronic myeloid leukaemia
CMP	Common myeloid progenitor
COSMIC	Catalogue of Somatic Mutations in Cancer
CR	Complete remission
CRISPR	Clustered regularly interspaced short palindromic repeats
CtIP	RB Binding Protein 8, Endonuclease
CXCL12	C-X-C motif chemokine 12
CXCL4	Chemokine (C-X-C motif) ligand 4
CXCR4	C-X-C chemokine receptor type 4
D-A	Disease-associated
dbSNP	Database of single nucleotide polymorphisms
D-C	Disease-causing

DDR	DNA damage repair
DDX41	Deadbox RNA helicase 41
DEB	diepoxybutane
DNA2	DNA replication helicase/nuclease2
DNMT3A	DNA methyltransferase 3A
DSB	Double-stranded break
ELANE	Elastase, neutrophil expressed
ELN	European LeukaemiaNET
EME	Essential Meiotic Structure-Specific Endonuclease
ERCC4	ERCC Excision Repair 4, Endonuclease Catalytic Subunit
ESP	Exome sequencing project (database)
ExAC	Exome aggregate consortium database
EXO1	Exonuclease 1
FA	Fanconi anaemia
FAAP	FA associated proteins
FAB	French-American-British
Fadb	FA mutation database
FAN1	FA associated nuclease 1
FANC	Fanconi anaemia genes
FEN1	Flap Structure-Specific Endonuclease 1
Flk	Mouse fetal liver kinase-2
FLT3	fms like tyrosine kinase 3
FPL	familial platelet disorder
GATA1	GATA binding protein 1
GATA2	GATA binding protein 2
gDNA	Genomic DNA
GEP	Gene expression profiling
GMP	Granulocyte-macrophage progenitors
GSEA	Geneset enrichment analysis
HCT	Haematopoietic stem cell transplantation
HGB	Haemoglobin proteins
HGMD	Human Gene Mutation Database
HOXA3	Homeobox A3
HRR	Homologous recombination repair
HSC	Haematopoietic stem cell
ICL	Interstrand crosslinks
IDH	Isocitrate dehydrogenase
IF	Immunofluorescence assays
ITD	Internal tandem duplications
Jagged	Jagged ligand
JAK	Janus kinase
Ki67	Ki67 antigen
LIG	DNA ligases
Lin	Lineage marker
LSC	Leukaemic stem cells
LSK	c-Kit+, Lin- and Sca-1+
MDS	Myelodysplastic syndrome
MEP	Megakaryocyte-erythrocyte progenitor
MLH	MutL Homolog

MLL	Mixed-lineage leukaemia
MLLT3	Protein AF-9
MMC	mitomycin C
MMR	Mismatch repair
MPN	Myeloproliferative neoplasm
MPP	Multipotent progenitor
MRE11A	MRE11 Homolog, Double Strand Break Repair Nuclease
MRN	MRE11-NBN-RAD50 complex
MSC	Mesenchymal stromal cells
MSH	MutS Homolog
MsigDB	Molecular Signatures Database
MUS81	MUS81 Structure-Specific Endonuclease Subunit
NBN	Nibrin
NBS	Nijmegen breakage syndrome
NER	Nucleotide excision repair
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining repair
NMP1	Nucleophosmin 1
NOS	Not otherwise specified
NOTCH	Notch receptor
OMIM	Online Mendelian Inheritance in Man compendium
PAH	Princess Alexandra Hospital
PALB2	Partner And Localizer Of BRCA2
PARP1	Poly(ADP-Ribose) Polymerase 1
PARPi	PARP1 inhibitors
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase chain reaction
POL	DNA polymerase
RAD17	RAD17 Checkpoint Clamp Loader Component
RAD50	RAD50 Double Strand Break Repair Protein
RAD51	RAD51 Recombinase
RAD51B	RAD51 paralogue B
RAD51C	RAD51 paralogue C
RAD51D	RAD51 paralogue D
RAD52	RAD52 Homolog, DNA Repair Protein
RAH	Royal Adelaide Hospital
REV1	REV1, DNA Directed Polymerase
REV3L	REV3 Like, DNA Directed Polymerase Zeta Catalytic Subunit
RFC	Replication Factor C
RMI	RecQ Mediated Genome Instability proteins
RNA Pol	RNA polymerases
ROS	Reactive oxygen species
RPA	Replication protein
RUNX1	Runt-related transcription factor 1
RUNX1T1	RUNX1 Translocation Partner 1
SACRB	South Australian Cancer Research Biobank
Sca-1	Mouse stem cells antigen-1
SCF	Stem cell factor
SDSA	Synthesis-dependent strand annealing

sgRNA	Single-guide RNA
SLX4	SLX4 Structure-Specific Endonuclease Subunit
SNP	Single nucleotide polymorphisms
SS	Seckel syndrome
SSB	Single-stranded break
ssDNA	Single-stranded DNA
SUNCR1	Succinate receptor 1
TAM	Transient abnormal myelopoiesis
TCGA	The Cancer Genome Atlas
TET2	Tet Methylcytosine Dioxygenase 2
TGF-β1	Tumour growth factor beta-1
TKD	Tyrosine kinase domain
TLS	Translesion synthesis
T-MN	Treatment-related myeloid neoplasms
TOF	Tetralogy of Fallot
TOP3A	Topoisomerase (DNA) III Alpha
TP53	Tumour protein 53
U2AF1	U2 Small Nuclear RNA Auxiliary Factor 1
UBA52	Ubiquitin A-52 Residue Ribosomal Protein Fusion Product 1
UBC	Ubiquitin C
UBE2T	Ubiquitin Conjugating Enzyme E2 T
UBZ4	Ubiquitin binding zinc finger 4
USP1	Ubiquitin Specific Peptidase 1
UV	Ultraviolet
WCC	White cell count
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization
WNT	WNT protein
XRCC	X-Ray Repair Cross Complementing proteins

Abstract

Acute myeloid leukaemia (AML) is an aggressive haematologic malignancy caused by somatically-acquired structural rearrangements, single-nucleotide variants (SNV) and insertions/deletions affecting a set of well-defined genes. However, the contribution of germline mutations to AML initiation and progression is incompletely established. Genomic instability has classically been shown to contribute to cancer, and DNA repair disorders such as Fanconi anaemia (FA) are associated with increased risk of AML. FA is caused by a severe dysfunction of haematopoietic stem cells (HSCs) due to bi-allelic mutations in one of the 21 FANC DNA repair genes and these patients have an increased risk of developing AML that is ≥800 times than the population average.

The STRINGdb interface was used to expand the FA DNA repair pathway to include interacting components involved in DNA repair and cell cycle checkpoints, generating an extended network of 58 genes which have been termed the extended FA/BRCA-homologous recombination repair (FA/BRCA-HRR) network. This was further organised into several functional subgroups. Rare deleterious variants across this FA/BRCA-HRR network were identified from an exome-wide next generation sequencing study of an AML cohort and healthy Australian controls. Whole exome sequencing was performed for 145 adult and 23 paediatric AML cases at diagnosis, as well as a 329 all-female healthy controls. A total of 199 variants were identified across this network in the adult AML cohort, with 32 variants identified in the paediatric. Based on Sanger sequencing of a subset of variants using matched tumour and non-tumour samples it was predicted that the majority of these rare variants are germline in origin.

Adult AML samples with variants across the FA/BRCA-HRR network displayed significantly increased abnormal karyotype (P=0.012). In the adult AML samples, there was a significant increase in frequency of rare and damaging variants in a number of FA/BRCA-HRR network genes, compared to the normal healthy population (n=33370) from the publically available Exome Aggregate Consortium database. Furthermore, burden testing revealed enrichment of rare deleterious variants affecting *FANCL* and *RMI1* in this adult AML cohort. In the smaller paediatric cohort, the mutation spectrum of the FA/BRCA-HRR network genes was significantly different to that of the adult samples. In particular, *FANCC*, *FANCL* and *FANCM* variants which were significantly over-represented in adult AML (compared to healthy controls), were absent in the paediatric AML samples. Moreover, *BRCA1/2* variants were observed at a strikingly increased frequency in the paediatric samples. In the paediatric AML,

a significant association of *BRCA1/2* variants with Down syndrome and trisomy 21 was also observed (*P*=0.045).

The mutation data were also cross-referenced to the disease databases for FA and breast cancer to determine known disease-causing mutations (D-C mutations). In the adult AML, a significant enrichment was observed for D-C mutations affecting the 19 FANC genes compared to the all-female healthy control cohort [*P*=0.018; Odds ratio=3.3 (1.3-8.6)]. Similarly, an over-representation of D-C mutations affecting 16 of the FANC genes was observed in the adult AML cohort, compared to that reported in a separate published study of large healthy populations [*P*=0.002; odds ratio=3.4 (1.7-7.0)]. The mutation data was also compared against cancer and disease databases to determine the presence of disease-causing (D-C) mutations in all 58 genes, and to identify other disease-associated (D-A) variants. This analysis revealed a number of mutations that were present in multiple disease samples, while being absent, or present at low frequency, in the control cohorts.

Gene expression profiling was performed for a small set of adult AML (n=57) using microarray. This analysis, comparing gene expression in mutant versus non-mutant AML, revealed differences in gene expression pattern associated with presence of rare variants affecting functional gene subgroups within the FA/BRCA-HRR network. Examination of the individual differentially-expressed genes, and gene set enrichment analysis (GSEA), suggested that there may be potential differences in leukaemic cell of origin for AML carrying rare variants affecting the different functional subgroups of the extended FA/BRCA-HRR network. GSEA also suggested potential up-regulation of gene-sets associated with base excision repair and homologous recombination repair, as well as replicative stress, in samples carrying rare variants affecting the genes encoding the FA core & ID2 proteins. It is speculated that this may be indicative of increased basal replicative stress and DNA damage in the samples carrying these rare variants, with compensatory up-regulation of these repair pathways.

Based on the data presented it is hypothesised that rare germline variants affecting the genes in the FA/BRCA-HRR network result in subtle changes to the effectiveness of the FA DNA repair pathway in HSC, with a resultant modest increased pre-disposition to AML. An important question raised by this study relates to the cellular phenotype associated with rare, heterozygous deleterious variants affecting genes in the FA/BRCA-HRR network, and the FANC genes in particular. To investigate this further clonal cell lines were generated from a non-cancer cell line (MCF10A) carrying heterozygous or bi-allelic damaging mutations in *FANCL*. Changes to DNA repair capacity have been shown in cells with *FANCL* heterozygous mutations to be statistically different to the wild type cells.

Table of contents

List of figures

List of tables

List of appendices

Chapter 1 Introduction

1.1 Haematopoiesis

Haematopoiesis is the life-long generation of blood cells, a process that is critical for the survival of organisms possessing a circulatory system. It provides organisms, including mammals and arthropods, with the specialised blood cells which make up the basis of the innate and adaptive immune system, as well as cells from erythroid lineage that are central to the transport of oxygen.

1.1.1 The haematopoietic stem cell

Primitive haematopoiesis occurs initially in the yolk sac and switches to the foetal liver during foetal development. The switch to definitive haematopoiesis occurs just before birth and continues into adulthood in the bone marrow (Ciriza et al., 2013). During early development when haematopoietic stem cells (HSC) are derived from foetal liver, the HSC compartment rapidly expands to construct the circulatory system. This is a critical step in embryogenesis and HSC rely heavily on competent DNA repair pathways during this process. HSCs are one of the most studied adult stem cells (ASC) and possess the classical characteristics of ASC, including self-renewal and the capacity to differentiate into multiple lineage specific progenitors that drive tissue regeneration.

Based on extensive studies in the murine system, HSC is classified into long-term HSC (LT-HSC) and short-term HSC (ST-HSC). LT-HSC divide symmetrically for self-renewal and can engraft for long periods of time upon transplantation. Conversely, ST-HSCs are able to differentiate into lineage specific progenitors and only engraft for a short period of time on transplantation. A major challenge in the study of HSC has been the isolation of sufficient numbers of cells. Approximately 0.01% of nucleated cells in murine bone marrow are HSC, based on engraftment assays, with ST-HSC being more common (1 in 2000 cells) than LT-HSC (1 in 10000 cells) (Morrison et al., 1995). The enrichment of murine HSC is performed using the expression of the cell surface makers c-Kit$^+$, Lin$^-$ and Sca-1$^+$ which are commonly referred to as the LSK phenotype. The LSK phenotype has become an accepted standard for enrichment of HSC (Okada et al., 1992). HSC can also be enriched using CD150$^+$, CD48$^-$ and CD41$^-$ (Kiel et al., 2007). The LSK phenotype selects for both LT-HSC and ST-HSC fractions with further enrichment possible using the expression of CD34 and Flk-2; LT-HSC (CD34$^-$ and Flk-2$^-$), ST-HSC (CD34$^+$ and Flk2$^-$) and multipotent progenitor (MPP) (CD34$^+$ & Flk-2$^+$) (Kiel et al.,

2007, Yang et al., 2005). The enrichment of HSC can also be performed through cell cycle kinetics using Ki67 and BrdU staining since LT-HSC are more quiescent and less proliferative than ST-HSC (Challen et al., 2009).The expression of cell surface markers beyond the LSK phenotype to classify common lymphoid progenitors (CLP), myeloid progenitors, common myeloid progenitor (CMP), megakaryocyte-erythrocyte progenitor (MEP) and granulocyte-macrophage progenitors (GMP) in murine haematopoiesis has also been determined (**Figure 1.1**) (Kondo et al., 1997, Akashi et al., 2000, Seita and Weissman, 2010). The understanding of the human HSC compartment is less established and studies in human HSC have shown minor differences between murine and human HSC. Human HSC are Lin^- and KIT^+ ($CD117^+$) but are also $CD34^+$ unlike murine LT-HSC (Notta et al., 2011). Interestingly, the most important and widely used cell surface marker in the identification and/or enrichment of human HSC is CD34, which along with $Thy1^+$ ($CD90^+$), $CD38^-$, $CD45RA^-$ and $CD49f^+$, has become the standard for human HSC (Craig et al., 1993, Terstappen et al., 1991, Lansdorp et al., 1990, Seita and Weissman, 2010).

The maturation of HSCs is a tightly regulated process which results in an organised hierarchy that progressively becomes more restrictive towards a specific cell type in which the HSC loses its stem-cell characteristics. The LT-HSC differentiate into ST-HSC which further differentiates into one of two multipotent progenitors, either the common myeloid progenitor (CMP) or the common lymphoid progenitor (CLP) and these progenitors give rise to an array of terminally differentiated cells (**Figure 1.1**). The decision to self-renew or terminally differentiate is coordinated by cell-cell interactions, growth factors and the bone marrow microenvironment (niche) (Ho et al., 2015). The concept of the bone marrow niche was first proposed by Schofield in 1978 (Schofield, 1978) and has since been established as a complex ecosystem with various cellular and acellular factors (Schepers et al., 2015). Two aspects of the bone marrow niche have the greatest effect on HSC biology and function, the endosteal and perivascular niches (Levesque and Winkler, 2011, Li and Li, 2006). The endosteal niche consist mainly of osteoblasts (bone-forming cells) and cells that are committed to the osteogenic lineage and is physically closer to the bone than the perivascular niche. Dormant and quiescent LT-HSC are found in the endosteal niche wherein quiescence is maintained by secreted factors and chemokines such as stem cell factor (SCF), tumour growth factor beta-1 (TGF-β1), chemokine (C-X-C motif) ligand 4 (CXCL4) and C-X-C chemokine receptor type 4 (CXCR4) (Schepers et al., 2015). The perivascular niche contains multipotent mesenchymal stromal cells (MSC) and chemokines such as CXCL12 and cellular factors which influence HSC homing and

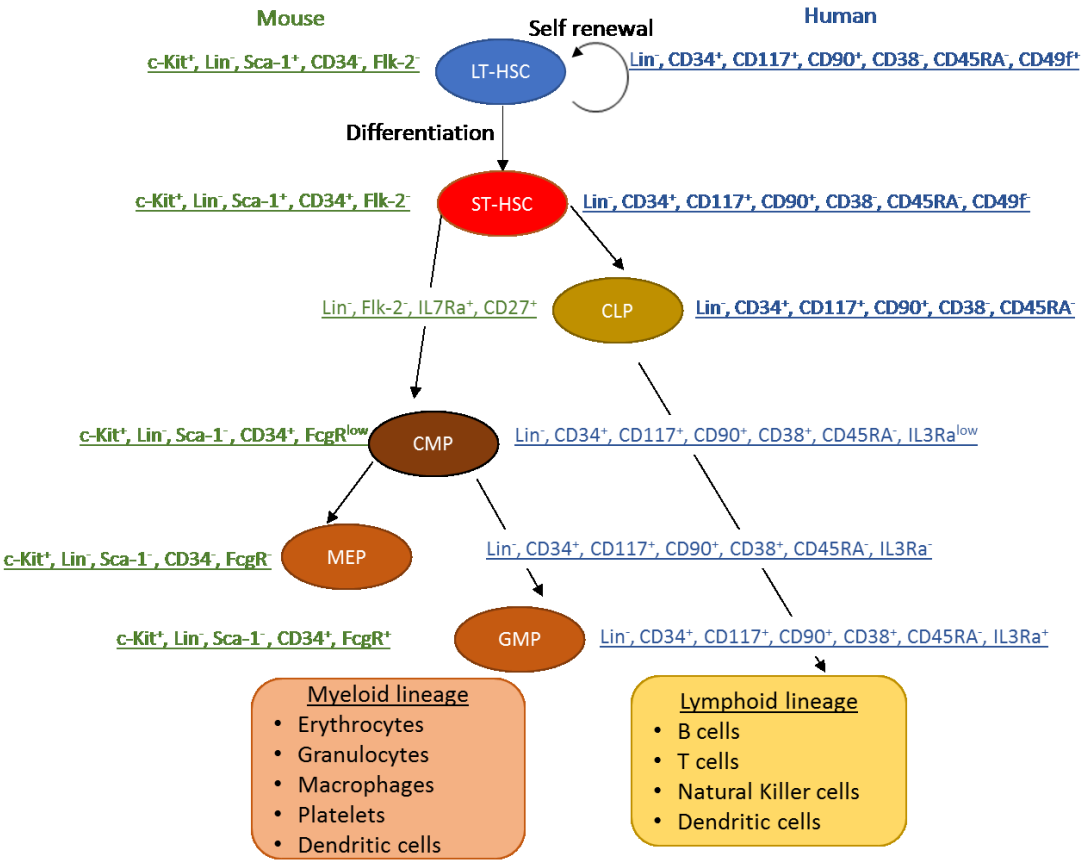


Figure 1.1. Haematopoiesis in mouse and human (adapted from (Seita and Weissman, 2010, Ho et al., 2015)). The maturation process of HSC to common myeloid and lymphoid progenitors (CMP and CLP), finally ending in mature blood cells. Cell surface markers used to enrich for various haematopoietic stem and progenitor cells (HSPC) at different stages of haematopoiesis in mouse and humans are shown.

lodgement. As such, actively cycling ST-HSCs are located in the perivascular niche. Key signalling pathways such as Jagged-NOTCH and WNT/β-catenin cascade also contribute to the maintenance of the bone marrow niche (reviewed in, (Ho et al., 2015)).

1.2 Acute myeloid leukaemia

Acute myeloid leukaemia (AML) is a complex haematologic malignancy with remarkable heterogeneity in morphology, cytogenetics, molecular characteristics, response to treatment and survival outcomes. It is a clonal disorder caused by the malignant transformation of HSC or committed myeloid progenitor cells that results in leukaemic stem cells (LSC) which are essential for malignant haematopoiesis. Upon transformation, LSCs acquire self-renewal capabilities with increased proliferation and blocked differentiation. This results in the accumulation of early myeloid precursor cells (myeloblasts) with markers of the myeloid lineage consisting of granulocytes (neutrophils, eosinophils and basophils), monocytes, macrophages, erythrocytes, megakaryocytes, and mast cells (Biondi and Rambaldi, 1996). AML can occur *de novo*, or secondary to myelodysplastic syndrome (MDS), myeloproliferative neoplasm (MPN) or as treatment-related myeloid neoplasms that arise as a result of cytotoxic therapy to a preceding malignancy (Granfeldt Ostgard et al., 2015). Clinical manifestation typically includes neutropenia, anaemia and thrombocytopenia (Meyers et al., 2005, Nebgen et al., 2016). Incidence rates of AML dramatically increase with age with AML having a median age of diagnosis of 69 years old. In Australia, approximately 4 in every 100000 individuals are diagnosed with AML with an estimate of approximately 1000 new cases annually and a corresponding mortality of 950 deaths annually (Cancer Australia 2017: http://www.aihw.gov.au/publication-detail/?id=60129558547).

1.3 Classification of AML

1.3.1 French-British-American and World Health Organization classification of AML

Different classification systems have been used to describe AML over the years, the first of which was by the French-American-British Cooperative Group (FAB) and was proposed in 1976 (Bennett et al., 1976). The FAB system uses cellular morphology and apparent maturity of the leukaemic blast when viewed using light microscopy as its main criteria. Eight subtypes (M0-M7) were defined with increased numbers matching the apparently increased maturity of the leukaemic blast (Shi et al., 2004). Additionally, for an AML classification, bone marrow biopsies require at least 30% of the bone marrow to be leukaemic blasts. A flaw in this

classification system however is that the FAB system does not account or correlate with patient prognosis and outcomes.

In 2002, the World Health Organization (WHO) proposed an updated classification system which integrates morphology with recurring chromosomal rearrangements and smaller molecular variations as criteria (Vardiman et al., 2002). The WHO classification system is regularly updated to incorporate advancement of the field (Arber et al., 2016). As of the latest revision in 2016 (Arber et al., 2016), there are 4 broad categories for the classification of AML as shown in **Table 1.1**. The first and broadest category is AML with recurrent genetic abnormalities including somatically acquired mutations in nucleophosmin 1 (*NPM1*), CCAAT/enhancer-binding protein alpha (*CEBPA*) and runt-related transcription factor 1 (*RUNX1*), and recurring chromosomal translocations in genes such as *RUNX1*, mixed-lineage leukaemia (*MLL*) and, GATA binding protein 2 (*GATA2*) (described in **Section 1.6** of this chapter). This category makes up approximately 50% of all adult AMLs. The second category is treatment-related myeloid neoplasms (T-MN) which are secondary AMLs developed by patients who have undergone treatment or therapy for another malignancy or disease. The third category is AMLs not otherwise specified (NOS). The WHO uses the FAB subtypes to further sub-divide this group of AMLs which have neither discerning karyotypic abnormalities nor clinical evidence consistent with treatment related malignancies. The fourth category is myeloid proliferations related to Down syndrome. This category includes transient abnormal myelopoiesis (TAM) which occurs within days after birth and progresses into myeloid leukaemia later (within 3 years of life). This category has been associated with inherited mutations associated with *GATA1* and the Janus kinase (JAK) and two Signal Transducer and Activator of Transcription (JAK-STAT) pathway (Bhatnagar et al., 2016, Kiyoi et al., 2007).

1.3.2 Proposed molecular classification of AML

A recent large-scale and multi-centre study performed by the Cancer Genome Project (Wellcome Trust Sanger Institute) and EMBL-EBI analysed 1540 AML patient samples tabulating the driver mutations in 111 cancer genes (Papaemmanuil et al., 2016). This data was then correlated with patient cytogenetic and clinical data. Owing to the size of the study, Papaemmanuil and colleagues were able to determine mutation co-occurrence and exclusivity patterns of AML driver mutations. A total of 5234 driver mutations were identified, involving 76 genes with 73% of the mutations being missense mutations (3824 out of 5234). Based on their findings, Papaemmanuil and colleagues proposed a genomic classification approach for

Table 1.1. World Health Organisation (WHO) classification of AML (adapted from (Arber et al., 2016)).

Classifications	Criteria for classifications
Acute myeloid leukaemia (AML) with recurrent genetic abnormalities	AML with recurrent genetic abnormalities
	AML with t(8;21)(q22;q22.1); *RUNX1-RUNX1T1*
	AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*
	APL with *PML-RARA*
	AML with t(9;11)(p21.3;q23.3); *MLLT3-MLL*
	AML with t(6;9)(p23;q34.1);*DEK-NUP214*
	AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM*
	AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);*RBM15-MKL1*
	Provisional entity: AML with *BCR-ABL1*
	AML with mutated *NPM1*
	AML with biallelic mutations of *CEBPA*
	Provisional entity: AML with mutated *RUNX1*
Treatment-related myeloid neoplasm	Therapy-related myeloid neoplasms
AML NOS (Not Otherwise Specified)	AML with minimal differentiation
	AML without maturation
	AML with maturation
	Acute myelomonocytic leukaemia
	Acute monoblastic/monocytic leukaemia
	Pure erythroid leukaemia
	Acute megakaryoblastic leukaemia
	Acute basophilic leukaemia
	Acute panmyelosis with myelofibrosis
Myeloid proliferations related to Down syndrome	Transient abnormal myelopoiesis (TAM)
	Myeloid leukaemia associated with Down syndrome

AML as shown in **Table 1.2**. This proposed classification incorporates both the mutations and cytogenetics for the identification of AML and expands upon the WHO classification.

1.4 Risk stratification of AML

1.4.1 ELN classification

In 2010, the European LeukaemiaNET (ELN) risk stratification scheme was proposed by Dohner and colleagues (Dohner et al., 2010). More recently, this classification has undergone a revision and as of 2017, the risk associated with highly penetrance germline predisposition mutations are now included (Dohner et al., 2017). Risk stratification recommended by the ELN for AML is grouped into favourable (17%), intermediate (68%) and adverse (15%) (Dohner et al., 2017). These categorical groups were established using standardised genetic abnormalities, clinical characteristics and patient outcome. The ELN classification stratifies patients depending on two prognostic indicators, patient factors and AML factors that are considered together to determine the outcome groups as detailed in. The main patient factor that determines outcome is the age of the patient, with poorer outcome and response to treatment increasing with age. AML factors include cytogenetic abnormalities and molecular genetics as described below in **Section 1.6.1** of this chapter. Key mutations include those in *NPM1* and tumour protein 53 (*TP53*), internal tandem duplications of fms like tyrosine kinase 3 (*FLT3*-ITD), as well as translocation and fusion genes involving *RUNX1* and *MLL*.

The favourable risk category includes *RUNX1-RUNX1T1* translocations, inversion of chromosome 16 (inv(16)), bi-allelic mutation of *CEBPA* and *NPM1* mutation without *FLT3*-ITD. The intermediate risk category includes *FLT3*-ITD mutation with *NPM1* mutation, the translocation involving *MLLT3-MLL* and other cytogenetic abnormalities that were not classified as favourable or adverse. The adverse risk category includes individuals with a "complex karyotype" which is determined by a patient carrying 3 or more karyotypic abnormalities, as well as aneuploidy such as monosomy 5 and monosomy 7.

1.4.2 Grimwade classification

An alternative to the ELN classification is the classification proposed by Grimwade and colleagues (Grimwade et al., 2010), another widely accepted risk stratification method. However, unlike the ELN classification that takes into account recurrent somatic mutation such

Table 1.2. Proposed Genomic Classification of AML. (adapted from (Papaemmanuil et al., 2016))

Proposed genomic Subgroup	No. of patients (%)
AML with *NPM1* mutation	418 (27)
[1]AML with mutated chromatin, RNA-splicing genes, or both	275 (18)
[2]AML with *TP53* mutations, chromosomal aneuploidy, or both	199 (13)
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB–MYH11*	81 (5)
AML with biallelic *CEBPA* mutations	66 (4)
AML with t(15;17)(q22;q12); *PML–RARA*	60 (4)
AML with t(8;21)(q22;q22); *RUNX1–RUNX1T1*	60 (4)
[3]AML with MLL fusion genes; t(9;11)(p21;q23)	44 (3)
AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); GATA2, *MECOM*(*EVI1*)	20 (1)
AML with *IDH2*-p.R172 mutations and no other class-defining lesions	18 (1)
AML with t(6;9)(p23;q34); *DEK–NUP214*	15 (1)
AML with driver mutations but no detected class-defining lesions	166 (11)
AML with no detected driver mutations	62 (4)
AML meeting criteria for ≥2 genomic subgroups	56 (4)

[1]Requires one or more driver mutations in *RUNX1*, *ASXL1*, *BCOR*, *STAG2*, *EZH2*, *SRSF2*, *SF3B1*, *U2AF1*, *ZRSR2*, or *MLL*-PTD or the presence of other class-defining lesions (inv(16), t(15;17), t(8;21), t(6;9), *MLL* fusion genes, or complex karyotype) in conjunction with *NPM1* mutations.
or driver mutations in *TP53*, *NPM1*, or *CEBPA* biallelic — two or more chromatin–spliceosome mutations are required.
[2]Requires *TP53* mutation, complex karyotype, or in the absence of other class-defining lesions, one or more of the following: −7/7q, −5/5q, −4/4q, −9q, −12/12p, −17/−17p, −18/18q, −20/20q, +11/11q, +13, +21, or +22.
[3]*MLL* has multiple fusion partners and the clinical implications depend on the specific fusion partner.

Table 1.3 ELN risk classification of AML (adapted from (Dohner et al., 2017))

Risk category	Genetic abnormality
Favourable	• t(8;21)(q22;q22.1); *RUNX1-RUNX1T1* • inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11 • Mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low} (low allelic ratio <0.5) • Biallelic mutated *CEBPA*
Intermediate	• Mutated *NPM1* and *FLT3*-ITD^{high} (high allelic ration >0.5) • Wild-type *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low} (low allelic ratio <0.5), without adverse-risk genetic lesions • t(9;11)(p21.3;q23.3); *MLLT3-MLL*
Adverse	• t(6;9)(p23;q34.1); *DEK-NUP214* • t(v;11q23.3); *MLL* rearranged • t(9;22)(q34.1;q11.2); *BCR-ABL1* • inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM*(*EVI1*) • -5 or del(5q); -7; -17/abnormal(17p) • Complex karyotype (3 or more abnormalities), monosomal karyotype • Wild-type *NPM1* and *FLT3*-ITD^{high} (high allelic ration >0.5) • Non-translocation mutations in *RUNX1* (if co-occur with favourable risk subtypes) • Mutated *ASXL1* (if co-occur with favourable risk subtypes) • Mutated *TP53* (Significantly associated with complex and monosomal karyotype)

as *NPM1* and *FLT3*-ITD, the Grimwade classification is based solely on karyotypic abnormalities. It includes three categories, like the newest revision of ELN, favourable, intermediate and adverse, as shown in **Table 1.4.** The favourable category includes the *RUNX1* and *CBF* translocations. The adverse category is also similar to that of the ELN classification except for the exclusion of *NPM1* and *FLT3*-ITD mutational status. The intermediate risk group accounts for cases that do not classify as either favourable or adverse.

1.5 Treatment of AML

Despite an increased understanding of the genomic and molecular landscape of AML, treatment options for patients have not improved since the introduction of anthracyline and cytarabine based regimens in the 1970s, and this treatment remain highly toxic with severe relapse rates (Mims and Stuart, 2013). The key breakthrough in treatment is the use of all-*trans* retinoic acid (ATRA) to treat acute promyelocytic leukaemia (APL), a subtype of AML (Degos and Wang, 2001). The induction therapy for APL does not include chemotherapeutic compounds but consists only of ATRA, and consolidation therapy with anthracycline described below, along with ATRA. Other targeted therapy include inhibitors that target mutated *FLT3* and *IDH1/2* that are currently in clinical trials (Garcia and Stone, 2017, Dang and Su, 2017).

Treatment options for AML are limited, with standard therapy consisting of chemotherapy using a combination of drugs. Allogenic haematopoietic stem cell transplantation (HCT) is an available treatment option for patients defined as "high-risk", which corresponds to being classified in the ELN "adverse" group. Induction therapy consists of 3 days of anthracycline and 7 days of cytarabine commonly referred to as the "7+3" regimen (Tefferi and Letendre, 2012). After the initial induction therapy, 60% to 80% of patients achieve complete remission (CR), which is defined by the presence of less than 5% blast in the bone marrow (biopsy) of the patient. Patients that did not achieve CR with the induction therapy may be given an additional intensive anthracycline regimen.

Once the patient achieves CR, they then undergo consolidation therapy. The intensity of the consolidation therapy is based on the individuals' age, intensive therapy for young patients (≤60 years of age) and less intensive therapies for the elderly (>60 years) (Estey, 2001). Younger patients with favourable and intermediate risk group are given 2 to 4 cycles of intense anthracycline regime while patients in the intermediate risk group undergo allogenic HCT

Table 1.4. Grimwade classification of AML (Adapted from (Grimwade et al., 2010))

Risk category	Genetic abnormality
Favourable	• t(15;17)(q22;q21) • t(8;21)(q22;q22) [Irrespective of additional cytogenetic abnormalities] • inv(16)(p13q22)/t(16;16)(p13;q22)
Intermediate	• Entities not classified as favourable or adverse
Adverse	• abnormal(3q) excluding t(3;5)(q21-25;q31-35) • inv(3)(q21q26)/t(3;3)(q21;q26) • add(5q), del(5q), -5 • +7, add(7q)/del(7q) [Excluding cases with favourable karyotype] • t(6;11)(q27;q23) • t(10;11)(p11~13;q23) • t(11q23) [excluding t(9;11)(p21~22;q23) and t(11;19)(q23;p13) • t(9;22)(q34;q11) • -17/abn(17p) • Complex (≥4 unrelated abnormalities)

(when stem cell donors are available). The patients in the adverse risk group are not given consolidation therapy but undergo allogenic HCT (when stem cell donors are available) upon achieving CR. Older patients in the favourable risk group are given a less intensive consolidation therapy regimen compared to the younger patients. Unfortunately, older patients in the intermediate and adverse risk groups have limited treatment options. Established intensive anthracycline regime and allogenic HCT is considered on a case by case scenario. However, as these patients that are not fit for intensive consolidation therapy, a number of low dose treatment approaches with azacitidine, decitabine and low-dose cytarabine are available.

1.6 AML genomics: somatically acquired mutations and pathogenesis of AML

The earliest next-generation sequencing (NGS) study of AML was carried out by Ley and colleagues, who later formed the Cancer Genome Atlas (TCGA) (Ley et al., 2008). This study involved the whole genome sequencing (WGS) of a single AML patient at diagnosis, remission and relapse, and reported 8 somatic mutations with similar frequencies at diagnosis and relapse (Ley et al., 2008). They also reported a second WGS study of a patient who had 64 somatic mutations that were subsequently tested in an additional 188 samples (Ley et al., 2010). This study also identified the recurrent *IDH1* and *IDH2* mutations (Mardis et al., 2009) which identified a novel pathway commonly affected in adult AML pathogenesis. The next major TCGA publication studying AML identified DNA methyltransferase 3A (*DNMT3A*) to be frequently mutated (62 out of 281 patients), with a particular *DNMT3A* mutation (p.R882H) found mutated in 37 out of the 62 patients (Ley et al., 2010).

The advancement in NGS technology also allowed Patel and colleagues to use an NGS approach to address the effects of recurrent mutations on treatment outcome for AML. The approach involved the sequencing of recurrently mutated genes (18 genes) in a large cohort of patients (n=398) who were undergoing treatment using a single drug, daunorubicin at high-dose and standard-dose (Patel et al., 2012). The cohort consisted only of patients younger than 60 years of age. The results from this study suggested that mutational analysis could be used for risk stratification, treatment decision and informative prognosis. In 2013, the TCGA analysed the genomes of 200 clinically annotated primary AML cases using WGS (n=50) and whole-exome sequencing (WES) (n=150). This study reported an average of 13 somatic mutations per AML case, and identified 23 which were significantly mutated (Cancer Genome Atlas

Research, 2013). These large scale genomic studies demonstrated the power of NGS for mutation discovery and in identifying key genes driving leukaemogenesis.

1.6.1 Classes of somatically acquired mutations in AML

In an attempt to better classify AML pathogenesis, Kelly and Gilliland (2002) proposed the "two-hit" model wherein two classes of acquired genetic lesions (Class I and Class II mutations) cooperate to give rise to AML (Kelly and Gilliland, 2002). Class I mutations confer a proliferative and survival advantage and consist mainly of signalling genes. Class II mutations affect mainly transcription factors and were proposed to impair differentiation. Advancement of NGS technologies has resulted in the discovery of new classes of recurrent mutations. As a result, Murati and colleagues expanded the classification of mutations classes to five. While retaining Class I and II, three new classes were added, epigenetic regulators (Class III), tumour suppressors (Class IV) and RNA maturation regulators (Class V) (Murati et al., 2012). Currently, the contribution of somatically acquired mutations in AML has reached saturation and the most recurrently mutated genes based on mutation classes as proposed by Murati and colleagues with the key genes for each classes are summarised in **Table 1.5**. Deep sequencing studies have identified specific classes of mutations that are found enriched in AML founder clones, suggesting that these mutations occur early in leukaemogenesis. Class III mutations in *DNMT3A*, *IDH1*, *IDH2* and *TET2* are suggested to be founder mutations while the Class I mutations in *FLT3*-ITD, *FLT3*-TKD are to be cooperative mutations (Sun et al., 2017).

1.7 Contribution of germline mutations to predisposition and pathogenesis of AML

1.7.1 Familial and genome wide association studies in AML

It is well established that selected germline mutations can confer a high risk of tumour initiation with classic examples including mutations the breast cancer associated (*BRCA*) genes. Mutations in these genes greatly increase the likelihood of developing breast and ovarian cancer (Petrucelli et al., 2010). Germline mutations contribute to rare cases of familial AML, occurring in a set of genes predominantly encoding haematopoietic transcription factors (section 1.7.2). The largest study to determine familial aggregation of AML was performed in by Goldin and colleagues (Goldin et al., 2012).

Table 1.5. Recurrent somatic mutations in AML

	Genes/Complex	Frequency in TCGA (1)	Role in AML pathogenicity	References
Class I	*FLT3*	28.0%	• Internal tandem duplication (*FLT3*-ITD) in the juxtamembrane domain of the receptor or mutation in the intracellular tyrosine kinase domain (*FLT3*-TKD) leads to constitutive activation	2, 3, 4,5
	NRAS *KRAS*	7.5% 4.5%	• Mutations at codons 12, 13 or 16 leads to constitutive activation that affect the MAPK/ERK pathway	1, 6
	c-KIT	8.0%	• Mutations in c-KIT lead to constitutive activation	7, 8
Class II	*RUNX1*	10.0%	• Commonly observed as translocations to form fusion proteins • t(8;21) results in *RUNX1-RUNXT1* which acts as a transcription repressor to block differentiation	9, 10
	CEBPA	6.0%	• Mutations lead to block in differentiation of granulocytes	10
	NPM1	27.0%	• Commonly observed as a 4-base pair frameshift insertion at positions 956-959 of exon 12 • Results in stronger nuclear export signal leading to cytoplasmic localisation	1, 11
	RARA	9.0%	• Commonly observed as 8 types of fusion proteins with *PML-RARA* being the most common (98%) • t(15;17)(q24;q21) *PML-RARA* binds to promoter region as a transcription repressor that blocks differentiation	12
Class III	*DNMT3A*	26.0%	• Commonly mutated at residue R882 which may cause the loss of function of DNMT3A, leading to aberrant methylation of the genome • Associated with poor prognosis • Often observed as a founder mutation in AML	1, 13, 14, 15
	TET2	9.0%	• Commonly observed as heterozygous mutations in the C-terminal catalytic domain • Aberrant methylation of the genome • Does not co-occur with *IDH* mutations	16, 17, 18

	IDH1/2	9.5%/ 10.0%	• Commonly observed as IDH1-R132 and IDH2-R172 and R140 resulting in block in differentiation • Mutant *IDH* converts the α-KG (alpha-ketoglutarate) which is a substrate required by enzymes such as *TET2* to 2-HG (2-hydroxy-glutarate)	16, 17, 18
	MLL (KMT2A)	5.5%	• Commonly observed as translocations (over 70 partners) or partial tandem duplication (PTD) • *MLL* fusion proteins have functional N-terminal domains of MLL, hence retaining the methyl transferase activity	19, 20
Class IV	*TP53*	8.0%	• *TP53* mutations have complex karyotype resulting from chromothripsis that results in complex somatic rearrangements with alternating copy numbers of one or more chromosomes • Loss of *TP53* also results in uncontrolled cell cycle checkpoint resulting in accumulation of mutations	21, 22
Class v	Spliceosome complex *(SF3B1, U2AF1, or SRSF2)*	14.0%	• Mutations alter splice site recognition • Affects methylation, DNA repair and apoptosis	23, 24
	Cohesin complex *(STAG1, STAG2, SMC1A, SMC3, and RAD21)*	13.0%	• Mutations in the sub units are mutually exclusive • Mutations result in dysregulated gene transcription in key haematopoietic genes such as *RUNX1*	1, 25, 26

(1) (Cancer Genome Atlas Research, 2013); (2) (Nakao et al., 1996); (3) (Abu-Duhier et al., 2001); (4) (Hatzimichael et al., 2013); (5) (Marcucci et al., 2011); (6) (Bowen et al., 2005); (7) (Ashman and Griffith, 2013); (8) (Qin et al., 2014); (9) (Lam and Zhang, 2012); (10)(Paz-Priel and Friedman, 2011); (11) (Heath et al., 2017); (12) (De Braekeleer et al., 2014); (13) (Shih et al., 2012); (14) (Yuan et al., 2016); (15) (Yang et al., 2015); (16) (Goyama and Kitamura, 2017); (17) (Inoue et al., 2016b); (18) (Dang et al., 2016); (19) (Meyer et al., 2009); (20) (Milne, 2017); (21) (Mrozek, 2008); (22) (Stengel et al., 2017); (23) (Ilagan et al., 2015); (24) (Inoue et al., 2016a); (25) (Leeke et al., 2014); (26) (Thol et al., 2014)

It included approximately 7000 AML patients with 25000 first-degree relatives matched with 28000 population controls and 90000 first-degree relatives. They reported that overall, AML did not aggregate in the relatives of patients, however, the relatives had moderately increased risk of developing haematologic and solid malignancies. Interestingly, Goldin and colleagues determined that the first-degree relatives of younger AML patients have a statistically higher risk (3-fold) of developing AML/MDS. This suggest that germline mutations may contribute a stronger role in AML which occurs at a younger age. This is consistent with studies in monozygotic twins that also revealed a very high concordance (25%) in infant twins with AML (Greaves et al., 2003), unlike for solid malignancies such as retinoblastoma (2% concordance) (Buckley et al., 1996).

Genome wide association studies (GWAS), have also been used to determine low risk germline alleles in AML (Manolio et al., 2009). The first landmark GWAS was performed by the Wellcome Trust Cast Control Consortium, consisting of 14000 cases distributed across seven common diseases (bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, type 1 and 2 diabetes) and 3000 controls (Wellcome Trust Case Control, 2007). They examined over 500000 SNPs and identified 24 SNPs which associated independently with the seven diseases (one with bipolar disorder, one with coronary artery disease, nine with Crohn's disease, three with rheumatoid arthritis, seven with type 1 diabetes and three with type 2 diabetes). They stated that the power of this study averaged across the SNPs with a minor allele frequency above 5% was estimated to be 43% for alleles with a relative risk of 1.3 and increasing to 80% for alleles with a relative risk of 1.5.

According to the largest GWAS catalogue available (NHGRI-EBI Catalogue: https://www.ebi.ac.uk/gwas/) as of 2017, there was only one reported GWAS in AML (Choi et al., 2013). Choi and colleagues performed a genome-wide SNP-array consisting of 247 NK-AML (118 for discovery and 129 for validation). They genotyped approximately a total of 870000 autosomal SNPs across their AML cohorts and the association with overall survival (OS). They determined four SNPs (rs2826063, rs12791420, rs11623492 and rs2575369) were associated with poorer OS in ethnically matched cohorts (discovery and validation). In a GWAS in t-AML by Knight and colleagues consisting of 80 t-AML cases (discovery), 70 t-AML cases (validation) and 150 controls, they identified 3 SNPs enriched in a subset of cases with loss of chromosomes 5 and 7 to be associated with t-AML (Knight et al., 2009).

1.7.2 Dominantly inherited mutations associated with AML predisposition

A number of familial predisposition syndromes have been described as being linked to inherited mutations in key haematopoietic transcription factors such as *RUNX1*, *GATA2* and *CEBPA*. Most recently, these genes have been incorporated into the 2017 revision of ELN classification (Dohner et al., 2017)

1.7.2.1 Germline mutations in RUNX1

The previously mentioned *RUNX1*, that has somatically acquired translocations in AML is one the best classified germline predisposition genes. *RUNX1*, also referred to as core binding factor subunit-alpha 2 (CBFA2), is a transcription factor that contains the DNA binding domain of the CBF transcription factor when it heterodimerises with CBFB to form CBF. CBF is primarily involved in haematopoiesis and the decision of HSC to divide, differentiate or remain quiescent (Link et al., 2010). The contribution of germline *RUNX1* mutations to familial haematological malignancies was first reported by Song and colleagues (Song et al., 1999). Autosomal dominant germline mutations have since then been reported in familial platelet disorder that increases the predisposition to AML (FPD/AML) (Beri-Dexheimer et al., 2008, Owen et al., 2008, Jongmans et al., 2010). It is noteworthy that the majority of the germline mutations are clustered to the N-terminal region of the protein which contains the DNA binding domain, thus, these mutations cause the loss of DNA binding function while retaining the dimerization capabilities (Schmit et al., 2015).

1.7.2.2 Germline mutations in CEBPA

Another previously mentioned gene is *CEBPA*. The first case linking germline CEBPA mutations with AML was reported by Smith and colleagues, describing a case of two siblings who developed AML within a span of two weeks (Smith et al., 2004). Their father had AML during childhood (age 10) and all three (father and two sons) had heterozygous germline deletion mutation in *CEBPA* at amino acid residue 158, which resulted in premature protein termination. A more recent study by Tawana and colleagues (2015) consisted of 10 *CEBPA* mutated families representing 24 members (Tawana et al., 2015). They observed that germline mutations are usually located in the N-terminus while the somatic mutations affect mainly the C-terminal. Similar to FPD/AML, familial AML derived from germline *CEBPA* mutations arise at an early age.

1.7.2.3 Germline mutations in GATA2

GATA2 is another transcription factor that is now classified as a high risk predisposition gene. While somatic mutations in *GATA2* have been associated with chronic myeloid leukaemia (CML) (Zhang et al., 2008), Hahn and colleagues reported *GATA2* as a high risk candidate predisposition gene for MDS/AML (Hahn et al., 2011). They identified the GATA2 mutation p.T354M in 3 families with AML and the deletion mutation p.T355del in a family with MDS. They have also functionally shown that the aforementioned *GATA2* mutations had reduced transactivation abilities. Similar to germline *RUNX1* and *CEBPA* mutations, familial AML derived from germline *GATA2* mutations arise at an earlier age. *GATA2* mutations are now well defined as associated with syndromic features (Collin et al., 2015).

1.7.2.4 Germline mutations in DDX41

The portable ATP-dependent deadbox RNA helicase, *DDX41* has recently been classified as a high risk predisposition gene for late-onset AML (Tawana and Fitzgibbon, 2016). Polprasert and colleagues first proposed that germline *DDX41* mutations predispose to the acquisition of a secondary hit in *DDX41* in myeloid neoplasms (MDS/AML, CML, non-Hodgkin lymphoma and Hodgkin lymphoma) (Polprasert et al., 2015). They reported that more than 65% of familial cases with *DDX41* mutations harboured the p.D140 frameshift along with an acquired mutation on the other allele of *DDX41*. This hypothesis was confirmed by Lewinsohn and colleagues who observed the same p.D140 frameshift mutation residing within a splicing site in 3 families associated with late-onset AML (Lewinsohn et al., 2016). A single copy of *DDX41* was sufficient for normal baseline haematopoiesis and the mean age of onset of 62 suggest a late-onset myeloid neoplasm. This is in stark contrast to germline *RUNX1*, *CEBPA* and *GATA2* mutations which were observed in younger patients.

1.7.3 Recessively inherited syndromes associated in AML predisposition

AML can arise from a number of syndromes associated with cancer predisposition. Since there are multiple subtypes of cancer within cell specific types meant that an extensive list of syndromes are associated with cancer. With respect to AML and haematopoietic malignancies, a number of these syndromes are autosomal recessive DNA repair disorders. Patients with these DNA repair disorders, such as Fanconi anaemia (FA), ataxia telangiectasia (A-T), Nijmegen breakage syndrome (NBS), Seckel syndrome (SS) and Bloom's syndrome (BS), have increased cancer predisposition as well as develop cancer at an early age (Kutler, 2003, Stankovic et al.,

1998, German, 1969, Tanaka et al., 2012, Kruger et al., 2007). In particular, FA has a predisposition to development of MDS and progression to AML.

1.7.3.1 Fanconi anaemia

FA is a rare inherited autosomal recessive disorder driven by progressive bone marrow failure and genomic instability, with a high risk of progression to AML and other malignancies. Patients with FA have an 800 fold increased risk of developing AML compared to the unaffected members of the population (Rosenberg et al., 2008). FA is a paediatric disease, first described by Guido Fanconi, a Swiss paediatrician in 1927, in three brothers from the same family who had physical abnormalities and died from severe pancytopenia at the age of 7 years (Velleuer, 2006). However, a further sixty-five years were needed to successfully clone and molecularly characterise the first FA gene, *FANCC* (Strathdee et al., 1992). Further studies have shown that FA results from biallelic mutations of the ever growing list of FANC proteins which participate in the FA DNA repair pathway. As of 2016, a total of 21 FANC genes had been identified.

1.7.3.2 Clinical features of FA

FA is classified into different subtypes based on the causative gene with the most common subtypes being FA-A, FA-C and FA-G, which make up approximately 85% of all FA cases (Schneider et al., 2015). However, the subtypes FA-D1 and FA-D2 have the most severe clinical phenotype and genomic instability (Hirsch et al., 2004, Kalb et al., 2007). Clinical features of FA include an array of congenital abnormalities such as short stature, skin pigmentation, and underdeveloped organs such as heart and kidney (Soulier, 2013). With the growing understanding of the pathology of FA and improved treatment, survival of FA patients has improved. However, persistent genomic instability leads to accumulation of mutations, which predispose FA patients to MDS/AML and solid malignancies. **Table 1.6** shows disease progression of FA patients and the various types of cancers reported by the International Fanconi Anaemia Registry (IFAR) from 1982-2001 (Kutler, 2003). Of the 754 reported cases, 16% developed haematopoietic malignancies and half of these were AML.

Table 1.6. Malignancies identified in the patients enrolled in the International Fanconi Anaemia Registry (1982-2001). Adapted from (Kutler, 2003).

Tumour type	No. cases
AML	60
MDS	53
ALL	5
CMML	1
Burkitt lymphoma	1
Liver adenoma	11
Hepatocellular carcinoma	6
Liver adenocarcinoma	1
Medulloblastoma	4
Astrocytoma	1
Wilms' tumour	4
Renal cell carcinoma	1
Nephroblastoma	1
Head and neck squamous cell carcinoma	19
Cancer of the vulva	8
Cervical cancer	6
Cutaneous cancer	3
Cancer of the anus	2
Oesophageal cancer	1
Breast cancer	3
Basal cell carcinoma	2
Neuroblastoma	1
Desmoid tumour	1
Gonadoblastoma	1
Melanoma	1
Neurilemmona	1
Osteogenic sarcoma	1

1.7.3.2.1 Pathogenesis of FA

Pathogenesis of FA results from chromosomal instability and hypersensitivity to interstrand crosslinking agents (ICL agents), such as mitomycin C and diepoxybutane (DEB), due to the loss of both alleles of a FANC gene. This unique characteristic forms the basis of a powerful diagnostic test for FA to measure chromosomal breakage (Deans and West, 2009). In recent years, it has been established that the pathogenesis of FA is caused by HSC dysfunction that leads to ineffective haematopoiesis and progressive bone marrow failure (Garaycoechea and Patel, 2014, Pontel et al., 2015). HSC are highly reliant on the FA DNA repair pathway for the maintenance of genomic integrity, particularly in the case of exposure to endogenous ICL and during repeated rounds of haematopoiesis due to replication and infections (Kaschutnig et al., 2015, Walter et al., 2015, Pontel et al., 2015).

1.7.3.2.2 Heterozygous mutations in FANC genes

While bi-allelic mutations in FANC genes can lead to FA with varied severity, the effect of heterozygous mutations on carriers is a heavily debated topic (Berwick et al., 2007, Tischkowitz et al., 2008, Lhota et al., 2016, Esteban-Jurado et al., 2016, Nielsen et al., 2016). Retrospective pedigree analysis of British families of FA patients by Tischkowitz and colleagues revealed that there was no increased risk or incidence of haematologic malignancies (Tischkowitz et al., 2008). However, a separate study by Berwick and colleagues observed an increased frequency of breast cancer in grandmothers of FA patients (Berwick et al., 2007). It is important to note that the distribution of FA subtypes in such studies are biased towards the more common subtypes (FA-A, FA-C and FA-G), and may not be a fair representation of other subtypes and their potential as cancer predisposition genes.

The advent of the genomics era has facilitated multi-centre cancer genomic studies using NGS technology to determine the mutational landscape of cancers. Mutations in *FANCA, FANCC, FANCG, FANCI, FANCL* and *FANCM* have been identified in cohorts of patients with familial pancreatic cancer, breast cancer, colorectal cancer, metastatic prostate cancer and paediatric cancer respectively (Nielsen et al., 2016, Esteban-Jurado et al., 2016, Lhota et al., 2016, Zhang et al., 2015, Hart et al., 2016, Lu et al., 2015). *FANCC* mutations have also been reported at increased frequency in children with sporadic AML(Awan et al., 1998). Several of the FANC

genes are also known high-risk familial cancer genes: the tumour suppressor and cancer predisposition genes *BRCA1* (*FANCS*), *BRCA2* (*FANCD1*), *BRIP1* (*FANCJ*), and *PALB2* (*FANCN*) (Nielsen et al., 2016).

Even though the cancer predisposition risk of heterozygous carriers is still being debated, on a molecular and cellular level, heterozygous phenotypes have been reported to show reduced function of the FA DNA repair pathway. Haploinsufficiency has been strongly established for the breast cancer associated genes *BRCA1* and *BRCA2* (Buchholz et al., 2002, Jeng et al., 2007, Feilotter et al., 2014, Pathania et al., 2014, Sedic and Kuperwasser, 2016, Tan et al., 2017). Dominant negative mutations have also been reported in *RAD51* (*FANCR*), *BRCA1* and *BRCA2* (Ameziane et al., 2015, Wang et al., 2015, Vaclova et al., 2016). Evidence for heterozygous phenotypes for the remaining FANC genes, however, is limited due to the rarity of FA as a disease. During the early 1980s, before the establishment of the different subtypes of FA, Auerbach and colleagues showed that lymphocytes of heterozygous parents and obligate heterozygous carriers were more sensitive to ICL induced by DEB than individuals (with wildtype FANC genes (Auerbach et al., 1981). More recently, studies have reported that lymphocytes from heterozygous carriers are sensitive to bleomycin and ionizing radiation (Barquinero et al., 2001), which suggests that there is a phenotype for heterozygous FANC gene mutations that should be studied further and potentially utilised for clinical decisions such as treatment and genetic counselling.

1.7.3.2 Ataxia telangiectasia

Ataxia telangiectasia (A-T) is a neurodegenerative disease with immunological defects, radiation sensitivity, sterility, accelerated aging, and cancer susceptibility, caused by the loss of function of the ATM kinase (van Os et al., 2017). ATM is widely considered as one of the three DNA signalling kinases (along with ATR and DNA-PKc) involved in regulating DNA repair pathways and cell cycle (Marechal and Zou, 2013). A study performed by Suarez and colleagues consisting of 279 A-T patients showed that 24.5% of individuals developed cancer, with the majority of these being lymphoid malignancies (Suarez et al., 2015). A recent systematic analysis of heterozygous carriers of *ATM* mutations observed a threefold increased risk of developing breast cancer in female carriers, as well as cancer of the digestive tract (van Os et al., 2016).

1.7.3.3 Seckel syndrome

Seckel syndrome (SS) is an autosomal recessive disorder wherein patients present with microcephaly. Similar to FA, SS has multiple subtypes (SCKL1-SCKL10) that are caused by the loss of function in genes involved in DNA repair and cell cycle regulation. The first mutations were identified in *ATR*, one of the three key DNA damage response signalling kinases (O'Driscoll et al., 2003). Though the incidence of cancer is lower than in A-T (Hayani et al., 1994), it has been shown that loss of ATR function in mouse models results in embryonic replicative stress and an accelerated aging phenotype (Murga et al., 2009).

1.7.3.4 Nijmegen breakage syndrome

Nijmegen breakage syndrome (NBS) is an autosomal recessive chromosomal instability disorder caused by the loss of function of the gene nibrin (*NBN*). NBS patients present with congenital defects such as microcephaly, intrauterine growth retardation and short stature, immunodeficiency, increased risk for cancer, and premature ovarian failure in females (Wegner et al., 1988, Saar et al., 1997). On a molecular level, cells from A-T, SS and NBS are phenotypically similar due to the three genes (*ATM*, *ATR* and *NBN*) all participating in DNA repair pathways, particularly during replication. The cancer incidence in NBS patients has been reported to be as high as 40%, with a predilection for lymphoma, similar to A-T (Chrzanowska et al., 2012). Heterozygous carriers are asymptomatic but have been reported to have increased cancer risk (Steffen et al., 2004, Seemanova et al., 2007).

1.7.3.5 Bloom's syndrome

Bloom's syndrome (BS) is caused by loss of function mutations in the RECQL-helicase BLM. BS patients present with congenital defects, partial immunodeficiency, sensitivity to sunlight, and increased risk of developing any of multiple cancer at an earlier age (Cunniff et al., 2017). Based on the Bloom's Syndrome Registry, 46% of BS patients develop cancer with the majority of them being epithelial carcinoma (52.5%), while 11.3% developed AML (Bloom's Syndrome Registry: http://weill.cornell.edu/bsr/data_from_registry/).

1.8 DNA damage and repair pathways

As discussed above in **Section 1.7.2**, ineffective or compromised DNA repair can give rise to numerous syndromes. Various types of DNA damage can occur endogenously during DNA

replication and cellular metabolism, as well as through exposure to genotoxic compounds such as chemotherapeutic and crosslinking compounds and exogenous environmental factors such as UV light and radiation (Ciccia and Elledge, 2010, Jeggo et al., 2016). The severity of the damage ranges from the simple, such as single-stranded breaks (SSB), to bulky adducts that distort the helical structure, to extremely detrimental double-stranded breaks (DSB), which, if not faithfully repaired, can trigger cellular apoptotic cascades (**Figure 1.2**). The FA DNA repair pathway functions to remove interstrand crosslinks (ICLs) and is one of the many DNA damage repair (DDR) pathways used by cells to remedy genetic lesions. The FA DNA repair pathway is described below.

1.8.1 Interstrand cross-link repair (FA DNA repair pathway)

Most FANC genes were initially 'orphan' genes, not known to participate in other biological pathways. The unique feature of FANC genes is their involvement in the FA DNA repair pathway (**Figure 1.3**). This pathway is activated during the G2/S phase of mitotic cell division when the cell encounters ICL. This pathway also plays a vital role in clearing genotoxic damage caused by chemical ICL agents, endogenous ICL agents such as aldehyde and reactive oxygen species (ROS) and exposure to ultraviolet (UV) light and stabilisation of the stalled replication fork (Naim and Rosselli, 2009) (Lachaud et al., 2016).

The FANC genes are broadly classified into three major groups in the order of their participation in the pathway. The first group is the FANC core complex (*FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL* and *FANCM*), which assembles as an ubiquitin E3 ligase complex and is responsible for the activation of the FANCI-FANCD2 (ID2) complex (Hodson and Walden, 2012). The loss of function of any of the 8 proteins will destabilise the entire complex, resulting in the loss of E3 ligase function and the breakdown of the FA DNA repair pathway. The second group is the ID2 complex. It consists of *FANCD2* and *FANCI* that heterodimerise and, when monoubiquitinated, act as a platform for the recruitment of the third group that carry out the DNA repair process. The third group consists of the remaining downstream FANC genes (*BRCA2, BRIP1, PALB2, RAD51C, SLX4*, *ERCC4,* and *BRCA1*), which carries out the downstream DNA repair processes (nucleolytic incision repair (NIR), and homologous recombination repair (HRR)).

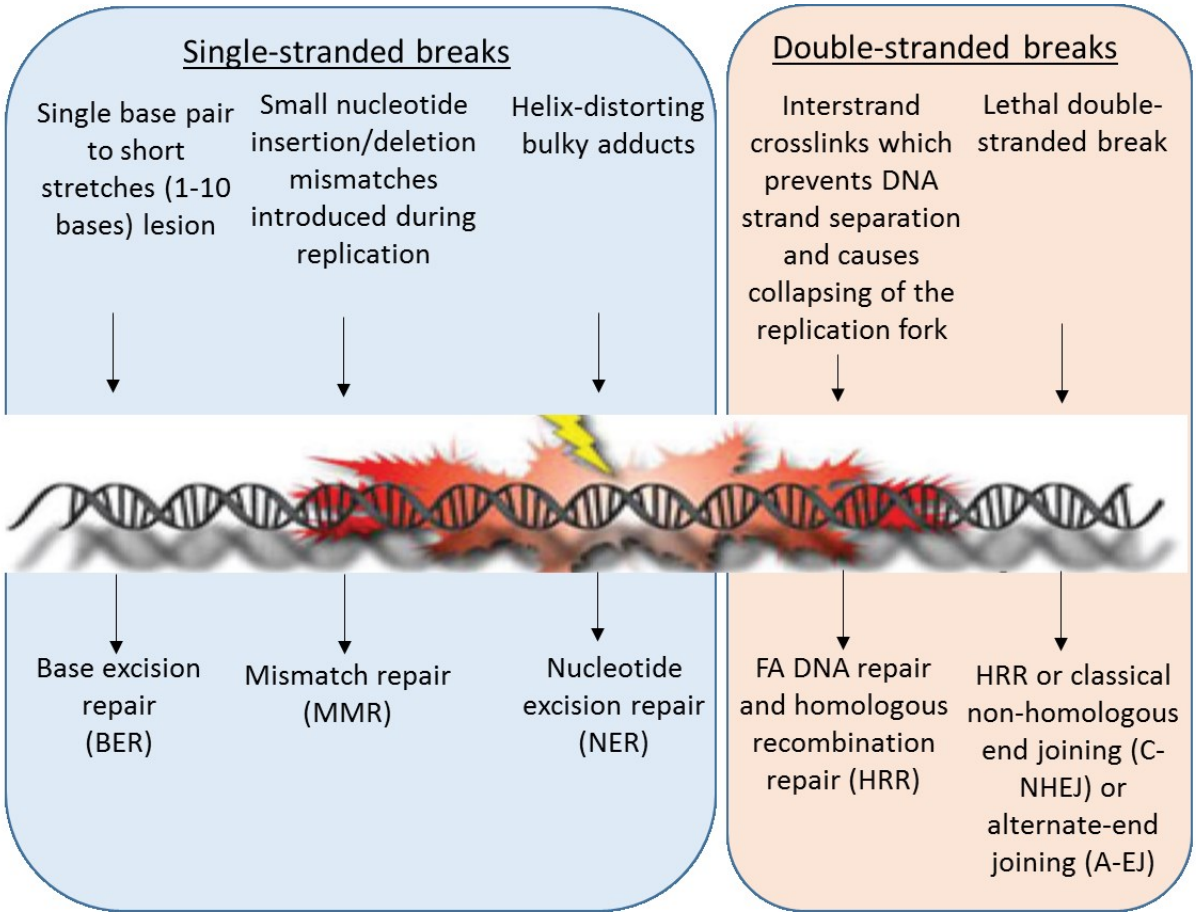


Figure 1.2. Types of DNA lesions and respective pathways (adapted from (Ghosal and Chen, 2013)). Single stranded breaks caused by hydrolysis, oxidation and metabolic by-products such as ROS are repaired by base excision repair (BER). Mismatch repair (MMR) is predominantly involved in the removal of erroneous and small nucleotide insertion/deletion mismatches that are introduced during replication. Bulky adducts and helical distortion caused by UV are repaired by NER. ICLs caused by chemotherapeutic compounds and stalled replication forks are repaired by the FA DNA repair pathway and homologous recombination repair (HRR) pathway. Double stranded breaks are caused by irradiation, chemotherapeutic compounds, meiosis and lymphocyte maturation. Pathways involved in DSB repair are HRR, classical non-homologous end joining (C-NHEJ) and alternate-end joining (A-EJ) repair pathways.

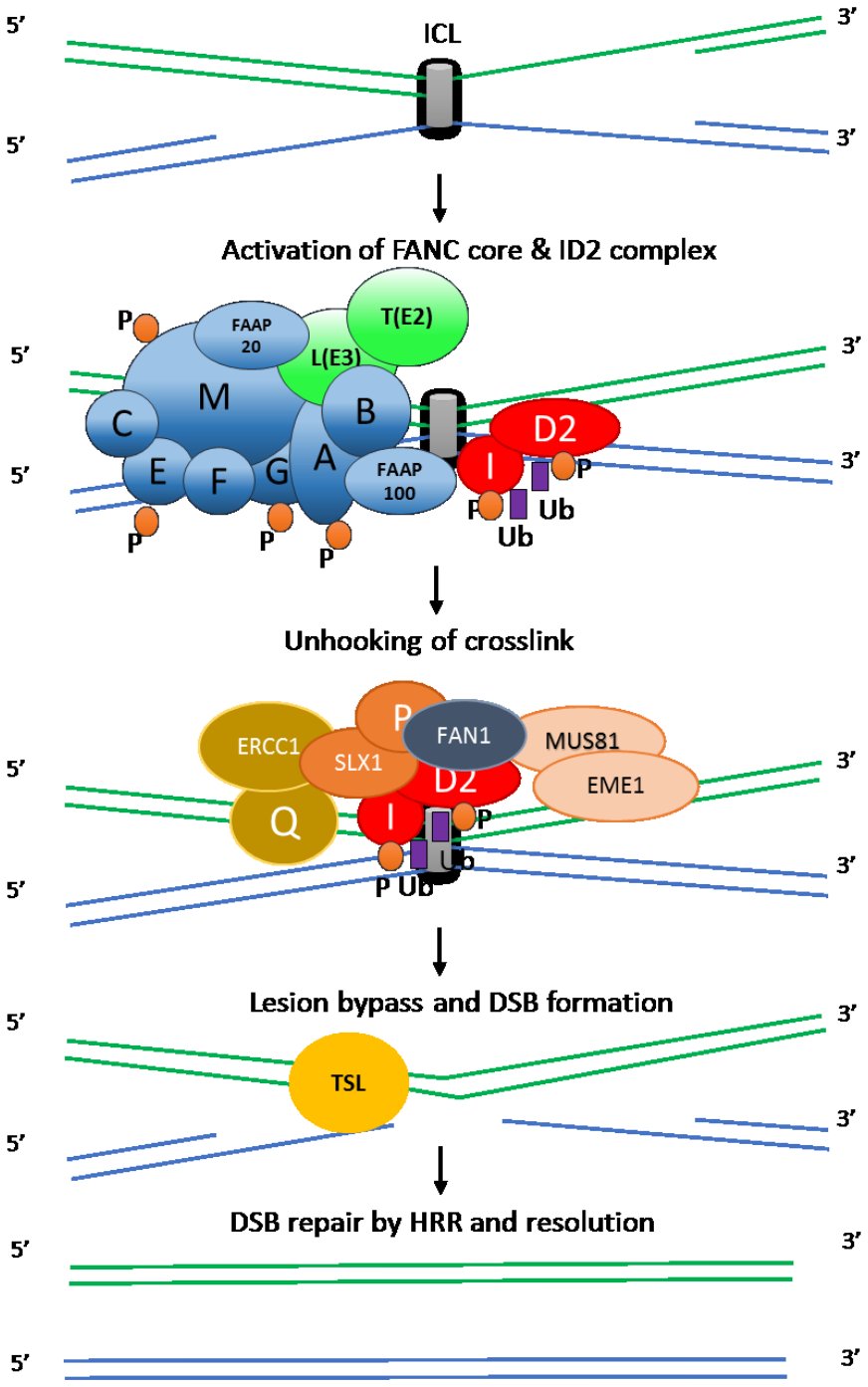


Figure 1.3. Interstrand cross-link (ICL) repair (adapted from (Garaycoechea and Patel, 2014)). The Fanconi Anaemia DNA repair pathway is activated upon ICL damage. The FANC core complex consisting of FANCA/B/C/E/F/G/L/M binds to the lesion, along with the ID2 complex (FANCI and FANCD2). The ID2 complex is activated through phosphorylation by ATR and monoubiquitination by the FANC core complex. Structure-specific endonucleases ERCC1-FANCQ, SLX1-FANCP, MUS81-EME1 and FAN1 are involved in the unhooking of the ICL. Translesion synthesis bypasses the lesion to generate a DSB. DSB is repaired by downstream FANC genes and HRR.

1.8.1.1 FANC core complex

Of the genes involved in the core complex, FANCM plays a key role in initiating the FA pathway by recognising genetic lesions and activating the ATR/CHEK1 checkpoint during the G2/S phase of replication (Collis et al., 2008). Phosphorylation of CHEK1 by ATR prevents late origin activation, halting replication. FANCM, with FANCA, FANCE, FANCD2 and FANCI are phosphorylated by checkpoint kinases. FANCM has a translocase/helicase motif and works in conjunction with FAAP24 (FA associated proteins 24), FAAP100, MHF1 (histone fold protein 1) and MHF2 (Singh et al., 2010). These stabilizing proteins aid FANCM by stalling the replication fork and facilitate recruitment of the remaining members of the core complex (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG and FANCL). FANCL contains a Zinc finger domain with E3 ligase activity and works in conjunction with UBE2T (FANCT), which has E2 ligase activity, to monoubiquitinate FANCD2 and FANCI (de Winter and Joenje, 2009). Once the FANC core complex is fully assembled, it performs monoubiquitination of the FANCD2 and FANCI. A recent study by Swuec and colleagues demonstrated the homo-dimeric modules within the FANC core complex and the contributions of these modules in the stabilisation and monoubiquitination of the ID2 complex (Swuec et al., 2017).

1.8.1.2 ID2 complex

FANCD2 and FANCI proteins are paralogues that exist as heterodimers, hence the name ID2 complex. The ID2 complex has been successfully isolated and the crystal structure has also been determined (Joo et al., 2011). The monoubiquitination of the ID2 complex is the hallmark of FA pathway activation and is often used in conjunction with the chromosomal breakage test using ICL agents in diagnostic tests. The monoubiquitination of FANCD2 and FANCI has been shown to be required for the nucleolytic incision at the ICL and the recruitment of the RAD51 recombinase for homologous recombination repair.

1.8.1.3 Nucleolytic incision and lesion bypass

The next step is the nucleolytic incision of the ICL and the exact mechanism surrounding the unhinging step is still unclear (Kim and D'Andrea, 2012). It has been proposed that a number of structure-specific endonucleases are involved (Palovcak et al., 2017). This is also called the

'unhooking step'. The ID2 complex recruits a series of structure-specific endonucleases that are involved in this step. SLX4 (FANCP) and FA associated nuclease 1 (FAN1) contain a unique ubiquitin-binding domain called UBZ4 (ubiquitin binding zinc finger 4) that recognize and binds to the ubiquitin of the activated FANCD2 (Ciccia et al., 2008). SLX1 is the catalytic subunit of SLX4 and the SLX1-SLX4 complex has 5'-endonuclease activity, while FAN1 has 5'-3'-exonuclease activity (Coulon et al., 2004) (Kratz et al., 2010). FAN1 has also recently been shown to contribute to replication fork stability (Lachaud et al., 2016). Other nucleases that interact with SLX4 include MUS81-EM1 and XPF-ERCC1 (ERCC4-ERCC1). It has been suggested that the various nucleases function in different order, with MUS81-EME1 being the first nuclease to carry out the first cleavage. Crossan and colleagues have shown that disruption of SLX4 in a murine model resulted in lowered leukocytes and platelet counts, and other congenital defects that arise with FA when compared to the wild-type mice (Crossan et al., 2011).

Once unhooking is complete, translesion DNA synthesis (TLS) is carried out to extend the nascent strand beyond the ICL using the Y family TSL polymerases. The Y family TSL polymerases (POLI, POLH, POLK, and REV1) lack the 3'-5' proofreading activity and allow for base pair mismatches (Waters et al., 2009). The B family polymerase REV3L is also involved (Sale et al., 2012). The monoubiquitination of PCNA by the E2-E3 ligase complex RAD6-RAD18 triggers the recruitment of REV1 and REV3L to the site of lesion. REV1 regulates the TLS repair by switching out different TSL polymerases (Haynes et al., 2015).

Once TLS has been performed, the BRCA/RAD51 driven homologous recombination repair (HRR) described below which recruits the RAD51 recombinase to the DSB to complete the repair process (Gravells et al., 2013). This is facilitated by the remainder of the FANC members, which consist mainly of breast cancer susceptibility genes (FANCD1/BRCA2, FANCJ/BRIP1, and PALB2/FANCN) as well as the RAD51 paralogues RAD51B, RAD51C/FANCO, RAD51D, XRCC2 and XRCC3 (D'Andrea, 2013, Godin et al., 2016).

1.8.2 Homologous recombination repair

HRR is a highly complex and tightly regulated process involved in the removal of DSBs during replication. HRR can also be triggered by ICL repair. It is considered to be the most important DDR pathway. HRR can be categorised into pre-synapsis, synapsis and post-synapsis. Pre-

synapsis involves the recognition and assembly of the repair components, synapsis involves the strand invasion and synthesis, and post-synapsis involves the resolution of the DNA heteroduplex (**Figure 1.4**).

The signalling of DSB repair is similar to that of ICL but involves the serine/threonine kinase ATM, as well as ATR. BRCA1 and PARP1 have also been implicated in being early recruited factors to the DSB site, along with MRE11A, NBN and RAD50 (the MRN complex). The MRN complex is essential for the recruitment of the ATM kinase (Lee and Paull, 2005) which auto-phosphorylates and phosphorylates the MRN complex, along with CtIP (RBBP8), BLM helicase, DNA replication helicase/nuclease2 (DNA2) and exonuclease 1 (EXO1) to perform strand resection (Shiotani and Zou, 2009, Liu and Huang, 2016). This creates a stretch of SSB overhang which allows for the binding of the replication protein A (RPA) to the single-stranded DNA (ssDNA), in order to prevent the degradation of the ssDNA (Zou and Elledge, 2003), as well as ATR and its heterodimeric partner ATRIP (Cortez et al., 2001). ATR and ATRIP are then activated and this leads to the phosphorylation of CHEK1 which induces cell cycle arrest at G2/S phase, allowing for the repair process to complete before cell cycle progression. CHEK1 also phosphorylates the recombinase RAD51 to recruit it to the site (Bahassi et al., 2008).

RAD51 recruitment involves BRCA1, BRCA2, BARD1 and PALB2 with stabilisation of the recombinase being performed by the RAD51 paralogues mentioned above (Liu et al., 2010). It is at this repair step where the FA DNA repair and HRR pathways overlap. The RAD52 homologue in yeast performs the role of BRCA2 but it has been implicated that in humans, RAD52 is still able to perform this task in BRCA2 null cells (Sugawara et al., 2003). The binding of RAD51 to the site of damage displaces RPA. RAD51 and its paralogues next form nucleoprotein filaments (RAD51 filaments) on the ssDNA to perform homology search and strand invasion. This process is assisted by RAD54 which stabilises the RAD51 filament and promotes strand invasion (Mazin et al., 2003), while also functioning to dissociate RAD51 from the heteroduplex DNA thus allowing for DNA polymerases to bind to the site (Li and Heyer, 2009). DNA polymerases then synthesise and extend the invading strand using the homology from the sister chromatid as a template. The capture of the second end of the DSB will result in the formation of the classical double Holliday junction. The resolution of the Holliday junction is performed by the structure-specific endonucleases BLM complex that produces no crossovers, while the resolution by MUS81-EME1 and SLX1-SLX4 will result in crossover between the two homologous chromatids (Matos and West, 2014). If the second end of the DSB

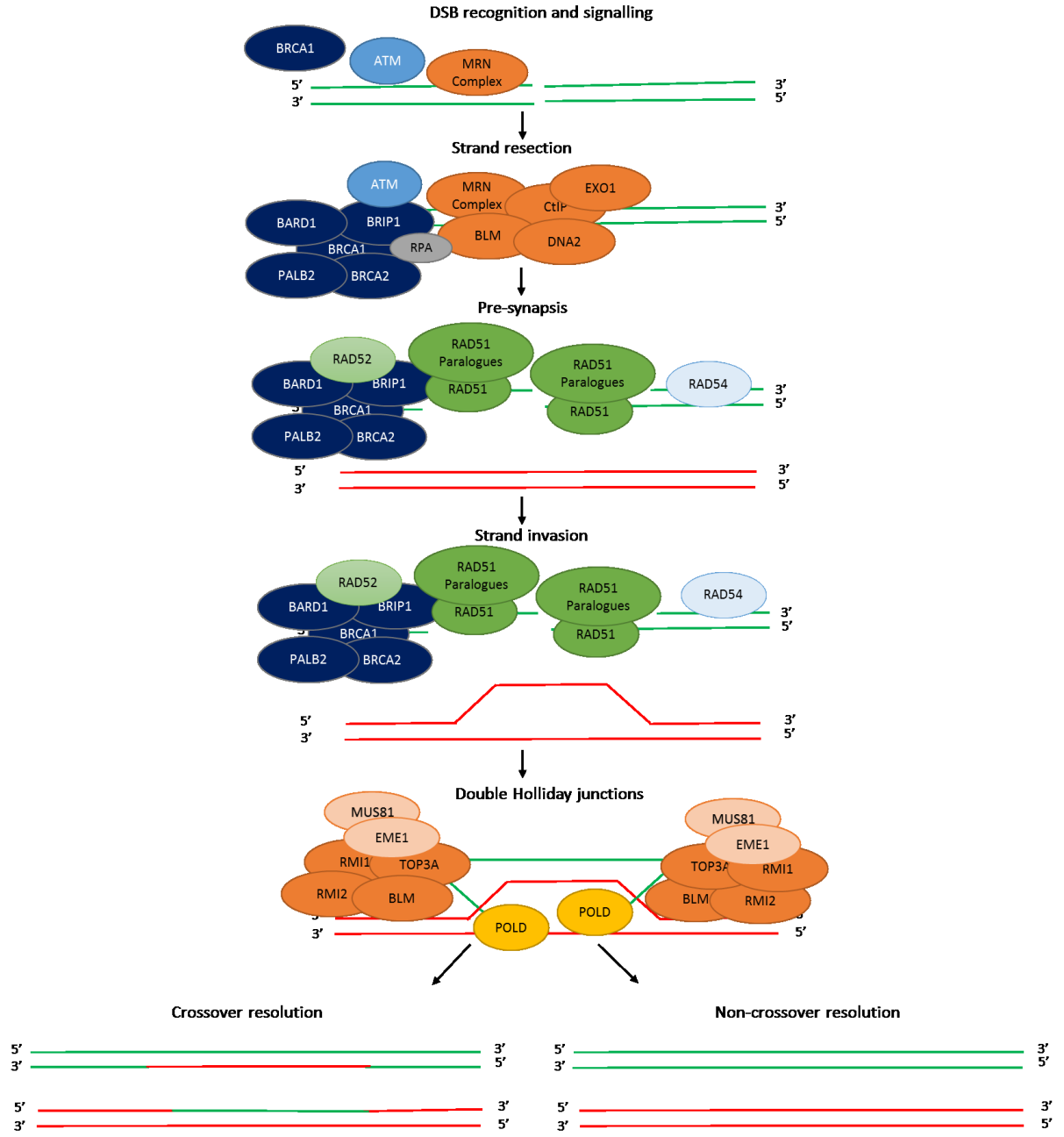


Figure 1.4. Homologous recombination repair (HRR). DSB is recognised by the MRN complex, ATM and BRCA1. MRN complex, ATM and breast cancer associated proteins (BRCA1/2, BARD1, PALB2 and BRIP1) localise to the lesion. BLM, CtIP, DNA2, EXO1 and MRE11A perform strand resection to generate overhangs. RPA coats the single stranded DNA. BRCA2 and RAD52 facilitate the recruitment of RAD51 to the site. The RAD51 paralogues (RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3) stabilise RAD51. RAD51 displaces RPA and form RAD51 filaments for strand invasion. Strand invasion occurs and repair synthesis is performed by POLD. Double Holliday junctions are resolved by the BLM complex and MUS81-EME1 endonuclease.

was not captured, synthesis-dependent strand annealing (SDSA) pathway of resolution will take place. RAD52 has also been shown to promote SDSA (Nimonkar et al., 2009).

1.8.3 Remaining DNA repair pathways

As previously shown in **Figure 1.2**, ICLs are just one type of DNA lesion. Other DNA repair pathways are involved in the repair of other DNA lesions. Maintaining genomic stability is paramount for cells to prevent malignant transformation, hence, these DNA repair pathways overlap with one another. Base excision repair (BER) is predominantly involved in the removal of non-helix-distorting single base lesions that occur on one strand of the DNA from endogenous deamination, oxidation and methylation, as well as exogenous factors such as treatment with cytotoxic compounds, alkylating agents and exposure to radiation (**Figure 1.5**). Mismatch repair (MMR) is predominantly involved in the removal of erroneous and small nucleotide insertion/deletion mismatches that are introduced during replication (**Figure 1.6**). Nucleotide excision repair (NER) is involved in the removal of helix-distorting bulky adducts, lesions caused by ultraviolet radiation (**Figure 1.7**), and also participate in ICL repair (detailed in **Section 1.8.1**). Classical non-homologous end joining (C-NHEJ) repair and alternate-EJ (A-EJ) (**Figure 1.8**) are tightly associated with HRR, and to a degree with ICL repair due to the involvement of all of these pathways in rectifying DSBs.

1.8.4 Targeting DNA repair defects in cancer

Defects in DNA repair have been targeted in the treatment of solid malignancies through synthetic lethality (Pearl et al., 2015). A popular model of synthetic lethality was demonstrated in *Brca1*$^{-/-}$ and *Brca2*$^{-/-}$ mouse embryonic stem cells in 2005 (Farmer et al., 2005). *Brca1*$^{-/-}$ and *Brca2*$^{-/-}$ cells were found to be sensitive to PARP1 inhibitors (PARPi) and currently, two models are proposed describing how PARPi selectively kill *BRCA1/2* deficient cells. The first, is the classical model in which PARPi prevent PARP1 from participating in BER, causing an accumulation SSBs that become DSBs when not repaired due to defective HRR. Consequently, *BRCA1/2* deficient cells become reliant on NHEJ, leading to genomic instability and apoptosis (Patel et al., 2011). The second model involves the trapping of PARP1 on the DNA, which prevents other repair factors from binding to the lesion to perform the repair processes and the trapped PARP1 on the DNA itself becoming a larger lesion, leading to genomic instability and

apoptosis (Murai et al., 2012). There are currently multiple clinical trials using synthetic lethality of DNA repair defects through PARPi, in combination with chemotherapeutic compounds, to treat breast and ovarian cancer (Matulonis and Monk, 2017).

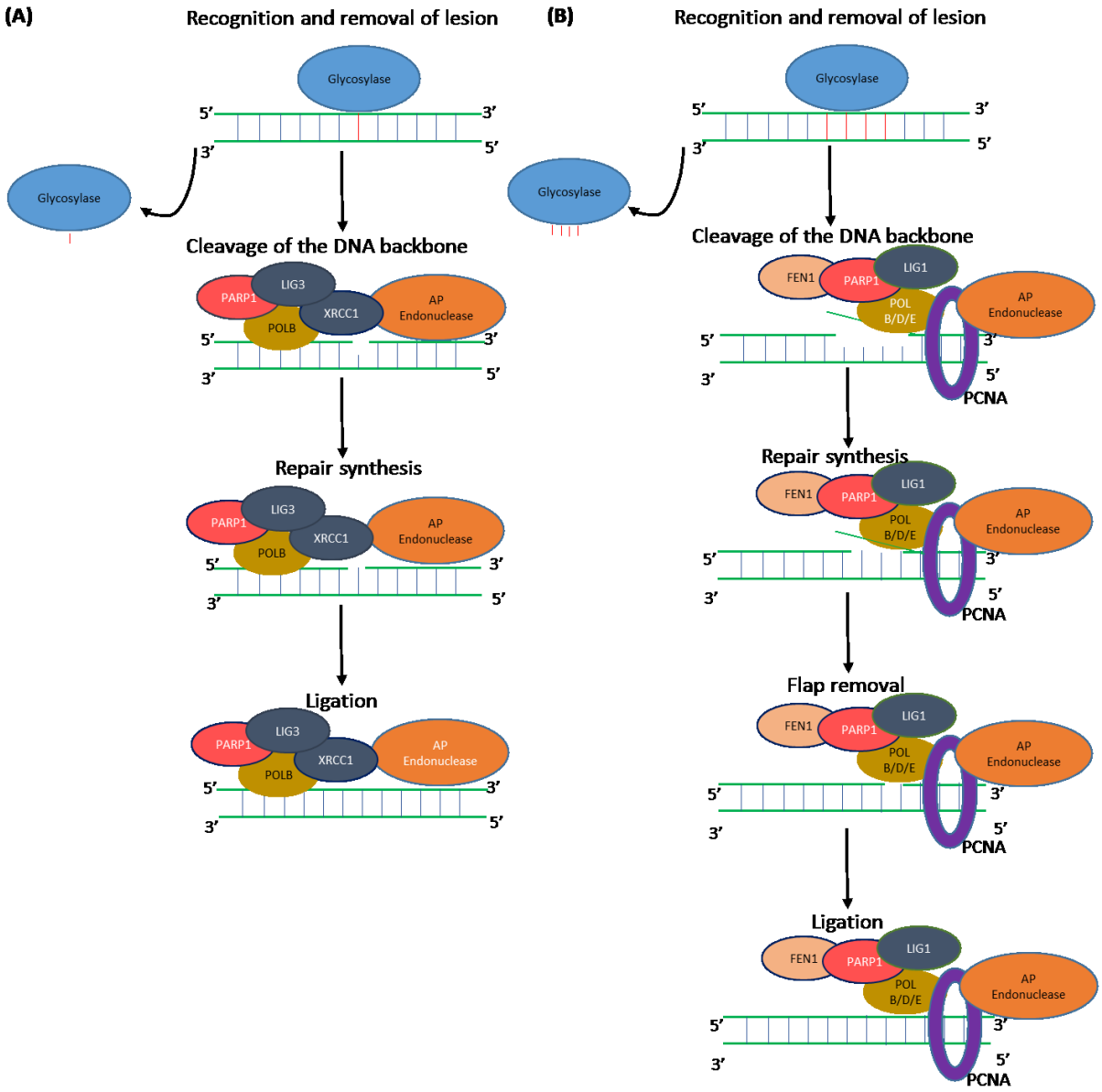


Figure 1.5. Base excision repair (BER). (A) Short-patch BER involved in the repair of single nucleotide error. The lesion is recognised by a glycosylase and the single base is removed, leaving an apurinic/apyrimidinic site (AP site). AP endonuclease removes the DNA backbone. Repair synthesis is performed by POLB, and the strand is ligated by LIG3 and XRCC1. **(B)** Long-patch BER involved in the repair of a stretch of bases. Similar to the short-patch BER, the damaged bases are recognised and removed by glycosylase, additional factors such as PCNA functions as a clamp in this process and FEN1 is used to remove the flap. The repair synthesis is performed by POLB/D/E and the strand is ligated by LIG1.

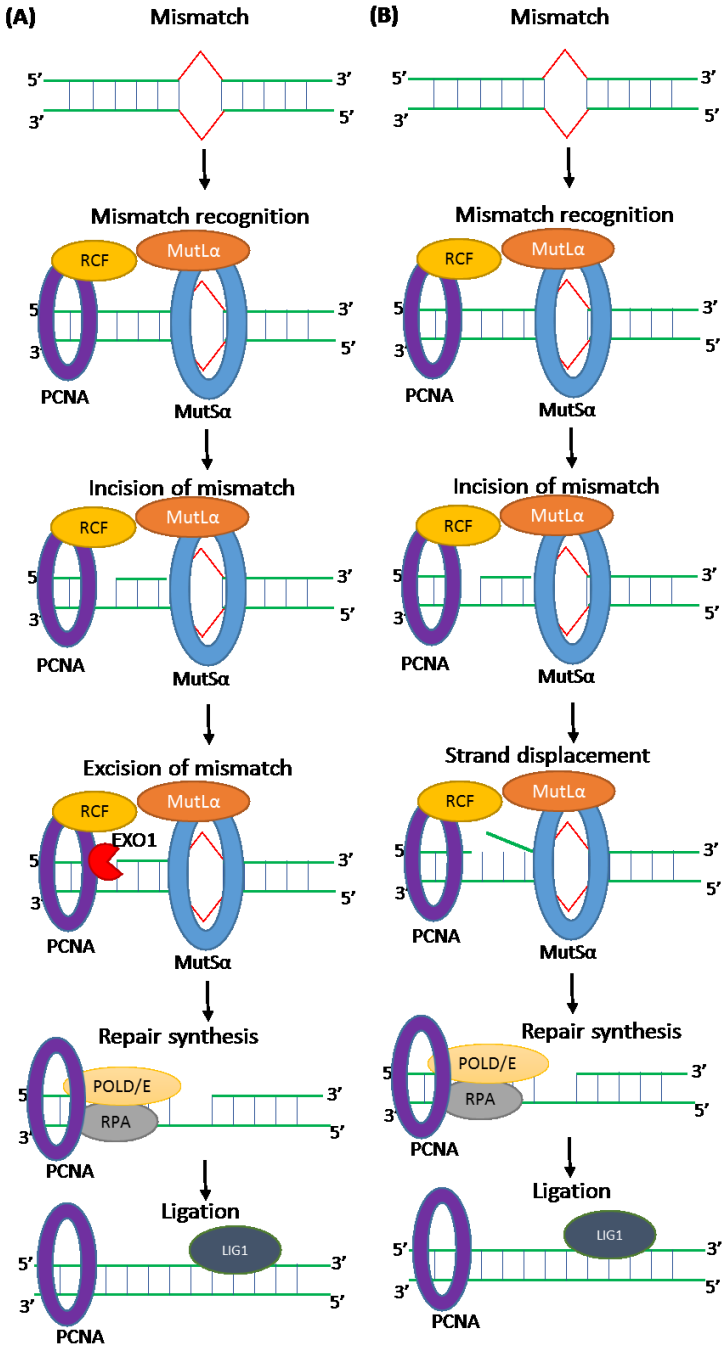


Figure 1.6. Mismatch repair (MMR). **(A)** EXO1-dependent MMR in a MutSα complex and MutLα complex driven pathway during replication. Mutsα complex recognises the mismatch and recruits MutLα complex. PCNA is loaded onto the DNA by RCF and activate MutLα complex. MutLα complex nicks the DNA, allowing EXO1 to bind and excise the strand. POLD/E binds to lesion along with RPA to carry out repair synthesis. The lesion is ligated by LIG1. **(B)** EXO1-independent MMR in a MutSα and MutLα complex driven pathway during replication. Similar to EXO1-dependent MMR but MutLα complex performs multiple nicks and the strand containing the mismatch is displaced and degraded. Repair synthesis and resolution are performed using similar components as in EXO1-dependent MMR.

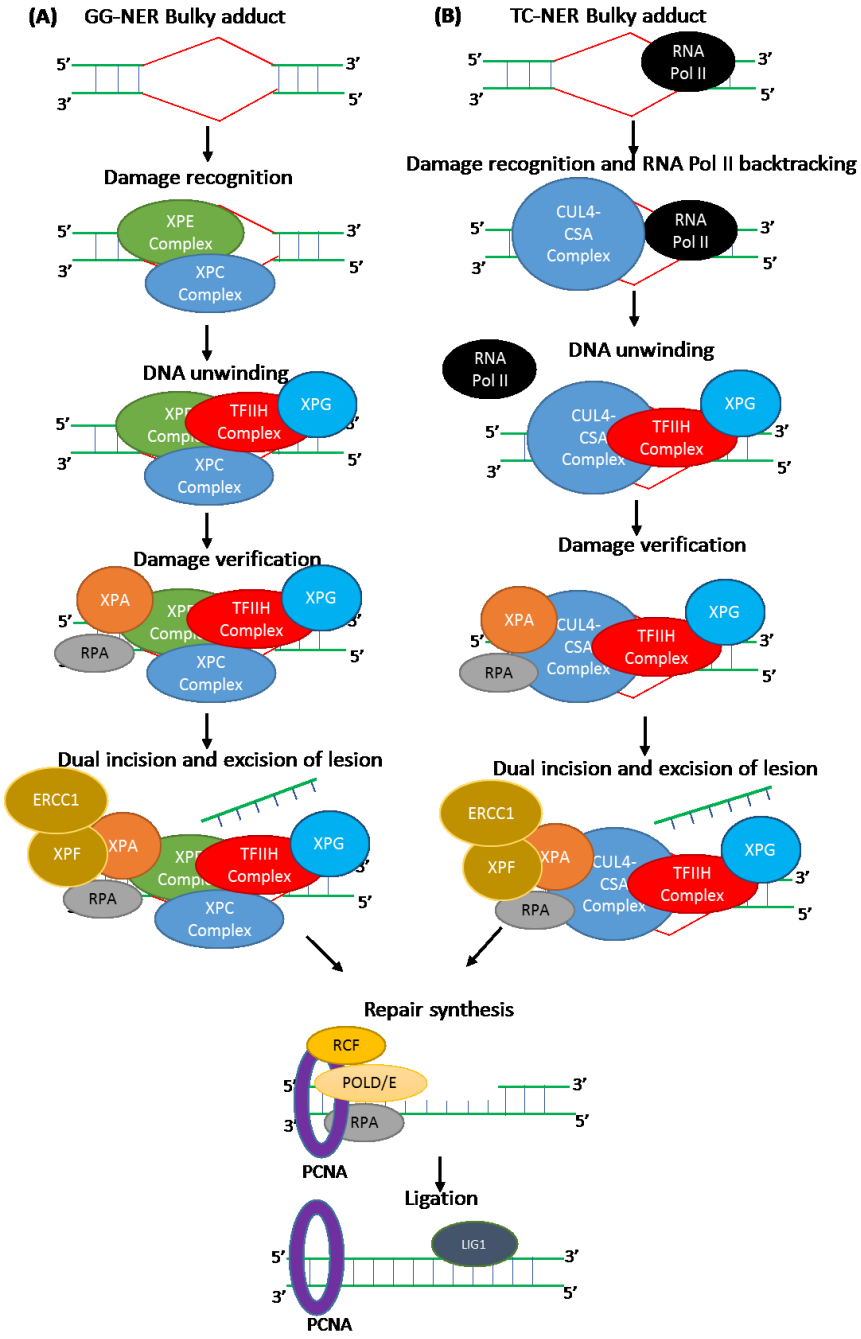


Figure 1.7. Nucleotide excision repair (NER). **(A)** Genomic-NER (GG-NER) lesion recognition involves the XPC complex which scans the genome and binds to the lesion, recruiting XPE complex, TFIIH complex and XPG. The helicase activity from the XPE complex loosens the DNA structure, allowing RPA and XPA to bind to the lesion. The 5' and 3' region of the lesion are nicked by XPA and XPG respectively. POLD/E, along with PCNA, perform the repair synthesis and ligation of the strand is performed by LIG1. **(B)** Transcription complex-NER (TC-NER) recognises the stalled RNA Poll II as the lesion. CUL4-CSA complex binds to the lesion and causes the backtracking and removal of RNA Pol II. Downstream repair processes are identical to GG-NER.

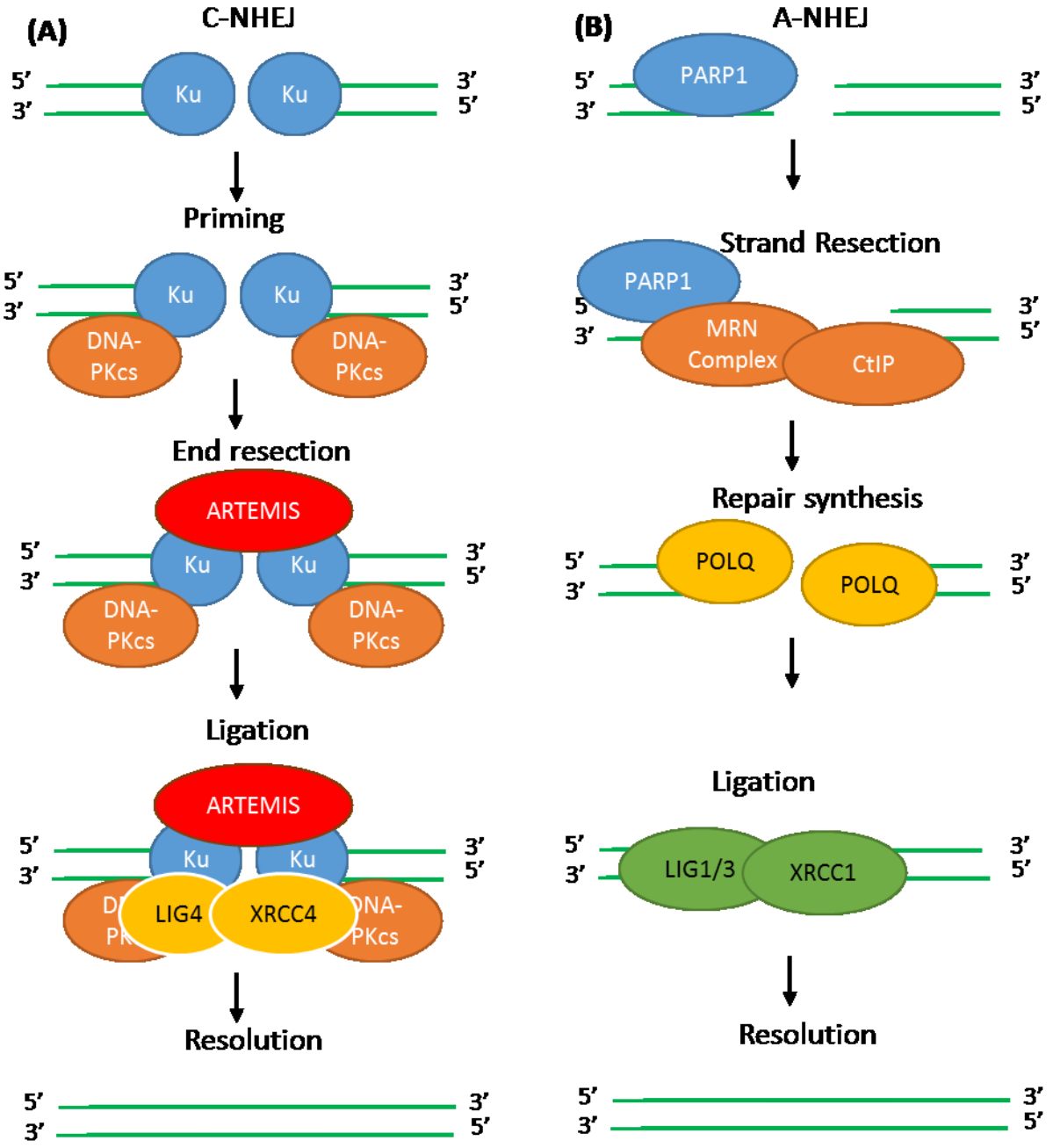


Figure 1.8. Non-homologous end joining (NHEJ) repair. (A) Classical-NHEJ (C-NHEJ) involves the recognition of the DSB by Ku70/80 heterodimers. DNA-PKcs is recruited to the site of lesion, as well as the endonuclease ARTEMIS, which primes the ends and perform small end resection. LIG4 and XRCC4 ligate the DSB break together. **(B)** Alternate-end joining (A-EJ) involves the recognition of the DSB by PARP1 and strand resection is performed by the MRN complex and CtIP to generate microhomology overhangs (20-25 bases). POLQ performs the repair synthesis and LIG1/3 and XRCC1 are involved in the ligation of the strands.

1.9 The research question

1.9.1 What is the role of *FANC* genes in AML pathogenesis?

As described in this chapter, the FA DNA repair pathway is critical for the removal of ICL from DNA, and recent studies have shown a non-redundant tumour suppressor function for this pathway in HSC through maintenance of genomic integrity. FA, caused by biallelic germline deleterious mutations in any of FA-causing genes (FANC genes), is associated with a profoundly increased risk of developing AML. Disease-causing FANC gene variants exist in the population at a frequency of 4.5% (Rogers et al., 2014), and rare damaging variants in certain FANC genes have been observed to be over-represented in cancer cohorts. However to date variants affecting FANC genes have not been extensively analysed in AML.

This project has the following hypotheses:

- Rare heterozygous damaging variants affecting FANC genes, and their interacting partners, will be increased in frequency in AML cohorts, consistent with a role in increasing AML pre-disposition.
- Such rare gene variants lead to impairment of FA DNA repair pathway capacity, increased genomic instability, and increased risk of HSC malignant transformation.

1.9.2 Aims of the project

In order to address these hypotheses, this study has the following Specific Aims:

1. To analyse whole exome sequencing data from a cohort of AML cases to identify rare deleterious, and disease related, gene variants affecting FANC genes, and genes encoding other factors that closely interact with the FA DNA repair pathway
2. To analyse the frequency of such predicted deleterious and disease-causing gene mutations in AML compared to healthy populations.
3. To analyse the gene expression signatures of AML carrying variants in different functional subgroups of the extended FA DNA repair pathway
4. To establish models to test the effect of heterozygous deleterious FANC gene mutations on FA DNA repair pathway activity.

Chapter 2 Materials and Methods

2.1 Whole exome sequencing

Whole exome sequencing (WES) was performed in two separate batches at the Diamentina Institute by A/Prof Leo and his team. The initial sequencing was performed using the Illumina TruSeq Exome Enrichment Kit v2.0 (Illumina) on 101 adult AML samples collected at diagnosis, 5 of which were removed due to multiple failed runs, resulting in 96 exomes. A control cohort consisting of 329 samples from healthy females were sequenced and analysed along with the first batch of AML samples. The second batch of AML samples were sequenced using the Illumina Nextera Rapid (FC-140-1003, Illumina) and consisted of 49 adult AML and 23 paediatric AML samples collected at diagnosis. Both batches of sequencing were performed on the HiSeq 2000 configured for paired-end reads. Details of variant calling and filtering of the exome data are presented in the supplementary methods of the attached manuscripts in **Chapter 4 and 5**.

2.1.1 Ethical approval

Primary AML samples for the WES were obtained from two tissue banks in Australia; the Australian Leukaemia & Lymphoma Group (ALLG) tissue bank at the Princess Alexandra Hospital (PAH, Brisbane, QLD, Australia), and the South Australian Cancer Research Biobank (SACRB) at the Royal Adelaide Hospital (RAH) and SA Pathology (Adelaide, SA, Australia). The use of the samples for this research study was approved by the PAH, the RAH, the University of Adelaide, the University of South Australia and the University of Queensland Human Research Ethics Committees (HREC/04/QPAH/172 & HREC/13/RAH/612). ALLG AML samples were collected with informed consent. SACRB AML patient samples obtained from 1998 onwards were collected with signed informed consent for research purposes. The RAH HREC waived the requirement for informed consent for SACRB specimens collected before 1998.

2.2 Validation of variants identified from the WES

2.2.1 Culture of bone marrow mononuclear cells (BMMNC), mesenchymal stromal cells (MSC) and T-cells of patient samples and genomic DNA extraction

Patient bone marrow samples were thawed in a 37ºC water bath and resuspended in 20mL of resuspension media consisting of Iscove's Modified Dulbecco's Medium (IMDM; Sigma: I3390) with 20% fetal bovine serum (FBS; JRH Biosciences:12003) supplemented with an additional 2mM L-glutamine (Sigma: G7513), 100U/mL penicillin and 100µg/mL streptomycin (Sigma: P4333) and 50U/mL DNAse I (Sigma: DN25-1G). The resuspension media was added dropwise at a steady rate. The samples were spun for 5 minutes, supernatant was discarded and the cell pellets were resuspended in the recovery media consisting of IMDM (Sigma: I3390) with 10% FBS (JRH Biosciences: 12003), supplemented with 2 mM L-glutamine (Sigma: G7513), 100U/mL penicillin and 100µg/mL streptomycin (Sigma: P4333), 10ng/mL hIL-3 (Peprotech: 200-03), 10ng/mL hTPO (Peprotech: 300-18), 50ng/mL hSCF (Peprotech: 300-07), 10ng/mL hFL (Peprotech: 300-19) and 20ng/mL hIL-6 (Peprotech: 200-06). Viability of each sample was determined using 0.4% Trypan Blue (Sigma T8154) and haemocytometer. The samples were recovered incubated in a Sanyo humidified incubator (Sanyo: MCO-20AIC) overnight with 5% CO_2 at 37ºC. After recovery, suspension cells consisting mainly of blast cells were lysed and genomic DNA was extracted using the QIAmp DNA Blood Mini Kit (QIAGEN: 51106) as per manufacturer's recommended protocol.

Mesenchymal stromal cells (MSC) which had adhered to the tissue culture flask after recovery of the patient bone marrow samples, were cultured in Alpha-Modification Minimum Essential Medium Eagle (A-MEM; Sigma: M4526) with 20% FBS (JRH Biosciences: 12003) supplemented with 100µM L-ascorbic acid (Wako: 013-12061), 2mM L-glutamine (Sigma: G7513), 100U/mL penicillin and 100µg/mL streptomycin (Sigma: P4333). Once sufficient numbers of MSC were obtained, cells were detached using 1xTrypsin (Sigma: 59418C), followed by extraction of genomic DNA using the QIAmp DNA Blood Mini Kit (QIAGEN: 51106) as per manufacturer's recommended protocol.

For bone marrow samples that did not grow MSC, an additional vial of BMMNC sample was thawed as previously mentioned and CD3 (T-cell) enrichment was performed. Following overnight recovery, BMMNC were pelleted as described above and resuspended using 80µL of MACS Buffer (Miltenyi Biotec: 130-091-222) and 2.5mL MACS BSA (Miltenyi Biotec: 130-091-376) per 10^7 cells. 20µL of CD3 MicroBeads (Miltenyi Biotec: 130-050-101) was added

per 10^7 cells and incubated at 4°C in the dark for 15-30 minutes. The cells were washed using 2mL of MACS Buffer (Miltenyi Biotec: 130-091-222) per 10^7 cells. The cells were centrifuged for 4 minutes, the supernatant was discarded and the cells resuspended in 500µL of MACS Buffer (Miltenyi Biotec: 130-091-222). The LD column (Miltenyi Biotec: 130-042-901) was placed in the magnetic field of the separator (Miltenyi Biotec Midi MACS separation unit: 130-042-302) and was rinsed with 2mL of MACS buffer (Miltenyi Biotec: 130-091-222). A pre-separation filter (Miltenyi Biotec: 130-041-407) was attached to the top of the LD column (Miltenyi Biotec: 130-042-901) and pre-wet by adding 0.5mL of MACS buffer (Miltenyi Biotec: 130-091-222). The cell suspension was added to the filter and allowed to flow through the LD column and into a collection tube. The LD column was removed from the magnetic field separator and placed over a clean collection tube. To elute the cells, 3mL of MACS buffer was added to the LD column, and the plunger for the LD column was inserted and pushed through. This elution step was repeated. The eluted cells were sorted for CD3 positivity by probing with CD3 antibody (ABCAM: ab16669) and sorted using Flow cytometry (BD FACSAria™ II cell sorter).

2.2.2 Validation of variants with Polymerase Chain Reaction (PCR)

Genomic sequences of the FA/BRCA-HRR network genes were obtained from the USCS genome browser (hg 19; http://genome.ucsc.edu/), and primers were designed using Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) (**Appendix A**).

PCRs were performed using the HotStarTaq DNA Polymerase (QIAGEN: 203203) as per the manufacturer's protocol and DNA extracted from samples as described above.

The PCRs were performed using the Eppendorf Nexus GX2 thermocycler. The cycling conditions were as follow: initial denaturing at 95°C for 5 minutes, followed by 35 cycles of denaturing at 94°C for 1 minute, annealing using the annealing temperatures shown in **Appendix A** for 1 minute, extension at 72°C for 1 minute 30 seconds, and a final extension at 72°C for 10 minutes.

The amplified PCR products were visualised on 2% agarose gels (Amresco: 0710-500G) to ensure a single PCR product was obtained from each PCR reaction. The PCR products were purified using the illustra™ ExoProStar™ kit (GE Healthcare Life Sciences: US78220). The purified PCR products were sequenced at the Southpath and Flinders Sequencing Facility

(Adelaide, Australia). The sequence were visualised using Chromas (v2.6, Technelysium Pty Ltd) and compared to the corresponding reference sequences for each gene to determine the presence of variants observed in the WES data.

2.3 FA DNA repair immunofluorescence assay in cell lines

2.3.1 MCF10A culture conditions

MCF10A (ATCC® CRL-10317™) is a cell line derived from normal breast epithelial cells. The cell line was cultured using Dulbecco's Modified Eagle's medium (DMEM; Sigma: D5546), with 5% horse serum (Sigma: H1138) supplemented with 20ng/mL epidermal growth factor (EGF; R&D Systems: 236-ED-200), 500ng/mL hydrocortisone (Sigma: H0888), 10µg/mL insulin (Sigma: I6634), 500ng/mL Cholera toxin (Sigma: C8052-0.5mg), 100U/mL penicillin and 100µg/mL streptomycin (Sigma: P4333). To detach the cells, multiple washes with 1x phosphate buffered saline (PBS; Sigma: P5493), followed by the addition of 5xTrypsin (Sigma: 59418C) and incubation for 5 minutes in 5% CO_2 at 37ºC incubator, and sterilised disposable cell scrapers were used. The cells were cultured in a Sanyo humidified incubator (Sanyo: MCO-20AIC) with 5% CO_2 at 37ºC.

2.3.2 PD20i and PD20i;RVD2 culture

The cell lines PD20i (Catalogue ID: GM16633) and PD20i;RVD2 (Catalogue ID: GM16634) were obtained from Corriell Institute Depository. The cells were cultured using A-MEM (Sigma: M4526) with 10 % FBS (JRH Biosciences: 12003) supplemented with 2mM L-glutamine (Sigma: G7513), 100 U/mL penicillin and 100µg/mL streptomycin (Sigma: P4333). To detach, 1xTrypsin (Sigma: 59418C) was used. The cells were cultured in a Sanyo humidified incubator (Sanyo: MCO-20AIC) with 5% CO_2 at 37ºC.

2.3.3 Immunofluorescence assay

The immunofluorescence assays (IF) were performed using 8-well chambered slides (Nunc® Lab-Tek® Chamber Slide™ system; Sigma: C7182). Approximately 10^4 cells (PD20i, PD20i;RVD2 or MCF10A) were seeded into each well of the chambered slide and incubated in 5% CO_2 at 37ºC for 48 hours.

The growth media was aspirated from the wells and cells were treated with 40ng/mL of mitomycin C (MMC) (Enzo Lifescience: BML-GR311-0002) for 24 hours in 5% CO_2 at 37^oC.

The media (with or without MMC) was removed from each well. Each well was washed using 500µL of 1x PBS (Sigma: P5493) on a rocking shelf for 5 minutes and then aspirated. The cells were fixed using 200µL of 4% paraformaldehyde (Sigma: P6148) containing 5mM $MgCl_2$ (Sigma: M1028) for exactly 20 minutes and then aspirated. Each well was washed using 500µL of 1x PBS (Sigma: P5493) on a rocking shelf for 5 minutes and then aspirated. The cells were permeabilised using 500µL of 0.3% Triton™ X-100 (Sigma: T8787) in PBS (Sigma: P5493) for exactly 10 minutes and then aspirated. Each well was washed using 500µL of 1x PBS (Sigma: P5493) on a rocking shelf for 5 minutes and then aspirated.

The cells were blocked for unspecific binding using 500µL of blocking buffer consisting of 1x PBS (Sigma: P5493), 10% FBS (JRH Biosciences: 12003) and 0.1% NP40 (US Biosciences: N3500) and incubated at room temperature for 1 hour. The blocking buffer was aspirated from the wells. Primary antibodies for FANCD2 rabbit polyclonal antibody (Novus: NB100-182) and γ-H2AX (MERCK-Millipore: 05-636-I) mouse monoclonal antibody were diluted using the above mentioned blocking buffer (at 1:1000 ratio), added to the wells and incubated for 3 hours on a rocking shelf at room temperature. The primary antibody from each well was aspirated. Each well was washed 3 times using 500µL of 1x PBS (Sigma: P5493) and 2% FBS (JRH Biosciences: 12003) on a rocking shelf for 5 minutes and then aspirated.

The secondary antibodies goat α-mouse Alexa488 (Cell Signalling: 4408) and goat α-rabbit Alexa647 (Cell Signalling: 4414) were diluted using the above mentioned blocking buffer (at 1:1000 ratio), added to the wells and incubated at 4^oC on a rocking shelf overnight. Each well was washed 3 times using 500µL of 1x PBS (Sigma: P5493) and 2% FBS (JRH Biosciences: 12003) on a rocking shelf for 5 minutes and then aspirated. DAPI anti-fade, ProLong™ Gold Antifade Mountant with DAPI (ThermoFisher Scientific: P36935) was added to each well to prevent rapid bleaching of the fluorophores. The cells were imaged using the ZEISS LSM 700 confocal microscope. The images were processed using ZEN Blue (version 2.3) program from ZEISS.

2.4 Generation of the *FANCL* heterozygous mutant cell lines

2.4.1 CRISPR-Cas9 plasmid

The CRISPR-Cas9 plasmid pSpCas9(BB)-2A-Puro (PX459) V2 (**Figure 2.1**) as described by Ran and colleagues (Ran et al., 2013) was kindly provided by Dr Dawei Liu of the Goodall Laboratory (Centre for Cancer Biology, Adelaide, Australia).

2.4.2 Amplification of the pSpCas9(BB)-2A-Puro (PX459) V2 construct

The plasmid was amplified in a bacteria system. All glassware (bottles and conical flasks) were autoclaved prior to use. Luria Broth bacterial powder (Miller) (Sigma: L3522) was used for both bacterial agar and broth. NEB® 10-beta Competent E. coli (High Efficiency) (NEB: NEBC3019) was used for all experiments and the transformations were performed using the recommended manufacturer's protocol. Antibiotic selection was performed using 100µg/mL of ampicillin (Sigma: A0166). The following morning, bacteria clones from the transformed plates were selected and cultured in 5mL of LB broth with appropriate antibiotic and were cultured in a shaking incubator at 37^{o}C for 8 hours. The 5mL cultures were transferred into 50mL cultures in 250mL conical flasks with appropriate antibiotic and cultured in a shaking incubator at 37^{o}C overnight. Midiprep was performed using QIAGEN Plasmid Midi Kit (QIAGEN: 12145) as per the manufacturer's recommended protocol.

2.4.3 Restriction enzyme digestion and preparation of the pSpCas9(BB)-2A-Puro(PX459) V2 construct

Restriction endonuclease digestion was performed using BbsI (NEB: R0539S) on 6µg of the purified plasmid and incubated for 4 hours at 37^{o}C. To dephosphorylate the 5' end of the construct, 1µL of Calf Intestinal Alkaline Phosphatase (CIP) (NEB: M0290S) was added to the digestion reaction and incubated for a further 1 hour at 37^{o}C. The digested and dephosphorylated construct was loaded into a 1% agarose gel (Amresco: 0710-500G), and ran at 30V overnight for gel extraction purification (PureLink™ Quick Gel Extraction Kit (ThermoFisher Scientific: K210012)). Ethanol purification using sodium acetate was performed to concentrate the purified construct.

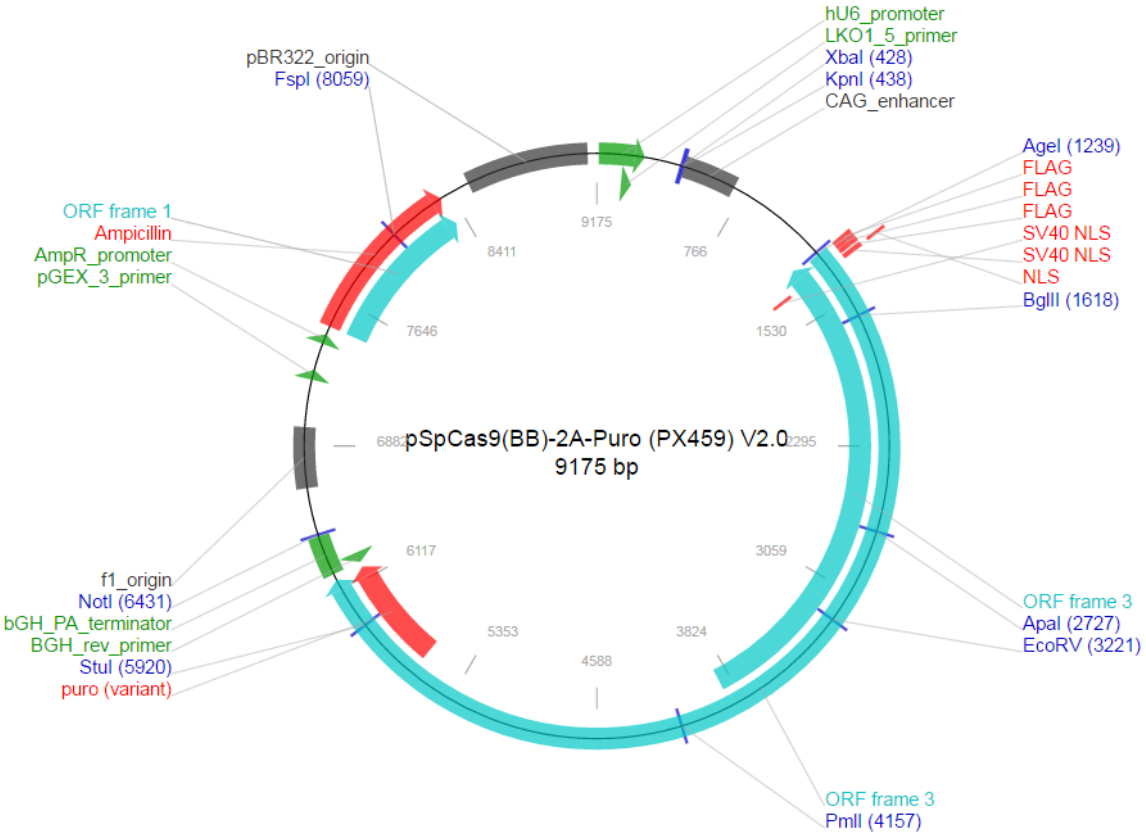


Figure 2.1 pSpCas9(BB)-2A-Puro (PX459) V2 vector map. (taken from Addgene: https://www.addgene.org/62988/)

2.4.4 Designing single guide RNA (sgRNA) for the pSpCas9(BB)-2A-Puro (PX459) V2 construct

The sgRNA specific for exon 1 of FANCL (Ref Seq: NM_018062) was generated using thehttp://crispr.mit.edu/ online tool. The top 3 pairs of sgRNA sequences with a guide score greater than 90 were selected to minimise off-target gene editing. The 3 pairs of sgRNA sequences were additionally modified to be able to be cloned into the CRISPR-Cas9 construct (**Table 2.1**). For the 5' ends of the sgRNA sequences of the forward oligo of the sgRNA, an additional sequence of CACC was added for cloning using the BbsI restriction site in the construct. An additional nucleotide G was also added to the 5' end of the sgRNA if the sequence did not began with a G to ensure efficient transcription of the sgRNA by RNA Polymerase III. Similarly for the reverse oligo, an additional sequence of AAAC was added to the 5' region and an additional nucleotide C was also added to the 3' region of the sgRNA.

2.4.5 Annealing the forward and reverse oligos of the 3 pairs of sgRNA

The T4 polynucleotide kinase (NEB: M0201S) was used as per the manufacturer's protocol to phosphorylate the forward and reverse strands of the 3 pairs of sgRNA followed by the annealing. The phosphorylation and annealing processes were performed using the Eppendorf Nexus GX2 thermocycler using the following cycling conditions. One Cycle of 37^{o}C for 30 minutes, 94^{o}C for 10 minutes, and ramp down to 25^{o}C at a rate of 1%.

2.4.6 Cloning the sgRNA sequences into the pSpCas9(BB)-2A-Puro (PX459) V2 construct

Ligation reaction was performed using T4 DNA Ligase (NEB: M0202) consisting of 15ng of vector and 2µL of annealed sgRNA, according to the manufacturer's protocol in a final volume of 15µL with H_2O for a total of 3 ligation reactions (1 for each sgRNA). The ligation reactions were incubated at 16^{o}C overnight in the Eppendorf Nexus GX2 thermocycler. Each ligation reaction was transformed into the above mentioned NEB® 10-beta Competent E. coli (High Efficiency) (NEB: NEBC3019) as per the manufacturer's protocol. Ten colonies from each plate were picked and cultured in 5mL of LB broth and ampicillin. Miniprep was performed using QIAprep Spin Miniprep Kit (QIAGEN: 27104) following the manufacturer's

Table 2.1 Single guide RNA (sgRNA) targeting exon 1 of FANCL.

Primer	Sequence	Target	Ref Seq
FANCL sgRNA 1 FW3	CACC G ACACGGTTTTCGACCGGTTC	Exon 1	NM_018062
FANCL sgRNA 1 RV3	AAAC GAACCGGTCGAAAACCGTGT C	Exon 1	NM_018062
FANCL sgRNA 2 FW3	CACC G CACGGTTTTCGACCGGTTCT	Exon 1	NM_018062
FANCL sgRNA 2 RV3	AAAC AGAACCGGTCGAAAACCGTG C	Exon 1	NM_018062
FANCL sgRNA 3 FW3	CACC G CCGGTCGAAAACCGTGTATG	Exon 1	NM_018062
FANCL sgRNA 3 RV3	AAAC CATACACGGTTTTCGACCGG C	Exon 1	NM_018062

Red font represents the additional sequences added to the sgRNA as restriction sites for BsbI. Blue font represents the additional base added to ensure efficient transcription of the sgRNA by RNA Polymerase III.

recommended protocol. The purified constructs were sequenced at the Southpath and Flinders Sequencing Facility (Adelaide, Australia) using the primers LK0 1.5 (GACTATCATATGCTTACCGT) and pSPCas9 RV (CACGCGCTAAAAACGGACTA). The constructs containing their respective sgRNA were transformed into the above mentioned NEB® 10-beta Competent E. coli (High Efficiency) (NEB: NEBC3019) as per the manufacturer's protocol. Antibiotic selection was performed using 100µg/mL of ampicillin (Sigma A0166). Bacteria clones from the transformed plates were selected and cultured in 5mL of LB broth with appropriate antibiotic the following morning and were cultured in a shaking incubator at 37^{o}C for 8 hours. The 5mL cultures were transferred into 50mL cultures in 250mL conical flasks with appropriate antibiotic and cultured in a shaking incubator at 37^{o}C overnight. Midiprep was performed using QIAGEN Plasmid Midi Kit (QIAGEN: 12145) as per the manufacturer's recommended protocol.

2.4.7 Transfection of MCF10A with the pSpCas9(BB)-2A-Puro (PX459) V2 construct

The MCF10A cell line was cultured using the media and conditions as described in **Section 2.3.1**. Cells were grown to 80% confluency and 10^6 cells from the culture were seeded into a 10cm petri dish. The cells were transfected with a pSpCas9(BB)-2A-Puro (PX459) V2 construct containing one sgRNA two days later. The transfection reagent used was Lipofectamine 3000 (ThermoFisher Scientific: L3000015) as per the manufacturer's protocol. Fresh media containing 2µg/mL of puromycin (ThermoFisher Scientific: A1113802) was added to the plate 24 hours after transfection. The cells were selected using puromycin for 48 hours. Fresh media was added to the cells (without puromycin) and left to recover for 7 days. The cells were detached and serial dilution was performed to obtain cell concentrations of 0.5 cells/well, 1.0 cells/well and 2.0 cells/well. Each dilution was seeded into two 96-well plates each. The clones were expanded and scaled up to 24-well and eventually 6-well plates upon reaching confluency. Each individual clone was frozen and gDNA was also made using QIAmp DNA Blood Mini Kit (QIAGEN: 51106) as per manufacturers recommended protocol.

2.5 Gene expression analysis

Gene expression profiling was performed using the Ilumina_HumanHT_12_v4 chip, (47247 probes) for a total of 57 samples. RNA was extracted from BMMNC at diagnosis (n=57). The

$CD34^+$ cell fraction from healthy control BMMNC (n=13) was used as control. The data was background corrected using negative control probes. Probes with sequences considered to be poor matches for the corresponding target or with no matches were removed. Detection Above Background (DABG) was also performed and probes with detection p-values >0.05 in more than half of the samples were removed leaving a total of 15484 probes for down-stream analysis For baseline expression levels, 9 healthy $CD34^+$ cell populations were used. The initial analysis and data processing was performed with the help of Stephen Pederson and Mahmoud Bassal. Detailed methods are shown in the attached manuscripts in **Chapters 4** and **5**.

Chapter 3 Defects in the FA/BRCA-HRR pathway and the impact on AML

3.1 Introduction

As described in **Chapter 1** (**Section 1.7.2.1**), bi-allelic loss of any of the 19 FANC genes in the Fanconi Anaemia (FA) DNA repair pathway is known to cause FA, an autosomal recessive disorder with progressive bone marrow failure, and a profoundly increased risk of AML (800 fold) (Rosenberg et al., 2008). FA pathogenicity is due to haematopoietic stem cell (HSC) dysfunction, which begins *in utero* and manifests as bone marrow failure in adulthood (Ceccaldi et al., 2012, Pontel et al., 2015). Its mode of pathogenicity centres on HSC dysfunction associated with increased sensitivity to endogenous/exogenous DNA toxins, replicative stress or repeated infections (Garaycoechea et al., 2012, Cherubini et al., 2011, Walter et al., 2015, Kaschutnig et al., 2015). This establishes the link between maintenance of HSC genomic integrity and suppression of MDS/AML. Previous studies in AML have shown that loss of function in FANC genes can occur through epigenetic changes such as promoter methylation, or large exonic deletions occurs (Hess et al., 2008, Tischkowitz et al., 2004), potentially contributing to genomic instability in AML. IDH1/2 mutations have recently been shown to impair homologous recombination repair (HRR) resulting in increased sensitivity to PARP1 inhibitors (PARPi) (Sulkowski et al., 2017).

However, the contribution of heterozygous germline FANC gene variants to AML initiation has not been clearly established. Some FANC genes are classical tumour suppressor genes associated with very high risk of solid cancers (*BRCA1* and *BRCA2*) (Petrucelli et al., 2010). In recent pan-cancer studies using large cohorts of more common cancers, deleterious mutations in a number of FANC genes were observed to be enriched (described in **Chapter 1; Section 1.7.2.1.3**), suggesting potential roles of FANC mutations in cancer predisposition. The cancer risk associated with heterozygous FANC mutations across individual FANC genes has been difficult to assess. In familial studies in the early 2000s, increased cancer incidence was not observed for FA carriers, but grandmothers of the probands had increased breast cancer risk, suggesting that there may potentially be risk for breast cancer with advanced age (Berwick et al., 2007, Tischkowitz et al., 2008). These studies are necessarily biased to the more common subtypes of FA, namely FA-A, FA-C and FA-G making up approximately 85% of FA cases (Dong et al., 2015). Furthermore, FA carriers occur at low frequency in the population (estimated at 0.5%) (Rosenberg et al., 2011) compared to the reported frequency of individuals

with heterozygous damaging variants across all FANC genes (reported at a frequency of 4.5% in European-Americans) (Rogers et al., 2014). This suggests that for most of the genes reported to cause FA, only rarely are mutations identified in the bi-allelic state, possibly due to lethality. However, given the frequency of heterozygous damaging variants in the population, and the profound increase in risk of AML for FA patients with severe FA pathway deficiency, variants in this group of genes could contribute significantly to the AML patient population.

The 19 FANC genes participate in the FA DNA repair pathway which is critical for repair of interstrand crosslinks (ICL), and operates in conjunction with HRR to resolve stalled replication forks that result from ICLs induced by ionization radiation, aldehydes and reactive oxygen species (ROS) (described in **Chapter 1**). These genes interact with other DNA damage sensors and repair pathways. The aim of the work presented in this chapter was to investigate rare and predicted deleterious variants in the FANC genes with their key interacting partners, and the co-occurrence or exclusivity with common AML mutations and karyotypic abnormalities in AML. As such, the results in this chapter describe the generation of a network of functionally related proteins built from the FANC genes, and the identification and characterisation of damaging variants across this network in adult AML and healthy control cohorts.

3.2 Results

3.2.1 Extended FA/BRCA-HRR network

As a first step to investigate the involvement of the FA DNA repair pathway in AML, an extended network of genes was constructed with the FANC genes as the seed network using the protein/gene interaction predictor STRINGdb v9.05 (Szklarczyk et al., 2011). This extended network is referred to as the extended FA/BRCA-HRR network. At the time of construction (2013), only 16 genes were officially associated with FA, and hence were used as the seed network. As of 2017, the total number of FANC genes has continued to increase to 21 with *FANCR* (*RAD51*), *FANCS* (*BRCA1*), *FANCT* (*UBE2T*), *FANCU* (*XRCC2*) and *FANCV* (*REV7*) recently associated with FA (Ameziane et al., 2015, Sawyer et al., 2015, Hira et al., 2015, Park et al., 2016, Bluteau et al., 2016). The results presented in this study were done with respect to 19 FANC genes (excluding the more recently classified *FANCU* and *FANCV*).

A two-step approach was utilised in constructing the FA/BRCA-HRR network. The first step utilised STRINGdb to identify interacting partners of each of the 16 FANC proteins. The second

step utilised the ability of STRINGdb to generate interactions within a list of query proteins. Thus, each of the 16 FANC proteins (query) were individually searched in STRINGdb and the top 50 proteins that interacted with each of the 16 FANC proteins with confidence level of 0.9 (90%) based on experimental evidence, knowledge-based and text mining evidences were compiled into a single list (**Appendix B**). This list was then used as the query file that was input into STRINGdb providing the interactions of high confidence within the query list as output. The output from STRINGdb was then further manually clustered to show the functional subgroups with high confidence interactions based on literature evidences, as detailed in **Table 3.1** and **Figure 3.1**. Thus in this FA/BRCA-HRR network, all the proteins have a direct relationship, based on experimental and literature evidence, with the 16 FANC proteins. Importantly, this analysis highlighted several interactions of a number of FANC proteins beyond the canonical ICL repair pathway. Each of the various functional subgroups is explained and summarised below.

3.2.1.1 FANC core complex

The genes of the classical FANC core complex as described in **Chapter 1** (**Section 1.8.1.1**), were grouped together as they are directly involved in the repair of ICLs and function to activate the heterodimeric ID2 complex, which is considered to be the master regulator of the FA pathway. This consisted of the eight FANC proteins (**Table 3.1**), three FA associated proteins (FAAP), which function to stabilise the complex, and two additional proteins, MHF1 and MHF2. The FANC core complex can be further subdivided into the FANCM anchor complex (Walden and Deans, 2014) and the minimal ubiquitination complex (Rajendra et al., 2014). The FANCM anchor complex consists of FANCM, MHF1, MHF2 and FAAP100. They have been shown to form a complex at the site of ICL to recruit the remaining members of the core complex (Collis et al., 2008). In a separate study, the minimal ubiquitination complex has been shown to be required for the ubiquitination of the ID2 complex (Hodson and Walden, 2012). The FANC core complex and ID2 complex (described below in **Section 3.2.1.2**) are referred to as a combined subgroup due to their direct interactions and cooperation in ICL repair.

Table 3.1 Functional subgroups and combined subgroups of the extended FA/BRCA-HRR network

Functional Subgroup	**Genes**
FANC core complex	*FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM, MHF1 (APITD1), MHF2 (STRA13), FAAP20 (C1orf86), FAAP24 (C19orf40), FAAP100 (C17ORF70)*
FANCM anchor complex	*FANCM ,MHF1 (APITD1), MHF2 (STRA13), FAAP24 (C19orf40)*
Minimal FANCD2 monoubiquitination complex	*FANCB, FANCL, FAAP100 (C17ORF70)*
ID2 Complex	*FANCD2, FANCI*
FANC core & ID2 complex	*FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM, MHF1 (APITD1), MHF2 (STRA13), FAAP20 (C1orf86), FAAP24 (C19orf40), FAAP100 (C17ORF70), FANCI, FANCD2*
Structure-specific endonucleases	*FANCP (SLX4), FANCQ (ERCC4), SLX1A, MUS81, ERCC1, FAN1, EME1, EME2*
BLM complex	*RMI1, RMI2, BLM, TOP3A,*
ATM/ATR checkpoint proteins	*RPA1, RPA2, RPA3, ATM, ATR, ATRIP, CHEK1 ,RAD9A, RAD17, CHEK2*
RPA proteins	*RPA1, RPA2, RPA3*
MRN complex	*MRE11A, NBN, RAD50*
Breast cancer associated proteins	*FANCD1 (BRCA2), FANCJ (BRIP1), FANCN (PALB2), BRCA1, BARD1*
RAD51 paralogues	*FANCO (RAD51C), RAD51, RAD51B, RAD51D, XRCC2, XRCC3*
Mismatch Repair (MMR) Genes	*MLH1, MLH3*
Ubiquitination modifiers	*UAF1, USP1, UBA52, UBE2T, UBC*
Combined Subgroup	**Genes**
FA/BRCA-HRR Network	*FANCA, FANCB, FANCC, FANCD1 (BRCA2), FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ (BRIP1), FANCL, FANCM, FANCN (PALB2), FANCO (RAD51C), FANCP (SLX4), FANCQ (ERCC4), FANCR (RAD51), FANCS (BRCA1), FANCT (UBE2T) MHF1 (APITD1), MHF2 (STRA13), FAAP20 (C1orf86), FAAP24 (C19orf40), FAAP100 (C17ORF70), SLX1A, MUS81, ERCC1, FAN1, EME1, EME2, MRE11A, NBN, RAD50, BRCA2, BRCA1, BARD1, RAD51, RAD51B, RAD51D, XRCC2, XRCC3, RMI1, RMI2, BLM, TOP3A, RPA1, RPA2, RPA3, ATM, ATR, ATRIP, CHEK1, RAD9A, RAD17, CHEK2, MLH1, MLH3*
BLM & ATM/ATR checkpoint	*RMI1, RMI2, BLM, TOP3A, RPA1, RPA2, RPA3, RPA1, RPA2, RPA3, ATM, ATR, ATRIP, CHEK1 ,RAD9A, RAD17, CHEK2*
KEGG Homologous Recombination Repair (HRR) Pathway	*MRE11A, NBN, RAD50, BRCA2, BRIP1, FANCN (PALB2), BRCA1, BARD1, FANCO (RAD51C), RAD51, RAD51B, RAD51D, XRCC2, XRCC3*
FANC 19	*FANCA, FANCB, FANCC, FANCD1 (BRCA2), FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ (BRIP1), FANCL, FANCM, FANCN (PALB2), FANCO (RAD51C), FANCP (SLX4), FANCQ (ERCC4), FANCR (RAD51), FANCS (BRCA1), FANCT (UBE2T)*

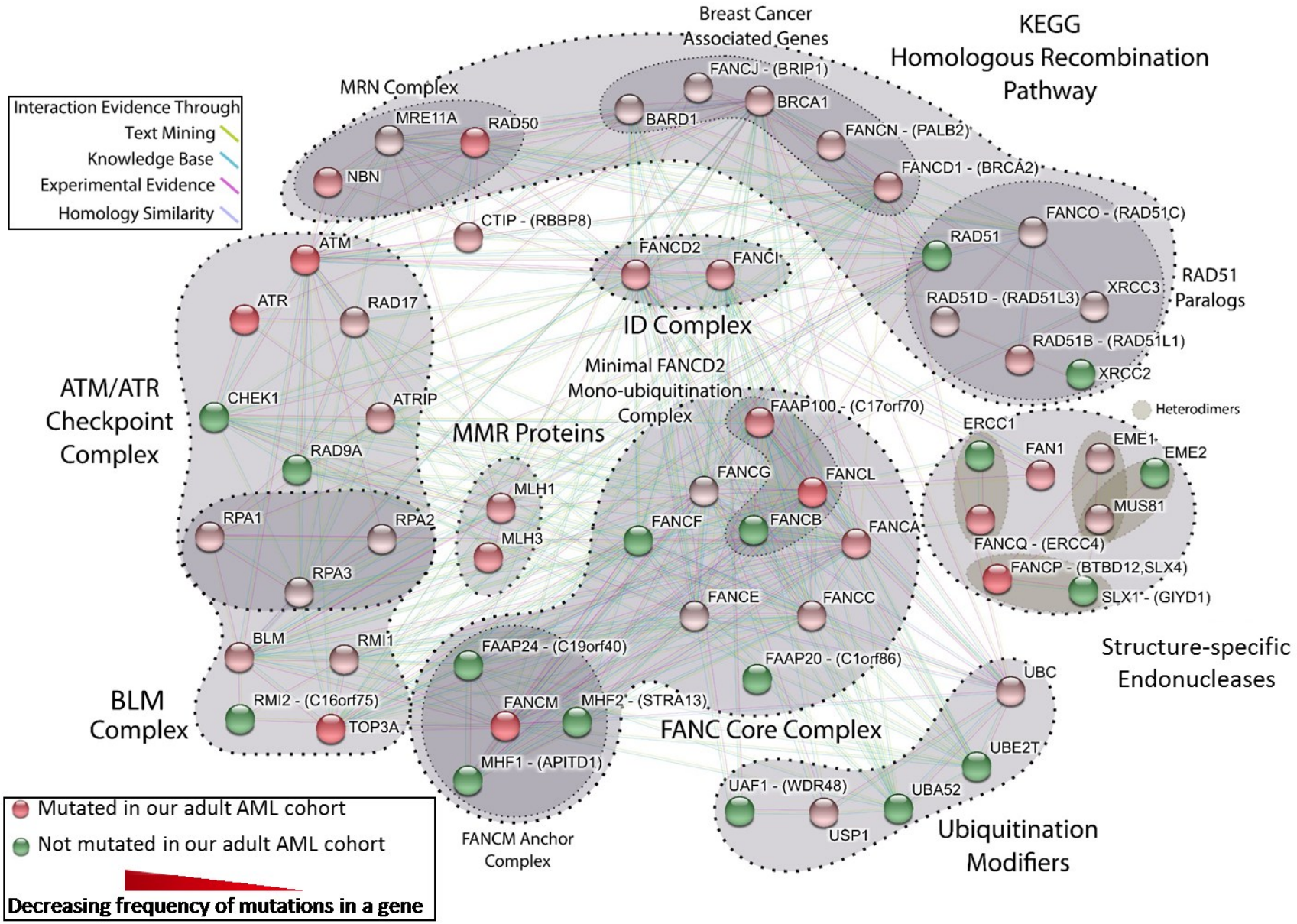

Interaction Evidence Through
Text Mining
Knowledge Base
Experimental Evidence
Homology Similarity
Breast Cancer Associated Genes
KEGG Homologous Recombination Pathway
MRN Complex
MRE11A
RAD50
NBN
FANCJ - (BRIP1)
BRCA1
BARD1
FANCN - (PALB2)
FANCD1 - (BRCA2)
CTIP - (RBBP8)
ATM
ATR
RAD17
FANCD2
FANCI
ID Complex
FANCO - (RAD51C)
RAD51
RAD51D - (RAD51L3)
XRCC3
RAD51 Paralogs
RAD51B - (RAD51L1)
XRCC2
ATM/ATR Checkpoint Complex
CHEK1
ATRIP
RAD9A
Minimal FANCD2 Mono-ubiquitination Complex
MMR Proteins
FAAP100 - (C17orf70)
Heterodimers
ERCC1
FAN1
EME1
EME2
MUS81
FANCQ - (ERCC4)
FANCP - (BTBD12,SLX4)
SLX1 - (GIYD1)
MLH1
MLH3
FANCG
FANCL
FANCF
FANCB
FANCA
RPA1
RPA2
RPA3
FANCE
FANCC
Structure-specific Endonucleases
BLM
RMI1
RMI2 - (C16orf75)
TOP3A
BLM Complex
FAAP24 - (C19orf40)
FAAP20 - (C1orf86)
FANCM
MHF2 - (STRA13)
MHF1 - (APITD1)
FANC Core Complex
UBC
UBE2T
UAF1 - (WDR48)
UBA52
USP1
Ubiquitination Modifiers
FANCM Anchor Complex
Mutated in our adult AML cohort
Not mutated in our adult AML cohort
Decreasing frequency of mutations in a gene

Figure 3.1. The extended FA/BRCA-HRR network generated using STRINGdb (v9.05). Proteins were clustered based on functional groups and involvement in the various pathways. Colour of the lines connecting each circle were based on the types of interaction evidences as shown in the figure legend. Green circles represent genes that have no alternate variants identified in the AML whole exome sequencing (WES) dataset, while red circles represent the genes that have one or more variants. The intensity of the colour signifies the total number of variants identified for each gene (darker shades representing genes with the most number of variants and lighter shades representing genes with fewer variants).

3.2.1.2 ID2 complex and structure-specific endonucleases

The FANCI/D2 heterodimer is recruited to sites of ICLs and stalled replication forks along with the FANC core complex (Joo et al., 2011). It is activated when phosphorylated by ATR (Matsuoka et al., 2007) and monoubiquitinated by the E3 ligase FANCL, of the FANC core complex as described in **Chapter** 1 (**Section 1.8.1.1**). The ID2 complex functions as a scaffold for the recruitment of the remaining repair factors, as well as in stabilisation and protection of stalled replication forks (Schlacher et al., 2012). Ubiquitin molecules of the ID2 complex are recognised by the ubiquitin-binding zinc-finger 4 (UBZ4) domain of the nuclease scaffold protein, SLX4 (FANCP) (Yamamoto et al., 2011). SLX4 then recruits a series of endonucleases and interacting partners to perform the incision of the crosslink. Such endonucleases include heterodimeric MUS81-EME1 and ERCC1-ERCC4 (FANCQ), as well as SLX1-SLX4 (Klein Douwel et al., 2014, Kottemann and Smogorzewska, 2013). The structure-specific nuclease Fanconi Anaemia nuclease 1 (FAN1) is also recruited to the ICL (Pizzolato et al., 2015), and it has recently been shown to be recruited by FANCD2 to resume replication and not necessarily for the resolution of ICL (Lachaud et al., 2016) . These structure specific-endonucleases participate in the "unhinging" of crosslinks, causing a double-stranded break (DSB) to occur.

3.2.1.3 BLM complex

In addition to its canonical function, the FANC core complex has been shown to interact with the BLM complex and is required for activation of the BLM pathway (Pichierri et al., 2004). The BLM complex consists of BLM and its interacting partners TOP3A, RMI1 and RMI2. The BLM complex is canonically known to be involved in the resolution of Holliday junctions during HRR. BLM mutations result in the development of Bloom syndrome (BS) described in **Chapter 1** (**Section 1.7.2.5**). The BLM complex also interacts with the replication proteins RPA1/2/3 (Xue et al., 2013). Consistent with BLM complex role, cells from BS patients have significantly increased frequency in sister chromatid exchange (McDaniel and Schultz, 1992).

3.2.1.4 MRN complex

The MRN complex, consisting of MRE11A, NBN and RAD50, is involved in the initial detection of DNA damage resulting from DSB (Lamarche et al., 2010). The MRN complex is one of the first factors recruited to the site of DNA damage and is upstream of ATM/ATR

kinases. This complex is essential for the rapid recruitment of the kinase ATM to the site of damage (Lee and Paull, 2005). It is also involved in the activation of the kinase ATR through DNA end resection (Lee and Dunphy, 2013).

3.2.1.5 ATM/ATR Checkpoint proteins

This functional subgroup consists of proteins which interact with the ATM and ATR kinases. There are 3 main kinases, ATM, ATR and DNA-PKc which are involved in signalling and phosphorylation of cell cycle checkpoint and DNA repair proteins (Marechal and Zou, 2013). DNA-PKc is primarily involved in activation of non-homologous recombination (NHEJ) repair (Davis et al., 2014); while ATM is primarily activated by DSB and is involved in HRR (Bensimon et al., 2010). ATR on the other hand is activated from a variety of damages including single-stranded breaks (SSB), DSB, as well as DNA lesions that interfere with replication (Bensimon et al., 2010). Despite being characterised as responding to different types of DNA damage, evidence has mounted suggesting cross talk between the two kinases (ATM and ATR) (Cuadrado et al., 2006).

ATM has been shown to exist as either heterodimers or oligomers. Upon recruitment by the above mentioned MRN complex (Uziel et al., 2003) and autophosphorylation, it dissociates and becomes a monomer. ATM activation promotes ATR activation by enhancing DNA end resection by a number of repair factors to form single-stranded DNA (ssDNA) as described in **Chapter 1** (**Section 1.8.2**) (Shiotani and Zou, 2009). ATR and its heterodimeric partner ATRIP (Cortez et al., 2001) bind to RPA bound ssDNA (Zou and Elledge, 2003). Activation of ATR can occur from phosphorylation dependent protein-protein interactions involving ATR-TOPB1, TOPBP1-RAD9A and RAD17-CLSPN (Marechal and Zou, 2013). The best classified effectors of ATM and ATR are the CHEK2 and CHEK1 kinases respectively (Marechal and Zou, 2013). Both ATM and ATR have also been shown to be involved in the recruitment of cell cycle regulators and DNA repair factors such as BRCA1 (Deng, 2006) and TP53 (Saito et al., 2002, Tibbetts et al., 1999).

Mutations in *ATM* cause ataxia telangiectasia (A-T), which is a neurodegenerative disorder with increased predisposition to various types of cancers as described in **Chapter 1** (**Section 1.7.2.2**). Mutations in *ATR* on the other hand lead to Seckel syndrome (SS) which presents with growth retardation and intellectual disabilities (Casper et al., 2004, Harsha Vardhan et al.,

2007). Clinically, both A-T and SS patients show some overlapping phenotypic characteristics (described in **Chapter 1**; **Section 1.7.2.3**).

3.2.1.6 Breast cancer associated proteins and RAD51 paralogues

The breast cancer associated proteins BRCA1 (FANCS) and BRCA2 (FANCD1) are classical tumour suppressors, mutations in which cause familial cancer with high penetrance. Germline mutations in these genes greatly increase the overall risk of breast and ovarian cancer in females, as well as prostate and breast cancer in males (Stan et al., 2013). *BRCA2* was one of the first genes to be cloned and classified as a FANC gene (Howlett et al., 2002). *BRCA1* has also recently been classified as a FANC gene but with a single case being reported, the validity of the classification is yet to be determined (Sawyer et al., 2015). *Brca1/2* null mice die at the embryonic stage, when both alleles are lost and this may be a reason why bi-allelic germline mutations in these genes in humans are observed only rarely (Evers and Jonkers, 2006). PALB2 (FANCN) is involved in the nuclear localisation and stability of BRCA2 (Xia et al., 2006), and *PALB2* mutations have been shown to cause familial breast cancer (Antoniou et al., 2014). BRIP1 (FANCJ) interacts with the C-Terminal domain of BRCA1 (Cantor et al., 2001), and mutations in *BRIP1* have been shown to increase risk of familial ovarian cancer (Rafnar et al., 2011). BRCA1-associated RING domain protein 1 (BARD1) is the heterodimeric partner of BRCA1. The BRCA1-BARD1 complex becomes stabilised upon heterodimerisation and functions mainly as a ubiquitin ligase (Irminger-Finger and Jefford, 2006). BARD1 has also been reported to be associated with familial breast and ovarian cancer, although with lower penetrance (Ratajska et al., 2012). These breast cancer associated proteins are central to HRR as described in **Chapter 1** (**Section 1.8.2**).

As previously described in **Chapter 1** (**Section 1.8.2**), the RAD51 recombinase and its paralogues (RAD51A, RAD51B, RAD51C, XRCC2 and XRCC3), play a pivotal role in HRR by generating the RAD51 filaments which are required for homology search and strand invasion (Suwaki et al., 2011, Godin et al., 2016). The successful recruitment and assembly of RAD51 to the site of damage is mediated by the RAD51 paralogues, as well as BRCA1 and BRCA2. *RAD51* (*FANCR*) has also been recently identified as a FANC gene, along with *RAD51C* (*FANCO*) and *XRCC2* (*FANCU*) (Ameziane et al., 2015, Sawyer et al., 2015, Hira et al., 2015, Park et al., 2016, Bluteau et al., 2016, Vaz et al., 2010). It is becoming increasingly evident that RAD51 and its paralogues are critical to HRR, hence it is not surprising that their loss of

function can result in DNA repair disorders such as FA. The RAD51 paralogues and BRCA proteins make up the KEGG-HRR combined subgroup along with the MRN complex due to their close interactions and participation in HRR.

3.2.1.7 Mismatch repair proteins

MLH1 and *MLH3* are central to the mismatch repair pathway (MMR), dysfunction of which leads to microsatellite instability (Mangold et al., 2005). Mutations in both genes are strongly associated with hereditary nonpolyposis colorectal cancer (HNPCC) (Wu et al., 2001). The key MRR genes *MutS*, *MutL* and *MutH* were first identified in prokaryotes (*S. pneumoniae* and *E. coli*) (Tiraby and Fox, 1973, Siegel and Bryson, 1967). The eukaryotic homologue, MutS homologues (MSH2/3/4/5/6) forms two main heterodimers, MSH2-MSH6 (MutSα) that recognises single base pair mismatches, and the heterodimer MSH2-MSH3 (MutSβ) that recognises small insertions/deletion loops (IDLs) (Silva et al., 2009, Jun et al., 2006). Similar to the MutS homologues, the MutL homologue, MLH1 also forms the following heterodimers; MLH1-PMS2 (MutLα), MLH1-PMS1 (MutLβ), and MLH1-MLH3 (MutLγ) (Jun et al., 2006). The main MMR pathway is driven by MutSα and MutLα (**Figure 1.6**). MutSα recognises mismatches during replication and utilises ATP to undergo a conformational change that allows it to form a clamp and move along the DNA. PCNA is also loaded onto the DNA by the replication factor C (RFC), allowing MutLα to bind and nick the nascent strand. In the EXO1-dependent pathway, the resulting DNA segment from the nick is excised by the EXO1 exonuclease, in cooperation with the single-stranded DNA-binding protein RPA. In the EXO1-independent pathway, strand displacement repair synthesis is carried out. The downstream repair synthesis involves the previously mentioned POLD, POLE, and LIG1.

3.2.1.8 Ubiquitination modifiers

As the monoubiquitination of the ID2 complex is central to the FA DNA repair pathway, a number of ubiquitin proteins also form a functional group in the extended FA/BRCA-HRR network. UBE2T (FANCT) (Hira et al., 2015) is a ubiquitin-conjugating E2 ligase that directly interacts with FANCL to ubiquitinate the ID2 complex (Machida et al., 2006). It has also been shown that upon completion of the ICL repair, the timely removal of the ubiquitin from the ID complex is critical and this is performed by the deubiquitinase USP1 and its interacting partner

U2AF1 (Kim and D'Andrea, 2012). Finally, *UBA52* and *UBC* are key genes involved in stress-induced protein degradation through ubiquitination (Flick and Kaiser, 2012).

3.2.2 AML cohort and whole exome sequencing

To identify variants across the genes in the extended FA/BRCA-HRR network in AML, whole exome sequencing (WES) was performed on a cohort of primary AML samples sourced from two Australian tissue banks; the Australian Leukaemia & Lymphoma Group (ALLG) tissue bank at the Princess Alexandra Hospital (PAH, Brisbane, QLD, Australia), and the South Australian Cancer Research Biobank (SACRB, Royal Adelaide Hospital and SA Pathology, Adelaide, Australia).

The detailed characteristics of the adult AML cohort (n=145) are shown in (**Table 3.2**). The cohort consisted of more male patients (60.7%) than female. The majority of the patients were *de novo* AML (where data were available). The cohort was a representative AML cohort with respect to characteristics such as cytogenetic risks, FAB subgroups and karyotypic abnormalities which were within the range observed for other AML cohorts.

WES was performed, initially on 96 diagnostic AML samples (70 samples from SACRB and 26 samples from ALLG). The Illumina TruSeq Exome Enrichment Kit v2.0 (Illumina) was used for the first batch of AML samples (n=96), and for an all-female healthy cohort (n=329) which will be referred to as the Australian control cohort from this point forward (Duncan et al., 2011). Two criteria were set for the first batch of AML samples that were sequenced (n=96); patients were required to have had a diagnosis bone marrow blast percentage of >50%, and all M3 FAB subclass patients were excluded. A second WES experiment was performed with an additional 49 independent diagnostic adult AML samples from the ALLG tissue bank. Illumina Nextera Rapid (FC-140-1003, Illumina) was used for this second batch of adult AML samples. Thus, the total number of AML exomes analysed was 145. Massive parallel sequencing was performed using the Illumina HiSeq 2000 configured for paired-end reads. The first batch of exomes had a mean coverage of 57x (26-102x), while for the second batch mean coverage was 47.6x (22-144x).

A cohort of paediatric AML samples (n=23) sourced from the Queensland Tumour Bank (in collaboration with Dr Andrew Moore), was also included in the second batch of sequencing.

Table 3.2. Cohort characteristics of the adult AML patients used in the WES analysis (n=145)

Characteristics	*All Cases (n=145)
Age	54 (16-89)
Male - n (%)	88 (60.7%)
Female - n (%)	57 (39.3%)
WCC x 10^9/L - median (range)	19 (1.1-315.6)
BM Blast % - median (range)	80.8 (50-100)
Primary/Secondary AML - n/total (%)	
De Novo	80/88 (90.9%)
Secondary	8/88 (9.7%)
Other/Unknown	57
Transplant – n/total (%)	
Yes	25/96 (26.0%)
No	71/96 (74.0%)
Unknown	49
FAB – n/total (%)	
M0	4/89 (4.5%)
M1	34/89 (38.2%)
M2	19/89 (21.3%)
M3	0/89 (0.0%)
M4	17/89 (19.1%)
M5	14/89 (15.7%)
M6	0/89 (0.0%)
M7	1/89 (1.1%)
Unknown	56
Grimwade Cytogenetic Risk – n/total (%)	
Good	6/128 (4.7%)
Intermediate	97/128 (75.8%)
Poor	25/128 (19.5%)
Unknown	17
Simple Karyotype - n/total (%)	
Normal	77/142 (54.2%)
Abnormal	45/142 (31.7%)
Complex	20/142 (14.1%)
Unknown	3
Cytogenetics – n/total (%)	
t(15;17)	0/142 (0.0%)
CBF	10/142 (7.0%)
MLL	8/142 (5.6%)
tri(8)	12/142 (8.5%)
mono(5) / del(5q)	6/142 (4.2%)
mono(7) / del(7q)	10/142 (7.0%)
tri(21)	3/142 (2.1%)
Mutations - n/total (%)	
FLT3-ITD	49 / 145 (33.8%)
FLT3-TKD	10 / 145 (6.9%)
NPM1	54 / 145 (37.2%)
DNMT3A	45 / 145 (31.0%)
IDH1	16 / 145 (11.0%)
IDH2	23 / 145 (15.9%)
TET2	26 / 145 (17.9%)
NRAS/KRAS	17 / 145 (11.7%)

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown

FAB: French-American-British classification (Neame et al., 1986)

Grimwade classification (Grimwade et al., 2010)

The paediatric AML data was analysed using the same pipeline as the adult AML data. The results from the paediatric exome data are presented in the manuscript in **Chapter 5**.

The sequencing, base calling, alignment and initial analysis of the data was performed at the Diamantina Institute, University of Queensland by A/Prof Paul Leo, and the methods for this are detailed in the supplementary methods of the manuscripts presented in **Chapter 4 and 5**.

In order to generate a list of rare variants, a minor allele frequency of 0.001 was used to filter against dbSNP147, 1000 genome project and the Exome Sequencing Project (ESP). To remove common single nucleotide polymorphisms (SNP), variants identified in >3% of the Australian control cohort (n=329) were also excluded. Finally, a pathogenicity prediction filter (Combined Annotation Dependent Depletion, CADD v1.2, (Kircher et al., 2014)) was applied to the dataset and only variants with a CADD score $\geq$ 10 (corresponding to a >90% likelihood of being a deleterious change to protein function) were selected for downstream analyses. CADD was chosen for pathogenicity filtering because it incorporates a number of different prediction algorithms to determine pathogenicity (Kircher et al., 2014).

3.2.3 FA/BRCA-HRR mutation status

In the AML cohort, 101 out of 145 patients (72%) carried one or more variants across the extended FA/BRCA-HRR network of 58 genes (**Table 3.3),** with a total of 199 variants (151 unique individual variants) identified. Of the 199 variants, 113 were unique to the AML patients and absent in the control cohort (n=329), while 37 were unique to the AML and were absent in the non-Finnish Europeans (n=33370) in the Exome Aggregate Consortium (ExAC) database. It was noted that out of the 37 variants, 25 were absent in the entire ExAC database (n=60706).

The number of variants across each subgroup of the extended FA/BRCA-HRR is summarised in **Table 3.4,** and for each individual gene in **Table 3.5**. The complete list of 199 variants in the 101 patients is detailed in **Appendix C**. The majority of variants were identified across the FANC 19 subgroup and the ATM/ATR checkpoint subgroup, which are the largest subgroups in the extended network.

Table 3.3. Summary of the variants identified from the WES of 145 adult AML patients

Type of mutation	Total no. of variants in WES of AML	Unique individual variants	Total no. of variants absent in healthy control (n=329)	Total no. of variants absent in non-Finnish Europeans in ExAC database (n=33370)
Missense	173	127	97	29
Frameshift In/del	8	7	5	4
Nonsense	10	9	8	4
Splicing	8	8	3	0
Total	**199**	**151**	**113**	**37**

Table 3.4. Number of variants identified from the WES of 145 adult AML patients across various subgroups in the extended FA/BRCA-HRR network

Functional Subgroup	**Total number of variants**
FANC core complex	49
FANCM anchor complex	12
Minimal FANCD2 monoubiquitination complex	21
ID2 complex	8
FANC core & ID2 complex	54
Structure-specific endonucleases	25
BLM complex	21
ATM/ATR checkpoint proteins	40
RPA proteins	5
MRN complex	10
Breast cancer associated proteins	11
RAD51 paralogues	8
Mismatch repair (MMR) proteins	18
Ubiquitination modifiers	4
Combined Subgroup	
FA/BRCA-HRR Network	199
BLM & ATM/ATR checkpoint	60
KEGG Homologous Recombination Repair (HRR) Pathway	29
FANC 19	69

Table 3.5. Number of variants identified in the extended FA/BRCA-HRR network genes

Gene	No. of Variants	No. of Unique Variants
ATM	21	18
ATR	9	8
ATRIP	5	3
BARD1	1	1
BLM	4	3
BRCA1	4	2
CHEK1	0	0
CHEK2	3	3
EME1	2	2
EME2	0	0
ERCC1	0	0
FAAP100 (C17ORF70)	10	5
FAAP20 (C1orf86)	0	0
FAAP24 (C19orf40)	0	0
FAN1	7	4
FANCA	5	5
FANCB	9	9
FANCC	4	2
FANCD1 (BRCA2)	3	3
FANCD2	4	4
FANCE	1	1
FANCF	0	0
FANCG	4	1
FANCI	4	4
FANCJ (BRIP1)	1	1
FANCL	11	9
FANCM	12	8
FANCN (PALB2)	2	2
FANCO (RAD51C)	1	1
FANCP (SLX4)	7	7
FANCQ (ERCC4)	6	5
MHF1 (APITD1)	0	0
MHF2 (STRA13)	0	0
MLH1	6	4
MLH3	12	5
MRE11A	1	1
MUS81	3	2
NBN	3	3
RAD9A	0	0
RAD17	2	2
RAD50	6	6
RAD51	0	0
RAD51B	5	3
RAD51D	1	1
RBBP8 (CTIP)	3	3
RMI1	2	2
RMI2	0	0
RPA1	2	2
RPA2	1	1
RPA3	2	2
SLX1A	0	0
TOP3A	15	8
UAF1	0	0
UBA52	0	0
UBC	2	2
UBE2T	0	0
USP1	2	2
XRCC2	0	0
XRCC3	1	1

3.2.3.1 Sanger validation of FA/BRCA-HRR variants

To validate the rare variants, and determine somatic/germline status, Sanger sequencing was performed on genomic DNA (gDNA) isolated from diagnosis bone marrow mononuclear cells (BMMNC) and where available, Sanger sequencing was also performed on gDNA from mesenchymal stromal cells (MSC) or $CD3^+$ T-cells as non-disease material. The sequencing primers used are listed in **Appendix A**. In total, 33 variants of the FA/BRCA-HRR were tested using this approach with all of these variants validating in the BMMNC samples. While germline/somatic status could not be determined for 13 variants (due to the lack of non-disease materials), 18 variants were determined to be germline in nature. Furthermore, two variants were confirmed to be somatically acquired; FANCM-p.Q1333fs (c.3998delA) and NBN-p.R43X (c.C127T) as shown in **Appendix D**. The high level of validation by Sanger sequencing showed the robustness of the WES sequencing platform and bioinformatics analyses that were used. Based on this analysis, and the median variant allele frequencies (VAF) which was calculated to be 46.8% (**Appendix C**), discussion of variants is focused on the implications of germline variants, although it is acknowledged that a small percentage of these variants (in the order of 6%) are likely somatically acquired in the tumour.

3.2.4 Frequency of FA/BRCA-HRR network variants in AML and healthy controls

Having obtained the mutation data of the extended FA/BRCA-HRR network in AML, it was important to determine the frequency of rare variants in these genes in AML and healthy individuals. It has been shown in population studies, ethnic groups such as Asians, Ashkenazi Jews and Finnish Europeans have distinct variants that are present at high frequencies (Lim et al., 2016, Kimchi-Sarfaty et al., 2007). These populations are also under-represented in the databases commonly used to filter for rare disease-associated variants (such as dbSNP, 1000 genome project, ESP, ExAC) leading to a contamination of disease-associated rare variants with ethnic population SNPs (Bustamante et al., 2011). In order to account for ethnically-biased variants, only Caucasian adult AML samples (n=131) were included for this analysis.

Therefore, the variants across genes of the extended FA/BRCA-HRR network for the Caucasian AML cohort were compared to that of the non-Finnish Europeans in the ExAC database. Thus, the total number of alleles that were examined for the AML cohort and the non-Finnish European control cohort in ExAC were 262 and 66740 respectively. The results presented below in **Table 3.6** relate only to the non-FANC genes since the results for the 19 FANC genes

Table 3.6. Frequency of FA/BRCA-HRR network gene variants (excluding the 19 FANC genes) in AML

	AML cohort; Allele count (n=262)		ExAC cohort; Allele count (n=66740)		
Gene	Allele count	Frequency	Allele count	Frequency	**P*-value
ATM	16	0.061069	1572	0.023554	**0.0006**
ATR	7	0.026718	940	0.014085	0.1041
ATRIP	5	0.019084	489	0.007327	0.0457
BARD1	1	0.003817	275	0.00412	1.0000
BLM	4	0.015267	568	0.008511	0.2923
C17ORF70	7	0.026718	298	0.004465	**0.0002**
CHEK2	3	0.01145	531	0.007956	0.4697
EME1	2	0.007634	340	0.005094	0.3869
FAN1	6	0.022901	488	0.007312	0.0137
MLH1	6	0.022901	931	0.01395	0.1879
MLH3	10	0.038168	702	0.010518	**0.0006**
MRE11A	1	0.003817	789	0.011822	0.3826
MUS81	2	0.007634	309	0.00463	0.3435
NBN	3	0.01145	308	0.004615	0.1233
RAD17	2	0.007634	268	0.004016	0.0909
RAD50	5	0.019084	812	0.012167	0.2572
RAD51B	5	0.019084	363	0.005439	0.0153
RAD51D	1	0.003817	198	0.002967	0.542
RAD9A	0	0	291	0.00436	0.634
RBBP8	3	0.01145	460	0.006892	0.4339
RMI1	2	0.007634	485	0.007267	0.7163
RPA1	2	0.007634	324	0.004855	0.3646
RPA2	1	0.003817	120	0.001798	0.3778
RPA3	1	0.003817	40	0.000599	0.1484
TOP3A	10	0.038168	423	0.006338	**<0.0001**
UBC	1	0.003817	48	0.000719	0.1747
USP1	2	0.007634	390	0.005844	0.6683
XRCC3	1	0.003817	112	0.001678	0.358

*Fisher's exact test was used to calculate the *P*-value

Red bolded, *P*<0.0017 (Bonferroni corrected), statistically significant *P*-value

are presented in **Chapter 4**. Fisher's exact test was performed to determine differences in the number of mutated alleles for each of the remaining non-FANC genes of the FA/BRCA-HRR network. Genes that were significantly over represented in the AML samples (compared to the controls in ExAC) were *ATM*, *C17ORF70*, *MLH3* and *TOP3A* as presented in **Table 3.6**. Burden testing provides a powerful statistical analytical method for detection of rare SNPs associated with disease traits, while accounting for gene sizes to minimize false positive enrichment associated with large genes (see **Chapter 4; Section 4.2; Supplementary Methods**). With the help of A/Prof Leo (Queensland University of Technology) the burden test was used to determine the mutation burden of the FA/BRCA-HRR network genes in AML. Following filtering of the WES for the AML (n=131) and Australian control (n=323) cohorts to remove SNPs that are associated with ethnic groups, *FANCL* and *RMI1* were observed to be significantly enriched for rare variants in the AML cohort (**Figure 3.2**). It was noted that *RAD9A* approaches significance as well. Enrichment of *FANCL* is discussed further in **Chapter 4**, and the role of *RMI1* is discussed further in the Final Discussion **Chapter 6**.

3.2.5 Analysis of common AML mutations in the WES

3.2.5.1 Recurrent AML mutations in the WES cohort

The co-occurrence or exclusivity of the variants in the extended FA/BRCA-HRR network genes with frequent somatic mutations described in AML (**Chapter 1; Section 1.6.1**), was next examined. The mutational status for the 43 genes associated with recurrent somatic mutations in AML (Cancer Genome Atlas Research, 2013) was determined for the AML cohort using the WES data. An alternative pipeline was implemented such that all the variants across these 43 genes (**Appendix E**) were identified without filtering for minor allele frequency and CADD score. This prevented the loss of somatic mutations due to low allelic frequency. Only missense variants that have been reported in haematopoietic malignancies in the Catalogue of Somatic Mutations in Cancer (COSMIC) and the previously published TCGA AML dataset (Cancer Genome Atlas Research, 2013)) were annotated as mutations. All nonsense and frameshift insertion/deletion (in/del) variants were also included in this analysis. Results of this analysis is shown in **Appendix F**. These results were compared to the published TCGA AML cohort to show similar frequency of mutations in the 43 genes (**Chapter 4; Section 4.2 supplementary figure S1**).

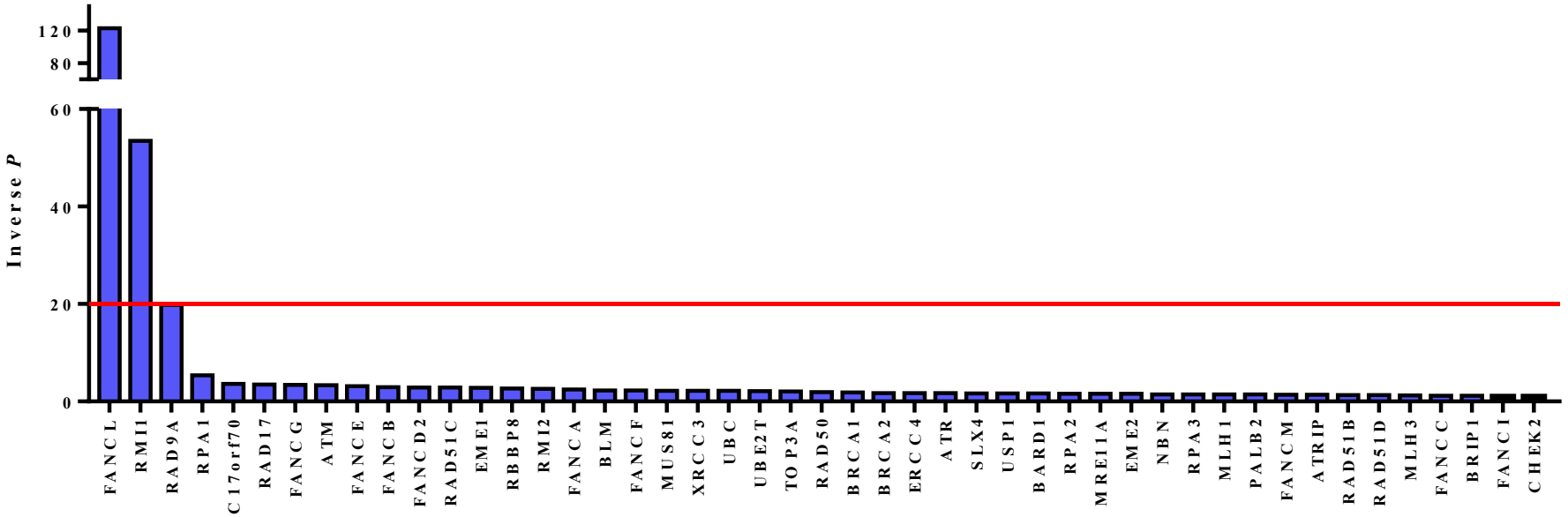


Figure 3.2. Graphical representation of mutational Burden test results for genes in the extended FA/BRCA-HRR network. The inverse p-values from statistical Burden test of individual genes are plotted on the y-axis. Statistically significant threshold is represented by the red line (*P*=0.05).

For a detailed genomic, molecular and cytogenetic analysis, the results from next generation sequencing were integrated with molecular testing and conventional cytogenetic analysis to identify potential cooperation and exclusivity. A visual representation of these data is shown in **Figure 3.3** and the results of statistical tests are shown in **Table 3.7**. Overall, samples that carried variants in the extended FA/BRCA-HRR network (n=58 genes) displayed statistically increased abnormal cytogenetics (*P*=0.012) and decreased frequency of normal cytogenetics (*P*=0.0115) (**Figure 3.3; Table 3.7**). Similar statistical analyses were performed for all 17 subgroups (**Table 3.1**); results are shown only for the subgroups that had more than 10 samples (n=12) in **Appendices G-R** (a similar visual representative figure is shown in **Appendix S)**. The FANC 19 mutant subgroup, showed a statistically decreased number of *FLT3*-ITD mutations (*P*=0.048) (**Appendix L**). Patients carrying variants in the KEGG-HRR subgroup (n=27 patients) displayed statistically increased abnormal karyotype cytogenetics (*P*=0.01) and decreased frequency of normal karyotype cytogenetics (*P*=0.03, **Appendix R**), thus this subgroup is likely responsible for the association of the rare variants across the whole network with normal/abnormal cytogenetics. The KEGG-HRR subgroup was also associated with an increased number of *MLL* translocations (*P*=0.04).

3.2.6 Recurring variants identified in the extended FA/BRCA-HRR network genes from the WES of AML samples

Rare variants identified in more than one sample as shown in **Figure 3.4** are of particular interest. Such variants are also summarised in **Table 3.8**. A number of variants identified recurrently in the AML cohort were associated with Asian ethnicity (i.e. C17ORF70 variant p.A694P identified in WES-9, WES-203, and WES-207). This is indicative of the under-representation of non-Caucasians in the SNP databases and the inability to effectively filter out ethnic variants when looking for rare variants associated with disease-cohorts. Therefore, variants that occur at high frequency in non-Caucasian AML samples will not be discussed further. Several recurrent variants observed in the 19 FANC genes were of interest and are discussed in the manuscript presented in **Chapter 4** (**Section 4.2**). Recurrent variants were also observed for the master transducers of DNA damage signalling, ATM and ATR (ATM: p.V410A, p.S333F and p.F582L, and *ATR*: p.H90Y). ATM variants p.V410A and p.F582L were identified in 2 samples (WES-30 and WES-55, and WES-28 and WES-62 respectively), while the ATM variant p.S333F was identified in 3 samples (WES-14, WES-22 and WES-245).

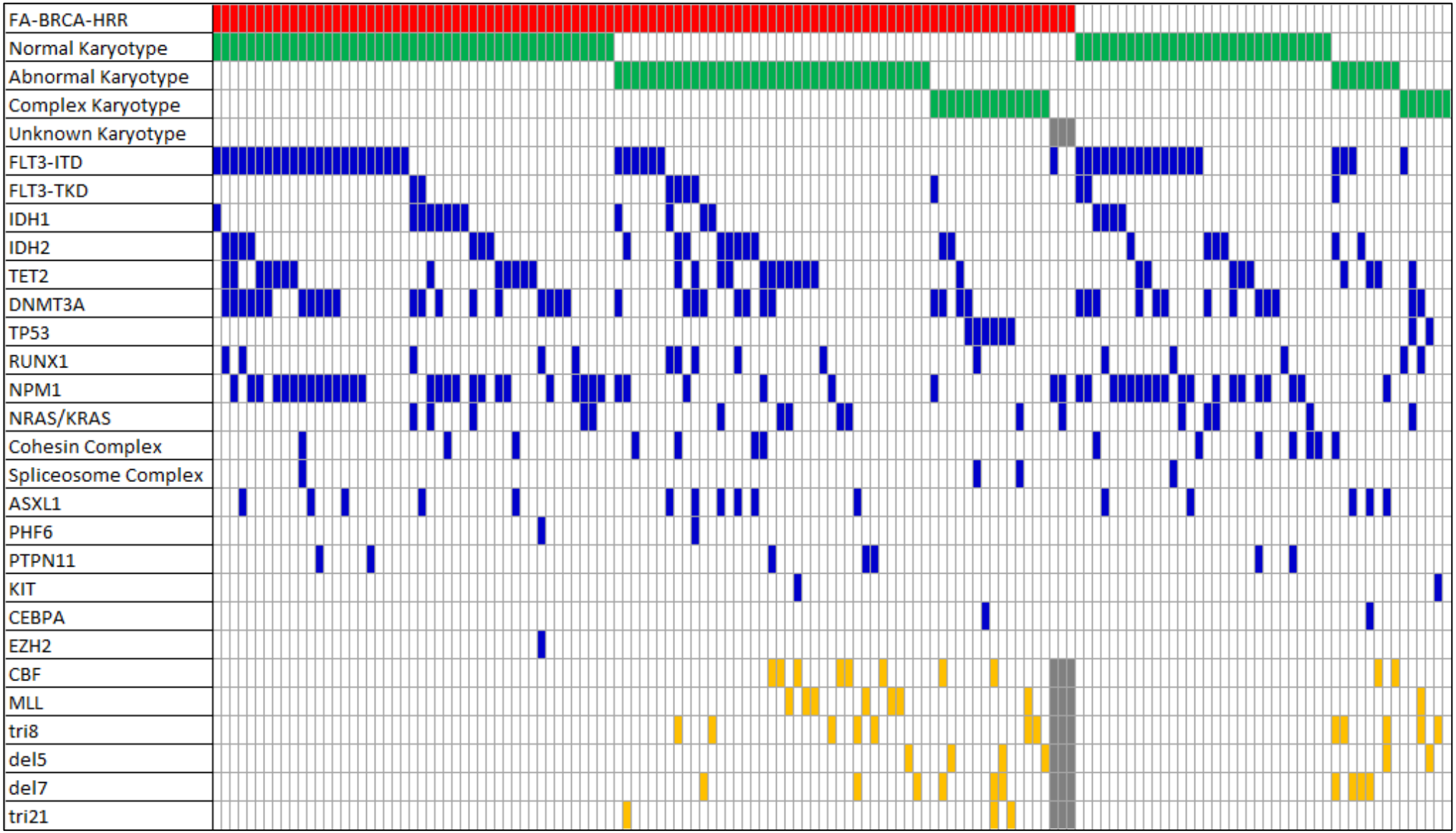

FA-BRCA-HRR
Normal Karyotype
Abnormal Karyotype
Complex Karyotype
Unknown Karyotype
FLT3-ITD
FLT3-TKD
IDH1
IDH2
TET2
DNMT3A
TP53
RUNX1
NPM1
NRAS/KRAS
Cohesin Complex
Spliceosome Complex
ASXL1
PHF6
PTPN11
KIT
CEBPA
EZH2
CBF
MLL
tri8
del5
del7
tri21

Figure 3.3. Graphical representation of the extended FA/BRCA-HRR network variants, common AML mutations and karyotypic abnormalities across the adult WES AML cohort (n=145). Red cells represent samples with mutations in the extended FA/BRCA-HRR network; Green cells represent the simple karyotype classification; Blue cells represent samples with recurring AML mutations; Yellow cells represent samples with karyotypic abnormalities; Grey cells represent samples with unknown karyotypes. Cohesin complex and spliceosome complex genes are defined as in (Cancer Genome Atlas Research, 2013).

Table 3.7. Cohort characteristics of the extended FA/BRCA-HRR network mutant subgroup samples

Characteristics	*All Cases (n=145)	FA/BRCA-HRR mutant	FA/BRCA-HRR non-mutant	[1]*P*-value
Age	54 (16-89)	54 (16-84)	54.5 (17-89)	0.999^
Male - n (%)	88 (60.7%)	64 (63.4%)	24 (54.5%)	0.358
Female - n (%)	57 (39.3%)	37 (36.6%)	20 (45.5%)	0.358
WCC x 10^9/L - median (range)	19 (1.1-315.6)	20.3 (1.17-315.6)	14.65 (1.07-313.3)	0.780^
BM Blast % - median (range)	80.8 (50-100)	81 (50-100)	79.75 (50-95)	0.522^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	55/62 (88.7%)	25/26 (96.2%)	0.434
Secondary	8/88 (9.1%)	7/62 (11.3%)	1/26 (3.8%)	0.434
Other/Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	17/68 (25.0%)	8/28 (28.6%)	0.799
No	71/96 (74.0%)	51/68 (75.0%)	20/28 (71.4%)	0.799
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	4/65 (6.2%)	0 (0%)	0.571
M1	34/89 (38.2%)	23/65 (35.4%)	11/24 (26.2%)	0.462
M2	19/89 (21.3%)	14/65 (21.5%)	5/24 (11.9%)	1
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	13/65 (20.0%)	4/24(9.52%)	1
M5	14/89 (15.7%)	10/65 (15.4%)	4/24 (9.52%)	1
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	1/65 (0.99%)	0 (0%)	1
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	5/89 (5.6%)	1/39 (2.6%)	0.666
Intermediate	97/128 (75.8%)	66/89 (74.2%)	31/39 (79.5%)	0.655
Poor	25/128 (19.5%)	18/89 (20.2%)	7/39 (17.9%)	1
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	44/98 (44.9%)	30/44 (68.2%)	**0.0115**
Abnormal	45/142 (31.7%)	40/98 (40.8%)	8/44 (18.2%)	**0.012**
Complex	20/142 (14.1%)	14/98 (14.3%)	6/44 (13.6%)	1
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	8/98 (8.16%)	2/44 (4.55%)	0.724
MLL	8/142 (5.6%)	7/98 (7.14%)	1/44 (2.27%)	0.435
tri(8)	12/142 (8.5%)	7/98 (7.14%)	5/44 (11.4%)	0.515
mono(5) / del(5q)	6/142 (4.2%)	4/98 (4.08%)	2/44 (4.55%)	1
mono(7) / del(7q)	10/142 (7.0%)	6/98 (6.12%)	4/44 (9.09%)	0.500
tri(21)	3/142 (2.1%)	3/98 (3.06%)	0 (0%)	0.552
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	30/101 (29.7%)	19/44 (43.2%)	0.129
FLT3-TKD	10/145 (6.9%)	6/101 (5.94%)	2/44 (4.55%)	1
NPM1	54/145 (37.2%)	33/101 (32.7%)	18/44 (40.9%)	0.450
DNMT3a	45/145 (31.0%)	34/101 (33.7%)	12/44 (27.3%)	0.561
IDH1	16/145 (11.0%)	12/101 (11.9%)	4/44 (9.09%)	0.777
IDH2	23/145 (15.9%)	18/101 (17.8%)	6/44 (13.6%)	0.632
TET2	26/145 (17.9%)	16/101 (15.8%)	7/44 (15.9%)	1
NRAS/KRAS	17/145 (11.7%)	12/101 (11.9%)	5/44 (11.4%)	1

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown.[1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-American-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

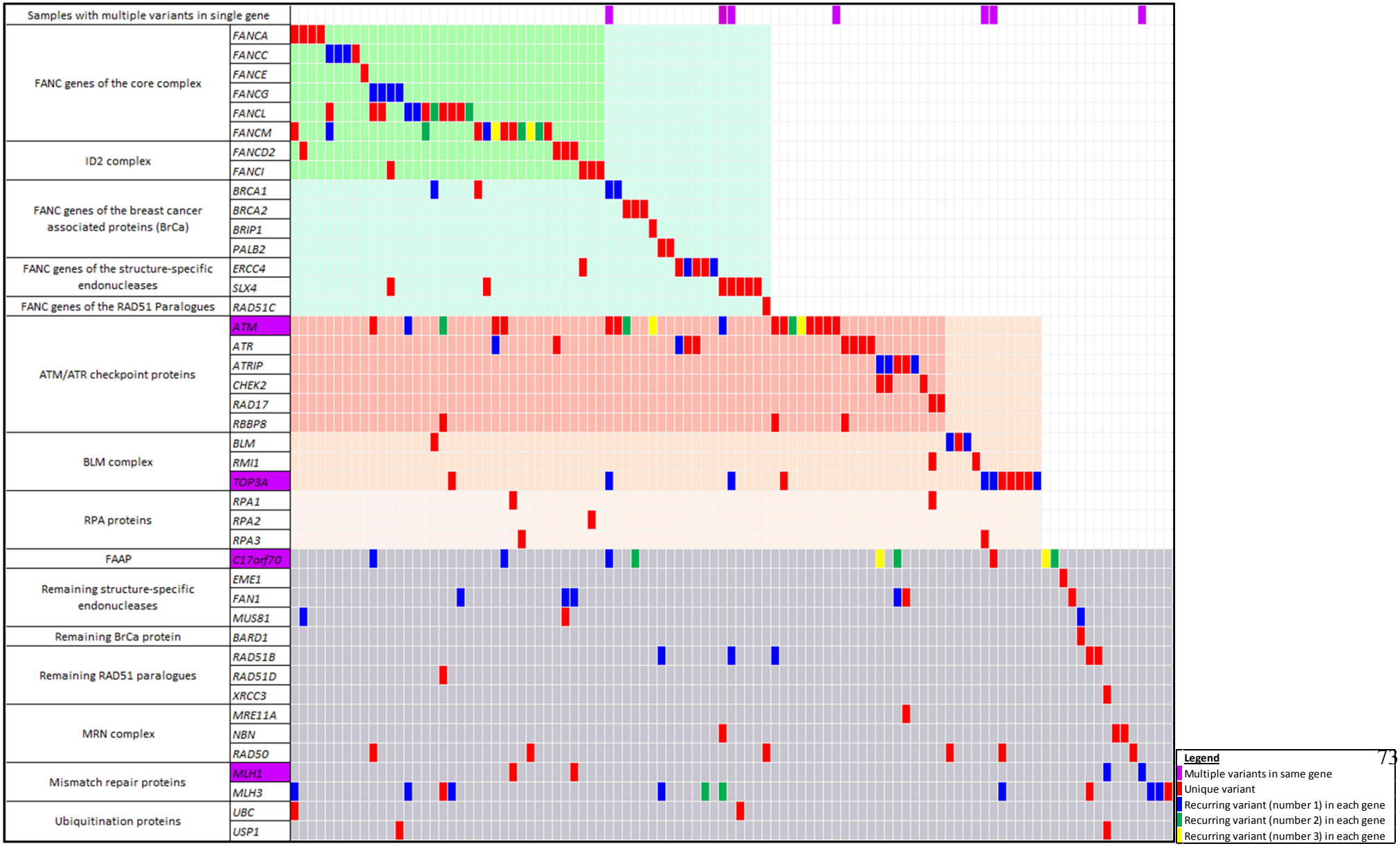

Samples with multiple variants in single gene
FANC genes of the core complex
FANCA
FANCC
FANCE
FANCG
FANCL
FANCM
ID2 complex
FANCD2
FANCI
FANC genes of the breast cancer associated proteins (BrCa)
BRCA1
BRCA2
BRIP1
PALB2
FANC genes of the structure-specific endonucleases
ERCC4
SLX4
FANC genes of the RAD51 Paralogues
RAD51C
ATM/ATR checkpoint proteins
ATM
ATR
ATRIP
CHEK2
RAD17
RBBP8
BLM complex
BLM
RMI1
TOP3A
RPA proteins
RPA1
RPA2
RPA3
FAAP
C17orf70
Remaining structure-specific endonucleases
EME1
FAN1
MUS81
Remaining BrCa protein
BARD1
Remaining RAD51 paralogues
RAD51B
RAD51D
XRCC3
MRN complex
MRE11A
NBN
RAD50
Mismatch repair proteins
MLH1
MLH3
Ubiquitination proteins
UBC
USP1
Legend
Multiple variants in same gene
Unique variant
Recurring variant (number 1) in each gene
Recurring variant (number 2) in each gene
Recurring variant (number 3) in each gene

Figure 3.4. Graphical representation of FA/BRCA-HRR extended network variants across the adult WES AML cohort. Only the 101 samples which had >1 variant in the extended FA/BRCA-HRR network are shown. Each column represents an individual sample and each row shows mutational status in the various genes and subgroups in the extended FA/BRCA-HRR network. Samples that carried variants in the 19 FANC genes and FANC core and ID2 genes are emphasised using the 2 different light-green shaded regions; BLM complex, ATM/ATR checkpoint and RPA protein subgroups are emphasised using the 3 light pink regions; and the remaining subgroups are emphasised using light grey region. Recurring variants are coloured blue for the first recurrent variant in a gene, green for the second recurrent variant in a gene, and yellow for the third recurrent variant in a gene. Variants that were in only one sample are coloured in red.

Table 3.8. Frequency of recurring variants in the extended FA/BRCA-HRR network

WES-ID	Gene	Ref Seq	Nucleotide change	Amino acid change	Frequency in our AML cohort	Frequency in ExAC (Non-Finnish European)	Frequency in control cohort
WES-30, WES-55	ATM	NM_000051	c.T1229C	p.V410A	0.014	0.003	0.003
WES-14, WES-22, WES-245	ATM	NM_000051	c.C998T	p.S333F	0.021	0.001	0.009
WES-28, WES-62	ATM	NM_000051	c.T1744C	p.F582L	0.014	0.001	0
WES-26, WES-38	ATR	NM_001184	c.C268T	p.H90Y	0.014	0.007	0.006
WES-6, WES-35, WES-72	ATRIP	NM_032166	c.G2024A	p.R675Q	0.021	0.01	0.015
WES-66, WES-248	BLM	NM_000057	c.T11C	p.V4A	0.014	0.003	0.003
WES-13, WES-60, WES-207	BRCA1	NM_007298	c.G1644A	p.M548I	0.021	0.002	0.009
WES-32, WES-235, WES-237	C17orf70	NM_025161	c.C791T	p.A264V	0.021	0.004	0.006
WES-35, WES-213	C17orf70	NM_025161	c.A1433G	p.Q478R	0.014	0.002	0
WES-9, WES-203, WES-207	C17orf70	NM_025161	c.G2080C	p.A694P	0.021	0.003	0
WES-74, WES-208	ERCC4	NM_005236	c.C2288T	p.P763L	0.014	0	0
WES-18, WES-235, WES-249, WES-250	FAN1	NM_001146094	c.G718A	p.E240K	0.028	0.005	0.009
WES-20, WES-79, WES-86	FANCC	NM_001243743	c.A584T	p.D195V	0.021	0.004	0.006
WES-1, WES-64, WES-83, WES-203	FANCG	NM_004629	c.C890T	p.T297I	0.028	0.001	0
WES-8, WES-60	FANCL	NM_018062	c.1099_1100insATTA	p.T367fs	0.014	0.003	0.012
WES-30, WES-216	FANCL	NM_018062	c.1007_1009del	p.336_337del	0.014	0	0.003
WES-81, WES-86	FANCM	NM_020937	c.A2859C	p.K953N	0.014	0.002	0.003
WES-26, WES-202	FANCM	NM_020937	c.C527T	p.T176I	0.014	0.006	0.006
WES-78, WES-89, WES-231	FANCM	NM_020937	c.C1576G	p.L526V	0.021	0.001	0.006
WES-88, WES-218	MLH1	NM_001258274	c.A1129G	p.K377E	0.014	0.005	0.009
WES-88, WES-218	MLH1	NM_001258274	c.A1129C	p.K377T	0.014	0.005	0.009
WES-30, WES-41, WES-67, WES-209, WES-211, WES-226, WES-246	MLH3	NM_014381	c.G2221T	p.V741F	0.048	0.007	0.003
WES-37, WES-55	MLH3	NM_014381	c.C3315A	p.D1105E	0.014	0.003	0.009
WES-21, WES-58	MUS81	NM_025128	c.G416A	p.R139Q	0.014	0.005	0
WES-25, WES-226, WES-227	RAD51B	NM_133510	c.A728G	p.K243R	0.021	0.01	0.012
WES-63, WES-217, WES-223, WES-227	TOP3A	NM_004618	c.G1375A	p.D459N	0.028	0.005	0.009
WES-63, WES-207, WES-223, WES-227	TOP3A	NM_004618	c.G1381A	p.A461T	0.028	0.002	0

Variants discussed in **Chapter 3** are shown in Red.

All three of these variants were observed at increased frequencies in the AML cohort, have previous disease-associations and are discussed further below in **Section 3.2.7.2**. The MLH3 variant p.V741F was identified in 7 AML samples (WES-30, WES-41, WES-67, WES-209, WES-211, WES-226 and WES-246) and in only on Australian healthy control. More importantly, this variant has shown to be a pathogenic mutation in colon cancer (Kim et al., 2007, Liu et al., 2003). The variant was reported in a 71 year old man with colon cancer, and in his two sisters who developed gastric and breast cancer respectively (Kim et al., 2007). Kim and colleagues reported that this variant was absent in their control cohort and a sporadic colon cancer cohort. A separate study reported this same variant in an endometrial cancer family (Liu et al., 2003). Thus, there is strong evidence for this variant to be a familial cancer risk allele.

Other recurring variants of interest include the TOP3A variants p.A461T and p.D459N which were identified together in the same 4 samples (WES-63, WES-217, WES-223 and WES-227). The variant p.A461T was absent in the Australian control cohort and has a frequency of 0.002 in the non-Finnish European population in ExAC (**Table 3.7**). For C17ORF70, which is a direct interacting partner of FANCL and a member of the minimal-ubiquitination complex (Rajendra et al., 2014), two recurrent variants; p.A264V in three samples, and p.Q478R (absent in the healthy Australian controls) in two separate samples were observed (**Table 3.8**).

3.2.7 Disease-causing and disease-associated mutations in the extended FA/BRCA-HRR network

In addition to well classified cancer predisposing genes such as *BRCA1* and *BRCA2*, mutations in a number of genes in the extended FA/BRCA-HRR network are known to directly cause recessive disorders such as the DNA repair disorders described in **Chapter 1** (**Section 1.7.2**), ataxia telangiectasia (A-T), Seckel syndrome, Bloom's syndrome (BS) and Nijmegen breakage syndrome (NBN) caused by mutations in *ATM*, *ATR*, *BLM* and *NBN* respectively.

The WES data was cross-tabulated against the disease-specific databases for Fanconi Anaemia (FAdb curated by Rockefeller University; http://www2.rockefeller.edu/fanconi/), and breast cancer (kConFAB*: BRCA1/2*, *BRIP1* and *PALB2*, http://www.kconfab.org/Index.shtml; and Breast Cancer Information Core, BIC: *BRCA1/2*, https://research.nhgri.nih.gov/bic/), and also against more general disease databases (such as COSMIC, the Human Gene Mutation Database (HGMD) and Online Mendelian Inheritance in Man (OMIM) compendium) to determine

whether the variants that were identified from the WES data of the AML cohort had previously been reported in cancer or human diseases. **Table 3.9** shows all the variants that were identified from this analysis. Other variants which were not reported in COSMIC, but mutations at the same amino acid residue as reported in COSMIC are shown in **Appendix T.** A number of variants were reported in HGMD with uncertain pathogenicity and are listed in **Appendix U**. Several key variants are discussed below.

3.2.7.1 Disease-causing mutations in the FA/BRCA-HRR network genes

For the purposes of this study, disease-causing (D-C) mutations have been defined as *(i)* mutations for which there is evidence in the disease databases causally linking them to the disease/syndrome, or *(ii)* mutations for which there is an OMIM entry linking them to disease. For example, the FANC mutations reported in the FAdb and the BRCA mutations reported in kConFAB and BIC as pathogenic in FA or breast/ovarian cancer. The D-C mutations in the 19 FANC genes are reported in detail the attached manuscript in **Chapter 4** (**section 4.2**).

The BLM variant p.R899X which affects a residue in the RecQ C-terminal BLM helicase domain (Kitano, 2014) was identified in WES-60 and is reported to cause Bloom's syndrome (German et al., 2007). This variant has also been reported to be associated with breast cancer risk in Slavic populations (Prokofyeva et al., 2013). It has also has been reported in colon cancer (COSMIC, **Table 3.9**) and in an AML sample in a published multi-cancer study (Lu et al., 2015). Thus there is an accumulation of evidence for this variant as a D-C mutation.

3.2.7.2 Disease-associated mutations in the FA/BRCA-HRR network genes

For the purposes of this study, disease-associated (D-A) mutations have been defined as mutations that have been reported in disease databases as being associated with predisposition to a disease which the gene is not classically associated with. For example, two the variants in ATM, p.F582L (identified in two AML samples) and p.K2431E have been reported to increase breast cancer susceptibility (**Table 3.9)**. Several *FANCM* and *FANCL* variants have been reported in a study of the familial congenital heart syndrome, Tetralogy of Fallot (TOF), but are not reported in FA or breast cancer. These include the FANCM variant p.Q1701X (Grunert et al., 2014), identified in WES-46, and two FANCL variants p.L38F and p.T224A that were not confirmed as pathogenic in TOF (**Appendix U).** A separate FANCL D-A variant, p.P17R (identified in WES-41), was shown in a recent study by Nicchia and colleagues to be a loss of function mutation, but this variant is not found in the FA database (Nicchia et al., 2015).

Table 3.9. Disease-causing and disease-associated mutations in the FA/BRCA-HRR network

WES ID	Gene Name	Variant	#COSMIC ID	Type of Cancer in COSMIC (No. of samples)	Type of disease (FAdb ID)	*HGMD Classification	Disease-causing (D-C) or Disease-associated (D-A)	#References
WES-14, WES-22, WES-245	*ATM*	S333F	COSM5020963	Haemangioblastoma (1)	-	N/A	-	-
WES-28, WES-62	*ATM*	F582L	-	-	Breast Cancer Susceptibility	Damaging	D-A	(Sommer et al., 2002)
WES-55	*ATM*	K2431E	-	-	Breast Cancer Susceptibility	Damaging	D-A	(Sommer et al., 2002)
WES-9	*ATM*	N1650S	COSM5009679	Colorectal cancer (1)	-	N/A	-	-
WES-30, WES-55	*ATM*	V410A	COSM5945737 & COSM21825	Glioma (2), Lymphoma (1), Non-Hodgkin Lymphoma (1), CLL (1), Melanoma (2)	-	N/A	-	-
WES-203	*ATM*	H1380Y	COSM24627	B-cell Lymphoma (1)	-	N/A	-	-
WES-96	*ATM*	K1454N	COSM22501	Lymphoma (1)	-	N/A	-	-
WES-101	*BLM*	P332S	COSM3505421	Malignant melanoma (2)	-	N/A	-	-
WES-60	*BLM*	R899X	COSM3690618	Colorectal cancer (1)	Bloom Syndrome	Damaging	D-C	(German et al., 2007)
WES-46	*BRCA1*	R1203X	-	-	Breast Cancer (BRCA1_001405)	N/A	D-C	kConFab
WES-13, WES-60, WES-207	*BRCA1*	M548I	COSM4985686	Haemangioblastoma (1) & Rhabdomyosarcoma (1)	-	N/A	-	-
WES-245	*BRCA2*	R2034C	COSM4987322 & COSM696739	Rhabdomyosarcoma (1) & ^Lung SCC (1)	Leukaemia risk	Disease polymorphism	D-A	(Rudd et al., 2006)
WES-223	*C17orf70*	P387L	COSM3378395	Pancreatic cancer (2)	-	N/A		-
WES-100	*CHEK2*	R474H	COSM4103242	Stomach (1) & Ovarian (1) cancer	Non-Hodgkin lymphoma	Damaging	D-A	(Aloraifi et al., 2015)

WES-35	*CHEK2*	I157T	COSM3693990	Colorectal cancer (1)	Li-Fraumeni syndrome; Increased risk of CLL	Disease functional polymorphism	D-A	(Bell et al., 1999, Rudd et al., 2006)
WES-37	*ERCC4*	R340Q	COSM4848860 & COSM1375969	Cervical (1), ^Breast (1) & ^Colorectal cancer (2)	-	N/A	-	-
WES-21	*FANCA*	T1131A	-	-	Fanconi anaemia (FANCA_000241)	N/A	D-C	(Adachi et al., 2002)
WES-224	*FANCC*	R548X	COSM1206492	Colorectal cancer (1)	Fanconi anaemia (FANCC_000005)	Damaging	D-C	(Murer-Orlando et al., 1993)
WES-20, WES-79, WES-86	*FANCC*	D195V	-	-	Fanconi anaemia	N/A	D-C	(Verlander et al., 1994)
WES-21	*FANCD2*	R926X	COSM4484163, COSM39696 & COSM1732713	[1]SCC (1), Colorectal cancer (1), ^Glioma (2) & ^Pancreatic cancer (1)	Fanconi anaemia (FANCD2_000019)	N/A	D-C	(Kalb et al., 2007)
WES-249	*FANCD2*	E906Lfs*4	-	-	Fanconi anaemia (FANCD2_000016)	N/A	D-C	(Kalb et al., 2007)
WES-1, WES-64, WES-83, WES-203	*FANCG*	T297I	COSM150601	Gastric cancer (1)	Fanconi anaemia (FANCG_000038)	N/A	D-C	(Nakanishi et al., 2001)
WES-225	*FANCI*	D251D	COSM4057754	Gastric cancer (1) & Ewing sarcoma (1)	-	N/A	-	-
WES-41	*FANCL*	P17R	-	-	Fanconi anaemia	Damaging	D-A	(Nicchia et al., 2015)
WES-8, WES-60	*FANCL*	T367fs	COSM5608709 & COSM70732	Malignant melanoma (1) & ^Ovarian cancer (1)	-	N/A	-	-
WES-30, WES-216	*FANCL*	336_337del	-	-	Fanconi anaemia (FANCL_000002)	N/A	D-C	(Ali et al., 2009)
WES-46	*FANCM*	Q1701X	COSM126686 & COSM1369862	[2]HNSCC (1) & ^Colorectal cancer (1)	Tetralogy of Fallot; Enriched in triple negative breast cancer; Paediatric B-ALL	N/A	D-A	(Grunert et al., 2014, Kiiski et al., 2014, Zhang et al., 2015)
WES-209	*FANCM*	T1941T	COSM955830	Endometrium cancer (1)	-	N/A	-	-
WES-88, WES-218	*MLH1*	K377T	COSM25915, COSM26083, COSM1422600	Colorectal (1), ^Colorectal (1), ^Colorectal (1)	Colorectal cancer	Damaging	D-A	(Han et al., 1995)

WES-94	*MLH1*	Y320X	COSM5967443	Endometrium cancer (1)	Colorectal cancer	Damaging	D-A	(Mangold et al., 2005)
WES-215	*MRE11A*	R380H	COSM4703143 & COSM2061340	Colorectal (1) & ^Colorectal cancer (1)	-	N/A	-	-
WES-31	*NBN*	R43X	-	-	Cancer	Damaging	D-A	(Lu et al., 2015, LaDuca et al., 2014)
WES-54	*NBN*	D95N	-	-	Acute lymphoblastic leukaemia	Damaging	D-A	(Mosor et al., 2006)
WES-27	*RAD51B*	V207L	COSM3815166	Breast cancer (1)	-	N/A	-	-
WES-33	*RAD51C*	A126T	-	-	Reduced activity	Functional polymorphism	D-A	(Meindl et al., 2010)
WES-94	*RPA1*	G434R	COSM1381221	Colorectal cancer (1)	-	N/A	-	-
WES-78	*RPA3*	P58P	COSM4722505	Gastric (1) & colorectal cancer (1)	-	N/A	-	-
WES-23	*SLX4*	E787K	-	-	Breast and/or ovarian cancer risk	Damaging	D-A	(de Garibay et al., 2013)
WES-84	*TOP3A*	P802L	COSM2741009	Kidney Renal Papillary Cell Carcinoma (1)	-	N/A	-	-

[1]Squamous cell carcinoma;

[2]Head and neck squamous cell carcinoma;

^Samples with different mutation at the same amino acid as identified in our study;

*N/A are variants that were absent in HGMD.

[#]References from FAdb, kConFab and HGMD are listed. References for COSMIC are available through the COSMIC IDs.

Variants discussed in detail in **Chapter 3** are in Red font.

The D-A variants identified in the AML also include a number of variants that are reported in COSMIC. For example, the ATM variant p.S333F, identified in 3 AML samples, was reported as a somatic mutation in a haemangioblastoma sample. The amino acid residues from Y332 to R337 in ATM have all been reported to be mutated in COSMIC, suggesting that it is potentially a mutational hotspot. The ATM variant p.V410A (**Table 3.9)** identified in 2 AML samples was reported in glioma (2 samples), lymphoma (1 sample), Non-Hodgkin lymphoma (1 sample), CLL (1 sample) and melanoma (2 samples). The ATM variants p.V410A and p.S333F have also been reported in ocular telangiectasia and colorectal cancer (**Appendix U)**. A number of CHEK2 variants are listed in COSMIC (**Table 3.9)**; p.I157T (WES-35), which is reported in colon and ovarian cancer, has also been associated with Li-Fraumeni syndrome and CLL (Bell et al., 1999, Rudd et al., 2006). A second CHEK2 COSMIC and WES variant R474H (WES-100) has been reported to be associated with Non-Hodgkin lymphoma (Aloraifi et al., 2015). Finally, two COSMIC variants affecting the deacetylation site Lys^{377} of MLH1 (p.K377T and p.K377E) have been reported multiple times in colorectal cancer were also identified in the AML samples. D-A variants also include a number of variants affecting an amino acid residue frequently mutated in COSMIC (**Appendix T**). For all of these variants, it will be important to determine the germline/somatic status in the AML samples, if material is available.

3.2.8 Gene expression profiling of FA/BRCA-HRR mutant AML

To reveal differences in the biology of AML subtypes based on the mutation status of the FA/BRCA-HRR genes, gene expression profiling (GEP) and Gene Set Enrichment Analysis (GSEA) were performed using groups of samples with and without mutations in each of the subgroups in the extended FA/BRCA-HRR network (see **Table 3.1** for subgroups). GEP was performed for a total of 57 AML samples and 13 samples from healthy controls (BMMNC) using the Ilumina_HumanHT_12_v4 chip as detailed in **Chapter 4** (**Section 4.2; Supplementary methods**). The initial analysis and data processing was performed with the help from Stephen Pederson and Mahmoud Bassal.

Since only a subset of the samples with WES data had been included in the GEP, smaller subgroups of the FA/BRCA-HRR network (where n<14) were excluded from the analysis. The 5 groups included were the extended FA/BRCA-HRR network (n=42), FANC 19 (n=22), FANC core & ID2 (n=14), BLM & ATM/ATR Checkpoint proteins (n=23) and KEGG-HRR (n=14). Lists of differentially expressed genes for each subgroup relative to the non-mutant

AML group were determined as outlined in **Chapter 4** (**Section 4.2; Supplementary methods**). The top 50 differentially expressed genes with a 1.5 fold change ($\log_2 \pm 0.6$) and adjusted p-value <0.05 between the mutant subgroup and the relevant non-mutant samples are shown in **Appendix V**. The differentially expressed gene list (ranked using t-statistic) for each comparison was used to perform GSEA using *GSEA Desktop v3.0 Beta 2* from the Broad Institute. For this, the gene expression signature of each gene set was compared to that of the publically available gene sets from the MSigDB. For the subgroups included in this analysis the top 50 (positively and negatively enriched) gene sets with a false discovery rate of $\leq$25% for each of the subgroups is shown in **Appendix W and Appendix X** and the GSEA plots which are discussed further in this chapter are shown in **Figures 3.5-3.7**.

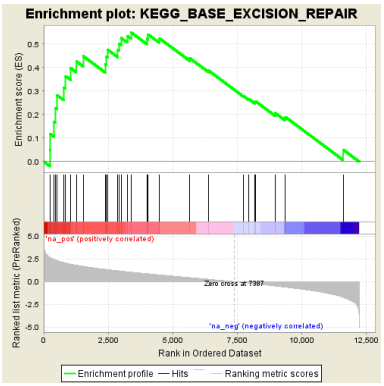
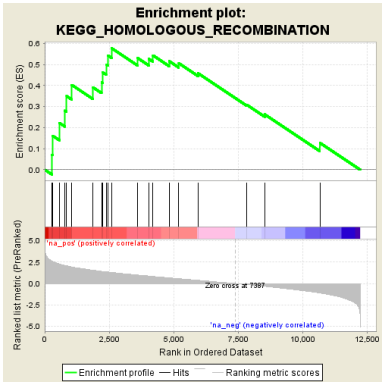
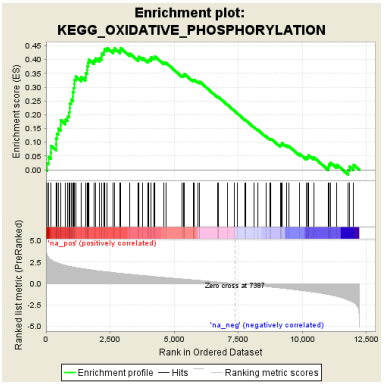
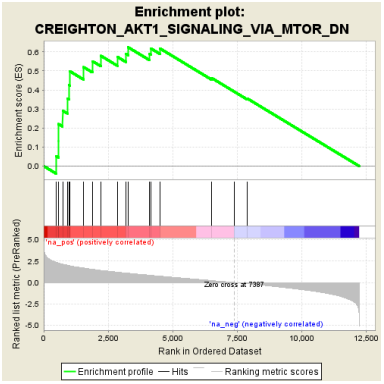
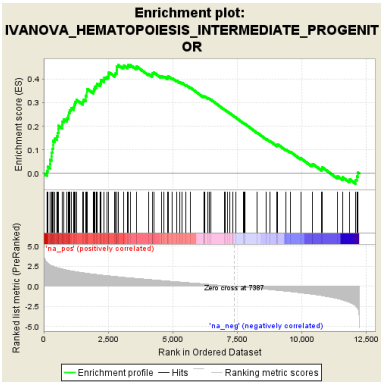
(A) 	**(B)** 	**(C)** 
NES=1.79.3; FDR q-val=0.054; NOM p-val= 0.001	NES=1.726; FDR q-val=0.077; NOM p-val=0.004	NES=1.796; FDR q-val=0.055; NOM p-val=0.000
(D) 	**(E)** 	
NES=1.870; FDR q-val=0.050; NOM p-val=0.000	NES=1.908; FDR q-val=0.044; NOM p-val=0.000	

Figure 3.5. GSEA Plots of positively and negatively correlated gene signatures for the FANC 19 mutant subgroup of AML patients. (A) A gene signature relating to increased BER. **(B)** A gene signature relating to increased HRR. **(C)** A gene signature relating to increased oxidative phosphorylation. **(D)** A gene signature relating to increased AKT signalling. **(E)** A gene signature that is observed in haematopoietic intermediate progenitors. (NES = normalised enrichment score; FDR q-val = false discovery rate q-value; NOM p-val = nominal p-value).

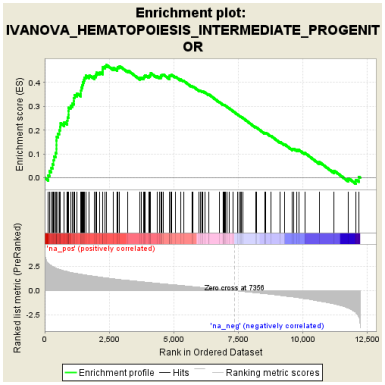
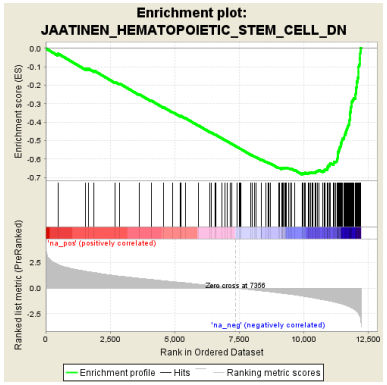
(A) 	**(B)** 
NES=1.975; FDR q-val=0.067; NOM p-val=0.000	NES=-3.756; FDR q-val=0.000; NOM p-val=0.000

Figure 3.6. GSEA Plots of positively and negatively correlated gene signatures for the BLM & ATM/ATR checkpoint subgroup of AML patients. (A) A gene signature that is observed in haematopoietic intermediate progenitors. **(B)** A gene signature of genes downregulated in $CD133^{+}$ stem cell compartment. (NES = normalised enrichment score; FDR q-val = false discovery rate q-value; NOM p-val = nominal p-value).

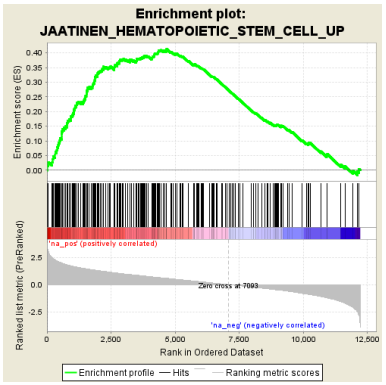
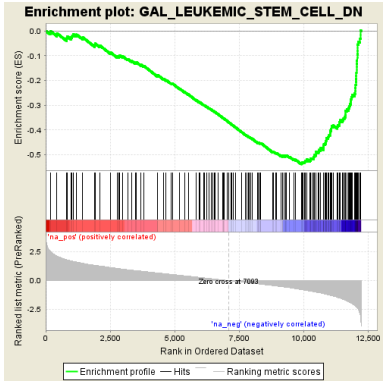
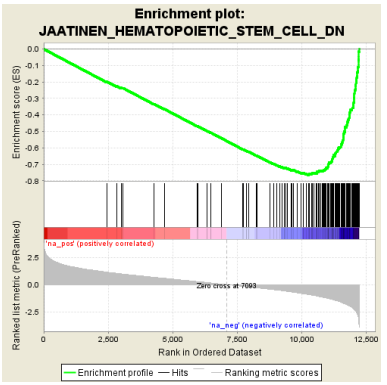
(A) 	**(B)** 	**(C)** 
NES=1.954; FDR q-val=0.102; NOM p-val=0.000	NES=-2.908; FDR q-val=0.000; NOM p-val=0.000	NES=-4.038; FDR q-val=0.100; NOM p-val=0.000

Figure 3.7. GSEA Plots of positively and negatively correlated gene signatures for the KEGG-HRR mutant subgroup of AML patients. (A) A gene signature that is observed in $CD133^+$ stem cells. **(B)** A gene signature observed in leukaemic stem cells. **(C)** A gene signature of genes downregulated in $CD133^+$ stem cell compartment. (NES = normalised enrichment score; FDR q-val = false discovery rate q-value; NOM p-val = nominal p-value).

3.3 Discussion

The resolution of ICL and stabilisation of replication forks is critical for maintaining genomic integrity (Joo et al., 2011, Longerich et al., 2014, Walden and Deans, 2014) and this process has been demonstrated to be critically important in HSC (see **Section 3.1**). The FANC genes which are critical for this process were used to build an extended FA/BRCA-HRR network of genes, which allowed for rare predicted damaging variants to be characterised across a cohort of AML samples. Case-control comparisons have also been used to define genes in the FA/BRCA-HRR network that were associated with rare predicted damaging variants occurring at increased frequency in AML in comparison to healthy controls. In this discussion, the focus is on the characteristics of the groups of patient samples carrying rare damaging variants affecting functional subgroups in the FA/BRCA-HRR network. The subgroups of the extended FA/BRCA-HRR network that were large enough to permit meaningful analysis were FANC 19, FANC core & ID2, BLM & ATM/ATR checkpoint, and KEGG-HRR. The interpretations below need to be considered in the context of the small size of the subgroups used; however, a number of interesting characteristics that warrant discussion and further investigation were observed. These findings are discussed further in **Chapter 4**. **Chapter 6** provides an overall synthesis of the results and presents models to explain the mechanism underlying a potential predisposition phenotype associated with rare variants affecting selected genes in this network.

3.3.1 The extended FA/BRCA-HRR network and its subgroups

3.3.1.1 The extended FA/BRCA-HRR network

For the group of samples carrying variants across the extended FA/BRCA-HRR network of genes, presence of a FA/BRCA-HRR network variant was statistically associated with increased karyotypic abnormalities, as shown in **Table 3.7** and **Figure 3.3**. As explained previously, it is predicted that the majority of these rare variants are germline, hence, it is proposed that overall, the data is consistent with a model in which subtle impairment of this DNA repair network through germline heterozygous mutations may result an increase in genomic instability.

In the gene expression analysis, succinate receptor 1 (*SUCNR1*) and aldehyde dehydrogenase 1 family member A1 (*ALDH1A1*) were observed as the most differentially expressed genes (*SUCNR1*: $\log_2$ of 1.302) (*ALDH1A1*: $\log_2$ of -1.566) for this large group of samples in comparison to the remaining 15 samples that had no variants in the extended FA/BRCA-HRR

network. SUCNR1, also referred to as GPR91, is a G-protein-coupled receptor that has the citric acid cycle intermediate and oncometabolite, succinate, as a ligand (de Castro Fonseca et al., 2016, Ariza et al., 2012). The activation of SUCNR1 can result in the activation of multiple signalling pathways including the MAP/ERK and PI3K/AKT signalling pathways (Gilissen et al., 2016). According to the gene expression database, BloodSpot (Bagger et al., 2016), *SUCNR1* is expressed at high levels in normal haematopoietic intermediate progenitors (GMP and CMP) with lower expression in the primitive HSPC and mature leukocytes. Conversely, *ALDH1A* is highly expressed in HSPC to maintain genomic integrity (Tomita et al., 2016), while displaying down-regulation in mature leukocyte populations. GSEA also points to differential expression of genes associated with stem and progenitor cell populations in the different AML subgroups, and these results are discussed further below.

3.3.1.2 FANC 19 subgroup.

The samples with mutations in the 19 FANC genes (FANC 19 subgroup) was associated with a decreased frequency of the *FLT3*-ITD mutation. For this subgroup *SUCNR1* and *ALDH1A1* were also two of the most differentially expressed genes (*SUCNR1*: $\log_2$ of 1.248) (*ALDH1A1*: $\log_2$ of -0.879) in comparison to the remaining AML samples. The results of the GSEA for the FANC 19 mutant subgroup (**Figure 3.5**) were very similar to that of the FANC core & ID2 complex mutant subgroup discussed in **Chapter 4**. For both groups, positive enrichment for gene-sets associated with BER and HRR signatures were observed, which may suggest a compensatory mechanism involving up-regulation of DNA repair pathways in response to partial loss/defect in the FA DNA repair pathway. Interestingly, the normalised enrichment scores (NES) for the HRR gene set signature was lower with the FANC 19 mutant subgroup than that of the FANC core & ID2 complex mutant subgroup. This may be related to a slightly reduced capacity of HRR for the FANC 19 mutant subgroup which contains samples with variants affecting *BRCA1*, *BRCA2*, *BRIP1*, *PALB2*, *RAD51C* and *RAD51,* which are central to HRR. For this FANC 19 subgroup, GSEA also showed positive enrichment for gene-sets associated with AKT activation, which was not observed for the other subgroups. Hyperactivation of AKT is indicative of survival and proliferation signalling, as well as changes in metabolism (Wang et al., 2017). Positive enrichment was also observed for gene-sets associated with oxidative phosphorylation suggesting that there may be an increased energy requirement in the FANC 19 mutant subgroup, possibly due to the increased use of DNA damage repair (DDR) pathways. Finally, the enrichment of gene-sets associated with

intermediate progenitors for this subgroup maybe suggestive of leukaemic initiation from cells at a particular stage of haematopoietic differentiation. This concept is discussed further in **Chapter 4** with focus on the FANC core & ID2 mutant samples, versus non-mutant samples.

3.3.1.3 BLM & ATM/ATR checkpoint subgroup.

BLM & ATM/ATR checkpoint subgroup was another large subgroup within the extended FA/BRCA-HRR network, and for these samples, *SUCNR1* and *ALDH1A1* were also observed as two of the most differentially expressed genes (*SUCNR1*: $\log_2$ of 1.134); *ALDH1A1*: $\log_2$ of -1.317) in comparison to the remaining AML samples. *ELANE*, which displays a profound increase in mRNA expression in intermediate progenitors, was also upregulated selectively in this subgroup of mutant samples ($\log_2$ of 1.04). Gene expression analysis showed that there were several similarities between the BLM & ATM/ATR subgroup and the FANC 19 subgroup, some of which is explained by the presence of 9 samples that are in common to both of these subgroups. A gene-set that is differentially down-regulated in cord blood stem cells (i.e. shows decreased expression in the $CD133^+$ stem cell compartment relative to $CD133^-$ cord blood progenitors) was also negatively enriched for this subgroup (**Figure 3.6,** and the FANC19 subgroup of samples). This is consistent with a gene expression pattern for these groups of samples that is more reflective of progenitor cell gene expression versus that of stem cells.

3.3.1.4 KEGG-HRR subgroup.

For the KEGG-HRR sub-group consisting of samples with mutations affecting breast cancer associated genes, the MRN complex and the RAD51 paralogues genes, a significant increase in karyotypic abnormalities was observed. This was not observed for the FANC 19, FANC core & ID2 complex mutant samples, or the BLM & ATM/ATR checkpoint mutant samples. This group also showed a significant increase in samples with *MLL* translocations.

Interestingly, for this KEGG-HRR mutant subgroup, the two most differentially expressed genes relative to non-mutant samples, were the erythrocyte markers, haemoglobin subunit gamma 1 and 2 (*HBG1*: log2 of -2.5; *HBG2*: log2 of -2.4). CD14 and CD36, both of which are highly expressed on monocytes (Bloodspot; (Bagger et al., 2016)), were also differentially down-regulated in this subgroup (*CD14*: $\log_2$ of -1.48; *CD36*: $\log_2$ of -1.31). This relative down-regulation of lineage genes, together with up-regulation of the *HOXA3* gene ($\log_2$ of 1.123),

which is strongly up-regulated in HSC (Bloodspot; (Bagger et al., 2016)), suggests that the gene expression program for this subgroup may be reflective of cells at the uncommitted progenitor or stem cell stage. The enrichment of stem cell associated gene-sets with the gene expression for this group revealed by GSEA (**Figure 3.7)**, is also consistent with this proposal. Thus, the leukaemic clones for KEGG-HRR mutant subgroup of AML are possibly derived from a more a primitive progenitor cell than the previously mentioned subgroups.

3.3.2 Summary, strengths and limitations of this analysis

Analysis of individual genes and GSEA suggests a model summarised in **Table 3.10**, with alternative cell types as the potential cell of origin for leukaemia associated with variants from different subgroups. Thus, it is speculated that the KEGG-HRR mutant AML is derived from the primitive uncommitted HSPC, with AML for the FANC 19, and BLM & ATM/ATR checkpoint subgroups arising from cells with characteristics of intermediate progenitors. Such a model would be consistent with the critical role of DNA repair genes in the primitive HSC compartment (see **Section 3.1**). As discussed in **Chapter 4**, analysis of gene expression for the subgroup of AML with FANC core & ID2 complex mutations also revealed unique characteristics, some of which may also relate to cell of origin. Further studies to confirm the expression differences between subgroups are now needed, and could most readily be performed from an RNA-sequencing analysis (RNA-Seq) of AML samples with mutation status for these genes determined. It is important to note that some of the differences observed in the gene expression profiles for these individual subgroups may relate to karyotypic abnormalities and common AML mutations that are differentially associated with each subgroup.

A strength of this study was the size of the adult AML cohort sequenced by WES, and the ability to link mutation status for these genes to AML clinical and molecular characteristics, and gene expression patterns. The unique access to an ethnically-matched healthy Australian control cohort, sequenced and analysed using the same WES platform, provided an important tool for further filtering, removing variants that are associated with SNPs preferentially enriched due to geographical or ethnic cohort bias. While this healthy cohort also allowed Burden analysis to be performed, approaches were needed to account for it being an all-female cohort (i.e. mutation data on the X chromosome was removed, including variants for FANCB). However, as the AML samples that were sequenced were not paired (i.e. non-disease sample was not available), it was not possible to determine germline/somatic status of the variants using

bioinformatics analysis (discussed further in **Chapter 6**). The multi-centre nature of this study was associated with some difficulties for access to clinical information, which was more readily available for the Adelaide samples than for those from interstate, namely the survival data and the family history. Samples of Asian descent were excluded for a number of the analyses as it was observed that these exhibited a higher background of variants than the Caucasian samples, a phenomenon related to ethnic bias associated with the samples in publically available SNP databases used for data filtering. Given this, it is important to consider that conclusions from these analyses can only be applied to the Caucasian population.

In conclusion, the results in this chapter have described the generation of a network of functionally related proteins built from the FANC genes, and the identification and characterisation of damaging variants across this network in adult AML and healthy control cohorts. The data in the chapter support that hypothesis that rare DNA repair gene variants associated with the FA/BRCA-HRR network are enriched in AML compared to controls. Even though no single variant is strongly associated with risk of AML, it is predicted that incomplete penetrance occurs in families, and only large families or collection of families may only reveal this association.

Table 3.10. Proposed leukaemic cell of origin in FA/BRCA-HRR network subgroups

Proposed leukaemic cell of origin	Gene Expression Analysis		FA/BRCA-HRR network subgroup
	Key differentially Expressed genes	GSEA signatures	
HSC compartment	*CD34*↑	BER↑ HRR↑ Replicative stress↑	FANC core & ID2
Uncommitted HSPC	*CD14*↓ *CD36*↓ *HBG1*↓ *HBG2*↓	Stem cell↓	KEGG-HRR
Intermediate progenitor compartment	*SUCNR1*↑ *ALDH1A*↓	BER↑ HRR↑ ATP synthesis↑	FANC 19
	SUCNR1↑ *ALDH1A*↓	Stem cell↓	BLM & ATM/ATR checkpoint

Chapter 4 Manuscript - Rare variants in Fanconi Anaemia complex genes increase risk for Acute Myeloid Leukaemia

4.1 Summary

In this chapter, a comprehensive analysis of the variants identified for FANC genes is presented in the attached manuscript (**Section 4.2**). The emphasis of the results presented in this manuscript is rare variants identified in the FANC core and ID2 complex genes, the products of which are critical for ICL repair and stabilisation of the replication fork (Schlacher et al., 2012). This focused study stemmed from the genomic analysis of the DNA repair gene network built from the Fanconi Anaemia (FA) DNA repair pathway (**Chapter 3**).

As of 2017, the list of FANC genes now numbers 21 with the inclusion of a subgroup consisting of the breast cancer associated genes (*BRCA1*, *BRCA2*, *BRIP1* and *PALB2*), as well as RAD51 recombinase and its paralogues (*RAD51B*, *RAD51C* and *XRCC2*) (Ameziane et al., 2015, Sawyer et al., 2015, Hira et al., 2015, Park et al., 2016, Bluteau et al., 2016). The products of the *BRCA1* and *BRCA2* genes, which are central to homologous recombination repair (HRR), have well-characterised roles as classical dominant, tumour suppressor genes, with tumour formation commonly associated with loss of both wild-type allele (Roy et al., 2011). Recent studies have also shown a haploinsufficient phenotype associated with *BRCA1* and *BRCA2*, which may also contribute to the predisposition to solid cancers (Buchholz et al., 2002, Jeng et al., 2007, Feilotter et al., 2014, Pathania et al., 2014, Sedic and Kuperwasser, 2016, Tan et al., 2017). In contrast, heterozygous FANC core and ID2 complex mutations, or FA carriers, do not have well defined association with cancer risk. The results in this chapter raise an important question regarding the significance of heterozygous damaging and disease-causing variants in the FANC core and ID2 complex genes; these issues form the basis of the discussion in the manuscript.

A key aspect of the manuscript is the identification of a number of rare disease-causing mutations of the FANC core and ID2 complex genes that were present in the AML cohort. Several of these variants are of specific interest and are discussed in detail. A number of these variants have also been validated by Sanger sequencing, (**Appendix D**) which has shown that the large majority were germline variants; hence, discussion is focused on the significance of such germline variants for AML predisposition, while acknowledging that some of these

variants may be somatically acquired. Approaches are also described to investigate enrichment of these rare FANC core and ID2 complex mutations, and known disease-causing mutations, in the AML cohort compared to healthy control cohorts. Finally, gene expression approaches are described to further investigate the characteristics of AML association with FANC core and ID2 complex mutations.

The limited evidence for a haploinsufficient phenotype associated with FANC core and ID2 complex genes is outlined in the manuscript. However to address this issue definitively a cell line model of *FANCL* haploinsufficiency was generated to allow a suite of assays to be performed to measure basal and induced DNA repair response to ICL-induced damage (such as FANCD2 monoubiquitination and cell cycle analysis).

4.2 Manuscript - Rare variants in Fanconi Anaemia complex genes increase risk for Acute Myeloid Leukaemia

Statement of Authorship

Title of Paper	Rare variants in Fanconi Anemia complex genes increase risk for Acute Myeloid Leukemia
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	PLOS Genetics

Principal Author

Name of Principal Author (Candidate)	Kyaw Ze Ya Maung		
Contribution to the Paper	Performed the basis of the research, analysed and interpreted the whole exome data and wrote the publication		
Overall percentage (%)	20%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Paul Leo		
Contribution to the Paper	20%. Performed and analysed the raw whole exome data and wrote the methodology of the sequencing experiment.		
Signature		Date	

Statement of Authorship

Title of Paper	Rare variants in Fanconi Anemia complex genes increase risk for Acute Myeloid Leukemia
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	PLOS Genetics

Principal Author

Name of Principal Author (Candidate)	Kyaw Ze Ya Maung		
Contribution to the Paper	Performed the basis of the research, analysed and interpreted the whole exome data and wrote the publication		
Overall percentage (%)	20%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	30. 7. 2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Paul Leo		
Contribution to the Paper	20%. Performed and analysed the raw whole exome data and wrote the methodology of the sequencing experiment.		
Signature		Date	

Statement of Authorship

Title of Paper	Rare variants in Fanconi Anemia complex genes increase risk for Acute Myeloid Leukemia
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	PLOS Genetics

Principal Author

Name of Principal Author (Candidate)	Kyaw Ze Ya Maung		
Contribution to the Paper	Performed the basis of the research, analysed and interpreted the whole exome data and wrote the publication		
Overall percentage (%)	20%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Paul Leo		
Contribution to the Paper	20%. Performed and analysed the raw whole exome data and wrote the methodology of the sequencing experiment.		
Signature		Date	25/7/17

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	20/7/17

Name of Co-Author	James Gray		
Contribution to the Paper	3%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Sarah Bray		
Contribution to the Paper	2%. Supervisor of the student and edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Stephen Pederson		
Contribution to the Paper	1%. Analysed the raw microarray data.		
Signature		Date	

Name of Co-Author	Deepak Singal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	3%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	20 JULY 2017

Name of Co-Author	Debora Casolari		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Sarah Bray		
Contribution to the Paper	2%. Supervisor of the student and edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Stephen Pederson		
Contribution to the Paper	1%. Analysed the raw microarray data.		
Signature		Date	

Name of Co-Author	Deepak Singal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	3%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	21/07/17

Name of Co-Author	Sarah Bray		
Contribution to the Paper	2%. Supervisor of the student and edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Stephen Pederson		
Contribution to the Paper	1%. Analysed the raw microarray data.		
Signature		Date	

Name of Co-Author	Deepak Singal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	3%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Sarah Bray		
Contribution to the Paper	2%. Supervisor of the student and edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	27-07-2017

Name of Co-Author	Stephen Pederson		
Contribution to the Paper	1%. Analysed the raw microarray data.		
Signature		Date	

Name of Co-Author	Deepak Singhal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	3%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Sarah Bray		
Contribution to the Paper	2%. Supervisor of the student and edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Stephen Pederson		
Contribution to the Paper	1%. Analysed the raw microarray data.		
Signature		Date	20/07/2017

Name of Co-Author	Deepak Singal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	3%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Sarah Bray		
Contribution to the Paper	2%. Supervisor of the student and edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Stephen Pederson		
Contribution to the Paper	1%. Analysed the raw microarray data.		
Signature		Date	

Name of Co-Author	Deepak Singhal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	21/07/2022

Name of Co-Author	Saumya Samaraweera		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	21/7/17

Name of Co-Author	Tran Nguyen		
Contribution to the Paper	1%. Sequenced the cDNA of FANC gene variants.		
Signature		Date	21/07/2017

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	

Name of Co-Author	Adam Ewing		
Contribution to the Paper	1%. Obtained access to the TCGA WES dataset.		
Signature		Date	

Name of Co-Author	Amanda Smith		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	

Name of Co-Author	Matt Brown		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Saumya Samaraweera		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Tran Nguyen		
Contribution to the Paper	1%. Sequenced the cDNA of FANC gene variants.		
Signature		Date	

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	27/7/17

Name of Co-Author	Adam Ewing		
Contribution to the Paper	1%. Obtained access to the TCGA WES dataset.		
Signature		Date	

Name of Co-Author	Amanda Smith		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	

Name of Co-Author	Matt Brown		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Saumya Samaraweera		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Tran Nguyen		
Contribution to the Paper	1%. Sequenced the cDNA of FANC gene variants.		
Signature		Date	

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	

Name of Co-Author	Adam Ewing		
Contribution to the Paper	1%. Obtained access to the TCGA WES dataset.		
Signature		Date	27/7/2017

Name of Co-Author	Amanda Smith		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	

Name of Co-Author	Matt Brown		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	26/07/2017

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	

Name of Co-Author	Luen Bik To		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Ian Lewis		
Contribution to the Paper	2%. Provided clinical information and corrdinated the collection of the patient samples NB. Signature for Matt Brown		
Signature		Date	26.07.17

Name of Co-Author	Saumya Samaraweera		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Tran Nguyen		
Contribution to the Paper	1%. Sequenced the cDNA of FANC gene variants.		
Signature		Date	

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	

Name of Co-Author	Adam Ewing		
Contribution to the Paper	1%. Obtained access to the TCGA WES dataset.		
Signature		Date	

Name of Co-Author	Amanda Smith NB. Not an Author on this manuscript		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	

Name of Co-Author	Matt Brown		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort. See Previous page		
Signature		Date	

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	20/1/17

Name of Co-Author	Luen Bik To		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Ian Lewis		
Contribution to the Paper	2%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	

Name of Co-Author	Luen Bik To		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	25.7.17

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Ian Lewis		
Contribution to the Paper	2%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	

Name of Co-Author	Luen Bik To		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	20/7/17

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Ian Lewis		
Contribution to the Paper	2%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	

Name of Co-Author	Luen Bik To		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	25/7/17

Name of Co-Author	Ian Lewis		
Contribution to the Paper	2%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	

Name of Co-Author	Luen Bik To		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Ian Lewis		
Contribution to the Paper	2%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	20 JULY 2017

Name of Co-Author	Andrew Deans		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	21 July 2017

Name of Co-Author	Anna Brown		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Richard D'Andrea		
Contribution to the Paper	15%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Thomas Gonda		
Contribution to the Paper	15%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Andrew Deans		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Anna Brown		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	24/7/17

Name of Co-Author	Richard D'Andrea		
Contribution to the Paper	15%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Thomas Gonda		
Contribution to the Paper	15%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Andrew Deans		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Anna Brown		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Richard D'Andrea		
Contribution to the Paper	15%. Contributing to the editing and writing of the manuscript.		
Signature		Date	31/7/17

Name of Co-Author	Thomas Gonda		
Contribution to the Paper	15%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Andrew Deans		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Anna Brown		
Contribution to the Paper	3%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Richard D'Andrea		
Contribution to the Paper	15%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Thomas Gonda		
Contribution to the Paper	15%. Contributing to the editing and writing of the manuscript.		
Signature		Date	20th July 2017

Statement of Authorship (additional authors for revised manuscript)

Name of Co-Author	Vinay Tergaonkar		
Contribution to the Paper	Provided input on how to improve the gene expression analysis including the design of experiments and acquisition of new data. Provided critical review of the manuscript with important comments/revisions, and approval of the final manuscript for publication.		
Signature		Date	13/12/2017

Name of Co-Author	Gokhan Cildir		
Contribution to the Paper	Provided input on how to improve the gene expression analysis including the design of experiments and acquisition of new data. Provided critical review of the manuscript with important comments/revisions, and approval of the final manuscript for publication.		
Signature		Date	13/12/2017

Rare variants in Fanconi Anaemia complex genes increase risk for Acute Myeloid Leukaemia

Kyaw Ze Ya Maung[1,2,3,4*], Paul J. Leo[5*], Mahmoud Bassal[4,6,7], Debora A. Casolari[2,4,7], James X Gray[1,2,3], Sarah C. Bray[2,4,7], Stephen Pederson[8], Deepak Singhal[2,4,7], Saumya E. Samaraweera[4,7], Tran Nguyen[4,7], Gökhan Cildir[4], Mhairi Marshall[5], Adam Ewing[9], Emma L. Duncan[5,10,11], Matthew A. Brown[5], Russell Saal[5], Vinay Tergaonkar[4,12], Luen Bik To[2,7], Paula Marlton[12], Devinder Gill[13], Ian Lewis[2,4,7], Andrew J Deans[14], Anna L Brown[4,15], Richard J D'Andrea[2,4,6,7**], Thomas J Gonda[16, 17**]

*MKZY and PL contributed equally to this work

**RJD and TJG contributed equally as senior authors

Affiliations:

[1]Basil Hetzel Institute for Translational Health Research, The Queen Elizabeth Hospital, Woodville, South Australia, Australia

[2]School of Medicine, University of Adelaide, Adelaide South Australia

[3]Department of Haematology and Oncology, the Queen Elizabeth Hospital, Woodville, South Australia, Australia

[4]Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, South Australia

[5]Institute of Health and Biomedical Innovation, Queensland University of Technology and University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia

[6]School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia, Australia

[7]Departments of Haematology, SA Pathology and Royal Adelaide Hospital, Adelaide, South Australia

[8]Adelaide Bioinformatics, School of Biological Sciences and School of Mathematics, University Of Adelaide, Adelaide, Australia

[9]Mater Research Institute, University of Queensland, Brisbane, Queensland, Australia

[10]Department of Endocrinology, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia

[11]School of Medicine, University of Queensland, Brisbane, Queensland, Australia

[12]Institute of Molecular and Cell Biology, NUS School of Medicine, Singapore

[13]Clinical Haematology, Princess Alexandra Hospital Brisbane, Queensland, Australia

[14]St Vincent's Institute of Medical Research, Melbourne, Victoria, Australia

[15]Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia

[16]Sansom Institute for Health Research Operations, University of South Australia, Australia

[17]School of Pharmacy, University of Queensland, Brisbane, Queensland, Australia

Correspondence:

Professor Richard D'Andrea: School of Pharmacy and Medical Sciences, University of South Australia, GPO Box 2471, Adelaide 5001, Australia. Email: richard.dandrea@unisa.edu.au

Professor Tom Gonda: Sansom Institute for Health Research, University of South Australia, City East Campus, GPO Box 2471, Adelaide 5001, Australia E-mail: Thomas.Gonda@unisa.edu.au

Keywords: Leukaemia, DNA repair, acute myeloid leukaemia, cancer risk, Fanconi Anaemia

Running Title: Genetic variants affecting FANC genes in AML

Word count: Abstract: 281 words; Text: 5243 words

ABSTRACT

The bone marrow failure syndrome Fanconi Anaemia (FA) is caused by inherited bi-allelic mutations affecting 22 FANC DNA repair genes, and is associated with profoundly increased risk of developing Acute Myeloid Leukaemia (AML). We performed a whole exome sequencing study and a focused analysis of rare, deleterious FANC gene mutations in 131 Australian Caucasian adult AML cases and 323 healthy controls. We identified 53 rare, predicted deleterious, FANC gene variants in 45 of the AML cases. For the FANC core complex genes *FANCM*, *FANCL* and *FANCC*, we observed significantly increased mutation frequency in AML relative to publicly-available non-cancer databases. We observed an elevated frequency of known disease-causing (D-C) FANC mutations in the AML cohort in comparison to that reported for healthy European-Americans (OR=3.4, 1.7-7.0; *P*=0.0008), and compared to a gender-matched healthy Australian cohort (OR=3.3, 1.3-8.6; *P*=0.018). For AML with FANC core and ID2 variants, gene expression analysis indicated activation of DNA synthesis and repair, and ATR replicative stress signatures. We show increased expression of *POLD2*, *POLD3*, *POLA2*, and *RPA1* and *RPA2* in these AML samples, consistent with increased replicative stress. CRISPR-mediated generation of independent clonal cell lines with heterozygous or bi-allelic *FANCL* deletion showed intermediate levels of FANCD2 foci formation in response to mitomycin C for the heterozygous clones, compared to wild type clones. Taken together, these data provide compelling evidence that rare heterozygous germline FANC gene variants confer a subtle phenotype associated with altered FA DNA repair pathway activity, which may explain the observed modest increased AML risk. Thus, these findings raise important biological, ethical and clinical considerations and highlight the need for further prospective studies to confirm the predictive significance for AML development and to guide the counselling process.

AUTHOR SUMMARY

In a cohort of sporadic adult AML we report an increased prevalence of rare deleterious mutations affecting selected genes associated with Fanconi Anaemia (FA). Our data suggests that there is a modest increased risk of AML for individuals carrying a mutation reported to cause FA. Based on these findings, and functional studies in a *FANCL*-deficient cell line model, we propose that rare damaging variants in a subset of genes which cause FA is sufficient to reduce FA pathway DNA repair efficiency in normal blood cell precursors, hence predisposing to malignant transformation. This study has important implications for genetic counselling of FA families, and also raises the possibility that AML patients carrying such mutations may respond to current therapeutic agents that target defects in DNA repair pathways.

INTRODUCTION

Acute Myeloid Leukaemia (AML) is an aggressive haematological malignancy caused by somatically-acquired structural rearrangements, single-nucleotide variants (SNV) and insertions/deletions (indels) affecting a well-defined set of genes (1). While AML is recognized as a heterogeneous disease current treatment still relies primarily on cytotoxic chemotherapy with anthracycline and cytarabine, although a number of targeted therapies are at varying stages of clinical development for specific subtypes (2). The contribution of germline variants to AML initiation and progression is incompletely established. High-risk variants affecting transcription factors, e.g. *RUNX1*, *CEBPA*, *ETV6*, *GATA2*, and *DDX41*, account for a proportion of myelodysplastic syndrome and AML associated with a family history, but the contribution of other germline variants conferring low-intermediate risk has not yet been determined. Such variants are more difficult to identify from pedigree analysis due to low penetrance (3-5).

Fanconi Anaemia (FA) is predominantly a rare recessive bone marrow failure (BMF) syndrome caused by bi-allelic germline mutations in any of the 22 FANC genes (exceptions are *FANCB* and *FANCR* which are X-linked and autosomal dominant respectively) (6). Progression from BMF to AML occurs in approximately 10% of patients (7), with the associated risk being ≥800 times higher than the population average (8). FA pathogenesis is due to hematopoietic stem cell (HSC) dysfunction, which begins in utero with disease manifesting early in life as BMF (9). The proteins encoded by the FANC genes participate, in conjunction with the BRCA1/2-mediated homologous recombination repair pathway (FA/BRCA-HRR), in the removal of interstrand crosslinks (ICL) and the protection and resolution of stalled replication forks, an essential step for faithful DNA replication (10), (11). This group of proteins can be broadly classified into subgroups based on their functions in the pathway; these include the FA core complex, the ID2 complex (FANCI/FANCD2), and downstream homologous repair (HRR) proteins (10, 12). Several of these downstream FANC genes are well-established high-risk

familial cancer genes; e.g. the tumour suppressor and cancer pre-disposition genes *BRCA1/FANCS, BRCA2/FANCD1, BRIP1/FANCJ, RAD51C/FANCO* and *PALB2/FANCN* (13). Moreover germline mutations in *BRCA1* and *BRCA2* are found in ~20% of patients with therapy-related leukaemias (14, 15), highlighting that carriers in these FA families should be carefully monitored after cancer treatment.

Ubiquitination of the ID2 complex by the FANC core complex occurs in response to stalled replication, such as that caused by nucleotide depletion and by interstrand crosslink (ICL) inducing chemotherapies or endogenous aldehydes (16, 17). This is the critical first step for activation of the FA pathway, repair of the damage via HRR, and restart of replication. The importance of FA core and ID2 complex function in HSC is highlighted by: ***(i)*** the DNA damage and collapse of the haematopoietic system that are direct consequences of replicative stress associated with inducing HSCs to exit quiescence in *Fanca*-deficient mice (18); and ***(ii)*** the increased genomic instability, and leukaemic transformation, caused by reactive aldehydes when HSC are deficient in both the FA pathway and alcohol dehydrogenase activity (i.e. *Fancd2*$^{-/-}$*Adh5*$^{-/-}$) (16).

Given the extreme susceptibility of FA patients to AML, a long-standing question in the field has been the risk of AML associated with heterozygous disease-causing (D-C) FANC gene mutations. This question is clearly of great significance for individuals known to carry D-C mutations in any of the FANC genes [estimated frequency from population studies to be 4.3%; ref (19)], and for FA carriers [frequency from FA databases estimated to be approximately 0.6%; ref (20)]. Previous studies of FA families have not identified an overall increased incidence of AML or other cancers in FA carriers (21, 22), although a significantly higher than expected rate of breast cancer was observed among carrier grandmothers with *FANCC* mutations (21). There are several limitations to these previous FA familial studies that need to

be considered. Firstly, such familial studies are necessarily biased towards analysis of 3 out of the 22 FANC genes (i.e. *FANCA*, *FANCC* and *FANCG)* as these represent the vast majority of FA cases. Secondly, such studies address whether there is a high-risk of AML development for carriers of D-C FANC gene variants, while modest or intermediate risk variants in these genes would be predicted to induce AML with low penetrance in FA families, hence masking the familial pre-disposition. Thirdly, the age profile for AML also contributes to difficulty in detecting increased AML in FA families; AML is a disease that occurs with dramatically increased frequency in individuals > 70 year old, so unless pre-disposition is associated with early onset of disease, detection of familial patterns of inheritance can be masked due to death from other causes. Finally, it is also important to consider that severe loss of function for some FANC genes may not be tolerated at the germline level (i.e. those genes rarely found mutated in FA), while rare heterozygous mutations in these FA genes may still confer a cancer risk for carriers. Evidence for this is provided by the embryonic lethality of *Brca1* and *Brca2* null mice (23). In contrast to the FA familial studies, cancer cohort studies have revealed that heterozygous variants affecting the FA core and ID2 complex genes are associated with modestly increased cancer risk; deleterious mutations in *FANCA, FANCC, FANCG, FANCI, FANCL* and *FANCM* have been identified in cohorts of patients with familial pancreatic cancer, breast cancer, colorectal cancer, metastatic prostate cancer and paediatric cancer (13, 24-29). The recent observation that the FA core complex activity requires dimerization (30), suggests that dominant negative activity of particular variants could also be possible and potentially contribute to pathogenesis. FANC core complex mutant mice are viable and do not develop malignancy, but there is ample evidence that deficiency for these genes combined with other mutations results in pre-leukaemia or leukaemia (16, 31). Here, we have used a case-control approach to investigate the frequency and nature of rare gene variants across the 22 FANC genes, identified through whole-exome sequencing (WES) in ethnically matched Australian adult AML and control cohorts.

RESULTS

FANC mutations in AML

We performed WES on 131 samples from adult Caucasian AML patients (cohort characteristics shown in **Suppl. Table S1)** from two major Australian centres and identified novel and rare somatic and germline coding and splicing variants (Tier 1 mutations, ref (1), MAF<0.001). Variants were also identified in WES data from an ethnically-matched all-female control cohort of 323 unaffected individuals (32). The frequencies of known recurrent somatic AML mutations in our AML cohort and that from the TCGA (1) showed excellent concordance (r=0.886, **Suppl. Fig. S1**), providing validation of our sequencing and mutation calling methods, and demonstrating high sensitivity for variant detection. Following CADD filtering (33) for pathogenicity (CADD>10), we identified a total of 53 rare FANC gene mutations in 45 patients (34% of patients; 40 non-synonymous SNV, 6 indels, 5 stop-gain and 2 splicing; **Suppl. Table S2**). For FANC variants that were tested by Sanger sequencing results were consistent with heterozygosity in the tumour material. Where matched tumour and normal samples were sequenced, >90% of FANC variants were germline in origin (10 out of 11 variants, data not shown). These results are consistent with reports that somatic FANC gene mutations are extremely rare in AML [<1%, ref (1)]. As mutation can result in selective loss of allelic expression via various mechanisms, we next used Sanger sequencing of cDNA generated from diagnosis AML samples (n=13) to determine whether mutant and wild-type (WT) alleles are both expressed. For the majority of mutations tested both WT and mutant sequences were detected. For the mutations BRCA1 p.Y856H and FANCM p.K515N, we observed minimal or negligible expression of the variant **(Suppl. Fig. S2)**. Thus, for rare deleterious FANC core and ID2 gene variants, some of which are known D-C variants, haploinsufficiency may result. However, we cannot exclude that the WT allele could be expressed more highly to compensate for loss of expression of the mutant allele. The distribution of the FANC gene mutations across all AML samples is indicated in **Suppl. Fig. S3** and associations with clinical, cytogenetic and

188 molecular features of AML are shown in **Suppl. Table S1.** FANC variants are not themselves
189 mutually exclusive, with some patients having a second variant affecting a different gene.
190 Interestingly, monosomy 7/del(7q) and *FLT3*-ITD mutations were under-represented in the
191 FANC-mutant AML group (P=0.028 and P=0.05, respectively).

192

193 **D-C FANC mutations and FANC mutation frequency in AML**

194 For a stringent analysis of pathogenic FANC gene mutations, we next defined known D-C
195 mutations as those associated with FA or breast cancer. We identified these by cross-
196 referencing the AML FANC variants to those listed as confirmed pathogenic in the FA
197 (FAMutdb: http://www2.rockefeller.edu/fanconi/) and breast cancer (kConFab:
198 http://www.kconfab.org/Index.shtml and BIC: https://research.nhgri.nih.gov/bic/) databases
199 (**Table 1**), and then compared the frequency against our healthy cohort (32). As this healthy
200 control cohort was all-female, and there have been reports of over-representation of variants for
201 some FANC genes in females (34, 35), we limited this analysis to female AML patients only.
202 The overall frequency of female patients with rare deleterious FANC gene variants was similar
203 to the control female cohort; we identified 19 variants in 18 female patients (35% of female
204 patients) compared to 151 variants in 105 healthy female controls (32%). This analysis revealed
205 an elevated frequency of female AML cases carrying D-C mutations compared to the healthy
206 female cohort (13.7% AML compared to 4.5% controls; Odds Ratio (OR) =3.3, 1.3-8.6;
207 P=0.018). Similarly, we observed an increased frequency of D-C mutations across all of our
208 AML cases when comparing the 16 FANC genes analysed in a previous population study (19)
209 (6.9% of AML compared to 2.1% European-Americans in the ESP database, n=4298
210 individuals, OR=3.4, 1.7-7.0; P=0.002).

211 For independent validation of this finding, we used a pipeline and filtering analysis similar to
212 that described above to determine rare deleterious FANC variants for the TCGA AML cohort

213 (1). Overall, we observed a trend towards increased frequency of patients carrying D-C variants
214 in the TCGA AML cohort (4.9%) compared to the European-Americans in the ESP database,
215 although this did not reach significance (*P=0.068*).

216

217 AML is recognized as a highly heterogeneous disease and it is plausible that FANC D-C
218 mutations pre-dispose to transformation via oncogenic mechanisms that may be frequent in
219 particular subtypes. The potential for analysis restricted to AML subtypes is limited by cohort
220 size, however separate analysis of the Australian normal karyotype AML (NK-AML) group
221 also showed increased frequency of D-C mutations for NK-AML vs European-Americans in
222 the ESP database (*P*=0.0041, cohort characteristics shown in **Suppl. Table S3**). Further
223 differences in frequency of these variants within AML subtypes could be revealed through a
224 multi-centre study, and would be of interest given that a number of recurrent chromosomal
225 rearrangements have been reported to impact on HRR activity (36).

226

227 **Enrichment of rare FA core complex variants in AML.**

228 Comparison of the variant frequency for each of the 22 FANC genes in the Australian AML
229 cohort relative to that observed for similarly defined variants in the non-Finnish European
230 population in the Exome Aggregation Consortium (ExAC) database (MAF<0.001, CADD>10;
231 **Table 2**) revealed a significant increase in variants of the FA core complex genes *FANCM*,
232 *FANCL* and *FANCC* in AML. The mutations found in these three genes and the domains
233 affected are represented in **Fig. 1A**. Several of the FANC core and ID2 gene variants are of
234 significant interest based on known disease association or clinical characteristics (summarized
235 in **Suppl. Table S4**). Analysis of the NK-AML showed a significant increase of *FANCM*
236 variants in the AML cohort (*P*=0.0267) **(Suppl. Table S5).**

237 Taken together these data are consistent with a potential increased risk of AML associated with

238 rare, deleterious, heterozygous variants affecting multiple FANC genes, including those
239 encoding proteins that comprise the FANC core and ID2 complexes. An elevated frequency of
240 deleterious variants affecting *FANCC* (and *FANCO*) in AML was also observed using burden
241 analysis by Lu et al (28) who investigated truncating FANC variants in TCGA cohorts. We
242 performed similar analyses to compare frequency of all rare damaging (missense and truncated)
243 variants identified in both the TCGA and Australian AML cohorts to the healthy controls. After
244 adjusting for variants on the X and Y chromosome this confirmed the elevated frequency of
245 *FANCC* and *FANCO* damaging variants for the TCGA cohort (P=0.037, P=0.01 respectively),
246 and further showed an elevated frequency of FANCL variants in the Australian AML cohort
247 (P=0.01, **Fig. 1B**). To further investigate the potential significance of our findings, we have
248 undertaken studies to identify functional consequence of heterozygous germline FANC
249 mutations.

250

251 **Gene expression profiling indicates activation of DNA synthesis and repair in FANC core**
252 **and ID2 mutant AML**

253 First, we used microarray gene expression profiling to identify genes that were differentially
254 expressed between AML samples with or without rare damaging variants affecting the FANC
255 genes. Given the above data suggesting enrichment of D-C FANC core and ID2 complex
256 mutations in the AML cohort, we focused on comparison of the FANC core and ID2 mutant
257 AML group (n=14) to non-FANC core and ID2 mutant AML (n=43). The top-ranked
258 differentially expressed genes are listed in **Supp. Table S6**. Gene set enrichment analysis
259 (GSEA) of the differential expression profiles associated with the FANC core and ID2 complex
260 mutant subgroup revealed an enrichment for gene-sets associated with DNA synthesis and
261 repair (i.e. Base Excision Repair; BER), and ATR activation and replication stress (**Fig 2A;**
262 **Supp. Table S7**). The ATR kinase cascade is activated in response to a number of events which
263 includes ICL lesions and stalled replication forks, raising the possibility that FANC core and

ID2 mutant AML cells may accumulate stalled replication forks, which leads to increased activity of other DNA repair pathways involved in fork protection and repair (37). To measure changes in expression of individual GSEA leading-edge genes involved in these processes, we performed qRT-PCR for a number of genes associated with DNA synthesis and fork protection (**Suppl. Table S8**). As shown in **Fig. 2B** the increased expression of *POLD2*, *POLD3*, *POLA2*, and *RPA1* and *RPA2* is apparent in the FANC core and ID2-mutant samples compared to non-mutants, consistent with up-regulation of DNA synthesis and/or repair in the mutant samples. The GSEA analysis also indicated a reduced growth factor response signature associated with FANC core and ID2-mutant samples **(Fig. 2A; Supp. Table S7**). Given the heterogeneity of AML we also performed the gene expression profiling and GSEA in NK-AML samples. **Suppl. Table S9** shows the top-ranked differentially expressed genes from this comparison (mutant n=6 and non-mutant n=22). Although the number of samples in the FANC core and ID2 mutant NK-AML group is small, we still observed positive enrichment for DNA replication-associated signatures (e.g. lagging strand synthesis), consistent with the results above (**Suppl. Table S10**).

Changes to FA DNA repair pathway activity in a *FANCL* heterozygous cell line model.

To investigate further the changes associated with heterozygous damaging FANC variants, and to validate the findings from our AML gene expression profiling, we used CRISPR-Cas9 mediated mutagenesis to generate a cell line model of *FANCL* heterozygosity. MCF10A was the cell line of choice over other commonly used, or leukaemia-derived cell lines, because it was derived from non-tumorigenic tissue and is diploid (38). For these experiments, we compared the responses to DNA cross-linking agent mitomycin C (MMC) treatment in 3 independent MCF10A clones with small heterozygous deletions in exon 1 of *FANCL*, a clone with bi-allelic *FANCL* mutations, and two independent WT clones (**Suppl. Fig. S4**). These deletions resulted in either amino acid deletions or frameshifts and premature termination. The cell cycle distribution of the heterozygous clones did not differ from the WT, while the biallelic

mutant *FANCL* MCF10A clone displayed increased accumulation of cells in G2/M phase following MMC treatment (**Suppl. Fig. S5**), consistent with delayed repair of ICLs. We next measured formation of FANCD2 foci by immunofluorescence as a readout for activation of the ID2 complex through FANCL-mediated monoubiquitination. As shown in **Fig 3A** and **3B**, following 48 hours treatment with MMC the two WT MCF10A clones (P1D3 and P1A4) displayed robust FANCD2-foci formation in $\gamma H2AX^+$ cells, while this was not observed for the bi-allelic mutant clone (P1B4), consistent with loss of FANCL activity. Importantly, the frequency of $\gamma H2AX^+$ cells with FANCD2 foci for the three heterozygous clones (P2D3, P2B5 and P4A4) following MMC treatment was intermediate between the WT and bi-allelic clones; and for clone P4A4 there was a statistically significant reduction in these FANCD2 foci-positive cells relative to the WT clone (P1A4) (**Fig. 3B**). We did not observe a difference in the frequency of γH2AX-positive cells relative to WT clones for any of the MMC-treated heterozygous samples (data not shown). While a profound reduction in FANCD2-ubiquitination (FANCD2-Ub) in the bi-allelic *FANCL* mutant clone was readily apparent by western blot analysis, consistent with loss of FANCL activity, for the heterozygous clones the total FANCD2-Ub levels were not detectably different to the WT clones (**Suppl. Fig. S6**), reflecting the insensitivity of western blot analysis for detecting small differences. Overall these data are consistent with heterozygous damaging mutations in the FANC core gene *FANCL* leading to a subtle defect in FA DNA repair pathway function, detectable using sensitive and specific quantitation of the events downstream of FANC core and ID2 activity (i.e. FANCD2 foci formation).

DISCUSSION

The elevated frequency in this AML cohort of deleterious heterozygous variants affecting FA core complex genes (*FANCM, FANCL* and *FANCC*), and D-C mutations in AML cases, is suggestive of a role in AML pathogenesis. Importantly, the MCF10A *FANCL* heterozygous

316 model described here has provided proof-of-principle that deleterious heterozygous mutations
317 can impair the cellular response to DNA damaging events, such as ICL treatment. Based on
318 these findings, we propose that germline heterozygous damaging and D-C variants in the FANC
319 core and ID2 complex genes, *FANCC*, *FANCD2*, *FANCL* and *FANCM* may result in partial
320 impairment of the FA DNA repair pathway in HSC. We predict that such HSC would have
321 reduced capacity to maintain genome integrity under physiological stress conditions associated
322 with increased replicative index or high levels of endogenous/exogenous ICL, and hence may
323 be associated with increased stalled replication. For example, a key role of the FA DNA repair
324 pathway has been demonstrated in HSC particularly during emergency haematopoiesis (18), or
325 when HSC are confronted with cross-linking toxins (16). This raises the possibility that the
326 level of individual risk associated with these rare variants may be significantly affected by other
327 factors such as environmental exposure, infectious agents and/or other genetic variants that
328 affect metabolism of aldehydes. The data obtained using an isogenic WT, heterozygous and bi-
329 allelic CRISPR-*FANCL* model is consistent with the effects of heterozygous mutations being
330 subtle, however over time even subtle impairment of FA pathway function in HSC may
331 contribute to acquisition of somatic driver mutations and increased AML risk. Indeed, small
332 differences in mutation rate between individuals can have large effects on relative probability
333 of developing cancer (39). Interestingly, a recent study has identified heterozygous somatic and
334 germline mutations in FANC genes in paediatric ALL cases without apparent loss of the WT
335 allele, consistent with these variants contributing to leukemic initiation (40).

336

337 These findings, and findings from other recent cancer cohort studies (26, 28), have important
338 implications for FA families and raise important ethical and clinical issues, particularly
339 regarding the potential identification of individuals who have modest-intermediate risk of AML
340 and other cancers, and also for selection of sibling donors for stem cell transplantation of FA
341 patients. While previous studies of FA families have not found an overall increased incidence

of AML in FA carriers (21, 22) a major difference between these studies and the case-control study described here is the number of FANC genes analysed; as 3 of the 22 FANC genes (*FANCA*, *FANCC* and *FANCG)* represent the vast majority of FA cases, familial FA studies are biased towards identification of AML in these families. Furthermore, based on our data, the risk associated with deleterious FANC gene variants is modest and predicted to cause AML with low penetrance in FA families, potentially masking familial pre-disposition. Our study did not reveal early onset of AML in the FANC-mutant AML group, so the late onset of adult AML is likely to also contribute to masking of any familial trends. It is still premature to propose genetic counselling and risk reduction strategies for confirmed carriers of FANC core and ID2 damaging variants or D-C mutations, and we suggest that further analysis focused on assessing familial cancer risk for families across all FA complementation groups, or in cases of severe FA phenotypes, will help to clarify these issues. While based on our current studies, the increased risk of AML associated with D-C FANC gene mutations is modest (estimated at ~ 3-4 fold), it is likely that the effects of individual FANC gene variants will be highly context-dependent with significant potential for risk to be modified by environmental influences. We also cannot exclude that particular variants will confer dominant negative activity with a greater effect on FA DNA repair pathway function than that observed in the cell line model described here, leading to more severe genomic instability, and hence higher individual risk of AML.

While it is well established that *BRCA1/2* haploinsufficiency is associated with an impaired DNA damage response, telomere erosion, genomic instability and premature senescence (41), this is the first study to investigate in detail FA DNA repair pathway function associated with heterozygous FANC core and ID2 damaging variants. Gene expression profiling showed that AML with heterozygous FANC core and ID2 variants display up-regulation of gene signatures associated with DNA synthesis and repair, ATR activation and replication stress, and with elevated levels of *POLD2*, *POLD3*, *POLA2*, and *RPA1* and *RPA2* gene expression in these

368 samples. Whether such changes are reflective of a replicative stress or DNA repair phenotype
369 will require further investigation. DNA repair phenotypes have been reported in cells from FA
370 carriers compared to controls; for example differences have been observed for DNA damage
371 induced by bleomycin (42). However, we note that not all studies have revealed such effects
372 (43, 44) reinforcing the hypothesis that the phenotype and aggressiveness may be variant
373 specific. Lymphocytes from FA heterozygotes also display a four-fold increase in chromosome
374 breakage when compared with control lymphocytes (45).

375

376 The MCF10A CRISPR-*FANCL* model allowed for a functional analysis of the response to DNA
377 cross-linking agents in the presence of a heterozygous loss of function FANCL mutation. In
378 this model we observed intermediate levels of FANCD2 foci following MMC treatment
379 compared to the WT and bi-allelic clones. This is consistent with an impaired capacity to
380 respond to ICL-induced damage. Interestingly, defective FANCD2 foci formation has also been
381 observed in solid tumours with heterozygous *FANCM* mutations (46). While in this study we
382 focused on FA DNA repair pathway activation in response to ICL-induced DNA damage,
383 recent reports indicate a DNA damage-independent role for the FANC genes, such as *FANCD2*
384 and *BRCA2* in replication fork protection (11, 47-49). It will now be of great interest to measure
385 the cellular response of FA DNA repair pathway heterozygous cells under conditions of
386 replicative stress, for example nucleotide depletion. Previous studies have shown that
387 replicative stress results in HSC DNA damage for *Fanca*$^{-/-}$ and *Fancd2*$^{-/-}$ mouse models (16,
388 50), so even subtle changes to FA pathway function in HSC may impact genomic stability and
389 leukaemia risk.

390

391 An interesting observation from our AML cohort analysis is that heterozygosity for rare FANC
392 gene variants in adult AML is not associated with characteristic cytogenetic changes that have
393 been described in FA AML [e.g. over-representation of +1q, -7/7q and -3q in FA (51, 52)].

We speculate that this may be because individuals with heterozygous FANC gene variants have a functional (albeit slightly impaired) FA DNA repair pathway and an intact haematopoietic system. Conversely in FA patients the AML clone develops in the context of complete loss of FA DNA repair pathway function, and is secondary to a failing haematopoietic system. In fact, in contrast to the *de novo* AML cases studied here, AML from FA is considered as secondary AML (53). Thus, it is likely that the specific changes observed in FA AML facilitate the development of AML clones in this context, allowing aberrant growth and survival, and malignant transformation in the absence of FA pathway function.

Recent studies have revealed that the FA pathway, together with Poly-ADP ribose polymerase (PARP1), play a key role protecting replication forks and preventing genomic instability during replication (11, 54, 55). Furthermore, boosting replicative stress and DNA damage with the chemotherapeutic 5-fluorouracil (5-FU) increases load on the BER DNA repair pathway and is synthetic lethal with inhibition of PARP1 in AML and ALL cells, and in patient derived xenografts; this combination approach (5FU and PARP1 inhibitor) has thus been proposed as a potential therapeutic approach in AML (and ALL) (56). Given the heterogeneity of AML, it is now of interest to establish the factors that determine sensitivity of AML cells to this strategy. Our observation that AML with heterozygous FANC core and ID2 mutations displays increased expression of DNA synthesis and DNA repair genes, and signatures associated with replicative stress, raise the important question of whether these AML cases may display heightened sensitivity to this combination, potentially providing a therapeutic option.

METHODS

AML samples

AML samples were obtained from the Australian Leukaemia & Lymphoma Group (ALLG) tissue bank at the Princess Alexandra Hospital (PAH, Brisbane, QLD, Australia) and from the

SA Cancer Research Biobank (SACRB) at the Royal Adelaide Hospital (RAH) and SA Pathology (Adelaide, SA, Australia). The use of the samples for this research study was approved by the PAH, the RAH, the University of Adelaide, the University of South Australia and the University of Queensland Human Research Ethics Committees (HRECs: HREC/05/QRCH/77, HREC/04/QPAH/172 and HREC/13/RAH/612). ALLG AML specimens were collected with informed consent. AML patient samples obtained from 1998 onwards from SACRB were collected with signed informed consent for research purposes. The RAH HREC waived the requirement for informed consent for SACRB specimens collected before 1998.

WES and bioinformatics analysis

WES was performed by hybridization capture followed by massively-parallel, paired-end, short-read sequencing using Illumina platforms. Methods and analysis are detailed in Supplementary Information.

Healthy control and independent AML extension data

Control WES data were obtained from a cohort of unaffected Australian Caucasian females (n=799), and additional AML WES data was obtained from the published TCGA study [n=102; ref (1)].

Variant calling and statistical analyses

Briefly, rare variants were identified using an established filtering and analysis pipeline and defined as those with a population frequency/minor allele frequency (MAF) of <0.001 as reported in dbSNP147, 1000 genomes (April 2012 version, any ethnicity), the 6500 NHLBI-ESP project (any ethnicity) and/or the Exome Aggregation consortium (ExAC) database. We filtered for predicted deleterious FANC gene variants with a high probability of pathogenicity using the Combined Annotation Dependent Depletion algorithm (CADD score >10) (33). Burden testing was employed for comparisons with a healthy control cohort. Detailed methods

are provided in Supplementary Information.

Cell culture

MCF10A cells were cultured using Dulbecco's Modified Eagle's Medium nutrient mixture F-12 HAM (DMEM-F12; Sigma), with 5% horse serum (Sigma) supplemented with 20ng/mL epidermal growth factor (EGF; R&D Systems), 500ng/mL hydrocortisone (Sigma), 10μg/mL insulin (Sigma), 100ng/mL Cholera toxin (Sigma), 100U/mL penicillin and 100μg/mL streptomycin (Sigma).

MCF10A CRISPR model of *FANCL* heterozygousity

Three different single-guide RNA (sgRNA) were designed using the http://crispr.mit.edu/ online tool and cloned individually into the pSpCas9(BB)-2A-Puro (PX459) V2 construct that contained Cas9. The CRISPR-Cas9 construct was transiently transfected into the MCF10A cell line, and transfected cells were selected for 2 days using antibiotic selection (puromycin). Single cells were isolated using limiting dilution and expanded to generate clonal lines. A total of 50 viable clones were obtained and gDNA extracted from each. T7 endonuclease assay and Sanger sequencing were used to determine the presence of deletions at the *FANCL* locus. TOPO TA (Thermo Fisher Scientific) cloning was performed following manufacturer's protocol and DNA from colonies of transformed DH5α cells was sequenced to confirm *FANCL* mutation and purity of each selected MCF10A CRISPR clone.

Immunofluorescence assay

In each well of an 8-well chambered slide (Nunc® Lab-Tek® Chamber Slide™ system), 7500 cells were seeded and left to recover for 72 hours. The cells were treated with 40ng/mL of MMC or DMSO (vehicle) for 48 hours. The wells were washed using 1x PBS. The cells were fixed using 4% paraformaldehyde for 20 minutes and washed using 1x PBS. The cells were permeabilised using 0.3% Triton™ X-100 solution for 10 minutes and washed using 1x PBS.

The wells were blocked using the blocking solution (10% FBS and 0.1% NP40 in 1x PBS) for 1 hour. Primary antibodies for FANCD2 rabbit polyclonal antibody (Novus: NB100-182) and γ-H2AX mouse monoclonal antibody (MERCK-Millipore: 05-636-I) were diluted 1:1000 in blocking solution, added to the wells and incubated for 3 hours on a rocking shelf at room temperature. The secondary antibodies goat α-mouse Alexa488 (Cell Signalling: 4408) and goat α-rabbit Alexa647 (Cell Signalling: 4414) were diluted 1:1000 in blocking buffer, added to the wells and incubated at 4ºC on a rocking shelf overnight. ProLong™ Gold Antifade Mountant with DAPI (ThermoFisher Scientific: P36935) was added to each well as mounting media and nuclear staining. The images were processed using ZEN Blue (version 2.3) program from ZEISS.

Quantification of FANCD2 foci

Images exported from ZEN Blue were quantitated using ImageJ (https://imagej.nih.gov/ij/). The images were converted to 16-bit images, and the background threshold was adjusted and set to minimize background (0.01-0.02% for the upper limit and the maximum for the lower limit). The "analyse particles" function was used to quantify the number of FAND2 foci in the images using 20 micron per pixel-square as minimum size of a focus. In each image, cells positive for γH2AX foci were first identified visually and then these cells were scored as positive or negative for FANCD2 foci using a criteria of greater than or equal to 3 FANCD2 foci per cell as determined by the ImageJ software. Four images were analysed for each condition, representing an average of 15 cells per image.

ACKNOWLEDGEMENTS

We would like to acknowledge the Leukaemia Foundation of Australia for their support of the Australian Leukaemia and Lymphoma Group (ALLG) tissue bank. SACRB is supported by the Cancer Council SA Beat Cancer Project, Medvet Laboratories Pty Ltd and the Government of South Australia. We would also like to acknowledge the assistance from the Detmold Family Cytometry Facility, Hanson Institute, SA Pathology.

The authors thank Brooke Gardiner and colleagues at UQ Centre for Clinical Genomics (UQ Diamantina Institute) for carrying out exome capture and sequencing. We also thank Megan Ellis (ALLG tissue bank) and Carolyn Butcher (SA Cancer Research Biobank) for invaluable assistance in identifying and preparing samples, Diana Iarossi, Michelle Perugini, Silke Danner and Sarah Moore for assistance with molecular and clinical annotation of AML samples, Crystal McGirr for general technical assistance and Charlotte Hodson for help with analysis of mutation effect on function.

REFERENCES

1. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. The New England journal of medicine. 2013;368(22):2059-74.

2. Kadia TM, Ravandi F, Cortes J, Kantarjian H. Toward Individualized Therapy in Acute Myeloid Leukemia: A Contemporary Review. JAMA Oncol. 2015;1(6):820-8.

3. Godley LA. Inherited predisposition to acute myeloid leukemia. Seminars in hematology. 2014;51(4):306-21.

4. Hahn CN, Chong CE, Carmichael CL, Wilkins EJ, Brautigan PJ, Li XC, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. Nature genetics. 2011;43(10):1012-7.

5. Lewinsohn M, Brown AL, Weinel LM, Phung C, Rafidi G, Lee MK, et al. Novel germ line DDX41 mutations define families with a lower age of MDS/AML onset and lymphoid malignancies. Blood. 2016;127(8):1017-23.

6. Wang AT, Kim T, Wagner JE, Conti BA, Lach FP, Huang AL, et al. A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. Molecular cell. 2015;59(3):478-90.

7. Alter BP. Fanconi anemia and the development of leukemia. Best practice & research Clinical haematology. 2014;27(3-4):214-21.

8. Rosenberg PS, Alter BP, Ebell W. Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. Haematologica. 2008;93(4):511-7.

9. Garaycoechea JI, Patel KJ. Why does the bone marrow fail in Fanconi anemia? Blood. 2014;123(1):26-34.

10. Walden H, Deans AJ. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. Annual review of biophysics. 2014;43:257-78.

11. Schlacher K. PARPi focus the spotlight on replication fork protection in cancer. Nature cell biology. 2017;19(11):1309-10.

12. Boisvert RA, Howlett NG. The Fanconi anemia ID2 complex: dueling saxes at the crossroads. Cell cycle. 2014;13(19):2999-3015.

13. Nielsen FC, van Overeem Hansen T, Sorensen CS. Hereditary breast and ovarian cancer: new genes in confined pathways. Nature reviews Cancer. 2016;16(9):599-612.

14. Schulz E, Valentin A, Ulz P, Beham-Schmid C, Lind K, Rupp V, et al. Germline mutations in the DNA damage response genes BRCA1, BRCA2, BARD1 and TP53 in patients with therapy related myeloid neoplasms. Journal of medical genetics. 2012;49(7):422-8.

15. Churpek JE, Marquez R, Neistadt B, Claussen K, Lee MK, Churpek MM, et al. Inherited mutations in cancer susceptibility genes are common among survivors of breast cancer who develop therapy-related leukemia. Cancer. 2016;122(2):304-11.

16. Pontel LB, Rosado IV, Burgos-Barragan G, Garaycoechea JI, Yu R, Arends MJ, et al. Endogenous Formaldehyde Is a Hematopoietic Stem Cell Genotoxin and Metabolic Carcinogen. Molecular cell. 2015;60(1):177-88.

17. Deans AJ, West SC. DNA interstrand crosslink repair and cancer. Nature reviews Cancer. 2011;11(7):467-80.

18. Walter D, Lier A, Geiselhart A, Thalheimer FB, Huntscha S, Sobotta MC, et al. Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. Nature. 2015;520(7548):549-52.

19. Rogers KJ, Fu W, Akey JM, Monnat RJ, Jr. Global and disease-associated genetic variation in the human Fanconi anemia gene family. Human molecular genetics. 2014;23(25):6815-25.

20. Rosenberg PS, Tamary H, Alter BP. How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel. American journal of medical genetics Part A. 2011;155A(8):1877-83.

21. Berwick M, Satagopan JM, Ben-Porat L, Carlson A, Mah K, Henry R, et al. Genetic heterogeneity among Fanconi anemia heterozygotes and risk of cancer. Cancer research. 2007;67(19):9591-6.

22. Tischkowitz M, Easton DF, Ball J, Hodgson SV, Mathew CG. Cancer incidence in relatives of British Fanconi Anaemia patients. BMC cancer. 2008;8:257.

23. Sharan SK, Morimatsu M, Albrecht U, Lim DS, Regel E, Dinh C, et al. Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. Nature. 1997;386(6627):804-10.

24. Esteban-Jurado C, Franch-Exposito S, Munoz J, Ocana T, Carballal S, Lopez-Ceron M, et al. The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. European journal of human genetics : EJHG. 2016;24(10):1501-5.

25. Lhota F, Zemankova P, Kleiblova P, Soukupova J, Vocka M, Stranecky V, et al. Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. Clin Genet. 2016;90(4):324-33.

26. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. The New England journal of medicine. 2015;373(24):2336-46.

27. Hart SN, Ellingson MS, Schahl K, Vedell PT, Carlson RE, Sinnwell JP, et al. Determining the frequency of pathogenic germline variants from exome sequencing in patients with castrate-resistant prostate cancer. BMJ Open. 2016;6(4):e010332.

28. Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MD, et al. Patterns and functional implications of rare germline variants across 12 cancer types. Nature communications. 2015;6:10086.

29. Awan A, Malcolm Taylor G, Gokhale DA, Dearden SP, Will A, Stevens RF, et al. Increased frequency of Fanconi anemia group C genetic variants in children with sporadic acute myeloid leukemia. Blood. 1998;91(12):4813-4.

30. van Twest S, Murphy VJ, Hodson C, Tan W, Swuec P, O'Rourke JJ, et al. Mechanism of Ubiquitination and Deubiquitination in the Fanconi Anemia Pathway. Molecular cell. 2017;65(2):247-59.

31. Du W, Amarachintha S, Wilson AF, Pang Q. Hyper-active non-homologous end joining selects for synthetic lethality resistant and pathological Fanconi anemia hematopoietic stem and progenitor cells. Scientific reports. 2016;6:22167.

32. Duncan EL, Danoy P, Kemp JP, Leo PJ, McCloskey E, Nicholson GC, et al. Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. PLoS genetics. 2011;7(4):e1001372.

33. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics. 2014;46(3):310-5.

34. de la Hoya M, Fernandez JM, Tosar A, Godino J, Sanchez de Abajo A, Vidart JA, et al. Association between BRCA1 mutations and ratio of female to male births in offspring of families with breast cancer, ovarian cancer, or both. Jama. 2003;290(7):929-31.

35. Moslehi R, Singh R, Lessner L, Friedman JM. Impact of BRCA mutations on female fertility and offspring sex ratio. Am J Hum Biol. 2010;22(2):201-5.

36. Esposito MT, Zhao L, Fung TK, Rane JK, Wilson A, Martin N, et al. Synthetic lethal targeting of oncogenic transcription factors in acute leukemia by PARP inhibitors. Nature medicine. 2015;21(12):1481-90.

37. Shigechi T, Tomida J, Sato K, Kobayashi M, Eykelenboom JK, Pessina F, et al. ATR-ATRIP kinase complex triggers activation of the Fanconi anemia DNA repair pathway. Cancer research. 2012;72(5):1149-56.

38. Soule HD, Maloney TM, Wolman SR, Peterson WD, Jr., Brenz R, McGrath CM, et al. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer research. 1990;50(18):6075-86.

39. Araten DJ, Golde DW, Zhang RH, Thaler HT, Gargiulo L, Notaro R, et al. A quantitative measurement of the human somatic mutation rate. Cancer research. 2005;65(18):8111-7.

40. Gayle P. Pouliot, Melissa A. Burns, James Degar , Chau Vo , Lisa Moreau , Justine E. Roderick, Bose Kochupurakkal, Sofie Peirs, Bjorn Menten, Mignon L. Loh, Stephen P. Hunger, Lewis B. Silverman, Stephen E. Sallan, Marian H., Harris KES, Donna S. Neuberg, David M. Weinstock, Andrew P. Weng, Michelle Kelliher, Pieter Van Vlierberghe, Alan D. D'Andrea, and Alejandro Gutierrez. The Fanconi Pathway in Pediatric T-Cell Acute Lymphoblastic Leukemia. 2017.

600 41. Sedic M, Kuperwasser C. BRCA1-hapoinsufficiency: Unraveling the molecular and cellular basis for
601 tissue-specific cancer. Cell cycle. 2016;15(5):621-7.

602 42. Barquinero JF, Barrios L, Ribas M, Egozcue J, Caballin MR. Cytogenetic sensitivity of three Fanconi
603 anemia heterozygotes to bleomycin and ionizing radiation. Cancer genetics and cytogenetics. 2001;124(1):80-3.

604 43. Pathania S, Bade S, Le Guillou M, Burke K, Reed R, Bowman-Colin C, et al. BRCA1 haploinsufficiency
605 for replication stress suppression in primary cells. Nature communications. 2014;5:5496.

606 44. Mohseni-Meybodi A, Mozdarani H, Mozdarani S. DNA damage and repair of leukocytes from Fanconi
607 anaemia patients, carriers and healthy individuals as measured by the alkaline comet assay. Mutagenesis.
608 2009;24(1):67-73.

609 45. Auerbach AD, Adler B, Chaganti RS. Prenatal and postnatal diagnosis and carrier detection of Fanconi
610 anemia by a cytogenetic method. Pediatrics. 1981;67(1):128-35.

611 46. Villalona-Calero MA, Duan W, Zhao W, Shilo K, Schaaf LJ, Thurmond J, et al. Veliparib Alone or in
612 Combination with Mitomycin C in Patients with Solid Tumors With Functional Deficiency in Homologous
613 Recombination Repair. Journal of the National Cancer Institute. 2016;108(7).

614 47. Chaudhury I, Sareen A, Raghunandan M, Sobeck A. FANCD2 regulates BLM complex functions
615 independently of FANCI to promote replication fork recovery. Nucleic acids research. 2013;41(13):6444-59.

616 48. Lachaud C, Moreno A, Marchesi F, Toth R, Blow JJ, Rouse J. Ubiquitinated Fancd2 recruits Fan1 to
617 stalled replication forks to prevent genome instability. Science. 2016;351(6275):846-9.

618 49. Xing M, Wang X, Palmai-Pallag T, Shen H, Helleday T, Hickson ID, et al. Acute MUS81 depletion leads
619 to replication fork slowing and a constitutive DNA damage response. Oncotarget. 2015;6(35):37638-46.

620 50. Kaschutnig P, Bogeska R, Walter D, Lier A, Huntscha S, Milsom MD. The Fanconi anemia pathway is
621 required for efficient repair of stress-induced DNA damage in haematopoietic stem cells. Cell cycle.
622 2015;14(17):2734-42.

623 51. Rochowski A, Olson SB, Alonzo TA, Gerbing RB, Lange BJ, Alter BP. Patients with Fanconi anemia
624 and AML have different cytogenetic clones than de novo cases of AML. Pediatric blood & cancer. 2012;59(5):922-
625 4.

626 52. Quentin S, Cuccuini W, Ceccaldi R, Nibourel O, Pondarre C, Pages MP, et al. Myelodysplasia and
627 leukemia of Fanconi anemia are associated with a specific pattern of genomic abnormalities that includes cryptic
628 RUNX1/AML1 lesions. Blood. 2011;117(15):e161-70.

629 53. Peffault de Latour R, Soulier J. How I treat MDS and AML in Fanconi anemia. Blood.
630 2016;127(24):2971-9.

631 54. Berti M, Ray Chaudhuri A, Thangavel S, Gomathinayagam S, Kenig S, Vujanovic M, et al. Human
632 RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. Nature structural &
633 molecular biology. 2013;20(3):347-54.

634 55. Schlacher K, Wu H, Jasin M. A distinct replication fork protection pathway connects Fanconi anemia
635 tumor suppressors to RAD51-BRCA1/2. Cancer cell. 2012;22(1):106-16.

636 56. Falzacappa MV, Ronchini C, Faretta M, Iacobucci I, Di Rora AG, Martinelli G, et al. The Combination
637 of the PARP Inhibitor Rucaparib and 5FU Is an Effective Strategy for Treating Acute Leukemias. Molecular
638 cancer therapeutics. 2015;14(4):889-98.

639

Table 1. Details of D-C mutations in the Australian and TCGA AML cohorts

Type of Mutation	AML ID	Sex	Age	Gene	Amino acid change	[1]LOH (Y, N)	[2]FA DB ID	Freq. in Australian AML cohort (n=131)	Freq. in Australian control cohort (n=323)	Freq. in ExAC (non-Finnish European, n=33370)
Missense	WES-21	F	72	*FANCA*	p.T1131A	N	FANCA_000241	0.76%	0.30%	0.0%
Missense	WES-20	F	79	*FANCC*	p.D195V	N	FANCC_000019	2.30%	0.98%	0.4%
	WES-79	M	73			N				
	WES-86	M	83			N				
Nonsense	WES-224	F	59	*FANCC*	p.R548X	N	FANCC_000005	0.76%	0.0%	0.0%
Nonsense	WES-21	F	72	*FANCD2*	p.R926X	N	FANCD2_000019	0.76%	0.0%	0.0%
Splicing	WES-249	F	N/A	*FANCD2*	c.2715+1G>A	N	FANCD2_000016	0.76%	0.0%	0.0%
Missense	WES-64	F	46	*FANCG*	p.T297I	N	FANCG_000038	0.76%	0.0%	0.01%
Indel	WES-30	M	56	*FANCL*	p.336_337del	N	FANCL_000002	1.53%	0.30%	0.0%
	WES-216	F	49			N				
Nonsense	WES-46	F	59	*BRCA1*	p.R1203X[3]	N	BRCA1_001405	0.76%	0.0%	0.0%

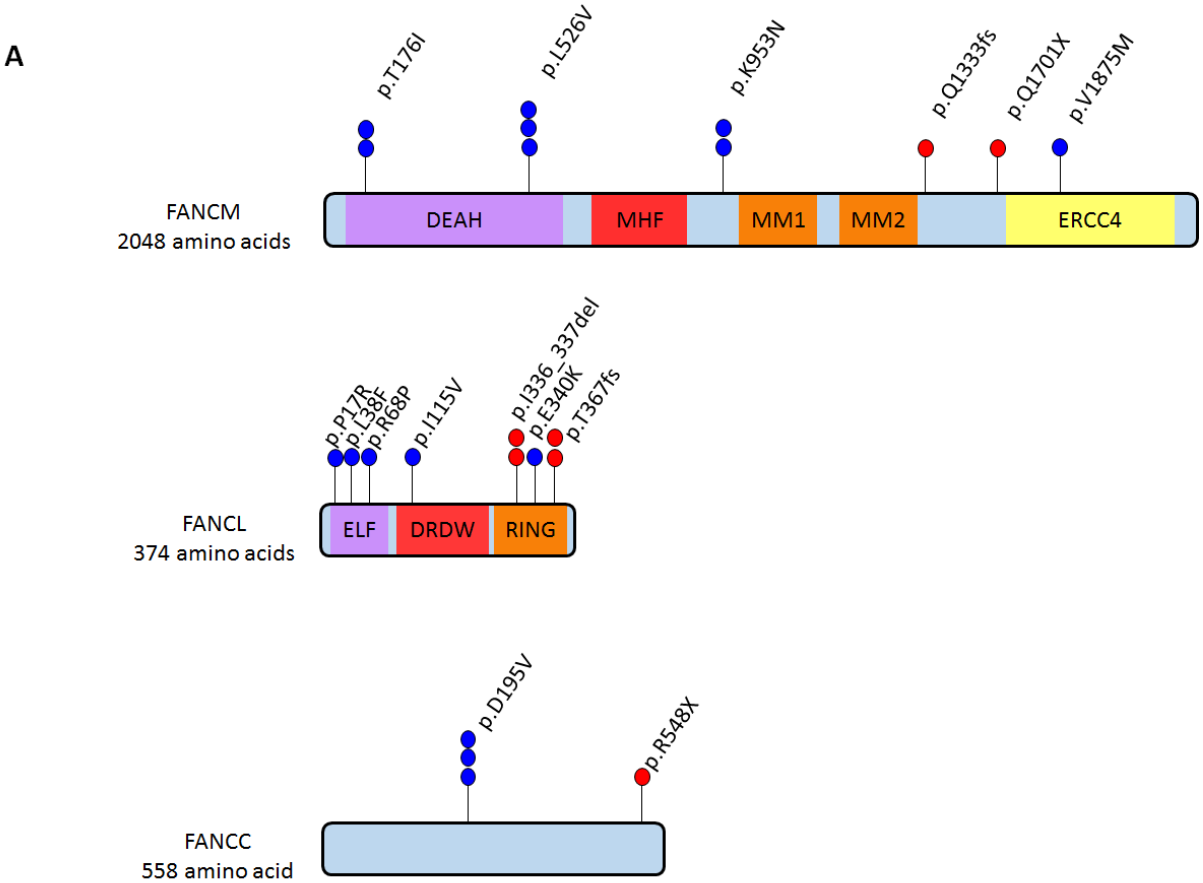
[1] Based on VAF > 55%; [2]FA DB ID represents the ID of mutations identified in the Rockefeller University FA database; [3] Confirmed germline.

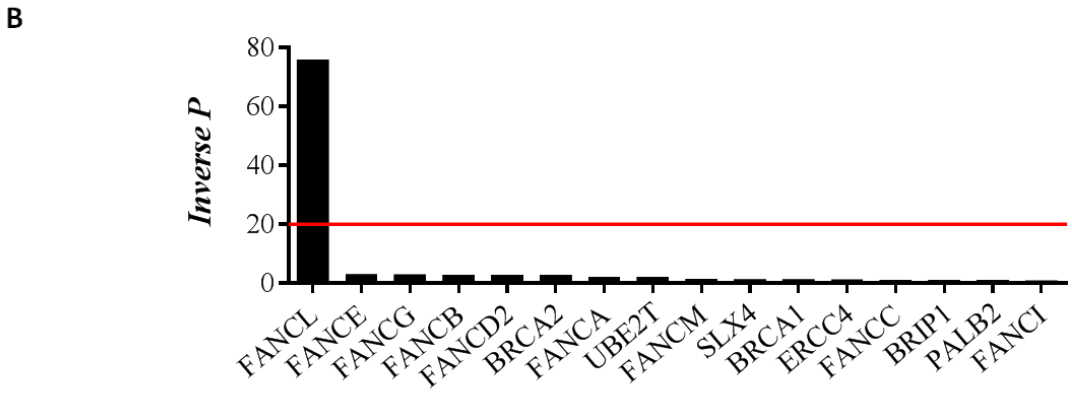
Table 2. Frequency of FANC variants in ExAC and Australian AML cohorts.

	[1]ExAC (n=66740)			[1]AML (n=262)			
Gene	**Count**	**ExAC Frequency**	**Frequency (%)**	**Count**	**AML Frequency**	**Frequency (%)**	**[2]P-value**
FANCA	653	0.0098	0.9784	2	0.0076	0.7634	0.9693
FANCB	145	0.0022	0.2173	0	0.0000	0.0000	0.9289
FANCC	**211**	**0.0032**	**0.3162**	**4**	**0.0153**	**1.5267**	**0.0036***
BRCA2	1276	0.0191	1.9119	3	0.0115	1.1450	0.4970
FANCD2	546	0.0082	0.8181	2	0.0076	0.7634	0.8061
FANCE	302	0.0045	0.4525	1	0.0038	0.3817	0.7712
FANCF	178	0.0027	0.2667	0	0.0000	0.0000	0.8136
FANCG	219	0.0033	0.3281	1	0.0038	0.3817	0.6967
FANCI	762	0.0114	1.1417	3	0.0115	1.1450	0.7746
BRIP1	406	0.0061	0.6083	1	0.0038	0.3817	0.9419
FANCL	**321**	**0.0048**	**0.4810**	**9**	**0.0344**	**3.4351**	**< 0.0001***
FANCM	**792**	**0.0119**	**1.1867**	**10**	**0.0382**	**3.8168**	**0.0003***
PALB2	322	0.0048	0.4825	2	0.0076	0.7634	0.8353
RAD51C	157	0.0024	0.2352	1	0.0038	0.3817	0.8805
SLX4	689	0.0103	1.0324	6	0.0229	2.2901	0.0892
ERCC4	669	0.0100	1.0024	4	0.0153	1.5267	0.5899
RAD51	94	0.0014	0.1408	0	0.0000	0.0000	0.8266
BRCA1	619	0.0093	0.9275	3	0.0115	1.1450	0.9651
UBE2T	108	0.0016	0.1618	0	0.0000	0.0000	0.9046
XRCC2	97	0.0015	0.1453	0	0.0000	0.0000	1.0000
MAD2L2	59	0.0009	0.0884	0	0.0000	0.0000	1.0000
RFWD3	281	0.0042	0.4210	1	0.0038	0.3817	1.0000

[1] n represents the total number of alleles in each cohort. The non-Finnish European cohort in ExAC consists of 33370 individuals. [2] Fisher's exact test was used to determine the statistical difference between the Australian AML cohort and the non-Finnish European cohort in ExAC. (* $P<0.05$).

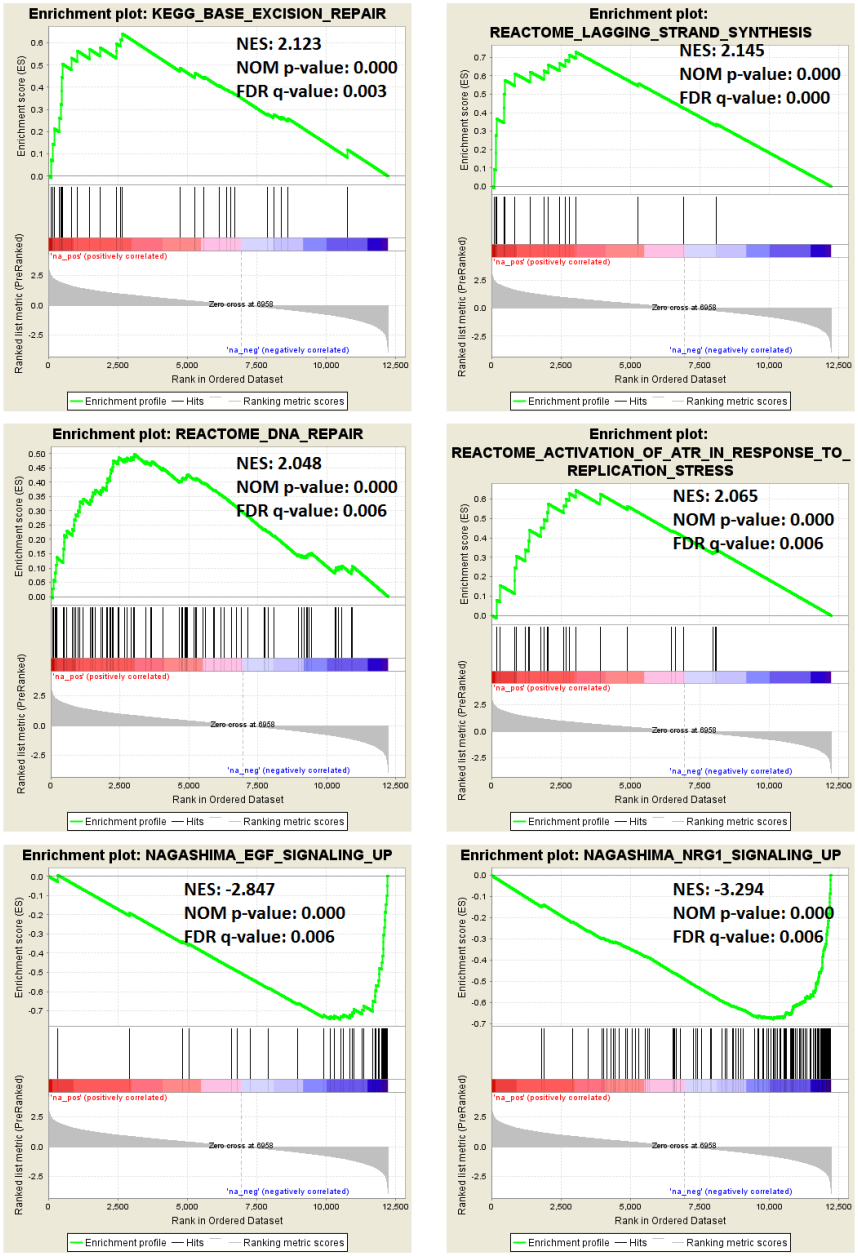
Maung el al, Figure 1



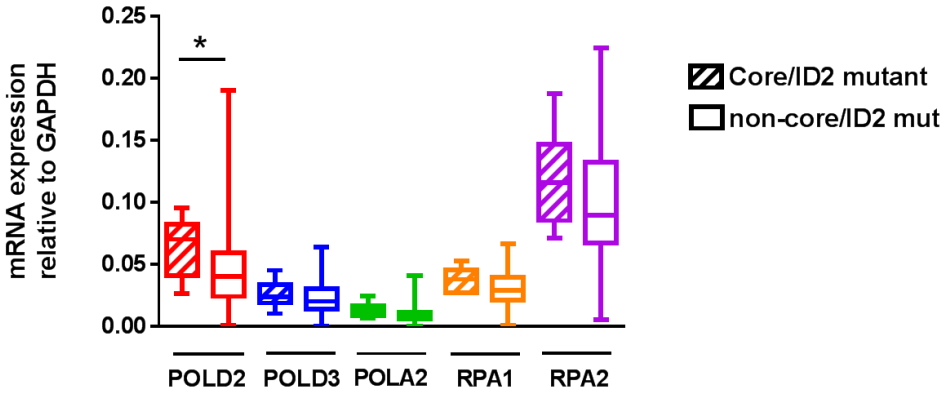


Maung et al, Figure 2

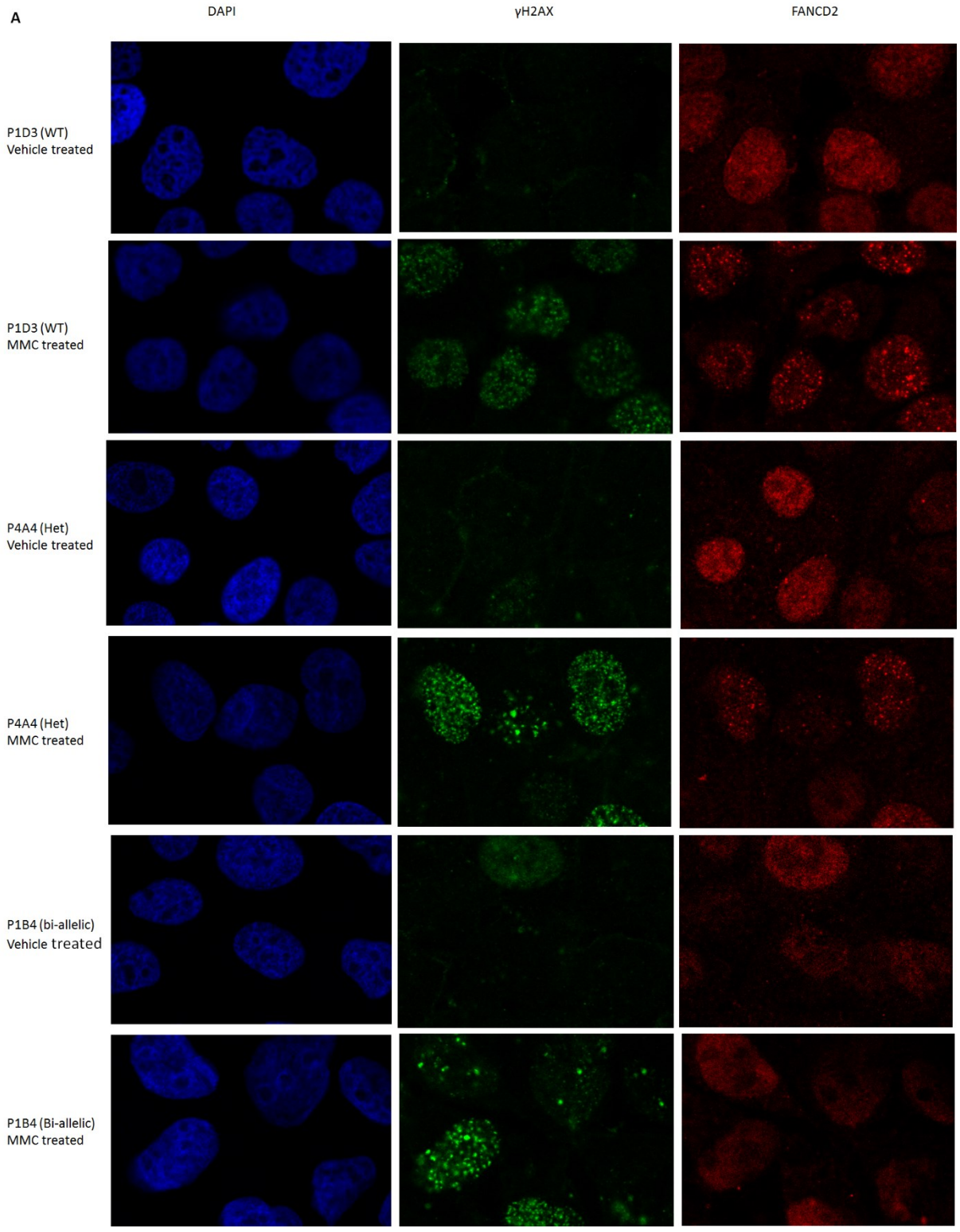
A



B



Maung et al, Figure 3



B Percentage of γH2AX^{+} cells with 3≥ FANCD2 foci

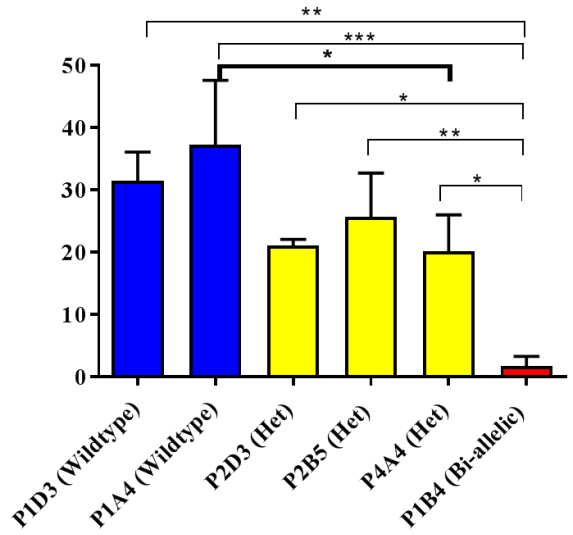


FIGURE LEGENDS

Figure 1. (A). Mutations identified in FANC genes in the Australian AML cohort. Schematics of protein structures showing mutations in FANCM, FANCL and FANCC identified in diagnostic AML samples. Conserved domains are indicated. Blue circles indicate amino acid substitutions, red circles indicate truncating mutations. **(B). Burden analysis.** Inverse *P* value is plotted for all FANC genes (line represents *P*=0.05).

Figure 2. (A). Gene Set Enrichment Analysis (GSEA) of FANC core and ID2 mutant AML samples. GSEA plots show enrichment, in the FANC core and ID2 mutant AML samples, of signatures corresponding to Base Excision Repair, DNA repair and synthesis, ATR response to replication stress and reduced growth factor response. NES, normalized enrichment score; NOM p-value, nominal p-value; FDR q-val, false discovery rate q value. Detailed descriptions of the top gene-sets identified by GSEA are shown in **Suppl. Table S7**. **(B). Expression of selected genes in FANC core and ID2 mutant AML samples.** mRNA expression of *POLD2*, *POLD3*, *POLA2*, *RPA1* and *RPA2* relative to *GAPDH* was determined by qRT-PCR for FANC core and ID2 mutant AML (n=9) and non-FANC core and ID2 mutant AML samples (n=37). **P*<0.05 (Mann-Whitney t test).

Figure 3. (A). A *FANCL* heterozygous cell line model. Immunofluorescent images captured at 63x magnification for three representative MCF10A CRISPR-*FANCL* clones (P1D3-wild type, P4A4-heterozygous, and P1B4-biallelic) treated with mitomycin C and vehicle (DMSO) are shown. Cells were probed with DAPI (blue) or antibodies for γH2AX (green), and FANCD2 (red). **(B).** Percentage of γH2AX positive cells with FANCD2 foci ≥ 3 in the 6 MCF10A CRISPR-*FANCL* cell lines. P1B4 shows a statistically lower percentage of $\gamma H2AX^+$ cells with FANCD2 foci compared to the remaining five clones. P4A4 shows a statistically lower percentage of $\gamma H2AX^+$ cells with FANCD2 foci compared to clone P1A4. For statistical comparison, One way ANOVA with Tukey multiple comparison was performed. * *P*<0.05, ** *P*<0.01, *** *P*<0.001.

SUPPORTING INFORMATION LEGENDS

S1_Suppl info.docx: This file contains aadditional detailed materials and methods, Supplementary Figures S1, S2, S3, S4, S5 and S6 with their respective figure legends, and

Supplementary Tables S1, S3, S4, S5, S6, S7, S8, S9 and S10.

S2_Table.xlsx: This file contains a table listing all the FANC variants identified in the Australian AML cohort.

Maung et al, Response to Reviewers.

We thank the Reviewer's for their acknowledgement of the amount of work, and the interest, of the findings in our initial manuscript (ms # PGENETICS-D-17-01431). We now submit an extensively revised version of this that addresses the Editor's and Reviewer's comments, including reduction of sections associated with speculative discussion. In this revised version, we include new data demonstrating a phenotype in a CRISPR-generated MCF10A cell line model of *FANCL* heterozygosity. Most importantly, for three independent het FANCL deletion clones we show a consistent reduction, relative to wild type (WT) controls for ICL-induced FANCD2 foci. This experimental validation of altered FA pathway function upon heterozygous mutation of the core FA pathway gene *FANCL* is novel (as highlighted by Reviewer 4), and further increases the impact of this study. We appreciate that that our findings challenge dogma in the FA field that is based on previous studies of AML in FA families, hence we now put our studies better into context by inclusion in our introduction of a more extensive discussion of the differences and limitations associated with these previously published studies.

Response to Editor's comments.

"1. Major criticism involves a lack of a relationship between the observed FA variants and specific AML molecular subtypes which have different routes of pathogenesis and have different gene expression profiles."

We acknowledge that AML is a highly heterogeneous disease with multiple oncogenic mechanisms. This creates a major challenge for AML genomics studies as very large multi-centre cohorts are required to derive meaningful analysis from smaller subgroups. We point out that statistical analysis of our cohort has revealed association of FANC core and ID2 gene mutations with AML molecular and cytogenetic markers, ie. monosomy 7/del 7 and FLT3-ITD (see **Suppl Table S1** and **Fig S2**). In response to the reviewer's comment we have now also included the analyses for the normal karyotype AML subgroup (NK-AML) with and without FANC core/ID gene mutations (**Suppl Table S3**); this analysis removes samples with chromosomal rearrangements, a number of which have been reported to impact HRR activity (Ref 36 in the ms). Importantly we still observe a significant increased frequency of D-C mutations in this NK-AML group relative to European Americans in the ESP database. We have also modified our conclusions (page 11) to reflect the limitations associated with an analysis of a cohort of this size.

"2. Absence of validation of gene expression profiling and the speculative nature of the role of FA variants in the pathogenesis of AML as a consequence."

We agree that the discussion of the GSEA was highly speculative. We have now substantially shortened this section and focused our discussion on the significance of the detection of signatures associated with DNA synthesis and repair, ATR activation and replication stress. Importantly, we have validated the changes in expression of individual POL and RPA genes

involved in DNA synthesis and repair (see **Fig 2B**). Finally, our quantitative image analysis of the MCF10A *FANCL* clones has revealed reduced FANCD2 foci formation for the heterozygous clones in response to ICL treatment. Thus, we now present a compelling case for a heterozygous phenotype as a result of *FANCL* mutation, consistent with the hypothesis that rare deleterious variants in the FANC core and ID2 genes may lead to increased genomic instability over time, and modest increased risk of AML.

Response to Reviewer #1.

"The strongest point against the assumed genetic risk is the clinical observation that in FA there are is not an increased frequency of familial leukaemia"

This issue is now more directly addressed in the ms, including on page 6-7 where we provide an explanation of the key differences between our study and previous studies. Importantly, we now also outline a number of key limitations associated with previous FA familial studies. We believe that our studies are not inconsistent with previous findings, but rather our AML cohort analysis has detected a modest AML risk that is difficult to detect in familial FA studies. The key points that we now emphasise in the text are:

1. Previous studies of FA families are necessarily biased towards detection of AML risk associated with D-C variants affecting the *three FANC genes* (of 22) that are most commonly affected in FA [ie 84% of FA is caused by *FANCA* (64%), *FANCG* (8%) *and FANCC* (12%) mutations]. In addition, it must be considered that deleterious variants in the genes that rarely cause FA may not be tolerated in bi-allelic combination at the germline level (ie due to embryonic lethality), but may still confer an increased risk of AML development when present in the het state; with this in mind cancer cohort studies that explore the frequency of deleterious variants across all 22 FANC genes are clearly warranted.

2. Previous studies of FA families have determined that there is not a high risk of AML development, for FA carriers ie there is not an AML susceptibility phenotype that is sufficiently penetrant to be revealed from investigations of AML in FA families. Our case-control study provides evidence of modest increased risk associated with heterozygous disease causing (D-C) gene variants in the FANC genes (ie core and ID2 genes), consistent with the subtle phenotype in the MCF10A cell line model. AML associated with such modest-risk variants is predicted to occur with low penetrance in FA families, and thus is likely to be uncommon.

3. AML occurs most predominantly in individuals > 70yo, so this age profile is an important factor that contributes to masking of an increased incidence of AML in FA families. In our study age was not significantly different for FANC-mutant versus non-mutant AML, thus our data suggests that there is not an early onset of AML associated with deleterious FANC germline mutations that would allow AML in FA families to be detected more readily.

"..the hypothesis behind this study (i.e. that because individuals with FA get AML, therefore

genetic variants in FA genes could play a role in sporadic AML is a little bit out of time.

We now acknowledge that the Question of AML in FA families has been raised previously by researchers in the FA field (page 6), and has been investigated via familial studies. As discussed above we have now emphasised the issues/limitations of these earlier studies, and the importance of investigating FANC genes in cancer cohort analysis. It is important to emphasise that while other AML cohort studies of selected FANC genes have been reported (eg Condie et al and Tischkowitz et al referred to by the reviewer and now cited in the ms; both FANCA) there has not been a comprehensive analysis of all FANC genes in adult-onset AML, hence our analysis. Our study complements other studies in cancer cohorts that have revealed significantly increased odds of finding rare deleterious FANC gene variants cf healthy controls (eg Lu et al analysis of the TCGA cancer data, Ref 28 in the ms).

The reviewer also makes the point here that much new has been learned about FA and the underlying genetic and cellular defect. We agree and emphasise that this includes recent studies from FA mouse models that show that severe FANC gene defects, in combination with environmental or infectious exposure, can lead to bone marrow failure (refs 16 and 18 in ms). These studies provide further justification for our detailed and focused study of the germline variants across all FANC genes in AML.

"The AML group and the control group are both too small to make far reaching associations about genetic risk. One could argue that "AML" is a heterogeneous disease, and "lumping" all AMLs together is not timely anymore". See above response #1 to Ed.

"It can be assumed that the analysis in the landmark paper in the NEJM will have included a detailed look at FA genes (for exactly the same reason), but no increased mutation rate in FA genes was reported (germline was not tested, but presumably not all germline mutations will have reverted to WT in the number of cases of leukaemia tested, like in the study here)"

We assume the reviewer is referring to the Ley NEJM paper (ref 1 the ms) which describes the TCGA AML data. This landmark paper focused on somatic variants, and as such *did not examine disease-causing (D-C) germline FANC gene variants*. We include analysis of the TCGA AML data showing increased D-C FANC variants compared to the ESP dataset (European Americans)(page 11), and an increase in *FANCC* rare damaging variants cf non-Finnish Europeans in the ExAC database. Finally, we also show an increase in rare deleterious FANC mutations (*FANCC* and *FANCO*) in the TCGA cohort by Burden analysis, consistent with a previous analysis, restricted to truncation variants (Lu et al).

"the letter in Blood from Awan et al is cited from 1998 (5), which reports SSCP variations in DNA analysis of AML samples. The follow up paper from the same group in 2003 Barber et al (1), where FANCC is actually sequenced, is not cited. In this study no dramatic increase of FANCC mutations are found in childhood AML, but only the variant of unknown significance S26F in four AML cases (Check this paper – ie cohort size etc)……as reported from the same group for the FANCG gene (3), there is no evidence of an increased frequency of mutation

carrier status in childhood AML, in line with the data presented in the NEJM paper, where also no increase in mutation frequency of other FA genes than BRCA2 have been reported (6)."

The reviewer is raising the issue that studies of selected FANC genes in childhood AML cohorts have not revealed an increase in the incidence of mutation. Given that recent studies suggest important differences in the aetiology for childhood and adult AML (Tarlock K and Meshinchi S. *Pediatr Clin North Am*. 2015;62(1):75-93), and as there are inconsistencies between the childhood AML studies relating to *FANCC*, we have now removed discussion of these papers from the manuscript. With regard to the most recent comprehensive NEJM childhood pan-cancer study, while this included 16 of the 22 FANC genes, the vast majority of leukaemia cases in this study will be ALL (numbers of ALL v AML were not disclosed). This study is cited in our manuscript given that we identified a FANCM variant (p.Q1701X) that these authors report in a childhood B-ALL case (also identified as enriched in patients with triple-negative breast cancer (Kiiski et al., 2014).

"The analysis of microarray data is difficult to follow……To divide by FA-gene mutated and non-mutated status, and assume a link from carrier status of any FA gene variant to gene expression patterns is ignoring a lot of published important work in this context, and simply does not make sense. "

See response #2 to Ed. The inclusion of the separate analysis of the NK-AML group (FANC core/ ID2 mutant vs non-mutant; see **Suppl Tables S8 and S9**), removes a number of confounding factors associated with the heterogeneous nature of AML and the effects of transcription factor fusions on HRR. We have extensively modified the discussion of the microarray analysis, and now include QRT-PCR for selected genes in AML (**Fig 2B**).

"What do the authors actually mean when they talk of a "higher level of replicative stress in the FANC core and ID-mutant AML group"?

"Replicative stress" occurs when cells are replicating under non-optimal conditions, for example in the presence of limited nucleotides. This is associated with stalled replication forks, recruitment of FANC proteins and BRCA2 to sites of stalling, and activation of the ATR checkpoint pathway (Zeman and Cimprich, Nat Cell Biol, 16:2-9, 2014).

"The pre-leukaemic cytogenetic changes associated with FA have been described in detail (8,9), and it is difficult to conceive that heterozygosity for FA variants might result in other cytogenetic changes that have never been described in FA, given the proposed mechanistic speculations by the authors ("heterozygous deleterious mutations may confer partially reduced activity of the FA pathway resulting in genomic instability that is relatively subtle compared to that seen with bi-allelic loss-of-function), is not different"

We do not agree that changes identical to those seen in FA AML (eg Rochowski et al, 2012 and Quentin et al, 2011) will necessarily be observed in sporadic AML associated with FANC

heterozygous mutations. A critical distinction is that individuals with heterozygous deleterious FANC gene mutations have a functional FA pathway (albeit our data suggests with subtle impairment) and an intact haematopoietic system, while in FA patients an AML clone that develops is formed in the context of severely reduced or nil FA pathway function and failure of the haematopoietic system (ie considered as AML secondary to BMF, ref 53 in ms). Thus, it is likely that the genetic changes that are seen selectively in FA AML may facilitate the development of an AML clone in this BMF context, and allow growth and survival of the AML clone in the absence of FA pathway function. This is now discussed on page 18-19 of the ms.

"The strongest point against the assumed genetic risk is the clinical observation that in FA there is not an increased frequency of familial leukaemia …..even the cases of multiple cancers in FA family members, these have been T-cell ALL, not AML, and solid tumours" "Described association with other cancers, like breast cancer, cannot be supporting a role for AML causation. There is very little evidence supporting a common clinical or cellular phenotype of FA-mutation carrier status (except for BRCA2, BRCA1 and PALB2 mutations, which are associated with familial cancer)".

We have now extensively addressed this issue; see responses above. Most importantly we now show a subtle phenotype associated with heterozygous *FANCL* mutation in the MCF10A cell line model

"There is an enormous amount of speculation relating to mechanisms …"discussion of R-loops has now been removed and other speculative sections reduced extensively.

Reviewer #2

"The correlation with FANC expression was described in the manuscript but is perhaps underplayed as it could give further indication of the mechanism of leukaemogenesis in AML arising from FANC mutations" See response above.

"the authors showed the clinical, molecular and cytogenetic parameters associated with the patient cohorts, but failed to mention in the main article text that whilst FANC mutations are not themselves mutually exclusive, indeed those patients with one mutation usually have a second, but they are almost mutually exclusive from FLT3 and several other frequently seen mutations in AML" This is now briefly discussed in the RESULTS, page 8-9 of the ms

"it would be interesting to determine the impact of FANC mutations on survival or to perform a multi-variant analysis of all known prognostic factors". Agree that this is of sig interest. We are acquiring the complete set of patient survival data that will allow us to perform this analysis and this may form the basis of a separate report.

Reviewer #3.

Minor points have been now corrected in the ms.

Reviewer #4.

"it is not clear whether these mutations are driver or passenger mutations and whether these heterozygous mutations have any functional significance".

This terminology is applied to somatic cancer variants and it is only rarely that FANC gene somatic mutations occur in AML; as such the major question that arises from our study is whether these mutations confer a phenotype consistent with cancer *pre-disposition*, hence the inclusion of the MCF10A CRISPR FANCL cell line model. This has provided proof-of-principle that het FANC core/ID2 gene mutations lead to impaired function of the FA pathway.

"… the gene expression analysis data and interpretation is largely speculative" Addressed above.

"Is the frequency of FA gene mutation greater in the AML cohort than the unaffected cohort?" The overall frequency of patients with rare deleterious FANC gene mutations was similar compared to the healthy female controls although our analysis also revealed an elevated frequency of female AML cases carrying D-C mutations (page 9).

"For this analysis, why was the control cohort all-female as opposed to a matched gender cohort?"

We have utilised an available all-female cohort that has been sequenced and analysed using the same bioinformatics pipeline. As discussed on page 9 for some FANC genes there have been reports of gender over-representation (ref 34 and 35 in the ms) hence we restricted our comparison to the female AML cases.

"The authors describe an increased frequency of FA gene disease-causing mutations among females compared to the control all-female cohort (page 10, line 153). Was a similar analysis performed for males?" See above

"The authors state that the frequency of patients carrying disease-causing FA gene variants in the TCGA AML cohort was not significantly different to that of European-Americans in the ESP database. Can the authors speculate as to why there is a difference between the findings from the Australian and TCGA AML cohorts?" We expect that this reflects the heterogeneity of AML and differences in characteristics between the Australian and US cohorts, ie differences in ethnicity backgrounds.

"FA is not strictly recessive (Wang et al., 2015, RAD51/FANCR, Molecular Cell)." Now corrected in the text.

"There are now 22 FA genes, and not 19 as reported in this manuscript. The authors should

ideally need to update their analysis for all 22 genes" Our analysis has now been updated to include all 22 FANC genes

"Abstract, page 3, line 61: The latter part of the sentence "Taken together with recent studies this data strongly suggests that rare heterozygous germline FANC mutations confer modest increased AML risk that maybe modified by environmental or infectious exposure, and/or variants affecting aldehyde metabolism" is too speculative to be included in the abstract" We have removed this point from the abstract, but we strongly believe that while speculative, this point needs to be included in the DISCUSSION.

"There are two FANCD genes: D1 and D2. ID should be replaced with ID2 throughout the manuscript." Corrected

Page 3, line 68: Fanconi anaemia and Fanconi syndrome are distinct clinical entities. Remove 'the' and 'syndrome' from line 68. Corrected

Page 5, lines 97-99 and 99-101: Insert the relevant citations. Corrected

Page 6, line 122: The most widely mutated FA genes are FANCA, FANCC, and FANCG. Corrected

Spell out and/or explain CADD filtering for pathogenicity. Ref 33 now included in the RESULTS section.

Page 10, line 155: 16 or 19 FA genes? Correct as stated. Requirement to match to the previous study.

Page 10, line 174: How can the authors make this conclusion based on the distribution of mutations? This statement has now been removed.

Page 10, lines 175-178. This statement is largely speculative and not supported by the data presented in Table 3. What literature evidence is the author referring to? This statement has been re-worded to better align with the data in the Table. Table has been moved to Suppl Information (now **Suppl Table S4**)

Page 13, line 207. How does the absence of mutant allele expression confirm haploinsufficiency? The discussion of haploinsufficiency has been modified and now better covers the limitations of mutant-allele detection in tumour material.

"Page 14. For the gene expression profiling analysis, while I realize that these results are potentially interesting, there is no experimental validation for any of this data and these findings remain largely speculative. For example, experimental validation of increased activity of the base excision repair pathway upon disruption of the FA pathway would be quite novel" See above. Expression differences for individual genes are shown in the AML samples (**Fig 2B**).

SUPPLEMENTARY INFORMATION

Rare variants in Fanconi Anaemia complex genes increase risk for Acute Myeloid Leukaemia

Maung *et al*

TABLE OF CONTENTS:

1. SUPPLEMENTARY MATERIALS AND METHODS

Control Cohort. For case-control studies comparing mutation frequencies to that in healthy individuals we used WES data generated from a normal healthy cohort (n=799) ref (1). This control cohort is all female of Caucasian ethnic origin with mean age of 67yrs (46-86yrs). These samples have been sequenced on the Illumina TrueSeq Enrichment kit v2.0 (Illumina, San Diego, CA, USA) and analysed using the same methods as described below. For the Australian AML patients, 323 of these controls with a mean coverage of (27.5x) were used for the burden analysis against the AML cohort. In addition, a cohort of 49 germline samples, sequenced using Illumina Nextera Rapid exome capture, were used to control for potential bias between the two capture kits, but were not used in the burden analysis (see methods below). For the replication study with TCGA AML sequence data, we used the other 476 samples from the healthy cohort, these had a mean coverage of 23.9x.

Whole Exome Sequencing of AML samples. Genomic DNA was extracted from patient AML samples or mesenchymal stromal cells using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Genomic DNA was sonicated and DNA sequencing libraries were constructed using a preparation kit for paired-end sequencing (Illumina) as per the manufacturer's protocol. Liquid-phase hybridization for exome capture was performed using the Illumina TruSeq Exome Enrichment Kit v2.0 (Illumina) (n=89) or Illumina Nextera Rapid (FC-140-1003, Illumina) (n=35). Efficiency of sequence capture was assessed using quantitative real-time PCR with standard control primers as recommended by the manufacturer and quality of each DNA sequencing library was assessed using a 2100 Bioanalyzer and DNA 1000 chip kits (Agilent Technologies, Santa Clara, CA, USA). DNA concentration was standardized to 10 nM for sequencing. Massive parallel sequencing was performed using the Illumina HiSeq 2000 configured for paired-end reads. The 89 Illumina TrueSeq AML exomes had a mean coverage of 57x (26-102x), while the 35 Illumina Nextera AML exomes had a mean coverage of 47.6x (22-144x). Exome sequencing and variant calling was performed at the UQ Centre for Clinical Genomics.

Base calling and variant filtering. Initial base calling was performed using the CASAVA 1.7 data analysis pipeline software (Illumina). Sequence data were aligned to the current build hg19 of the human genome using the Novoalign alignment tool [V2.07.09 1] ref (2). Sequence alignment files were converted using SAMtools [v0.1.14] (ref(3)) and Picard tools (v1.42). SNPs and indels for all 131 AML samples were simultaneously called using the best practice protocols described for Genome Analysis Toolkit (GATK v3.5-2 for Australian AML samples

and GATK v3.2-2 for TCGA replication samples). This genotyping included sample level Indel realignment and variant quality score recalibration (VSQ). Genotypes were annotated using ANNOVAR (4) using Refseq, Ensembl and UCSC transcript definitions and the Ensembl Variant Effect Predictor (VEP). Sequencing data were analysed and filtered using custom scripts employing R and Bioconductor. Good quality SNPs and indels (VSQ: FILTER=PASS) were retained. We describe additional sample and genotype level filtering applied for the burden analyses below.

For several of the subsequent analyses we defined rare variants as those with a population frequency/minor allele frequency (MAF) <0.001 reported in dbSNP147, 1000 genomes (April 2012 version, any ethnicity) and the 6500 NHLBI-ESP project (any ethnicity). We include monomorphic variants and those reported in dbSNP without a population frequency in this category. Deleterious variants were defined if ANNOVAR (using any of the 3 transcript definitions), or the VEP, predicted a mutation to be non-synonymous, splicing (including splice donor and splice acceptor variants defined by VEP), stop-gain, stop-loss, a frameshift or non-frameshift substitution or deletion, an initiator codon variant, a stop retained variant or an incomplete terminal codon variant. We additionally excluded loci where coverage was low and/or the missing rate was high across the cohort. These loci were identified by first filtering genotypes with less than 7 reads supporting heterozygous calls, and 2 for homozygous calls, before calculating the missing rate. If the missing rate exceeded 80%, all genotypes at that loci were excluded. This approach removes low coverage regions where genotypes can be unreliable and discrimination between homozygous and heterozygous calls is poor. We are aware that sequencing artefacts can occur even with this filtering, often appearing as novel or rare mutations (in a population sense) occurring at high frequency in a genotyped cohort. To help identify these, we applied a Hardy-Weinberg p-value filter of 10^{-6} to the controls and additionally excluded loci where the allele frequency in the control genotypes greatly exceeded that of the population frequency filter. We defined this as 6 standard deviations away from the population threshold of 0.001. Under the binominal approximation, this is defined by ("control allele count"- n*p)/sqrt(n*p*(1-p))<=6. For example, if n=400 and *P*=0.001, this restricts our analysis to loci with 3 or less alleles reported in the control cohort. We additionally excluded loci where the mean genotype quality (GQ) score was <50 in samples where an alternative allele was predicted. To control for potential biases generated by different exome capture methods and different cohorts, we compared both genotype counts between control groups and modelled the background allele frequency between the different cohorts. If the genotype counts differed between Illumina and Nextera control cohorts under the binomial approximation with *P*<0.001, we excluded that loci. We modelled the background allele frequency for all loci using samples

with homozygous reference genotypes, essentially using the alternative allele frequency to estimate a position dependent background sequencing error rate within each cohort. Differences in that background error rate were used to identify loci where potential batch effects might occur. We chose to exclude loci where the background error rates differed by 6 standard deviation in loci where there was a least 100 reads from the cohorts. Control cohorts on Illumina and Nextera Captures were compared in the analysis for the Australian AML cases and background error rates between control and TCGA was made in the replication experiment. One loci in the FANC gene group failed this filtering in the analysis for the Australian AML cohort, but none in the TCGA cohort analysis. For further pathogenicity filtering we used the Combined Annotation Dependent Depletion algorithm (CADD score >10) ref (5). Selected FANC gene variants were validated in matched diagnostic and germline material by conventional Sanger sequencing to establish somatic status.

Analysis of FANC variant enrichment in AML cases vs controls. To compare the frequency of uncommon, deleterious protein coding variants affecting the FANC gene group (22 genes) and individual FANC genes, we performed burden analysis between the Australian AML patient cohort (n=131) and the ethnically matched (Caucasian) healthy female control population (n=323). Burden analysis (6) was performed using the SkatMeta program in R with no allele weighting. Because the control cohort was all female, we excluded genes on the X or Y chromosomes to remove any gender bias. Ethnicity was confirmed by converting the control and AML samples to PLINK genotype format and merging with 1000 genome control samples of known ethnicities. After removing regions with known long-range LD, principle component analysis using shellfish (http://www.stats.ox.ac.uk/~davison/software/shellfish/shellfish.php) was then preformed to remove ethnic outliers (greater than 6 standard deviation from the weighted mean of Caucasian 1000 genome samples). Residual population stratification was controlled for by using the first 4 principle component eigenvectors as covariates in the burden tests. However, we found that results did change significantly if Caucasian samples where used in burden test without additional population correction using the eigenvectors, suggesting that the remaining 131 AML cases and 323 controls were ethnically matched.

For comparison of the TCGA AML cohort with healthy individuals, we compared variants identified in the published sequences from an ethnically-matched (reported as Caucasian) cohort of 102 AML cases from the TCGA consortium (7). Genotyping was performed using the haplotype callers (version 3.0), again with GATK best practices and otherwise annotated and analysed using the same pipeline and quality control procedures as above.

Identification of Disease-causing (D-C) mutations. D-C mutations were identified by cross-

referencing to the FA database (Rockefeller University, FAMutdb; http://www.rockefeller.edu/fanconi/), the Kathleen Cuningham Foundation Consortium for research into Familial Breast (kConFab; http://www.kconfab.org/Index.shtml), and the National Institute of Health (NIH)'s Breast Cancer Information Core (BIC) database (https://research.nhgri.nih.gov/bic/). Mutations annotated as FA haplotype or highlighted in red in the FA database (breast cancer associated), annotated as pathogenic in the kConFab database, or annotated as Class 5 mutation in the BIC database were classified as D-C.

Odds ratio calculation. Odds ratios were calculated using Stata/IC 12.1 software (StataCorp, College Station, TX, USA).

Gene expression profiles of FA/BRCA-HR mutant AML. For gene expression analysis we initially analysed a dataset consisting 139 Illumina HumanHT12_V4_0_R2_15002873_B Bead Arrays. Probe-level data was loaded using the R package lumi and Illumina Probe identifiers were also converted to nuID labels for compatibility with the analysis packages. Array quality on this set of 139 (130 AML and 9 CD34+) arrays was assessed using the R package arrayQualityMetrics. Distances between arrays were calculated by taking the mean of the absolute values of the differences in log2 intensities between a pair of arrays. Utilizing this method, 3 samples were consistently called as outliers and were removed leaving a total of 136 arrays for downstream analysis. The data was background-corrected using the Negative Control Probes to estimate the parameters of a Robust Multi-Array Average (RMA) like model in the R package mbcb. Background corrected data was log2-transformed to help minimise the impact of increasing variance with increasing signal, and the dataset was normalised using Robust Spline Normalisation (RSN). Probes with sequences considered to be poor matches for the corresponding target, or with no matches were removed. Detection Above Background (DABG) was also performed and probes with detection p-values >0.05 in more than half of the samples were removed leaving a total of 15,484 probes for down-stream analysis. For baseline expression levels, 9 healthy $CD34^+$ cell populations were used. Of the 136 arrays analysed, WES mutation data for the FA/BRCA-HRR network was available for 57 samples and these were utilized for further differential gene expression analysis. For this, annotations were transformed from categorical to binary variables and a model matrix created. Probe-level weights were assigned as the number of beads from which the intensity estimate was obtained, and array-level weights were subsequently estimated using these probe-level weights along with the model matrix. On performing this analysis, we found 115 genes to be significantly differentially expressed at $P<0.05$. These genes were further filtered using a fold-change cut-off of 1.5 (Supplementary Table S5). To determine whether differentially expressed genes show

enrichment for other gene signatures we used gene set enrichment analysis GSEA (8, 9). Results of GSEA are shown in Supplementary Table S6 and Figure. 2.

qRT-PCR. RNA was extracted from AML patient's bone marrow mononuclear cells (BMMNC) and cDNA synthesised with the QuantiTect Reverse Transcription Kit (Qiagen) following manufacturer's instructions. qRT-PCR was performed with Power SYBR Green Master Mix (ThermoFisher Scientific), respective primer pair at 125nM and cDNA diluted 1/10 using the Applied Biosystems ViiA™ 7 Real-Time PCR System (Life Technologies). Results were analyzed with QuantStudio Real-Time PCR Software v1.1 (Life Technologies).

Cell cycle analysis. $2x10^5$ cells were seeded in each well of a 6-well plate and left to recover overnight. 40ng/mL of MMC or vehicle (DMSO) were added to the wells and incubated for 48 hours. The cells were harvested and resuspended in 300µL of cold 1x PBS. Cells were fixed with 700µL of 100% ethanol added to each tube dropwise (with vortexing). The tubes were incubated at 4ºC for 30 minutes, spun down and resuspended in 100µL of propidium iodide solution (4µg/mL). The tubes were incubated at 37ºC for 30 minutes (in the dark). Fluorescence was measured in a Beckman Coulter Gallios, using Gallios Cytometry List Mode Data Acquisition and Analysis Software version 1.2 (Beckman Coulter). Results were analysed with FCS Express 4 Flow Research Edition software (De Novo Software) using the Multicycle setting.

Western blot. $2x10^5$ cells were seeded in each well of a 6-well plate and left to recover overnight. 40ng/mL of MMC or vehicle (DMSO) were added to the wells and incubated for 48 hours. The cells were harvested and lysed with NP40 Lysis buffer (ThermoFisher Scientific) supplemented with cOmplete™ protease inhibitor (Roche), PhosSTOP™ (Roche) and Pefabloc® (Roche) at the recommended concentrations. Protein was quantified using the DC Protein Assay kit (BioRad) as per manufacturer's protocol. 50µg of lysate were loaded for SDS-PAGE. Protein was transferred from the gels into PVDF membranes using a semi dry transfer apparatus (BioRad). The membranes were blocked for 1h at room temperature with 5% skim milk in 0.1% Tris-buffered saline + 0.1% TritonX-100 (TBS-t) and then incubated with primary antibodies overnight at 4ºC. Antibodies were diluted as follows: anti-FANCD2 (Novus: NB100-182) 1:5000 in 5% skim milk in 0.1% TBS-t, and anti-HSP90 (Cell Signalling: 4875, clone E289) 1:2000 in 5% BSA in 0.1% TBS-t. After incubation with secondary antibodies for 1h at room temperature membranes were scanned in a Typhoon FLA 9000 (GE Healthcare) scanner.

2. SUPPLEMENTARY FIGURES

Figure S1. Correlation between percentage of patients in the Australian and TCGA AML cohorts with somatic mutations in genes recurrently mutated in AML (r=0.886).

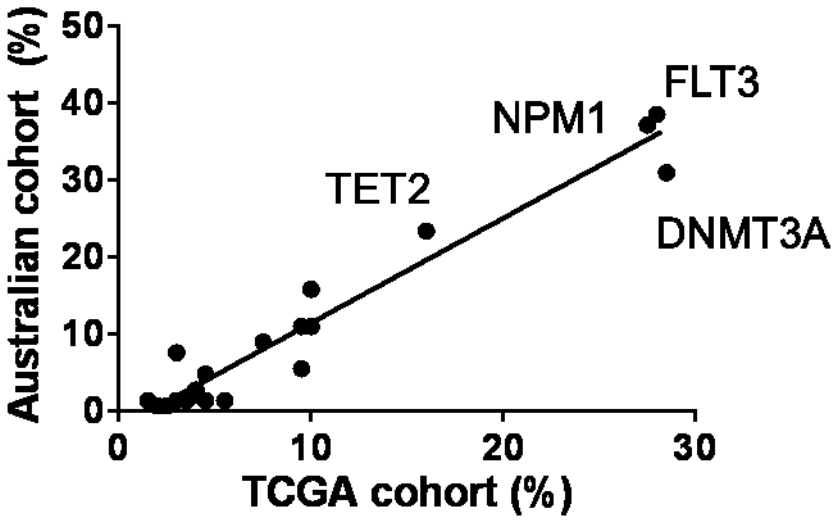
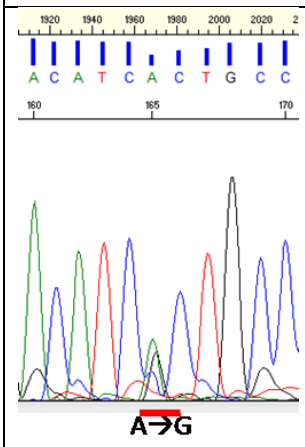
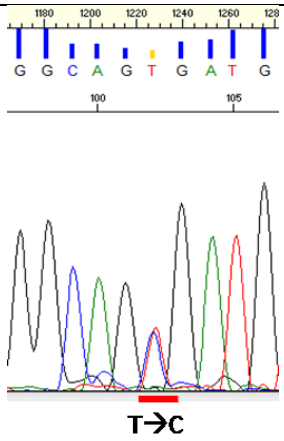
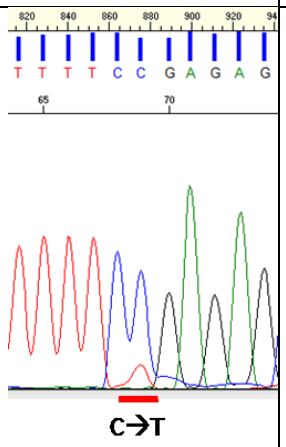
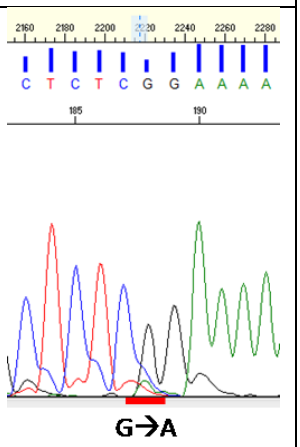
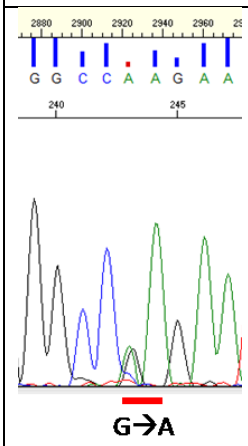
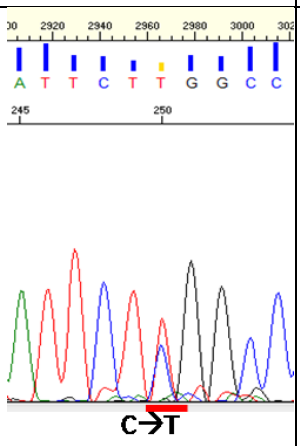
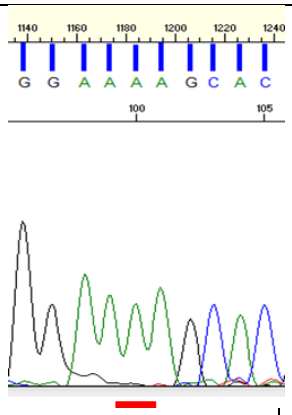
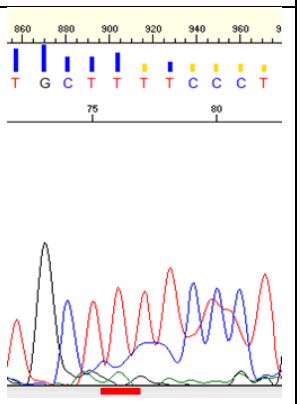
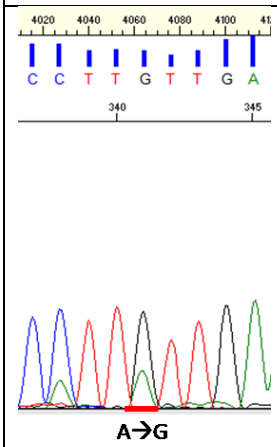
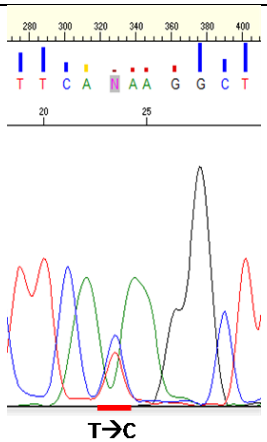
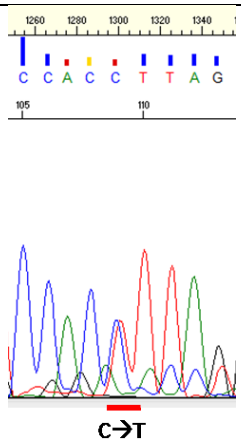
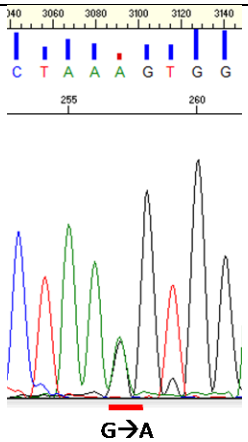
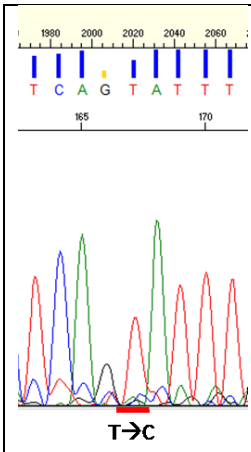
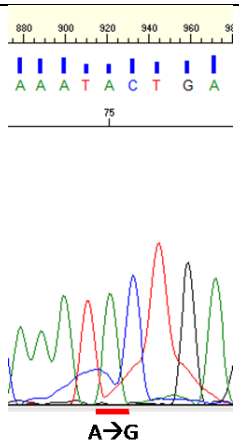


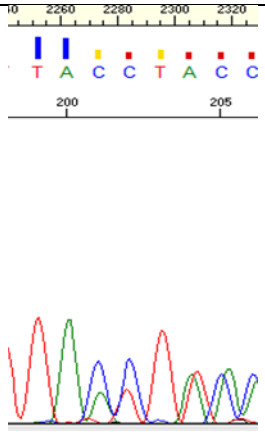
Figure S2. Expression of mutant FANC variants by Sanger sequencing. Sanger sequencing traces of cDNA of selected FANC gene variants in AML samples (n=13).

Forward sequences	Reverse sequences	Forward sequences	Reverse sequences
			
FANCA - c.A3391G:p.T1131A		**FANCD2 - c.C2776T:p.R926X**	
			
SLX4 - c.G2437A:p.E813K		**FANCM - c.A1545C:p.K515N**	
			
FANCL - c.A343G:p.I115V		**FANCL - c.C112T:p.L38F**	

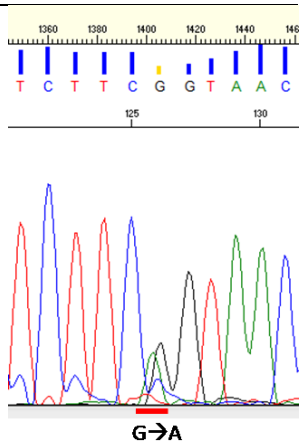




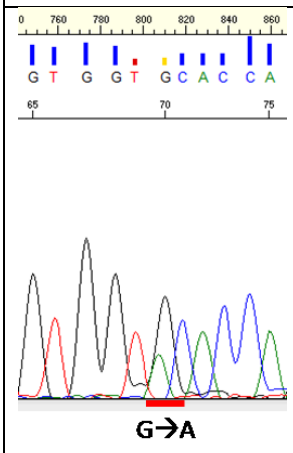
BRCA1 - c.T2566C:p.Y856H

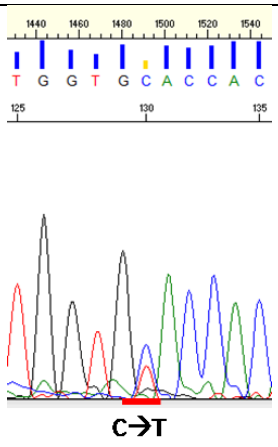


FANCL - c.1099_1100insATTA:p.T367fs

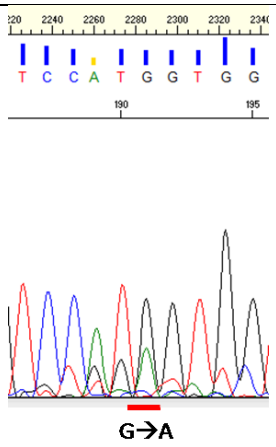


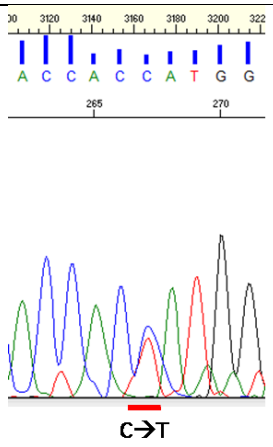
BRCA1 - c.C3607T:p.R1203X



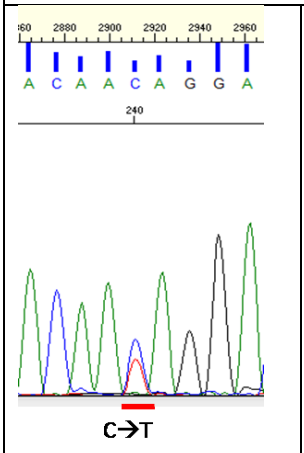


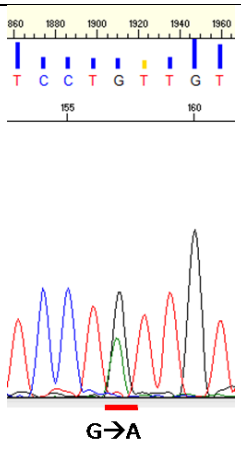
RAD51C - c.G376A:p.A126T



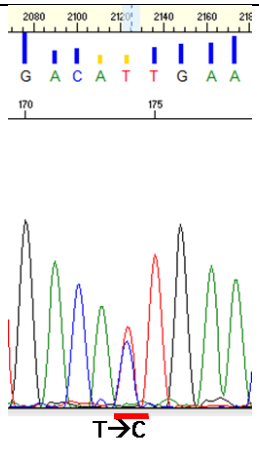


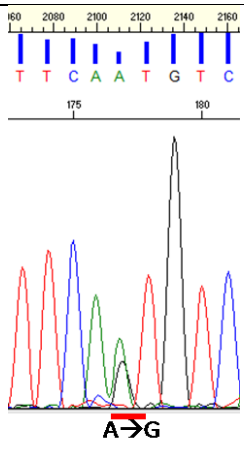
BRCA1 - c.G4956A:p.M1652I





FANCM - c.C5101T:p.Q1701X





ERCC4 - c.T2117C:p.I706T

Figure S3. Association of FANC mutations with recurrent somatic AML mutations and cytogenetics in the Australian AML Cohort. The columns in the figure represent each patient in the Australian AML cohort (n=131). The first line in the figure represents the 22 FANC genes as a group. The subsequent lines represent individual FANC genes, AML recurrently mutated genes, karyotype and cytogenetics characteristics. Patients carrying 1 or more FANC mutations (n= 44) are represented with red boxes. Patients with specific common AML mutations are represented with blue boxes. Patients with normal, abnormal (1 or 2 abnormalities) and complex (3 or more abnormalities) karyotypes are represented with green boxes. Patients with specific karyotypic abnormalities are represented by orange boxes. Del7/7q and *FLT3*-ITD are under-represented (blue highlight) in the FANC-mutant AML patient group (P=0.028 and P=0.05, respectively).

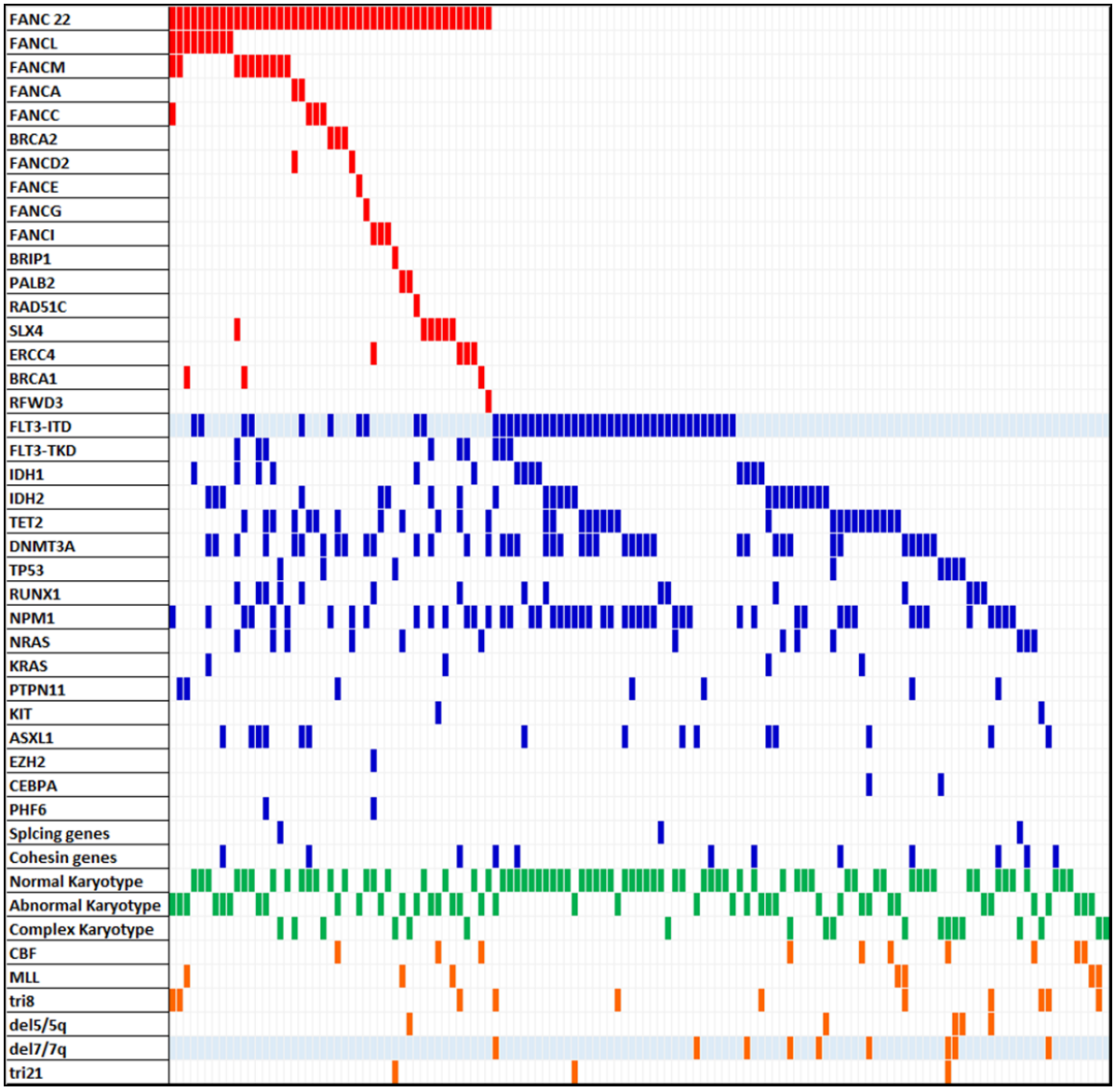


Figure S4. MCF10A CRISPR clones. MCF10A clones generated using CRISPR-Cas9. Blue font represent wild type sequence (identical to the reference sequence) and red font represent the alternate sequence obtained due to frameshift deletions from CRISPR-Cas9. Clone P1B4 had a premature stop mutation on 1 allele and an arginine was deleted on the other allele. Clone P2B5 had a four nucleotide deletion resulting in a premature stop codon on one allele. Clone P2D3 had a deletion of two nucleotides resulting in a premature stop codon on one allele. Clone P4A4 had two consecutive amino acids deleted (glutamine and asparagine) on one allele and was wild type on the second allele. Clones P1A4 and P1D3 are shown to be wild type.

Clone ID	FANCL	DNA		Protein
	REF SEQ	CAGAACCGGTCGAAAA	+15	LPQNRSKTVYEGFISAQGRDFHLRIV
P1B4	Allele 1	CAGAA--GGTCGAAAA	+15	LPQKVENRVStop
	Allele 2	CAGAA----TTCGAAAA	+15	LPQN-SKTVYEGFISAQGRDFHLRIV
P2B5	Allele 1	CAGAACCGGTCGAAAA	+15	LPQNRSKTVYEGFISAQGRDFHLRIV
	Allele 2	CAGAA----TCGAAAA	+15	LPQNRKPCMRDSSRLREETSTLGStop
P2D3	Allele 1	CAGAACCGGTCGAAAA	+15	LPQNRSKTVYEGFISAQGRDFHLRIV
	Allele 2	CAGAA--GGTCGAAAA	+15	LPQKVENRVStop
P4A4	Allele 1	CAGAACCGGTCGAAAA	+15	LPQNRSKTVYEGFISAQGRDFHLRIV
	Allele 2	C------GGTCGAAAA	+15	LP--RSKTVYEGFISAQGRDFHLRIV
P1A4	Allele 1	CAGAACCGGTCGAAAA	+15	LPQNRSKTVYEGFISAQGRDFHLRIV
	Allele 2	CAGAACCGGTCGAAAA	+15	LPQNRSKTVYEGFISAQGRDFHLRIV
P1D3	Allele 1	CAGAACCGGTCGAAAA	+15	LPQNRSKTVYEGFISAQGRDFHLRIV
	Allele 2	CAGAACCGGTCGAAAA	+15	LPQNRSKTVYEGFISAQGRDFHLRIV

Figure S5. Cell cycle analysis (n=4). (**A**) Cell cycle analysis for each of the six MCF10A *FANCL* CRISPR clones treated with 40ng/mL of MMC (t-clones) and vehicle (un-clones). Cell cycle analysis showing the G0/G1 phase (**B**), S phase (**C**) and G2/M phase (**D**) for each of the six MCF10A *FANCL* CRISPR clones treated with 40ng/mL of MMC (t-clones) and vehicle (un-clones).

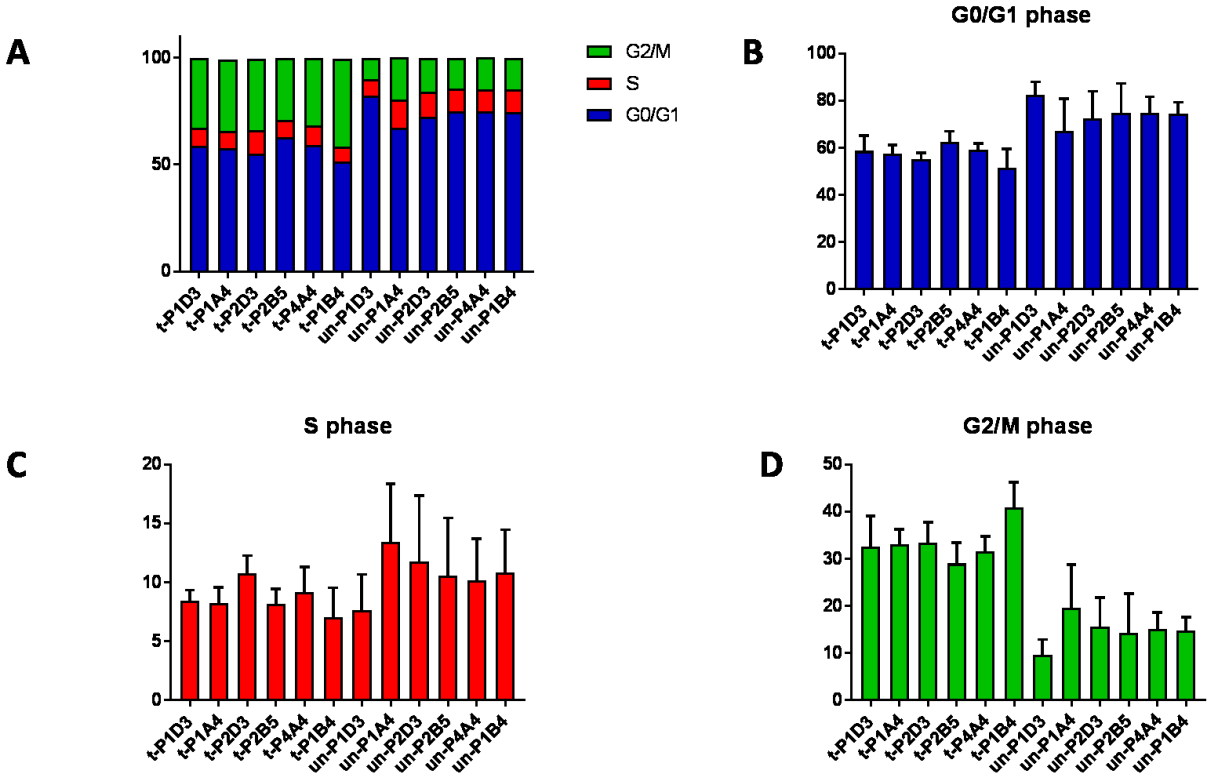
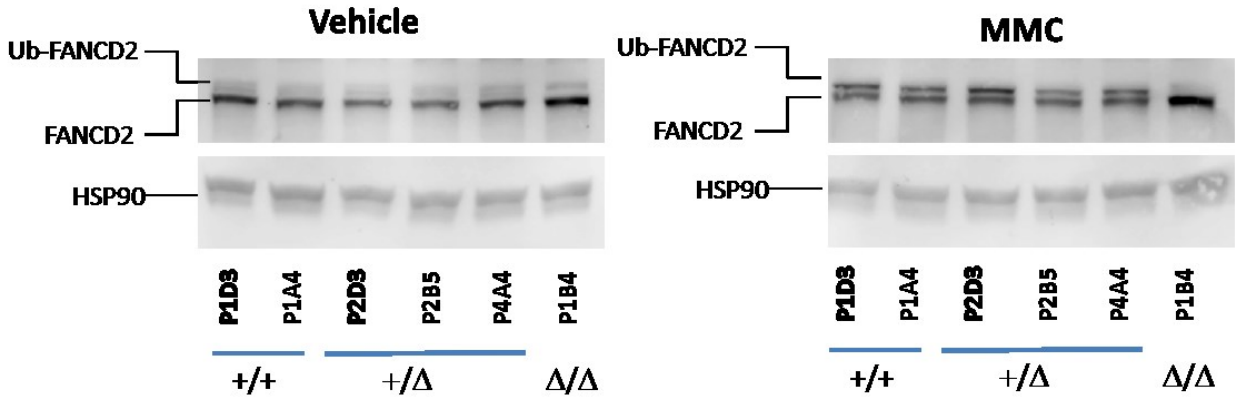


Figure S6. Western blot analysis MCF10A clones. Western blots for FANCD2 and HSP90 (loading control) of the six MCF10A CRISPR clones treated with vehicle or MMC (40ng/mL) for 48 hours.



3. SUPPLEMENTARY TABLES

Table S1. Characteristics of the Australian AML Cohort

	All Cases (n=131)	Mutant Group (n=45)	Non-Mutant Group (n=86)	[1]P value
Age - median (range)	55 (17-89)	58 (18-84)	54 (17-89)	0.8055^
Male – n / total (%)	80 / 131 (61.1%)	29 / 45 (64.4%)	51 / 86 (59.3%)	0.7062
Female – n / total (%)	51 / 131 (38.9%)	16 / 45 (35.6%)	35 / 86 (40.7%)	0.7062
WCC x10^9/L - median (range)	19 (1.07-313.3)	26 (1.22-227)	15.45 (1.07-313.3)	0.8793^
BM Blast % - median (range)	80.75 (50-100)	77 (50-100)	83.5 (50-99)	0.2869^
Primary/Secondary AML – n / total (%)				
De Novo	75 / 83 (90.4%)	26/30 (86.7%)	49/53 (92.5%)	0.4514
Secondary	8 / 83 (9.6%)	4 /30(13.3%)	4/53 (7.5%)	0.4514
Unknown	48			
Transplant – n / total (%)				
Yes	23 / 89 (25.8%)	10 / 32 (31.2%)	13 / 57 (22.8%)	0.4521
No	66 / 89 (74.2%)	22 / 32 (68.8%)	44 / 57 (77.2%)	0.4521
Unknown	42			
[2]FAB – n / total (%)				
M0	4 / 87 (4.6%)	0 / 32 (0%)	4 / 55 (7.3%)	0.2932
M1	31 / 87 (35.6%)	9 / 32 (28.1%)	22 / 55 (40%)	1
M2	17 / 87 (19.5%)	7 / 32 (21.9%)	10 / 55 (18.2%)	0.7808
M3	0 / 87 (0%)	0 / 32 (0%)	0 / 55 (0%)	1
M4	17 / 87 (19.5%)	8 / 32 (25%)	9 / 55 (16.4%)	0.4032
M5	13 / 87 (14.9%)	5 / 32 (15.6%)	8 / 55 (14.5%)	1
M6	0 / 87 (0%)	0 / 32 (0%)	0 / 55 (0%)	1
M7	1 / 87 (1.1%)	1 / 32 (3.1%)	0 / 55 (0%)	0.3678
Not classified	4 / 87 (4.6%)	2 / 32 (6.3%)	2 / 55 (3.6%)	0.6196
Unknown	44			
[3]ELN Cytogenetic Risk – n/total (%)				
Good	11 / 54 (20.4%)	5 / 21 (23.8%)	6 / 33 (18.2%)	0.7329
Intermediate-1	11 / 54 (20.4%)	3 / 21 (14.3 %)	8 / 33 (24.2%)	0.4974
Intermediate-2	13 / 54 (24.1%)	9 / 21 (42.9%)	4 / 33 (12.1%)	**0.02***
Adverse	19 / 54 (35.2%)	4 / 21 (19%)	15 / 33 (45.5%)	0.0786
Unknown	77			
[4]Grimwade Cytogenetic Risk – n/total (%)				
Good	7 / 122 (5.7%)	2 / 40 (5%)	5 / 82 (6.1%)	1
Intermediate	88 / 122 (72.1%)	31 / 40 (77.5%)	57 / 82 (69.5%)	0.3970
Poor	27 / 122 (22.1%)	7 /40 (17.5%)	20 / 82 (24.4%)	0.4886
Unknown	9			
Simple Karyotype – n/total (%)				
Normal	70 / 129 (54.3%)	20 / 44 (44.5%)	50 / 85 (58.8%)	0.1921
Abnormal	40 / 129 (31.0%)	18 / 44 (40.9%)	22 / 85 (25.9%)	0.1079
Complex	19 / 129 (14.7%)	6 / 44 (13.6%)	13 / 85 (15.3%)	1
Unknown	2			

Cytogenetics – n/total (%)				
t(15;17)	0 / 129 (0%)	0 (0%)	0 (0%)	1
CBF	10 / 129 (7.75%)	3 / 44 (6.82%)	7 / 85 (8.24%)	1
MLL	7 / 129 (5.43%)	3 / 44 (6.82%)	4 / 85 (4.71%)	0.6895
tri(8)	11 / 129 (8.53%)	3 / 44 (6.82%)	8 / 85 (9.41%)	0.7481
mono(5) / del(5q)	5 / 129 (3.88%)	1 / 44 (2.27%)	4 / 85 (4.71%)	0.6605
mono(7) / del(7q)	9 / 129 (6.98%)	0 / 44 (0%)	9 / 85 (10.59%)	**0.0275***
tri(21)	3 / 129 (2.33%)	1 / 44 (2.27%)	2 / 85 (2.35%)	1
Mutations – n/total (%)				
FLT3-ITD	44 / 131 (33.6%)	10 / 45 (22.2%)	34 / 86 (39.5%)	0.0531
FLT3-TKD	7 / 131 (5.34%)	4 / 45 (8.9%)	3 / 86 (3.5%)	0.2315
NPM1	46 / 131 (35.1%)	12 / 45 (26.7%)	34 / 86 (39.5%)	0.1784
DNMT3A	41 / 131 (31.3%)	15 / 45 (33.3%)	26 / 86 (30.2%)	0.8429
IDH1	14 / 131 (10.7%)	6 / 45 (13.3%)	8 / 86 (9.3%)	0.5549
IDH2	24 / 131 (18.3%)	9 / 45 (20%)	15 / 86 (17.4%)	0.8128
TET2	21 / 131 (16%)	7 / 45 (15.6%)	14 / 86 (16.3%)	1

[1]P values are calculated by Fisher's exact test except for: ^ determined by Student's *t*-test.

[2]FAB: French-America-British classification (10)

[3]ELN: European LeukaemiaNet (11)

[4]Grimwade classification (12)

**P*<0.05

Table S2. List of FANC variants in AML (Australian cohort)

ID	chr	start	end	REF	ALT	TYPE	Gene	Ref Seq	Exon	Nucleotide Change	Amino acid change	Consequence	CADD Score	No. of Reads (Ref)	No. of Reads (Alt)	% of Alt Reads	Disease Causing variant (FAmutDB)
WES-8	chr2	58386928	58386928	-	TAAT	indel	FANCL	NM_018062	exon14	c.1099_1100insATTA	p.T367fs	frameshift_variant	38	11	9	45.0	Yes
WES-13	chr17	41222975	41222975	C	T	snp	BRCA1	NM_007298	exon14	c.G1644A	p.M548I	missense_variant	22.3	17	16	48.5	No
WES-14	chr2	58459232	58459232	G	A	snp	FANCL	NM_018062	exon2	c.C112T	p.L38F	missense_variant	25.3	20	7	25.9	No
WES-18	chr2	58449108	58449108	T	C	snp	FANCL	NM_018062	exon5	c.A343G	p.I115V	missense_variant	14.1	14	6	30.0	No
WES-20	chr9	97912307	97912307	T	A	snp	FANCC	NM_001243743	exon7	c.A584T	p.D195V	missense_variant	25.2	37	35	48.6	Yes
WES-21	chr16	89813256	89813256	T	C	snp	FANCA	NM_000135	exon34	c.A3391G	p.T1131A	missense_variant	23.5	47	31	39.7	Yes
WES-21	chr3	10116274	10116274	C	T	snp	FANCD2	NM_033084	exon29	c.C2776T	p.R926X	stop_gained	46	14	10	41.7	Yes
WES-23	chr16	3641280	3641280	C	T	snp	SLX4	NM_032444	exon12	c.G2359A	p.E787K	missense_variant	21.6	26	27	50.9	No
WES-26	chr14	45606290	45606290	C	T	snp	FANCM	NM_020937	exon2	c.C527T	p.T176I	missense_variant	20.2	24	18	42.9	No
WES-28	chr17	59938933	59938933	A	G	snp	BRIP1	-	-	-	-	splice_region_variant	15.04	13	20	60.6	No
WES-30	chr2	58388668	58388670	ATA	-	indel	FANCL	NM_018062	exon12	c.1007_1009del	p.336_337del	missense_variant	22.8	14	15	51.7	Yes
WES-33	chr17	56772522	56772522	G	A	snp	RAD51C	NM_002876	exon2	c.G376A	p.A126T	missense_variant	21.7	38	40	51.3	No
WES-34	chr16	14041570	14041570	T	C	snp	ERCC4	NM_005236	exon11	c.T2117C	p.I706T	missense_variant	27.2	69	59	46.1	No
WES-37	chr16	14026059	14026059	G	A	snp	ERCC4	NM_005236	exon6	c.G1019A	p.R340Q	missense_variant	34	6	3	33.3	No
WES-38	chr16	14028081	14028081	C	T	snp	ERCC4	NM_005236	exon7	c.C1135T	p.P379S	missense_variant	32	39	21	35.0	No
WES-40	chr14	45665603	45665603	G	A	snp	FANCM	NM_020937	exon21	c.G5569A	p.V1857M	missense_variant	27.7	29	26	47.3	No
WES-41	chr2	58468399	58468399	G	C	snp	FANCL	NM_018062	exon1	c.C50G	p.P17R	missense_variant	22.2	51	35	40.7	No
WES-45	chr16	74657852	74657852	T	C	snp	RFWD3	NM_018124	exon13	c.A2299G	p.M767V	missense_variant	10.21	53	39	42.4	No
WES-46	chr17	41243941	41243941	G	A	snp	BRCA1	NM_007294	exon10	c.C3607T	p.R1203X	stop_gained	35	52	45	46.4	Yes (kConFab)
WES-46	chr14	45658326	45658326	C	T	snp	FANCM	NM_020937	exon20	c.C5101T	p.Q1701X	stop_gained	35	46	47	50.5	No
WES-47	chr16	89877182	89877182	G	C	snp	FANCA	NM_001018112	exon5	c.C455G	p.A152G	missense_variant	13.73	83	47	36.2	No
WES-55	chr16	3642715	3642715	C	G	snp	SLX4	NM_032444	exon11	c.G2312C	p.S771T	missense_variant	15.99	22	20	47.6	No
WES-57	chr16	3641202	3641202	C	T	snp	SLX4	NM_032444	exon12	c.G2437A	p.E813K	missense_variant	28.2	28	33	54.1	No
WES-59	chr15	89804921	89804921	A	G	snp	FANCI	NM_018193	exon5	c.A394G	p.I132V	missense_variant	10.54	43	29	40.3	No
WES-60	chr17	41222975	41222975	C	T	snp	BRCA1	NM_007298	exon14	c.G1644A	p.M548I	missense_variant	22.3	16	13	44.8	No
WES-60	chr2	58386928	58386928	-	TAAT	indel	FANCL	NM_018062	exon14	c.1099_1100insATTA	p.T367fs	frameshift_variant	38	11	9	45.0	Yes
WES-64	chr9	35076755	35076755	G	A	snp	FANCG	NM_004629	exon7	c.C890T	p.T297I	missense_variant	21.6	72	80	52.6	Yes
WES-73	chr16	3647893	3647893	G	A	snp	SLX4	NM_032444	exon6	c.C1271T	p.A424V	missense_variant	27.5	18	20	52.6	No
WES-78	chr14	45628478	45628478	C	G	snp	FANCM	NM_020937	exon9	c.C1576G	p.L526V	missense_variant	21.1	20	28	58.3	No
WES-79	chr9	97912307	97912307	T	A	snp	FANCC	NM_001243743	exon7	c.A584T	p.D195V	missense_variant	25.2	49	47	49.0	Yes
WES-81	chr14	45644816	45644816	A	C	snp	FANCM	NM_020937	exon14	c.A2859C	p.K953N	missense_variant	22.2	24	16	40.0	No
WES-81	chr16	3633255	3633255	G	A	snp	SLX4	NM_032444	exon14	c.C4996T	p.R1666X	stop_gained	36	40	37	48.1	No
WES-86	chr9	97912307	97912307	T	A	snp	FANCC	NM_001243743	exon7	c.A584T	p.D195V	missense_variant	25.2	39	40	50.6	Yes
WES-86	chr2	58456962	58456962	C	G	snp	FANCL	NM_018062	exon3	c.G203C	p.R68P	missense_variant	22.5	8	13	61.9	No
WES-86	chr14	45644816	45644816	A	C	snp	FANCM	NM_020937	exon14	c.A2859C	p.K953N	missense_variant	22.2	21	11	34.4	No
WES-89	chr14	45628478	45628478	C	G	snp	FANCM	NM_020937	exon9	c.C1576G	p.L526V	missense_variant	21.1	14	7	33.3	No
WES-94	chr14	45645955	45645955	A	-	indel	FANCM	NM_020937	exon14	c.3998delA	p.Q1333fs	frameshift_variant	22.3	21	20	48.8	No
WES-201	chr16	14015897	14015897	A	G	snp	ERCC4	NM_005236	exon2	c.A217G	p.I73V	missense_variant	22.2	34	40	54.1	No
WES-201	chr15	89807213	89807213	G	C	snp	FANCI	NM_018193	exon8	c.G625C	p.E209Q	missense_variant	25.8	100	84	45.7	No
WES-202	chr14	45606290	45606290	C	T	snp	FANCM	NM_020937	exon2	c.C527T	p.T176I	missense_variant	20.2	34	25	42.4	No
WES-216	chr2	58388668	58388670	ATA	-	indel	FANCL	NM_018062	exon12	c.1007_1009del	p.336_337del	missense_variant	22.8	56	54	49.1	Yes
WES-222	chr16	23614892	23614892	A	C	snp	PALB2	NM_024675	exon13	c.T3449G	p.L1150R	missense_variant	15.08	65	54	45.4	No
WES-224	chr9	97864024	97864024	G	A	snp	FANCC	NM_001243743	exon15	c.C1642T	p.R548X	stop_gained	37	68	40	37.0	Yes
WES-225	chr15	89807836	89807836	C	T	snp	FANCI	NM_018193	exon9	c.C753T	p.D251D	splice_region_variant	14.29	41	36	46.8	No
WES-226	chr16	23652433	23652433	T	C	snp	PALB2	NM_024675	exon1	c.A46G	p.K16E	missense_variant	22.3	15	17	53.1	No
WES-227	chr16	3640664	3640664	C	T	snp	SLX4	NM_032444	exon12	c.G2975A	p.G992E	missense_variant	16.23	26	23	46.9	No
WES-231	chr2	58388659	58388659	C	T	snp	FANCL	NM_018062	exon12	c.G1018A	p.E340K	missense_variant	22.5	89	60	40.3	No
WES-231	chr14	45628478	45628478	C	G	snp	FANCM	NM_020937	exon9	c.C1576G	p.L526V	missense_variant	21.1	41	42	50.6	No

WES-236	chr13	32907277	32907277	T	G	snp	BRCA2	NM_000059	exon10	c.T1662G	p.C554W	missense_variant	22.9	55	39	41.5	No
WES-245	chr13	32914592	32914592	C	T	snp	BRCA2	NM_000059	exon11	c.C6100T	p.R2034C	missense_variant	22.2	58	64	52.5	No
WES-247	chr6	35427531	35427531	T	C	snp	FANCE	NM_021922	exon7	c.T1310C	p.M437T	missense_variant	23.4	25	29	53.7	No
WES-249	chr3	10115047	10115047	G	A	snp	FANCD2	NM_033084	exon28	c.2715+1G>A	-	splice_donor_variant	27	118	108	47.8	Yes

Table S3. Characteristics of the Australian cohort Normal Karyotype AML

	All Cases (n=70)	Mutant Group (n=20)	Non-Mutant Group (n=50)	[1]P value
Age - median (range)	52.5 (17-84)	55.5 (18-84)	52 (17-81)	0.9856^
Male – n / total (%)	51 / 70 (72.9%)	13 / 20 (65%)	38 / 50 (76%)	0.3825
Female – n / total (%)	19 / 70 (27.1%)	7 / 20 (35%)	12 / 50 (24%)	0.3825
WCC x10^9/L - median (range)	23.3 (1.22-284)	27.3 (1.22-132)	22.1 (1.3-284)	0.4521^
BM Blast % - median (range)	82 (50-98)	74.5 (50-92)	87.5 (50-98)	0.0527^
Primary/Secondary AML – n / total (%)				
De Novo	75 / 41 (90.4%)	12 / 13 (92.3%)	26 / 28 (92.9%)	1
Secondary	8 / 41 (9.6%)	1 / 13(7.7%)	2 / 28 (7.1%)	1
Unknown	36			
Transplant – n / total (%)				
Yes	14 / 42 (33.3%)	5 / 14 (35.7%)	9 / 28 (32.1%)	0.7228
No	28 / 42 (66.7%)	9 / 14 (64.3%)	19 / 28 (67.9%)	0.7228
Unknown	28			
[2]FAB – n / total (%)				
M0	0 / 41 (0%)	0 / 14 (0%)	0 / 27 (7.3%)	1
M1	17 / 41 (41.5%)	3 / 14 (21.4%)	14 / 27 (40%)	0.0958
M2	11 / 41 (26.9%)	6 / 14 (42.9%)	5 / 27 (18.2%)	0.1403
M3	0 / 41 (0%)	0 / 14 (0%)	0 / 27 (0%)	1
M4	9 / 41 (22%)	3 / 14 (21.4%)	6 / 27 (16.4%)	1
M5	2 / 41 (4.9%)	1 / 14 (7.1%)	1 / 27 (14.5%)	1
M6	0 / 41 (0%)	0 / 14 (0%)	0 / 27 (0%)	1
M7	0 / 41 (0%)	0 / 14 (3.1%)	0 / 27 (0%)	0.3678
Not classified	2 / 41 (4.9%)	1 / 14 (6.3%)	1 / 27 (3.6%)	1
Unknown	29			
Mutations – n/total (%)				
FLT3-ITD	35 / 70 (50.0%)	8 / 20 (40%)	27 / 50 (54%)	0.4279
FLT3-TKD	3 / 70 (4.3%)	1 / 20 (5%)	2 / 50 (4%)	1
NPM1	41 / 70 (58.8%)	10 / 20 (50%)	31 / 50 (62%)	0.4252
DNMT3A	27 / 70 (38.6%)	7 / 20 (35%)	20 / 50 (40%)	0.7896
IDH1	10 / 70 (14.3%)	4 / 20 (20%)	6 / 50 (12%)	0.4558
IDH2	11 / 70 (15.7%)	3 / 20 (15%)	8 / 50 (16%)	1
TET2	16 / 70 (22.9%)	5 / 20 (25%)	11 / 50 (22%)	0.7628
ASXL1	6/ 70 (8.6%)	3 / 20 (15%)	3 / 50 (6%)	0.3431

[1]P values are calculated by Fisher's exact test except for: ^ determined by Student's *t*-test.

[2]FAB: French-America-British classification (10)

[3]ELN: European LeukaemiaNet (11)

[4]Grimwade classification (12)

Table S4. Selected FA Core and ID2 gene variants of interest in the Australian AML cohort

Gene	AML Mutation	Comments
FANCA	p.R1144W	Known FA compound heterozygous mutation (13).
FANCC	p.R548X	Occurs in multiple FA patients; severe phenotype (14); detected in both Australian and TCGA cohorts.
	p.D195V	Reported mutation in FA (15).
FANCD2	p.R926X	Known FA compound heterozygous variant (16). Recurrent somatic truncations at this residue in solid tumours (COSMIC).
	c.2715+1G>A	Known FA compound heterozygous variant (16).
FANCL	p.336_337del	Known RING domain null mutation (17).
	p.T367fs	Frame shift occurs at the same position as reported in the FA database (ID: FANCL_000003). Present in two patients in the AML cohort with early onset of disease (27 and 46 years-old).
	p.E340K	Ring domain charge-reversal, affects the binding of *FANCL* to *FANCT* (18).
	p.L38F	Patient presents with anaemia and neutropenia prior to AML diagnosis.
	p.I115V	Patient presents with neutropenia and MDS prior to AML diagnosis.
	p.P17R	Patient presents with mild neutropenia prior to AML diagnosis.
FANCM	p.Q1701X	Enriched in patients with triple-negative breast cancer (19). Also identified in a paediatric B-ALL case (20).
FANCM	p.V1857M	Reported in a breast cancer family (21).

Table S5. Frequency of FANC variants in ExAC and normal karyotype Australian AML cohorts.

	[1]ExAC (n=66740)			[1]AML (n=140)			
Gene	**Count**	**ExAC Frequency**	**Frequency (%)**	**Count**	**AML Frequency**	**Frequency (%)**	**[2]P-value**
FANCA	653	0.009784	0.978424	1	0.007143	0.714286	1
FANCB	145	0.002173	0.217261	0	0	0	1
FANCC	211	0.003162	0.316152	2	0.014286	1.428571	0.0738
BRCA2	1276	0.019119	1.911897	2	0.014286	1.428571	1
FANCD2	546	0.008181	0.8181	0	0	0	0.6339
FANCE	302	0.004525	0.452502	0	0	0	1
FANCF	178	0.002667	0.266707	0	0	0	1
FANCG	219	0.003281	0.328139	1	0.007143	0.714286	0.3698
FANCI	762	0.011417	1.141744	2	0.014286	1.428571	1
BRIP1	406	0.006083	0.608331	0	0	0	1
FANCL	321	0.00481	0.480971	2	0.014286	1.428571	0.1472
FANCM	**792**	**0.011867**	**1.186695**	**5**	**0.035714**	**3.571429**	**0.0267***
PALB2	322	0.004825	0.482469	0	0	0	1
RAD51C	157	0.002352	0.235241	0	0	0	1
SLX4	689	0.010324	1.032364	3	0.021429	2.142857	0.1775
ERCC4	669	0.010024	1.002397	2	0.014286	1.428571	0.654
RAD51	94	0.001408	0.140845	0	0	0	1
BRCA1	619	0.009275	0.92748	1	0.007143	0.714286	1
UBE2T	108	0.001618	0.161822	0	0	0	1
XRCC2	97	0.001453	0.14534	0	0	0	1
MADL2	59	0.000884	0.088403	0	0	0	1
RFWD3	281	0.00421	0.421037	0	0	0	1

[1] n represents the total number of alleles in each cohort. The non-Finnish European cohort in ExAC consists of 33370 individuals.

[2] Fisher's exact test was used to determine the statistical difference between the Australian AML cohort and the non-Finnish European cohort in ExAC. (***** $P<0.05$)

Table S6. Differentially expressed genes in FANC core and ID2 mutant vs non-mutant AML cases*

Probe_Id	Symbol	logFC	P.Value	Gene description
ILMN_1732799	CD34	2.1116	0.0061	Homo sapiens CD34 molecule
ILMN_2341229	CD34	1.8166	0.0101	Homo sapiens CD34 molecule
ILMN_1808122	LOC652377	1.2892	0.0091	N/A
ILMN_2195462	C1QTNF4	1.2777	0.0112	Homo sapiens C1q and TNF related 4
ILMN_2233539	SLC39A8	1.1438	0.0032	Homo sapiens solute carrier family 39 member 8
ILMN_1681601	SUCNR1	1.0616	0.0113	Homo sapiens succinate receptor 1
ILMN_1680902	LOC284998	1.0578	0.0073	Homo sapiens long intergenic non-protein coding RNA 1114
ILMN_1808590	GUCY1A3	1.0285	0.0258	Homo sapiens guanylate cyclase 1 soluble subunit alpha
ILMN_1673605	PRSSL1	0.9916	0.0097	Homo sapiens PRSS57 protease, serine 57
ILMN_1677723	ANGPT1	0.9705	0.0344	Homo sapiens angiopoietin 1
ILMN_1762957	LOC648868	0.9439	0.0157	Homo sapiens TCR gamma alternate reading frame protein
ILMN_1670305	SERPING1	0.9307	0.0066	Homo sapiens serpin family G member 1
ILMN_1751868	TCTEX1D1	0.9274	0.0406	Homo sapiens Tctex1 domain containing 1
ILMN_2086890	ANGPT1	0.9140	0.0268	Homo sapiens angiopoietin 1
ILMN_3247023	FLJ22536	0.9129	0.0192	Homo sapiens cancer susceptibility 15 (non-protein coding)
ILMN_1670692	LPAR4	0.8686	0.0033	Homo sapiens lysophosphatidic acid receptor 4
ILMN_1809496	COPG2	0.8252	0.0001	Homo sapiens coatomer protein complex subunit gamma 2
ILMN_1696380	GHRL	0.8023	0.0043	Homo sapiens ghrelin and obestatin prepropeptide
ILMN_1667315	STAG3L1	0.7963	0.0039	Homo sapiens stromal antigen 3-like 1 (pseudogene)
ILMN_1812070	ABCB1	0.7958	0.0424	Homo sapiens ATP binding cassette subfamily B member 1
ILMN_1760778	ENG	0.7927	0.0008	Homo sapiens homodimeric transmembrane protein endoglin

ILMN_3280998	LOC100131831	-0.9243	0.0038	N/A
ILMN_1769575	JAM3	-0.9358	0.0365	Homo sapiens junctional adhesion molecule 3
ILMN_1771800	PRKCA	-0.9485	0.0060	Homo sapiens protein kinase C alpha
ILMN_1656011	RGS1	-0.9554	0.0208	Homo sapiens regulator of G protein signalling 1
ILMN_1807662	IGF2R	-0.9835	0.0120	Homo sapiens insulin like growth factor 2 receptor
ILMN_1665761	BCL11B	-0.9934	0.0403	Homo sapiens zinc finger protein B-cell CLL/lymphoma 11B
ILMN_2109489	GZMB	-0.9945	0.0487	Homo sapiens granzyme B
ILMN_1810274	HOXB2	-0.9993	0.0419	Homo sapiens nuclear protein homeobox B2
ILMN_2083469	IRS2	-1.0009	0.0008	Homo sapiens signalling molecule insulin receptor substrate 2
ILMN_2339955	NR4A2	-1.0132	0.0138	Homo sapiens nuclear receptor subfamily 4 group A member 2
ILMN_1652379	SUCLG2	-1.0296	0.0076	Homo sapiens beta subunit of succinyl-CoA synthetase
ILMN_1660462	MCOLN2	-1.0466	0.0108	Homo sapiens mucolipin 2
ILMN_1789733	CLIP3	-1.0722	0.0294	Homo sapiens cytoplasmic linker protein CAP-Gly domain containing linker protein 3
ILMN_1728106	TNF	-1.1766	0.0116	Homo sapiens proinflammatory cytokine tumour necrosis factor
ILMN_1702691	TNFAIP3	-1.2112	0.0001	Homo sapiens TNF alpha induced protein 3
ILMN_1782419	GNG11	-1.2435	0.0159	Homo sapiens G protein subunit gamma 11
ILMN_1651826	BASP1	-1.2651	0.0163	Homo sapiens membrane-bound protein brain abundant membrane attached signal protein 1
ILMN_2169801	TPSAB1	-1.6353	0.0144	Homo sapiens tryptase alpha/beta 1

*Differentially-expressed genes were identified from comparison of gene expression in patient samples with (n=14) or without (n=43) mutations in FANC core and ID2 complex genes. Table shows genes with an expression change of ≥ 1.5-fold at an adjusted P value of <0.05.

Table S7. Gene set enrichment analysis of differential gene expression between FANC core and ID2 mutant vs non-mutant AML cases*

Gene Set	NES	NOM p-val	FDR q-val
REACTOME_DNA_STRAND_ELONGATION	2.3923	0	0.0000
KEGG_DNA_REPLICATION	2.3201	0	0.0000
REACTOME_EXTENSION_OF_TELOMERES	2.2070	0	0.0017
REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX	2.2041	0	0.0013
REACTOME_LAGGING_STRAND_SYNTHESIS	2.1541	0	0.0027
KEGG_BASE_EXCISION_REPAIR	2.1228	0	0.0034
HONMA_DOCETAXEL_RESISTANCE	2.0967	0	0.0050
JAATINEN_HEMATOPOIETIC_STEM_CELL_UP	2.0893	0	0.0046
REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS	2.0649	0	0.0061
REACTOME_G2_M_CHECKPOINTS	2.0641	0	0.0057
REACTOME_BASE_EXCISION_REPAIR	2.0497	0	0.0063
REACTOME_DNA_REPAIR	2.0483	0	0.0057
SONG_TARGETS_OF_IE86_CMV_PROTEIN	2.0438	0	0.0058
KEGG_PURINE_METABOLISM	2.0136	0	0.0081
DUTERTRE_ESTRADIOL_RESPONSE_24HR_UP	2.0047	0	0.0082
NIKOLSKY_BREAST_CANCER_11Q12_Q14_AMPLICON	1.9904	0	0.0099
ALCALAY_AML_BY_NPM1_LOCALIZATION_DN	1.9714	0	0.0124
BUYTAERT_PHOTODYNAMIC_THERAPY_STRESS_DN	1.9552	0	0.0147
LUI_THYROID_CANCER_CLUSTER_3	1.9346	0	0.0186
HOLLEMAN_ASPARAGINASE_RESISTANCE_B_ALL_UP	1.9255	0	0.0202
NAGASHIMA_NRG1_SIGNALING_UP	-3.2940	0	0.0000
PICCALUGA_ANGIOIMMUNOBLASTIC_LYMPHOMA_DN	-3.2504	0	0.0000
ZWANG_CLASS_3_TRANSIENTLY_INDUCED_BY_EGF	-2.8813	0	0.0000
NAGASHIMA_EGF_SIGNALING_UP	-2.8467	0	0.0000
UZONYI_RESPONSE_TO_LEUKOTRIENE_AND_THROMBIN	-2.7533	0	0.0000
DIRMEIER_LMP1_RESPONSE_EARLY	-2.7491	0	0.0000
PRAMOONJAGO_SOX4_TARGETS_UP	-2.7240	0	0.0000
PHONG_TNF_TARGETS_UP	-2.7155	0	0.0000
VILIMAS_NOTCH1_TARGETS_UP	-2.6482	0	0.0000
AMIT_EGF_RESPONSE_40_HELA	-2.6290	0	0.0000
JAATINEN_HEMATOPOIETIC_STEM_CELL_DN	-2.6257	0	0.0000
GALINDO_IMMUNE_RESPONSE_TO_ENTEROTOXIN	-2.6226	0	0.0000
LINDSTEDT_DENDRITIC_CELL_MATURATION_B	-2.5918	0	0.0000
SEKI_INFLAMMATORY_RESPONSE_LPS_UP	-2.5577	0	0.0001
DAUER_STAT3_TARGETS_UP	-2.5362	0	0.0001
MITSIADES_RESPONSE_TO_APLIDIN_UP	-2.5175	0	0.0001
GRAHAM_CML_QUIESCENT_VS_NORMAL_DIVIDING_UP	-2.4925	0	0.0001
BURTON_ADIPOGENESIS_1	-2.4872	0	0.0001
NOJIMA_SFRP2_TARGETS_UP	-2.4775	0	0.0001
ZHOU_INFLAMMATORY_RESPONSE_LIVE_UP	-2.4651	0	0.0001

Abbreviations: NES, normalized enrichment score; NOM, nominal; FDR, false discovery rate.

*Table shows the top 20 positive and negative gene-sets based on the NES.

Table S8. Genes and primers for qRT-PCR validation of GSEA

Gene	Forward primer	Reverse primer
POLD3	TGTCTGTCACGGAACCAAAG	CTACTCGCTTCCCCCTTTTT
POLD2	CTTGGAGATCCTGGAGTGGA	AAGTAGACATGCGGGCACTC
POLA2	GCAGCGAACTCAAGGAACAT	TTCAGCTTCCCGTTGCTATC
RPA2	TCTGCCACTTTGGTTGATGA	GTTGGTTGGAGCCTTCTCTG
RPA1	AGTGGACCATTTGTGCTCGT	GCTGTAGCTCGGATTTCACC
GAPDH	CGCTCTCTGCTCCTCCTGTT	GTTGACTCCGACCTTCACCTTCC

Table S9. Differentially expressed genes in FANC core and ID2 mutant vs non-mutant normal karyotype AML cases*

Probe ID	Symbol	logFC	P.Value	Gene description
ILMN_1732799	CD34	2.185403	0.0184	Homo sapiens CD34 molecule
ILMN_1789166	SHD	2.051213	0.0049	Homo sapiens src homology 2 domain containing transforming protein D
ILMN_2094875	ABCB1	1.940798	0.0004	Homo sapiens ATP binding cassette subfamily B member 1
ILMN_2341229	CD34	1.883016	0.0235	Homo sapiens CD34 molecule
ILMN_1786720	PROM1	1.656472	0.0303	Homo sapiens prominin 1
ILMN_3247023	FLJ22536	1.656119	0.0009	Homo sapiens cancer susceptibility 15 (non-protein coding)
ILMN_1812070	ABCB1	1.655002	0.0009	Homo sapiens ATP binding cassette subfamily B member 1
ILMN_1694817	TRH	1.632	0.0278	Homo sapiens thyrotropin releasing hormone
ILMN_1677723	ANGPT1	1.596347	0.0091	Homo sapiens angiopoietin 1
ILMN_2195462	C1QTNF4	1.517277	0.0247	Homo sapiens C1q and TNF related 4
ILMN_1808122	LOC652377	1.461842	0.0189	N/A
ILMN_2086890	ANGPT1	1.440354	0.0093	Homo sapiens angiopoietin 1
ILMN_2111229	BZRAP1	1.39009	0.0043	Homo sapiens TSPO associated protein 1
ILMN_1756439	SCRN1	1.355693	0.0463	Homo sapiens secernin 1
ILMN_1751868	TCTEX1D1	1.328454	0.0159	Homo sapiens Tctex1 domain containing 1
ILMN_1659024	TMCC2	1.190715	0.0499	Homo sapiens transmembrane and coiled-coil domain family 2
ILMN_2215824	ANKRD20A1	1.189385	0.0055	Homo sapiens ankyrin repeat domain 20 family member A1
ILMN_1681601	SUCNR1	1.152819	0.0328	Homo sapiens succinate receptor 1
ILMN_2359287	ITGA6	1.13852	0.0018	Homo sapiens integrin subunit alpha 6
ILMN_1670452	ANKRD20A1	1.135451	0.0025	Homo sapiens ankyrin repeat domain 20 family member A1
ILMN_1688231	TREM1	-1.21439	0.0202	Homo sapiens triggering receptor expressed on myeloid cells 1

ILMN_2064725	METTL7B	-1.22858	0.0368	Homo sapiens methyltransferase like 7B
ILMN_1784300	TUBA4A	-1.23926	0.0054	Homo sapiens tubulin alpha 4a
ILMN_1654331	HOXB4	-1.25778	0.0012	Homo sapiens momeobox B4
ILMN_1806165	HSPA6	-1.28035	0.0443	Homo sapiens heat shock protein family A (Hsp70) member 6
ILMN_2169490	TDRD9	-1.30075	0.0040	Homo sapiens tudor domain containing 9
ILMN_1778321	SLC2A6	-1.3104	0.0076	Homo sapiens solute carrier family 2 member 6
ILMN_1683798	LOC404266	-1.31351	0.0490	N/A
ILMN_1674908	HOXB5	-1.35775	0.0481	Homo sapiens homeobox B5
ILMN_1652379	SUCLG2	-1.3896	0.0053	Homo sapiens beta subunit of succinyl-CoA synthetase
ILMN_2089329	SPRY2	-1.41628	0.0095	Homo sapiens sprouty RTK signalling antagonist 2
ILMN_2150851	SERPINB2	-1.42145	0.0424	Homo sapiens serpin family B member 2
ILMN_1728106	TNF	-1.51608	0.0097	Homo sapiens proinflammatory cytokine tumour necrosis factor
ILMN_1651826	BASP1	-1.54557	0.0191	Homo sapiens brain abundant membrane attached signal protein 1
ILMN_2150856	SERPINB2	-1.56445	0.0337	Homo sapiens serpin family B member 2
ILMN_1782352	VENTX	-1.61146	0.0165	Homo sapiens VENT homeobox
ILMN_1763455	VSTM1	-1.76002	0.0032	Homo sapiens V-set and transmembrane domain containing 1
ILMN_2066060	HLA-DRB6	-1.84161	0.0170	Homo sapiens major histocompatibility complex, class II, DR beta 6 (pseudogene)
ILMN_1810274	HOXB2	-1.95995	0.0007	Homo sapiens nuclear protein homeobox B2
ILMN_2169801	TPSAB1	-2.30224	0.0065	Homo sapiens tryptase alpha/beta 1

*Differentially-expressed genes were identified from comparison of gene expression in patient samples with (n=6) or without (n=22) mutations in FANC core and ID2 complex genes. Table shows genes with an expression change of ≥ 1.5-fold at an adjusted P value of <0.05.

Table S10. Gene set enrichment analysis of differential gene expression between FANC core and ID2 mutant vs non-mutant normal karyotype AML cases*

Gene Set	NES	NOM p-val	FDR q-val
JAATINEN_HEMATOPOIETIC_STEM_CELL_UP	2.6361	0.0000	0.0000
VALK_AML_CLUSTER_8	2.3297	0.0000	0.0008
LUI_THYROID_CANCER_CLUSTER_3	2.2804	0.0000	0.0009
HOLLEMAN_ASPARAGINASE_RESISTANCE_B_ALL_UP	2.1214	0.0000	0.0097
VERHAAK_AML_WITH_NPM1_MUTATED_DN	2.0238	0.0000	0.0305
REACTOME_FORMATION_OF_THE_TERNARY_COMPLEX_AND_SUBSEQUENTLY_THE_43S_COMPLEX	1.8492	0.0000	0.1734
ZHAN_MULTIPLE_MYELOMA_HP_UP	1.8152	0.0022	0.2081
VANHARANTA_UTERINE_FIBROID_UP	1.8129	0.0045	0.1860
NIKOLSKY_BREAST_CANCER_11Q12_Q14_AMPLICON	1.8046	0.0000	0.1810
REACTOME_PEPTIDE_CHAIN_ELONGATION	1.7869	0.0000	0.1922
REACTOME_EXTENSION_OF_TELOMERES	1.7526	0.0045	0.2384
PLASARI_TGFB1_SIGNALING_VIA_NFIC_10HR_UP	1.7442	0.0107	0.2361
REACTOME_PURINE_METABOLISM	1.7403	0.0023	0.2254
RHEIN_ALL_GLUCOCORTICOID_THERAPY_DN	1.7345	0.0000	0.2207
LUI_TARGETS_OF_PAX8_PPARG_FUSION	1.7246	0.0113	0.2256
BILANGES_SERUM_AND_RAPAMYCIN_SENSITIVE_GENES	1.7090	0.0024	0.2413
REACTOME_LAGGING_STRAND_SYNTHESIS	1.6983	0.0130	0.2493
CHNG_MULTIPLE_MYELOMA_HYPERPLOID_UP	1.6970	0.0025	0.2378
NAGASHIMA_NRG1_SIGNALING_UP	-2.8448	0.0000	0.0000
VERHAAK_AML_WITH_NPM1_MUTATED_UP	-2.6047	0.0000	0.0000
DIRMEIER_LMP1_RESPONSE_EARLY	-2.5243	0.0000	0.0000
PHONG_TNF_TARGETS_UP	-2.5211	0.0000	0.0000
GALINDO_IMMUNE_RESPONSE_TO_ENTEROTOXIN	-2.4581	0.0000	0.0000
PICCALUGA_ANGIOIMMUNOBLASTIC_LYMPHOMA_DN	-2.4303	0.0000	0.0000
NIKOLSKY_BREAST_CANCER_8Q23_Q24_AMPLICON	-2.3826	0.0000	0.0000
ZHANG_RESPONSE_TO_IKK_INHIBITOR_AND_TNF_UP	-2.3702	0.0000	0.0000
NAGASHIMA_EGF_SIGNALING_UP	-2.3502	0.0000	0.0000
KEGG_VIRAL_MYOCARDITIS	-2.3210	0.0000	0.0000
KIM_WT1_TARGETS_UP	-2.3049	0.0000	0.0000
UZONYI_RESPONSE_TO_LEUKOTRIENE_AND_THROMBIN	-2.2939	0.0000	0.0000
BASSO_CD40_SIGNALING_UP	-2.2820	0.0000	0.0000
SEKI_INFLAMMATORY_RESPONSE_LPS_UP	-2.2701	0.0000	0.0000
BILD_HRAS_ONCOGENIC_SIGNATURE	-2.2500	0.0000	0.0003
AMIT_SERUM_RESPONSE_120_MCF10A	-2.2487	0.0000	0.0003
GESERICK_TERT_TARGETS_DN	-2.2451	0.0000	0.0004
BURTON_ADIPOGENESIS_PEAK_AT_2HR	-2.2303	0.0000	0.0005
KEGG_LEISHMANIA_INFECTION	-2.2241	0.0000	0.0005
GAZDA_DIAMOND_BLACKFAN_ANEMIA_PROGENITOR_UP	-2.2193	0.0000	0.0007

Abbreviations: NES, normalized enrichment score; NOM, nominal; FDR, false discovery rate.

*Table shows the top 19 positive and 20 negative gene-sets based on the NES.

4. SUPPLEMENTARY REFERENCES

1. Duncan EL, Danoy P, Kemp JP, Leo PJ, McCloskey E, Nicholson GC, et al. Genome-Wide Association Study Using Extreme Truncate Selection Identifies Novel Genes Affecting Bone Mineral Density and Fracture Risk. PLoS Genetics. 2011;7(4):e1001372.

2. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. Briefings in Bioinformatics. 2010;11(5):473-83.

3. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

4. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Human mutation. 2011;32(8):894-9.

5. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics. 2014;46(3):310-5.

6. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies. The American Journal of Human Genetics. 2012;91(2):224-37.

7. Network TCGAR. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. The New England journal of medicine. 2013;368(22):2059-74.

8. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics. 2003;34(3):267-73.

9. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50.

10. Neame PB, Soamboonsrup P, Browman GP, Meyer RM, Benger A, Wilson WE, et al. Classifying acute leukemia by immunophenotyping: a combined FAB-immunologic classification of AML. Blood. 1986;68(6):1355-62.

11. Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood. 2010;115(3):453-74.

12. Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood. 2010;116(3):354-65.

13. Gille JJ, Floor K, Kerkhoven L, Ameziane N, Joenje H, de Winter JP. Diagnosis of Fanconi Anemia: Mutation Analysis by Multiplex Ligation-Dependent Probe Amplification and PCR-Based Sanger Sequencing. Anemia. 2012;2012:603253.

14. Lo ten Foe JR, Barel MT, Thuss P, Digweed M, Arwert F, Joenje H. Sequence variations in the Fanconi anaemia gene, FAC: pathogenicity of 1806insA and R548X and recognition of D195V as a polymorphic variant. Hum Genet. 1996;98(5):522-3.

15. Verlander PC, Lin JD, Udono MU, Zhang Q, Gibson RA, Mathew CG, et al. Mutation analysis of the Fanconi anemia gene FACC. American journal of human genetics. 1994;54(4):595-601.

16. Kalb R, Neveling K, Hoehn H, Schneider H, Linka Y, Batish SD, et al. Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. American journal of human genetics. 2007;80(5):895-910.

17. Ali AM, Kirby M, Jansen M, Lach FP, Schulte J, Singh TR, et al. Identification and characterization of mutations in FANCL gene: a second case of Fanconi anemia belonging to FA-L complementation group. Human mutation. 2009;30(7):E761-70.

18. Hodson C, Purkiss A, Miles JA, Walden H. Structure of the human FANCL RING-Ube2T complex reveals determinants of cognate E3-E2 selection. Structure. 2014;22(2):337-44.

19. Kiiski JI, Pelttari LM, Khan S, Freysteinsdottir ES, Reynisdottir I, Hart SN, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(42):15172-7.

20. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. The New England journal of medicine. 2015;373(24):2336-46.

21. Garcia MJ, Fernandez V, Osorio A, Barroso A, Fernandez F, Urioste M, et al. Mutational analysis of FANCL, FANCM and the recently identified FANCI suggests that among the 13 known Fanconi Anemia genes, only FANCD1/BRCA2 plays a major role in high-risk breast cancer predisposition. Carcinogenesis. 2009;30(11):1898-902.

Chapter 5 Manuscript - BRCA1/2 mutations in childhood AML

5.1 Summary

The results presented in **Chapter 3 and 4** have been centred on adult AML (21>years of age at diagnosis). This chapter focuses on paediatric AML which occurs at much lower frequency, and often associated with genetic syndromes. For example, as discussed in **Chapter 1**, children with the recessive DNA repair disorder Fanconi anaemia are at greatly elevated risk of developing AML (estimated 800 fold increased risk over the normal population) (Rosenberg et al., 2008). Risk of AML is also elevated for children with a number of other recessive DNA repair disorders, such as ataxia telangiectasia (A-T), Nijmegen breakage syndrome (NBS), Seckel syndrome (SS) and Bloom's syndrome (BS). These patients have increased cancer predisposition, as well as developing cancer at an early age (Kutler, 2003, Stankovic et al., 1998, German, 1969, Tanaka et al., 2012, Kruger et al., 2007) as described in **Chapter 1**, thus emphasising the importance of DNA repair pathways in maintaining genomic integrity, particularly in haematopoietic stem cells (HSC) with regards to AML. Another genetic disorder that shows increased AML risk is Down syndrome (DS), caused by trisomy 21. AML in DS displays specific characteristics and is associated with acquired mutations in the transcription factor GATA1 (Ahmed et al., 2004). A percentage of DS patients can develop transient abnormal myelopoiesis (TAM) that is associated with increased risk of AML (Bhatnagar et al., 2016); however, other factors that lead to AML development in DS are not clearly defined.

While overall survival for childhood AML is higher than that for adult AML, there are treatment-related toxicity from long-term treatment with genotoxic compounds that need to be considered for childhood AML patients (Creutzig et al., 2012). Over recent years, childhood AML has been shown to be distinctly different to adult AML with regard to acquired chromosome rearrangements and mutations, and the role of epigenetic de-regulation. In childhood AML certain translocations occur at higher frequency than in adults. For example *MLL* gene rearrangements are common in the childhood AML setting and are associated with poor prognosis (Tarlock and Meshinchi, 2015).

Despite several recent reports describing the spectrum of somatic changes in childhood AML (Valerio et al., 2014, Tarlock and Meshinchi, 2015), there has not been an extensive analysis to identify germline variants that contribute to development of AML in children. Recently, a landmark study of a large paediatric multi-cancer cohort identified rare pathogenic, or probably pathogenic germline mutations across 565 cancer-related genes, suggesting that a percentage

of childhood cancers (8.5% as reported) may be associated with predisposition alleles (Zhang et al., 2015). While this report has highlighted the role of germline variants in determining childhood cancer risk, AML cases were extremely rare in this cohort. In this chapter, analysis of variants affecting the extended FA/BRCA-HRR network (as defined in **Chapter 3)**, which includes many genes associated with recessive DNA repair syndromes, was undertaken in a childhood AML cohort. The data and conclusions are presented in the attached manuscript in **Section 5.2.**

5.2 Manuscript - BRCA1/2 mutations in childhood AML

Statement of Authorship

Title of Paper	BRCA1/2 mutations in childhood AML
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Blood Advances: Stimulus Report

Principal Author

Name of Principal Author (Candidate)	Kyaw Ze Ya Maung		
Contribution to the Paper	Performed the basis of the research, analysed and interpreted the whole exome data and wrote the publication		
Overall percentage (%)	25%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Paul Leo		
Contribution to the Paper	25%. Performed and analysed the raw whole exome data and wrote the methodology of the sequencing experiment.		
Signature		Date	

Statement of Authorship

Title of Paper	BRCA1/2 mutations in childhood AML
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Blood Advances: Stimulus Report

Principal Author

Name of Principal Author (Candidate)	Kyaw Ze Ya Maung		
Contribution to the Paper	Performed the basis of the research, analysed and interpreted the whole exome data and wrote the publication		
Overall percentage (%)	25%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	30. 7. 2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Paul Leo		
Contribution to the Paper	25%. Performed and analysed the raw whole exome data and wrote the methodology of the sequencing experiment.		
Signature		Date	

Statement of Authorship

Title of Paper	BRCA1/2 mutations in childhood AML
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Blood Advances: Stimulus Report

Principal Author

Name of Principal Author (Candidate)	Kyaw Ze Ya Maung		
Contribution to the Paper	Performed the basis of the research, analysed and interpreted the whole exome data and wrote the publication		
Overall percentage (%)	25%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Paul Leo		
Contribution to the Paper	25%. Performed and analysed the raw whole exome data and wrote the methodology of the sequencing experiment.		
Signature		Date	25/7/17

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	20/7/17

Name of Co-Author	Debora Casolari		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	2%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Deepak Singal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	

Name of Co-Author	Amanda Smith		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	31/7/17

Name of Co-Author	James Gray		
Contribution to the Paper	2%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Deepak Singhal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	

Name of Co-Author	Amanda Smith		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	2%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	20 JULY 2017

Name of Co-Author	Deepak Singal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	

Name of Co-Author	Amanda Smith		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	2%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Deepak Singhal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	21/07/2017

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	

Name of Co-Author	Amanda Smith		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	2%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Deepak Singal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	27/7/17

Name of Co-Author	Amanda Smith		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	

Name of Co-Author	Mahmoud Bassal		
Contribution to the Paper	3%. Contributed to analysing the whole exome data and the development of the STRING network.		
Signature		Date	

Name of Co-Author	Debora Casolari		
Contribution to the Paper	2%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	James Gray		
Contribution to the Paper	2%. Principal supervisor of the student and edited and gave feedback and input on the writing of the manuscript		
Signature		Date	

Name of Co-Author	Deepak Singal		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Mhairi Marshall		
Contribution to the Paper	1%. Assisted in the whole exome sequencing.		
Signature		Date	

Name of Co-Author	Amanda Smith		
Contribution to the Paper	1% Prepared the paediatric AML samples used for this study.		
Signature		Date	20/7/2017

Name of Co-Author	Matt Brown		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	26.07.17

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	26.07.17.

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Matt Brown		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	20/7/17

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Matt Brown		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	20/7/17

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Matt Brown		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Emma Duncan		
Contribution to the Paper	1%. Provided the samples for the healthy control cohort.		
Signature		Date	

Name of Co-Author	Russell Saal		
Contribution to the Paper	1%, Provided details of clinical characteristics of the samples.		
Signature		Date	

Name of Co-Author	Paula Marlton		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	

Name of Co-Author	Devinder Gill		
Contribution to the Paper	1%. Provided clinical information and corrdinated the collection of the patient samples		
Signature		Date	25/7/17

Name of Co-Author	Anna Brown		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	24/7/17

Name of Co-Author	Andrew Moore		
Contribution to the Paper	5%. Provided clinical information and corrdinated the collection of the patient samples.		
Signature		Date	

Name of Co-Author	Thomas Gonda		
Contribution to the Paper	12%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Richard D'Andrea		
Contribution to the Paper	18%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Anna Brown		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Andrew Moore		
Contribution to the Paper	5%. Provided clinical information and corrdinated the collection of the patient samples.		
Signature		Date	20/07/2017

Name of Co-Author	Thomas Gonda		
Contribution to the Paper	12%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Richard D'Andrea		
Contribution to the Paper	18%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Anna Brown		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Andrew Moore		
Contribution to the Paper	5%. Provided clinical information and corrdinated the collection of the patient samples.		
Signature		Date	

Name of Co-Author	Thomas Gonda		
Contribution to the Paper	12%. Contributing to the editing and writing of the manuscript.		
Signature		Date	20th July 2017

Name of Co-Author	Richard D'Andrea		
Contribution to the Paper	18%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Anna Brown		
Contribution to the Paper	1%. Edited and gave feedback and input on the writing of the manuscript.		
Signature		Date	

Name of Co-Author	Andrew Moore		
Contribution to the Paper	5%. Provided clinical information and corrdinated the collection of the patient samples.		
Signature		Date	

Name of Co-Author	Thomas Gonda		
Contribution to the Paper	12%. Contributing to the editing and writing of the manuscript.		
Signature		Date	

Name of Co-Author	Richard D'Andrea		
Contribution to the Paper	18%. Contributing to the editing and writing of the manuscript.		
Signature		Date	31/7/17

BRCA1/2 mutations in childhood AML

Kyaw Ze Ya Maung[1,2,3,4*], Paul Leo[5*], Mahmoud Bassal[4,6,7], Debora A. Casolari[2,4,7], James X. Gray[3], Deepak Singhal[2,7], Mhairi Marshall[5], Emma L. Duncan[5,8,9], Matt Brown[5], Amanda M Smith[10], Russell Saal[5], Paula Marlton[11], Devinder Gill[11], Anna L. Brown[4,12], Andrew S. Moore[10,13,14], Thomas J Gonda[4,6,15] and Richard J D'Andrea[4,6,7]

*MKZY and PL contributed equally to this work

Affiliations:

[1]Basil Hetzel Institute for Translational Health Research, The Queen Elizabeth Hospital, Woodville, South Australia, Australia

[2]School of Medicine, University of Adelaide, Adelaide South Australia

[3]Department of Haematology and Oncology, the Queen Elizabeth Hospital, Woodville, South Australia, Australia

[4]Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, South Australia

[5]Institute of Health and Biomedical Innovation, Queensland University of Technology and University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia

[6]School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia, Australia

[7]Departments of Haematology, SA Pathology and Royal Adelaide Hospital, Adelaide, South Australia

[8]Department of Endocrinology, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia

[9]School of Medicine, University of Queensland, Brisbane, Queensland, Australia

[10]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia

[11]Clinical Haematology, Princess Alexandra Hospital Brisbane, Queensland, Australia

[12]Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia

[13]Oncology Services Group, Children's Health Queensland Hospital and Health Service, Brisbane, Queensland, Australia

[14]UQ Child Health Research Centre, University of Queensland, Brisbane, Queensland, Australia

[15]School of Pharmacy, University of Queensland, Brisbane, Queensland, Australia

Correspondence:

Professor Richard D'Andrea: Centre for Cancer Biology, University of South Australia, GPO Box 2471, Adelaide 5001, Australia. Email: richard.dandrea@unisa.edu.au

Keywords: childhood leukaemia, DNA repair, acute myeloid leukaemia, cancer risk, BRCA2

Running Title: BRCA1/2 mutations in childhood AML

Word count: 1196words

Introduction

Leukaemia is the most common childhood cancer accounting for approximately 30% of all paediatric cancers,[1] with Acute myeloid leukaemia (AML) accounting for approximately 25% of these cases.[2] The overall survival of paediatric AML (~70%) is significantly better than adult AML (~30-60%). It has become increasingly recognised that there are distinct genetic and epigenetic differences between paediatric and adult AML.[3] The spectrum of recurrent somatic AML mutations varies between paediatric and adult AMLs,[3,4] and approximately 50% of adult AML cases are karyotypically normal compared to only 20% for paediatric cases.[5] Consistent with this, recessive DNA repair disorders that are prevalent in children, such as Ataxia Telangiectasia, Nijmegen breakage syndrome, Blooms Syndrome and Fanconi Anaemia, are associated with increased risk of AML.[6-8] In addition, the incidence of AML is markedly increased in children with Down Syndrome (DS),[5] where it can be preceded by transient abnormal myelopoiesis (TAM) that occurs in the foetus or a few days after birth, and resolves spontaneously within 3 months.[9] The early onset and increased number of abnormal/complex karyotype cases suggest that underlying increased genomic instability may lie at the root of paediatric AML, more so than adult AML. Therefore we undertook a survey of 58 genes associated with DNA repair.

Methodology

We undertook whole exome sequencing (WES) of 23 *de novo* paediatric AML collected with consent at diagnosis, and with Institutional human research ethics committee approval (**Supplementary Methods**). We performed a focused analysis of rare somatic and germline mutations in genes involved in the Fanconi Anaemia and BRCA driven homologous recombination repair pathway (FA/BRCA-HRR) (**Supplementary Table S1 & Figure S1**). WES and variant calling was performed as described in the **Supplementary Methods**. Variants were filtered against common SNP databases (dbSNP, 1000 genome project and ESP) for minor allele frequency of < 0.001, and a pathogenicity prediction filter (Combined Annotation Dependent Depletion; CADD v1.2) was next applied to the dataset (CADD> 10). WES data was also available for a healthy female Caucasian control cohort, which was sequenced and analysed using the same pipeline (**Supplementary methods**).

Results & Discussion

Mutations were identified in 21 out of 58 genes in the FA/BRCA-HRR network (**Supplementary Figure. S1**) with 13 out of 23 patients (56.5%) having mutations (**Table 1**). Ten patients (43.5%) carried one or more potentially pathogenic mutations in breast cancer associated genes (*BRCA1*, *BRCA2* and *BRIP1*) and RAD51 paralogue genes (*RAD51B* and *RAD51D*) (**Figure 1).** The most striking observation from this study was the frequency of samples carrying damaging mutations in the breast cancer associated genes *BRCA1* and *BRCA2* (26.1%, n=6). These BRCA-mutant patients were diagnosed earlier than the rest of the cohort, however this did not reach statistical significance (p=0.07). While the size of this cohort is small, we observed significantly increased frequency of mutations affecting *BRCA1/2* in paediatric AML compared to the healthy control cohort (*P*=0.0047), and compared to non-Finnish Europeans in ExAC (*P*<0.0001). The BRCA2 mutation R2034C was identified in 2 unrelated AML patients, and was absent in our control cohort (**Table 1**). This mutation is classified as benign in Clinvar, and has a frequency in ExAC of 0.5%, however it has been reported to segregate with disease in a Dutch breast cancer family (Leiden University, Family IDRUL023). Mutations affecting this residue have also been reported in Rhabdomyosarcoma (R2034C) and lung cancer (R2034H). The BRCA1 mutation Y856H, also previously reported in a breast cancer cohort,[10] was identified in a single sample and was absent in the control cohort, and ExAC. Similarly, the BRCA2 mutations S1733F and R2108C were absent from our control cohort. The BRCA2 mutation R2108C has previously been shown to decrease HRR activity.[11] For the majority of *BRCA1/2* variants VAF was above 45% (BRCA2-S1733F is 36.6%; BRCA1-Y856H is 28.9%) consistent with likely germline origin.

We identified *BRCA1/2* mutations at a significantly increased frequency in the samples with tri-21 (*P*=0.045; **Figure 1**). Of the 6 DS AMLs, 3 carried *BRCA1* or *BRCA2* mutations. While there are clear limitations to studies with small paediatric rare-cancer cohorts, these findings raise the possibility that deleterious *BRCA1/2* mutation increases risk of AML in DS. Longitudinal studies are needed in cohorts of DS patients to further test association of *BRCA1/2* mutation with development of TAM, and AML.

BRCA1/2, BRIP1, PALB2 and BARD1 are central to the repair of double-stranded breaks and are required for the recruitment of the RAD51 recombinase and its paralogues (RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3)[13,14]. Four damaging mutations were identified in the *RAD51* family genes (3 in *RAD51B* and 1 in *RAD51D*), with the RAD51B mutation V207L identified in 2 of the AML samples, and absent in the control cohort. One of these cases is a likely acquired mutation (VAF 10%, **Table 1**). Although the size of this cohort precludes exclusivity analysis, we observed that the samples that carried these RAD51-paralogue

mutations did not carry mutations in any of the *BRCA* and BRCA-associated genes (**Figure 1**), consistent with the cooperative role of the BRCA and RAD51 proteins in HRR.

The results from this study raise an intriguing question regarding BRCA mutations in paediatric AML. Two independent studies[15,16] have reported contradictory results relating to the association of familial paediatric cancer with germline *BRCA1/2* mutations. However, most recently Zhang et al reported increased pathogenic mutations in *BRCA2* in a cohort of 1120 paediatric cancers, compared to non-cancer controls.[12] It is important to note that in all 3 studies, the number of AML cases was limited.

Based on our findings, together with functional studies demonstrating haploinsufficient phenotypes in BRCA1/2-mutant cells,[17,18,20] and the dominant negative effects associated with missense *BRCA1/2* mutations,[19] we hypothesise that heterozygous germline deleterious *BRCA1/2* mutation may result in ineffective HRR in hematopoietic stem/progenitor cells, increased replicative stress and genomic instability that favours malignant transformation. Such an effect is likely to be relatively subtle, but we suggest it is sufficient to increase risk of early AML development, particularly in the context of trisomy 21. Moreover, it is possible that detrimental effects of heterozygous *BRCA2* mutations may be exacerbated by exposure to aldehydes in utero, or during childhood. [21] At this point in time, it is premature to propose peripheral blood BRCA1/2 screening in DS, or genetic counselling in families of childhood AML patients with BRCA1/2 mutations. However, these results highlight the need for further studies with multi-centre paediatric AML cohorts to confirm our findings.

The frequency of *BRCA1/2* and *RAD51* paralogue mutations in this cohort may provide a potential treatment option for this subset of AMLs. Given that selected heterozygous *BRCA2* mutations can confer a severe haploinsufficient phenotype through dominant negative activity,[19] and that HRR activity can be reduced therapeutically (i.e. HSP90 inhibitors),[22] combination treatments with PARP1 and/or HSP90 inhibitors with anthracyclines such as daunorubicin may be appropriate.[23] Such therapeutic approach is also supported by the report that a breast cancer signature detecting deficiency in the BRCA pathway predicts sensitivity to anthracycline/cyclophosphamide–based chemotherapy.[24] Finally, there is a strong case for analysing paediatric AML genomic data to detect of genomic signatures associated with compromised HRR activity. Such studies will be important to determine whether paediatric AML is more widely associated with ineffective HRR, and may provide an approach to identifying childhood AML patients suitable for alternative tailored therapy with the agents discussed above.

Conflict of Interest

All authors have revised and approved the manuscript and declare no competing financial interests.

Table 1. Mutations in the FA/BRCA-HRR network genes (n=58) identified from the WES of diagnostic paediatric AML (n=23)

AML ID	Age	Gender	Gene	Amino acid change	%of Alt Reads	CADD Score	[1]Frequency in ExAC	[2]Frequency in healthy controls	COSMIC ID	Type of cancer (No. of samples)
WES-252	15	M	ATM	p.V410A	11.3	23.4	0.003	0.003	COSM5945737 & COSM21825	Glioma (2), Lymphoma (1), Non-Hodgkin Lymphoma (1), CLL (1), Melanoma (2)
WES-257	4	F	BRCA2	p.R2034C	52.6	22.2	0.005	0	COSM4987322 & COSM696739	Rhabdomyosarcoma (1), ^Lung [3]SCC (1)
WES-259	8	M	RAD51B	p.V207L	57.5	20.7	0.002	0	COSM3815166	Breast cancer (1)
WES-260**	2	M	ATM	p.H1380Y	52.2	13.4	<0.001	0	COSM24627	B-cell Lymphoma (1)
			BRCA2	p.V2739I	51.1	11.4	0	0	-	-
			BRIP1	p.86_90del	40.0	22.7	0	0	-	-
			FANCE	p.R343Q	53.0	15.1	0	0	-	-
			MLH3	p.G1163D	41.4	22.7	<0.001	0	-	-
WES-261	2	M	RAD51D	-	48.3	24.5	0	0	-	-
WES-262	1	F	ATM	p.A1931A	51.7	19.0		0.009	-	-
			BRCA2	p.R2108C	51.5	18.8	<0.001	0	-	-
			BRIP1	p.L340F	53.6	22.9	0	0	-	-
			RAD50	p.S128N	49.5	25.8	<0.001	0	-	-
			RAD9A	p.H163R	100.0	22.2	<0.001	0	-	-
WES-265	3	F	RAD51B	p.V207L	10.0	20.7	0.002	0	COSM3815166	Breast cancer (1)
WES-266	10	F	ATR	p.R336W	49.1	22.6	0.00E+00	0	COSM205540	Renal cancer
			RAD51B	p.K243R	52.6	27.1	0.011	0.012	-	-

WES-267**	2	M	C17orf70	p.A694P	12.5	12.9	0.003	0	-	-
			EME1	p.A479P	48.1	21.4	<0.001	0	-	-
WES-268	2	M	BLM	p.E270K	54.1	14.1	0	0	-	-
			C17orf70	p.A694P	12.5	12.9	0.003	0	-	-
			FANCD2	p.E742D	49.4	10.55	0	0	-	-
			TOP3A	p.A461T	42.9	26.0	0.002	0	-	-
			TOP3A	p.D459N	45.2	23.4	0.005	0.009	-	-
WES-269**	1	F	BRCA2	p.R2034C	45.8	22.2	0.005	0	COSM4987322 & COSM696739	Rhabdomyosarcoma (1), ^Lung [3]SCC (1)
			FANCA	p.V384F	42.0	24.3	<0.001	0	-	-
WES-272*	1	F	ATM	p.A1931A	33.3	19.0		0.009	-	-
			ATRIP	p.R675Q	54.5	22.2	0.011	0.015	-	-
			BRCA2	p.S1733F	36.6	13.4	<0.001	0	-	-
			ERCC4	p.Q849E	55.0	24.1	0	0	-	-
			NBN	p.Q298R	55.1	21.8	0	0	-	-
WES-274**	2	M	BRCA1	p.Y856H	28.9	16.0	0	0	-	-

Footnote: [1]Frequncy of mutation in the non-Finnish Europeans (n=33370) of the Exome Aggregate Consortium (ExAC) database. [2]Frequency of mutations in the all-female healthy control cohort (n=329).[3]Squamous cell carcinoma; ^Samples with different mutation at the same amino acid as identified in our study; *Indicates patients with Trisomy 21; **Indicates patients with Down syndrome; BRCA and RAD51 paralogue variants discussed in detail are in **Bold** font.

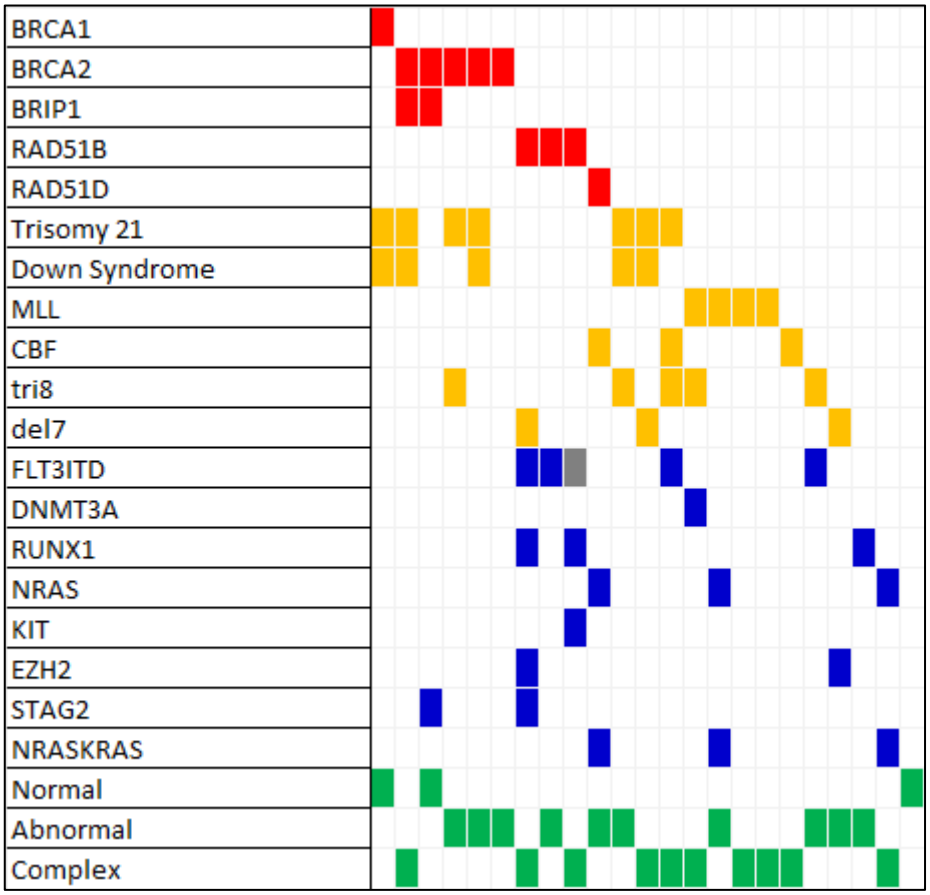


Figure 1. Mutational status of BRCA, BRCA-associated genes and RAD 51 paralogues. Mutations are shown in RED in this figure, along with the cytogenetic abnormalities (YELLOW), recurrent AML mutations (BLUE) and karyotype. Recurrent AML mutations were determined from WES analysis. GREY indicates unavailable data.

Bibliography

1. Pui CH, Carroll WL, Meshinchi S, Arceci RJ. Biology, risk stratification, and therapy of pediatric acute leukemias: an update. J Clin Oncol. 2011;29(5):551-565.

2. Bonaventure A, Harewood R, Stiller CA, et al. Worldwide comparison of survival from childhood leukaemia for 1995-2009, by subtype, age, and sex (CONCORD-2): a population-based study of individual data for 89 828 children from 198 registries in 53 countries. Lancet Haematol. 2017;4(5):e202-e217.

3. Tarlock K, Meshinchi S. Pediatric acute myeloid leukemia: biology and therapeutic implications of genomic variants. Pediatr Clin North Am. 2015;62(1):75-93.

4. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059-2074.

5. Creutzig U, van den Heuvel-Eibrink MM, Gibson B, et al. Diagnosis and management of acute myeloid leukemia in children and adolescents: recommendations from an international expert panel. Blood. 2012;120(16):3187-3205.

6. Stankovic T, Kidd AM, Sutcliffe A, et al. ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of mutant ATM and the risk of leukemia, lymphoma, and breast cancer. Am J Hum Genet. 1998;62(2):334-345.

7. German J. Bloom's syndrome. I. Genetical and clinical observations in the first twenty-seven patients. Am J Hum Genet. 1969;21(2):196-227.

8. Rosenberg PS, Alter BP, Ebell W. Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. Haematologica. 2008;93(4):511-517.

9. Saida S. Predispositions to Leukemia in Down Syndrome and Other Hereditary Disorders. Curr Treat Options Oncol. 2017;18(7):41.

10. Emi M, Matsushima M, Katagiri T, et al. Multiplex mutation screening of the BRCA1 gene in 1000 Japanese breast cancers. Jpn J Cancer Res. 1998;89(1):12-16.

11. Balia C, Galli A, Caligo MA. Effect of the overexpression of BRCA2 unclassified missense variants on spontaneous homologous recombination in human cells. Breast Cancer Res Treat. 2011;129(3):1001-1009.

12. Zhang J, Walsh MF, Wu G, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. N Engl J Med. 2015;373(24):2336-2346.

13. Cousineau I, Abaji C, Belmaaza A. BRCA1 regulates RAD51 function in response to DNA damage and suppresses spontaneous sister chromatid replication slippage: implications for sister chromatid cohesion, genome stability, and carcinogenesis. Cancer Res. 2005;65(24):11384-11391.

14. Godin SK, Sullivan MR, Bernstein KA. Novel insights into RAD51 activity and regulation during homologous recombination and DNA replication. Biochem Cell Biol. 2016:1-12.

15. Magnusson S, Borg A, Kristoffersson U, Nilbert M, Wiebe T, Olsson H. Higher occurrence of childhood cancer in families with germline mutations in BRCA2, MMR and CDKN2A genes. Fam Cancer. 2008;7(4):331-337.

16. Brooks GA, Stopfer JE, Erlichman J, Davidson R, Nathanson KL, Domchek SM. Childhood cancer in families with and without BRCA1 or BRCA2 mutations ascertained at a high-risk breast cancer clinic. Cancer Biol Ther. 2006;5(9):1098-1102.

17. Pathania S, Bade S, Le Guillou M, et al. BRCA1 haploinsufficiency for replication stress suppression in primary cells. Nat Commun. 2014;5:5496.

18. Sedic M, Skibinski A, Brown N, et al. Haploinsufficiency for BRCA1 leads to cell-type-specific genomic instability and premature senescence. Nat Commun. 2015;6:7505.

19. Vaclova T, Woods NT, Megias D, et al. Germline missense pathogenic variants in the BRCA1 BRCT domain, p.Gly1706Glu and p.Ala1708Glu, increase cellular sensitivity to PARP inhibitor Olaparib by a dominant negative effect. Hum Mol Genet. 2016.

20. Parmar K, D'Andrea AD. Aldehydes Pose a Threat to BRCA2 Mutation Carriers. Cell. 2017;169(6):979-981.

21. Tan SLW, Chadha S, Liu Y, et al. A Class of Environmental and Endogenous Toxins Induces BRCA2 Haploinsufficiency and Genome Instability. Cell. 2017;169(6):1105-1118 e1115.

22. Karras GI, Yi S, Sahni N, et al. HSP90 Shapes the Consequences of Human Genetic Variation. Cell. 2017;168(5):856-866 e812.

23. Spencer DM, Bilardi RA, Koch TH, et al. DNA repair in response to anthracycline-DNA adducts: a role for both homologous recombination and nucleotide excision repair. Mutat Res. 2008;638(1-2):110-121.

24. Mulligan JM, Hill LA, Deharo S, et al. Identification and validation of an anthracycline/cyclophosphamide-based chemotherapy response assay in breast cancer. J Natl Cancer Inst. 2014;106(1):djt335.

SUPPLEMENTARY INFORMATION

Rare damaging BRCA1/2 mutations in childhood AML

Maung *et al*

TABLE OF CONTENTS:

1. SUPPLEMENTARY MATERIALS AND METHODS

Whole Exome Sequencing of AML samples. The use of the samples for this research study was approved by the PAH, the RAH, the University of Adelaide, the University of South Australia and the University of Queensland Human Research Ethics Committees (HREC/05/QRCH/77, HREC/04/QPAH/172 & HREC/13/RAH/612). Genomic DNA was extracted from patient AML samples (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Genomic DNA was sonicated and DNA sequencing libraries were constructed using a preparation kit for paired-end sequencing (Illumina) as per the manufacturer's protocol. Liquid-phase hybridization for exome capture was performed using Illumina Nextera Rapid (FC-140-1003, Illumina) (n=23). Efficiency of sequence capture was assessed using quantitative real-time PCR with standard control primers as recommended by the manufacturer and quality of each DNA sequencing library was assessed using a 2100 Bioanalyzer and DNA 1000 chip kits (Agilent Technologies, Santa Clara, CA, USA). DNA concentration was standardized to 10 nM for sequencing. Massive parallel sequencing was performed using the Illumina HiSeq 2000 configured for paired-end reads. The 23 AML exomes had a mean coverage of 47.6x (22-144x). Exome sequencing and variant calling was performed at the UQ Centre for Clinical Genomics.

Control Cohort. For this study, we used WES data generated from a normal healthy cohort (n=329) from the Anglo-Australian Osteoporosis Genetics Consortium for access to control WES data.[1] This control cohort is all female of Caucasian ethnic origin with mean age of 67yrs (46-86yrs). These samples have been sequenced on the Illumina TrueSeq Enrichment kit v2.0 (Illumina, San Diego, CA, USA) and analysed using the same methods as described above. These controls had a mean coverage of (27.5x). In addition, a cohort of 49 germline samples, sequenced using Illumina Nextera Rapid exome capture, were used to control for potential bias between the Illumina Nextera Rapid (FC-140-1003, Illumina) and the Illumina TrueSeq Enrichment kit v2.0 (Illumina, San Diego, CA, USA).

Base calling and variant filtering. Initial base calling was performed using the CASAVA 1.7 data analysis pipeline software (Illumina). Sequence data were aligned to the current build hg19 of the human genome using the Novoalign alignment tool [V2.07.09 1] (Ref.[2]) Sequence alignment files were converted using SAMtools [v0.1.14] (Ref.[3]) and Picard tools (v1.42). SNPs and indels for all AML samples were simultaneously called using the best practice protocols described for Genome Analysis Toolkit (GATK v3.5-2 for Australian AML samples and GATK v3.2-2 for TCGA replication samples). This genotyping included sample level Indel

realignment and variant quality score recalibration (VSQ). Genotypes were annotated using ANNOVAR[4] using Refseq, Ensembl and UCSC transcript definitions and the Ensembl Variant Effect Predictor (VEP). Sequencing data were analysed and filtered using custom scripts employing R and Bioconductor. Good quality SNPs and indels (VSQ: FILTER=PASS) were retained.

We defined rare variants as those with a population frequency/minor allele frequency (MAF) <0.001 reported in dbSNP147, 1000 genomes (April 2012 version, any ethnicity), and the 6500 NHLBI-ESP project (any ethnicity). We include monomorphic variants and those reported in dbSNP without a population frequency in this category. Variants were defined as real calls if ANNOVAR (using any of the 3 transcript definitions), or the VEP, predicted a mutation to be non-synonymous, splicing (including splice donor and splice acceptor variants defined by VEP), stop-gain, stop-loss, a frameshift or non-frameshift substitution or deletion, an initiator codon variant, a stop retained variant or an incomplete terminal codon variant. We additionally excluded loci where coverage was low and/or the missing rate was high across the cohort. These loci were identified by first filtering genotypes with less than 7 reads supporting heterozygous calls, and 2 for homozygous calls, before calculating the missing rate. If the missing rate exceeded 80%, all genotypes at that loci were excluded. This approach removes low coverage regions where genotypes can be unreliable and discrimination between homozygous and heterozygous calls is poor. We are aware that sequencing artefacts can occur even with this filtering, often appearing as novel or rare mutations (in a population sense) occurring at high frequency in a genotyped cohort. To help identify these, we applied a Hardy-Weinberg p-value filter of 10^{-6} to the healthy controls and additionally excluded loci where the allele frequency in the control genotypes greatly exceeded that of the population frequency filter. We defined this as 6 standard deviations away from the population threshold of 0.001. Under the binominal approximation, this is defined by ("control allele count"- n*p)/sqrt(n*p*(1-p))<=6. For example, if n=400 and *P*=0.001, this restricts our analysis to loci with 3 or less alleles reported in the control cohort. We additionally excluded loci where the mean genotype quality (GQ) score was <50 in samples where an alternative allele was predicted. For further pathogenicity filtering we used the Combined Annotation Dependent Depletion algorithm (CADD score >10).[5] To control for potential biases generated from the use of different exome capture methods, we compared both genotype counts between control groups and modelled the background allele frequency between the different cohorts. If the genotype counts differed between Illumina and Nextera control cohorts under the binomial approximation with *P*<0.001, we excluded that loci. We modelled the background allele frequency for all loci using samples with homozygous reference genotypes, essentially using the alternative allele frequency to

estimate a position-dependent background sequencing error rate within each cohort. Differences in that background error rate were used to identify loci where potential batch effects might occur. We chose to exclude loci where the background error rates differed by 6 standard deviation in loci where there was a least 100 reads from the cohorts. One loci in the FANC gene group failed this filtering in the analysis for the AML cohort.

Generating the Extended FA/BRCA-HRR network

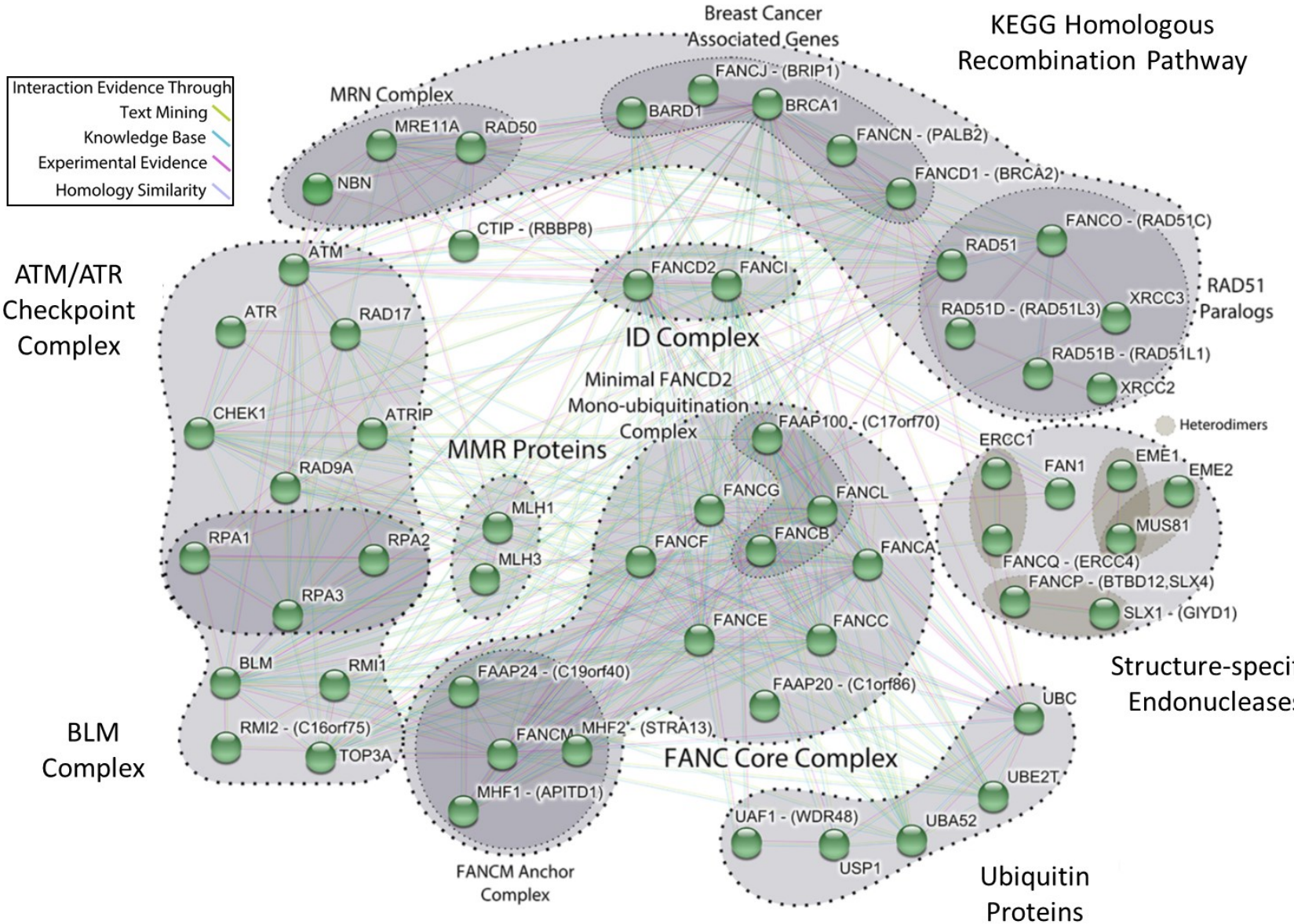
Given the extremely elevated risk of AML associated with Fanconi Anemia we investigated the involvement of an extended network of genes using the canonical FA DNA repair pathway as the seed. For this, the protein/gene interaction predictor STRINGdb v9.05 was used to construct a network from 16 FANC genes.[6] This extended network is referred to as FA/BRCA-HRR. While 16 FANC genes were used as the basis of this analysis, the total number of FANC genes has further increased to twenty-one with the addition of *FANCR* (*RAD51*), *FANCS* (*BRCA1*), *FANCT* (*UBE2T*), *FANCU* (*XRCC2*) and *FANCV* (*REV7*). All of these genes are included in our genomic analysis, except *FANCV*.

A two-step approach using STRINGdb was used. The first step used the ability of STRINGdb to predict interacting partners of the individual proteins. The second step utilised its ability to predict interactions within a list of query proteins. Thus, each of the 16 FANC proteins (query) were individually searched in STRINGdb. The top 50 proteins that interacted with each of the FANC proteins with confidence level of 0.9 (90%) based on experimental evidence, knowledge based and text mining evidence were compiled into a single list. This list was then used as the query file that was input into STRINGdb providing the interactions of high confidence within the query list (**Supplementary Figure S1**). This output from STRINGdb was then further manually clustered to show the functional subgroups with high confidence interactions detailed in **Supplementary Table S1 and Figure S1**. Thus, in this FA/BRCA-HRR network, all the proteins have direct relationships, based on experimental and literature evidence, with the 16 FANC proteins. Each of the various functional subgroups is summarised in **Supplementary Table S1**.

2. SUPPLEMENTARY TABLE

Table S1. Functional subgroups and combined subgroups of the extended FA/BRCA-HRR network

Functional Subgroup	**Genes**
FANC Core Complex	*FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM, MHF1 (APITD1), MHF2 (STRA13), FAAP20 (C1orf86), FAAP24 (C19orf40), FAAP100 (C17ORF70)*
FANCM Anchor Complex	*FANCM ,MHF1 (APITD1), MHF2 (STRA13), FAAP24 (C19orf40)*
Minimal FANCD2 monoubiquitination Complex	*FANCB, FANCL, FAAP100 (C17ORF70)*
ID2-Complex	*FANCD2, FANCI*
FANC core & ID Complex	*FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM, MHF1 (APITD1), MHF2 (STRA13), FAAP20 (C1orf86), FAAP24 (C19orf40), FAAP100 (C17ORF70), FANCI, FANCD2*
Structure-specific Endonucleases	*FANCP (SLX4), FANCQ (ERCC4), SLX1A, MUS81, ERCC1, FAN1, EME1, EME2*
BLM Complex	*RMI1, RMI2, BLM, TOP3A,*
ATM/ATR Checkpoint Proteins	*RPA1, RPA2, RPA3, ATM, ATR, ATRIP, CHEK1 ,RAD9A, RAD17, CHEK2*
RPA proteins	*RPA1, RPA2, RPA3*
MRN Complex	*MRE11A, NBN, RAD50*
BRCA Proteins	*FANCD1 (BRCA2), FANCJ (BRIP1), FANCN (PALB2), BRCA1, BARD1*
RAD51 Paralogues	*FANCO (RAD51C), RAD51, RAD51B, RAD51D, XRCC2, XRCC3*
Mismatch Repair (MMR) Genes	*MLH1, MLH3*
Ubiquitination Modifiers	*UAF1, USP1, UBA52, UBE2T, UBC*


Interaction Evidence Through
Text Mining
Knowledge Base
Experimental Evidence
Homology Similarity
Breast Cancer Associated Genes
KEGG Homologous Recombination Pathway
MRN Complex
MRE11A
RAD50
NBN
FANCJ - (BRIP1)
BARD1
BRCA1
FANCN - (PALB2)
FANCD1 - (BRCA2)
CTIP - (RBBP8)
ATM/ATR Checkpoint Complex
ATM
ATR
RAD17
CHEK1
ATRIP
RAD9A
RPA1
RPA2
RPA3
FANCD2
FANCI
ID Complex
Minimal FANCD2 Mono-ubiquitination Complex
MMR Proteins
MLH1
MLH3
FAAP100 - (C17orf70)
FANCG
FANCL
FANCF
FANCB
FANCA
FANCE
FANCC
FAAP24 - (C19orf40)
FAAP20 - (C1orf86)
FANCM
MHF2 - (STRA13)
MHF1 - (APITD1)
FANC Core Complex
FANCM Anchor Complex
BLM Complex
BLM
RMI1
RMI2 - (C16orf75)
TOP3A
FANCO - (RAD51C)
RAD51
RAD51D - (RAD51L3)
XRCC3
RAD51B - (RAD51L1)
XRCC2
RAD51 Paralogs
Heterodimers
ERCC1
FAN1
EME1
EME2
MUS81
FANCQ - (ERCC4)
FANCP - (BTBD12,SLX4)
SLX1 - (GIYD1)
Structure-specific Endonucleases
UBC
UBE2T
UBA52
UAF1 - (WDR48)
USP1
Ubiquitin Proteins

3. SUPPLEMENTARY FIGURE

Figure S1. The extended FA/BRCA-HRR network generated using STRINGdb (v9.05). Proteins were manually clustered based on functional groups and involvement in the various pathways. Colour of the lines connecting each circle is based on the types of interaction evidences.

4. REFERENCES

1. Duncan EL, Danoy P, Kemp JP, et al. Genome-Wide Association Study Using Extreme Truncate Selection Identifies Novel Genes Affecting Bone Mineral Density and Fracture Risk. *PLoS Genetics*. 2011;7(4):e1001372.
2. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in Bioinformatics*. 2010;11(5):473-483.
3. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
4. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat*. 2011;32(8):894-899.
5. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315.
6. Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 2011;39(Database issue):D561-568.

Chapter 6 Final discussion

6.1 FA/BRCA-HRR mutations in adult AML

In this project a comprehensive genomics approach was used to determine genetic variation across a DNA repair gene network in AML and correlate this to disease. Specifically, the mutation landscape of the genes that comprise the FA DNA repair pathway, involved in repair of DNA damage induced by interstrand crosslinks (ICL), was determined. The central role of this pathway in maintaining genomic integrity and survival of HSC, and the profoundly increased AML risk of FA patients (800 fold compared to normal individuals) provided the rationale for this focused genomics analysis.

A key starting point outlined in **Chapter 3** was the construction of the extended FA/BRCA-HRR network built from the FANC genes and extended to include genes with a close functional connection to the FANC genes. The approach of using STRINGdb to generate this extended FA/BRCA-HRR network provided a visual understanding of interactions between the FANC genes and other gene products known to be involved in DNA damage checkpoints and DNA repair, and highlighted a number of functional subgroups. This approach also allowed an analysis of the FA DNA repair pathway within the context of its broader role in a DNA repair network that includes the ATM/ATR checkpoint kinases, site-specific endonucleases, and the classical BRCA/RAD51 driven homologous recombination repair (HRR) used to repair lethal double-stranded break (DSB). It also provided a basis for the analysis in **Chapter 3** of variants affecting the individual functional subgroups, thus revealing potentially unique roles for these subgroups in AML initiation or progression.

The analysis of the mutation data generated from the WES showed that across this extended FA/BRCA-HRR network, rare variants that are predicted to be deleterious/pathogenic were detected in 70% of the AML patients in the cohort. The median variant allele frequency of the variants was 46.8% (**Appendix C**), consistent with the majority of the variants being heterozygous in the tumour sample. A limitation of this study was that matched non-tumour material was not available for these samples, hence germline/somatic status of variants could not be determined directly from the WES data. However, Sanger sequencing for a selected number of variants in matched samples showed that majority of variants are germline (90%, **Appendix D**); hence, it is predicted that the majority of identified variants in this study will be germline in origin. This prediction is in line with other AML genomics studies that have rarely found somatic mutations affecting DNA repair genes in AML (Kandoth et al., 2013). Based on

this, the focus of discussion throughout this thesis is on the role of these rare variants as potential germline AML predisposition alleles, while acknowledging that some of these variants will be somatically acquired mutations. Consistent with a potential role in AML predisposition, selected genes across this network showed an increased frequency of rare variants in the AML cohort compared to the healthy Australian control cohort and the large population studied in ExAC (**Table 3.9**; **Chapter 4; Section 4.2; Chapter 5; Section 5.2)**. With regards to the FANC genes, only three, *FANCC*, *FANCL* and *FANCM,* showed a significantly increased frequency of variants in the AML cohort. The significance of variants in these genes, all of which encode subunits of the FANC core complex is discussed in **Chapter 4** (**Section 4.2**). In this manuscript the focused analysis on the FANC genes was reported, highlighting a number of specific FANC core and ID2 complex gene variants for which it is important to now determine germline/somatic status. As discussed in the manuscript, germline variants in a number FANC core and ID2 complex genes have previously been reported to be enriched in cancer cohorts; *FANCC* variants in childhood AML, *FANCC* and *FANCE* in familial colorectal cancer, and specific *FANCL* (p.T367fs) and *FANCM* (p.Q1701X) variants in breast cancer (Awan et al., 1998, Esteban-Jurado et al., 2016, Lhota et al., 2016, Kiiski et al., 2014). To further support the mutation data, a cell line model using CRISPR-Cas9 to generate heterozygous and bi-allelic *FANCL* clones was performed. A key finding in the manuscript was shown with the immunofluorescence assay for FANCD2 monoubiquitination in the FANCL heterozygous clones having a consistent reduction of FANCD2 monoubiquitination compared to the wildtype clones. This result indicate that a phenotype for deleterious heterozygous mutations can be observed using highly sensitive assays.

Three approaches have been used throughout this thesis to investigate the significance of rare variants affecting individual genes in AML:

Firstly, the burden analysis was used to compare the AML cohort to an ethnically matched healthy Australian control cohort. This statistical test collapses rare variants at a gene level and tests for enrichment of variants in specific genomic regions in a case-control setting (Lee et al., 2012). Applying this test to the data showed that, among the 58 genes of the extended FA/BRCA-HRR network, rare variants in *RMI1* and *FANCL* were statistically enriched in the AML cohort (compared to the Australian control cohort). The roles and importance of increased mutation burden of *FANCL* for pathogenesis of AML is discussed in **Chapter 4**. RMI1 is a direct interacting partner of the BLM complex, that consists of BLM, TOP3A and RMI2, and performs a critical role in resolution of recombination intermediates arising from ICL and DSB. Depletion of RMI1 destabilises the BLM complex and increases sister chromatid exchange (SCE) (Yin et al., 2005, Xu et al., 2008), suggesting that rare *RMI1* variants in AML could be

associated with higher levels of SCE. The effect of the *RMI1* variants in the leukaemic samples could be investigated by visualising metaphase spreads which is a clinically available diagnostic protocol for FA or copy number analysis which is commonly used for AML diagnosis. In a cancer genomic study by Lu and colleagues, the Burden analysis was performed using the TCGA AML data and enrichment of germline truncation mutations in *FANCO* was observed (Lu et al., 2015). Taken together with the data presented in this thesis, these studies highlight that rare damaging germline variants affecting genes across the extended FA/BRCA-HRR network may be more frequently observed in AML, and, hence, potentially associated with increased AML pre-disposition.

The second approach was to focus on recurrent variants that were identified in the AML cohort. Selective or increased recurrence of a specific germline variant in cancer cohorts is an important indicator of potential risk alleles. In **Chapters 3 and 4**, a number of rare variants in *FANCL*, *FANCM*, *MLH3* and *TOP3A* are described to have been detected in multiple AML samples, and absent, or extremely rare, in healthy cohorts. These variants are worthy of further investigation to confirm germline/somatic status and to undertake using functional assays to define their pathogenicity and correlate their presence with AML predisposition. The functional effect of individual variants can be determined using rescue experiments with cell lines in which the endogenous gene has been knocked down or genetically targeted, or using cell lines known to have loss of function of the respective genes. In such experiments the ability to complement the cell line defect can be determined for the wild type and mutant version of the gene of interest. For the FANC core complex genes with variants of particular interest (*FANCL* and *FANCM*), activity of the FA DNA repair pathway can be determined using western blotting or immunofluorescence assay for FANCD2 monoubiquitination, and measurements of cell cycle length as detailed in **Chapter 4 (section 4.2)**. For the *TOP3A*, which is involved in the resolution of the Holliday junction during HRR, assays to measure DNA repair kinetics by measuring γH2AX by flow cytometry at several time-points after MMC treatment could be performed (Durdik et al., 2015). For the mismatch repair protein, *MLH3*, efficiency of MMR after the introduction of the mutant *MLH3* can be measured using the method outlined by Lei and colleagues (Lei et al., 2004).

Thirdly, it was particularly informative to investigate variants with previous links to disease, and such an analysis could be readily performed by cross-checking variants with a number of disease databases. For this, variants were defined as disease-causing (D-C) if they were *(i)* mutations for which there is evidence in the disease databases causally linking them to the disease/syndrome, or *(ii)* mutations for which there is an OMIM entry linking them to disease. Several variants identified in the extended FA/BRCA-HRR screen of AML samples were D-C

mutations. Several of these variants affecting the FANC genes are discussed in **Chapter 4**, which also included a case-control comparison of D-C FANC gene mutations for the AML cohort and a gender/ethnically-matched control cohort. This indicated a 3-fold increased odds of such variants being present in AML, consistent with these D-C alleles representing modest AML risk variants.

A second group of variants were classified as disease-associated (D-A). For this, variants were defined as D-A mutations if they had been reported in disease databases (COSMIC and HGMD) as being associated with predisposition to a disease which the gene is not classically associated with. An important example includes the FANCM variant, p.Q1701X, which has been associated with the familial syndrome, Tetralogy of Fallot (TOF), and also reported to be enriched in a familial breast cancer cohort (Kiiski et al., 2014). This, and a number of other FANC gene variants, were annotated in HGMD, but were not reported in the FA or breast cancer databases. It is possible that such heterozygous FANC variants may be important in conferring susceptibility to AML, just as they associate with breast cancer and TOF, while not being tolerated in the context of biallelic germline mutation in FA. D-A variants were also identified in ATM and CHEK2 and are discussed in Chapter 3.

6.2 FA/BRCA-HRR mutations in childhood AML

Even though only a small cohort of childhood samples were included in the WES analysis, the data obtained emphasises differences between childhood and adult AML, and raises important questions regarding the potential of *BRCA1/2* variants to confer an increased likelihood of childhood AML. The potential for germline variants to contribute to risk of childhood cancer was recently highlighted by a study that identified germline risk alleles in a significant percentage (approximately 20%) of childhood cancer cases (Zhang et al., 2015). As the number of AML cases in this cancer cohort was severely limited, the study outlined in **Chapter 5**, using a small paediatric AML cohort adds significantly to this field. The most striking observation when comparing the adult and childhood AML cohorts was that the spread of FANC gene variants differed significantly. Predicted damaging *BRCA1/2* variants occurred at dramatically increased frequency in childhood AML, compared to the Australian (P=0.003) and TCGA (P=0.009) adult AML cohort, and also the Australian (P=0.005) and non-Finnish European (P<0.0001) healthy control cohorts. In contrast, variants in the FANC core and ID2 genes, specifically *FANCC*, *FANCL* and *FANCM* occurred at elevated frequency in adult AML (P=0.028) and were absent in the childhood AML. As a result of the observation that *BRCA1/2* variants occurred in a high frequency (50%) of Down syndrome AML patients, a collaborative

study has been initiated to determine *BRCA1/2* mutation status of an Australian Down syndrome cohort that is the subject of extended follow-up.

6.3 Haploinsufficiency model for damaging FANC and BRCA gene variants

The results presented indicate that rare pathogenic and deleterious variants for selected genes in the extended FA/BRCA-HRR network are frequently present in AML, predicted to be most commonly of germline origin, and with a potentially differing spectrum in adult versus childhood AML. In the adult AML cohort, a significantly increased frequency of karyotypic abnormalities was observed for AML samples carrying a variant in the full FA/BRCA-HRR network. Given the predicted germline status of the majority of these variants, it is possible that for this group of AML the increased genomic instability is an important contributor to disease. As the majority of the variants are possibly heterozygous, a key question now is to determine whether these variants are associated with a phenotype that is consistent with increased cancer predisposition. As described in **Chapter 5**, phenotype is likely to be highly dependent on the type of mutation, particularly for *BRCA1/2* where missense mutations in specific domains of the proteins are more deleterious than truncation mutations and can be associated with dominant-negative activity (Vaclova et al., 2016). It is predicted that the D-C and D-A variants are most likely to result in impairment of DNA repair, from the recognition of the DNA lesion, through to the initiation and unhinging of crosslinks, to the eventual resolution of the lesion. Functional studies focused on these variants are now needed to further determine the impact of such mutations. Most recently, van Twist and colleagues have developed cell-free assays to determine the effects of mutations in the FANC core and ID2 genes on ubiquitination of FANCD2 (van Twest et al., 2017) and it will be of interest to test the FANC core and ID2 gene variants that were identified from this study using this system. It will be important that such studies focus on variants that are confirmed to be expressed in the tumour sample as for several of the variants presented in **Chapter 4**, Sanger sequencing of cDNA has shown a lack or very low signal expression for the variant (**Chapter 4; Section 4.2; Supplementary Figure S3**), consistent with silencing of the mutant allele.

In tumours where mutant allele is not expressed there may be concomitant reduction in protein levels and function due to expression from only the wild type allele. Such haploinsufficiency may result in an extremely subtle phenotype that is associated with a modest increase in risk of AML development. There is evidence supporting a haploinsufficient phenotype associated with heterozygous *BRCA1/2* variants (see **Chapter 5**), and although these studies have not been performed in haematopoietic cells, the results are consistent with a model in which such variants

lead to accumulation of AML initiating mutations in HSC (see **Chapter 5**). For the remaining FANC genes, there is more limited evidence of haploinsufficiency associated with heterozygous mutation, although further studies are needed (discussed in more detail in **Chapter 4**).

Given the identification of rare *FANCL* variants at increased frequency in the AML cohort, and increased mutation burden of *FANCL* in AML compared to healthy controls (see **Chapter 4**), the MCF10A model of *FANCL* haploinsufficiency, generated as part of this project, will provide an important model system allowing a range of sensitive assays to be used to investigate the cellular phenotype associated with haploinsufficiency. With regard to non-FANC genes there is strong evidence for a haploinsufficient tumour susceptibility phenotype in murine $Atm^{+/-}$ models, in which a two-fold increase in the formation of carcinogen-induced mammary tumours was observed for $Atm^{+/-}$ mice (Lu et al., 2006). Furthermore, when *Atm* haploinsufficiency was combined with haploinsufficiency in other genes (such as *Rad9A*, *Mrad9* and *p53*) distinct tumour phenotypes from the wild type strains were achieved (Kleiman et al., 2007, Smilenov et al., 2005, Umesako et al., 2005). Thus, the co-occurrence of rare heterozygous variants across this FA/BRCA-HRR network is potentially an important factor in determining risk of cancer development, and several patients were identified with more than one rare predicted damaging variant across this network. Haploinsufficient phenotypes have also been proposed for a number of other non-FANC genes in the extended FA/BRCA-HRR network based on murine and human *in vitro* models (O'Driscoll, 2008).

In summary, based on these studies it is proposed that haploinsufficiency for selected FANC and BRCA genes in particular may affect DNA repair capacity in haematopoietic stem and progenitor populations, leading to subtle increases in genomic instability and accumulation of AML initiating mutations. The difference in spectrum of variants between the adult and childhood AML cohorts suggests that *BRCA1/2* haploinsufficiency induces a more rapid accumulation of initiating mutations, hence the earlier age of onset. Consistent with this concept, it is well-established that both *BRCA1/2* and FA pathway, along with the other major DNA repair pathways, are critical for maintenance of genomic stability in HSC (Kenyon and Gerson, 2007, Pontel et al., 2015). A key role of the FA pathway has been demonstrated particularly under replicative stress, or when HSC are confronted with cross-linking toxins, and this raises the possibility that the level of risk associated with these rare variants may be affected significantly by other factors, such as environmental exposure, infectious agents and/or other genetic variants that affect metabolism of aldehydes (Pontel et al., 2015, Parmar and D'Andrea, 2017). Finally, based on the gene expression profiling and GSEA (presented in **Chapter 3**), it

is speculated that the specific gene affected (*BRCA1/2* or other FANC genes) may contribute to the primitive haemopoietic population at risk of malignant transformation.

6.4 Relevance to the field: germline predisposition and cancer risk

In recent years, with the advent of the genomics era and developments in next generation sequencing, AML classification has progressed considerably. The number of recurrent somatically mutated genes in AML has reached saturation and mutation status for several somatically mutated genes is now used in classification and prognosis as described in **Chapter 1** (Arber et al., 2016). However, only a handful of genes have been associated with germline predisposition. Rare mutations in genes such as *RUNX1*, *CEPBA*, *GATA2* and *DDX41* are associated with dominantly inherited predisposition to myeloid malignancies. D-C mutations in these genes are extremely rare and development of myeloid malignancy is highly penetrant and readily detectable in familial studies (see **Chapter 1; Section 1.7.1**). Mutations in this group of genes have now been incorporated into the most current revision of the ELN risk stratification (Dohner et al., 2017). In contrast, the studies in this thesis are consistent with modest risk of AML development associated with rare variants affecting the genes of the extended FA/BRCA-HRR network (estimated at approximately 3-fold increased risk for known D-C alleles), with potential for this to be higher with the influence of environmental or genetic factors. These findings need to be reconciled with the large study in 2012 by Goldin and colleagues showing that adult AML is not associated with a high level of familial risk (Goldin et al., 2012). It is likely that for AML a relatively small number of genes confer germline risk, including modest risk associated with the FANC core and ID2 genes, and the effect of this will be small even in a large cohort of first-degree relatives of AML patients. The identification of alleles that confer only a modest increased cancer risk is associated with a number of ethical issues. These have been reviewed in the literature and are discussed in the manuscripts included in **Chapters 4** and **5**.

6.5 Future directions

Genomic instability is recognised as a hallmark of cancer (Hanahan and Weinberg, 2011) and it is not surprising that genetic variation in DNA repair capacity, particularly in pathways known to be critical in HSC, is an important determinant of AML risk. This emphasised the need for a better understanding of variants in DNA repair genes. This better understanding means determining their interaction with endogenous and exogenous factors, and the potential

long-term consequences of subtle changes to pathway effectiveness, compared to the severe deficiencies that lead to high-risk syndromes such as FA. Finally, some patients in the subgroup of AML with heterozygous FANC gene variants may be candidates for targeted therapy. For example, PARP inhibitors (PARPi) were first shown to be effective in tumour cells with severe deficiency in *BRCA1/2*, and have since been shown to kill tumour cells with severe deficiency for other FANC genes (Kennedy et al., 2007). Recent studies have also shown that the loss of function in BRCA1/2 does not always result in PARPi sensitivity, but rather the main determinant of sensitivity is a "BRCAness" phenotype (a term used to describe a specific type of genomic instability including mutation landscape and expression of genes relating to DNA repair deficiencies) (Murata et al., 2016, Lord and Ashworth, 2017, Lord and Ashworth, 2016). As discussed in **Chapter 4**, further study is now needed to determine whether there is a therapeutic window for use of PARPi inhibitors in tumours with heterozygous FANC variants. A number of clinical trials testing PARPi in conjunction with chemotherapeutic compounds are currently underway in cervical, breast and ovarian cancer, and it will be of interest to determine whether responses are correlated with mutation status of FANC genes. The GSEA analysis outlined in **Chapter 4** is also suggestive that bulk leukaemic cells from patients with heterozygous FANC core & ID2 complex variants may exhibit elevated levels of replicative stress, and this raises the possibility that inhibitors targeting DNA repair pathways that are required during DNA replication (i.e. PARPi, Ku inhibitors, RAD52 inhibitors) may provide an option as therapy (Matulonis and Monk, 2017, Weterings et al., 2016, Kumar et al., 2017, Lok et al., 2013, Hengel et al., 2016). There is precedent for use of such agents in combination with chemotherapy in solid cancers (Matulonis and Monk, 2017).

In conclusion, the aims set forth at the conception of the study have been achieved. A detailed case control study in AML to identify are deleterious, and disease related, gene variants affecting FANC genes, and genes encoding other factors that closely interact with the FA DNA repair pathway. This also resulted in the identification of D-C and D-A mutations across these genes. Gene expression profiling of AMLs carrying these variants in different functional subgroups of the FA/BRCA-HRR network have also revealed novel insight on the activity of other biological pathways namely, DNA replication and synthesis are affected. Most importantly, a heterozygous phenotype have been shown in a cell line model for FANCL heterozygous mutation.

Responses to examiners' comments

Comments from Examiner 1

Note: The comments from the examiner are shown in quotation and blue font. The responses are shown in black font.

Chapter 1

"There is absolutely no background at all to the central question of the thesis: i.e. why do you think that there are germline predisposing mutations in AML, outside of the rare multiple case families with known mutations? What are the familial relative risks?" Given that the contribution of somatic mutations to the pathogenesis of AML has been established by the advancement in NGS, the next logical step utilising NGS data was to determine whether germline mutations would also contribute to AML pathogenesis. As reviewed and described in Chapter 1; Section 1.7, the contribution germline mutations to the pathogenesis of AML though not as well established as somatic mutations, is still a question of significant interest. The incorporation of germline mutations in the classification of AML as outlined by WHO most recently (Arber et al., 2016), shows the recognition by the scientific community of germline mutations as contributing factors to AML pathogenesis.

"Are there twin studies that have estimated heritability?" Only limited studies in Chapter 1; section 1.7.1 (page 16).

"What sort of GWAS have been done, how big and what have they found?" Only limited studies in Chapter 1; section 1.7.1 (page 16).

"Are there multiple case AML families with no known predisposing mutations?" Yes. The Australian Familial Haematological Cancer Study (Hamish Scott and Richard D'Andrea) has identified several of such families.

"What sort of power did your study have to find predisposing mutations of what effect size?" Unlike GWAS which examine known population SNPs with MAF>0.05 and small effect size (Wellcome Trust Case Control et al., 2007; Choi et al., 2013), our study focused on rare variants and attempted to identify rare variants/mutations that confers relatively higher risk and can be detected in smaller cohorts.

"Have people examined them with WES and WGS? What sort of similar WES and WGS studies have been done for other cancers e.g. melanoma, breast cancer, prostate cancer, colorectal

cancer etc, with what success?" Similar studies, as well as targeted sequencing studies have been performed in solid malignancies such as breast, prostate and colorectal cancer as described in Chapter 1; Section 1.7.3.2.2 (page 21). These studies have shown increased number of mutations FANC genes in solid malignancies (Zhang et al., 2015), as well as specific mutations/genes that are associated with specific types of malignancies (i.e FANCM-p.Q170X is associated with increased risk of developing triple negative breast cancer (Kiiski et al., 2014); mutations in FANCC, FANCE, BRIP1 and BRCA2 which have been shown to segregate within families with colorectal cancer (Esteban-Jurado et al., 2016).

"There needs to be extensive discussion of these issues, if necessary (for reasons of space) by cutting out some details on DNA repair mechanisms, or methods of classifying AML. Goldin et al JCO 2012 is an important reference." This has been addressed in Chapter 1; section 1.7.1 (page 13).

Chapter 3

"I am concerned that there were no controls sequenced in the second batch of cases. Could this have introduced biases, and contributed to the identification of 'mutations' in cases that were absent in controls. In addition, the cases in batch II were done on a different platform, and to higher depth than the controls which were all done in the first batch. This could lead to biases that might result in a high frequency of 'mutations' in cases than controls." This is incorrect. A second batch of controls (n=49) were sequenced along with the second batch of AML samples and analysed using the same bioinformatics pipeline. This was detailed in the supplementary information of the attached manuscript in Chapter 4 (page 134). The bioinformatics pipeline also took into consideration the differences in platform, also detailed in the supplementary information of the attached manuscript in Chapter 4. Hence, approached were included to ensure there were no bias in variants identified across the two batches of WES in both the AML and in healthy controls.

"It is not clear to me that the candidate looked for variants that were unique to their control series but absent in cases. This is an important control – are rare variants enriched in cases, or is there a similar but different spectrum of rare variants found also in controls?" The variants that were unique/present at a higher frequency in the controls (than in the AML) were analysed and are now presented as **Appendix Y** (i.e variants that were only present in the controls with a MAF<0.001 and CADD>10). These variants were analysed when comparing the number/frequency of D-C mutations in the AML vs controls, as well as the Burden test.

"Re Table 3.8., it is not surprising that there are multiple variants that occur more than once in the cases but less frequently in ExAC. But what about looking at the reverse? There are probably variants that occur more often in ExAC than in cases - does this mean that they are protective? No it is probably just chance but it would be a useful exercise to do this, and discuss the difference in the characteristics (if there are any) between the possible 'risk' and 'protective' variants." Yes, I acknowledge that there are variants that occur more frequently in ExAC than in the AML cohort in this study. However, this does not mean that such variants are protective in nature, they are most likely to be rare population SNPs that by chance are not represented in the small AML cohort.

"There is no discussion as to whether the cases had a family history of AML, or of any cancer, or were all sporadic cases of unknown family history. To find predisposing mutations it would have been ideal to enrich for familial cases, if they exist (and if they don't, the likelihood of finding predisposing mutations must be low)." There were limited family history available for the AML samples used in this study. For patients with D-C mutations, our clinical collaborator has checked for family history in the available clinical information.

"My most important criticism of this chapter is that there was no correction for multiple testing. For example, 28 tests were done (for 28 genes) in Table 3.6 so you need to use a Bonferroni corrected p value of 0.0017; not one of 0.05. The same applies to the other analyses, and only genes that reach significance with this correction for multiple testing should be further described and used as the basis for additional analysis, such as the expression and GSEA." This has now been addressed in this chapter, the four genes ATM, C17ORF70, MLH3 and TOP3A are statistically enriched in the AML cohort in comparison to the ExAC database after multiple correction (page 66).

"At the end of Chapter 3, I think you also need to discuss my major caveat - that there is no strong evidence that any of these 'germline mutations' are associated with risk. If you correct for multiple testing, few comparisons are significant; comparison of variants identified through different platforms and analysed in different ways, can lead to false positive associations; there is no independent validation; and no sign that these 'mutations' occur and segregate with disease in multiple case families. So given all of these caveats i think the GSEA analyses need to be considered with caution." No single variant is strongly associated with risk of AML, however analysis of the independent TCGA AML cohort as described in Chapter 4 (section 4.2; page

104) also showed enrichment of rare predicted damaging FANC variants in their AML cohort. The results from this study also showed that there is an increased frequency of D-C mutations in the AML compared to normal healthy controls (**Table 3.9**). As previously mentioned, the potential for false positive associated with identifying variants across two different platforms has been considered during the bioinformatics analysis. Due to the poor availability of family history of the AML samples in this cohort, analyses from a familial inheritance and segregation of variants could not be performed. However, we point out that given the modest risk associated with the variants, we predict incomplete penetrance in families, or large families or collection of families may only reveal this association. The GSEA analyses were considered with caution and were only used to supplement the results from the WES analysis, providing potential clues to the biological consequences associated with rare FANC gene variants.

Chapter 4

Please note that after obtaining the reviewers' comments from the initial submission, the attached manuscript has been edited extensively, including additional experiments which have been performed. The manuscript has since been resubmitted to the same journal (PLOS Genetics) for review, and is now included in Chapter 4; section 4.2 including the reviewers' comments and our responses.

"i think only FANCLL survives correction for multiple testing, so personally I'd focus entirely on that gene. What is the OR for FANCL? and the 95% CI? If it is about 3-4, then wouldn't you expect to see multiple AML cases among FANCL families? and that the cases you found with FANCL 'mutations' would have a family hx?" It is important to note that FANCL is one of the least common subtype of FA which in itself is already a very rare disorder. This meant that there is limited to no familial studies on FANCL mutations in the literature.

"The MCF10A cell line was a strange choice for CRISPR of FANCL. Are there no more appropriate lines for a study focused on AML? If not, this at least should be discussed. And was no characterization of the edited lines done at all? It seems odd to generate the clones and not do a basic characterization of the activity of the pathway." As described in Chapter 4; Section 4.2, MCF10A cell line is a non-tumour derived diploid cell line with intact FA DNA repair pathway. Characterisation of selected clones have been performed and the results are included in the revised version of the manuscript which has been resubmitted. Upon receiving the reviewers' comments from PLOS Genetics, a revised manuscript has been resubmitted as of 31st November 2017 and is attached in Chapter 4; Section 4.2.

Chapter 5

"what is the evidence that childhood AML has a genetic component?" As reviewed by Stieglitz and Loh (Stieglitz and Loh 2013), germline mutations contribute to childhood AML similarly to adult AML through mutations in familial AML genes such as *CEBPA*, *RUNX1* and *GATA2* (reviewed and described in Chapter 1; Section 1.7.1 and 1.7.2).

"did you correct for multiple testing in the analysis?" Multiple correction testing was not performed.

"the term 'mutation' is not appropriate unless known to be disease causing" This has been addressed in the attached manuscript. The term "rare variant" is now used to unless the variant is a known disease-causing (D-C) mutation.

"are the BRCA1/2 variants described really pathogenic for BC or unclassified variants?" The BRCA1/2 variants described are rare predicted pathogenic variants. If a variant has been reported as pathogenic in breast cancer or other diseases, this is stated in the text.

Final Discussion

"I think this should focus on the most reliable results eg genes that survive correction for multiple testing, and what should be done to validate this finding. The sorts of experiments proposed might tell you if the variant is non-functional but won't strengthen the case that it is associated with risk of AML. What is the best way to do that? Is there any sign of second hits in the carriers of the BRCA/1 variants? How could that be evaluated? Why is it important?" I agree that the experiments proposed may test whether the variants identified directly affected the function of the protein and different experiments would have to be designed to determine that association with the risk of AML. Such experiments would require analysis of germline variants for a much larger multi-centre AML cohort, as well as a large cohort of normal healthy controls. This is discussed further in detail in Chapter 6 (section 6.4; page 193). With regards to the *BRCA1/2* variants, I agree it would be important to determine the presence/absence of a second hit to the genes. The WES data show an absence of a second hit but the potential loss of heterozygosity (LOH) of the second allele of the genes cannot be confirmed from the WES data based on variant allele frequency. As shown in Supplementary Figure S3 in Section 4.2; Chapter 4, sequencing the cDNA of the AML sample would reveal if the variant identified from the WES is actually expressed (transcribed).

Comments from Examiner 2

"The methods for exome sequencing have very limited detail compared to the other methods described in chapter 2. The sequencing methods are included in the Chapter 4 and 5 manuscripts therefore this section of chapter 2 should acknowledge the Diamentina Institute for performing the sequencing and analysis and refer to chapter 4 & 5 methods. Alternatively, details on DNA quantification, quality control and input amount should be added to chapter 2. A brief description of library preparation and quality control steps and amount used for the HiSeq runs need to be included too." The Diamentina Institute has been acknowledged for performing the sequencing and initial analysis of the data in Chapter 2.

"The method for exome sequencing should be placed after details of the patient cohort and ethics. Details of the healthy control cohort (including Duncan et al reference) should be added too. Why was an all-female control cohort used? This is not explained." An all-female cohort was used due it being readily available to the team at the Dimentina Institute. Also see Chapter 4, section 4.2 and detailed response to reviewers' comments.

"Further explanation of figures 3.5 – 3.7 is needed in the results of chapter 3. What results are these figures showing and what does each result mean in the context of the aims of the project? This explanation first appears in the discussion sections of chapter 3 and should be first described in the results." Due to GSEA being a hypothesis generating tool, as well as the discussion of GSEA being of a more speculative nature, there is overlap in what is considered to be results and discussion. Hence, it was decided that the GSEA will be discussed in the discussion section and integrated with the conclusions from the findings from the WES data.

"The Chapter 3 discussion describes the results well but further in depth discussion of the subgroups should be included. Is there clinical relevance for any of the subgroups? Would they be expected to respond to standard and targeted therapies differently due to changes in specific DNA repair processes?" As described in Chapter 3; Section 3.2.5.1, aside from the association with abnormal karyotype for the FA/BRCA-HRR network mutant AML and KEGG-HRR subgroup mutant AML which also had an association with *MLL* translocations, there were no other clinical characteristics that associated with the various subgroups. It is possible that selected therapies could be used for AML with partially deficient DNA repair pathways. As discussed in Chapter 4, Section 4.2, a combinational treatment with PARP1 inhibitor and chemotherapeutic 5-fluorouracil (5-FU) has recently been shown by Falzacappa and colleagues to be effective in both *in vitro* and *in vivo* models of AML and ALL.

"The conclusion of chapter 3 should reflect the aim of the study outlined in the introduction of the chapter. Was the aim achieved? Was the hypothesis supported by the data?" The conclusions of Chapter 3 has been adjusted to reflect the aims of the project (page 90). The data in the chapter support that hypothesis that rare DNA repair gene variants associated with the FA/BRCA-HRR network are enriched in AML compared to controls.

"The role of common AML mutations and karyotypic abnormalities is only very briefly mentioned in the summary and throughout the discussion. It is likely that these features of AML influence the results and should be carefully considered and discussed in more detail – perhaps as a brief new section in the discussion." Similar comment were given by the editor and reviewers for the manuscript in Chapter 4; Section 4.2 and has been addressed in Chapter 4.

"A final brief conclusion that summarises the results of each of the overall project aims should be added after the future directions to end the thesis with a concluding paragraph." A brief conclusion summarising the results of each aim have been added after the future directions at the end of Chapter 6 (page 194).

Bibliography

ABU-DUHIER, F. M., GOODEVE, A. C., WILSON, G. A., CARE, R. S., PEAKE, I. R. & REILLY, J. T. 2001. Identification of novel FLT-3 Asp835 mutations in adult acute myeloid leukaemia. *Br J Haematol,* 113**,** 983-8.

ADACHI, D., ODA, T., YAGASAKI, H., NAKASATO, K., TANIGUCHI, T., D'ANDREA, A. D., ASANO, S. & YAMASHITA, T. 2002. Heterogeneous activation of the Fanconi anemia pathway by patient-derived FANCA mutants. *Hum Mol Genet,* 11**,** 3125-34.

AHMED, M., STERNBERG, A., HALL, G., THOMAS, A., SMITH, O., O'MARCAIGH, A., WYNN, R., STEVENS, R., ADDISON, M., KING, D., STEWART, B., GIBSON, B., ROBERTS, I. & VYAS, P. 2004. Natural history of GATA1 mutations in Down syndrome. *Blood,* 103**,** 2480-9.

AKASHI, K., TRAVER, D., MIYAMOTO, T. & WEISSMAN, I. L. 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature,* 404**,** 193-7.

ALI, A. M., KIRBY, M., JANSEN, M., LACH, F. P., SCHULTE, J., SINGH, T. R., BATISH, S. D., AUERBACH, A. D., WILLIAMS, D. A. & MEETEI, A. R. 2009. Identification and characterization of mutations in FANCL gene: a second case of Fanconi anemia belonging to FA-L complementation group. *Hum Mutat,* 30**,** E761-70.

ALORAIFI, F., MCDEVITT, T., MARTINIANO, R., MCGREEVY, J., MCLAUGHLIN, R., EGAN, C. M., CODY, N., MEANY, M., KENNY, E., GREEN, A. J., BRADLEY, D. G., GERAGHTY, J. G. & BRACKEN, A. P. 2015. Detection of novel germline mutations for breast cancer in non-BRCA1/2 families. *FEBS J,* 282**,** 3424-37.

AMEZIANE, N., MAY, P., HAITJEMA, A., VAN DE VRUGT, H. J., VAN ROSSUM-FIKKERT, S. E., RISTIC, D., WILLIAMS, G. J., BALK, J., ROCKX, D., LI, H., ROOIMANS, M. A., OOSTRA, A. B., VELLEUER, E., DIETRICH, R., BLEIJERVELD, O. B., MAARTEN ALTELAAR, A. F., MEIJERS-HEIJBOER, H., JOENJE, H., GLUSMAN, G., ROACH, J., HOOD, L., GALAS, D., WYMAN, C., BALLING, R., DEN DUNNEN, J., DE WINTER, J. P., KANAAR, R., GELINAS, R. & DORSMAN, J. C. 2015. A novel Fanconi anaemia subtype associated with a dominant-negative mutation in RAD51. *Nat Commun,* 6**,** 8829.

ANTONIOU, A. C., CASADEI, S., HEIKKINEN, T., BARROWDALE, D., PYLKAS, K., ROBERTS, J., LEE, A., SUBRAMANIAN, D., DE LEENEER, K., FOSTIRA, F., TOMIAK, E., NEUHAUSEN, S. L., TEO, Z. L., KHAN, S., AITTOMAKI, K., MOILANEN, J. S., TURNBULL, C., SEAL, S., MANNERMAA, A., KALLIONIEMI, A., LINDEMAN, G. J., BUYS, S. S., ANDRULIS, I. L., RADICE, P., TONDINI, C., MANOUKIAN, S., TOLAND, A. E., MIRON, P., WEITZEL, J. N., DOMCHEK, S. M., POPPE, B., CLAES, K. B., YANNOUKAKOS, D., CONCANNON, P., BERNSTEIN, J. L., JAMES, P. A., EASTON, D. F., GOLDGAR, D. E., HOPPER, J. L., RAHMAN, N., PETERLONGO, P., NEVANLINNA, H., KING, M. C., COUCH, F. J., SOUTHEY, M. C., WINQVIST, R., FOULKES, W. D. & TISCHKOWITZ, M. 2014. Breast-cancer risk in families with mutations in PALB2. *N Engl J Med,* 371**,** 497-506.

ARBER, D. A., ORAZI, A., HASSERJIAN, R., THIELE, J., BOROWITZ, M. J., LE BEAU, M. M., BLOOMFIELD, C. D., CAZZOLA, M. & VARDIMAN, J. W. 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood,* 127**,** 2391-405.

ARIZA, A. C., DEEN, P. M. & ROBBEN, J. H. 2012. The succinate receptor as a novel therapeutic target for oxidative and metabolic stress-related conditions. *Front Endocrinol (Lausanne),* 3**,** 22.

ASHMAN, L. K. & GRIFFITH, R. 2013. Therapeutic targeting of c-KIT in cancer. *Expert Opin Investig Drugs,* 22**,** 103-15.

AUERBACH, A. D., ADLER, B. & CHAGANTI, R. S. 1981. Prenatal and postnatal diagnosis and carrier detection of Fanconi anemia by a cytogenetic method. *Pediatrics,* 67**,** 128-35.

AWAN, A., MALCOLM TAYLOR, G., GOKHALE, D. A., DEARDEN, S. P., WILL, A., STEVENS, R. F., BIRCH, J. M. & EDEN, T. 1998. Increased frequency of Fanconi anemia group C genetic variants in children with sporadic acute myeloid leukemia. *Blood,* 91**,** 4813-4.

BAGGER, F. O., SASIVAREVIC, D., SOHI, S. H., LAURSEN, L. G., PUNDHIR, S., SONDERBY, C. K., WINTHER, O., RAPIN, N. & PORSE, B. T. 2016. BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. *Nucleic Acids Res,* 44**,** D917-24.

BAHASSI, E. M., OVESEN, J. L., RIESENBERG, A. L., BERNSTEIN, W. Z., HASTY, P. E. & STAMBROOK, P. J. 2008. The checkpoint kinases Chk1 and Chk2 regulate the functional associations between hBRCA2 and Rad51 in response to DNA damage. *Oncogene,* 27**,** 3977-85.

BARQUINERO, J. F., BARRIOS, L., RIBAS, M., EGOZCUE, J. & CABALLIN, M. R. 2001. Cytogenetic sensitivity of three Fanconi anemia heterozygotes to bleomycin and ionizing radiation. *Cancer Genet Cytogenet,* 124**,** 80-3.

BELL, D. W., VARLEY, J. M., SZYDLO, T. E., KANG, D. H., WAHRER, D. C., SHANNON, K. E., LUBRATOVICH, M., VERSELIS, S. J., ISSELBACHER, K. J., FRAUMENI, J. F., BIRCH, J. M., LI,

F. P., GARBER, J. E. & HABER, D. A. 1999. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science,* 286**,** 2528-31.
BENNETT, J. M., CATOVSKY, D., DANIEL, M. T., FLANDRIN, G., GALTON, D. A., GRALNICK, H. R. & SULTAN, C. 1976. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol,* 33**,** 451-8.
BENSIMON, A., SCHMIDT, A., ZIV, Y., ELKON, R., WANG, S. Y., CHEN, D. J., AEBERSOLD, R. & SHILOH, Y. 2010. ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage. *Sci Signal,* 3**,** rs3.
BERI-DEXHEIMER, M., LATGER-CANNARD, V., PHILIPPE, C., BONNET, C., CHAMBON, P., ROTH, V., GREGOIRE, M. J., BORDIGONI, P., LECOMPTE, T., LEHEUP, B. & JONVEAUX, P. 2008. Clinical phenotype of germline RUNX1 haploinsufficiency: from point mutations to large genomic deletions. *Eur J Hum Genet,* 16**,** 1014-8.
BERWICK, M., SATAGOPAN, J. M., BEN-PORAT, L., CARLSON, A., MAH, K., HENRY, R., DIOTTI, R., MILTON, K., PUJARA, K., LANDERS, T., DEV BATISH, S., MORALES, J., SCHINDLER, D., HANENBERG, H., HROMAS, R., LEVRAN, O. & AUERBACH, A. D. 2007. Genetic heterogeneity among Fanconi anemia heterozygotes and risk of cancer. *Cancer Res,* 67**,** 9591-6.
BHATNAGAR, N., NIZERY, L., TUNSTALL, O., VYAS, P. & ROBERTS, I. 2016. Transient Abnormal Myelopoiesis and AML in Down Syndrome: an Update. *Curr Hematol Malig Rep,* 11**,** 333-41.
BIONDI, A. & RAMBALDI, A. 1996. Molecular diagnosis and monitoring of acute myeloid leukemia. *Leuk Res,* 20**,** 801-7.
BLUTEAU, D., MASLIAH-PLANCHON, J., CLAIRMONT, C., ROUSSEAU, A., CECCALDI, R., DUBOIS D'ENGHIEN, C., BLUTEAU, O., CUCCUINI, W., GACHET, S., PEFFAULT DE LATOUR, R., LEBLANC, T., SOCIE, G., BARUCHEL, A., STOPPA-LYONNET, D., D'ANDREA, A. D. & SOULIER, J. 2016. Biallelic inactivation of REV7 is associated with Fanconi anemia. *J Clin Invest,* 126**,** 3580-4.
BOWEN, D. T., FREW, M. E., HILLS, R., GALE, R. E., WHEATLEY, K., GROVES, M. J., LANGABEER, S. E., KOTTARIDIS, P. D., MOORMAN, A. V., BURNETT, A. K. & LINCH, D. C. 2005. RAS mutation in acute myeloid leukemia is associated with distinct cytogenetic subgroups but does not influence outcome in patients younger than 60 years. *Blood,* 106**,** 2113-9.
BUCHHOLZ, T. A., WU, X., HUSSAIN, A., TUCKER, S. L., MILLS, G. B., HAFFTY, B., BERGH, S., STORY, M., GEARA, F. B. & BROCK, W. A. 2002. Evidence of haplotype insufficiency in human cells containing a germline mutation in BRCA1 or BRCA2. *Int J Cancer,* 97**,** 557-61.
BUCKLEY, J. D., BUCKLEY, C. M., BRESLOW, N. E., DRAPER, G. J., ROBERSON, P. K. & MACK, T. M. 1996. Concordance for childhood cancer in twins. *Med Pediatr Oncol,* 26**,** 223-9.
BUSTAMANTE, C. D., BURCHARD, E. G. & DE LA VEGA, F. M. 2011. Genomics for the world. *Nature,* 475**,** 163-5.
CANCER GENOME ATLAS RESEARCH, N. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med,* 368**,** 2059-74.
CANTOR, S. B., BELL, D. W., GANESAN, S., KASS, E. M., DRAPKIN, R., GROSSMAN, S., WAHRER, D. C., SGROI, D. C., LANE, W. S., HABER, D. A. & LIVINGSTON, D. M. 2001. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell,* 105**,** 149-60.
CASPER, A. M., DURKIN, S. G., ARLT, M. F. & GLOVER, T. W. 2004. Chromosomal instability at common fragile sites in Seckel syndrome. *Am J Hum Genet,* 75**,** 654-60.
CECCALDI, R., PARMAR, K., MOULY, E., DELORD, M., KIM, J. M., REGAIRAZ, M., PLA, M., VASQUEZ, N., ZHANG, Q. S., PONDARRE, C., PEFFAULT DE LATOUR, R., GLUCKMAN, E., CAVAZZANA-CALVO, M., LEBLANC, T., LARGHERO, J., GROMPE, M., SOCIE, G., D'ANDREA, A. D. & SOULIER, J. 2012. Bone marrow failure in Fanconi anemia is triggered by an exacerbated p53/p21 DNA damage response that impairs hematopoietic stem and progenitor cells. *Cell Stem Cell,* 11**,** 36-49.
CHALLEN, G. A., BOLES, N., LIN, K. K. & GOODELL, M. A. 2009. Mouse hematopoietic stem cell identification and analysis. *Cytometry A,* 75**,** 14-24.
CHERUBINI, G., NAIM, V., CARUSO, P., BURLA, R., BOGLIOLO, M., CUNDARI, E., BENIHOUD, K., SAGGIO, I. & ROSSELLI, F. 2011. The FANC pathway is activated by adenovirus infection and promotes viral replication-dependent recombination. *Nucleic Acids Res,* 39**,** 5459-73.
CHOI, H., JUNG, C., SOHN, S. K., KIM, S., KIM, H. J., KIM, Y. K., KIM, T., ZHANG, Z., SHIN, E. S., LEE, J. E., MOON, J. H., KIM, S. H., KIM, K. H., MUN, Y. C., KIM, H., PARK, J., KIM, J. & KIM, D. 2013. Genome-wide genotype-based risk model for survival in acute myeloid leukaemia patients with normal karyotype. *Br J Haematol,* 163**,** 62-71.
CHRZANOWSKA, K. H., GREGOREK, H., DEMBOWSKA-BAGINSKA, B., KALINA, M. A. & DIGWEED, M. 2012. Nijmegen breakage syndrome (NBS). *Orphanet J Rare Dis,* 7**,** 13.
CICCIA, A. & ELLEDGE, S. J. 2010. The DNA damage response: making it safe to play with knives. *Mol Cell,* 40**,** 179-204.
CICCIA, A., MCDONALD, N. & WEST, S. C. 2008. Structural and functional relationships of the XPF/MUS81 family of proteins. *Annu Rev Biochem,* 77**,** 259-87.

CIRIZA, J., THOMPSON, H., PETROSIAN, R., MANILAY, J. O. & GARCIA-OJEDA, M. E. 2013. The migration of hematopoietic progenitors from the fetal liver to the fetal bone marrow: lessons learned and possible clinical applications. *Exp Hematol,* 41**,** 411-23.

COLLIN, M., DICKINSON, R. & BIGLEY, V. 2015. Haematopoietic and immune defects associated with GATA2 mutation. *Br J Haematol,* 169**,** 173-87.

COLLIS, S. J., CICCIA, A., DEANS, A. J., HOREJSI, Z., MARTIN, J. S., MASLEN, S. L., SKEHEL, J. M., ELLEDGE, S. J., WEST, S. C. & BOULTON, S. J. 2008. FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. *Mol Cell,* 32**,** 313-24.

CORTEZ, D., GUNTUKU, S., QIN, J. & ELLEDGE, S. J. 2001. ATR and ATRIP: partners in checkpoint signaling. *Science,* 294**,** 1713-6.

COULON, S., GAILLARD, P. H., CHAHWAN, C., MCDONALD, W. H., YATES, J. R., 3RD & RUSSELL, P. 2004. Slx1-Slx4 are subunits of a structure-specific endonuclease that maintains ribosomal DNA in fission yeast. *Mol Biol Cell,* 15**,** 71-80.

CRAIG, W., KAY, R., CUTLER, R. L. & LANSDORP, P. M. 1993. Expression of Thy-1 on human hematopoietic progenitor cells. *J Exp Med,* 177**,** 1331-42.

CREUTZIG, U., VAN DEN HEUVEL-EIBRINK, M. M., GIBSON, B., DWORZAK, M. N., ADACHI, S., DE BONT, E., HARBOTT, J., HASLE, H., JOHNSTON, D., KINOSHITA, A., LEHRNBECHER, T., LEVERGER, G., MEJSTRIKOVA, E., MESHINCHI, S., PESSION, A., RAIMONDI, S. C., SUNG, L., STARY, J., ZWAAN, C. M., KASPERS, G. J., REINHARDT, D. & GROUP, A. M. L. C. O. T. I. B. S. 2012. Diagnosis and management of acute myeloid leukemia in children and adolescents: recommendations from an international expert panel. *Blood,* 120**,** 3187-205.

CROSSAN, G. P., VAN DER WEYDEN, L., ROSADO, I. V., LANGEVIN, F., GAILLARD, P. H., MCINTYRE, R. E., SANGER MOUSE GENETICS, P., GALLAGHER, F., KETTUNEN, M. I., LEWIS, D. Y., BRINDLE, K., ARENDS, M. J., ADAMS, D. J. & PATEL, K. J. 2011. Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. *Nat Genet,* 43**,** 147-52.

CUADRADO, M., MARTINEZ-PASTOR, B., MURGA, M., TOLEDO, L. I., GUTIERREZ-MARTINEZ, P., LOPEZ, E. & FERNANDEZ-CAPETILLO, O. 2006. ATM regulates ATR chromatin loading in response to DNA double-strand breaks. *J Exp Med,* 203**,** 297-303.

CUNNIFF, C., BASSETTI, J. A. & ELLIS, N. A. 2017. Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. *Mol Syndromol,* 8**,** 4-23.

D'ANDREA, A. D. 2013. BRCA1: a missing link in the Fanconi anemia/BRCA pathway. *Cancer Discov,* 3**,** 376-8.

DANG, L. & SU, S. M. 2017. Isocitrate Dehydrogenase Mutation and (R)-2-Hydroxyglutarate: From Basic Discovery to Therapeutics Development. *Annu Rev Biochem,* 86**,** 305-331.

DANG, L., YEN, K. & ATTAR, E. C. 2016. IDH mutations in cancer and progress toward development of targeted therapeutics. *Ann Oncol,* 27**,** 599-608.

DAVIS, A. J., CHEN, B. P. & CHEN, D. J. 2014. DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. *DNA Repair (Amst),* 17**,** 21-9.

DE BRAEKELEER, E., DOUET-GUILBERT, N. & DE BRAEKELEER, M. 2014. RARA fusion genes in acute promyelocytic leukemia: a review. *Expert Rev Hematol,* 7**,** 347-57.

DE CASTRO FONSECA, M., AGUIAR, C. J., DA ROCHA FRANCO, J. A., GINGOLD, R. N. & LEITE, M. F. 2016. GPR91: expanding the frontiers of Krebs cycle intermediates. *Cell Commun Signal,* 14**,** 3.

DE GARIBAY, G. R., DIAZ, A., GAVINA, B., ROMERO, A., GARRE, P., VEGA, A., BLANCO, A., TOSAR, A., DIEZ, O., PEREZ-SEGURA, P., DIAZ-RUBIO, E., CALDES, T. & DE LA HOYA, M. 2013. Low prevalence of SLX4 loss-of-function mutations in non-BRCA1/2 breast and/or ovarian cancer families. *Eur J Hum Genet,* 21**,** 883-6.

DE WINTER, J. P. & JOENJE, H. 2009. The genetic and molecular basis of Fanconi anemia. *Mutat Res,* 668**,** 11-9.

DEANS, A. J. & WEST, S. C. 2009. FANCM connects the genome instability disorders Bloom's Syndrome and Fanconi Anemia. *Mol Cell,* 36**,** 943-53.

DEGOS, L. & WANG, Z. Y. 2001. All trans retinoic acid in acute promyelocytic leukemia. *Oncogene,* 20**,** 7140-5.

DENG, C. X. 2006. BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. *Nucleic Acids Res,* 34**,** 1416-26.

DOHNER, H., ESTEY, E., GRIMWADE, D., AMADORI, S., APPELBAUM, F. R., BUCHNER, T., DOMBRET, H., EBERT, B. L., FENAUX, P., LARSON, R. A., LEVINE, R. L., LO-COCO, F., NAOE, T., NIEDERWIESER, D., OSSENKOPPELE, G. J., SANZ, M., SIERRA, J., TALLMAN, M. S., TIEN, H. F., WEI, A. H., LOWENBERG, B. & BLOOMFIELD, C. D. 2017. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood,* 129**,** 424-447.

DOHNER, H., ESTEY, E. H., AMADORI, S., APPELBAUM, F. R., BUCHNER, T., BURNETT, A. K., DOMBRET, H., FENAUX, P., GRIMWADE, D., LARSON, R. A., LO-COCO, F., NAOE, T., NIEDERWIESER, D., OSSENKOPPELE, G. J., SANZ, M. A., SIERRA, J., TALLMAN, M. S., LOWENBERG, B., BLOOMFIELD, C. D. & EUROPEAN, L. 2010. Diagnosis and management of acute

myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood,* 115**,** 453-74.
DONG, H., NEBERT, D. W., BRUFORD, E. A., THOMPSON, D. C., JOENJE, H. & VASILIOU, V. 2015. Update of the human and mouse Fanconi anemia genes. *Hum Genomics,* 9**,** 32.
DUNCAN, E. L., DANOY, P., KEMP, J. P., LEO, P. J., MCCLOSKEY, E., NICHOLSON, G. C., EASTELL, R., PRINCE, R. L., EISMAN, J. A., JONES, G., SAMBROOK, P. N., REID, I. R., DENNISON, E. M., WARK, J., RICHARDS, J. B., UITTERLINDEN, A. G., SPECTOR, T. D., ESAPA, C., COX, R. D., BROWN, S. D., THAKKER, R. V., ADDISON, K. A., BRADBURY, L. A., CENTER, J. R., COOPER, C., CREMIN, C., ESTRADA, K., FELSENBERG, D., GLUER, C. C., HADLER, J., HENRY, M. J., HOFMAN, A., KOTOWICZ, M. A., MAKOVEY, J., NGUYEN, S. C., NGUYEN, T. V., PASCO, J. A., PRYCE, K., REID, D. M., RIVADENEIRA, F., ROUX, C., STEFANSSON, K., STYRKARSDOTTIR, U., THORLEIFSSON, G., TICHAWANGANA, R., EVANS, D. M. & BROWN, M. A. 2011. Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. *PLoS Genet,* 7**,** e1001372.
DURDIK, M., KOSIK, P., GURSKY, J., VOKALOVA, L., MARKOVA, E. & BELYAEV, I. 2015. Imaging flow cytometry as a sensitive tool to detect low-dose-induced DNA damage by analyzing 53BP1 and gammaH2AX foci in human lymphocytes. *Cytometry A,* 87**,** 1070-8.
ESTEBAN-JURADO, C., FRANCH-EXPOSITO, S., MUNOZ, J., OCANA, T., CARBALLAL, S., LOPEZ-CERON, M., CUATRECASAS, M., VILA-CASADESUS, M., LOZANO, J. J., SERRA, E., BELTRAN, S., BREA-FERNANDEZ, A., RUIZ-PONTE, C., CASTELLS, A., BUJANDA, L., GARRE, P., CALDES, T., CUBIELLA, J., BALAGUER, F. & CASTELLVI-BEL, S. 2016. The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. *Eur J Hum Genet,* 24**,** 1501-5.
ESTEY, E. H. 2001. Therapeutic options for acute myelogenous leukemia. *Cancer,* 92**,** 1059-73.
EVERS, B. & JONKERS, J. 2006. Mouse models of BRCA1 and BRCA2 deficiency: past lessons, current understanding and future prospects. *Oncogene,* 25**,** 5885-97.
FARMER, H., MCCABE, N., LORD, C. J., TUTT, A. N., JOHNSON, D. A., RICHARDSON, T. B., SANTAROSA, M., DILLON, K. J., HICKSON, I., KNIGHTS, C., MARTIN, N. M., JACKSON, S. P., SMITH, G. C. & ASHWORTH, A. 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature,* 434**,** 917-21.
FEILOTTER, H. E., MICHEL, C., UY, P., BATHURST, L. & DAVEY, S. 2014. BRCA1 haploinsufficiency leads to altered expression of genes involved in cellular proliferation and development. *PLoS One,* 9**,** e100068.
FLICK, K. & KAISER, P. 2012. Protein degradation and the stress response. *Semin Cell Dev Biol,* 23**,** 515-22.
GARAYCOECHEA, J. I., CROSSAN, G. P., LANGEVIN, F., DALY, M., ARENDS, M. J. & PATEL, K. J. 2012. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature,* 489**,** 571-5.
GARAYCOECHEA, J. I. & PATEL, K. J. 2014. Why does the bone marrow fail in Fanconi anemia? *Blood,* 123**,** 26-34.
GARCIA, J. S. & STONE, R. M. 2017. The Development of FLT3 Inhibitors in Acute Myeloid Leukemia. *Hematol Oncol Clin North Am,* 31**,** 663-680.
GERMAN, J. 1969. Bloom's syndrome. I. Genetical and clinical observations in the first twenty-seven patients. *Am J Hum Genet,* 21**,** 196-227.
GERMAN, J., SANZ, M. M., CIOCCI, S., YE, T. Z. & ELLIS, N. A. 2007. Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry. *Hum Mutat,* 28**,** 743-53.
GHOSAL, G. & CHEN, J. 2013. DNA damage tolerance: a double-edged sword guarding the genome. *Transl Cancer Res,* 2**,** 107-129.
GILISSEN, J., JOURET, F., PIROTTE, B. & HANSON, J. 2016. Insight into SUCNR1 (GPR91) structure and function. *Pharmacol Ther,* 159**,** 56-65.
GODIN, S. K., SULLIVAN, M. R. & BERNSTEIN, K. A. 2016. Novel insights into RAD51 activity and regulation during homologous recombination and DNA replication. *Biochem Cell Biol***,** 1-12.
GOLDIN, L. R., KRISTINSSON, S. Y., LIANG, X. S., DEROLF, A. R., LANDGREN, O. & BJORKHOLM, M. 2012. Familial aggregation of acute myeloid leukemia and myelodysplastic syndromes. *J Clin Oncol,* 30**,** 179-83.
GOYAMA, S. & KITAMURA, T. 2017. Epigenetics in normal and malignant hematopoiesis: An overview and update 2017. *Cancer Sci,* 108**,** 553-562.
GRANFELDT OSTGARD, L. S., MEDEIROS, B. C., SENGELOV, H., NORGAARD, M., ANDERSEN, M. K., DUFVA, I. H., FRIIS, L. S., KJELDSEN, E., MARCHER, C. W., PREISS, B., SEVERINSEN, M. & NORGAARD, J. M. 2015. Epidemiology and Clinical Significance of Secondary and Therapy-Related Acute Myeloid Leukemia: A National Population-Based Cohort Study. *J Clin Oncol,* 33**,** 3641-9.
GRAVELLS, P., HOH, L., SOLOVIEVA, S., PATIL, A., DUDZIEC, E., RENNIE, I. G., SISLEY, K. & BRYANT, H. E. 2013. Reduced FANCD2 influences spontaneous SCE and RAD51 foci formation in uveal melanoma and Fanconi anaemia. *Oncogene,* 32**,** 5338-46.
GREAVES, M. F., MAIA, A. T., WIEMELS, J. L. & FORD, A. M. 2003. Leukemia in twins: lessons in natural history. *Blood,* 102**,** 2321-33.

GRIMWADE, D., HILLS, R. K., MOORMAN, A. V., WALKER, H., CHATTERS, S., GOLDSTONE, A. H., WHEATLEY, K., HARRISON, C. J., BURNETT, A. K. & NATIONAL CANCER RESEARCH INSTITUTE ADULT LEUKAEMIA WORKING, G. 2010. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood,* 116**,** 354-65.

GRUNERT, M., DORN, C., SCHUELER, M., DUNKEL, I., SCHLESINGER, J., MEBUS, S., ALEXI-MESKISHVILI, V., PERROT, A., WASSILEW, K., TIMMERMANN, B., HETZER, R., BERGER, F. & SPERLING, S. R. 2014. Rare and private variations in neural crest, apoptosis and sarcomere genes define the polygenic background of isolated Tetralogy of Fallot. *Hum Mol Genet,* 23**,** 3115-28.

HAHN, C. N., CHONG, C. E., CARMICHAEL, C. L., WILKINS, E. J., BRAUTIGAN, P. J., LI, X. C., BABIC, M., LIN, M., CARMAGNAC, A., LEE, Y. K., KOK, C. H., GAGLIARDI, L., FRIEND, K. L., EKERT, P. G., BUTCHER, C. M., BROWN, A. L., LEWIS, I. D., TO, L. B., TIMMS, A. E., STOREK, J., MOORE, S., ALTREE, M., ESCHER, R., BARDY, P. G., SUTHERS, G. K., D'ANDREA, R. J., HORWITZ, M. S. & SCOTT, H. S. 2011. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet,* 43**,** 1012-7.

HAN, H. J., MARUYAMA, M., BABA, S., PARK, J. G. & NAKAMURA, Y. 1995. Genomic structure of human mismatch repair gene, hMLH1, and its mutation analysis in patients with hereditary non-polyposis colorectal cancer (HNPCC). *Hum Mol Genet,* 4**,** 237-42.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HARSHA VARDHAN, B. G., MUTHU, M. S., SARASWATHI, K. & KOTEESWARAN, D. 2007. Bird-headed dwarf of Seckel. *J Indian Soc Pedod Prev Dent,* 25 Suppl**,** S8-9.

HART, S. N., ELLINGSON, M. S., SCHAHL, K., VEDELL, P. T., CARLSON, R. E., SINNWELL, J. P., BARMAN, P., SICOTTE, H., ECKEL-PASSOW, J. E., WANG, L., KALARI, K. R., QIN, R., KRUISSELBRINK, T. M., JIMENEZ, R. E., BRYCE, A. H., TAN, W., WEINSHILBOUM, R., WANG, L. & KOHLI, M. 2016. Determining the frequency of pathogenic germline variants from exome sequencing in patients with castrate-resistant prostate cancer. *BMJ Open,* 6**,** e010332.

HATZIMICHAEL, E., GEORGIOU, G., BENETATOS, L. & BRIASOULIS, E. 2013. Gene mutations and molecularly targeted therapies in acute myeloid leukemia. *Am J Blood Res,* 3**,** 29-51.

HAYANI, A., SUAREZ, C. R., MOLNAR, Z., LEBEAU, M. & GODWIN, J. 1994. Acute myeloid leukaemia in a patient with Seckel syndrome. *J Med Genet,* 31**,** 148-9.

HAYNES, B., SAADAT, N., MYUNG, B. & SHEKHAR, M. P. 2015. Crosstalk between translesion synthesis, Fanconi anemia network, and homologous recombination repair pathways in interstrand DNA crosslink repair and development of chemoresistance. *Mutat Res Rev Mutat Res,* 763**,** 258-66.

HEATH, E. M., CHAN, S. M., MINDEN, M. D., MURPHY, T., SHLUSH, L. I. & SCHIMMER, A. D. 2017. Biological and clinical consequences of NPM1 mutations in AML. *Leukemia,* 31**,** 798-807.

HENGEL, S. R., MALACARIA, E., FOLLY DA SILVA CONSTANTINO, L., BAIN, F. E., DIAZ, A., KOCH, B. G., YU, L., WU, M., PICHIERRI, P., SPIES, M. A. & SPIES, M. 2016. Small-molecule inhibitors identify the RAD52-ssDNA interaction as critical for recovery from replication stress and for survival of BRCA2 deficient cells. *Elife,* 5.

HESS, C. J., AMEZIANE, N., SCHUURHUIS, G. J., ERRAMI, A., DENKERS, F., KASPERS, G. J., CLOOS, J., JOENJE, H., REINHARDT, D., OSSENKOPPELE, G. J., ZWAAN, C. M. & WAISFISZ, Q. 2008. Hypermethylation of the FANCC and FANCL promoter regions in sporadic acute leukaemia. *Cell Oncol,* 30**,** 299-306.

HIRA, A., YOSHIDA, K., SATO, K., OKUNO, Y., SHIRAISHI, Y., CHIBA, K., TANAKA, H., MIYANO, S., SHIMAMOTO, A., TAHARA, H., ITO, E., KOJIMA, S., KURUMIZAKA, H., OGAWA, S., TAKATA, M., YABE, H. & YABE, M. 2015. Mutations in the gene encoding the E2 conjugating enzyme UBE2T cause Fanconi anemia. *Am J Hum Genet,* 96**,** 1001-7.

HIRSCH, B., SHIMAMURA, A., MOREAU, L., BALDINGER, S., HAG-ALSHIEKH, M., BOSTROM, B., SENCER, S. & D'ANDREA, A. D. 2004. Association of biallelic BRCA2/FANCD1 mutations with spontaneous chromosomal instability and solid tumors of childhood. *Blood,* 103**,** 2554-9.

HO, M. S., MEDCALF, R. L., LIVESEY, S. A. & TRAIANEDES, K. 2015. The dynamics of adult haematopoiesis in the bone and bone marrow environment. *Br J Haematol,* 170**,** 472-86.

HODSON, C. & WALDEN, H. 2012. Towards a molecular understanding of the fanconi anemia core complex. *Anemia,* 2012**,** 926787.

HOWLETT, N. G., TANIGUCHI, T., OLSON, S., COX, B., WAISFISZ, Q., DE DIE-SMULDERS, C., PERSKY, N., GROMPE, M., JOENJE, H., PALS, G., IKEDA, H., FOX, E. A. & D'ANDREA, A. D. 2002. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science,* 297**,** 606-9.

ILAGAN, J. O., RAMAKRISHNAN, A., HAYES, B., MURPHY, M. E., ZEBARI, A. S., BRADLEY, P. & BRADLEY, R. K. 2015. U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res,* 25**,** 14-26.

INOUE, D., BRADLEY, R. K. & ABDEL-WAHAB, O. 2016a. Spliceosomal gene mutations in myelodysplasia: molecular links to clonal abnormalities of hematopoiesis. *Genes Dev,* 30**,** 989-1001.

INOUE, S., LEMONNIER, F. & MAK, T. W. 2016b. Roles of IDH1/2 and TET2 mutations in myeloid disorders. *Int J Hematol,* 103**,** 627-33.
IRMINGER-FINGER, I. & JEFFORD, C. E. 2006. Is there more to BARD1 than BRCA1? *Nat Rev Cancer,* 6**,** 382-91.
JEGGO, P. A., PEARL, L. H. & CARR, A. M. 2016. DNA repair, genome stability and cancer: a historical perspective. *Nat Rev Cancer,* 16**,** 35-42.
JENG, Y. M., CAI-NG, S., LI, A., FURUTA, S., CHEW, H., CHEN, P. L., LEE, E. Y. & LEE, W. H. 2007. Brca1 heterozygous mice have shortened life span and are prone to ovarian tumorigenesis with haploinsufficiency upon ionizing irradiation. *Oncogene,* 26**,** 6160-6.
JONGMANS, M. C., KUIPER, R. P., CARMICHAEL, C. L., WILKINS, E. J., DORS, N., CARMAGNAC, A., SCHOUTEN-VAN MEETEREN, A. Y., LI, X., STANKOVIC, M., KAMPING, E., BENGTSSON, H., SCHOENMAKERS, E. F., VAN KESSEL, A. G., HOOGERBRUGGE, P. M., HAHN, C. N., BRONS, P. P., SCOTT, H. S. & HOOGERBRUGGE, N. 2010. Novel RUNX1 mutations in familial platelet disorder with enhanced risk for acute myeloid leukemia: clues for improved identification of the FPD/AML syndrome. *Leukemia,* 24**,** 242-6.
JOO, W., XU, G., PERSKY, N. S., SMOGORZEWSKA, A., RUDGE, D. G., BUZOVETSKY, O., ELLEDGE, S. J. & PAVLETICH, N. P. 2011. Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway. *Science,* 333**,** 312-6.
JUN, S. H., KIM, T. G. & BAN, C. 2006. DNA mismatch repair system. Classical and fresh roles. *FEBS J,* 273**,** 1609-19.
KALB, R., NEVELING, K., HOEHN, H., SCHNEIDER, H., LINKA, Y., BATISH, S. D., HUNT, C., BERWICK, M., CALLEN, E., SURRALLES, J., CASADO, J. A., BUEREN, J., DASI, A., SOULIER, J., GLUCKMAN, E., ZWAAN, C. M., VAN SPAENDONK, R., PALS, G., DE WINTER, J. P., JOENJE, H., GROMPE, M., AUERBACH, A. D., HANENBERG, H. & SCHINDLER, D. 2007. Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. *Am J Hum Genet,* 80**,** 895-910.
KANDOTH, C., MCLELLAN, M. D., VANDIN, F., YE, K., NIU, B., LU, C., XIE, M., ZHANG, Q., MCMICHAEL, J. F., WYCZALKOWSKI, M. A., LEISERSON, M. D. M., MILLER, C. A., WELCH, J. S., WALTER, M. J., WENDL, M. C., LEY, T. J., WILSON, R. K., RAPHAEL, B. J. & DING, L. 2013. Mutational landscape and significance across 12 major cancer types. *Nature,* 502**,** 333-339.
KASCHUTNIG, P., BOGESKA, R., WALTER, D., LIER, A., HUNTSCHA, S. & MILSOM, M. D. 2015. The Fanconi anemia pathway is required for efficient repair of stress-induced DNA damage in haematopoietic stem cells. *Cell Cycle,* 14**,** 2734-42.
KELLY, L. M. & GILLILAND, D. G. 2002. Genetics of myeloid leukemias. *Annu Rev Genomics Hum Genet,* 3**,** 179-98.
KENNEDY, R. D., CHEN, C. C., STUCKERT, P., ARCHILA, E. M., DE LA VEGA, M. A., MOREAU, L. A., SHIMAMURA, A. & D'ANDREA, A. D. 2007. Fanconi anemia pathway-deficient tumor cells are hypersensitive to inhibition of ataxia telangiectasia mutated. *J Clin Invest,* 117**,** 1440-9.
KENYON, J. & GERSON, S. L. 2007. The role of DNA damage repair in aging of adult stem cells. *Nucleic Acids Res,* 35**,** 7557-65.
KIEL, M. J., RADICE, G. L. & MORRISON, S. J. 2007. Lack of evidence that hematopoietic stem cells depend on N-cadherin-mediated adhesion to osteoblasts for their maintenance. *Cell Stem Cell,* 1**,** 204-17.
KIISKI, J. I., PELTTARI, L. M., KHAN, S., FREYSTEINSDOTTIR, E. S., REYNISDOTTIR, I., HART, S. N., SHIMELIS, H., VILSKE, S., KALLIONIEMI, A., SCHLEUTKER, J., LEMINEN, A., BUTZOW, R., BLOMQVIST, C., BARKARDOTTIR, R. B., COUCH, F. J., AITTOMAKI, K. & NEVANLINNA, H. 2014. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci U S A,* 111**,** 15172-7.
KIM, H. & D'ANDREA, A. D. 2012. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes Dev,* 26**,** 1393-408.
KIM, J. C., ROH, S. A., YOON, Y. S., KIM, H. C. & PARK, I. J. 2007. MLH3 and EXO1 alterations in familial colorectal cancer patients not fulfilling Amsterdam criteria. *Cancer Genet Cytogenet,* 176**,** 172-4.
KIMCHI-SARFATY, C., MARPLE, A. H., SHINAR, S., KIMCHI, A. M., SCAVO, D., ROMA, M. I., KIM, I. W., JONES, A., ARORA, M., GRIBAR, J., GURWITZ, D. & GOTTESMAN, M. M. 2007. Ethnicity-related polymorphisms and haplotypes in the human ABCB1 gene. *Pharmacogenomics,* 8**,** 29-39.
KIRCHER, M., WITTEN, D. M., JAIN, P., O'ROAK, B. J., COOPER, G. M. & SHENDURE, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet,* 46**,** 310-5.
KITANO, K. 2014. Structural mechanisms of human RecQ helicases WRN and BLM. *Front Genet,* 5**,** 366.
KIYOI, H., YAMAJI, S., KOJIMA, S. & NAOE, T. 2007. JAK3 mutations occur in acute megakaryoblastic leukemia both in Down syndrome children and non-Down syndrome adults. *Leukemia,* 21**,** 574-6.
KLEIMAN, N. J., DAVID, J., ELLISTON, C. D., HOPKINS, K. M., SMILENOV, L. B., BRENNER, D. J., WORGUL, B. V., HALL, E. J. & LIEBERMAN, H. B. 2007. Mrad9 and atm haploinsufficiency enhance spontaneous and X-ray-induced cataractogenesis in mice. *Radiat Res,* 168**,** 567-73.

KLEIN DOUWEL, D., BOONEN, R. A., LONG, D. T., SZYPOWSKA, A. A., RASCHLE, M., WALTER, J. C. & KNIPSCHEER, P. 2014. XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol Cell,* 54**,** 460-71.

KNIGHT, J. A., SKOL, A. D., SHINDE, A., HASTINGS, D., WALGREN, R. A., SHAO, J., TENNANT, T. R., BANERJEE, M., ALLAN, J. M., LE BEAU, M. M., LARSON, R. A., GRAUBERT, T. A., COX, N. J. & ONEL, K. 2009. Genome-wide association study to identify novel loci associated with therapy-related myeloid leukemia susceptibility. *Blood,* 113**,** 5575-82.

KONDO, M., WEISSMAN, I. L. & AKASHI, K. 1997. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell,* 91**,** 661-72.

KOTTEMANN, M. C. & SMOGORZEWSKA, A. 2013. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature,* 493**,** 356-63.

KRATZ, K., SCHOPF, B., KADEN, S., SENDOEL, A., EBERHARD, R., LADEMANN, C., CANNAVO, E., SARTORI, A. A., HENGARTNER, M. O. & JIRICNY, J. 2010. Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell,* 142**,** 77-88.

KRUGER, L., DEMUTH, I., NEITZEL, H., VARON, R., SPERLING, K., CHRZANOWSKA, K. H., SEEMANOVA, E. & DIGWEED, M. 2007. Cancer incidence in Nijmegen breakage syndrome is modulated by the amount of a variant NBS protein. *Carcinogenesis,* 28**,** 107-11.

KUMAR, A., BHATKAR, D., JAHAGIRDAR, D. & SHARMA, N. K. 2017. Non-homologous End Joining Inhibitor SCR-7 to Exacerbate Low-dose Doxorubicin Cytotoxicity in HeLa Cells. *J Cancer Prev,* 22**,** 47-54.

KUTLER, D. I. 2003. A20-year perspective on the International Fanconi Anemia Registry (IFAR).

LACHAUD, C., MORENO, A., MARCHESI, F., TOTH, R., BLOW, J. J. & ROUSE, J. 2016. Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to prevent genome instability. *Science,* 351**,** 846-9.

LADUCA, H., STUENKEL, A. J., DOLINSKY, J. S., KEILES, S., TANDY, S., PESARAN, T., CHEN, E., GAU, C. L., PALMAER, E., SHOAEPOUR, K., SHAH, D., SPEARE, V., GANDOMI, S. & CHAO, E. 2014. Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet Med,* 16**,** 830-7.

LAM, K. & ZHANG, D. E. 2012. RUNX1 and RUNX1-ETO: roles in hematopoiesis and leukemogenesis. *Front Biosci (Landmark Ed),* 17**,** 1120-39.

LAMARCHE, B. J., ORAZIO, N. I. & WEITZMAN, M. D. 2010. The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett,* 584**,** 3682-95.

LANSDORP, P. M., SUTHERLAND, H. J. & EAVES, C. J. 1990. Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow. *J Exp Med,* 172**,** 363-6.

LEE, J. & DUNPHY, W. G. 2013. The Mre11-Rad50-Nbs1 (MRN) complex has a specific role in the activation of Chk1 in response to stalled replication forks. *Mol Biol Cell,* 24**,** 1343-53.

LEE, J. H. & PAULL, T. T. 2005. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science,* 308**,** 551-4.

LEE, S., EMOND, M. J., BAMSHAD, M. J., BARNES, K. C., RIEDER, M. J., NICKERSON, D. A., TEAM, N. G. E. S. P.-E. L. P., CHRISTIANI, D. C., WURFEL, M. M. & LIN, X. 2012. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet,* 91**,** 224-37.

LEEKE, B., MARSMAN, J., O'SULLIVAN, J. M. & HORSFIELD, J. A. 2014. Cohesin mutations in myeloid malignancies: underlying mechanisms. *Exp Hematol Oncol,* 3**,** 13.

LEI, X., ZHU, Y., TOMKINSON, A. & SUN, L. 2004. Measurement of DNA mismatch repair activity in live cells. *Nucleic Acids Res,* 32**,** e100.

LEVESQUE, J. P. & WINKLER, I. G. 2011. Hierarchy of immature hematopoietic cells related to blood flow and niche. *Curr Opin Hematol,* 18**,** 220-5.

LEWINSOHN, M., BROWN, A. L., WEINEL, L. M., PHUNG, C., RAFIDI, G., LEE, M. K., SCHREIBER, A. W., FENG, J., BABIC, M., CHONG, C. E., LEE, Y., YONG, A., SUTHERS, G. K., POPLAWSKI, N., ALTREE, M., PHILLIPS, K., JAENSCH, L., FINE, M., D'ANDREA, R. J., LEWIS, I. D., MEDEIROS, B. C., POLLYEA, D. A., KING, M. C., WALSH, T., KEEL, S., SHIMAMURA, A., GODLEY, L. A., HAHN, C. N., CHURPEK, J. E. & SCOTT, H. S. 2016. Novel germ line DDX41 mutations define families with a lower age of MDS/AML onset and lymphoid malignancies. *Blood,* 127**,** 1017-23.

LEY, T. J., DING, L., WALTER, M. J., MCLELLAN, M. D., LAMPRECHT, T., LARSON, D. E., KANDOTH, C., PAYTON, J. E., BATY, J., WELCH, J., HARRIS, C. C., LICHTI, C. F., TOWNSEND, R. R., FULTON, R. S., DOOLING, D. J., KOBOLDT, D. C., SCHMIDT, H., ZHANG, Q., OSBORNE, J. R., LIN, L., O'LAUGHLIN, M., MCMICHAEL, J. F., DELEHAUNTY, K. D., MCGRATH, S. D., FULTON, L. A., MAGRINI, V. J., VICKERY, T. L., HUNDAL, J., COOK, L. L., CONYERS, J. J., SWIFT, G. W., REED, J. P., ALLDREDGE, P. A., WYLIE, T., WALKER, J., KALICKI, J., WATSON, M. A., HEATH, S., SHANNON, W. D., VARGHESE, N., NAGARAJAN, R., WESTERVELT, P., TOMASSON, M. H., LINK, D. C., GRAUBERT, T. A., DIPERSIO, J. F., MARDIS, E. R. & WILSON, R. K. 2010. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med,* 363**,** 2424-33.

LEY, T. J., MARDIS, E. R., DING, L., FULTON, B., MCLELLAN, M. D., CHEN, K., DOOLING, D., DUNFORD-SHORE, B. H., MCGRATH, S., HICKENBOTHAM, M., COOK, L., ABBOTT, R., LARSON, D. E., KOBOLDT, D. C., POHL, C., SMITH, S., HAWKINS, A., ABBOTT, S., LOCKE, D., HILLIER, L. W., MINER, T., FULTON, L., MAGRINI, V., WYLIE, T., GLASSCOCK, J., CONYERS, J., SANDER, N., SHI, X., OSBORNE, J. R., MINX, P., GORDON, D., CHINWALLA, A., ZHAO, Y., RIES, R. E., PAYTON, J. E., WESTERVELT, P., TOMASSON, M. H., WATSON, M., BATY, J., IVANOVICH, J., HEATH, S., SHANNON, W. D., NAGARAJAN, R., WALTER, M. J., LINK, D. C., GRAUBERT, T. A., DIPERSIO, J. F. & WILSON, R. K. 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature,* 456**,** 66-72.

LHOTA, F., ZEMANKOVA, P., KLEIBLOVA, P., SOUKUPOVA, J., VOCKA, M., STRANECKY, V., JANATOVA, M., HARTMANNOVA, H., HODANOVA, K., KMOCH, S. & KLEIBL, Z. 2016. Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. *Clin Genet,* 90**,** 324-33.

LI, X. & HEYER, W. D. 2009. RAD54 controls access to the invading 3'-OH end after RAD51-mediated DNA strand invasion in homologous recombination in Saccharomyces cerevisiae. *Nucleic Acids Res,* 37**,** 638-46.

LI, Z. & LI, L. 2006. Understanding hematopoietic stem-cell microenvironments. *Trends Biochem Sci,* 31**,** 589-95.

LIM, R. M., SILVER, A. J., SILVER, M. J., BORROTO, C., SPURRIER, B., PETROSSIAN, T. C., LARSON, J. L. & SILVER, L. M. 2016. Targeted mutation screening panels expose systematic population bias in detection of cystic fibrosis risk. *Genet Med,* 18**,** 174-9.

LINK, K. A., CHOU, F. S. & MULLOY, J. C. 2010. Core binding factor at the crossroads: determining the fate of the HSC. *J Cell Physiol,* 222**,** 50-6.

LIU, H. X., ZHOU, X. L., LIU, T., WERELIUS, B., LINDMARK, G., DAHL, N. & LINDBLOM, A. 2003. The role of hMLH3 in familial colorectal cancer. *Cancer Res,* 63**,** 1894-9.

LIU, J., DOTY, T., GIBSON, B. & HEYER, W. D. 2010. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. *Nat Struct Mol Biol,* 17**,** 1260-2.

LIU, T. & HUANG, J. 2016. DNA End Resection: Facts and Mechanisms. *Genomics Proteomics Bioinformatics,* 14**,** 126-30.

LOK, B. H., CARLEY, A. C., TCHANG, B. & POWELL, S. N. 2013. RAD52 inactivation is synthetically lethal with deficiencies in BRCA1 and PALB2 in addition to BRCA2 through RAD51-mediated homologous recombination. *Oncogene,* 32**,** 3552-8.

LONGERICH, S., KWON, Y., TSAI, M. S., HLAING, A. S., KUPFER, G. M. & SUNG, P. 2014. Regulation of FANCD2 and FANCI monoubiquitination by their interaction and by DNA. *Nucleic Acids Res,* 42**,** 5657-70.

LORD, C. J. & ASHWORTH, A. 2016. BRCAness revisited. *Nat Rev Cancer,* 16**,** 110-20.

LORD, C. J. & ASHWORTH, A. 2017. PARP inhibitors: Synthetic lethality in the clinic. *Science,* 355**,** 1152-1158.

LU, C., XIE, M., WENDL, M. C., WANG, J., MCLELLAN, M. D., LEISERSON, M. D., HUANG, K. L., WYCZALKOWSKI, M. A., JAYASINGHE, R., BANERJEE, T., NING, J., TRIPATHI, P., ZHANG, Q., NIU, B., YE, K., SCHMIDT, H. K., FULTON, R. S., MCMICHAEL, J. F., BATRA, P., KANDOTH, C., BHARADWAJ, M., KOBOLDT, D. C., MILLER, C. A., KANCHI, K. L., ELDRED, J. M., LARSON, D. E., WELCH, J. S., YOU, M., OZENBERGER, B. A., GOVINDAN, R., WALTER, M. J., ELLIS, M. J., MARDIS, E. R., GRAUBERT, T. A., DIPERSIO, J. F., LEY, T. J., WILSON, R. K., GOODFELLOW, P. J., RAPHAEL, B. J., CHEN, F., JOHNSON, K. J., PARVIN, J. D. & DING, L. 2015. Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun,* 6**,** 10086.

LU, S., SHEN, K., WANG, Y., SANTNER, S. J., CHEN, J., BROOKS, S. C. & WANG, Y. A. 2006. Atm-haploinsufficiency enhances susceptibility to carcinogen-induced mammary tumors. *Carcinogenesis,* 27**,** 848-55.

MACHIDA, Y. J., MACHIDA, Y., CHEN, Y., GURTAN, A. M., KUPFER, G. M., D'ANDREA, A. D. & DUTTA, A. 2006. UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation. *Mol Cell,* 23**,** 589-96.

MANGOLD, E., PAGENSTECHER, C., FRIEDL, W., MATHIAK, M., BUETTNER, R., ENGEL, C., LOEFFLER, M., HOLINSKI-FEDER, E., MULLER-KOCH, Y., KELLER, G., SCHACKERT, H. K., KRUGER, S., GOECKE, T., MOESLEIN, G., KLOOR, M., GEBERT, J., KUNSTMANN, E., SCHULMANN, K., RUSCHOFF, J. & PROPPING, P. 2005. Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer. *Int J Cancer,* 116**,** 692-702.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R., CHAKRAVARTI, A., CHO, J. H., GUTTMACHER, A. E., KONG, A., KRUGLYAK, L., MARDIS, E., ROTIMI, C. N., SLATKIN, M., VALLE, D., WHITTEMORE, A. S., BOEHNKE, M., CLARK, A. G., EICHLER, E. E., GIBSON, G.,

HAINES, J. L., MACKAY, T. F., MCCARROLL, S. A. & VISSCHER, P. M. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-53.
MARCUCCI, G., HAFERLACH, T. & DOHNER, H. 2011. Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications. *J Clin Oncol,* 29**,** 475-86.
MARDIS, E. R., DING, L., DOOLING, D. J., LARSON, D. E., MCLELLAN, M. D., CHEN, K., KOBOLDT, D. C., FULTON, R. S., DELEHAUNTY, K. D., MCGRATH, S. D., FULTON, L. A., LOCKE, D. P., MAGRINI, V. J., ABBOTT, R. M., VICKERY, T. L., REED, J. S., ROBINSON, J. S., WYLIE, T., SMITH, S. M., CARMICHAEL, L., ELDRED, J. M., HARRIS, C. C., WALKER, J., PECK, J. B., DU, F. Y., DUKES, A. F., SANDERSON, G. E., BRUMMETT, A. M., CLARK, E., MCMICHAEL, J. F., MEYER, R. J., SCHINDLER, J. K., POHL, C. S., WALLIS, J. W., SHI, X. Q., LIN, L., SCHMIDT, H., TANG, Y. Z., HAIPEK, C., WIECHERT, M. E., IVY, J. V., KALICKI, J., ELLIOTT, G., RIES, R. E., PAYTON, J. E., WESTERVELT, P., TOMASSON, M. H., WATSON, M. A., BATY, J., HEATH, S., SHANNON, W. D., NAGARAJAN, R., LINK, D. C., WALTER, M. J., GRAUBERT, T. A., DIPERSIO, J. F., WILSON, R. K. & LEY, T. J. 2009. Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome. *New England Journal of Medicine,* 361**,** 1058-1066.
MARECHAL, A. & ZOU, L. 2013. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb Perspect Biol,* 5.
MATOS, J. & WEST, S. C. 2014. Holliday junction resolution: regulation in space and time. *DNA Repair (Amst),* 19**,** 176-81.
MATSUOKA, S., BALLIF, B. A., SMOGORZEWSKA, A., MCDONALD, E. R., 3RD, HUROV, K. E., LUO, J., BAKALARSKI, C. E., ZHAO, Z., SOLIMINI, N., LERENTHAL, Y., SHILOH, Y., GYGI, S. P. & ELLEDGE, S. J. 2007. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science,* 316**,** 1160-6.
MATULONIS, U. A. & MONK, B. J. 2017. PARP inhibitor and chemotherapy combination trials for the treatment of advanced malignancies: does a development pathway forward exist? *Ann Oncol,* 28**,** 443-447.
MAZIN, A. V., ALEXEEV, A. A. & KOWALCZYKOWSKI, S. C. 2003. A novel function of Rad54 protein. Stabilization of the Rad51 nucleoprotein filament. *J Biol Chem,* 278**,** 14029-36.
MCDANIEL, L. D. & SCHULTZ, R. A. 1992. Elevated sister chromatid exchange phenotype of Bloom syndrome cells is complemented by human chromosome 15. *Proc Natl Acad Sci U S A,* 89**,** 7968-72.
MEINDL, A., HELLEBRAND, H., WIEK, C., ERVEN, V., WAPPENSCHMIDT, B., NIEDERACHER, D., FREUND, M., LICHTNER, P., HARTMANN, L., SCHAAL, H., RAMSER, J., HONISCH, E., KUBISCH, C., WICHMANN, H. E., KAST, K., DEISSLER, H., ENGEL, C., MULLER-MYHSOK, B., NEVELING, K., KIECHLE, M., MATHEW, C. G., SCHINDLER, D., SCHMUTZLER, R. K. & HANENBERG, H. 2010. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet,* 42**,** 410-4.
MEYER, C., KOWARZ, E., HOFMANN, J., RENNEVILLE, A., ZUNA, J., TRKA, J., BEN ABDELALI, R., MACINTYRE, E., DE BRAEKELEER, E., DE BRAEKELEER, M., DELABESSE, E., DE OLIVEIRA, M. P., CAVE, H., CLAPPIER, E., VAN DONGEN, J. J., BALGOBIND, B. V., VAN DEN HEUVEL-EIBRINK, M. M., BEVERLOO, H. B., PANZER-GRUMAYER, R., TEIGLER-SCHLEGEL, A., HARBOTT, J., KJELDSEN, E., SCHNITTGER, S., KOEHL, U., GRUHN, B., HEIDENREICH, O., CHAN, L. C., YIP, S. F., KRZYWINSKI, M., ECKERT, C., MORICKE, A., SCHRAPPE, M., ALONSO, C. N., SCHAFER, B. W., KRAUTER, J., LEE, D. A., ZUR STADT, U., TE KRONNIE, G., SUTTON, R., IZRAELI, S., TRAKHTENBROT, L., LO NIGRO, L., TSAUR, G., FECHINA, L., SZCZEPANSKI, T., STREHL, S., ILENCIKOVA, D., MOLKENTIN, M., BURMEISTER, T., DINGERMANN, T., KLINGEBIEL, T. & MARSCHALEK, R. 2009. New insights to the MLL recombinome of acute leukemias. *Leukemia,* 23**,** 1490-9.
MEYERS, C. A., ALBITAR, M. & ESTEY, E. 2005. Cognitive impairment, fatigue, and cytokine levels in patients with acute myelogenous leukemia or myelodysplastic syndrome. *Cancer,* 104**,** 788-93.
MILNE, T. A. 2017. Mouse models of MLL leukemia: recapitulating the human disease. *Blood,* 129**,** 2217-2223.
MIMS, A. & STUART, R. K. 2013. Developmental therapeutics in acute myelogenous leukemia: are there any new effective cytotoxic chemotherapeutic agents out there? *Curr Hematol Malig Rep,* 8**,** 156-62.
MORRISON, S. J., UCHIDA, N. & WEISSMAN, I. L. 1995. The biology of hematopoietic stem cells. *Annu Rev Cell Dev Biol,* 11**,** 35-71.
MOSOR, M., ZIOLKOWSKA, I., PERNAK-SCHWARZ, M., JANUSZKIEWICZ-LEWANDOWSKA, D. & NOWAK, J. 2006. Association of the heterozygous germline I171V mutation of the NBS1 gene with childhood acute lymphoblastic leukemia. *Leukemia,* 20**,** 1454-6.
MROZEK, K. 2008. Cytogenetic, molecular genetic, and clinical characteristics of acute myeloid leukemia with a complex karyotype. *Semin Oncol,* 35**,** 365-77.
MURAI, J., HUANG, S. Y., DAS, B. B., RENAUD, A., ZHANG, Y., DOROSHOW, J. H., JI, J., TAKEDA, S. & POMMIER, Y. 2012. Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer Res,* 72**,** 5588-99.
MURATA, S., ZHANG, C., FINCH, N., ZHANG, K., CAMPO, L. & BREUER, E. K. 2016. Predictors and Modulators of Synthetic Lethality: An Update on PARP Inhibitors and Personalized Medicine. *Biomed Res Int,* 2016**,** 2346585.

MURATI, A., BRECQUEVILLE, M., DEVILLIER, R., MOZZICONACCI, M. J., GELSI-BOYER, V. & BIRNBAUM, D. 2012. Myeloid malignancies: mutations, models and management. *BMC Cancer,* 12**,** 304.

MURER-ORLANDO, M., LLERENA, J. C., JR., BIRJANDI, F., GIBSON, R. A. & MATHEW, C. G. 1993. FACC gene mutations and early prenatal diagnosis of Fanconi's anaemia. *Lancet,* 342**,** 686.

MURGA, M., BUNTING, S., MONTANA, M. F., SORIA, R., MULERO, F., CANAMERO, M., LEE, Y., MCKINNON, P. J., NUSSENZWEIG, A. & FERNANDEZ-CAPETILLO, O. 2009. A mouse model of ATR-Seckel shows embryonic replicative stress and accelerated aging. *Nat Genet,* 41**,** 891-8.

NAIM, V. & ROSSELLI, F. 2009. The FANC pathway and mitosis: a replication legacy. *Cell Cycle,* 8**,** 2907-11.

NAKANISHI, K., MORAN, A., HAYS, T., KUANG, Y., FOX, E., GARNEAU, D., MONTES DE OCA, R., GROMPE, M. & D'ANDREA, A. D. 2001. Functional analysis of patient-derived mutations in the Fanconi anemia gene, FANCG/XRCC9. *Exp Hematol,* 29**,** 842-9.

NAKAO, M., YOKOTA, S., IWAI, T., KANEKO, H., HORIIKE, S., KASHIMA, K., SONODA, Y., FUJIMOTO, T. & MISAWA, S. 1996. Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia,* 10**,** 1911-8.

NEAME, P. B., SOAMBOONSRUP, P., BROWMAN, G. P., MEYER, R. M., BENGER, A., WILSON, W. E., WALKER, I. R., SAEED, N. & MCBRIDE, J. A. 1986. Classifying acute leukemia by immunophenotyping: a combined FAB-immunologic classification of AML. *Blood,* 68**,** 1355-62.

NEBGEN, D. R., RHODES, H. E., HARTMAN, C., MUNSELL, M. F. & LU, K. H. 2016. Abnormal Uterine Bleeding as the Presenting Symptom of Hematologic Cancer. *Obstet Gynecol,* 128**,** 357-63.

NICCHIA, E., GRECO, C., DE ROCCO, D., PECILE, V., D'EUSTACCHIO, A., CAPPELLI, E., CORTI, P., MARRA, N., RAMENGHI, U., PILLON, M., FARRUGGIA, P., DUFOUR, C., PALLAVICINI, A., TORELLI, L. & SAVOIA, A. 2015. Identification of point mutations and large intragenic deletions in Fanconi anemia using next-generation sequencing technology. *Mol Genet Genomic Med,* 3**,** 500-12.

NIELSEN, F. C., VAN OVEREEM HANSEN, T. & SORENSEN, C. S. 2016. Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat Rev Cancer,* 16**,** 599-612.

NOTTA, F., DOULATOV, S., LAURENTI, E., POEPPL, A., JURISICA, I. & DICK, J. E. 2011. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science,* 333**,** 218-21.

O'DRISCOLL, M. 2008. Haploinsufficiency of DNA Damage Response Genes and their Potential Influence in Human Genomic Disorders. *Curr Genomics,* 9**,** 137-46.

O'DRISCOLL, M., RUIZ-PEREZ, V. L., WOODS, C. G., JEGGO, P. A. & GOODSHIP, J. A. 2003. A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. *Nat Genet,* 33**,** 497-501.

OKADA, S., NAKAUCHI, H., NAGAYOSHI, K., NISHIKAWA, S., MIURA, Y. & SUDA, T. 1992. In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. *Blood,* 80**,** 3044-50.

OWEN, C. J., TOZE, C. L., KOOCHIN, A., FORREST, D. L., SMITH, C. A., STEVENS, J. M., JACKSON, S. C., POON, M. C., SINCLAIR, G. D., LEBER, B., JOHNSON, P. R., MACHETA, A., YIN, J. A., BARNETT, M. J., LISTER, T. A. & FITZGIBBON, J. 2008. Five new pedigrees with inherited RUNX1 mutations causing familial platelet disorder with propensity to myeloid malignancy. *Blood,* 112**,** 4639-45.

PALOVCAK, A., LIU, W., YUAN, F. & ZHANG, Y. 2017. Maintenance of genome stability by Fanconi anemia proteins. *Cell Biosci,* 7**,** 8.

PAPAEMMANUIL, E., GERSTUNG, M., BULLINGER, L., GAIDZIK, V. I., PASCHKA, P., ROBERTS, N. D., POTTER, N. E., HEUSER, M., THOL, F., BOLLI, N., GUNDEM, G., VAN LOO, P., MARTINCORENA, I., GANLY, P., MUDIE, L., MCLAREN, S., O'MEARA, S., RAINE, K., JONES, D. R., TEAGUE, J. W., BUTLER, A. P., GREAVES, M. F., GANSER, A., DOHNER, K., SCHLENK, R. F., DOHNER, H. & CAMPBELL, P. J. 2016. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med,* 374**,** 2209-21.

PARK, J. Y., VIRTS, E. L., JANKOWSKA, A., WIEK, C., OTHMAN, M., CHAKRABORTY, S. C., VANCE, G. H., ALKURAYA, F. S., HANENBERG, H. & ANDREASSEN, P. R. 2016. Complementation of hypersensitivity to DNA interstrand crosslinking agents demonstrates that XRCC2 is a Fanconi anaemia gene. *J Med Genet*.

PARMAR, K. & D'ANDREA, A. D. 2017. Aldehydes Pose a Threat to BRCA2 Mutation Carriers. *Cell,* 169**,** 979-981.

PATEL, A. G., SARKARIA, J. N. & KAUFMANN, S. H. 2011. Nonhomologous end joining drives poly(ADP-ribose) polymerase (PARP) inhibitor lethality in homologous recombination-deficient cells. *Proc Natl Acad Sci U S A,* 108**,** 3406-11.

PATEL, J. P., GONEN, M., FIGUEROA, M. E., FERNANDEZ, H., SUN, Z., RACEVSKIS, J., VAN VLIERBERGHE, P., DOLGALEV, I., THOMAS, S., AMINOVA, O., HUBERMAN, K., CHENG, J., VIALE, A., SOCCI, N. D., HEGUY, A., CHERRY, A., VANCE, G., HIGGINS, R. R., KETTERLING, R. P., GALLAGHER, R. E., LITZOW, M., VAN DEN BRINK, M. R., LAZARUS, H. M., ROWE, J. M., LUGER, S., FERRANDO, A., PAIETTA, E., TALLMAN, M. S., MELNICK, A., ABDEL-WAHAB,

O. & LEVINE, R. L. 2012. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med,* 366**,** 1079-89.

PATHANIA, S., BADE, S., LE GUILLOU, M., BURKE, K., REED, R., BOWMAN-COLIN, C., SU, Y., TING, D. T., POLYAK, K., RICHARDSON, A. L., FEUNTEUN, J., GARBER, J. E. & LIVINGSTON, D. M. 2014. BRCA1 haploinsufficiency for replication stress suppression in primary cells. *Nat Commun,* 5**,** 5496.

PAZ-PRIEL, I. & FRIEDMAN, A. 2011. C/EBPalpha dysregulation in AML and ALL. *Crit Rev Oncog,* 16**,** 93-102.

PEARL, L. H., SCHIERZ, A. C., WARD, S. E., AL-LAZIKANI, B. & PEARL, F. M. 2015. Therapeutic opportunities within the DNA damage response. *Nat Rev Cancer,* 15**,** 166-80.

PETRUCELLI, N., DALY, M. B. & FELDMAN, G. L. 2010. Hereditary breast and ovarian cancer due to mutations in BRCA1 and BRCA2. *Genet Med,* 12**,** 245-59.

PICHIERRI, P., FRANCHITTO, A. & ROSSELLI, F. 2004. BLM and the FANC proteins collaborate in a common pathway in response to stalled replication forks. *EMBO J,* 23**,** 3154-63.

PIZZOLATO, J., MUKHERJEE, S., SCHARER, O. D. & JIRICNY, J. 2015. FANCD2-associated nuclease 1, but not exonuclease 1 or flap endonuclease 1, is able to unhook DNA interstrand cross-links in vitro. *J Biol Chem,* 290**,** 22602-11.

POLPRASERT, C., SCHULZE, I., SEKERES, M. A., MAKISHIMA, H., PRZYCHODZEN, B., HOSONO, N., SINGH, J., PADGETT, R. A., GU, X., PHILLIPS, J. G., CLEMENTE, M., PARKER, Y., LINDNER, D., DIENES, B., JANKOWSKY, E., SAUNTHARARAJAH, Y., DU, Y., OAKLEY, K., NGUYEN, N., MUKHERJEE, S., PABST, C., GODLEY, L. A., CHURPEK, J. E., POLLYEA, D. A., KRUG, U., BERDEL, W. E., KLEIN, H. U., DUGAS, M., SHIRAISHI, Y., CHIBA, K., TANAKA, H., MIYANO, S., YOSHIDA, K., OGAWA, S., MULLER-TIDOW, C. & MACIEJEWSKI, J. P. 2015. Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. *Cancer Cell,* 27**,** 658-70.

PONTEL, L. B., ROSADO, I. V., BURGOS-BARRAGAN, G., GARAYCOECHEA, J. I., YU, R., ARENDS, M. J., CHANDRASEKARAN, G., BROECKER, V., WEI, W., LIU, L., SWENBERG, J. A., CROSSAN, G. P. & PATEL, K. J. 2015. Endogenous Formaldehyde Is a Hematopoietic Stem Cell Genotoxin and Metabolic Carcinogen. *Mol Cell,* 60**,** 177-88.

PROKOFYEVA, D., BOGDANOVA, N., DUBROWINSKAJA, N., BERMISHEVA, M., TAKHIROVA, Z., ANTONENKOVA, N., TURMANOV, N., DATSYUK, I., GANTSEV, S., CHRISTIANSEN, H., PARK-SIMON, T. W., HILLEMANNS, P., KHUSNUTDINOVA, E. & DORK, T. 2013. Nonsense mutation p.Q548X in BLM, the gene mutated in Bloom's syndrome, is associated with breast cancer in Slavic populations. *Breast Cancer Res Treat,* 137**,** 533-9.

QIN, Y. Z., ZHU, H. H., JIANG, Q., JIANG, H., ZHANG, L. P., XU, L. P., WANG, Y., LIU, Y. R., LAI, Y. Y., SHI, H. X., JIANG, B. & HUANG, X. J. 2014. Prevalence and prognostic significance of c-KIT mutations in core binding factor acute myeloid leukemia: a comprehensive large-scale study from a single Chinese center. *Leuk Res,* 38**,** 1435-40.

RAFNAR, T., GUDBJARTSSON, D. F., SULEM, P., JONASDOTTIR, A., SIGURDSSON, A., JONASDOTTIR, A., BESENBACHER, S., LUNDIN, P., STACEY, S. N., GUDMUNDSSON, J., MAGNUSSON, O. T., LE ROUX, L., ORLYGSDOTTIR, G., HELGADOTTIR, H. T., JOHANNSDOTTIR, H., GYLFASON, A., TRYGGVADOTTIR, L., JONASSON, J. G., DE JUAN, A., ORTEGA, E., RAMON-CAJAL, J. M., GARCIA-PRATS, M. D., MAYORDOMO, C., PANADERO, A., RIVERA, F., ABEN, K. K., VAN ALTENA, A. M., MASSUGER, L. F., AAVIKKO, M., KUJALA, P. M., STAFF, S., AALTONEN, L. A., OLAFSDOTTIR, K., BJORNSSON, J., KONG, A., SALVARSDOTTIR, A., SAEMUNDSSON, H., OLAFSSON, K., BENEDIKTSDOTTIR, K. R., GULCHER, J., MASSON, G., KIEMENEY, L. A., MAYORDOMO, J. I., THORSTEINSDOTTIR, U. & STEFANSSON, K. 2011. Mutations in BRIP1 confer high risk of ovarian cancer. *Nat Genet,* 43**,** 1104-7.

RAJENDRA, E., OESTERGAARD, V. H., LANGEVIN, F., WANG, M., DORNAN, G. L., PATEL, K. J. & PASSMORE, L. A. 2014. The genetic and biochemical basis of FANCD2 monoubiquitination. *Mol Cell,* 54**,** 858-69.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-308.

RATAJSKA, M., ANTOSZEWSKA, E., PISKORZ, A., BROZEK, I., BORG, A., KUSMIEREK, H., BIERNAT, W. & LIMON, J. 2012. Cancer predisposing BARD1 mutations in breast-ovarian cancer families. *Breast Cancer Res Treat,* 131**,** 89-97.

ROGERS, K. J., FU, W., AKEY, J. M. & MONNAT, R. J., JR. 2014. Global and disease-associated genetic variation in the human Fanconi anemia gene family. *Hum Mol Genet,* 23**,** 6815-25.

ROSENBERG, P. S., ALTER, B. P. & EBELL, W. 2008. Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. *Haematologica,* 93**,** 511-7.

ROSENBERG, P. S., TAMARY, H. & ALTER, B. P. 2011. How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel. *Am J Med Genet A,* 155A**,** 1877-83.

ROY, R., CHUN, J. & POWELL, S. N. 2011. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer,* 12**,** 68-78.

RUDD, M. F., SELLICK, G. S., WEBB, E. L., CATOVSKY, D. & HOULSTON, R. S. 2006. Variants in the ATM-BRCA2-CHEK2 axis predispose to chronic lymphocytic leukemia. *Blood,* 108**,** 638-44.

SAAR, K., CHRZANOWSKA, K. H., STUMM, M., JUNG, M., NURNBERG, G., WIENKER, T. F., SEEMANOVA, E., WEGNER, R. D., REIS, A. & SPERLING, K. 1997. The gene for the ataxia-telangiectasia variant, Nijmegen breakage syndrome, maps to a 1-cM interval on chromosome 8q21. *Am J Hum Genet,* 60**,** 605-10.

SAITO, S., GOODARZI, A. A., HIGASHIMOTO, Y., NODA, Y., LEES-MILLER, S. P., APPELLA, E. & ANDERSON, C. W. 2002. ATM mediates phosphorylation at multiple p53 sites, including Ser(46), in response to ionizing radiation. *J Biol Chem,* 277**,** 12491-4.

SALE, J. E., LEHMANN, A. R. & WOODGATE, R. 2012. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol,* 13**,** 141-52.

SAWYER, S. L., TIAN, L., KAHKONEN, M., SCHWARTZENTRUBER, J., KIRCHER, M., UNIVERSITY OF WASHINGTON CENTRE FOR MENDELIAN, G., CONSORTIUM, F. C., MAJEWSKI, J., DYMENT, D. A., INNES, A. M., BOYCOTT, K. M., MOREAU, L. A., MOILANEN, J. S. & GREENBERG, R. A. 2015. Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. *Cancer Discov,* 5**,** 135-42.

SCHEPERS, K., CAMPBELL, T. B. & PASSEGUE, E. 2015. Normal and leukemic stem cell niches: insights and therapeutic opportunities. *Cell Stem Cell,* 16**,** 254-67.

SCHLACHER, K., WU, H. & JASIN, M. 2012. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell,* 22**,** 106-16.

SCHMIT, J. M., TURNER, D. J., HROMAS, R. A., WINGARD, J. R., BROWN, R. A., LI, Y., LI, M. M., SLAYTON, W. B. & COGLE, C. R. 2015. Two novel RUNX1 mutations in a patient with congenital thrombocytopenia that evolved into a high grade myelodysplastic syndrome. *Leuk Res Rep,* 4**,** 24-7.

SCHNEIDER, M., CHANDLER, K., TISCHKOWITZ, M. & MEYER, S. 2015. Fanconi anaemia: genetics, molecular biology, and cancer - implications for clinical management in children and adults. *Clin Genet,* 88**,** 13-24.

SCHOFIELD, R. 1978. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells,* 4**,** 7-25.

SEDIC, M. & KUPERWASSER, C. 2016. BRCA1-hapoinsufficiency: Unraveling the molecular and cellular basis for tissue-specific cancer. *Cell Cycle,* 15**,** 621-7.

SEEMANOVA, E., JAROLIM, P., SEEMAN, P., VARON, R., DIGWEED, M., SWIFT, M. & SPERLING, K. 2007. Cancer risk of heterozygotes with the NBN founder mutation. *J Natl Cancer Inst,* 99**,** 1875-80.

SEITA, J. & WEISSMAN, I. L. 2010. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med,* 2**,** 640-53.

SHI, J., SHAO, Z. H., LIU, H., BAI, J., CAO, Y. R., HE, G. S., TU, M. F., WANG, X. L., HAO, Y. S., YANG, T. Y. & YANG, C. L. 2004. Transformation of myelodysplastic syndromes into acute myeloid leukemias. *Chin Med J (Engl),* 117**,** 963-7.

SHIH, A. H., ABDEL-WAHAB, O., PATEL, J. P. & LEVINE, R. L. 2012. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat Rev Cancer,* 12**,** 599-612.

SHIOTANI, B. & ZOU, L. 2009. Single-stranded DNA orchestrates an ATM-to-ATR switch at DNA breaks. *Mol Cell,* 33**,** 547-58.

SIEGEL, E. C. & BRYSON, V. 1967. Mutator gene of Escherichia coli B. *J Bacteriol,* 94**,** 38-47.

SILVA, F. C., VALENTIN, M. D., FERREIRA FDE, O., CARRARO, D. M. & ROSSI, B. M. 2009. Mismatch repair genes in Lynch syndrome: a review. *Sao Paulo Med J,* 127**,** 46-51.

SINGH, T. R., SARO, D., ALI, A. M., ZHENG, X. F., DU, C. H., KILLEN, M. W., SACHPATZIDIS, A., WAHENGBAM, K., PIERCE, A. J., XIONG, Y., SUNG, P. & MEETEI, A. R. 2010. MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. *Mol Cell,* 37**,** 879-86.

SMILENOV, L. B., LIEBERMAN, H. B., MITCHELL, S. A., BAKER, R. A., HOPKINS, K. M. & HALL, E. J. 2005. Combined haploinsufficiency for ATM and RAD9 as a factor in cell transformation, apoptosis, and DNA lesion repair dynamics. *Cancer Res,* 65**,** 933-8.

SMITH, M. L., CAVENAGH, J. D., LISTER, T. A. & FITZGIBBON, J. 2004. Mutation of CEBPA in familial acute myeloid leukemia. *N Engl J Med,* 351**,** 2403-7.

SOMMER, S. S., BUZIN, C. H., JUNG, M., ZHENG, J., LIU, Q., JEONG, S. J., MOULDS, J., NGUYEN, V. Q., FENG, J., BENNETT, W. P. & DRITSCHILO, A. 2002. Elevated frequency of ATM gene missense mutations in breast cancer relative to ethnically matched controls. *Cancer Genet Cytogenet,* 134**,** 25-32.

SONG, W. J., SULLIVAN, M. G., LEGARE, R. D., HUTCHINGS, S., TAN, X., KUFRIN, D., RATAJCZAK, J., RESENDE, I. C., HAWORTH, C., HOCK, R., LOH, M., FELIX, C., ROY, D. C., BUSQUE, L., KURNIT, D., WILLMAN, C., GEWIRTZ, A. M., SPECK, N. A., BUSHWELLER, J. H., LI, F. P., GARDINER, K., PONCZ, M., MARIS, J. M. & GILLILAND, D. G. 1999. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet,* 23**,** 166-75.

SOULIER, J. 2013. Pathogenesis of bone marrow failure and leukemia

progression in Fanconi anemia.

STAN, D. L., SHUSTER, L. T., WICK, M. J., SWANSON, C. L., PRUTHI, S. & BAKKUM-GAMEZ, J. N. 2013. Challenging and complex decisions in the management of the BRCA mutation carrier. *J Womens Health (Larchmt),* 22**,** 825-34.

STANKOVIC, T., KIDD, A. M., SUTCLIFFE, A., MCGUIRE, G. M., ROBINSON, P., WEBER, P., BEDENHAM, T., BRADWELL, A. R., EASTON, D. F., LENNOX, G. G., HAITES, N., BYRD, P. J. & TAYLOR, A. M. 1998. ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of mutant ATM and the risk of leukemia, lymphoma, and breast cancer. *Am J Hum Genet,* 62**,** 334-45.

STEFFEN, J., VARON, R., MOSOR, M., MANEVA, G., MAURER, M., STUMM, M., NOWAKOWSKA, D., RUBACH, M., KOSAKOWSKA, E., RUKA, W., NOWECKI, Z., RUTKOWSKI, P., DEMKOW, T., SADOWSKA, M., BIDZINSKI, M., GAWRYCHOWSKI, K. & SPERLING, K. 2004. Increased cancer risk of heterozygotes with NBS1 germline mutations in Poland. *Int J Cancer,* 111**,** 67-71.

STENGEL, A., KERN, W., HAFERLACH, T., MEGGENDORFER, M., FASAN, A. & HAFERLACH, C. 2017. The impact of TP53 mutations and TP53 deletions on survival varies between AML, ALL, MDS and CLL: an analysis of 3307 cases. *Leukemia,* 31**,** 705-711.

STRATHDEE, C. A., GAVISH, H., SHANNON, W. R. & BUCHWALD, M. 1992. Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature,* 358**,** 434.

SUAREZ, F., MAHLAOUI, N., CANIONI, D., ANDRIAMANGA, C., DUBOIS D'ENGHIEN, C., BROUSSE, N., JAIS, J. P., FISCHER, A., HERMINE, O. & STOPPA-LYONNET, D. 2015. Incidence, presentation, and prognosis of malignancies in ataxia-telangiectasia: a report from the French national registry of primary immune deficiencies. *J Clin Oncol,* 33**,** 202-8.

SUGAWARA, N., WANG, X. & HABER, J. E. 2003. In vivo roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. *Mol Cell,* 12**,** 209-19.

SULKOWSKI, P. L., CORSO, C. D., ROBINSON, N. D., SCANLON, S. E., PURSHOUSE, K. R., BAI, H., LIU, Y., SUNDARAM, R. K., HEGAN, D. C., FONS, N. R., BREUER, G. A., SONG, Y., MISHRA-GORUR, K., DE FEYTER, H. M., DE GRAAF, R. A., SUROVTSEVA, Y. V., KACHMAN, M., HALENE, S., GUNEL, M., GLAZER, P. M. & BINDRA, R. S. 2017. 2-Hydroxyglutarate produced by neomorphic IDH mutations suppresses homologous recombination and induces PARP inhibitor sensitivity. *Sci Transl Med,* 9.

SUN, Q. Y., DING, L. W., TAN, K. T., CHIEN, W., MAYAKONDA, A., LIN, D. C., LOH, X. Y., XIAO, J. F., MEGGENDORFER, M., ALPERMANN, T., GARG, M., LIM, S. L., MADAN, V., HATTORI, N., NAGATA, Y., MIYANO, S., YEOH, A. E., HOU, H. A., JIANG, Y. Y., TAKAO, S., LIU, L. Z., TAN, S. Z., LILL, M., HAYASHI, M., KINOSHITA, A., KANTARJIAN, H. M., KORNBLAU, S. M., OGAWA, S., HAFERLACH, T., YANG, H. & KOEFFLER, H. P. 2017. Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD). *Leukemia,* 31**,** 1-10.

SUWAKI, N., KLARE, K. & TARSOUNAS, M. 2011. RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis. *Semin Cell Dev Biol,* 22**,** 898-905.

SWUEC, P., RENAULT, L., BORG, A., SHAH, F., MURPHY, V. J., VAN TWEST, S., SNIJDERS, A. P., DEANS, A. J. & COSTA, A. 2017. The FA Core Complex Contains a Homo-dimeric Catalytic Module for the Symmetric Mono-ubiquitination of FANCI-FANCD2. *Cell Rep,* 18**,** 611-623.

SZKLARCZYK, D., FRANCESCHINI, A., KUHN, M., SIMONOVIC, M., ROTH, A., MINGUEZ, P., DOERKS, T., STARK, M., MULLER, J., BORK, P., JENSEN, L. J. & VON MERING, C. 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res,* 39**,** D561-8.

TAN, S. L. W., CHADHA, S., LIU, Y., GABASOVA, E., PERERA, D., AHMED, K., CONSTANTINOU, S., RENAUDIN, X., LEE, M., AEBERSOLD, R. & VENKITARAMAN, A. R. 2017. A Class of Environmental and Endogenous Toxins Induces BRCA2 Haploinsufficiency and Genome Instability. *Cell,* 169**,** 1105-1118 e15.

TANAKA, A., WEINEL, S., NAGY, N., O'DRISCOLL, M., LAI-CHEONG, J. E., KULP-SHORTEN, C. L., KNABLE, A., CARPENTER, G., FISHER, S. A., HIRAGUN, M., YANASE, Y., HIDE, M., CALLEN, J. & MCGRATH, J. A. 2012. Germline mutation in ATR in autosomal- dominant oropharyngeal cancer syndrome. *Am J Hum Genet,* 90**,** 511-7.

TARLOCK, K. & MESHINCHI, S. 2015. Pediatric acute myeloid leukemia: biology and therapeutic implications of genomic variants. *Pediatr Clin North Am,* 62**,** 75-93.

TAWANA, K. & FITZGIBBON, J. 2016. Inherited DDX41 mutations: 11 genes and counting. *Blood,* 127**,** 960-1.

TAWANA, K., WANG, J., RENNEVILLE, A., BODOR, C., HILLS, R., LOVEDAY, C., SAVIC, A., VAN DELFT, F. W., TRELEAVEN, J., GEORGIADES, P., UGLOW, E., ASOU, N., UIKE, N., DEBELJAK, M., JAZBEC, J., ANCLIFF, P., GALE, R., THOMAS, X., MIALOU, V., DOHNER, K., BULLINGER, L., MUELLER, B., PABST, T., STELLJES, M., SCHLEGELBERGER, B., WOZNIAK, E., IQBAL, S., OKOSUN, J., ARAF, S., FRANK, A. K., LAURIDSEN, F. B., PORSE, B., NERLOV, C., OWEN, C., DOKAL, I., GRIBBEN, J., SMITH, M., PREUDHOMME, C., CHELALA, C., CAVENAGH, J. & FITZGIBBON, J. 2015. Disease evolution and outcomes in familial AML with germline CEBPA mutations. *Blood,* 126**,** 1214-23.

TEFFERI, A. & LETENDRE, L. 2012. Going beyond 7 + 3 regimens in the treatment of adult acute myeloid leukemia. *J Clin Oncol,* 30**,** 2425-8.

TERSTAPPEN, L. W., HUANG, S., SAFFORD, M., LANSDORP, P. M. & LOKEN, M. R. 1991. Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34+CD38- progenitor cells. *Blood,* 77**,** 1218-27.

THOL, F., BOLLIN, R., GEHLHAAR, M., WALTER, C., DUGAS, M., SUCHANEK, K. J., KIRCHNER, A., HUANG, L., CHATURVEDI, A., WICHMANN, M., WIEHLMANN, L., SHAHSWAR, R., DAMM, F., GOHRING, G., SCHLEGELBERGER, B., SCHLENK, R., DOHNER, K., DOHNER, H., KRAUTER, J., GANSER, A. & HEUSER, M. 2014. Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood,* 123**,** 914-20.

TIBBETTS, R. S., BRUMBAUGH, K. M., WILLIAMS, J. M., SARKARIA, J. N., CLIBY, W. A., SHIEH, S. Y., TAYA, Y., PRIVES, C. & ABRAHAM, R. T. 1999. A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes Dev,* 13**,** 152-7.

TIRABY, J. G. & FOX, M. S. 1973. Marker discrimination in transformation and mutation of pneumococcus. *Proc Natl Acad Sci U S A,* 70**,** 3541-5.

TISCHKOWITZ, M., EASTON, D. F., BALL, J., HODGSON, S. V. & MATHEW, C. G. 2008. Cancer incidence in relatives of British Fanconi Anaemia patients. *BMC Cancer,* 8**,** 257.

TISCHKOWITZ, M. D., MORGAN, N. V., GRIMWADE, D., EDDY, C., BALL, S., VORECHOVSKY, I., LANGABEER, S., STOGER, R., HODGSON, S. V. & MATHEW, C. G. 2004. Deletion and reduced expression of the Fanconi anemia FANCA gene in sporadic acute myeloid leukemia. *Leukemia,* 18**,** 420-5.

TOMITA, H., TANAKA, K., TANAKA, T. & HARA, A. 2016. Aldehyde dehydrogenase 1A1 in stem cells and cancer. *Oncotarget,* 7**,** 11018-32.

UMESAKO, S., FUJISAWA, K., IIGA, S., MORI, N., TAKAHASHI, M., HONG, D. P., SONG, C. W., HAGA, S., IMAI, S., NIWA, O. & OKUMOTO, M. 2005. Atm heterozygous deficiency enhances development of mammary carcinomas in p53 heterozygous knockout mice. *Breast Cancer Res,* 7**,** R164-70.

UZIEL, T., LERENTHAL, Y., MOYAL, L., ANDEGEKO, Y., MITTELMAN, L. & SHILOH, Y. 2003. Requirement of the MRN complex for ATM activation by DNA damage. *EMBO J,* 22**,** 5612-21.

VACLOVA, T., WOODS, N. T., MEGIAS, D., GOMEZ-LOPEZ, S., SETIEN, F., GARCIA BUENO, J. M., MACIAS, J. A., BARROSO, A., URIOSTE, M., ESTELLER, M., MONTEIRO, A. N., BENITEZ, J. & OSORIO, A. 2016. Germline missense pathogenic variants in the BRCA1 BRCT domain, p.Gly1706Glu and p.Ala1708Glu, increase cellular sensitivity to PARP inhibitor Olaparib by a dominant negative effect. *Hum Mol Genet*.

VALERIO, D. G., KATSMAN-KUIPERS, J. E., JANSEN, J. H., VERBOON, L. J., DE HAAS, V., STARY, J., BARUCHEL, A., ZIMMERMANN, M., PIETERS, R., REINHARDT, D., VAN DEN HEUVEL-EIBRINK, M. M. & ZWAAN, C. M. 2014. Mapping epigenetic regulator gene mutations in cytogenetically normal pediatric acute myeloid leukemia. *Haematologica,* 99**,** e130-2.

VAN OS, N. J., ROELEVELD, N., WEEMAES, C. M., JONGMANS, M. C., JANSSENS, G. O., TAYLOR, A. M., HOOGERBRUGGE, N. & WILLEMSEN, M. A. 2016. Health risks for ataxia-telangiectasia mutated heterozygotes: a systematic review, meta-analysis and evidence-based guideline. *Clin Genet,* 90**,** 105-17.

VAN OS, N. J. H., HAAXMA, C. A., VAN DER FLIER, M., MERKUS, P., VAN DEUREN, M., DE GROOT, I. J. M., LOEFFEN, J., VAN DE WARRENBURG, B. P. C., WILLEMSEN, M. & GROUP, A. T. S. 2017. Ataxia-telangiectasia: recommendations for multidisciplinary treatment. *Dev Med Child Neurol,* 59**,** 680-689.

VAN TWEST, S., MURPHY, V. J., HODSON, C., TAN, W., SWUEC, P., O'ROURKE, J. J., HEIERHORST, J., CRISMANI, W. & DEANS, A. J. 2017. Mechanism of Ubiquitination and Deubiquitination in the Fanconi Anemia Pathway. *Mol Cell,* 65**,** 247-259.

VARDIMAN, J. W., HARRIS, N. L. & BRUNNING, R. D. 2002. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood,* 100**,** 2292-302.

VAZ, F., HANENBERG, H., SCHUSTER, B., BARKER, K., WIEK, C., ERVEN, V., NEVELING, K., ENDT, D., KESTERTON, I., AUTORE, F., FRATERNALI, F., FREUND, M., HARTMANN, L., GRIMWADE, D., ROBERTS, R. G., SCHAAL, H., MOHAMMED, S., RAHMAN, N., SCHINDLER, D. & MATHEW, C. G. 2010. Mutation of the RAD51C gene in a Fanconi anemia-like disorder. *Nat Genet,* 42**,** 406-9.

VELLEUER, S. L. A. E. 2006. Guido Fanconi (1892–1979): a jack of all trades.

VERLANDER, P. C., LIN, J. D., UDONO, M. U., ZHANG, Q., GIBSON, R. A., MATHEW, C. G. & AUERBACH, A. D. 1994. Mutation analysis of the Fanconi anemia gene FACC. *Am J Hum Genet,* 54**,** 595-601.

WALDEN, H. & DEANS, A. J. 2014. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu Rev Biophys,* 43**,** 257-78.

WALTER, D., LIER, A., GEISELHART, A., THALHEIMER, F. B., HUNTSCHA, S., SOBOTTA, M. C., MOEHRLE, B., BROCKS, D., BAYINDIR, I., KASCHUTNIG, P., MUEDDER, K., KLEIN, C., JAUCH, A., SCHROEDER, T., GEIGER, H., DICK, T. P., HOLLAND-LETZ, T., SCHMEZER, P., LANE, S. W., RIEGER, M. A., ESSERS, M. A., WILLIAMS, D. A., TRUMPP, A. & MILSOM, M. D.

2015. Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature,* 520**,** 549-52.

WANG, A. T., KIM, T., WAGNER, J. E., CONTI, B. A., LACH, F. P., HUANG, A. L., MOLINA, H., SANBORN, E. M., ZIERHUT, H., CORNES, B. K., ABHYANKAR, A., SOUGNEZ, C., GABRIEL, S. B., AUERBACH, A. D., KOWALCZYKOWSKI, S. C. & SMOGORZEWSKA, A. 2015. A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. *Mol Cell,* 59**,** 478-90.

WANG, Q., CHEN, X. & HAY, N. 2017. Akt as a target for cancer therapy: more is not always better (lessons from studies in mice). *Br J Cancer,* 117**,** 159-163.

WATERS, L. S., MINESINGER, B. K., WILTROUT, M. E., D'SOUZA, S., WOODRUFF, R. V. & WALKER, G. C. 2009. Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol Mol Biol Rev,* 73**,** 134-54.

WEGNER, R. D., METZGER, M., HANEFELD, F., JASPERS, N. G., BAAN, C., MAGDORF, K., KUNZE, J. & SPERLING, K. 1988. A new chromosomal instability disorder confirmed by complementation studies. *Clin Genet,* 33**,** 20-32.

WELLCOME TRUST CASE CONTROL, C. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* 447**,** 661-78.

WETERINGS, E., GALLEGOS, A. C., DOMINICK, L. N., COOKE, L. S., BARTELS, T. N., VAGNER, J., MATSUNAGA, T. O. & MAHADEVAN, D. 2016. A novel small molecule inhibitor of the DNA repair protein Ku70/80. *DNA Repair (Amst),* 43**,** 98-106.

WU, Y., BERENDS, M. J., SIJMONS, R. H., MENSINK, R. G., VERLIND, E., KOOI, K. A., VAN DER SLUIS, T., KEMPINGA, C., VAN DDER ZEE, A. G., HOLLEMA, H., BUYS, C. H., KLEIBEUKER, J. H. & HOFSTRA, R. M. 2001. A role for MLH3 in hereditary nonpolyposis colorectal cancer. *Nat Genet,* 29**,** 137-8.

XIA, B., SHENG, Q., NAKANISHI, K., OHASHI, A., WU, J., CHRIST, N., LIU, X., JASIN, M., COUCH, F. J. & LIVINGSTON, D. M. 2006. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol Cell,* 22**,** 719-29.

XU, D., GUO, R., SOBECK, A., BACHRATI, C. Z., YANG, J., ENOMOTO, T., BROWN, G. W., HOATLIN, M. E., HICKSON, I. D. & WANG, W. 2008. RMI, a new OB-fold complex essential for Bloom syndrome protein to maintain genome stability. *Genes Dev,* 22**,** 2843-55.

XUE, X., RAYNARD, S., BUSYGINA, V., SINGH, A. K. & SUNG, P. 2013. Role of replication protein A in double holliday junction dissolution mediated by the BLM-Topo IIIalpha-RMI1-RMI2 protein complex. *J Biol Chem,* 288**,** 14221-7.

YAMAMOTO, K. N., KOBAYASHI, S., TSUDA, M., KURUMIZAKA, H., TAKATA, M., KONO, K., JIRICNY, J., TAKEDA, S. & HIROTA, K. 2011. Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proc Natl Acad Sci U S A,* 108**,** 6492-6.

YANG, L., BRYDER, D., ADOLFSSON, J., NYGREN, J., MANSSON, R., SIGVARDSSON, M. & JACOBSEN, S. E. 2005. Identification of Lin(-)Sca1(+)kit(+)CD34(+)Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood,* 105**,** 2717-23.

YANG, L., RAU, R. & GOODELL, M. A. 2015. DNMT3A in haematological malignancies. *Nat Rev Cancer,* 15**,** 152-65.

YIN, J., SOBECK, A., XU, C., MEETEI, A. R., HOATLIN, M., LI, L. & WANG, W. 2005. BLAP75, an essential component of Bloom's syndrome protein complexes that maintain genome integrity. *EMBO J,* 24**,** 1465-76.

YUAN, X. Q., PENG, L., ZENG, W. J., JIANG, B. Y., LI, G. C. & CHEN, X. P. 2016. DNMT3A R882 Mutations Predict a Poor Prognosis in AML: A Meta-Analysis From 4474 Patients. *Medicine (Baltimore),* 95**,** e3519.

ZHANG, J., WALSH, M. F., WU, G., EDMONSON, M. N., GRUBER, T. A., EASTON, J., HEDGES, D., MA, X., ZHOU, X., YERGEAU, D. A., WILKINSON, M. R., VADODARIA, B., CHEN, X., MCGEE, R. B., HINES-DOWELL, S., NUCCIO, R., QUINN, E., SHURTLEFF, S. A., RUSCH, M., PATEL, A., BECKSFORT, J. B., WANG, S., WEAVER, M. S., DING, L., MARDIS, E. R., WILSON, R. K., GAJJAR, A., ELLISON, D. W., PAPPO, A. S., PUI, C. H., NICHOLS, K. E. & DOWNING, J. R. 2015. Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med,* 373**,** 2336-46.

ZHANG, S. J., MA, L. Y., HUANG, Q. H., LI, G., GU, B. W., GAO, X. D., SHI, J. Y., WANG, Y. Y., GAO, L., CAI, X., REN, R. B., ZHU, J., CHEN, Z. & CHEN, S. J. 2008. Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia. *Proc Natl Acad Sci U S A,* 105**,** 2076-81.

ZOU, L. & ELLEDGE, S. J. 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science,* 300**,** 1542-8.

Appendices

Appendix A. Oligonucleotides used for validation of FA-BRCA-HRR variants by PCR & Sanger sequencing

WES ID	Gene	Type of Mutation	Changes	Primer Set	Forward	Reverse	Annealing Temperature (Forward / Reverse)	Amplicon (in base pairs)	GC% (Forward / Reverse)
WES-47	FANCA	Missense	c.C455G	FANCA F1 & R1	TCCCGTGGGTATTCTCTCAG	TCAACAGAACATTGCCTGGA	60/60	383	55/45
WES-32	FANCD1	Non FS Del	c.6399_6401del	FANCD1-1	CAGCAAGTGGAAAGCAAGTTT	AAGCCTGTTCTTTTCCCAAA	59.5/58.9	387	43/40
WES-21	FANCD2	Nonsense	c.C2776T	FANCD2-F2 & R2	AGGGACTTGGGCTAGAGGAA	CCTCAGTGTCACAGTGTTCTTTG	60/60	386	55/47.8
WES-59	FANCI	Missense	c.A394G	FANCI F1 & R1	CAGTTCTGGATCTCGGTCAA	TTTTCCGCCACTTGTATTCC	60/60	458	50/45
WES-1	FANCI	Missense	c.C1656G	FANCI F3 & R3	GTTTTGCTCTACGCTTCATTGT	GACAGTGGCTACCATACTGGAT	58.6/58.1	472	40.9/50
WES-18	FANCL	Missense	c.A343G	FANCL F2 & R2	GGGAATGAGTCAGCCAGATT	CCCTCTCTTTAATTCACAGCAA	59.1/58.5	402	50/40.9
WES-41	FANCL	Missense	c.C50G	FANCL F1 & R1	TGGACTTGAGGGCAATCTTC	CCTAGCCCGTCACAGACTTC	60/60	410 or 343	50/60
WES-94	FANCM	FS Del	c.3998delA	FANCM F2 & R2	AGGATTCTGTAGTCCAGATTCTGA	GCAGATACATTGGACCACTGC	58.4/60.5	497	41.7/52.4
WES-9	FANCM	Missense	c.A1545C	FANCM F1 & R1	AACGTGATGAGACCCGAGTT	AGTCAGGATTCCAACCCAGTT	59.6/59.8	429	50/47.6
WES-46	FANCM	Nonsense	c.C5101T	FANCM-3	GAGTCTTGCAAAGGCCAATC	AGGGTGGTGTGGTAGACTGG	60/60	454	50/60
WES-55	FANCP	Missense	c.G2312C	SLX4 F1 & R1	AGGCTGCAGTAAGCCATGAT	CTGGTCATGGACTTGGGATT	60/60	492	50/50
WES-73	FANCP	Missense	c.C1271T	SLX4 F2 & R2	TTCACACAAGGCAGTGGAAG	CTCCCAGGGTCACTCTTCTG	60/60	390	50/60
WES-81	FANCP	Nonsense	c.C4996T	SLX4 F3 & R3	TCCTGAGATCCACCTGTTCC	GAGGATACATGAGGCCACTGA	60/60	486	N/A
WES-34	FANCQ	Missense	c.T2117C	ERCC4-1	GTAGGTGGCCAGGAACAGAA	CACCTCGGGAAGTGAGAGAG	60/60	258	45/60
WES-74, WES-208	FANCQ	Missense	c.C2288T	ERCC4-3	TGCGTGAATTTCGAAGTGAG	TGTGGCTTGCTTTGTTTCAG	45/45	448	60/60
WES-46	FANCS	Nonsense	c.C3607T	BRCA1 F1 & R1	TGCATCTCAGGTTTGTTCTGA	ACAGACACTCGGTAGCAACG	59.4/59	335	42.9/55
WES-2	ATM	Missense	c.G7618A	ATM F1 & R1	GAATGGGGACCAAGATGATG	CAAAGTGTGATGGGGGTGAT	60.1/60.6	417	50/50
WES-26	ATM	Missense	c.G6860C	ATM F3 & R3	CATGCAGACAGAGAGGTCCTT	ACAGCTGGCATCCAACTTCT	59.5/59.9	309	52.4/50
WES-61	ATM	Missense	c.A5116G	ATM F6 & R6	GAAGCTGCTTGGGAGAAGTG	CAGGGACCTTGTCTGGAATG	60.1/60.5	381	55/55
WES-84	ATM	FS deletion	c.2905delC	ATM F7 & R7	CTCCCAAAGTGCTGGGATTA	CCTTGAGCATCCCTTGTGTT	60.1/60.1	428	50/50
WES-34	ATR	Missense	c.A2437G	ATR F1 & R1	GGTGACAGAGCAAGACCCTA	CTTCAGAGTCCAAGGATTCCA	57.9/59.3	421	55/47.6

WES-74	ATR	Missense	c.A6259G	ATR F2 & R2	TTTTTGTGAAAACGGTATGTGG	GATTTGTATTTGTGGCCCTGA	59.8/59.8	414	36.4/42.9
WES-58	BARD1	Missense	c.G1868A	BARD1 F1 & R1	AAGAAGCAGGCCAAAGAAAT	ACAGGGCTTCACCGTGTTAG	58.1/60.2	408	40/55
WES-87	FAN1	Missense	c.A2525G	FAN1 F2 & R2	TGAGAACCACTGCTTTGTGG	AGATGGTTTCTGCTGGCTGT	59.9/59.9	477	50/50
WES-94	MLH1	Nonsense	c.C960G	MLH1 F1 & R1	ACAGAACATGAGTGGCAGCA	CCGATAACCTGAGAACACCAA	60.5/60	341	50/47.6
WES-14	MLH3	Missense	c.A1234G	MLH3 F1 & R1	ACTTCCGATGAGAGGAGCAA	CTGGTTCCACACGGATTTTC	60/60.4	365	50/50
WES-27	MLH3	Missense	c.T3173G	MLH3 F2 & R2	CGGTAGAAGATGCCACAGGT	CTTGTCCAGCATTCCCATCT	60.1/60.1	316	55/50
WES-31	NBN	Nonsense	c.C127T	NBN F1 & R1	AACCTTTGATAGCCTTCAGTGAG	TTTGTGATTTCAACCCCCTTA	59/59.3	387	43.5/38.1
WES-14	RAD51D	Missense	c.C553T	RAD51D F1 & R1	ACCTGAGTCCTTGCATCCAG	ATTGCACATCTGCATTTCCA	60.3/60.1	302	55/40
WES-4	RMI1	FS deletion	c.1419_1434del	RMI1 F1 & R1	CCAGCAGTTCAGATAGCCATT	ACCATCAGACACCTTTGCAGT	59.4/59.6	323	47.6/47.6
WES-4	RPA1	Missense	c.C1165T	RPA1 F1 & R1	GGGAAATGCTCTTTCCCCTA	GCCTTCAAGAAACACGGAAG	60.4/59.9	495	50/50
WES-94	RPA1	Missense	c.G1300A	RPA1 F2 & R2	ATGACCGTGACCTGTGTGAA	TTCCCTCTGCCAGTGATTCT	60/59.8	425	50/50
WES-63	RPA3	Missense	c.T83C	RPA3 F1 & R1	GAGCGCTAGTCTTCGCTGAT	GACGGGCACTGGAATTTAGA	59.9/60.1	448	55/50

Appendix B. Output of interacting proteins from queries of individual FANC 16 proteins from STRINGdb

Interacts with FANCA	Interacts with FANCB	Interacts with FANCC	Interacts with BRCA2	Interacts with FAND2	Interacts with FANCE	Interacts with FANCF	Interacts with FANCG	Interacts with FANCI	Interacts with BRIP1	Interacts with FANCL	Interacts with FANCM	Interacts with PALB2	Interacts with RAD51C	Interacts with SLX4	Interacts with ERCC4	Combined list used to generate the final FA/BRCA-HRR network		
FANCC	FANCM	FANCE	RAD51	FANCI	FANCC	FANCA	FANCA	FANCD2	BRCA1	FANCA	FANCL	BRCA2	RAD51L1	GIYD1	ERCC1	FANCC	SHFM1	RAP1
FANCF	FANCL	FANCF	BRCA1	FANCE	FANCD2	FANCG	FANCF	FANCL	MLH1	FANCM	FANCA	BRCA1	XRCC3	ERCC4	XPC	FANCF	XRCC3	C16orf75
FANCE	FANCD2	FANCG	FANCD2	BRCA2	FANCA	FANCC	FANCC	FAN1	TOPBP1	FANCG	C19orf40	RAD51	RAD51	EME1	XPA	FANCE	TP53	MORF4
FANCL	FANCG	FANCD2	PALB2	FANCC	FANCM	FANCE	BRCA2	UBC	RAP1	FANCF	FANCE	FANCI	BRCA2	MUS81	ERCC5	FANCL	HMG20B	BRIP1
FANCM	C17orf70	FANCM	SHFM1	BRCA1	FANCF	FANCM	FANCE	C19orf40	PALB2	FANCC	RMI1	FANCD2	MND1	PLK1	ERCC2	FANCM	ATM	RAD51L1
TOP3A	APITD1	FANCL	FANCG	MRE11A	FANCG	FANCL	FANCL	ATM		C17orf70	C17orf70	MORF4	RAD51L3		RAD23B	TOP3A	CHEK2	RAD51L3
FANCB	C19orf40	ZBTB32	XRCC3	ATM	CHEK1	FANCD2	FAH	FANCA		FANCB	FANCF	BRIP1	BRCA1		BTBD12	FANCB	PLK1	GIYD1
BLM	RAD51	STAT1	TP53	FAN1	FANCL	RMI1	FANCM	FANCM		FANCD2	FANCB		MRE11A		CKN2	BLM	CDK2	ERCC4
C17orf70	FANCC	EIF2AK2	HMG20B	FANCA	BRCA2	APITD1	XRCC3	PALB2		UBE2T	FANCG		RAD50		GTF2H1	C17orf70	C11orf30	EME1
FANCD2	FANCF	HSPA4	ATM	FANCM	TOP3A	TOP3A	C17orf70	C17orf70		FANCI	FANCC		ATM		DDB1	FANCD2	RAD50	MUS81
C19orf40	FANCE	FANCI	BARD1	FANCB	FANCI	BLM	FANCB	USP1		C19orf40	APITD1		SPO11		ERCC3	C19orf40	MLH1	ERCC1
BRCA1	FANCI	TOP3A	CHEK2	USP1	BLM	FANCI	FANCD2	WDR48		FANCE	BLM		PSMC3IP		POLD1	BRCA1	MRE11A	XPC
FAH	TOP3A	BLM	PLK1	FANCL	C17orf70	BRCA1	BLM	FANCC		STRA13	FANCD2		CDK4		POLR2A	FAH	RAD51C	XPA
STRA13	UBA52	BRCA1	CDK2	UBC	C19orf40	C17orf70	TOP3A	FANCF		APITD1	TOP3A				DDB2	STRA13	BCCIP	ERCC5
FANCI		C17orf70	FANCE	C19orf40	BRCA1	C19orf40	C19orf40	FANCE		BLM	STRA13				CCNH	FANCI	FANCA	ERCC2
SMARCA4		C19orf40	C11orf30	FANCG	FANCB	FANCB	STRA13	BRCA2		TOP3A	C16orf75				CDK7	SMARCA4	BUB1B	RAD23B
RMI1		FANCB	RAD50	ATRIP	UBE2T	STRA13	CYP2E1	BRCA1		BRCA1	FANCI				XAB2	RMI1	CDK4	BTBD12
APITD1		SPTAN1	MLH1	UBE2T	BARD1	UBA52	APITD1	FANCG		UBA52	UBE2T				ERCC8	APITD1	USP11	CKN2
AKT1		UBA52	MRE11A	RAD18	RFC4	RFC4	BRCA1	UBE2T		CHEK1	UBA52				CETN2	AKT1	H2AFX	GTF2H1
HES1		RMI1	UBC	RAD50	RFC2	CHEK1	FANCI	ATRIP		ATRIP	UBC				MNAT1	HES1	MND1	DDB1
SPTAN1		BARD1	RAD51C	FANCF	STRA13	UBE2T	UBC	UBA52		RFC2					GTF2H3	SPTAN1	RBBP8	ERCC3
BRCA2		CHEK1	FANCI	WDR48	UBA52	RFC2	PRDX3	RFC2		UBC					RPA3	BRCA2	STAT5A	POLD1
RPA1		RFC4	BCCIP	PALB2	UBC	BARD1	BARD1	FANCB		RMI1					POLR2L	RPA1	SMAD3	POLR2A
CHEK1		UBE2T	CHEK1	C17orf70	RMI1	UBC	UBA52	TOPBP1		BARD1						CHEK1	FLNA	DDB2

UBA52	STRA13	FANCA	UBB	APITD1	HUS1	CHEK1	RAD17	TOPBP1	UBA52	PSMD3	CCNH
UBE2T	RFC2	BUB1B	XRCC3	HUS1	RAD9A	ATRIP	RAD9A	UBB	UBE2T	SPO11	CDK7
UBC	APITD1	CDK4	PCNA	ATRIP	TOPBP1	UBE2T	HUS1		UBC	MLH3	XAB2
RFC4	HSPA1A	USP11	UBA52	RAD17		RAD9A	UBB		RFC4	MSH4	ERCC8
RFC2	TOPBP1	BLM	MEN1	UBB					RFC2	PARP1	CETN2
BARD1	RAD17	H2AFX	H2AFX	RAD9A					BARD1	RPA3	MNAT1
ATRIP	UBC	MND1	KAT5	TOPBP1					ATRIP	DMC1	GTF2H3
HUS1	RAD9A	RBBP8	RFC2						HUS1	PSMC3IP	POLR2L
TOPBP1	ATRIP	STAT5A	BARD1						TOPBP1	TEX15	
RAD17	HUS1	SMAD3	RAD17						RAD17	FAN1	
RAD9A	RFC3	FLNA	RFC4						RAD9A	USP1	
	UBB	PSMD3	TOPBP1						FANCG	RAD18	
		TOP3A	RFC3						RAD51	WDR48	
		SPO11	RFC5						ZBTB32	PCNA	
		MLH3	TNKS						STAT1	MEN1	
		MSH4	HUS1						EIF2AK2	KAT5	
		PARP1	RAD9A						HSPA4	RFC5	
		RPA3	ZBTB32						HSPA1A	TNKS	
		DMC1	CHEK1						RFC3	NBN	
		PSMC3IP	NBN						UBB	CYP2E1	
		TEX15							PALB2	PRDX3	

Appendix C. List of FA/BRCA-HRR variants identified from the WES of 145 AML patients at diagnosis

WES ID	Chr	start	end	Gene	Consequence	Ref Seq	Exon	*Nucleotide change	Amino acid change	CADD Score	No. of Reads (Ref)	No. of Reads (Alt)	VAF
WES-1	9	35076755	35076755	FANCG	Missense	NM_004629	7	c.C890T	p.T297I	21.6	30	47	61.0
WES-1	15	89826439	89826439	FANCI	Missense	NM_018193	17	c.C1656G	p.S552R	24.8	18	16	47.1
WES-1	16	3640815	3640815	SLX4	Missense	NM_032444	12	c.G2824C	p.E942Q	14.65	23	19	45.2
WES-2	11	1.08E+08	1.08E+08	ATM	Splicing	NM_000051	-	108202273 G>A	-	11.17	7	5	41.7
WES-2	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	51	c.G7618A	p.V2540I	15.92	19	36	65.5
WES-4	5	68682299	68682299	RAD17	Splicing	-	-	68682299 C>T	-	10.7	9	7	43.8
WES-4	9	86617320	86617335	RMI1	FS Del	NM_024945	3	c.1419_1434del	p.S473fs	35	11	10	47.6
WES-4	17	1783909	1783909	RPA1	Missense	NM_002945	12	c.C1165T	p.R389W	35	49	36	42.4
WES-6	22	29121061	29121061	CHEK2	Missense	NM_007194	4	c.A496G	p.N166D	26.1	47	27	36.5
WES-6	3	48506279	48506279	ATRIP	Missense	NM_032166	11	c.G2024A	p.R675Q	22.2	24	28	53.8
WES-8	2	58386928	58386928	FANCL	FS Ins	NM_018062	14	c.1099_1100insATTA	p.T367fs	38	11	9	45.0
WES-9	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	33	c.A4949G	p.N1650S	11.08	16	15	48.4
WES-9	17	79514028	79514028	C17orf70	Missense	NM_025161	5	c.G2080C	p.A694P	12.9	18	23	56.1
WES-9	14	45628447	45628447	FANCM	Missense	NM_020937	9	c.A1545C	p.K515N	22.2	36	32	47.1
WES-13	11	1.08E+08	1.08E+08	ATM	Splicing	NM_000051	-	108202273 G>A	-	11.18	4	14	77.8
WES-13	17	41222975	41222975	BRCA1	Missense	NM_007298	14	c.G1644A	p.M548I	22.3	17	16	48.5
WES-14	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	8	c.C998T	p.S333F	25.1	4	4	50.0
WES-14	2	58459232	58459232	FANCL	Missense	NM_018062	2	c.C112T	p.L38F	25.3	20	7	25.9
WES-14	14	75515125	75515125	MLH3	Missense	NM_014381	2	c.A1234G	p.K412E	19.29	29	28	49.1
WES-14	17	33433488	33433488	RAD51D	Missense	NM_001142571	6	c.C553T	p.R185W	35	15	16	51.6
WES-14	18	20602153	20602153	RBBP8	Missense	NM_002894	18	c.G2516A	p.R839Q	22.5	31	20	39.2
WES-16	9	86616761	86616761	RMI1	Missense	NM_024945	3	c.C860T	p.P287L	13.88	12	13	52.0
WES-18	15	31197584	31197584	FAN1	Missense	NM_001146094	2	c.G718A	p.E240K	12.62	41	36	46.8

WES-18	2	58449108	58449108	FANCL	Missense	NM_018062	5	c.A343G	p.I115V	14.1	14	6	30.0
WES-20	9	97912307	97912307	FANCC	Missense	NM_001243743	7	c.A584T	p.D195V	25.2	37	35	48.6
WES-21	16	89813256	89813256	FANCA	Missense	NM_000135	34	c.A3391G	p.T1131A	23.5	47	31	39.7
WES-21	3	10116274	10116274	FANCD2	Nonsense	NM_033084	29	c.C2776T	p.R926X	46	14	10	41.7
WES-21	11	65629482	65629482	MUS81	Missense	NM_025128	4	c.G416A	p.R139Q	17.49	3	26	89.7
WES-22	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	8	c.C998T	p.S333F	25.1	5	9	64.3
WES-23	16	3641280	3641280	SLX4	Missense	NM_032444	12	c.G2359A	p.E787K	21.6	26	27	50.9
WES-24	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	31	c.T4709C	p.V1570A	15.35	2	2	50.0
WES-25	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	50	c.T7475G	p.L2492R	22	8	7	46.7
WES-25	14	68353893	68353893	RAD51B	Missense	NM_133510	7	c.A728G	p.K243R	27.1	14	6	30.0
WES-25	18	20572836	20572836	RBBP8	Missense	NM_002894	11	c.C1046T	p.S349F	29	17	7	29.2
WES-26	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	47	c.G6860C	p.G2287A	17.09	29	16	35.6
WES-26	3	1.42E+08	1.42E+08	ATR	Missense	NM_001184	3	c.C268T	p.H90Y	15.97	26	23	46.9
WES-26	14	45606290	45606290	FANCM	Missense	NM_020937	2	c.C527T	p.T176I	20.2	24	18	42.9
WES-27	14	75513186	75513186	MLH3	Missense	NM_014381	2	c.T3173G	p.V1058G	25.3	70	50	41.7
WES-27	14	68353784	68353784	RAD51B	Missense	NM_133510	7	c.G619T	p.V207L	20.7	14	17	54.8
WES-28	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	11	c.T1744C	p.F582L	10.28	21	25	54.3
WES-28	17	59938933	59938933	BRIP1	Splicing	-	-	59938933 A>G	-	15.04	13	20	60.6
WES-30	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	9	c.T1229C	p.V410A	23.4	12	11	47.8
WES-30	2	58388668	58388670	FANCL	Missense	NM_018062	12	c.1007_1009del	p.336_337del	22.8	14	15	51.7
WES-30	14	75514138	75514138	MLH3	Missense	NM_014381	2	c.G2221T	p.V741F	16.6	0	39	100.0
WES-31	8	90994994	90994994	NBN	Nonsense	NM_002485	2	c.C127T	p.R43X	15.87	26	6	18.8
WES-32	13	32914891	32914893	BRCA2	Missense	NM_000059	11	c.6399_6401del	p.2133_2134del	13.71	13	0	0.0
WES-32	17	79517729	79517729	C17orf70	Missense	NM_025161	3	c.C791T	p.A264V	12.51	48	35	42.2
WES-33	5	1.32E+08	1.32E+08	RAD50	Missense	NM_005732	16	c.T2525C	p.V842A	20.5	35	23	39.7
WES-33	17	56772522	56772522	RAD51C	Missense	NM_002876	2	c.G376A	p.A126T	21.7	38	40	51.3

WES-34	3	1.42E+08	1.42E+08	ATR	Missense	NM_001184	11	c.A2437G	p.M813V	19.44	21	16	43.2
WES-34	16	14041570	14041570	ERCC4	Missense	NM_005236	11	c.T2117C	p.I706T	27.2	69	59	46.1
WES-35	3	48506279	48506279	ATRIP	Missense	NM_032166	11	c.G2024A	p.R675Q	22.2	27	24	47.1
WES-35	17	79514675	79514675	C17orf70	Missense	NM_025161	5	c.A1433G	p.Q478R	25	10	7	41.2
WES-35	22	29121087	29121087	CHEK2	Missense	NM_007194	4	c.T470C	p.I157T	22.7	66	49	42.6
WES-37	16	14026059	14026059	ERCC4	Missense	NM_005236	6	c.G1019A	p.R340Q	34	6	3	33.3
WES-37	14	75509146	75509146	MLH3	Missense	NM_014381	3	c.C3315A	p.D1105E	16.99	58	29	33.3
WES-38	3	1.42E+08	1.42E+08	ATR	Missense	NM_001184	3	c.C268T	p.H90Y	15.97	78	48	38.1
WES-38	17	48453136	48453136	EME1	Missense	NM_001166131	2	c.T567A	p.N189K	10.06	28	38	57.6
WES-38	16	14028081	14028081	ERCC4	Missense	NM_005236	7	c.C1135T	p.P379S	32	39	21	35.0
WES-40	14	45665603	45665603	FANCM	Missense	NM_020937	21	c.G5569A	p.V1857M	27.7	29	26	47.3
WES-41	2	58468399	58468399	FANCL	Missense	NM_018062	1	c.C50G	p.P17R	22.2	51	35	40.7
WES-41	14	75514138	75514138	MLH3	Missense	NM_014381	2	c.G2221T	p.V741F	16.6	11	19	63.3
WES-41	17	18188600	18188600	TOP3A	Missense	NM_004618	15	c.A1733C	p.E578A	22.9	97	88	47.6
WES-43	17	18205933	18205933	TOP3A	Missense	NM_004618	6	c.G604A	p.D202N	24.3	22	32	59.3
WES-46	17	41243941	41243941	BRCA1	Nonsense	NM_007294	10	c.C3607T	p.R1203X	35	52	45	46.4
WES-46	14	45658326	45658326	FANCM	Nonsense	NM_020937	20	c.C5101T	p.Q1701X	35	46	47	50.5
WES-47	16	89877182	89877182	FANCA	Missense	NM_001018112	5	c.C455G	p.A152G	13.73	83	47	36.2
WES-48	17	18186148	18186148	TOP3A	Missense	NM_004618	16	c.G1885A	p.E629K	24.5	13	13	50.0
WES-52	3	1.42E+08	1.42E+08	ATR	Missense	NM_001184	30	c.A5257G	p.I1753V	22.5	16	15	48.4
WES-54	8	90993640	90993640	NBN	Missense	NM_002485	3	c.G283A	p.D95N	25.8	26	16	38.1
WES-55	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	9	c.T1229C	p.V410A	23.4	7	9	56.3
WES-55	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	49	c.A7291G	p.K2431E	20.5	6	12	66.7
WES-55	14	75509146	75509146	MLH3	Missense	NM_014381	3	c.C3315A	p.D1105E	16.99	27	26	49.1
WES-55	8	90992986	90992986	NBN	Missense	NM_002485	4	c.G456A	p.M152I	28	7	11	61.1
WES-55	16	3642715	3642715	SLX4	Missense	NM_032444	11	c.G2312C	p.S771T	15.99	22	20	47.6
WES-56	5	1.32E+08	1.32E+08	RAD50	FS Ins	NM_005732	13	c.2157dupA	p.L719fs	36	9	0	0.0

WES-57	16	3641202	3641202	SLX4	Missense	NM_032444	12	c.G2437A	p.E813K	28.2	28	33	54.1
WES-57	12	1.25E+08	1.25E+08	UBC	Missense	NM_021009	2	c.G595A	p.G199R	22.2	74	66	47.1
WES-58	2	2.16E+08	2.16E+08	BARD1	Missense	NM_001282548	4	c.G458A	p.G153E	32	13	6	31.6
WES-58	11	65629482	65629482	MUS81	Missense	NM_025128	4	c.G416A	p.R139Q	17.49	23	29	55.8
WES-59	15	89804921	89804921	FANCI	Missense	NM_018193	5	c.A394G	p.I132V	10.54	43	29	40.3
WES-59	1	28240815	28240815	RPA2	Missense	ENSG00000117748	-	c.G140A	p.G47E	10.34	10	7	41.2
WES-60	15	91328183	91328183	BLM	Nonsense	NM_001287248	14	c.C1570T	p.R524X	36	31	38	55.1
WES-60	17	41222975	41222975	BRCA1	Missense	NM_007298	14	c.G1644A	p.M548I	22.3	16	13	44.8
WES-60	2	58386928	58386928	FANCL	FS Ins	NM_018062	14	c.1099_1100insATTA	p.T367fs	38	11	9	45.0
WES-61	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	34	c.A5116G	p.K1706E	22.1	22	13	37.1
WES-62	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	11	c.T1744C	p.F582L	10.28	15	23	60.5
WES-63	7	7679967	7679967	RPA3	Missense	NM_002947	5	c.T83C	p.V28A	22.8	30	24	44.4
WES-63	17	18194242	18194242	TOP3A	Missense	NM_004618	12	c.G1381A	p.A461T	26	69	36	34.3
WES-63	17	18194248	18194248	TOP3A	Missense	NM_004618	12	c.G1375A	p.D459N	23.4	67	34	33.7
WES-64	9	35076755	35076755	FANCG	Missense	NM_004629	7	c.C890T	p.T297I	21.6	72	80	52.6
WES-64	1	62916290	62916290	USP1	Missense	NM_001017416	9	c.A1996G	p.I666V	18.04	76	46	37.7
WES-66	15	91290633	91290633	BLM	Missense	NM_000057	2	c.T11C	p.V4A	23.4	28	13	31.7
WES-66	5	1.32E+08	1.32E+08	RAD50	Missense	NM_005732	4	c.G379A	p.V127I	24.5	31	27	46.6
WES-67	14	75514138	75514138	MLH3	Missense	NM_014381	2	c.G2221T	p.V741F	16.6	24	21	46.7
WES-68	17	18196087	18196087	TOP3A	Missense	NM_004618	11	c.C1153T	p.P385S	17.31	45	38	45.8
WES-72	3	48506279	48506279	ATRIP	Missense	NM_032166	11	c.G2024A	p.R675Q	22.2	23	11	32.4
WES-73	16	3647893	3647893	SLX4	Missense	NM_032444	6	c.C1271T	p.A424V	27.5	18	20	52.6
WES-74	3	1.42E+08	1.42E+08	ATR	Missense	NM_001184	37	c.A6259G	p.M2087V	21.8	21	7	25.0
WES-74	16	14041741	14041741	ERCC4	Missense	NM_005236	11	c.C2288T	p.P763L	25.7	45	36	44.4
WES-75	3	1.42E+08	1.42E+08	ATR	Missense	NM_001184	25	c.A4405G	p.T1469A	13.92	16	24	60.0
WES-78	14	45628478	45628478	FANCM	Missense	NM_020937	9	c.C1576G	p.L526V	21.1	20	28	58.3
WES-78	7	7678701	7678701	RPA3	Splicing	NM_002947	6	c.C174A	p.P58P	15.76	23	23	50.0

WES-79	9	97912307	97912307	FANCC	Missense	NM_001243743	7	c.A584T	p.D195V	25.2	49	47	49.0
WES-81	14	45644816	45644816	FANCM	Missense	NM_020937	14	c.A2859C	p.K953N	22.2	24	16	40.0
WES-81	16	3633255	3633255	SLX4	Nonsense	NM_032444	14	c.C4996T	p.R1666X	36	40	37	48.1
WES-83	9	35076755	35076755	FANCG	Missense	NM_004629	7	c.C890T	p.T297I	21.6	62	58	48.3
WES-83	2	58392880	58392880	FANCL	Missense	NM_018062	8	c.A670G	p.T224A	23.1	31	26	45.6
WES-84	11	1.08E+08	1.08E+08	ATM	FS Del	NM_000051	19	c.2905delC	p.L969fs	35	41	16	28.1
WES-84	17	18181411	18181411	TOP3A	Missense	NM_004618	18	c.C2405T	p.P802L	14.95	44	40	47.6
WES-85	3	1.42E+08	1.42E+08	ATR	Missense	NM_001184	43	c.C7300G	p.P2434A	22.1	21	21	50.0
WES-86	9	97912307	97912307	FANCC	Missense	NM_001243743	7	c.A584T	p.D195V	25.2	39	40	50.6
WES-86	2	58456962	58456962	FANCL	Missense	NM_018062	3	c.G203C	p.R68P	22.5	8	13	61.9
WES-86	14	45644816	45644816	FANCM	Missense	NM_020937	14	c.A2859C	p.K953N	22.2	21	11	34.4
WES-87	15	31220792	31220792	FAN1	Missense	NM_014967	11	c.A2525G	p.Y842C	11.71	17	6	26.1
WES-88	3	37089130	37089130	MLH1	Missense	NM_001258274	17	c.A1129G	p.K377E	21.8	28	25	47.2
WES-88	3	37089131	37089131	MLH1	Missense	NM_001258274	17	c.A1130C	p.K377T	21.5	28	25	47.2
WES-89	14	45628478	45628478	FANCM	Missense	NM_020937	9	c.C1576G	p.L526V	21.1	14	7	33.3
WES-94	14	45645955	45645955	FANCM	FS Del	NM_020937	14	c.3998delA	p.Q1333fs	22.3	21	20	48.8
WES-94	3	37083774	37083774	MLH1	Nonsense	NM_001258274	16	c.C960G	p.Y320X	41	5	14	73.7
WES-94	17	1787164	1787164	RPA1	Missense	NM_002945	13	c.G1300A	p.G434R	22.9	14	36	72.0
WES-96	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	29	c.A4362C	p.K1454N	18.78	6	33	84.6
WES-97	17	48458123	48458123	EME1	Splicing	NM_001166131	9	c.1576-1G>A	-	20.5	32	28	46.7
WES-99	14	75506696	75506696	MLH3	Missense	NM_014381	5	c.G3488A	p.G1163D	22.7	16	4	20.0
WES-100	22	29090060	29090060	CHEK2	Missense	NM_007194	13	c.G1421A	p.R474H	35	29	33	53.2
WES-101	15	91308570	91308570	BLM	Missense	NM_001287248	9	c.C994T	p.P332S	23	32	29	47.5
WES-101	15	31200357	31200357	FAN1	FS Ins	NM_001146094	3	c.1271_1272insTAAAT	p.R424fs	36	17	14	45.2
WES-201	16	14015897	14015897	ERCC4	Missense	NM_005236	2	c.A217G	p.I73V	22.2	34	40	54.1
WES-201	15	89807213	89807213	FANCI	Missense	NM_018193	8	c.G625C	p.E209Q	25.8	100	84	45.7
WES-202	14	45606290	45606290	FANCM	Missense	NM_020937	2	c.C527T	p.T176I	20.2	34	25	42.4

WES-202	5	1.32E+08	1.32E+08	RAD50	Missense	NM_005732	7	c.G980A	p.R327H	24	51	30	37.0
WES-203	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	28	c.C4138T	p.H1380Y	13.39	45	32	41.6
WES-203	17	79514028	79514028	C17orf70	Missense	NM_025161	5	c.G2080C	p.A694P	12.9	8	2	20.0
WES-203	9	35076755	35076755	FANCG	Missense	NM_004629	7	c.C890T	p.T297I	21.6	33	27	45.0
WES-203	2	58459236	58459236	FANCL	Missense	NM_018062	2	c.C108G	p.F36L	22.4	43	21	32.8
WES-203	5	1.32E+08	1.32E+08	RAD50	Missense	NM_005732	14	c.A2264G	p.Q755R	22.3	25	15	37.5
WES-205	16	89877455	89877455	FANCA	Missense	NM_001018112	4	c.C308T	p.S103L	15.16	43	27	38.6
WES-206	3	1.42E+08	1.42E+08	ATR	Splicing	NM_001184	7	c.1350-4A>G	-	11.54	25	28	52.8
WES-206	3	10074646	10074646	FANCD2	Missense	NM_033084	3	c.G195C	p.Q65H	19.74	47	58	55.2
WES-207	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	39	c.T5793C	p.A1931A	18.98	30	32	51.6
WES-207	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	53	c.C7919T	p.T2640I	17.95	54	52	49.1
WES-207	17	41222975	41222975	BRCA1	Missense	NM_007298	14	c.G1644A	p.M548I	22.3	36	38	51.4
WES-207	17	79514028	79514028	C17orf70	Missense	NM_025161	5	c.G2080C	p.A694P	12.9	6	3	33.3
WES-207	17	18194242	18194242	TOP3A	Missense	NM_004618	12	c.G1381A	p.A461T	26	16	12	42.9
WES-207	17	18194248	18194248	TOP3A	Missense	NM_004618	12	c.G1375A	p.D459N	23.4	17	14	45.2
WES-208	16	14041741	14041741	ERCC4	Missense	NM_005236	11	c.C2288T	p.P763L	25.7	36	29	44.6
WES-209	16	89836383	89836383	FANCA	Missense	NM_000135	26	c.T2366C	p.V789A	22.8	26	15	36.6
WES-209	14	45667953	45667953	FANCM	Missense	NM_020937	22	c.C5823T	p.T1941T	19.21	52	40	43.5
WES-209	14	75514138	75514138	MLH3	Missense	NM_014381	2	c.G2221T	p.V741F	16.6	100	77	43.5
WES-209	12	1.25E+08	1.25E+08	UBC	Splicing	NM_021009	2	c.G504A	p.E168E	17.08	72	13	15.3
WES-211	14	75514138	75514138	MLH3	Missense	NM_014381	2	c.G2221T	p.V741F	16.6	128	106	45.3
WES-211	5	1.32E+08	1.32E+08	RAD50	Missense	NM_005732	17	c.C2750T	p.T917I	22.1	75	69	47.9
WES-211	17	18188537	18188537	TOP3A	Missense	NM_004618	15	c.A1796G	p.K599R	17.34	40	48	54.5
WES-213	17	79514675	79514675	C17orf70	Missense	NM_025161	5	c.A1433G	p.Q478R	25	4	8	66.7
WES-215	3	48506305	48506305	ATRIP	Missense	NM_032166	11	c.G2050A	p.D684N	11.46	7	7	50.0
WES-215	15	31206173	31206173	FAN1	Missense	NM_014967	5	c.G1690A	p.A564T	26	19	34	64.2
WES-215	11	94197365	94197365	MRE11A	Missense	NM_005591	11	c.G1139A	p.R380H	22.2	83	82	49.7

WES-216	2	58388668	58388670	FANCL	Missense	NM_018062	12	c.1007_1009del	p.336_337del	22.8	56	54	49.1
WES-217	17	18194248	18194248	TOP3A	Missense	NM_004618	12	c.G1375A	p.D459N	23.4	7	10	58.8
WES-218	3	37089130	37089130	MLH1	Missense	NM_001258274	17	c.A1129G	p.K377E	21.8	94	77	45.0
WES-218	3	37089131	37089131	MLH1	Missense	NM_001258274	17	c.A1130C	p.K377T	21.5	95	79	45.4
WES-218	1	62916282	62916282	USP1	Missense	NM_001017416	9	c.T1988C	p.V663A	20.7	72	93	56.4
WES-218	14	1.04E+08	1.04E+08	XRCC3	Missense	NM_005432	4	c.A37C	p.I13L	15.46	107	96	47.3
WES-219	14	68290310	68290310	RAD51B	Missense	NM_133510	2	c.A50G	p.D17G	23	22	21	48.8
WES-221	3	1.42E+08	1.42E+08	ATR	Missense	NM_001184	4	c.A992G	p.D331G	12.83	78	69	46.9
WES-221	18	20581624	20581624	RBBP8	Missense	NM_002894	15	c.T2219C	p.L740S	22.2	102	93	47.7
WES-222	16	23614892	23614892	PALB2	Missense	NM_024675	13	c.T3449G	p.L1150R	15.08	65	54	45.4
WES-223	17	79508395	79508395	C17orf70	Missense	NM_025161	8	c.G2454C	p.Q818H	13.06	15	19	55.9
WES-223	17	79517360	79517360	C17orf70	Missense	NM_025161	3	c.C1160T	p.P387L	13.14	6	6	50.0
WES-223	17	18194242	18194242	TOP3A	Missense	NM_004618	12	c.G1381A	p.A461T	26	26	13	33.3
WES-223	17	18194248	18194248	TOP3A	Missense	NM_004618	12	c.G1375A	p.D459N	23.4	23	12	34.3
WES-224	9	97864024	97864024	FANCC	Nonsense	NM_001243743	15	c.C1642T	p.R548X	37	68	40	37.0
WES-225	15	89807836	89807836	FANCI	Splicing	NM_018193	9	c.C753T	p.D251D	14.29	41	36	46.8
WES-226	14	75514138	75514138	MLH3	Missense	NM_014381	2	c.G2221T	p.V741F	16.6	121	110	47.6
WES-226	16	23652433	23652433	PALB2	Missense	NM_024675	1	c.A46G	p.K16E	22.3	15	17	53.1
WES-226	14	68353893	68353893	RAD51B	Missense	NM_133510	7	c.A728G	p.K243R	27.1	54	40	42.6
WES-227	14	68353893	68353893	RAD51B	Missense	NM_133510	7	c.A728G	p.K243R	27.1	41	20	32.8
WES-227	16	3640664	3640664	SLX4	Missense	NM_032444	12	c.G2975A	p.G992E	16.23	26	23	46.9
WES-227	17	18194242	18194242	TOP3A	Missense	NM_004618	12	c.G1381A	p.A461T	26	22	12	35.3
WES-227	17	18194248	18194248	TOP3A	Missense	NM_004618	12	c.G1375A	p.D459N	23.4	21	10	32.3
WES-231	2	58388659	58388659	FANCL	Missense	NM_018062	12	c.G1018A	p.E340K	22.5	89	60	40.3
WES-231	14	45628478	45628478	FANCM	Missense	NM_020937	9	c.C1576G	p.L526V	21.1	41	42	50.6
WES-235	3	48506912	48506912	ATRIP	Missense	NM_032166	12	c.G2254A	p.A752T	15.91	21	13	38.2
WES-235	17	79517729	79517729	C17orf70	Missense	NM_025161	3	c.C791T	p.A264V	12.51	25	19	43.2

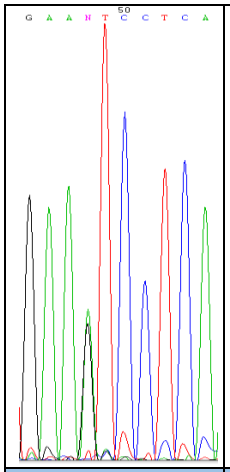
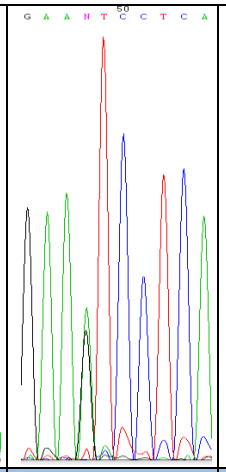
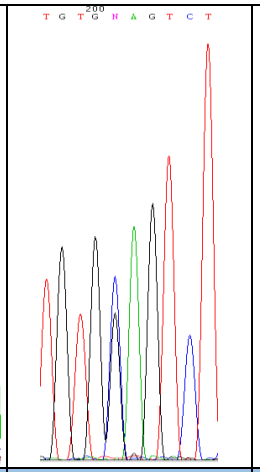
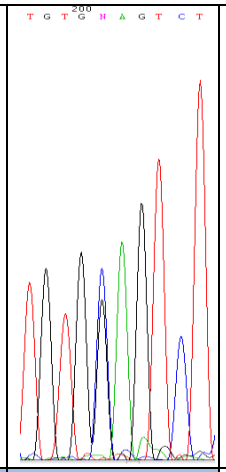
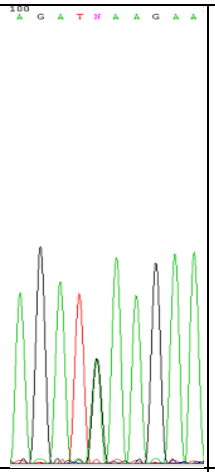
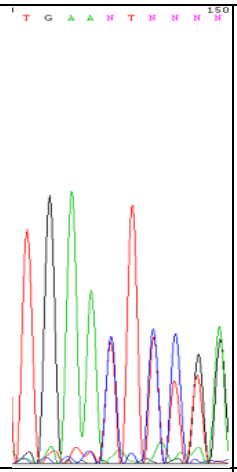
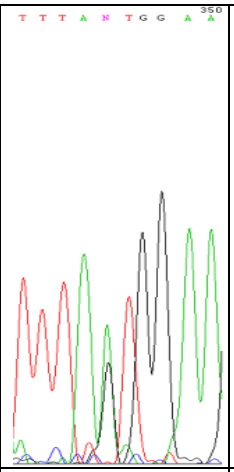
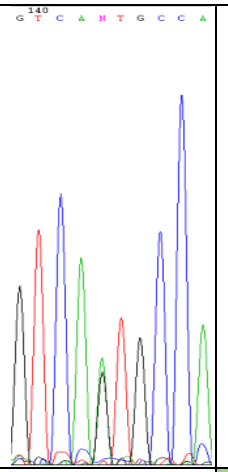
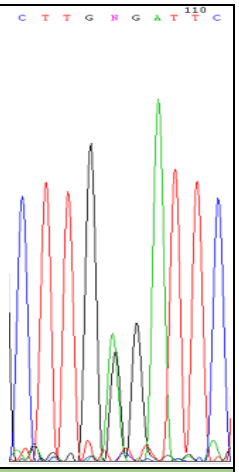
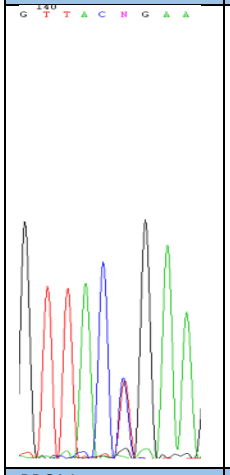
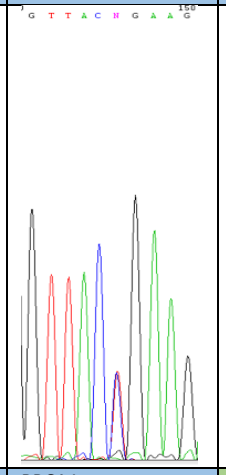
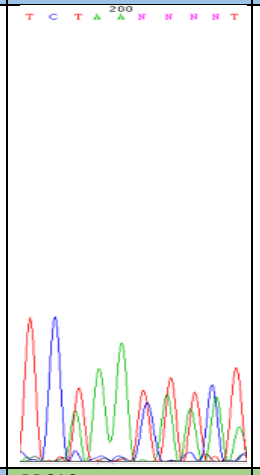
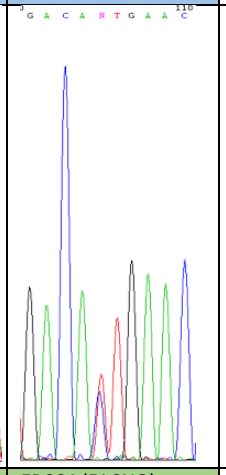
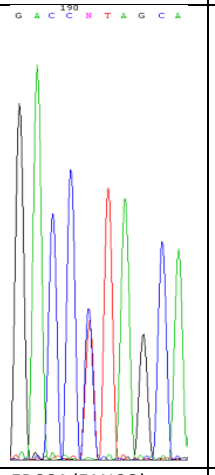
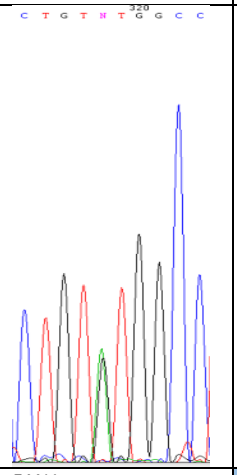
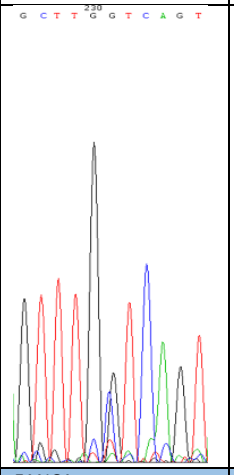
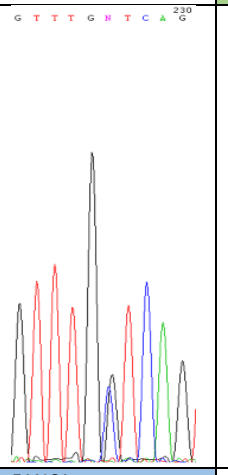
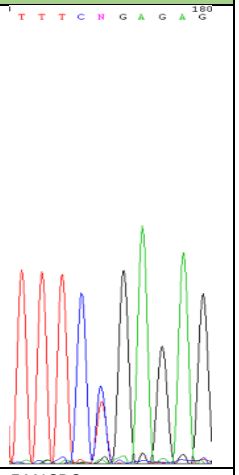
WES-235	15	31197584	31197584	FAN1	Missense	NM_001146094	2	c.G718A	p.E240K	12.62	51	40	44.0
WES-236	13	32907277	32907277	BRCA2	Missense	NM_000059	10	c.T1662G	p.C554W	22.9	55	39	41.5
WES-237	17	79517729	79517729	C17orf70	Missense	NM_025161	3	c.C791T	p.A264V	12.51	17	8	32.0
WES-239	5	68709909	68709909	RAD17	Missense	NM_133340	16	c.C1308G	p.D436E	16.05	26	23	46.9
WES-245	11	1.08E+08	1.08E+08	ATM	Missense	NM_000051	8	c.C998T	p.S333F	25.1	80	79	49.7
WES-245	13	32914592	32914592	BRCA2	Missense	NM_000059	11	c.C6100T	p.R2034C	22.2	58	64	52.5
WES-246	14	75514138	75514138	MLH3	Missense	NM_014381	2	c.G2221T	p.V741F	16.6	105	88	45.6
WES-247	6	35427531	35427531	FANCE	Missense	NM_021922	7	c.T1310C	p.M437T	23.4	25	29	53.7
WES-248	15	91290633	91290633	BLM	Missense	NM_000057	2	c.T11C	p.V4A	23.4	48	48	50.0
WES-249	15	31197584	31197584	FAN1	Missense	NM_001146094	2	c.G718A	p.E240K	12.62	81	99	55.0
WES-249	3	10115047	10115047	FANCD2	Splicing	NM_033084	28	c.2715+1G>A		27	118	108	47.8
WES-249	3	37061870	37061870	MLH1	Missense	NM_001258274	12	c.C231T	p.H77H	15.22	97	66	40.5
WES-250	15	31197584	31197584	FAN1	Missense	NM_001146094	2	c.G718A	p.E240K	12.62	47	45	48.9
WES-250	3	10081411	10081411	FANCD2	Missense	NM_033084	9	c.A577G	p.T193A	21.9	41	29	41.4
WES-250	11	65631970	65631970	MUS81	Missense	NM_025128	11	c.C1062G	p.F354L	16.19	30	21	41.2

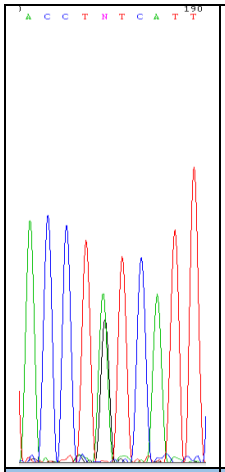
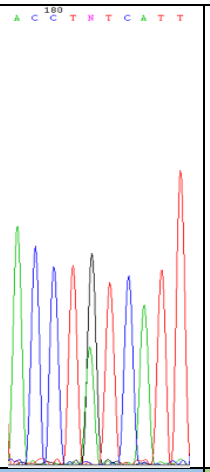
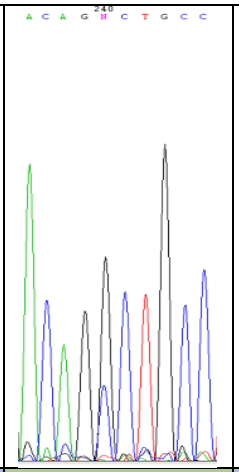
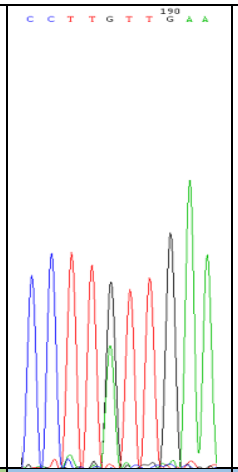
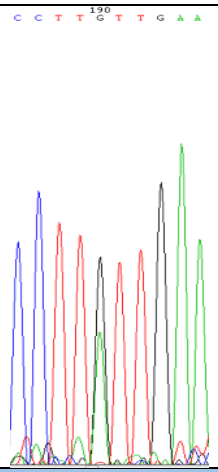
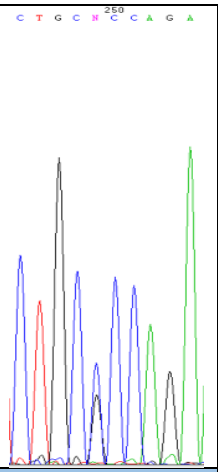
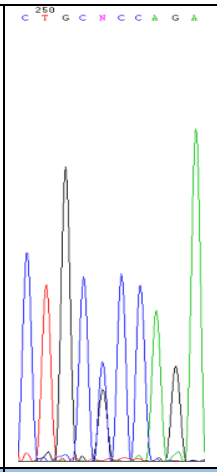
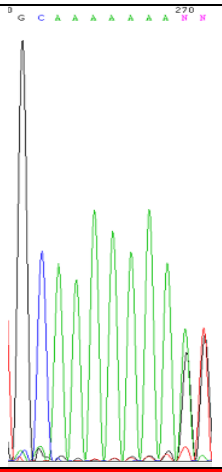
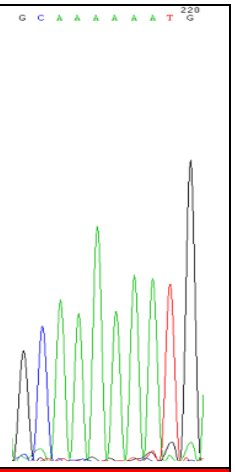
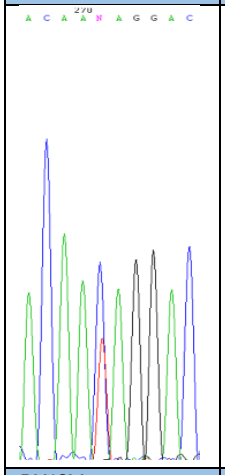
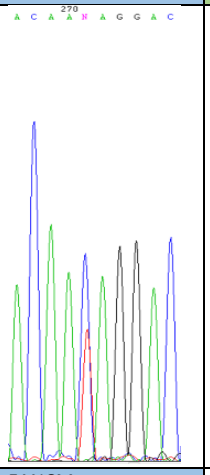
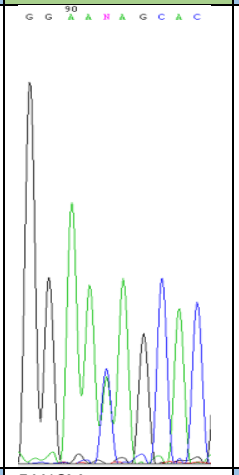
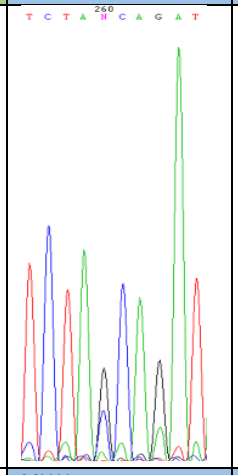
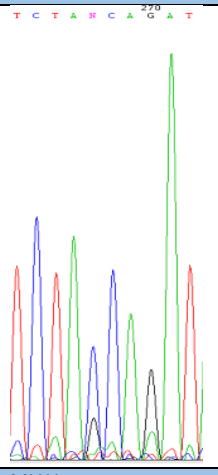
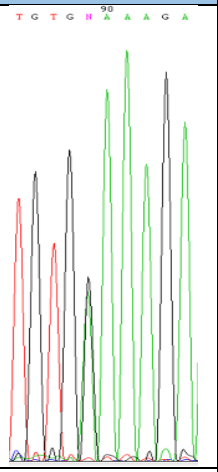
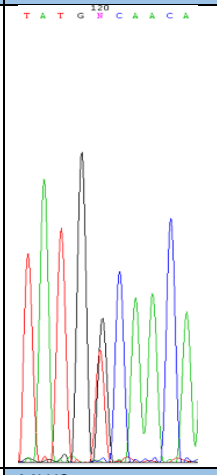
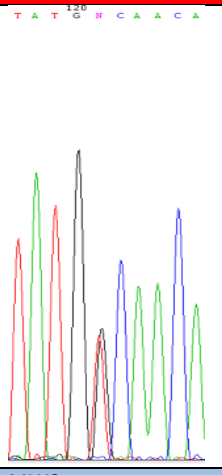
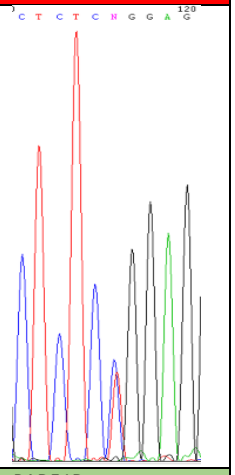
*Splicing mutations where the exon is not annotated, the genomic coordinate of the variant were listed.

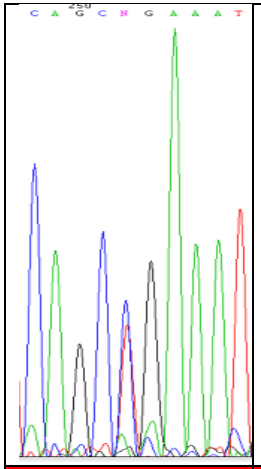
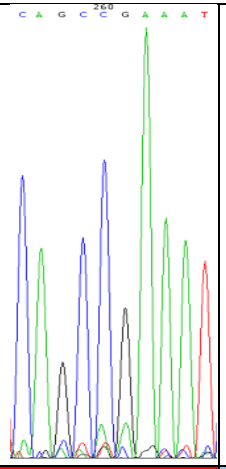
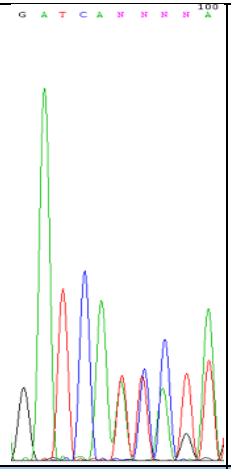
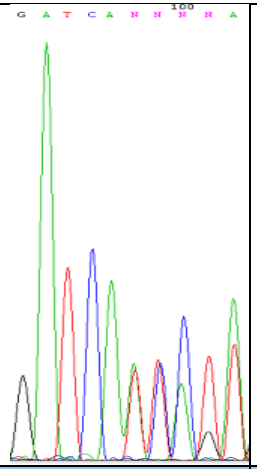
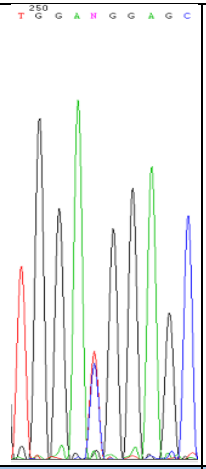
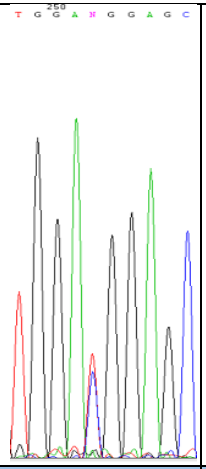
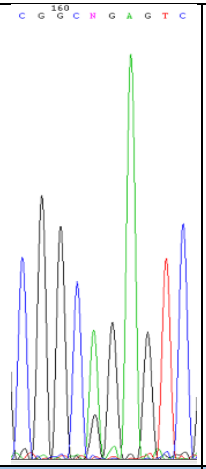
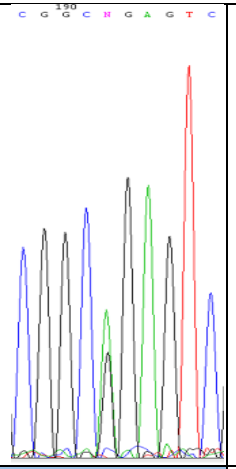
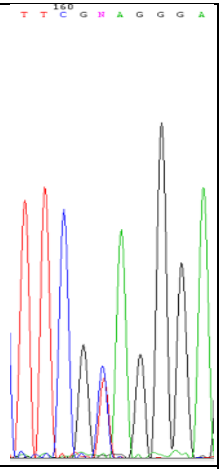
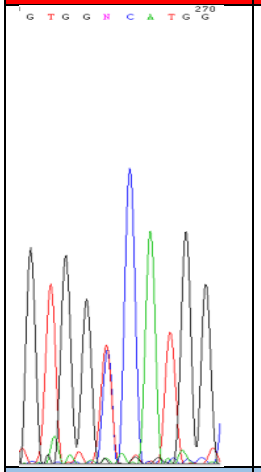
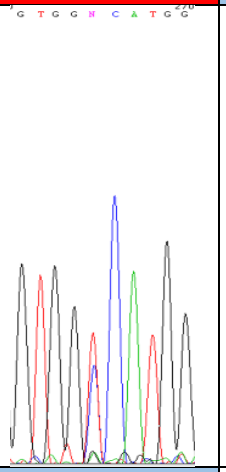
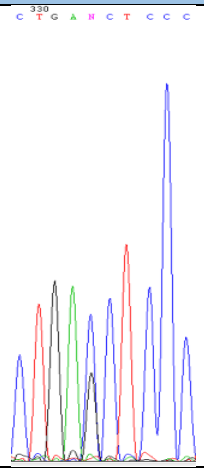
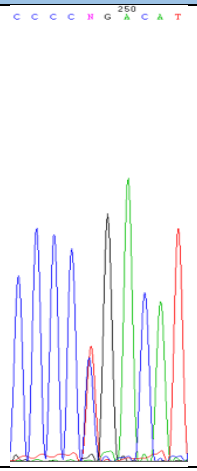
VAF – Variant allele frequency

Appendix D. Sanger validation of FA/BRCA-HRR variants.

Results from the Sanger sequencing of gDNA of disease and non-disease (when available) samples are shown as sequencing traces. Cells that are shaded in Blue represent paired (diagnosis BMMNC (Dx) and MSC) traces; cells that are shaded in Green represent MSC only samples; and somatically acquired mutations are shown in Red.

								
ATM	ATM	ATM	ATM	ATM	ATM	ATR	ATR	BARD1
c.G7618A (DX)	c.G7618A (MSC)	c.G6860C (DX)	c.G6860C (MSC)	c.A5116G (DX)	c.2905delC (DX)	c.A2437G (DX)	c.A6259G (DX)	c.G1868A (MSC)
WES-2	WES-2	WES-26	WES-26	WES-61	WES-84	WES-34	WES-74	WES-58
								
BRCA1	BRCA1	BRCA2	ERCC4 (FACNQ)	ERCC4 (FANCQ)	FAN1	FANCA	FANCA	FANCD2
c.C3607T (DX)	c.C3607T (MSC)	c.6399_6401del (MSC)	p.I706T (MSC)	c.C2288T (DX)	c.A2525G (DX)	c.C455G (DX)	c.C455G (MSC)	c.C2776T (DX)
WES-46	WES-46	WES-32	WES-34	WES-74	WES-87	WES-47	WES-47	WES-21

								
FANCI	FANCI	FANCI	FANCL	FANCL	FANCL	FANCL	FANCM	FANCM
c.A394G (DX)	c.A394G (MSC)	c.C1656G (MSC)	c.A343G (DX)	c.A343G (MSC)	c.C50G (DX)	c.C50G (MSC)	c.3998delA (DX)	c.3998delA (MSC)
WES-59	WES-59	WES-1	WES-18	WES-18	WES-41	WES-41	WES-94	WES-94
								
FANCM	FANCM	FANCM	MLH1	MLH1	MLH3	MLH3	MLH3	RAD51D
c.C5101T (DX)	c.C5101T (MSC)	c.A1545C (DX)	c.C960G (DX)	c.C960G (MSC)	c.A1234G (DX)	c.T3173G (DX)	c.T3173G (MSC)	c.C553T (MSC)
WES-46	WES-46	WES-9	WES-94	WES-94	WES-14	WES-27	WES-27	WES-14

								
NBN	NBN	RMI1	RMI1	RPA1	RPA1	RPA1	RPA1	RPA3
c.C127T (DX)	c.C127T (MSC)	c.1419_1434del (DX)	c.1419_1434del (MSC)	c.C1165T (DX)	c.C1165T (MSC)	c.G1300A (DX)	c.G1300A (MSC)	c.T83C (DX)
WES-31	WES-31	WES-4	WES-4	WES-4	WES-4	WES-94	WES-94	WES-63
								
SLX4 (FANCP)	SLX4 (FANCP)	SLX4 (FANCP)	SLX4 (FANCP)					
c.C1271T (DX)	c.C1271T (MSC)	c.G2312C (DX)	c.C4996T (DX)					
WES-73	WES-73	WES-55	WES-81					

Appendix E. The 43 recurrently mutated genes in AML as reported by TCGA

ASXL1	PRPF3
BRINP3	PRPF8
CEBPA	PTPN11
CSTF2T	RAD21
DDX1	RBMX
DDX23	RUNX1
DHX32	SF3B1
DNMT1	SMC1A
DNMT3A	SMC3
DNMT3B	SRSF6
EZH2	STAG2
FLT3	SUPT5H
HNRNPK	TET1
IDH1	TET2
IDH2	TET3
KIT	TP53
KRAS	TRA2B
METTL3	U2AF1
NPM1	U2AF1L4
NRAS	U2AF2
PHF6	WT1
PLRG1	

Appendix F. Mutations in the 43 recurrently mutated AML genes

Gene Name	Ref seq	Nucleotide change	Amino Acid Change	Type of Mutation	WES-ID	COSMIC Somatic Status	Comments	*Classification
ASXL1	NM_015338	c.G1205A	p.R402Q	missense	WES-235	Somatic	-	Exact missense
ASXL1	NM_015338	c.1926dupA	p.G642fs	fs ins	WES-35	-	3x Nonsense mutations reported	Truncation reported at codon
ASXL1	NM_015338	c.1927dupG	p.G642fs	fs ins	WES-29, WES-40, WES-47, WES-70, WES-72, WES-206	-	3x Nonsense mutations reported	Truncation reported at codon
ASXL1	NM_015338	c.C1896A	p.C632X	nonsense	WES-78	Somatic	-	Exact truncation
ASXL1	NM_015338	c.C2077T	p.R693X	nonsense	WES-12, WES-224	Somatic	-	Exact truncation
ASXL1	NM_015338	c.C2338T	p.Q780X	nonsense	WES-14	Somatic	-	Exact truncation
ASXL1	NM_015338	c.G3306T	p.E1102D	missense	WES-49, WES-82	Somatic	-	Exact missense
CEBPA	NM_001287435	c.G976A	p.G326S	missense	WES-52	Somatic	-	Exact missense
CEBPA	NM_001287435	c.867_876del	p.A289fs	fs del	WES-29	Unknown	-	Exact truncation
DNMT3A	NM_022552	c.G2645A	p.R882H	missense	WES-10, WES-16, WES-21, WES-33, WES-35, WES-41, WES-43, WES-47, WES-54, WES-6, WES-64, WES-68, WES-73, WES-78, WES-235, WES-214, WES-230, WES-241, WES-211, WES-240, WES-213, WES-232, WES-201, WES-245, WES-228, WES-206	Somatic	-	Exact missense
DNMT3A	NM_022552	c.C2644A	p.R882S	missense	WES-219	Somatic	-	Exact missense
DNMT3A	NM_022552	c.C2644T	p.R882C	missense	WES-45, WES-69, WES-237, WES-215, WES-216, WES-212, WES-209, WES-219	Somatic	-	Exact missense
DNMT3A	NM_022552	c.C2644T	p.R882C	missense	WES-45, WES-69, WES-237, WES-215, WES-216, WES-212, WES-209	Somatic	-	Exact missense
DNMT3A	NM_022552	c.G2186A	p.R729Q	missense	WES-25	Somatic	-	Exact missense
DNMT3A	NM_022552	c.2099delC	p.P700fs	fs del	WES-66	-	-	Truncation not reported
DNMT3A	NM_022552	c.C2074T	p.Q692X	nonsense	WES-91	-	-	Truncation not reported
DNMT3A	NM_022552	c.1902_1906del	p.I634fs	fs del	WES-81	-	-	Truncation not reported

DNMT3A	NM_022552	c.C1717T	p.Q573X	nonsense	WES-34, WES-39	-	-	Truncation not reported
DNMT3A	NM_022552	c.879_880insTGGG	p.E294fs	fs ins	WES-32	-	-	Truncation not reported
EZH2	NM_004456	c.195_196insGATA	p.Q66fs	fs ins	-	-		Truncation not reported
EZH2	NM_004456	c.60dupA	p.S21fs	fs ins	-	-		Truncation not reported
EZH2	NM_004456	c.217_218insTC	p.S73fs	fs ins	WES-201	Somatic	-	Exact truncation
HNRNPK	NM_031262	c.1297delC	p.R433fs	fs del	WES-209	-	-	Truncation not reported
HNRNPK	NM_031262	c.1176_1182del	p.V392fs	fs del	WES-50	-	-	Truncation not reported
IDH1	NM_001282387	c.G395A	p.R132H	missense	WES-30, WES-33, WES-43, WES-80,WES-203	Somatic	-	Exact missense
IDH1	NM_001282387	c.C394T	p.R132C	missense	WES-12, WES-16, WES-38, WES-40, WES-62, WES-7, WES-81, WES-89, WES-218, WES-228, WES-206	Somatic	snp:209113113	Exact missense
IDH1	NM_001282387	c.C394T	p.R132C	missense	WES-12, WES-16, WES-38, WES-40, WES-62, WES-81, WES-89, WES-228, WES-206	Somatic	snp:209113113	Exact missense
IDH1	NM_001282387	c.C394G	p.R132G	missense	WES-218	Somatic	snp:209113113	Exact missense
IDH1	NM_001282387	c.C394A	p.R132S	missense	WES-7	Somatic	snp:209113113	Exact missense
IDH2	NM_001289910	c.G263A	p.R88Q	missense	WES-100, WES-14, WES-22, WES-25, WES-31, WES-37, WES-47, WES-53, WES-59, WES-6, WES-65, WES-66, WES-68, WES-69, WES-73, WES-98, WES-225, WES-216, WES-217	Somatic	-	Exact missense
IDH2	NM_001289910	c.G359A	p.R120K	missense	WES-15, WES-17, WES-35, WES-41, WES-245	Somatic	-	Exact missense
KIT	NM_001093772	c.G2434T	p.D812Y	missense	-	Somatic	-	Exact missense
KIT	NM_001093772	c.A2435T	p.D812V	missense	WES-55	Somatic	-	Exact missense
KIT	NM_001093772	c.G1195A	p.V399I	missense	WES-210	Somatic	-	Exact missense
KIT	NM_000222	c.A1621C	p.M541L	missense	-	Somatic	Many controls and Patients have this variant	Exact missense
KRAS	NM_004985	c.G35T	p.G12V	missense	WES-57, WES-67, WES-216, WES-217	Somatic	snp:25398284	Exact missense
KRAS	NM_004985	c.G35T	p.G12V	missense	WES-216, WES-217	Somatic	snp:25398284	Exact missense
KRAS	NM_004985	c.G35A	p.G12D	missense	WES-57, WES-67	Somatic	snp:25398284	Exact missense

NPM1	NM_002520	c.859_860insTCTG	p.L287fs	fs ins	WES-100, WES-22, WES-25, WES-27, WES-33, WES-38, WES-4, WES-43, WES-44, WES-45, WES-46, WES-54, WES-57, WES-6, WES-64, WES-65, WES-7, WES-70, WES-73, WES-75, WES-80, WES-85, WES-89, WES-202, WES-235, WES-237, WES-214, WES-230, WES-241, WES-215, WES-236, WES-211, WES-240, WES-216, WES-213, WES-212, WES-209, WES-229, WES-248, WES-99, WES-203, WES-205	-	-	Exact truncation
NPM1	NM_002520	c.860_861insCTGC	p.L287fs	fs ins	WES-93, WES-208, WES-242, WES-249	-	-	Exact truncation
NPM1	NM_002520	c.861_862insTGCA	p.L287fs	fs ins	WES-10, WES-51, WES-91, WES-98, WES-218	-	indel:170837545	Exact truncation
NPM1	NM_002520	c.861_862insTGCA	p.L287fs	fs ins	WES-10, WES-91, WES-98	-	indel:170837545	Exact truncation
NPM1	NM_002520	c.861_862insTGCT	p.L287fs	fs ins	WES-51, WES-218	-	indel:170837545	Exact truncation
NPM1	NM_002520	c.862_863insGTCA	p.W288fs	fs ins	WES-34, WES-49	-	4 Nucleotides inserted are different (indel:170837546)	Truncation reported at codon
NPM1	NM_002520	c.862_863insGTCA	p.W288fs	fs ins	WES-34	-	4 Nucleotides inserted are different (indel:170837546)	Truncation reported at codon
NPM1	NM_002520	c.862_863insGCCA	p.W288fs	fs ins	WES-49	-	4 Nucleotides inserted are different (indel:170837546)	Truncation reported at codon
NPM1	NM_002520	c.863_864insTCAC	p.W288fs	fs ins	WES-26, WES-86	-	4 Nucleotides inserted are different (indel:170837547)	Truncation reported at codon
NPM1	NM_002520	c.863_864insTCAC	p.W288fs	fs ins	WES-26	-	4 Nucleotides inserted are different (indel:170837547)	Truncation reported at codon
NPM1	NM_002520	c.863_864insTCGC	p.W288fs	fs ins	WES-86	-	4 Nucleotides inserted are different (indel:170837547)	Truncation reported at codon
NRAS	NM_002524	c.A183T	p.Q61H	missense	WES-98	Somatic	-	Exact missense
NRAS	NM_002524	c.A182G	p.Q61R	missense	WES-101	Somatic	-	Exact missense
NRAS	NM_002524	c.C176A	p.A59D	missense	WES-81	Somatic	-	Exact missense
NRAS	NM_002524	c.G175A	p.A59T	missense	WES-202	Somatic	-	Exact missense
NRAS	NM_002524	c.G38T	p.G13V	missense	WES-39, WES-56, WES-242	Somatic	snp:115258744	Exact missense
NRAS	NM_002524	c.G38T	p.G13V	missense	WES-39	Somatic	snp:115258744	Exact missense
NRAS	NM_002524	c.G38A	p.G13D	missense	WES-56, WES-242	Somatic	snp:115258744	Exact missense

NRAS	NM_002524	c.G37T	p.G13C	missense	WES-226	Somatic	-	Exact missense
NRAS	NM_002524	c.G35A	p.G12D	missense	WES-89, WES-238, WES-249	Somatic	-	Exact missense
NRAS	NM_002524	c.G34T	p.G12C	missense	WES-13, WES-56, WES-69	Somatic	snp:115258748	Exact missense
NRAS	NM_002524	c.G34T	p.G12C	missense	WES-56	Somatic	snp:115258748	Exact missense
NRAS	NM_002524	c.G34A	p.G12S	missense	WES-13, WES-69	Somatic	snp:115258748	Exact missense
PHF6	NM_032458	c.864_865insCC	p.A288fs	fs ins	WES-201	-	-	Truncation not reported
PHF6	NM_032458	c.902_903insG	p.Y301_H302delinsX	nonsense	WES-78	-	-	Exact truncation
PHF6	NM_032458	c.903_904insCT	p.Y301fs	fs ins	WES-78	-	-	Exact truncation
PTPN11	NM_002834	c.G181T	p.D61Y	missense	WES-231	Somatic	-	Exact missense
PTPN11	NM_002834	c.A182T	p.D61V	missense	WES-229	Somatic	-	Exact missense
PTPN11	NM_002834	c.G1508A	p.G503E	missense	WES-32, WES-60, WES-211	Somatic	snp:112926888	Exact missense
PTPN11	NM_002834	c.G1508A	p.G503E	missense	WES-32	Somatic	snp:112926888	Exact missense
PTPN11	NM_002834	c.G1508C	p.G503A	missense	WES-211	Somatic	snp:112926888	Exact missense
PTPN11	NM_002834	c.G1530C	p.Q510H	missense	WES-239	Somatic	-	Exact missense
RAD21	NM_006265	c.C1432T	p.R478X	nonsense	WES-214	Somatic	-	Exact truncation
RUNX1	NM_001122607	c.G511A	p.D171N	missense	-	Somatic	-	Exact missense
RUNX1	NM_001001890	c.C877T	p.R293X	nonsense	WES-81	Somatic	-	Exact truncation
RUNX1	NM_001001890	c.C780A	p.Y260X	nonsense	WES-77	-	-	Truncation reported at codon
RUNX1	NM_001122607	c.666_667insTC	p.R223fs	fs ins	WES-5	-	-	Truncation reported at codon
RUNX1	NM_001122607	c.627dupG	p.R210fs	fs ins	WES-50	-	-	Truncation reported at codon
RUNX1	NM_001122607	c.548delT	p.L183fs	fs del	WES-76	-	-	Truncation not reported
RUNX1	NM_001122607	c.G530A	p.R177Q	missense	WES-87	Somatic	-	Exact missense
RUNX1	NM_001122607	c.G521A	p.R174Q	missense	WES-47, WES-5	Somatic	-	Exact missense
RUNX1	NM_001122607	c.C520T	p.R174X	nonsense	WES-78	Somatic	-	Exact truncation
RUNX1	NM_001122607	c.G416A	p.R139Q	missense	WES-35, WES-66	Somatic	-	Exact missense
RUNX1	NM_001122607	c.G404A	p.R135K	missense	WES-37	Somatic	-	Exact missense

RUNX1	NM_001122607	c.C341T	p.S114L	missense	WES-40, WES-201	Somatic	-	Exact missense
RUNX1	NM_001122607	c.C341T	p.S114L	missense	WES-40	Somatic	-	Exact missense
RUNX1	NM_001122607	c.G239A	p.R80H	missense	WES-12	Somatic	-	Exact missense
RUNX1	NM_001122607	c.T86C	p.L29S	missense	WES-4, WES-94	Somatic	-	Exact missense
SMC1A	NM_006306	c.2571_2574del	p.Q857fs	fs del	WES-17	-	-	Truncation not reported
SMC1A	NM_006306	c.2563_2566del	p.E855fs	fs del	WES-17	-	-	Truncation reported at codon
SMC1A	NM_006306	c.1285_1286insCCCCG	p.E429fs	fs ins	WES-37	-	-	Truncation not reported
SMC1A	NM_006306	c.1280_1283del	p.E427fs	fs del	WES-37	-	-	Truncation not reported
SMC1A	NM_006306	c.1276_1277insAGCAA	p.R426fs	fs ins	WES-37	-	-	Truncation not reported
STAG2	NM_001042751	c.C646T	p.R216X	nonsense	WES-14	Somatic	-	Exact truncation
STAG2	NM_001042751	c.C1018T	p.Q340X	nonsense	WES-215, WES-243, WES-218, WES-209, WES-228, WES-220, WES-207	-	-	Truncation not reported
STAG2	NM_001042751	c.1385_1386insACTT	p.K462fs	fs ins	WES-238	-	-	Truncation not reported
STAG2	NM_001042751	c.C3173A	p.S1058X	nonsense	WES-14	-	-	Truncation not reported
STAG2	NM_001042751	c.C3349T	p.Q1117X	nonsense	WES-229	-	-	Truncation not reported
TET2	NM_017628	c.C2440T	p.R814C	missense	-	Somatic	-	Exact missense
TET2	exon6			splicing	-	Somatic	-	Exact missense
TET2	NM_017628	c.232dupT	p.D77fs	fs ins	WES-21	-	-	Truncation not reported
TET2	NM_017628	c.495delT	p.S165fs	fs del	WES-46	-	-	Truncation not reported
TET2	NM_001127208	c.T4884G	p.Y1628X	nonsense	WES-205	Somatic	-	Exact truncation
TET2	NM_017628	c.759_762del	p.N253fs	fs del	WES-39	-	-	Truncation not reported
TET2	NM_017628	c.1208delA	p.Q403fs	fs del	WES-20	-	-	Truncation not reported
TET2	NM_017628	c.1575_1576del	p.L525fs	fs del	WES-32	-	-	Truncation not reported
TET2	NM_017628	c.1950_1968del	p.H650fs	fs del	WES-219	-	-	Truncation not reported
TET2	NM_017628	c.C2272T	p.Q758X	nonsense	WES-85	Somatic	-	Exact truncation
TET2	NM_017628	c.2399_2408del	p.H800fs	fs del	WES-93	-	-	Truncation not reported

TET2	NM_017628	c.T2599C	p.Y867H	missense	WES-37, WES-42, WES-66, WES-67, WES-78, WES-226	Somatic	-	Exact missense
TET2	NM_017628	c.C2737T	p.Q913X	nonsense	WES-21	Somatic	-	Exact truncation
TET2	NM_001127208	c.G3593A	p.W1198X	nonsense	WES-240	-	-	Truncation reported at codon
TET2	NM_001127208	c.C3646T	p.R1216X	nonsense	WES-208	Somatic	-	Exact truncation
TET2	NM_001127208	c.3812dupG	p.C1271fs	fs ins	WES-75	Somatic*	-	Exact truncation
TET2	NM_001127208	c.3822delG	p.Q1274fs	fs del	WES-29	Somatic*	-	Exact truncation
TET2	NM_001127208	c.G3845A	p.G1282D	missense	WES-221	Somatic	-	Exact missense
TET2	NM_001127208	c.4076_4077insTC	p.R1359fs	fs ins	WES-224	-	-	Truncation reported at codon
TET2	NM_001127208	c.4097delG	p.R1366fs	fs del	WES-85	-	-	Truncation reported at codon
TET2	NM_001127208	c.4220_4221del	p.G1407fs	fs del	WES-55	-	-	Truncation not reported
TET2	NM_001127208	c.4312dupA	p.E1437fs	fs ins	WES-92	-	-	Truncation not reported
TET2	NM_001127208	c.C4519T	p.Q1507X	nonsense	WES-45	Somatic	-	Exact truncation
TET2	NM_001127208	c.4588_4589insCT	p.P1530fs	fs ins	WES-44	-	-	Truncation not reported
TET2	NM_001127208	c.C4820G	p.S1607X	nonsense	WES-6	-	-	Truncation not reported
TET2	NM_001127208	c.5230delC	p.L1744X	nonsense	WES-217	-	-	Truncation not reported
TET2	NM_001127208	c.C5347T	p.Q1783X	nonsense	WES-215	-	-	Truncation not reported
TET2	NM_001127208	c.5544_5545del	p.S1848fs	fs del	WES-29	-	-	Truncation not reported
TP53	NM_001126116	c.G389T	p.G130V	missense	-	Somatic	-	Exact missense
TP53	NM_001126113	c.G853A	p.E285K	missense	WES-39	Somatic	-	Exact missense
TP53	NM_001126113	c.G818A	p.R273H	missense	WES-52	Somatic	-	Exact missense
TP53	NM_001126113	c.G713A	p.C238Y	missense	WES-97	Somatic	-	Exact missense
TP53	NM_001126113	c.A659G	p.Y220C	missense	WES-36, WES-79	Somatic	-	Exact missense
U2AF1	NM_001025204	c.A251G	p.Q84R	missense	WES-94	Somatic	-	Exact missense
U2AF1	NM_001025203	c.C101T	p.S34F	missense	WES-101, WES-50	Somatic	snp:44524456	Exact missense
U2AF1	NM_001025203	c.C101T	p.S34F	missense	WES-50	Somatic	snp:44524456	Exact missense
U2AF1	NM_001025203	c.C101A	p.S34Y	missense	WES-101	Somatic	snp:44524456	Exact missense

***The classification is as follow:**

Exact missense – The exact mutation from the WES has been reported in COSMIC/cBioportal

Exact truncation – The exact frameshift insertion/deletion or nonsense mutation from the WES has been reported in COSMIC/cBioportal

Truncation not reported – Truncation mutation identified from the WES which has not been reported in COSMIC/cBioportal

Truncation reported at codon – Missense mutation identified from the WES while a truncation mutation is reported by COSMIC/cBioportal at the same codon

Appendix G. Cohort characteristics of FANC core & ID2 mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=43)	Non-Mutant Group (n=107)	[1]P-value
Age	54 (16-89)	51.5 (17-84)	54.5 (16-89)	0.944^
Male - n (%)	88 (60.7%)	24 (63.2%)	64 (59.8%)	0.847
Female - n (%)	57 (39.3%)	14 (36.8%)	43 (40.2%)	0.847
WCC x 10^9/L - median (range)	19 (1.1-315.6)	27.7 (1.17-132)	17.7 (1.07-315.6)	0.569^
BM Blast % - median (range)	80.8 (50-100)	75 (50-100)	82 (50-99)	0.372^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	18/21 (85.7%)	62/67 (92.5%)	0.390
Secondary	8/88 (9.1%)	3/21 (14.3%)	5/67 (7.5%)	0.390
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	8/24 (33.3%)	17/72 (23.6%)	0.422
No	71/96 (74.0%)	16/24 (66.7%)	55/72 (76.4%)	0.422
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	0 (0%)	4/68 (6.3%)	0.571
M1	34/89 (38.2%)	5/21 (23.8%)	29/68 (42.6%)	0.134
M2	19/89 (21.3%)	7/21 (33.3%)	12/68 (17.6%)	0.138
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	5/21 (23.8%)	12/68 (17.6%)	0.536
M5	14/89 (15.7%)	4/21 (19.0%)	10/68 (14.7%)	0.733
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	0 (0%)	1/68 (1.5%)	1
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	0 (0%)	6/94 (6.06%)	0.3405
Intermediate	97/128 (75.8%)	28/34 (80%)	69/94 (69.7%)	0.3563
Poor	25/128 (19.5%)	6/34 (17.1%)	19/94 (19.2%)	0.4778
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	21/41 (51.2%)	56/101 (55.4%)	0.712
Abnormal	45/142 (31.7%)	16/41 (39.0%)	29/101 (28.7%)	0.240
Complex	20/142 (14.1%)	4/41 (9.8%)	16/101 (15.8%)	0.432
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0/41 (0%)	0 (0%)	1.000
CBF	10/142 (7.0%)	1/41 (2.4%)	9/101 (8.9%)	0.289
MLL	8/142 (5.6%)	2/41 (4.9%)	6/101 (5.9%)	1.000
tri(8)	12/142 (8.5%)	3/41 (7.3%)	9/101 (8.9%)	1.000
mono(5) / del(5q)	6/142 (4.2%)	0 (0%)	6/101 (5.9%)	0.192
mono(7) / del(7q)	10/142 (7.0%)	1/41 (2.4%)	9/101 (8.9%)	0.289
tri(21)	3/142 (2.1%)	0 (0%)	3/101 (3.0%)	0.564
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	12 / 43 (27.9%)	37 / 102 (36.3%)	0.442
FLT3-TKD	10/145 (6.9%)	4 / 43 (9.3%)	6 / 102 (5.9%)	0.483
NPM1	54/145 (37.2%)	13 / 43 (30.2%)	41 / 102 (40.2%)	0.347
DNMT3A	45/145 (31.0%)	17 / 43 (39.5%)	28 / 102 (27.5%)	0.172
IDH1	16/145 (11.0%)	6 / 43 (14.0%)	10 / 102 (9.8%)	0.563
IDH2	23/145 (15.9%)	7 / 43 (16.3%)	16 / 102 (15.7%)	1.000
TET2	26/145 (17.9%)	7 / 43 (16.3%)	19 / 102 (18.6%)	1.000
NRAS/KRAS	17/145 (11.7%)	4 / 43 (9.3%)	13 / 102 (12.7%)	0.778

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown.[1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix H. Cohort characteristics of FANC core complex mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=37)	Non-Mutant Group (n=108)	[1]*P*-value
Age	54 (16-89)	52 (17-84)	54 (16-89)	0.970^
Male - n (%)	88 (60.7%)	23 (62.2%)	65 (60.2%)	1
Female - n (%)	57 (39.3%)	14 (37.8%)	43 (39.8%)	1
WCC x 10^9/L - median (range)	19 (1.1-315.6)	27.7 (1.17-132)	17.7 (1.07-315.6)	0.569^
BM Blast % - median (range)	80.8 (50-100)	75 (50-100)	82 (50-99)	0.372^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	18/21 (85.7%)	62/67 (92.5%)	0.390
Secondary	8/88 (9.1%)	3/21 (14.3%)	5/67 (7.5%)	0.390
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	8/24 (33.3%)	17/72 (23.6%)	0.422
No	71/96 (74.0%)	16/24 (66.7%)	55/72 (76.4%)	0.422
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	0 (0%)	4/68 (5.9%)	0.569
M1	34/89 (38.2%)	5/21 (23.8%)	29/68 (42.6%)	0.134
M2	19/89 (21.3%)	7/21 (33.3%)	12/68 (17.6%)	0.138
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	5/21 (23.8%)	12/68 (17.6%)	0.536
M5	14/89 (15.7%)	4/21 (19.0%)	10/68 (14.7%)	0.733
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	0 (0%)	1/68 (1.5%)	1
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	0 (0%)	6/95 (6.3%)	0.338
Intermediate	97/128 (75.8%)	27/33 (81.8%)	70/95 (73.7%)	0.480
Poor	25/128 (19.5%)	6/33 (18.2%)	19/95 (20.0%)	1
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	17/37 (45.9%)	57/105 (54.3%)	0.446
Abnormal	45/142 (31.7%)	16/37 (43.2%)	32/105 (30.5%)	0.164
Complex	20/142 (14.1%)	4/37 (10.8%)	16/105 (15.2%)	0.594
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	1/37 (2.7%)	9/105 (8.6%)	0.454
MLL	8/142 (5.6%)	2/37 (5.4%)	6/105 (5.7%)	1
tri(8)	12/142 (8.5%)	3/37 (8.1%)	9/105 (8.6%)	1
mono(5) / del(5q)	6/142 (4.2%)	0 (0%)	6/105 (5.7%)	0.340
mono(7) / del(7q)	10/142 (7.0%)	1/37 (2.7%)	9/105 (8.6%)	0.454
tri(21)	3/142 (2.1%)	0 (0%)	3/105 (2.9%)	0.568
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	12 / 37 (32.4%)	37 / 108 (34.3%)	1
FLT3-TKD	10/145 (6.9%)	1 / 37 (2.7%)	7 / 108 (6.48%)	0.682
NPM1	54/145 (37.2%)	11 / 37 (29.7%)	40 / 108 (37%)	0.550
DNMT3A	45/145 (31.0%)	15 / 37 (40.5%)	31 / 108 (28.7%)	0.220
IDH1	16/145 (11.0%)	5 / 37 (13.5%)	11 / 108 (10.2%)	0.555
IDH2	23/145 (15.9%)	5 / 37 (13.5%)	19 / 108 (17.6%)	0.798
TET2	26/145 (17.9%)	6 / 37 (16.2%)	17 / 108 (15.7%)	1
NRAS/KRAS	17/145 (11.7%)	3 / 37 (8.1%)	14 / 108 (13.0%)	0.561

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown.[1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix I. Cohort characteristics of FANCM anchor complex mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=13)	Non-Mutant Group (n=132)	[1]*P*-value
Age	54 (16-89)	64 (19-84)	53 (16-89)	0.624^
Male - n (%)	88 (60.7%)	9 (69.2%)	79 (59.8%)	0.568
Female - n (%)	57 (39.3%)	4 (30.8%)	53 (40.2%)	0.568
WCC x 10^9/L - median (range)	19 (1.1-315.6)	40.6 (1.22-128.4)	18.45 (1.07-315.6)	0.848^
BM Blast % - median (range)	80.8 (50-100)	85 (60-100)	79 (50-99)	0.186^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	6/8 (75.0%)	74/81 (91.4%)	0.154
Secondary	8/88 (9.1%)	2/8 (25.0%)	6/81 (8.6%)	0.154
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	1/10 (10.0%)	24/86 (27.9%)	0.446
No	71/96 (74.0%)	9/10 (90.0%)	62/86 (72.1%)	0.446
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	0 (0%)	4/81 (49.4%)	1
M1	34/89 (38.2%)	3/8 (37.5%)	31/81 (38.3%)	1
M2	19/89 (21.3%)	2/8 (25.0%)	17/81 (21.0%)	0.678
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	1/8 (12.5%)	16/81 (19.8%)	1
M5	14/89 (15.7%)	2/8 (25.0%)	12/81 (14.8%)	0.607
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	0 (0%)	1/81 (1.2%)	1
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	0 (0%)	6/118 (5.1%)	1
Intermediate	97/128 (75.8%)	9/10 (90.0%)	88/118 (74.6%)	0.449
Poor	25/128 (19.5%)	1/10 (10.0%)	24/118 (20.3%)	0.686
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	6/13 (46.2%)	68/129 (52.7%)	0.774
Abnormal	45/142 (31.7%)	6/13 (46.2%)	42/129 (32.6%)	0.363
Complex	20/142 (14.1%)	1/13 (7.7%)	19/129 (14.7%)	0.694
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	1/13 (7.7%)	9/129 (7.0%)	1
MLL	8/142 (5.6%)	1/13 (7.7%)	7/129 (5.4%)	0.546
tri(8)	12/142 (8.5%)	2/13 (15.4%)	10/129 (7.8%)	0.302
mono(5) / del(5q)	6/142 (4.2%)	0 (0%)	6/129 (4.7%)	1
mono(7) / del(7q)	10/142 (7.0%)	0 (0%)	10/129 (7.8%)	0.599
tri(21)	3/142 (2.1%)	0 (0%)	3/129 (2.3%)	1
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	3 / 13 (23.1%)	46 / 132 (34.8%)	0.544
FLT3-TKD	10/145 (6.9%)	1 / 13 (7.7%)	7 / 132 (5.3%)	0.550
NPM1	54/145 (37.2%)	4 / 13 (30.8%)	47 / 132 (35.6%)	1
DNMT3A	45/145 (31.0%)	4 / 13 (30.8%)	42 / 132 (31.8%)	1
IDH1	16/145 (11.0%)	3 / 13 (23.1%)	13 / 132 (9.85%)	0.158
IDH2	23/145 (15.9%)	0 / 13 (0%)	24 / 132 (18.2%)	0.127
TET2	26/145 (17.9%)	2 / 13 (15.4%)	21 / 132 (15.9%)	1
NRAS/KRAS	17/145 (11.7%)	3 / 13 (23.1%)	14 / 132 (10.6%)	0.182

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix J. Cohort characteristics of minimal FANCD2 monoubiquitination complex mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=19)	Non-Mutant Group (n=126)	[1]*P*-value
Age	54 (16-89)	50 (17-83)	55 (16-89)	0.330^
Male - n (%)	88 (60.7%)	13 (68.4%)	75 (59.5%)	0.616
Female - n (%)	57 (39.3%)	6 (31.6%)	51 (40.5%)	0.616
WCC x 10^9/L - median (range)	19 (1.1-315.6)	19 (1.17-132)	19 (1.07-315.6)	0.737^
BM Blast % - median (range)	80.8 (50-100)	70.5 (50-100)	81.75 (50-99)	0.548^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	8/9 (88.9%)	72/79 (91.1%)	1.000
Secondary	8/88 (9.1%)	1/9 (11.1%)	7/79 (8.9%)	1.000
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	6/11 (54.5%)	19/85 (22.4%)	0.032
No	71/96 (74.0%)	5/11 (45.5%)	66/85 (77.6%)	0.032
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	0 (0%)	4/79 (5.1%)	1.000
M1	34/89 (38.2%)	3/10 (30.0%)	31/79 (39.2%)	0.736
M2	19/89 (21.3%)	3/10 (30.0%)	16/79 (20.3%)	0.440
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1.000
M4	17/89 (19.1%)	2/10 (20.0%)	15/79 (19.0%)	1.000
M5	14/89 (15.7%)	2/10 (20.0%)	12/79 (15.2%)	0.654
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1.000
M7	1/89 (1.1%)	0 (0%)	1/79 (1.3%)	1.000
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	0 (0%)	6/112 (5.4%)	1.000
Intermediate	97/128 (75.8%)	14/16 (87.5%)	83/112 (74.1%)	0.354
Poor	25/128 (19.5%)	2/16 (12.5%)	23/112 (20.5%)	0.736
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	6/19 (31.6%)	68/123 (55.3%)	0.082
Abnormal	45/142 (31.7%)	13/19 (68.4%)	35/123 (28.5%)	0.001
Complex	20/142 (14.1%)	0 (0%)	20/123 (16.3%)	0.075
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1.000
CBF	10/142 (7.0%)	1/19 (5.3%)	9/123 (7.3%)	1.000
MLL	8/142 (5.6%)	2/19 (10.5%)	6/123 (4.9%)	0.291
tri(8)	12/142 (8.5%)	2/19 (10.5%)	10/123 (8.1%)	0.664
mono(5) / del(5q)	6/142 (4.2%)	0 (0%)	6/123 (4.9%)	1.000
mono(7) / del(7q)	10/142 (7.0%)	1/19 (5.3%)	9/123 (7.3%)	1.000
tri(21)	3/142 (2.1%)	0 (0%)	3/123 (2.4%)	1.000
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	6 / 19 (31.6%)	43 / 126 (34.1%)	1.000
FLT3-TKD	10/145 (6.9%)	0 / 19 (0%)	8 / 126 (6.35%)	0.599
NPM1	54/145 (37.2%)	5 / 19 (26.3%)	46 / 126 (36.5%)	0.602
DNMT3A	45/145 (31.0%)	7 / 19 (36.8%)	39 / 126 (31%)	0.605
IDH1	16/145 (11.0%)	2 / 19 (10.5%)	14 / 126 (11.1%)	1.000
IDH2	23/145 (15.9%)	4 / 19 (21.1%)	20 / 126 (15.9%)	0.522
TET2	26/145 (17.9%)	1 / 19 (5.3%)	22 / 126 (17.5%)	0.310
NRAS/KRAS	17/145 (11.7%)	0 / 19 (0%)	17 / 126 (32.5%)	0.128

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix K. Cohort characteristics of breast cancer associated protein mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=11)	Non-Mutant Group (n=134)	[1]*P*-value
Age	54 (16-89)	48 (17-73)	54 (16-89)	0.106^
Male - n (%)	88 (60.7%)	6 (54.5%)	82 (61.2%)	0.752
Female - n (%)	57 (39.3%)	5 (45.5%)	52 (38.8%)	0.752
WCC x 10^9/L - median (range)	19 (1.1-315.6)	24.75 (2.1-111.1)	18.9 (1.07-315.6)	0.867^
BM Blast % - median (range)	80.8 (50-100)	86 (70-100)	80.75 (50-99)	0.132^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	6/6 (100.0%)	74/82 (90.2%)	1
Secondary	8/88 (9.1%)	0 (0%)	8/82 (9.8%)	1
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	2/6 (33.3%)	23/90 (25.6%)	0.649
No	71/96 (74.0%)	4/6 (66.7%)	67/90 (74.4%)	0.649
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	0 (0%)	4/83 (4.8%)	1
M1	34/89 (38.2%)	3/6 (50.0%)	31/83 (37.3%)	0.671
M2	19/89 (21.3%)	0 (0%)	19/83 (22.9%)	0.335
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	1/6 (16.7%)	16/83 (19.3%)	1
M5	14/89 (15.7%)	1/6 (16.7%)	13/83 (15.7%)	1
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	1/6 (16.7%)	0 (0%)	0.067
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	1/10 (10.0%)	5/118 (4.2%)	0.392
Intermediate	97/128 (75.8%)	7/10 (70.0%)	90/118 (76.3%)	0.704
Poor	25/128 (19.5%)	2/10 (20.0%)	23/118 (19.5%)	1
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	3/11 (27.3%)	71/131 (54.2%)	0.118
Abnormal	45/142 (31.7%)	6/11 (54.5%)	42/131 (32.1%)	0.183
Complex	20/142 (14.1%)	2/11 (18.2%)	18/131 (13.7%)	0.654
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	2/11 (18.2%)	8/131 (6.1%)	0.174
MLL	8/142 (5.6%)	3/11 (27.3%)	5/131 (3.8%)	0.016
tri(8)	12/142 (8.5%)	0 (0%)	12/131 (9.2%)	0.599
mono(5) / del(5q)	6/142 (4.2%)	1/11 (9.1%)	5/131 (3.8%)	0.389
mono(7) / del(7q)	10/142 (7.0%)	0 (0%)	10/131 (7.6%)	1
tri(21)	3/142 (2.1%)	1/11 (9.1%)	2/131 (1.5%)	0.216
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	3 / 11 (27.3%)	46 / 134 (34.3%)	0.750
FLT3-TKD	10/145 (6.9%)	0 / 11 (0%)	8 / 134 (6.0%)	1
NPM1	54/145 (37.2%)	2 / 11 (18.2%)	49 / 134 (36.6%)	0.328
DNMT3A	45/145 (31.0%)	2 / 11 (18.2%)	44 / 134 (32.8%)	0.503
IDH1	16/145 (11.0%)	0 / 11 (0%)	16 / 134 (11.9%)	0.611
IDH2	23/145 (15.9%)	1 / 11 (9.09%)	23 / 134 (17.2%)	0.692
TET2	26/145 (17.9%)	2 / 11 (18.2%)	21 / 134 (15.7%)	0.687
NRAS/KRAS	17/145 (11.7%)	2 / 11 (18.2%)	15 / 134 (11.2%)	0.619

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix L. Cohort characteristics of FANC 19 mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=55)	Non-Mutant Group (n=95)	[1]*P*-value
Age	54 (16-89)	54 (16-84)	54 (17-89)	0.481^
Male - n (%)	88 (60.7%)	36 (65.5%)	52 (57.8%)	0.386
Female - n (%)	57 (39.3%)	19 (34.5%)	38 (42.2%)	0.386
WCC x 10^9/L - median (range)	19 (1.1-315.6)	21.6 (1.22-227)	16.4 (1.07-315.6)	0.617^
BM Blast % - median (range)	80.8 (50-100)	78.5 (50-100)	82.5 (50-99)	0.493^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	28/32 (87.5%)	52/56 (92.9%)	0.455
Secondary	8/88 (9.1%)	4/32 (12.5%)	4/56 (7.1%)	0.455
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	11/35 (31.4%)	14/35 (40.0%)	0.469
No	71/96 (74.0%)	24/35 (68.6%)	47/35 (60.0%)	0.469
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	0 (0%)	4/57 (7.0%)	0.292
M1	34/89 (38.2%)	10/32 (31.3%)	24/57 (42.1%)	0.368
M2	19/89 (21.3%)	8/32 (25.0%)	11/57 (19.3%)	0.594
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	7/32 (21.9%)	10/57 (17.5%)	0.779
M5	14/89 (15.7%)	6/32 (18.8%)	8/57 (14.0%)	0.559
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	1/32 (3.1%)	0 (0%)	0.360
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	2/47 (4.3%)	4/81 (4.9%)	1
Intermediate	97/128 (75.8%)	37/47 (78.7%)	60/81 (74,1%)	0.67
Poor	25/128 (19.5%)	8/47 (17.0%)	17/81 (21.0%)	0.650
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	23/53 (43.4%)	51/89 (57.3%)	0.121
Abnormal	45/142 (31.7%)	23/53 (43.4%)	25/89 (28.1%)	0.069
Complex	20/142 (14.1%)	7/53 (13.2%)	13/89 (14.6%)	1
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	3/53 (5.7%)	7/89 (7.9%)	0.743
MLL	8/142 (5.6%)	4/53 (7.6%)	4/89 (4.5%)	0.471
tri(8)	12/142 (8.5%)	4/53 (7.6%)	8/89 (9.0%)	1
mono(5) / del(5q)	6/142 (4.2%)	2/53 (3.8%)	4/89 (4.5%)	1
mono(7) / del(7q)	10/142 (7.0%)	1/53 (1.9%)	9/89 (10.1%)	0.091
tri(21)	3/142 (2.1%)	1/53 (1.9%)	2/89 (2.3%)	1
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	13 / 55 (23.6%)	36 / 90 (40%)	0.048
FLT3-TKD	10/145 (6.9%)	5 / 55 (9.1%)	3 / 90 (3.3%)	0.270
NPM1	54/145 (37.2%)	15 / 55 (27.3%)	36 / 90 (40%)	0.153
DNMT3A	45/145 (31.0%)	18 / 55 (32.7%)	28 / 90 (31.1%)	0.855
IDH1	16/145 (11.0%)	8 / 55 (14.5%)	8 / 90 (8.9%)	0.413
IDH2	23/145 (15.9%)	9 / 55 (16.4%)	15 / 90 (16.7%)	1
TET2	26/145 (17.9%)	8 / 55 (14.5%)	15 / 90 (16.7%)	0.818
NRAS/KRAS	17/145 (11.7%)	8 / 55 (14.5%)	7 / 90 (7.8%)	0.261

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix M. Cohort characteristics of BLM complex mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=17)	Non-Mutant Group (n=128)	[1]*P*-value
Age	54 (16-89)	51.5 (17-79)	54 (16-89)	0.500^
Male - n (%)	88 (60.7%)	11 (64.7%)	77 (60.2%)	0.800
Female - n (%)	57 (39.3%)	6 (35.3%)	51 (39.8%)	0.800
WCC x 10^9/L - median (range)	19 (1.1-315.6)	58.6 (2.17-315.6)	18.8 (1.07-313.3)	0.300^
BM Blast % - median (range)	80.8 (50-100)	83.5 (60-99)	80.75 (50-100)	0.395^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	10/10 (100.0%)	70/78 (89.7%)	0.589
Secondary	8/88 (9.1%)	0 (0%)	8/78 (10.3%)	0.589
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	3/11 (27.3%)	22/85 (25.9%)	1.000
No	71/96 (74.0%)	8/11 (72.7%)	63/85 (74.1%)	1.000
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	1/11 (9.1%)	3/78 (3.8%)	0.416
M1	34/89 (38.2%)	4/11 (36.4%)	30/78 (38.5%)	1.000
M2	19/89 (21.3%)	2/11 (18.2%)	17/78 (21.8%)	1.000
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1.000
M4	17/89 (19.1%)	4/11 (36.4%)	13/78 (16.7%)	0.211
M5	14/89 (15.7%)	0 (0%)	14/78 (17.9%)	0.201
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1.000
M7	1/89 (1.1%)	0 (0%)	1/78 (1.3%)	1.000
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	0 (0%)	6/115 (5.2%)	1.000
Intermediate	97/128 (75.8%)	10/13 (76.9%)	87/115 (75.7%)	1.000
Poor	25/128 (19.5%)	3/13 (23.1%)	22/115 (19.1%)	0.717
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	6/16 (37.5%)	68/126 (54.0%)	0.289
Abnormal	45/142 (31.7%)	8/16 (50.0%)	40/126 (31.7%)	0.167
Complex	20/142 (14.1%)	2/16 (12.5%)	18/126 (14.3%)	1.000
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1.000
CBF	10/142 (7.0%)	1 /16(6.3%)	9/126 (7.1%)	1.000
MLL	8/142 (5.6%)	2/16 (12.5%)	6/126 (4.8%)	0.223
tri(8)	12/142 (8.5%)	0 (0%)	12/126 (9.5%)	0.361
mono(5) / del(5q)	6/142 (4.2%)	0 (0%)	6/126 (4.8%)	1.000
mono(7) / del(7q)	10/142 (7.0%)	2/16 (12.5%)	8/126 (6.4%)	0.313
tri(21)	3/142 (2.1%)	0 (0%)	3/126 (2.4%)	1.000
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	7 / 17 (41.2%)	42 / 128 (32.8%)	0.587
FLT3-TKD	10/145 (6.9%)	0 / 17 (0%)	8 / 128 (6.25%)	0.595
NPM1	54/145 (37.2%)	4 / 17 (23.5%)	47 / 128 (36.7%)	0.419
DNMT3A	45/145 (31.0%)	8 / 17 (47.1%)	38 / 128 (29.7%)	0.170
IDH1	16/145 (11.0%)	2 / 17 (11.8%)	14 / 128 (10.9%)	1.000
IDH2	23/145 (15.9%)	4 / 17 (23.5%)	20 / 128 (15.6%)	0.485
TET2	26/145 (17.9%)	2 / 17 (11.8%)	21 / 128 (16.4%)	1.000
NRAS/KRAS	17/145 (11.7%)	2 / 17 (11.8%)	15 / 128 (11.7%)	1.000

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix N. Cohort characteristics of ATM/ATR checkpoint proteins mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=34)	Non-Mutant Group (n=111)	[1]*P*-value
Age	54 (16-89)	56 (17-80)	52.5 (16-89)	0.348^
Male - n (%)	88 (60.7%)	23 (67.6%)	65 (58.6%)	0.424
Female - n (%)	57 (39.3%)	11 (32.4%)	46 (41.4%)	0.424
WCC x 10^9/L - median (range)	19 (1.1-315.6)	18.8 (1.17-315.6)	19.1 (1.07-313.3)	0.628^
BM Blast % - median (range)	80.8 (50-100)	80 (52-96)	81.5 (50-100)	0.919^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	22/24 (91.6%)	58/64 (90.6%)	1
Secondary	8/88 (9.1%)	2/24 (8.4%)	6/64 (9.4%)	1
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	5/26 (19.2%)	20/70 (28.6%)	0.439
No	71/96 (74.0%)	21/26 (80.8%)	50/70 (71.4%)	0.439
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	1/26 (3.8%)	3/63 (4.8%)	1
M1	34/89 (38.2%)	11/26 (42.3%)	23/63 (36.5%)	0.638
M2	19/89 (21.3%)	5/26 (19.2%)	14/63 (22.2%)	1
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	4/26 (15.4%)	13/63 (20.6%)	0.768
M5	14/89 (15.7%)	4/26 (15.4%)	10/63 (15.9%)	1
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	1/26 (3.8%)	0 (0%)	0.292
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	3/32 (9.4%)	3/96 (3.1%)	0.165
Intermediate	97/128 (75.8%)	22/32 (68.8%)	75/96 (78.1%)	0.342
Poor	25/128 (19.5%)	7/32 (21.9%)	18/96 (18.8%)	0.797
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	17/34 (50%)	57/108 (52.8%)	0.845
Abnormal	45/142 (31.7%)	13/34 (38.2%)	35/108 (32.4%)	0.539
Complex	20/142 (14.1%)	4/34 (11.8%)	16/108 (14.8%)	0.783
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	3/34 (8.8%)	7/108 (6.5%)	0.703
MLL	8/142 (5.6%)	2/34 (5.9%)	6 (5.6%)	1
tri(8)	12/142 (8.5%)	2/34 (5.9%)	10/108 (9.3%)	0.731
mono(5) / del(5q)	6/142 (4.2%)	2/34 (5.9%)	4/108 (3.7%)	0.630
mono(7) / del(7q)	10/142 (7.0%)	2/34 (5.9%)	8/108 (7.4%)	1
tri(21)	3/142 (2.1%)	2/34 (5.9%)	1/108 (0.93%)	0.143
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	10 / 34 (29.4%)	39 / 111 (35.1%)	0.679
FLT3-TKD	10/145 (6.9%)	2 / 34 (5.88%)	6 / 111 (5.41%)	1
NPM1	54/145 (37.2%)	11 / 34 (32.4%)	40 / 111 (36%)	0.838
DNMT3A	45/145 (31.0%)	10 / 34 (29.4%)	36 / 111 (32.4%)	0.835
IDH1	16/145 (11.0%)	5 / 34 (14.7%)	11 / 111 (9.91%)	0.531
IDH2	23/145 (15.9%)	7 / 34 (20.6%)	17 / 111 (15.3%)	0.443
TET2	26/145 (17.9%)	6 / 34 (17.6%)	17 / 111 (15.3%)	0.790
NRAS/KRAS	17/145 (11.7%)	2 / 34 (5.9%)	15 / 111 (13.5%)	0.529

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix O. Cohort characteristics of BLM & ATM/ATR checkpoint proteins mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=51)	Non-Mutant Group (n=94)	[1]*P*-value
Age	54 (16-89)	55.5 (17-80)	52 (16-89)	0.303^
Male - n (%)	88 (60.7%)	35 (68.6%)	53 (56.4%)	0.16
Female - n (%)	57 (39.3%)	16 (31.4%)	41 (43.6%)	0.16
WCC x 10^9/L - median (range)	19 (1.1-315.6)	38.4 (1.17-315.6)	17.95 (1.07-313.3)	0.548^
BM Blast % - median (range)	80.8 (50-100)	80 (52-99)	81.5 (50-100)	0.629^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	33/36 (91.7%)	47/52 (90.4%)	1
Secondary	8/88 (9.1%)	3/36 (8.3%)	5/52 (9.6%)	1
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	9/38 (23.7%)	16/58 (27.6%)	0.813
No	71/96 (74.0%)	29/38 (76.3%)	42/58 (44.7%)	0.813
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	2/37 (5.4%)	2/52 (3.8%)	1
M1	34/89 (38.2%)	15/37 (40.5%)	19/52 (36.5%)	0.825
M2	19/89 (21.3%)	8/37 (21.6%)	11/52 (21.1%)	1
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	7/37 (18.9%)	10/52 (19.2%)	1
M5	14/89 (15.7%)	4/37 (10.8%)	10/52 (19.2%)	0.380
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	1/37 (2.7%)	0 (0%)	0.416
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	3/46 (6.5%)	3/82 (3.7%)	0.666
Intermediate	97/128 (75.8%)	32/46 (69.6%)	65/82 (79.2%)	0.283
Poor	25/128 (19.5%)	11/46 (23.9%)	14/82 (17.1%)	0.362
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	22/50 (44%)	52/92 (56.5%)	0.164
Abnormal	45/142 (31.7%)	21/50 (42%)	27/92 (29.3%)	0.141
Complex	20/142 (14.1%)	7/50 (14%)	13/92 (14.1%)	1
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	4/50 (8%)	6/92 (6.52%)	0.742
MLL	8/142 (5.6%)	4/50 (8%)	4/92 (4.35%)	0.452
tri(8)	12/142 (8.5%)	2/50 (4%)	10/92 (10.9%)	0.214
mono(5) / del(5q)	6/142 (4.2%)	2/50 (4%)	4/92 (4.35%)	1
mono(7) / del(7q)	10/142 (7.0%)	4/50 (8%)	6/92 (6.52%)	0.742
tri(21)	3/142 (2.1%)	2/50 (4%)	1/92 (1.09%)	0.283
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	15 / 51 (29.4%)	34 / 94 (36.2%)	0.465
FLT3-TKD	10/145 (6.9%)	2 / 51 (3.92%)	6 / 94 (6.38%)	0.713
NPM1	54/145 (37.2%)	14 / 51 (27.5%)	37 / 94 (39.4%)	0.150
DNMT3A	45/145 (31.0%)	18 / 51 (35.3%)	28 / 94 (29.8%)	0.576
IDH1	16/145 (11.0%)	7 / 51 (13.7%)	9 / 94 (9.57%)	0.580
IDH2	23/145 (15.9%)	12 / 51 (23.5%)	12 / 94 (12.8%)	0.107
TET2	26/145 (17.9%)	8 / 51 (15.7%)	15 / 94 (16%)	1
NRAS/KRAS	17/145 (11.7%)	4 / 51 (7.8%)	13 / 94 (13.8%)	0.419

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix P. Cohort characteristics of Mismatch repair proteins mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=17)	Non-Mutant Group (n=128)	[1]*P*-value
Age	54 (16-89)	53 (20-84)	54 (16-89)	0.577^
Male - n (%)	88 (60.7%)	13 (76.5%)	75 (58.6%)	0.193
Female - n (%)	57 (39.3%)	4 (23.5%)	53 (41.4%)	0.193
WCC x 10^9/L - median (range)	19 (1.1-315.6)	24.25 (1.75-227)	19 (1.07-315.6)	0.573^
BM Blast % - median (range)	80.8 (50-100)	78.5 (60-93)	81.75 (50-100)	0.899^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	9/10 (90.0%)	71/78 (91.0%)	1
Secondary	8/88 (9.1%)	1/10 (10.0%)	7/78 (9.0%)	1
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	3/10 (30.0%)	22/86 (25.6%)	0.717
No	71/96 (74.0%)	7/10 (70.0%)	64/86 (74.4%)	0.717
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	0 (0%)	4/79 (5.1%)	1
M1	34/89 (38.2%)	2/10 (20.0%)	32/79 (40.5%)	0.306
M2	19/89 (21.3%)	2/10 (20.0%)	17/79 (21.5%)	1
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	3/10 (30.0%)	14/79 (17.7%)	0.395
M5	14/89 (15.7%)	3/10 (30.0%)	11/79 (13.9%)	0.189
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	0 (0%)	1/79 (1.3%)	1
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	2/16 (12.5%)	4/112 (3.6%)	0.163
Intermediate	97/128 (75.8%)	13/16 (81.3%)	84/112 (75.0%)	0.760
Poor	25/128 (19.5%)	1/16 (6.3%)	24/112 (21.4%)	0.194
Unknown	17	0 (0%)	0 (0%)	1
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	6/16 (37.5%)	68/126 (54%)	0.289
Abnormal	45/142 (31.7%)	8/16 (50.0%)	40/126 (31.7%)	0.167
Complex	20/142 (14.1%)	2/16 (12.5%)	18/126 (14.3%)	1
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	2/16 (12.5%)	8/126 (6.4%)	0.313
MLL	8/142 (5.6%)	2/16 (12.5%)	6/126 (4.8%)	0.223
tri(8)	12/142 (8.5%)	2/16 (12.5%)	10/126 (7.9%)	0.626
mono(5) / del(5q)	6/142 (4.2%)	0 (0%)	6/126 (4.8%)	1
mono(7) / del(7q)	10/142 (7.0%)	0 (0%)	10/126 (7.9%)	0.603
tri(21)	3/142 (2.1%)	0 (0%)	3/126 (2.4%)	1
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	5 / 17 (29.4%)	44 / 128 (34.4%)	0.790
FLT3-TKD	10/145 (6.9%)	2 / 17 (11.8%)	6 / 128 (4.69%)	0.256
NPM1	54/145 (37.2%)	7 / 17 (41.2%)	44 / 128 (34.4%)	0.599
DNMT3A	45/145 (31.0%)	4 / 17 (23.5%)	42 / 128 (32.8%)	0.582
IDH1	16/145 (11.0%)	2 / 17 (11.8%)	14 / 128 (10.9%)	1
IDH2	23/145 (15.9%)	3 / 17 (17.6%)	21 / 128 (16.4%)	1
TET2	26/145 (17.9%)	2 / 17 (11.8%)	21 / 128 (16.4%)	1
NRAS/KRAS	17/145 (11.7%)	3 / 17 (17.6%)	14 / 128 (10.9%)	0.423

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix Q. Cohort characteristics of Structure-specific endonucleases mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=23)	Non-Mutant Group (n=122)	[1]*P*-value
Age	54 (16-89)	56 (16-84)	53 (17-89)	1.000^
Male - n (%)	88 (60.7%)	15 (65.2%)	73 (59.8%)	0.816
Female - n (%)	57 (39.3%)	8 (34.8%)	49 (40.2%)	0.816
WCC x 10^9/L - median (range)	19 (1.1-315.6)	15.25 (2.1-227)	22.3 (1.07-315.6)	0.672^
BM Blast % - median (range)	80.8 (50-100)	81 (50-99)	80.25 (50-100)	0.617^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	12/14 (85.7%)	68/74 (91.9%)	0.608
Secondary	8/88 (9.1%)	2/14 (14.3%)	6/74 (8.1%)	0.608
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	3/16 (18.8%)	22/80 (27.5%)	0.55
No	71/96 (74.0%)	13/16 (81.2%)	58/80 (72.5%)	0.55
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	2/15 (13.3%)	2/64 (3.1%)	0.161
M1	34/89 (38.2%)	5/15 (33.3%)	29/64 (45.3%)	0.564
M2	19/89 (21.3%)	3/15 (20.0%)	16/64 (25.0%)	1
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	2/15 (13.3%)	15/64 (23.4%)	0.503
M5	14/89 (15.7%)	3/15 (20.0%)	11/64 (12.2%)	0.723
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	0 (0%)	1/64 (1.6%)	1
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	1/18 (5.6%)	5/110 (4.5%)	1
Intermediate	97/128 (75.8%)	10/18 (55.6%)	87/110 (79.1%)	0.040
Poor	25/128 (19.5%)	7/18 (38.9%)	18/110 (16.4%)	0.048
Unknown	17			
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	8/21 (38.1%)	66/121 (54.5%)	0.236
Abnormal	45/142 (31.7%)	8/21 (38.1%)	40/121 (33.1%)	0.628
Complex	20/142 (14.1%)	5/21 (23.8%)	15/121 (12.4%)	0.179
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	2/21 (9.5%)	8/121 (6.6%)	0.643
MLL	8/142 (5.6%)	2/21 (9.5%)	6/121 (5.0%)	0.337
tri(8)	12/142 (8.5%)	2/21 (9.5%)	10/121 (8.3%)	0.692
mono(5) / del(5q)	6/142 (4.2%)	1/21 (4.8%)	5/121 (4.1%)	1
mono(7) / del(7q)	10/142 (7.0%)	1/21 (4.8%)	9/121 (7.4%)	1
tri(21)	3/142 (2.1%)	1/21 (4.8%)	2/121 (1.7%)	0.384
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	4 / 23 (17.4%)	45 / 122 (36.9%)	0.093
FLT3-TKD	10/145 (6.9%)	3 / 23 (13%)	5 / 122 (4.1%)	0.105
NPM1	54/145 (37.2%)	7 / 23 (30.4%)	44 / 122 (36.1%)	0.642
DNMT3A	45/145 (31.0%)	8 / 23 (34.8%)	38 / 122 (31.1%)	0.808
IDH1	16/145 (11.0%)	2 / 23 (8.7%)	14 / 122 (11.5%)	1
IDH2	23/145 (15.9%)	2 / 23 (8.7%)	22 / 122 (18%)	0.368
TET2	26/145 (17.9%)	4 / 23 (17.4%)	19 / 122 (15.6%)	0.762
NRAS/KRAS	17/145 (11.7%)	4 / 23 (17.4%)	13 / 122 (10.7%)	0.476

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

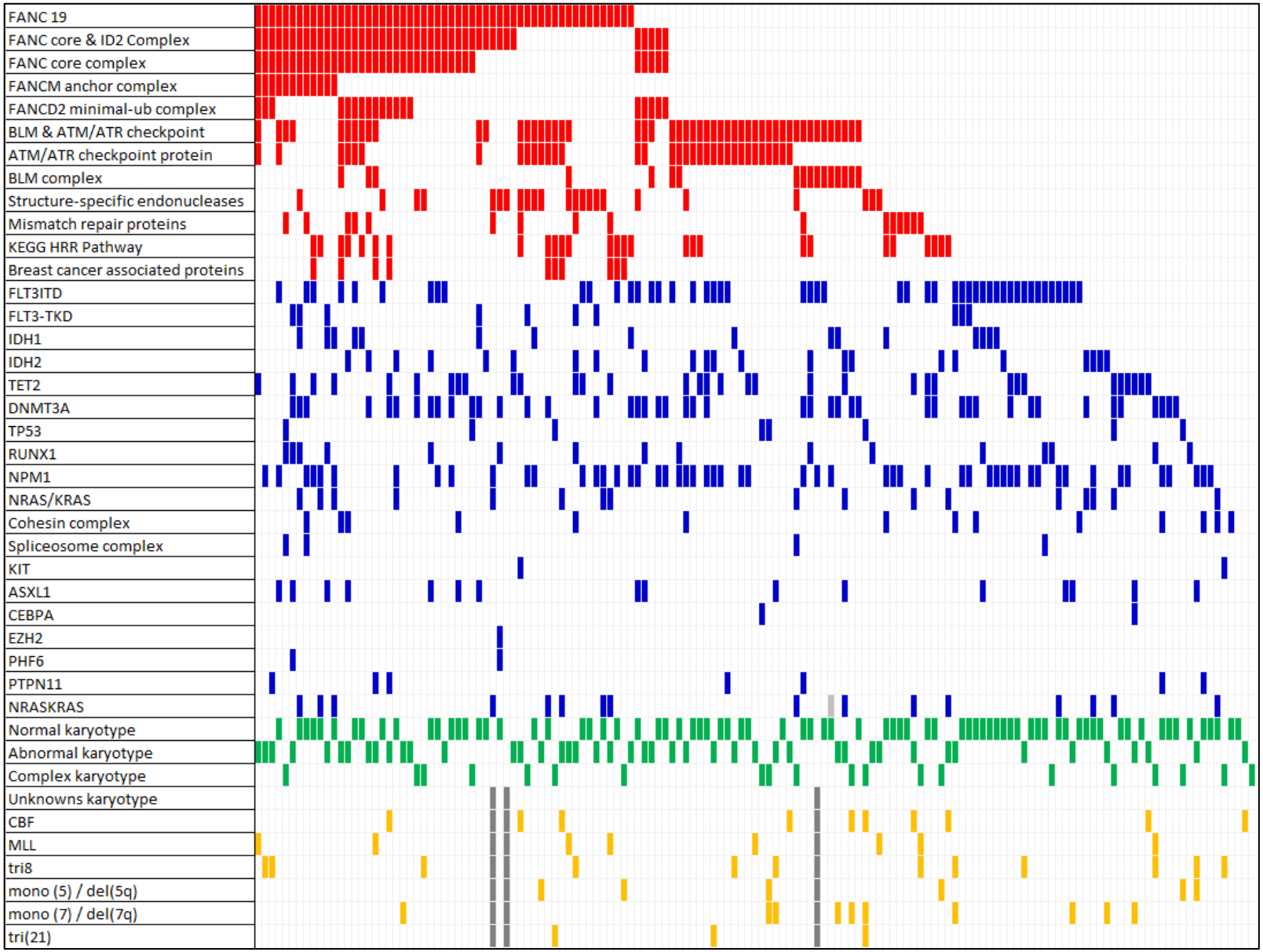
Appendix R. Cohort characteristics of KEGG-HRR mutant subgroup

Characteristics	*All Cases (n=145)	Mutant Group (n=27)	Non-Mutant Group (n=118)	[1]*P*-value
Age	54 (16-89)	52 (17-81)	54 (16-89)	0.339^
Male - n (%)	88 (60.7%)	15 (55.6%)	73 (61.9%)	0.663
Female - n (%)	57 (39.3%)	12 (44.4%)	45 (38.1%)	0.663
WCC x 10^9/L - median (range)	19 (1.1-315.6)	30.5 (1.75-250)	18.8 (1.07-315.6)	0.262^
BM Blast % - median (range)	80.8 (50-100)	84 (60-100)	79 (50-99)	0.082^
Primary/Secondary AML - n/total (%)				
De Novo	80/88 (90.9%)	14/15 (93.3%)	66/73 (90.4%)	1
Secondary	8/88 (9.1%)	1/15 (6.7%)	7/73 (9.6%)	1
Unknown	57			
Transplant – n/total (%)				
Yes	25/96 (26.0%)	5/15 (33.3%)	20/81 (24.7%)	0.527
No	71/96 (74.0%)	10/15 (66.7%)	61/81 (75.3%)	0.527
Unknown	49			
[2]FAB – n/total (%)				
M0	4/89 (4.5%)	1/15 (6.7%)	3/74 (4.1%)	0.529
M1	34/89 (38.2%)	7/15 (46.7%)	27/74 (36.5%)	0.563
M2	19/89 (21.3%)	2/15 (13.3%)	17/74 (23.0%)	0.529
M3	0/89 (0.0%)	0 (0%)	0 (0%)	1
M4	17/89 (19.1%)	2/15 (13.3%)	15/74 (20.3%)	0.503
M5	14/89 (15.7%)	2/15 (13.3%)	12/74 (16.2%)	0.725
M6	0/89 (0.0%)	0 (0%)	0 (0%)	1
M7	1/89 (1.1%)	1/15 (6.7%)	0 (0%)	0.169
Unknown	56			
[3]Grimwade Cytogenetic Risk – n/total (%)				
Good	6/128 (4.7%)	3/25 (12.0%)	3/103 (2.9%)	0.088
Intermediate	97/128 (75.8%)	19/25 (76.0%)	78/103 (75.7%)	1
Poor	25/128 (19.5%)	3/25 (12.0%)	22/103 (21.4%)	0.403
Unknown	17	0 (0%)	0 (0%)	1
Simple Karyotype - n/total (%)				
Normal	77/142 (54.2%)	9/27 (33.3%)	65/115 (56.5%)	0.034
Abnormal	45/142 (31.7%)	15/27 (55.6%)	33/115 (28.7%)	0.012
Complex	20/142 (14.1%)	3/27 (11.1%)	17/115 (14.8%)	0.766
Unknown	3			
Cytogenetics – n/total (%)				
t(15;17)	0/142 (0.0%)	0 (0%)	0 (0%)	1
CBF	10/142 (7.0%)	4/27 (14.8%)	6/115 (5.2%)	0.096
MLL	8/142 (5.6%)	4/27 (14.8%)	4/115 (3.5%)	0.043
tri(8)	12/142 (8.5%)	0 (0%)	12/115 (10.4%)	0.123
mono(5) / del(5q)	6/142 (4.2%)	2/27 (7.4%)	4/115 (3.5%)	0.320
mono(7) / del(7q)	10/142 (7.0%)	0 (0%)	10/115 (8.7%)	0.209
tri(21)	3/142 (2.1%)	1/27 (3.7%)	2/115 (1.7%)	0.472
Mutations - n/total (%)				
FLT3-ITD	49/145 (33.8%)	9 / 27 (33.3%)	40 / 118 (33.9%)	1
FLT3-TKD	10/145 (6.9%)	0 / 27 (0%)	8 / 118 (6.78%)	0.204
NPM1	54/145 (37.2%)	11 / 27 (40.7%)	40 / 118 (33.9%)	0.513
DNMT3A	45/145 (31.0%)	9 / 27 (33.3%)	37 / 118 (31.4%)	0.823
IDH1	16/145 (11.0%)	3 / 27 (11.1%)	13 / 118 (11%)	1
IDH2	23/145 (15.9%)	5 / 27 (18.5%)	19 / 118 (16.1%)	0.776
TET2	26/145 (17.9%)	6 / 27 (22.2%)	17 / 118 (14.4%)	0.380
NRAS/KRAS	17/145 (11.7%)	5 / 27 (18.5%)	12 / 118 (10.2%)	0.315

*Total number of samples in the cohort is 145, however, where clinical characteristics were not available, the sample is listed as unknown. [1]P-values are calculated by Fisher's exact test except for: ^ determined by Student's t-test. [2]FAB: French-America-British classification (Neame et al., 1986); [3]Grimwade classification (Grimwade et al., 2010)

Appendix S. Graphical visual representation the 12 subgroups of FA/BRCA-HRR extended network variants, common AML mutations and karyotypic abnormalities across the adult WES AML cohort (n=145).

Red cells represent samples with mutations in the extended FA/BRCA-HRR network; Green cells represent the simple karyotype classification; Blue cells represent samples with recurring AML mutations; Yellow cells represent samples with karyotypic abnormalities; Grey cells represent samples with unknown karyotypes.


FANC 19
FANC core & ID2 Complex
FANC core complex
FANCM anchor complex
FANCD2 minimal-ub complex
BLM & ATM/ATR checkpoint
ATM/ATR checkpoint protein
BLM complex
Structure-specific endonucleases
Mismatch repair proteins
KEGG HRR Pathway
Breast cancer associated proteins
FLT3ITD
FLT3-TKD
IDH1
IDH2
TET2
DNMT3A
TP53
RUNX1
NPM1
NRAS/KRAS
Cohesin complex
Spliceosome complex
KIT
ASXL1
CEBPA
EZH2
PHF6
PTPN11
NRASKRAS
Normal karyotype
Abnormal karyotype
Complex karyotype
Unknowns karyotype
CBF
MLL
tri8
mono (5) / del(5q)
mono (7) / del(7q)
tri(21)

Appendix T. COSMIC mutations occurring at the same codon as variants identified in the WES cohort.

Gene Names	Amino acid change from the WES	COSMIC ID	Somatic status in COSMIC	Primary tissues	Histology	No of Patients with variant	No. of independent samples in COSMIC	Comments
ATM	p.L2492R	COSM5547397 & COSM758331	Somatic & Unknown	Prostate & Haematopoietic and lymphoid tissue	Carcinoma & Lymphoid neoplasm	1	2	Reported as missense & frameshift
ATM	p.T2640I	COSM3358392	Somatic	Kidney	Carcinoma	1	1	p.T2640fs*6 reported instead
ATR	p.I1753V	COSM5633507	Somatic	Oesophagus	Carcinoma	1	1	Reported as I1753T
ATR	p.P2434A	COSM3125094	Somatic	Skin	Carcinoma	1	1	Reported as P2434S
BLM	p.V4A	COSM3505417	Somatic	Skin	Malignant melanoma	1	1	Reported as V4F
BRCA1	p.R1203X	COSM3402921	Somatic	Central nervous system	Glioma	1	1	p.R1203Q instead
BRCA2	p.C554W	COSM6074241	Unknown	Lung	Carcinoma	1	1	Reported as C554F
FANCD2	p.Q65H	COSM5494739	Somatic	Biliary tract	Carcinoma	1	1	Reported as Q65X
FANCI	p.E209Q	COSM3505239	Somatic	Skin	Malignant melanoma	1	1	p.E209K reported instead
FANCL	p.T224A	COSM3053711	Somatic	Lung	Carcinoma	1	1	p.T224K reported instead (COSM3053711)
FANCL	p.R68P	COSM1021961	Somatic	Endometrium & Large intestine	Carcinoma	1	1	p. R68Q reported instead (COSM1021961)
FANCL	p.F36L	COSM4991778	Somatic	Skin	Carcinoma	1	1	p.F36F reported instead
MLH1	p.K377E	COSM25915, COSM26083, COSM1422600	Somatic	Large intestine	Carcinoma	2	2	Reported as FS & p.K618T (1422600, 26083, 25915)
MLH3	p.V741F	COSM4595998 & COSM4595997	Unknown	Upper aerodigestive tract	Carcinoma	7	1	Reported as V741I
NBN	p.R43X	COSM1102344	Somatic	Endometrium, Large intestine & Oesophagus	Carcinoma	1	4	Reported 3 times as p.R43Q instead (COSM1102344)
RAD50	p.V127I	COSM3608463	Somatic	Skin	Malignant melanoma	1	1	Reported as V127V
RAD51B	p.D17G	COSM4843490	Somatic	Cervix	Carcinoma	1	1	p.D17Y reported instead (COSM4843490)
RAD51D	p.R185W	COSM4721160	Somatic	Large intestine	Carcinoma	1	1	Reported as p.R165C instead (COSM4721160)

RBBP8	p.R839Q	COSM3712642	Somatic	Upper aerodigestive tract	Carcinoma	1	1	Reported as p.R839* instead (COSM3712642)
RMI1	p.P287L	COSM3220409	Unknown	Large intestine	Carcinoma	1	1	p.P287S reported instead (COSM3220409)
SLX4	p.E813K	COSM285188	Somatic	Large intestine	Carcinoma	1	2	Reported twice as p.E813D instead (COSM285188)

Variants discussed in detail are in Red font.

Appendix U. FA/BRCA-HRR variants from the WES cohort with unconfirmed pathogenicity by HGMD.

Gene	Ref Seq	cDNA change	Amino acid change	Comments	References	Class*	WES ID
ATM	NM_000051	c.C998T	p.S333F	Colorectal cancer	(Tanskanen et al., 2015)	DM?	WES-14,WES- 22, WES-245
ATM	NM_000051	c.T4709C	p.V1570A	Breast Cancer susceptibility	(Dörk et al., 2001)	DM?	WES-24
ATM	NM_000051	c.G6860C	p.G2287A	Breast Cancer susceptibility	(Dörk et al., 2001)	DM?	WES-26
ATM	NM_000051	c.T1229C	p.V410A	Ocular telangiectasia	(Mauget-Faysse et al., 2003)	DM?	WES-30, WES-55
ATM	NM_000051	c.A4362C	p.K1454N	Breast Cancer susceptibility	(Tavtigian et al., 2009)	DM?	WES-96
ATM	NM_000051	c.C4138T	p.H1380Y	Breast Cancer susceptibility	(Atencio et al., 2001)	DM?	WES-203
BRCA1	NM_007298	c.G1644A	p.M548I	Breast and/or ovarian cancer	(Schoumacher et al., 2001, Woods et al., 2016, Figge et al., 2004)	DM?	WES-13, WES-60, WES-207
FANCC	NM_001243743	c.A584T	p.D195V	Fanconi anaemia	(Verlander et al., 1994)	DM?	WES-20, WES-79, WES-86
FANCL	NM_018062	c.C112T	p.L38F	Tetralogy of Fallot	(Grunert et al., 2014)	DM?	WES-14
FANCL	NM_018062	c.A670G	p.T224A	Tetralogy of Fallot	(Grunert et al., 2014)	DM?	WES-83
MLH1	NM_001258274	c.A1129G	p.K377E	Colorectal cancer, non-polyposis	Beck et al., 1997	DM?	WES-88, WES-218
MLH3	NM_014381	c.A1234G	p.K412E	Colorectal cancer, increased risk	(Liu et al., 2003)	DM?	WES-14
MLH3	NM_014381	c.G2221T	p.V741F	Endometrial cancer	(Liu et al., 2003)	DM?	WES-30, WES-41, WES-67, WES-209, WES-211, WES-226, WES-246
MRE11A	NM_005591	c.G1139A	p.R380H	Prostate cancer	(Leongamornlert et al., 2014)	DM?	WES-215
NBN	NM_002485	c.G456A	p.M152I	Ovarian cancer	(Ramus et al., 2015)	DM?	WES-55
RAD50	NM_005732	c.G980A	p.R327H	Breast cancer	(Tommiska et al., 2006)	DM?	WES-202

*DM?-Potentially damaging and/or pathogenic mutation with contradicting reports

Appendix V. Top 50 Differentially expressed genes with a >1.5 fold change for FA/BRCA-HRR network, FANC 19, FACN core & ID2, BLM & ATM/ATR checkpoint and KEGG-HRR subgroups

FA/BRCA-HRR network			FANC 19			FANC core & ID2			BLM & ATM/ATR checkpoint			KEGG-HRR		
Gene Name	**logFC**	**P.Value**	**Gene name**	**logFC**	**P.Value**	**Gene name**	**logFC**	**P.Value**	**Gene Name**	**logFC**	**P.Value**	**ILMN_Gene**	**logFC**	**P.Value**
SUCNR1	1.302	5.00E-06	SUCNR1	1.248	0	CD34	1.838	0.016	SUCNR1	1.134	0.001	HS.25318	1.206	0.006
LOC100132395	0.947	6.00E-05	CLEC11A	0.979	0.014	CD34	1.588	0.023	LOC389816	1.062	0.005	HOXA3	1.123	0.017
LOC100130000	0.862	9.00E-05	APOC2	0.928	0.024	LOC652377	1.073	0.03	ELANE	1.04	0.007	LOC100132395	1.102	0.002
ITPKA	0.841	4.00E-05	LOC648868	0.866	0.008	C1QTNF4	1.055	0.037	HSPA1B	1.015	0	C10ORF140	1.083	0.046
APOC2	0.819	2.00E-02	LOC284998	0.854	0.009	FLJ22536	1.01	0.009	CLEC14A	0.996	0.001	LOC100130000	1.079	0.001
ECM1	0.793	9.00E-04	LTC4S	0.842	0.006	SLC39A8	0.986	0.012	AIF1L	0.975	0.013	LOC649841	1.058	0.001
LOC100133080	0.774	2.00E-05	COPG2	0.742	0	LOC648868	0.958	0.015	EFCAB4A	0.954	0.001	SLC38A1	1.032	0.003
CLEC11A	0.773	2.00E-02	VTRNA1-1	0.707	0.02	LOC284998	0.942	0.017	F13A1	0.953	0.023	RAGE	1.021	0.01
SF1	0.77	5.00E-04	OLIG1	0.69	0.036	PRSSL1	0.937	0.012	NDN	0.948	0.046	LOC100133080	0.906	0.001
F13A1	0.735	3.00E-02	GHRL	0.689	0.003	TCTEX1D1	0.925	0.04	PALM	0.935	0.004	LOC400986	0.896	0.002
ZNF573	0.712	1.00E-04	STAG3L1	0.679	0.003	SUCNR1	0.887	0.035	CYP2S1	0.906	0.009	MTX3	0.884	0.003
ITPRIPL2	0.709	2.00E-04	TNFAIP8L2	0.674	0.006	COPG2	0.864	0	LRRC26	0.896	0.004	HS.549989	0.852	0.006
C5ORF20	0.701	3.00E-02	TMEM51	0.67	0.025	GHRL	0.829	0.003	AIF1L	0.86	0.031	HSPA1B	0.835	0.011
GOLPH4	0.696	8.00E-06	LPAR4	0.669	0.007	SERPING1	0.827	0.016	ITPKA	0.859	0.001	MLC1	0.832	0.005
MEF2C	0.688	2.00E-03	PRSSL1	0.662	0.035	ST3GAL4	0.803	0	LOC646723	0.847	0.032	C21ORF96	0.823	0.003
LOC646723	0.686	4.00E-02	USMG5	0.661	0.031	ENG	0.8	0.001	ALDH7A1	0.834	0.012	C10ORF114	0.819	0.012
TNFAIP8L2	0.681	9.00E-04	ACTA2	0.659	0.001	LPAR4	0.79	0.008	LOC100132395	0.819	0.005	TSC22D1	0.807	0.04
IL13RA1	0.671	2.00E-02	TSPAN4	0.655	0.011	STAG3L1	0.775	0.005	ALDH7A1	0.809	0.006	C20ORF94	0.799	0.001
LTC4S	0.659	1.00E-02	HSPA1B	0.649	0.015	TSPAN4	0.753	0.014	USMG5	0.751	0.015	ECM1	0.789	0.027
CLEC14A	0.657	6.00E-03	PIWIL4	0.645	0.012	CDKN2C	0.739	0.049	CD200	0.751	0.008	FAM91A2	0.781	0.001
NRXN2	0.655	2.00E-02	GPT2	0.643	0.009	RICS	0.732	0.032	SDK2	0.738	0.006	LRP5	0.763	0.007

FLT3	0.653	2.00E-04	FKBP2	0.641	0	CDH26	0.715	0.004	CSF1R	0.737	0.009	HS.505676	0.762	0.002
C21ORF96	0.648	5.00E-04	CDKN2C	0.633	0.043	CYLN2	0.677	0.017	VSIG4	0.734	0.004	LOC732450	0.757	0.001
CDC2L2	0.645	2.00E-05	LOC100132395	0.628	0.03	SORL1	0.675	0.005	LOC100130000	0.724	0.007	ZNF573	0.747	0.008
LOC284998	0.643	2.00E-02	CCDC71	0.624	0.001	ITM2C	0.67	0.033	NRXN2	0.714	0.04	GOLPH4	0.742	0.002
WDR49	0.643	1.00E-03	MGST1	0.62	0.003	SV2A	0.669	0.008	HGF	0.711	0.048	LOC653080	0.74	0.002
HS.549989	0.636	2.00E-03	TSPAN32	0.609	0.002	GPT2	0.652	0.027	EFNA1	0.71	0.027	SLC38A1	0.733	0.015
NFIC	0.634	3.00E-05	TARP	0.602	0.039	ITGA6	0.621	0.019	RHOBTB3	0.709	0.026	WIT1	0.73	0.022
GSN	0.634	2.00E-03	LOC100130000	0.591	0.027	ARHGAP10	0.616	0.02	DSE	0.703	0.003	CPXM1	0.725	0.034
LPPR3	0.634	1.00E-02	ITPKA	0.59	0.019	RAB37	0.608	0.021	PTGR1	0.699	0.003	LOC730417	0.725	0.034
HSPA1B	0.633	4.00E-03	RAB37	0.588	0.007	SESTD1	0.6	0.034	HS.10862	0.697	0.025	HOXA6	0.725	0.017
CCDC71	0.632	9.00E-05	CPXM1	0.588	0.033	LOC100132740	0.583	0.026	PTGR1	0.697	0.012	SERPINI2	0.715	0.038
GPX1	0.628	5.00E-04	LOC100133080	0.584	0.009	C21ORF128	0.582	0	MYOZ3	0.686	0.013	LOC644677	0.709	0.003
WDR49	0.626	5.00E-03	SIK1	-0.674	0.004	TRPC2	0.582	0.036	BZRAP1	0.677	0.028	HS.107418	0.709	0
RAB37	0.624	5.00E-04	SAMSN1	-0.685	0.011	PFKFB3	-0.701	0.006	FNBP1L	0.673	0.028	ACTA2	0.709	0.006
VSIG4	0.621	4.00E-03	TDRD9	-0.688	0.018	ITGAV	-0.702	0.012	LOC644615	0.667	0.002	RPAP2	0.705	0
STAG3L1	0.616	1.00E-03	TUFT1	-0.688	0.006	HOMER2	-0.718	0.026	LOC648868	0.666	0.045	DAD1L	0.702	0.015
AHDC1	0.615	5.00E-05	TKTL1	-0.688	0.024	DDEF2	-0.718	0.007	LOC149134	0.662	0.041	C21ORF96	0.695	0.001
CSF1R	0.614	9.00E-03	CXCL2	-0.693	0.05	NLRC5	-0.732	0.001	RAB13	0.659	0	LOC642678	0.693	0.007
HS.131041	0.611	5.00E-04	ETS1	-0.695	0.014	ZC3H12A	-0.733	0.007	ADCY6	0.646	0.003	ANKRD36	0.682	0.014
SNHG3-RCC1	0.609	3.00E-05	LRIG1	-0.697	0.013	EGR1	-0.733	0.015	CLEC12A	0.645	0.043	LOC730995	0.677	0.001
LOC642678	0.605	4.00E-04	PRDM1	-0.698	0.005	MAF	-0.735	0.017	NUDT7	0.639	0.005	TSC22D1	0.677	0.019
SPIN1	0.605	4.00E-06	PTGS2	-0.707	0.042	SIK1	-0.74	0.011	ZDHHC1	0.633	0.005	ANKRD36B	0.676	0.017
HS.443185	0.602	7.00E-04	ZAP70	-0.714	0.017	TGFBR3	-0.741	0.05	LOC651957	0.629	0.044	MACF1	0.675	0.006
FAM129A	0.601	3.00E-03	ZNF135	-0.717	0.011	ETS1	-0.75	0.027	LOC100133080	0.625	0.006	SLC39A10	0.675	0.001
LOC648868	0.6	3.00E-02	SETBP1	-0.719	0.006	STAT4	-0.751	0.027	TGIF1	0.621	0	PLGLB1	0.671	0.009
LOC647886	0.592	3.00E-05	ITPR3	-0.728	0.002	AZIN1	-0.754	0.015	TPM2	0.613	0.023	HS.452445	0.669	0.024
C12ORF24	0.59	7.00E-05	JAM3	-0.732	0.048	VNN3	-0.76	0.05	FAM129A	0.612	0.013	DBN1	0.667	0.04

LAT2	0.589	1.00E-03	SPRED1	-0.734	0.029	SLC25A24	-0.766	0.03	ME3	0.611	0.045	LOC100130887	0.665	0.004
CCR2	0.586	2.00E-02	SYTL4	-0.736	0.032	TDRD9	-0.77	0.027	ADA	0.61	0.005	DDX17	0.665	0.002
RAP1GAP	-0.915	5.00E-03	STAT4	-0.738	0.009	NR4A2	-0.77	0.024	CCL4L2	-0.768	0.015	CD36	-1.26	0.007
GBP1	-0.921	1.00E-04	LAMB2	-0.739	0.032	TSHZ3	-0.774	0.009	GZMA	-0.77	0.047	ANK1	-1.26	0.009
FHL2	-0.929	3.00E-04	GBP1	-0.745	0.011	MAFF	-0.775	0.007	CTLA4	-0.772	0.003	LOC728835	-1.26	0.006
ALAS2	-0.935	6.00E-03	HS.535044	-0.745	0	MGC3020	-0.778	0.004	LST1	-0.773	0.013	CLEC7A	-1.27	0.018
XK	-0.936	4.00E-03	FAM134B	-0.765	0.02	NAMPT	-0.783	0.03	LGALS3	-0.779	0.043	IFI30	-1.28	0.023
EPOR	-0.941	1.00E-04	TRO	-0.768	0.032	ECHDC2	-0.787	0.03	LEF1	-0.785	0.024	S100A8	-1.31	0.004
HBG2	-0.947	5.00E-02	LOC100127983	-0.776	0.007	TRIB1	-0.788	0.006	HBBP1	-0.795	0.018	EPSTI1	-1.31	0.001
TNFAIP3	-0.952	1.00E-05	SIK1	-0.779	0.001	PRKCA	-0.801	0.023	CDH1	-0.796	0.018	HBA1	-1.31	0.008
GZMH	-0.963	5.00E-03	TGFBR3	-0.786	0.012	LOC100131831	-0.807	0.014	GZMK	-0.798	0.028	CD36	-1.31	0.009
IFI27	-0.964	1.00E-02	RGS1	-0.79	0.023	ARL4C	-0.842	0.001	CD8A	-0.803	0.048	TNF	-1.31	0.006
RGS1	-0.964	8.00E-04	SNRPN	-0.795	0.018	NAMPT	-0.845	0.02	CD6	-0.804	0.018	FOLR3	-1.31	0.041
TRIM10	-0.967	6.00E-03	MARCKS	-0.797	0.039	MCOLN2	-0.858	0.038	ITGA2B	-0.815	0.003	CA2	-1.33	0.003
RHAG	-0.977	2.00E-02	BAMBI	-0.808	0.016	LRIG1	-0.874	0.009	HS.554324	-0.834	0.012	CTSL1	-1.34	0.001
HBQ1	-0.986	3.00E-03	NFIA	-0.817	0.023	BAMBI	-0.874	0.03	MAF	-0.834	0.001	SELENBP1	-1.34	0.013
CDH1	-0.987	3.00E-04	FAM89A	-0.853	0.002	CXCL2	-0.884	0.036	HEMGN	-0.835	0.047	CCL4L1	-1.36	0
TGFBR3	-0.997	1.00E-04	ATP9A	-0.857	0.024	IL18RAP	-0.885	0.031	IFI44L	-0.835	0.021	ALDH1A1	-1.37	0.006
GBP1	-0.997	3.00E-05	MAF	-0.858	0.001	PTGS2	-0.906	0.029	RHCE	-0.844	0.05	CA2	-1.37	0.002
JAM3	-1.004	1.00E-03	PTGS2	-0.858	0.008	IGF2R	-0.907	0.02	CA2	-0.849	0.022	CRISPLD2	-1.38	0.009
KEL	-1.004	4.00E-03	F2RL1	-0.862	0.014	JAM3	-0.911	0.04	BCL11B	-0.862	0.032	OSBP2	-1.38	0.012
CCL4L1	-1.007	1.00E-04	PTRF	-0.866	0.044	EGR2	-0.917	0.028	KEL	-0.866	0.042	RHOU	-1.39	0.001
GNLY	-1.014	2.00E-03	SOCS2	-0.875	0.028	IRS2	-0.929	0.002	FAM178B	-0.87	0.008	FPR1	-1.41	0.023
ANK1	-1.015	2.00E-03	ALDH1A1	-0.879	0.03	SUCLG2	-0.941	0.015	KCNH2	-0.886	0.008	NPL	-1.41	0
HBE1	-1.021	6.00E-03	MCOLN2	-0.88	0.011	BCL11B	-0.946	0.048	GZMH	-0.892	0.034	BEX2	-1.41	0.014
LOC728835	-1.029	7.00E-04	NR4A2	-0.887	0.009	PTGS2	-0.963	0.013	MAL	-0.901	0.024	LGALS3	-1.44	0.002
AHSP	-1.03	8.00E-03	GBP1	-0.895	0.002	IL8	-1.021	0.017	TNF	-0.917	0.02	KEL	-1.44	0.006

HBG1	-1.036	3.00E-02	MYH10	-0.902	0.021	NR4A2	-1.029	0.012	CCL4L1	-0.925	0.004	RHAG	-1.48	0.014
MYH10	-1.04	1.00E-03	ARL4C	-0.903	0	RGS1	-1.031	0.013	CD3D	-0.937	0.025	IFI27	-1.48	0.008
LGALS3	-1.043	9.00E-04	IFITM3	-0.904	0.032	HOXB2	-1.039	0.039	SERPINE2	-0.951	0.016	CD14	-1.48	0.029
GYPB	-1.05	2.00E-02	TNFAIP3	-0.973	0	CLIP3	-1.055	0.031	LOC728835	-0.955	0.01	S100P	-1.5	0.021
GNLY	-1.051	6.00E-03	KCNK12	-0.985	0.001	BASP1	-1.068	0.041	KLRF1	-0.963	0.003	XK	-1.5	0.002
SELENBP1	-1.062	3.00E-03	KLF9	-1.008	0.001	TNF	-1.095	0.018	FGFBP2	-0.966	0.018	MYL4	-1.5	0.012
OSBP2	-1.083	4.00E-03	SERPINE2	-1.12	0.004	GNG11	-1.139	0.027	MCOLN2	-0.967	0.005	HLA-DRB6	-1.53	0.016
MGC13057	-1.087	2.00E-04	GNG11	-1.211	0.005	TNFAIP3	-1.184	0	MGC13057	-0.969	0.006	C19ORF59	-1.54	0.01
HMBS	-1.091	9.00E-03				TPSAB1	-1.718	0.009	EPSTI1	-0.973	0.002	MAFB	-1.55	0.017
HBM	-1.1	2.00E-02							C7ORF28B	-0.989	0.007	NPL	-1.56	0
ALDH1A1	-1.11	8.00E-05							CD3D	-0.992	0.014	S100A12	-1.57	0.019
HEMGN	-1.111	3.00E-03							ALDH1A1	-1.007	0.003	HMBS	-1.61	0.009
BEX2	-1.139	3.00E-03							SERPINB2	-1.028	0.015	CTSL1	-1.62	0
HEMGN	-1.159	8.00E-04							PROK2	-1.033	0.02	CD14	-1.67	0.03
EPSTI1	-1.159	5.00E-06							LOC644936	-1.033	0.005	RETN	-1.71	0.003
MYL4	-1.163	4.00E-03							CES1	-1.038	0.034	MARCKS	-1.76	0
NFIA	-1.172	7.00E-05							PRF1	-1.053	0.003	LOC100131164	-1.78	0.006
CA1	-1.208	5.00E-03							MARCKS	-1.058	0.007	GYPB	-1.84	0.004
MARCKS	-1.216	1.00E-04							MYL4	-1.07	0.028	AHSP	-1.91	0.001
LOC100131164	-1.222	5.00E-03							VENTX	-1.078	0.021	EPB42	-1.93	0.003
GZMB	-1.226	3.00E-04							GZMB	-1.112	0.008	ALAS2	-2.05	0.002
EPB42	-1.252	5.00E-03							GNLY	-1.17	0.003	HBM	-2.07	0.002
ALAS2	-1.281	4.00E-03							GNLY	-1.229	0.008	CA1	-2.15	0.001
SERPINE2	-1.419	8.00E-06							SERPINB2	-1.271	0.005	HBG2	-2.38	0.001
ALDH1A1	-1.566	2.00E-06							ALDH1A1	-1.317	0.001	HBG1	-2.49	0

Appendix W. Top 50 positively correlated gene sets with FDR<25% for FA/BRCA-HRR network, FANC 19, FACN core & ID2, BLM & ATM/ATR checkpoint and KEGG-HRR subgroups.

FA/BRCA-HRR network					
Name of Gene set	Size	NES	NOM p-val	FDR q-val	FWER p-val
KIM_ALL_DISORDERS_DURATION_CORR_DN	117	1.866	0.000	0.207	0.622
REACTOME_SYNTHESIS_OF_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI	15	1.865	0.000	0.155	0.622
REACTOME_PEROXISOMAL_LIPID_METABOLISM	18	1.858	0.000	0.137	0.655
NIKOLSKY_BREAST_CANCER_20Q11_AMPLICON	22	1.856	0.001	0.118	0.670
MENSSEN_MYC_TARGETS	49	1.841	0.000	0.126	0.744
REACTOME_BASE_EXCISION_REPAIR	16	1.829	0.001	0.129	0.798
REACTOME_TRNA_AMINOACYLATION	38	1.792	0.000	0.187	0.929
FANC 19 subgroup					
KEGG_GLYCOSAMINOGLYCAN_DEGRADATION	15	2.013	0.000	0.063	0.096
WONG_MITOCHONDRIA_GENE_MODULE	186	1.999	0.000	0.042	0.129
REACTOME_TELOMERE_MAINTENANCE	52	1.980	0.001	0.039	0.178
KIM_ALL_DISORDERS_DURATION_CORR_DN	117	1.976	0.000	0.031	0.189
REACTOME_MEIOTIC_RECOMBINATION	44	1.933	0.000	0.048	0.328
REACTOME_PEROXISOMAL_LIPID_METABOLISM	18	1.932	0.000	0.041	0.332
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS_	73	1.918	0.000	0.044	0.394
IVANOVA_HEMATOPOIESIS_INTERMEDIATE_PROGENITOR	114	1.908	0.000	0.044	0.437
NIKOLSKY_BREAST_CANCER_20Q11_AMPLICON	22	1.893	0.001	0.050	0.517
YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTER_17	159	1.891	0.000	0.046	0.530
MOOTHA_VOXPHOS	73	1.881	0.000	0.048	0.577
CREIGHTON_AKT1_SIGNALING_VIA_MTOR_DN	19	1.870	0.000	0.051	0.630
YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTER_13	145	1.858	0.000	0.056	0.695
KEGG_SELENOAMINO_ACID_METABOLISM	17	1.855	0.001	0.054	0.705
REACTOME_TRNA_AMINOACYLATION	38	1.855	0.000	0.051	0.707
KEGG_PEROXISOME	56	1.851	0.001	0.050	0.724
REACTOME_DNA_STRAND_ELONGATION	27	1.846	0.000	0.051	0.756
SCHLOSSER_SERUM_RESPONSE_AUGMENTED_BY_MYC	82	1.839	0.000	0.052	0.790
NIKOLSKY_BREAST_CANCER_11Q12_Q14_AMPLICON	90	1.837	0.000	0.051	0.800
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	58	1.832	0.002	0.052	0.831
KEGG_PYRIMIDINE_METABOLISM	79	1.828	0.001	0.052	0.838
KEEN_RESPONSE_TO_ROSIGLITAZONE_UP	29	1.825	0.001	0.052	0.850
REACTOME_BASE_EXCISION_REPAIR	16	1.820	0.003	0.054	0.869
LU_EZH2_TARGETS_UP	210	1.817	0.000	0.053	0.879
KEGG_LYSOSOME	101	1.808	0.000	0.057	0.910
MOOTHA_HUMAN_MITODB_6_2002	345	1.805	0.000	0.058	0.919
DACOSTA_UV_RESPONSE_VIA_ERCC3_TTD_UP	52	1.804	0.000	0.056	0.919
PURBEY_TARGETS_OF_CTBP1_AND_SATB1_DN	121	1.799	0.000	0.057	0.932
KEGG_PARKINSONS_DISEASE	88	1.798	0.000	0.056	0.934
KEGG_OXIDATIVE_PHOSPHORYLATION	94	1.796	0.000	0.055	0.938

REACTOME_POST_TRANSLATIONAL_PROTEIN_MODIFICATION	120	1.794	0.000	0.055	0.942
KEGG_BASE_EXCISION_REPAIR	30	1.793	0.001	0.054	0.944
MOOTHA_MITOCHONDRIA	356	1.792	0.000	0.052	0.946
MOREAUX_B_LYMPHOCYTE_MATURATION_BY_TACI_DN	64	1.792	0.000	0.051	0.946
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	103	1.791	0.000	0.050	0.947
BIOCARTA_MITOCHONDRIA_PATHWAY	18	1.782	0.003	0.055	0.962
REACTOME_EXTENSION_OF_TELOMERES	24	1.782	0.001	0.053	0.962
KEGG_PURINE_METABOLISM	105	1.772	0.000	0.058	0.979
MULLIGHAN_MLL_SIGNATURE_1_UP	314	1.772	0.000	0.057	0.979
REACTOME_MITOCHONDRIAL_TRNA_AMINOACYLATION	19	1.767	0.001	0.059	0.985
MALONEY_RESPONSE_TO_17AAG_UP	29	1.764	0.001	0.060	0.987
KEGG_GALACTOSE_METABOLISM	19	1.764	0.003	0.058	0.987
CREIGHTON_AKT1_SIGNALING_VIA_MTOR_UP	27	1.756	0.001	0.062	0.989
SPIELMAN_LYMPHOBLAST_EUROPEAN_VS_ASIAN_UP	416	1.749	0.000	0.067	0.992
MENSSEN_MYC_TARGETS	49	1.746	0.004	0.067	0.993
DAIRKEE_TERT_TARGETS_UP	283	1.742	0.000	0.070	0.996
REACTOME_LAGGING_STRAND_SYNTHESIS	18	1.738	0.007	0.071	0.996
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	37	1.737	0.004	0.071	0.996
LI_DCP2_BOUND_MRNA	83	1.729	0.000	0.076	0.997
KEGG_HOMOLOGOUS_RECOMBINATION	21	1.726	0.004	0.077	0.997
FANC core & ID2 subgroup					
REACTOME_DNA_STRAND_ELONGATION	27	2.432	0.000	0.000	0.000
KEGG_DNA_REPLICATION	34	2.347	0.000	0.000	0.000
REACTOME_EXTENSION_OF_TELOMERES	24	2.196	0.000	0.001	0.005
KEGG_BASE_EXCISION_REPAIR	30	2.176	0.000	0.001	0.009
NIKOLSKY_BREAST_CANCER_11Q12_Q14_AMPLICON	90	2.169	0.000	0.001	0.009
REACTOME_LAGGING_STRAND_SYNTHESIS	18	2.165	0.000	0.001	0.010
DUTERTRE_ESTRADIOL_RESPONSE_24HR_UP	231	2.146	0.000	0.002	0.021
REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX	19	2.134	0.000	0.002	0.024
REACTOME_BASE_EXCISION_REPAIR	16	2.089	0.000	0.003	0.043
REACTOME_DNA_REPAIR	82	2.044	0.000	0.006	0.084
REACTOME_G2_M_CHECKPOINTS	29	2.027	0.000	0.007	0.113
REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS	25	2.014	0.002	0.008	0.139
SONG_TARGETS_OF_IE86_CMV_PROTEIN	52	2.005	0.000	0.009	0.158
OXFORD_RALA_OR_RALB_TARGETS_UP	40	1.973	0.000	0.013	0.248
FRASOR_RESPONSE_TO_SERM_OR_FULVESTRANT_DN	44	1.953	0.000	0.016	0.322
KEGG_HOMOLOGOUS_RECOMBINATION	21	1.940	0.002	0.018	0.371
KEGG_PURINE_METABOLISM	105	1.920	0.000	0.023	0.465
KAUFFMANN_DNA_REPAIR_GENES	181	1.902	0.000	0.029	0.562
OUELLET_CULTURED_OVARIAN_CANCER_INVASIVE_VS_LMP_UP	54	1.898	0.000	0.029	0.581
HONMA_DOCETAXEL_RESISTANCE	32	1.898	0.000	0.028	0.581
MISSIAGLIA_REGULATED_BY_METHYLATION_DN	100	1.888	0.000	0.029	0.621
CUI_TCF21_TARGETS_2_UP	244	1.888	0.000	0.028	0.622
DACOSTA_UV_RESPONSE_VIA_ERCC3_TTD_UP	52	1.886	0.000	0.028	0.633
REACTOME_SYNTHESIS_OF_DNA	76	1.881	0.000	0.028	0.654
REACTOME_DOUBLE_STRAND_BREAK_REPAIR	15	1.874	0.000	0.030	0.690

LUI_THYROID_CANCER_CLUSTER_3	26	1.866	0.000	0.032	0.726
MOREAUX_B_LYMPHOCYTE_MATURATION_BY_TACI_DN	64	1.855	0.000	0.035	0.765
BUYTAERT_PHOTODYNAMIC_THERAPY_STRESS_DN	441	1.837	0.000	0.042	0.827
TOOKER_GEMCITABINE_RESISTANCE_UP	68	1.832	0.000	0.044	0.862
MORI_LARGE_PRE_BII_LYMPHOCYTE_UP	74	1.825	0.000	0.047	0.887
WAKASUGI_HAVE_ZNF143_BINDING_SITES	50	1.821	0.003	0.047	0.898
REN_BOUND_BY_E2F	52	1.819	0.000	0.047	0.902
HOLLEMAN_ASPARAGINASE_RESISTANCE_B_ALL_UP	23	1.817	0.006	0.047	0.908
ALCALAY_AML_BY_NPM1_LOCALIZATION_DN	151	1.809	0.000	0.050	0.930
REACTOME_GLOBAL_GENOMIC_NER_GG_NER	31	1.809	0.003	0.049	0.930
KEGG_PYRIMIDINE_METABOLISM	79	1.786	0.000	0.062	0.973
MORI_IMMATURE_B_LYMPHOCYTE_DN	79	1.784	0.000	0.062	0.975
REACTOME_SYNTHESIS_OF_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI	15	1.780	0.005	0.063	0.980
KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	22	1.778	0.003	0.063	0.981
KEGG_NUCLEOTIDE_EXCISION_REPAIR	41	1.766	0.000	0.070	0.986
REACTOME_POST_TRANSLATIONAL_PROTEIN_MODIFICATION	120	1.765	0.000	0.069	0.986
SPIELMAN_LYMPHOBLAST_EUROPEAN_VS_ASIAN_UP	416	1.761	0.000	0.070	0.987
REACTOME_NUCLEOTIDE_EXCISION_REPAIR	44	1.755	0.003	0.075	0.993
LI_DCP2_BOUND_MRNA	83	1.753	0.001	0.074	0.994
ISHIDA_E2F_TARGETS	46	1.752	0.003	0.073	0.995
YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTER_13	145	1.748	0.000	0.075	0.996
REACTOME_TRANSCRIPTION_COUPLED_NER_TC_NER	39	1.748	0.002	0.073	0.996
REACTOME_S_PHASE	91	1.747	0.000	0.072	0.997
NIKOLSKY_BREAST_CANCER_20Q11_AMPLICON	22	1.738	0.003	0.078	0.998
KEGG_MISMATCH_REPAIR	22	1.738	0.000	0.077	0.998
BLM & ATM/ATR checkpoint subgroup					
JAATINEN_HEMATOPOIETIC_STEM_CELL_UP	222	1.995	0.000	0.094	0.130
IVANOVA_HEMATOPOIESIS_INTERMEDIATE_PROGENITOR	114	1.975	0.000	0.067	0.180
REACTOME_INHIBITION_OF_THE_PROTEOLYTIC_ACTIVITY_OF_APC_C_REQUIRED_FOR_THE_ONSET_OF_ANAPHASE_BY_MITOTIC_SPINDLE_CHECKPOINT_COMPONENTS	16	1.918	0.000	0.113	0.392
NIKOLSKY_BREAST_CANCER_1Q21_AMPLICON	24	1.906	0.000	0.101	0.443
REACTOME_SYNTHESIS_OF_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI	15	1.901	0.000	0.087	0.463
RHEIN_ALL_GLUCOCORTICOID_THERAPY_DN	320	1.877	0.000	0.102	0.580
KEGG_PROPANOATE_METABOLISM	27	1.871	0.003	0.095	0.608
REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_CYCLIN_B	16	1.871	0.001	0.083	0.609
REACTOME_APC_CDC20_MEDIATED_DEGRADATION_OF_NEK2A	19	1.852	0.000	0.095	0.703
GARGALOVIC_RESPONSE_TO_OXIDIZED_PHOSPHOLIPIDS_RED_DN	20	1.823	0.001	0.125	0.836
LUI_THYROID_CANCER_CLUSTER_3	26	1.817	0.001	0.124	0.865
BILANGES_SERUM_AND_RAPAMYCIN_SENSITIVE_GENES	63	1.787	0.001	0.168	0.944
RIZ_ERYTHROID_DIFFERENTIATION_12HR	22	1.778	0.001	0.174	0.963
EPPERT_LSC_R	27	1.759	0.008	0.204	0.989
KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	22	1.738	0.007	0.248	0.996
KEGG HRR subgroup					
GAZDA_DIAMOND_BLACKFAN_ANEMIA_PROGENITOR_DN	50	1.973	0.000	0.153	0.252

JAATINEN_HEMATOPOIETIC_STEM_CELL_UP	222	1.954	0.000	0.102	0.325
KANG_FLUOROURACIL_RESISTANCE_UP	17	1.869	0.003	0.235	0.732
NIKOLSKY_BREAST_CANCER_20Q11_AMPLICON	22	1.868	0.000	0.177	0.732
REACTOME_SYNTHESIS_OF_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI	15	1.863	0.002	0.150	0.749
FAELT_B_CLL_WITH_VH_REARRANGEMENTS_DN	42	1.843	0.001	0.158	0.826
REACTOME_CHOLESTEROL_BIOSYNTHESIS	17	1.836	0.002	0.147	0.852
MANALO_HYPOXIA_DN	235	1.826	0.000	0.145	0.883
HEIDENBLAD_AMPLIFIED_IN_PANCREATIC_CANCER	22	1.798	0.003	0.177	0.948
RICKMAN_METASTASIS_UP	226	1.795	0.000	0.165	0.951
MOREAUX_MULTIPLE_MYELOMA_BY_TACI_DN	142	1.768	0.000	0.204	0.977
KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	22	1.767	0.005	0.188	0.977
ZUCCHI_METASTASIS_UP	31	1.762	0.003	0.185	0.986
HORTON_SREBF_TARGETS	19	1.758	0.003	0.178	0.989
WALLACE_PROSTATE_CANCER_UP	17	1.744	0.006	0.195	0.995
GALE_APL_WITH_FLT3_MUTATED_UP	44	1.743	0.003	0.186	0.995
BIOCARTA_MITOCHONDRIA_PATHWAY	18	1.721	0.005	0.223	0.999
REACTOME_MITOCHONDRIAL_TRNA_AMINOACYLATION	19	1.720	0.003	0.212	0.999
SCHMIDT_POR_TARGETS_IN_LIMB_BUD_UP	19	1.717	0.005	0.207	0.999
PID_NCADHERIN_PATHWAY	24	1.705	0.011	0.222	1.000
ASGHARZADEH_NEUROBLASTOMA_POOR_SURVIVAL_DN	23	1.695	0.009	0.235	1.000
WHITFIELD_CELL_CYCLE_G1_S	105	1.692	0.001	0.231	1.000
SCHUHMACHER_MYC_TARGETS_UP	72	1.691	0.001	0.224	1.000
MOREAUX_B_LYMPHOCYTE_MATURATION_BY_TACI_DN	64	1.673	0.003	0.246	1.000
MENSSEN_MYC_TARGETS	49	1.669	0.006	0.247	1.000
BUYTAERT_PHOTODYNAMIC_THERAPY_STRESS_DN	441	1.668	0.000	0.239	1.000
KEGG_SELENOAMINO_ACID_METABOLISM	17	1.666	0.003	0.236	1.000

Appendix X. Top 50 negatively correlated gene sets with FDR<25% for FA/BRCA-HRR network, FANC 19, FACN core & ID2, BLM & ATM/ATR checkpoint and KEGG-HRR subgroups.

FA/BRCA-HRR network					
Name of Gene set	Size	NES	NOM p-val	FDR q-val	FWER p-val
JAATINEN_HEMATOPOIETIC_STEM_CELL_DN	159	-3.609	0	0	0
VALK_AML_CLUSTER_8	18	-3.144	0	0	0
VALK_AML_CLUSTER_7	20	-3.057	0	0	0
MOSERLE_IFNA_RESPONSE	23	-2.997	0	0	0
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_8D_UP	105	-2.868	0	0	0
GRAHAM_CML_QUIESCENT_VS_NORMAL_QUIESCENT_UP	73	-2.837	0	0	0
HECKER_IFNB1_TARGETS	63	-2.834	0	0	0
ROSS_AML_OF_FAB_M7_TYPE	57	-2.797	0	0	0
ALTEMEIER_RESPONSE_TO_LPS_WITH_MECHANICAL_VENTILATION	88	-2.785	0	0	0
PICCALUGA_ANGIOIMMUNOBLASTIC_LYMPHOMA_DN	110	-2.779	0	0	0
VILIMAS_NOTCH1_TARGETS_UP	38	-2.738	0	0	0
FARMER_BREAST_CANCER_CLUSTER_1	33	-2.732	0	0	0
BROWNE_INTERFERON_RESPONSIVE_GENES	54	-2.711	0	0	0
MAHADEVAN_RESPONSE_TO_MP470_UP	15	-2.700	0	0	0
LIANG_SILENCED_BY_METHYLATION_2	30	-2.698	0	0	0
IVANOVA_HEMATOPOIESIS_MATURE_CELL	211	-2.685	0	0	0
GRAHAM_CML_QUIESCENT_VS_NORMAL_DIVIDING_UP	42	-2.668	0	0	0
KEGG_GRAFT_VERSUS_HOST_DISEASE	27	-2.665	0	0	0
REICHERT_MITOSIS_LIN9_TARGETS	23	-2.659	0	0	0
WALLACE_PROSTATE_CANCER_RACE_UP	193	-2.652	0	0	0
YU_MYC_TARGETS_UP	37	-2.621	0	0	0
NAKAYAMA_SOFT_TISSUE_TUMORS_PCA2_UP	53	-2.615	0	0	0
SEITZ_NEOPLASTIC_TRANSFORMATION_BY_8P_DELETION_UP	43	-2.603	0	0	0
HAHTOLA_SEZARY_SYNDROM_DN	27	-2.571	0	0	0
TANG_SENESCENCE_TP53_TARGETS_DN	44	-2.545	0	0	0
GRAHAM_CML_DIVIDING_VS_NORMAL_QUIESCENT_UP	159	-2.534	0	0	0
BOWIE_RESPONSE_TO_TAMOXIFEN	17	-2.530	0	0	0
FARMER_BREAST_CANCER_CLUSTER_2	30	-2.518	0	0	0
KAN_RESPONSE_TO_ARSENIC_TRIOXIDE	76	-2.507	0	0	0
TIAN_TNF_SIGNALING_VIA_NFKB	21	-2.506	0	0	0
BIOCARTA_CTLA4_PATHWAY	16	-2.501	0	0	0
WILLIAMS_ESR1_TARGETS_UP	15	-2.491	0	0	0
KEGG_AUTOIMMUNE_THYROID_DISEASE	24	-2.488	0	0	0
REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_NON_LYMPHOID_CELL	39	-2.485	0	0	0
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_10D_UP	121	-2.472	0	7E-05	1E-03
ODONNELL_TFRC_TARGETS_DN	89	-2.469	0	7E-05	1E-03
KRASNOSELSKAYA_ILF3_TARGETS_UP	27	-2.468	0	7E-05	1E-03
SEKI_INFLAMMATORY_RESPONSE_LPS_UP	50	-2.466	0	7E-05	1E-03
KEGG_ALLOGRAFT_REJECTION	24	-2.444	0	1E-04	2E-03
GREENBAUM_E2A_TARGETS_UP	30	-2.435	0	1E-04	2E-03
BOSCO_TH1_CYTOTOXIC_MODULE	58	-2.434	0	1E-04	2E-03
WINZEN_DEGRADED_VIA_KHSRP	49	-2.434	0	1E-04	2E-03
SANA_TNF_SIGNALING_UP	52	-2.416	0	1E-04	2E-03
GRAHAM_NORMAL_QUIESCENT_VS_NORMAL_DIVIDING_DN	78	-2.391	0	2E-04	4E-03
FINETTI_BREAST_CANCER_KINOME_RED	15	-2.382	0	2E-04	4E-03
LINDSTEDT_DENDRITIC_CELL_MATURATION_A	40	-2.381	0	2E-04	4E-03
ACEVEDO_NORMAL_TISSUE_ADJACENT_TO_LIVER_TUMOR_UP	123	-2.374	0	3E-04	5E-03

SHAFFER_IRF4_MULTIPLE_MYELOMA_PROGRAM	31	-2.370	0	3E-04	5E-03
LINDSTEDT_DENDRITIC_CELL_MATURATION_B	37	-2.361	0	4E-04	7E-03
PID_IL12_2PATHWAY	49	-2.347	0	5E-04	9E-03
FANC 19 subgroup					
PICCALUGA_ANGIOIMMUNOBLASTIC_LYMPHOMA_DN	110	-3.494	0	0	0
NAGASHIMA_NRG1_SIGNALING_UP	122	-2.730	0	0	0
ZWANG_CLASS_3_TRANSIENTLY_INDUCED_BY_EGF	143	-2.728	0	0	0
DEURIG_T_CELL_PROLYMPHOCYTIC_LEUKEMIA_DN	242	-2.636	0	6.4E-04	1.0E-03
ACEVEDO_NORMAL_TISSUE_ADJACENT_TO_LIVER_TUMOR_UP	123	-2.619	0	5.1E-04	1.0E-03
DAUER_STAT3_TARGETS_UP	33	-2.619	0	4.3E-04	1.0E-03
GENTILE_UV_RESPONSE_CLUSTER_D2	34	-2.593	0	3.7E-04	1.0E-03
SEKI_INFLAMMATORY_RESPONSE_LPS_UP	50	-2.537	0	3.2E-04	1.0E-03
UZONYI_RESPONSE_TO_LEUKOTRIENE_AND_THROMBIN	26	-2.516	0	5.8E-04	2.0E-03
KEGG_GRAFT_VERSUS_HOST_DISEASE	27	-2.501	0	7.7E-04	3.0E-03
VALK_AML_CLUSTER_8	18	-2.492	0	7.0E-04	3.0E-03
GRAHAM_CML_QUIESCENT_VS_NORMAL_DIVIDING_UP	42	-2.489	0	6.4E-04	3.0E-03
KEGG_ALLOGRAFT_REJECTION	24	-2.478	0	5.9E-04	3.0E-03
HADDAD_T_LYMPHOCYTE_AND_NK_PROGENITOR_UP	58	-2.460	0	5.5E-04	3.0E-03
JAATINEN_HEMATOPOIETIC_STEM_CELL_DN	159	-2.453	0	6.7E-04	4.0E-03
AMIT_EGF_RESPONSE_60_HELA	33	-2.449	0	6.3E-04	4.0E-03
KEGG_AUTOIMMUNE_THYROID_DISEASE	24	-2.441	0	5.9E-04	4.0E-03
VILIMAS_NOTCH1_TARGETS_UP	38	-2.417	0	5.6E-04	4.0E-03
BURTON_ADIPOGENESIS_1	28	-2.387	0	9.3E-04	7.0E-03
GHANDHI_DIRECT_IRRADIATION_UP	52	-2.375	0	1.3E-03	1.0E-02
SEIDEN_ONCOGENESIS_BY_MET	72	-2.357	0	1.3E-03	1.1E-02
BIOCARTA_CTLA4_PATHWAY	16	-2.347	0	1.5E-03	1.3E-02
NAGASHIMA_EGF_SIGNALING_UP	40	-2.342	0	1.5E-03	1.4E-02
BROCKE_APOPTOSIS_REVERSED_BY_IL6	113	-2.335	0	1.7E-03	1.6E-02
GAVIN_FOXP3_TARGETS_CLUSTER_P7	53	-2.326	0	1.8E-03	1.8E-02
WINZEN_DEGRADED_VIA_KHSRP	49	-2.293	0	2.3E-03	2.4E-02
SHAFFER_IRF4_MULTIPLE_MYELOMA_PROGRAM	31	-2.292	0	2.2E-03	2.4E-02
GAURNIER_PSMD4_TARGETS	39	-2.289	0	2.3E-03	2.6E-02
LINDSTEDT_DENDRITIC_CELL_MATURATION_B	37	-2.282	0	2.3E-03	2.7E-02
PHONG_TNF_TARGETS_UP	43	-2.282	0	2.3E-03	2.7E-02
ZHANG_RESPONSE_TO_IKK_INHIBITOR_AND_TNF_UP	143	-2.281	0	2.2E-03	2.7E-02
BURTON_ADIPOGENESIS_12	26	-2.278	0	2.2E-03	2.8E-02
SCHOEN_NFKB_SIGNALING	17	-2.277	0	2.1E-03	2.8E-02
ONDER_CDH1_TARGETS_3_DN	19	-2.275	0	2.1E-03	2.8E-02
VALK_AML_CLUSTER_7	20	-2.271	0	2.2E-03	3.1E-02
VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_UP	70	-2.257	0	2.4E-03	3.5E-02
DUTERTRE_ESTRADIOL_RESPONSE_24HR_DN	302	-2.254	0	2.5E-03	3.6E-02
KEGG_TYPE_I_DIABETES_MELLITUS	28	-2.244	0	2.6E-03	3.8E-02
MAHADEVAN_RESPONSE_TO_MP470_UP	15	-2.240	0	2.7E-03	4.1E-02
GROSS_HYPOXIA_VIA_ELK3_ONLY_UP	23	-2.236	0	2.8E-03	4.3E-02
HOWLIN_CITED1_TARGETS_1_DN	28	-2.229	0	3.1E-03	4.9E-02
RIGGINS_TAMOXIFEN_RESISTANCE_DN	146	-2.223	0	3.2E-03	5.1E-02
PLASARI_TGFB1_TARGETS_1HR_UP	21	-2.215	0	3.2E-03	5.3E-02
WALLACE_PROSTATE_CANCER_RACE_UP	193	-2.207	0	3.4E-03	5.8E-02
GESERICK_TERT_TARGETS_DN	16	-2.206	0.003 1646	3.4E-03	5.9E-02
SMIRNOV_RESPONSE_TO_IR_2HR_UP	36	-2.200	0	3.4E-03	6.0E-02
AMIT_EGF_RESPONSE_40_HELA	34	-2.194	0	3.5E-03	6.4E-02
BOYAULT_LIVER_CANCER_SUBCLASS_G5_DN	24	-2.194	0	3.5E-03	6.4E-02
WIERENGA_STAT5A_TARGETS_GROUP2	42	-2.182	0	4.0E-03	7.5E-02
GABRIELY_MIR21_TARGETS	197	-2.179	0	4.0E-03	7.7E-02
FANC core & ID2 subgroup					
NAGASHIMA_NRG1_SIGNALING_UP	122	-3.357	0	0	0
PICCALUGA_ANGIOIMMUNOBLASTIC_LYMPHOMA_DN	110	-3.307	0	0	0

ZWANG_CLASS_3_TRANSIENTLY_INDUCED_BY_EGF	143	-2.993	0	0	0
NAGASHIMA_EGF_SIGNALING_UP	40	-2.916	0	0	0
DIRMEIER_LMP1_RESPONSE_EARLY	54	-2.866	0	0	0
UZONYI_RESPONSE_TO_LEUKOTRIENE_AND_THROMBIN	26	-2.836	0	0	0
PHONG_TNF_TARGETS_UP	43	-2.809	0	0	0
PRAMOONJAGO_SOX4_TARGETS_UP	40	-2.794	0	0	0
GRAHAM_CML_QUIESCENT_VS_NORMAL_DIVIDING_UP	42	-2.729	0	0	0
SEKI_INFLAMMATORY_RESPONSE_LPS_UP	50	-2.654	0	0	0
GALINDO_IMMUNE_RESPONSE_TO_ENTEROTOXIN	71	-2.628	0	0	0
AMIT_EGF_RESPONSE_40_HELA	34	-2.626	0	0	0
BURTON_ADIPOGENESIS_PEAK_AT_2HR	43	-2.616	0	0	0
BURTON_ADIPOGENESIS_1	28	-2.590	0	0	0
PODAR_RESPONSE_TO_ADAPHOSTIN_UP	113	-2.588	0	0	0
DAUER_STAT3_TARGETS_UP	33	-2.566	0	0	0
WIERENGA_STAT5A_TARGETS_GROUP2	42	-2.551	0	0	0
AMIT_SERUM_RESPONSE_40_MCF10A	24	-2.537	0	0	0
ZHANG_RESPONSE_TO_IKK_INHIBITOR_AND_TNF_UP	143	-2.524	0	0	0
LINDSTEDT_DENDRITIC_CELL_MATURATION_B	37	-2.522	0	0	0
ZWANG_CLASS_1_TRANSIENTLY_INDUCED_BY_EGF	287	-2.517	0	0	0
MITSIADES_RESPONSE_TO_APLIDIN_UP	318	-2.514	0	0	0
ZHOU_INFLAMMATORY_RESPONSE_LIVE_UP	228	-2.508	0	0	0
JAATINEN_HEMATOPOIETIC_STEM_CELL_DN	159	-2.488	0	0	0
VILIMAS_NOTCH1_TARGETS_UP	38	-2.481	0	0	0
BILD_HRAS_ONCOGENIC_SIGNATURE	153	-2.477	0	0	0
CHEN_HOXA5_TARGETS_9HR_UP	172	-2.472	0	0	0
ALTEMEIER_RESPONSE_TO_LPS_WITH_MECHANICAL_VENTILATION	88	-2.471	0	0	0
GARGALOVIC_RESPONSE_TO_OXIDIZED_PHOSPHOLIPIDS_BLUE_UP	95	-2.469	0	0	0
WINZEN_DEGRADED_VIA_KHSRP	49	-2.453	0	0	0
CROONQUIST_STROMAL_STIMULATION_UP	34	-2.445	0	0	0
AMIT_EGF_RESPONSE_60_MCF10A	29	-2.439	0	6E-05	0.001
NOJIMA_SFRP2_TARGETS_UP	21	-2.437	0	5E-05	0.001
BASSO_CD40_SIGNALING_UP	85	-2.436	0	5E-05	0.001
OSWALD_HEMATOPOIETIC_STEM_CELL_IN_COLLAGEN_GEL_UP	165	-2.429	0	5E-05	0.001
GESERICK_TERT_TARGETS_DN	16	-2.424	0	5E-05	0.001
TIAN_TNF_SIGNALING_NOT_VIA_NFKB	18	-2.410	0	5E-05	0.001
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES	22	-2.401	0	5E-05	0.001
SCHOEN_NFKB_SIGNALING	17	-2.358	0	9E-05	0.002
PID_AP1_PATHWAY	40	-2.349	0	9E-05	0.002
GHANDHI_DIRECT_IRRADIATION_UP	52	-2.334	0	9E-05	0.002
SMIRNOV_RESPONSE_TO_IR_2HR_UP	36	-2.328	0	8E-05	0.002
VERHAAK_AML_WITH_NPM1_MUTATED_UP	137	-2.321	0	8E-05	0.002
RHEIN_ALL_GLUCOCORTICOID_THERAPY_UP	55	-2.319	0	1E-04	0.003
LEE_EARLY_T_LYMPHOCYTE_DN	41	-2.312	0	1E-04	0.003
BROCKE_APOPTOSIS_REVERSED_BY_IL6	113	-2.307	0	1E-04	0.003
ZHOU_INFLAMMATORY_RESPONSE_FIMA_UP	217	-2.294	0	2E-04	0.006
SANA_TNF_SIGNALING_UP	52	-2.294	0	2E-04	0.006
CHEN_LVAD_SUPPORT_OF_FAILING_HEART_UP	69	-2.287	0	2E-04	0.007
DAZARD_UV_RESPONSE_CLUSTER_G2	20	-2.286	0	2E-04	0.007
BLM & ATM/ATR checkpoint subgroup					
JAATINEN_HEMATOPOIETIC_STEM_CELL_DN	159	-3.756	0	0	0
FARMER_BREAST_CANCER_CLUSTER_1	33	-3.166	0	0	0
WALLACE_PROSTATE_CANCER_RACE_UP	193	-3.078	0	0	0
BROWNE_INTERFERON_RESPONSIVE_GENES	54	-3.066	0	0	0
HECKER_IFNB1_TARGETS	63	-3.035	0	0	0
VALK_AML_CLUSTER_7	20	-3.008	0	0	0

REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_NON_LYMPHOID_CELL	39	-2.971	0	0	0
MCLACHLAN_DENTAL_CARIES_UP	173	-2.963	0	0	0
WIELAND_UP_BY_HBV_INFECTION	88	-2.937	0	0	0
MOSERLE_IFNA_RESPONSE	23	-2.842	0	0	0
KEGG_GRAFT_VERSUS_HOST_DISEASE	27	-2.829	0	0	0
VERHAAK_AML_WITH_NPM1_MUTATED_UP	137	-2.824	0	0	0
ICHIBA_GRAFT_VERSUS_HOST_DISEASE_D7_UP	76	-2.823	0	0	0
BOSCO_TH1_CYTOTOXIC_MODULE	58	-2.784	0	0	0
GAURNIER_PSMD4_TARGETS	39	-2.779	0	0	0
BOWIE_RESPONSE_TO_TAMOXIFEN	17	-2.745	0	0	0
VILIMAS_NOTCH1_TARGETS_UP	38	-2.719	0	0	0
HAHTOLA_SEZARY_SYNDROM_DN	27	-2.711	0	0	0
BIOCARTA_CTLA4_PATHWAY	16	-2.700	0	0	0
VALK_AML_CLUSTER_8	18	-2.689	0	0	0
POOLA_INVASIVE_BREAST_CANCER_UP	189	-2.677	0	0	0
LINDSTEDT_DENDRITIC_CELL_MATURATION_A	40	-2.651	0	0	0
REACTOME_GENERATION_OF_SECOND_MESSENGER_MOLECULES	22	-2.640	0	0	0
REACTOME_INTERFERON_GAMMA_SIGNALING	46	-2.634	0	0	0
KEGG_AUTOIMMUNE_THYROID_DISEASE	24	-2.611	0	0	0
GRAHAM_CML_QUIESCENT_VS_NORMAL_QUIESCENT_UP	73	-2.595	0	0	0
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_8D_UP	105	-2.584	0	0	0
BENNETT_SYSTEMIC_LUPUS_ERYTHEMATOSUS	27	-2.584	0	0	0
LIANG_SILENCED_BY_METHYLATION_2	30	-2.582	0	0	0
MAHADEVAN_RESPONSE_TO_MP470_UP	15	-2.576	0	0	0
BASSO_CD40_SIGNALING_UP	85	-2.555	0	0	0
ALTEMEIER_RESPONSE_TO_LPS_WITH_MECHANICAL_VENTILATION	88	-2.534	0	7E-05	0.001
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_10D_UP	121	-2.524	0	7E-05	0.001
QI_PLASMACYTOMA_UP	183	-2.522	0	7E-05	0.001
REACTOME_TCR_SIGNALING	43	-2.504	0	6E-05	0.001
ZHANG_INTERFERON_RESPONSE	21	-2.497	0	6E-05	0.001
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_16D_UP	116	-2.487	0	6E-05	0.001
REICHERT_MITOSIS_LIN9_TARGETS	23	-2.484	0	6E-05	0.001
FLECHNER_BIOPSY_KIDNEY_TRANSPLANT_REJECTED_VS_OK_UP	80	-2.481	0	6E-05	0.001
PID_CD8_TCR_PATHWAY	45	-2.461	0	6E-05	0.001
KEGG_ALLOGRAFT_REJECTION	24	-2.449	0	6E-05	0.001
WUNDER_INFLAMMATORY_RESPONSE_AND_CHOLESTEROL_UP	36	-2.444	0	5E-05	0.001
TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_HSC_DN	148	-2.440	0	1E-04	0.002
SMID_BREAST_CANCER_NORMAL_LIKE_UP	282	-2.423	0	2E-04	0.003
VALK_AML_CLUSTER_5	19	-2.419	0	1E-04	0.003
SMID_BREAST_CANCER_LUMINAL_B_DN	280	-2.404	0	1E-04	0.003
UROSEVIC_RESPONSE_TO_IMIQUIMOD	19	-2.394	0	1E-04	0.003
ROSS_AML_OF_FAB_M7_TYPE	57	-2.350	0	3E-04	0.006
PID_IL12_2PATHWAY	49	-2.339	0	3E-04	0.006
BOWIE_RESPONSE_TO_EXTRACELLULAR_MATRIX	16	-2.337	0	3E-04	0.006
KEGG HRR subgroup					
JAATINEN_HEMATOPOIETIC_STEM_CELL_DN	159	-4.038	0	0	0
MCLACHLAN_DENTAL_CARIES_UP	173	-3.608	0	0	0
ALTEMEIER_RESPONSE_TO_LPS_WITH_MECHANICAL_VENTILATION	88	-3.399	0	0	0
WIELAND_UP_BY_HBV_INFECTION	88	-3.306	0	0	0
WALLACE_PROSTATE_CANCER_RACE_UP	193	-3.255	0	0	0
HECKER_IFNB1_TARGETS	63	-3.066	0	0	0
GAURNIER_PSMD4_TARGETS	39	-3.056	0	0	0
VERHAAK_AML_WITH_NPM1_MUTATED_UP	137	-3.012	0	0	0
RUTELLA_RESPONSE_TO_CSF2RB_AND_IL4_DN	262	-2.988	0	0	0

FLECHNER_BIOPSY_KIDNEY_TRANSPLANT_REJECTED_VS_OK_UP	80	-2.973	0	0	0
SMIRNOV_CIRCULATING_ENDOTHELIOCYTES_IN_CANCER_UP	105	-2.933	0	0	0
VALK_AML_CLUSTER_7	20	-2.918	0	0	0
GAL_LEUKEMIC_STEM_CELL_DN	165	-2.908	0	0	0
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_8D_DN	111	-2.903	0	0	0
HESS_TARGETS_OF_HOXA9_AND_MEIS1_DN	54	-2.896	0	0	0
VALK_AML_CLUSTER_5	19	-2.875	0	0	0
KEGG_GRAFT_VERSUS_HOST_DISEASE	27	-2.873	0	0	0
FARMER_BREAST_CANCER_CLUSTER_1	33	-2.872	0	0	0
HALMOS_CEBPA_TARGETS_UP	30	-2.866	0	0	0
IVANOVA_HEMATOPOIESIS_MATURE_CELL	211	-2.843	0	0	0
PARK_APL_PATHOGENESIS_DN	41	-2.841	0	0	0
ZHANG_RESPONSE_TO_IKK_INHIBITOR_AND_TNF_UP	143	-2.813	0	0	0
HAHTOLA_MYCOSIS_FUNGOIDES_CD4_UP	53	-2.810	0	0	0
SEKI_INFLAMMATORY_RESPONSE_LPS_UP	50	-2.809	0	0	0
BASSO_CD40_SIGNALING_UP	85	-2.809	0	0	0
GALINDO_IMMUNE_RESPONSE_TO_ENTEROTOXIN	71	-2.799	0	0	0
BROWNE_INTERFERON_RESPONSIVE_GENES	54	-2.793	0	0	0
POOLA_INVASIVE_BREAST_CANCER_UP	189	-2.788	0	0	0
MARKEY_RB1_ACUTE_LOF_DN	169	-2.784	0	0	0
RHEIN_ALL_GLUCOCORTICOID_THERAPY_UP	55	-2.760	0	0	0
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_10D_DN	62	-2.760	0	0	0
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	182	-2.741	0	0	0
LENAOUR_DENDRITIC_CELL_MATURATION_DN	105	-2.723	0	0	0
SANA_TNF_SIGNALING_UP	52	-2.722	0	0	0
GRAHAM_CML_QUIESCENT_VS_NORMAL_DIVIDING_UP	42	-2.720	0	0	0
SEITZ_NEOPLASTIC_TRANSFORMATION_BY_8P_DELETION_UP	43	-2.718	0	0	0
BROWN_MYELOID_CELL_DEVELOPMENT_UP	109	-2.700	0	0	0
TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_HSC_DN	148	-2.690	0	0	0
KEGG_AUTOIMMUNE_THYROID_DISEASE	24	-2.681	0	0	0
LIAN_LIPA_TARGETS_6M	39	-2.678	0	0	0
LINDSTEDT_DENDRITIC_CELL_MATURATION_A	40	-2.671	0	0	0
ICHIBA_GRAFT_VERSUS_HOST_DISEASE_D7_UP	76	-2.662	0	0	0
LIAN_LIPA_TARGETS_3M	34	-2.660	0	0	0
KEGG_ALLOGRAFT_REJECTION	24	-2.657	0	0	0
HELLER_SILENCED_BY_METHYLATION_UP	178	-2.652	0	0	0
PICCALUGA_ANGIOIMMUNOBLASTIC_LYMPHOMA_DN	110	-2.636	0	0	0
MOSERLE_IFNA_RESPONSE	23	-2.635	0	0	0
NABA_MATRISOME_ASSOCIATED	211	-2.634	0	0	0
FULCHER_INFLAMMATORY_RESPONSE_LECTIN_VS_LPS_DN	338	-2.633	0	0	0
REACTOME_INTERFERON_GAMMA_SIGNALING	46	-2.632	0	0	0

Appendix Y. Rare FA/BRCA-HRR network variants unique to the control cohort.

Chr	Start	End	Gene	Consequence	Ref Seq	*Exon	Nucleotide change	Amino acid change	Controls with variant	No. of reads (Ref, Alt)	[1]D-C mutation	CADD
11	108115601	108115601	ATM	Missense	NM_000051	7	c.G749A	p.R250Q	AOGC-15-0030, S09-F44-P01	8,1; 8,5	-	13.35
11	108141988	108141988	ATM	Missense	NM_000051	20	c.T2932C	p.S978P	S09-F12-P01	13,16	-	26.5
11	108155038	108155038	ATM	Missense	NM_000051	26	c.G3831C	p.E1277D	AOGC-14-3234	13,11	-	16.5
11	108160516	108160516	ATM	Missense	NM_000051	29	c.A4424G	p.Y1475C	S02-F07-P01, S09-F45-P01	3,4; 1,1	-	16.92
11	108170506	108170506	ATM	Missense	NM_000051	34	c.A5071C	p.S1691R	S08-F09-P01	7,3	-	13.71
11	108183194	108183194	ATM	Missense	NM_000051	40	c.A5975C	p.K1992T	AOGC-15-0098	11,8	-	18.49
11	108186598	108186598	ATM	Missense	NM_000051	41	c.T6055C	p.Y2019H	AOGC-14-3376	39,31	-	21.7
11	108196797	108196797	ATM	Missense	NM_000051	47	c.G6820A	p.A2274T	AOGC-14-4788	11,11	-	22.2
11	108201023	108201023	ATM	Missense	NM_000051	50	c.T7390C	p.C2464R	S09-F43-P01	19,11	-	20.4
11	108201083	108201083	ATM	Missense	NM_000051	50	c.G7450A	p.V2484I	S02-F13-P01	18,8	-	16.82
3	142188205	142188205	ATR	Missense	NM_001184	38	c.A6526G	p.M2176V	AOGC-14-4224	10,11	-	21.4
3	142261533	142261533	ATR	Missense	NM_001184	17	c.A3424G	p.S1142G	AOGC-14-4329	5,5	-	22.7
3	142268421	142268421	ATR	Missense	NM_001184	15	c.A3071G	p.N1024S	AOGC-14-2976	16,12	-	14.62
3	142272098	142272098	ATR	Missense	NM_001184	13	c.T2776C	p.F926L	AOGC-14-3553, S12-F06-P01	12,9; 3,2	-	25.3
3	142274770	142274770	ATR	Missense	NM_001184	10	c.A2290G	p.K764E	S07-F23-P01, S09-F37-P01	4,8; 9,6	-	22.7
3	142277468	142277468	ATR	Missense	NM_001184	8	c.A1883G	p.Y628C	AOGC-15-0028	5,4	-	22.8
3	48491440	48491440	ATRIP	Splicing	NM_130384	2	c.248-3C>T	-	AOGC-14-3509	11,8	-	17.41
3	48491568	48491568	ATRIP	Missense	NM_032166	2	c.A373C	p.K125Q	AOGC-14-4600	9,13	-	19.37
3	48493268	48493268	ATRIP	Missense	NM_032166	3	c.A515G	p.Q172R	S11-F07-P01	14,18	-	17.19
3	48495799	48495799	ATRIP	Missense	NM_032166	4	c.C652G	p.P218A	AOGC-14-3160	23,17	-	25.2
3	48498706	48498706	ATRIP	Missense	NM_032166	5	c.C719T	p.P240L	AOGC-14-1504	10,13	-	26.5
3	48501625	48501625	ATRIP	Missense	NM_032166	8	c.G1172A	p.R391Q	S09-F09-P01	12,10	-	19.41
3	48506404	48506404	ATRIP	Missense	NM_032166	11	c.G2149T	p.V717F	AOGC-14-4663	19,12	-	22.2
3	48506476	48506476	ATRIP	Missense	NM_032166	11	c.T2221G	p.C741G	AOGC-14-2385, AOGC-14-2747	12,13; 6,1	-	22.2

2	215595196	215595196	BARD1	Missense	NM_001282549	4	c.A401G	p.Q134R	AOGC-14-1048	8,14	-	14.89
2	215609859	215609859	BARD1	Missense	NM_001282548	4	c.A425T	p.D142V	AOGC-14-2342	11,13	-	14.31
2	215610538	215610538	BARD1	Missense	NM_001282548	3	c.T308C	p.I103T	S09-F46-P01	13,14	-	25.5
2	215632365	215632365	BARD1	Missense	NM_000465	6	c.A1409G	p.N470S	AOGC-14-0782	21,14	-	10.31
2	215645985	215645985	BARD1	Missense	NM_000465	4	c.A613C	p.K205Q	AOGC-14-4329	14,9	-	22.5
15	91298049	91298049	BLM	Missense	NM_000057	5	c.A968G	p.K323R	AOGC-14-4249, S09-F31-P03	22,14; 13,4	-	19.34
15	91304466	91304466	BLM	Missense	NM_001287248	7	c.G738C	p.E246D	AOGC-14-2291	8,10	-	18.77
15	91312417	91312417	BLM	Missense	NM_001287248	11	c.C1237A	p.L413I	S01-F02-P01	13,4	-	27.7
15	91326134	91326134	BLM	Missense	NM_001287248	13	c.G1513C	p.E505Q	AOGC-14-2560	13,5	-	22.6
17	41215926	41215926	BRCA1	Missense	NM_007298	16	c.G1805C	p.G602A	S15-F03-P01	3,3	-	27.8
17	41244252	41244252	BRCA1	Missense	NM_007294	10	c.C3296T	p.P1099L	S09-F47-P01	21,21	-	22.5
17	41245120	41245120	BRCA1	Missense	NM_007294	10	c.A2428T	p.N810Y	AOGC-14-3610	21,18	-	10.5
17	41245465	41245465	BRCA1	Missense	NM_007294	10	c.G2083T	p.D695Y	S02-F06-P01	15,21	-	24
17	41246062	41246062	BRCA1	Missense	NM_007294	10	c.C1486T	p.R496C	S08-F02-P01	23,24	DM	14.2
17	41256266	41256266	BRCA1	Missense	NM_007298	5	c.A314G	p.Y105C	AOGC-14-4129	8,11	-	24.7
13	32906766	32906766	BRCA2	Missense	NM_000059	10	c.C1151T	p.S384F	S09-F08-P01	15,16	-	23
13	32906973	32906973	BRCA2	Missense	NM_000059	10	c.C1358A	p.P453Q	AOGC-14-1048	24,5	-	22.5
13	32907000	32907000	BRCA2	Missense	NM_000059	10	c.A1385G	p.E462G	S03-F01-P01	15,20	-	16.12
13	32907075	32907075	BRCA2	Missense	NM_000059	10	c.C1460A	p.A487E	S01-F21-P04	18,15	-	14.24
13	32907129	32907129	BRCA2	Missense	NM_000059	10	c.T1514C	p.I505T	AOGC-15-0075	11,10	-	12.53
13	32907401	32907401	BRCA2	Missense	NM_000059	10	c.G1786C	p.D596H	AOGC-14-1081, S09-F23-P01	3,7; 8,5	-	24.4
13	32914707	32914707	BRCA2	Missense	NM_000059	11	c.C6215G	p.S2072C	AOGC-14-2259	14,11	DM	28.4
13	32931904	32931904	BRCA2	Missense	NM_000059	16	c.A7643G	p.H2548R	S13-F07-P01	12,17	-	24.4
13	32937333	32937333	BRCA2	Missense	NM_000059	18	c.A7994G	p.D2665G	AOGC-14-1708	9,11	-	32
13	32945172	32945172	BRCA2	Missense	NM_000059	20	c.A8567C	p.E2856A	AOGC-14-2624, AOGC-14-2809	8,5; 16,9	-	26.6
13	32953549	32953549	BRCA2	Missense	NM_000059	22	c.G8850T	p.K2950N	S14-F22-P01	6,11	-	23.1
17	59857686	59857686	BRIP1	Missense	NM_032043	13	c.C1871T	p.S624L	AOGC-14-1569	31,15	-	34

17	59876546	59876546	BRIP1	Missense	NM_032043	9	c.C1255T	p.R419W	AOGC-15-0028	12,10	-	32
17	59885856	59885856	BRIP1	Missense	NM_032043	7	c.A890G	p.K297R	AOGC-14-1130	10,5	-	18.95
17	59926603	59926603	BRIP1	Missense	NM_032043	5	c.A394T	p.T132S	S02-F05-P01	7,9	-	19
17	79511073	79511073	C17orf70	Missense	NM_025161	7	c.C2384G	p.A795G	AOGC-14-2111	11,8	-	23.3
17	79516266	79516266	C17orf70	Missense	NM_025161	4	c.G1369A	p.E457K	AOGC-14-1509	32,21	-	24.9
17	79517700	79517700	C17orf70	Missense	NM_025161	3	c.A820G	p.I274V	AOGC-15-0026	21,16	-	14.17
19	33467575	33467575	C19orf40	Missense	-	-	C>T	-	AOGC-14-1643	4,5	-	15.55
11	125495891	125495891	CHEK1	Nonsense	NM_001114122		c.-773G>A	-	S09-F87-P01	2,2	-	15.5
22	29091178	29091178	CHEK2	Missense	NM_007194	12	c.G1312T	p.D438Y	AOGC-14-2342	18,6	-	34
22	29091774	29091774	CHEK2	Missense	NM_007194	11	c.G1183C	p.V395L	S07-F22-P01	9,17	-	23.6
22	29121326	29121326	CHEK2	Missense	NM_007194	3	c.A349G	p.R117G	AOGC-14-2821	32,35	-	26.2
22	29121360	29121360	CHEK2	Splicing	NM_145862	4	c.320-5T>A	-	AOGC-14-4261, S09-F41-P01, S13-F14-P01	14,12; 17,16; 11,13	-	12.79
22	29091857	29091857	CHEK2	FS del	NM_007194	11	c.1100delC	p.T367fs	S09-F14-P01, S09-F77-P01, S09-F86-P02, S14-F17-P01	5,3; 7,7; 15,6; 6,8	-	37
17	48452928	48452928	EME1	Missense	NM_001166131	2	c.A359C	p.K120T	AOGC-14-1821	29,13	-	10.71
17	48453487	48453487	EME1	Missense	NM_001166131	3	c.G836A	p.R279H	AOGC-14-2111, S09-F48-P01	10,16; 10,11	-	19.09
17	48457749	48457749	EME1	Missense	NM_001166131	8	c.G1462A	p.G488S	AOGC-15-0070	10,9	-	19.68
17	48457756	48457756	EME1	Missense	NM_001166131	8	c.G1469A	p.G490E	S09-F16-P01	14,10	-	19.06
16	1825067	1825067	EME2	Missense	NM_001257370	4	c.C503T	p.P168L	AOGC-14-2342	7,4	-	15.6
16	1825645	1825645	EME2	Missense	NM_001257370	6	c.G739T	p.V247L	AOGC-14-5396	18,15	-	12.21
16	14022028	14022028	ERCC4	Missense	NM_005236	4	c.A728G	p.H243R	S14-F09-P01	12,8	-	21.9
16	14029352	14029352	ERCC4	Missense	NM_005236	8	c.C1563G	p.S521R	AOGC-14-2410	32,31	-	16.06
16	14029516	14029516	ERCC4	Missense	NM_005236	8	c.G1727C	p.R576T	AOGC-14-4046, AOGC-14-4314, S04-F31-P01	3,7; 7,5; 7,9	-	21.8
16	14041848	14041848	ERCC4	Missense	NM_005236	11	c.C2395T	p.R799W	S04-F04-P01	15,17	-	35
16	14041959	14041959	ERCC4	Missense	NM_005236	11	c.G2506A	p.E836K	S13-F04-P01	20,19	-	23
15	31197015	31197015	FAN1	Missense	NM_001146094	2	c.T149G	p.M50R	AOGC-14-4249, S02-F07-P01	25,28; 26,35	-	22.3

15	31197602	31197602	FAN1	Missense	NM_001146094	2	c.C736T	p.R246W	AOGC-15-0111	22,18	-	23.3
15	31198050	31198050	FAN1	Missense	NM_001146094	2	c.T1184C	p.L395P	AOGC-14-3463	11,11	-	17.66
15	31200396	31200396	FAN1	Missense	NM_001146094	3	c.A1310G	p.E437G	AOGC-14-2809	16,7	-	17.55
15	31206263	31206263	FAN1	Missense	NM_014967	5	c.C1780G	p.H594D	AOGC-14-2088	18,14	-	22.3
15	31214555	31214555	FAN1	Missense	NM_014967	8	c.C2170T	p.P724S	AOGC-14-4541	7,8	-	21.6
15	31217412	31217412	FAN1	Missense	NM_014967	9	c.G2255A	p.R752H	AOGC-14-4013	18,16	-	21.9
16	89813075	89813075	FANCA	Missense	NM_000135	35	c.C3430T	p.R1144W	S04-F19-P01	20,20	FA	24
16	89836314	89836314	FANCA	Missense	NM_000135	26	c.T2435G	p.L812R	S02-F14-P01	20,13	-	23
16	89839745	89839745	FANCA	Missense	NM_000135	22	c.C1948G	p.L650V	AOGC-14-1504	13,10	-	15.01
16	89842176	89842176	FANCA	Missense	NM_000135	21	c.G1874C	p.C625S	AOGC-15-0003	25,17	-	23
16	89874721	89874721	FANCA	Missense	NM_001018112	6	c.C577G	p.L193V	AOGC-14-0460, AOGC-14-0719, AOGC-14-2410, AOGC-14-2762	23,21; 29,20; 31,23; 19,15	-	14.12
16	89877157	89877157	FANCA	Missense	NM_001018112	5	c.G480A	p.M160I	S09-F55-P01	13,6	-	13.81
16	89807250	89807252	FANCA	Del	NM_000135	38	c.3788_3790del	p.1263_1264del	AOGC-14-1382	16,10	FA	18.82
X	14862803	14862803	FANCB	Missense	NM_001018113	9	c.A1987G	p.T663A	AOGC-14-2259	7,8	-	14.69
X	14863011	14863011	FANCB	Missense	NM_001018113	8	c.T1894G	p.Y632D	AOGC-14-3592	12,14	-	21.7
9	97873764	97873764	FANCC	Missense	NM_001243743	13	c.A1310C	p.Q437P	AOGC-14-2345	4,7	-	13.87
9	97897645	97897645	FANCC	Missense	NM_001243743	8	c.A826G	p.I276V	S09-F58-P01	13,7	-	13.02
9	97897676	97897676	FANCC	Missense	NM_001243743	8	c.G795C	p.E265D	S09-F76-P01	9,12	-	19.56
9	97912259	97912259	FANCC	Missense	NM_001243743	7	c.C632G	p.P211R	AOGC-14-4260, S09-F05-P01	10,22; 14,10	FA	23
3	10080987	10080987	FANCD2	Missense	NM_033084	8	c.A516G	p.I172M	AOGC-15-0096	14,17	-	18.76
3	10089689	10089689	FANCD2	Missense	NM_033084	16	c.T1367G	p.L456R	AOGC-15-0096	29,8	FA	24.1
3	10105516	10105516	FANCD2	Missense	NM_033084	21	c.A1868C	p.Q623P	AOGC-15-0096	19,10	-	23.8
3	10119777	10119777	FANCD2	Missense	NM_033084	30	c.G2872A	p.V958M	AOGC-15-0030	12,10	-	16.89
6	35426216	35426216	FANCE	Missense	NM_021922	5	c.G1112A	p.R371Q	AOGC-14-4684, S09-F12-P01	29,13; 20,9	-	26.9
6	35430683	35430683	FANCE	Missense	NM_021922	9	c.C1501G	p.Q501E	S09-F82-P01	15,5	-	22.4
11	22646710	22646710	FANCF	Missense	NM_022725	1	c.G647C	p.R216P	S09-F75-P01	24,11	-	10.05
9	35074173	35074173	FANCG	Missense	NM_004629	14	c.C1801T	p.R601C	S02-F03-P01	20,14	-	15.12

9	35074978	35074978	FANCG	Missense	NM_004629	12	c.G1582A	p.G528S	AOGC-14-3359	27,21	-	22.4
9	35077292	35077292	FANCG	Missense	NM_004629	5	c.T615G	p.D205E	AOGC-14-1640	23,25	-	18.25
15	89807200	89807200	FANCI	Missense	NM_018193	8	c.G612C	p.K204N	AOGC-15-0060	13,5	-	19.73
15	89811698	89811698	FANCI	Missense	NM_018193	10	c.T824C	p.I275T	AOGC-14-3408	30,21	-	23
15	89820093	89820093	FANCI	Missense	NM_018193	13	c.G1264A	p.G422R	AOGC-14-1421	16,14	FA	34
15	89828438	89828438	FANCI	Missense	NM_018193	18	c.A1810G	p.M604V	S02-F07-P01	22,25	-	22.3
15	89848640	89848640	FANCI	Missense	NM_018193	29	c.T3075G	p.C1025W	AOGC-14-4541	15,13	-	25.2
15	89858561	89858561	FANCI	Missense	NM_018193	36	c.A3685G	p.I1229V	AOGC-14-3503	20,11	-	22.5
2	58425760	58425760	FANCL	Missense	NM_018062	7	c.C509G	p.P170R	AOGC-14-2959	7,4	-	26
14	45609863	45609863	FANCM	Missense	NM_020937	3	c.A710G	p.N237S	S09-F08-P01	9,19	-	14.84
14	45623211	45623211	FANCM	Missense	NM_020937	6	c.G1139A	p.R380K	S11-F07-P01	9,1	-	24.5
14	45623953	45623953	FANCM	Missense	NM_020937	7	c.T1237C	p.Y413H	AOGC-15-0011, S07-F21-P01	10,14; 12,6	-	23.5
14	45633616	45633616	FANCM	Missense	NM_020937	10	c.G1636A	p.G546S	AOGC-14-3848	12,14	-	32
14	45642364	45642364	FANCM	Missense	NM_020937	13	c.G2267A	p.R756H	S07-F10-P01	21,15	-	15.66
14	45642408	45642408	FANCM	Missense	NM_020937	13	c.G2311A	p.E771K	AOGC-15-0053	12,13	-	34
14	45644953	45644953	FANCM	Missense	NM_020937	14	c.C2996T	p.P999L	S15-F25-P01	17,11	-	28.8
14	45645949	45645949	FANCM	Missense	NM_020937	14	c.C3992T	p.P1331L	AOGC-14-4782	22,15	-	21.9
3	37035075	37035075	MLH1	Missense	NM_001258271	1	c.G37A	p.E13K	AOGC-14-3553	21,14	-	23
3	37059000	37059000	MLH1	Missense	NM_001258274	11	c.G71A	p.R24H	AOGC-14-1569	25,20	-	22.5
3	37059014	37059014	MLH1	Missense	NM_001258274	11	c.A85G	p.T29A	S09-F58-P01	15,20	-	15.37
3	37061929	37061929	MLH1	Missense	NM_001258274	12	c.A290G	p.N97S	S07-F05-P01	4,9	-	22.2
3	37067255	37067255	MLH1	Missense	NM_001258274	13	c.G443A	p.R148Q	AOGC-14-2928	22,27	-	20.6
3	37089131	37089131	MLH1	Missense	NM_001258274	17	c.A1130G	p.K377R	S09-F43-P01	15,13	-	21.5
3	37089154	37089154	MLH1	Missense	NM_001258274	17	c.T1153C	p.F385L	AOGC-14-3965	17,18	-	22.3
3	37090471	37090471	MLH1	Missense	NM_001258274	19	c.A1343G	p.Q448R	AOGC-14-2063	12,14	-	22.2
3	37092025	37092025	MLH1	Missense	NM_001258274	20	c.C1429T	p.H477Y	S07-F13-P01	7,11	-	22.5
14	75498858	75498858	MLH3	Nonsense	NM_014381	7	c.A3668T	p.Q1223L	AOGC-15-0003	15,10	-	23.4

14	75498886	75498886	MLH3	Splicing	NM_014381	8	c.3644-4A>G	-	AOGC-14-2063	8,8	-	12.7
14	75506636	75506636	MLH3	Missense	NM_014381	5	c.A3548G	p.K1183R	AOGC-08-0067, AOGC-15-0028	7,6; 7,1	-	28.1
14	75513222	75513222	MLH3	Missense	NM_014381	2	c.G3137A	p.R1046Q	AOGC-14-1821	16,10	-	16.54
14	75515329	75515329	MLH3	Missense	NM_014381	2	c.G1030A	p.V344M	AOGC-14-1095	16,12	-	19.14
14	75515646	75515646	MLH3	Missense	NM_014381	2	c.A713C	p.Y238S	AOGC-14-1118	5,6	-	19.85
14	75513205	75513205	MLH3	FS del	NM_014381	2	c.3154delC	p.L1052fs	AOGC-14-3408	28,23	-	22.1
11	94180441	94180441	MRE11A	Missense	NM_005591	15	c.G1727A	p.R576Q	AOGC-14-2348	47,52	-	17.49
11	94192594	94192594	MRE11A	Missense	NM_005591	13	c.G1480A	p.E494K	AOGC-14-0681	13,10	-	25.7
11	94192608	94192608	MRE11A	Missense	NM_005591	13	c.A1466G	p.H489R	AOGC-14-5041	12,9	-	17.15
11	94204878	94204878	MRE11A	Missense	NM_005591	8	c.A707C	p.D236A	S02-F24-P01	5,7	-	22.3
11	65631344	65631344	MUS81	Missense	NM_025128	10	c.T1031G	p.I344S	AOGC-14-0469	24,20	-	22
8	90967510	90967510	NBN	Missense	NM_002485	11	c.1397+1->ACA)	-	S08-F03-P01	6,8	-	24.7
8	90967766	90967766	NBN	FS del	NM_002485	10	c.1142delC	p.P381fs	AOGC-14-4093	7,9	-	35
16	23632683	23632683	PALB2	Nonsense	NM_024675	10	c.G3113A	p.W1038X	AOGC-14-1241	15,9	P	20.3
16	23635370	23635370	PALB2	Missense	NM_024675	8	c.G2794A	p.V932M	AOGC-14-2005, AOGC-14-5294, S14-F17-P01	7,6; 0,29; 9,6	-	25.8
16	23641346	23641346	PALB2	Missense	NM_024675	5	c.C2129T	p.T710M	S07-F21-P01	26,20	-	23.4
16	23649280	23649280	PALB2	Splicing	-	-	G>T	-	S12-F06-P01	11,10	-	12.13
16	23649405	23649405	PALB2	Missense	NM_024675	2	c.C94G	p.L32V	AOGC-08-0067	8,10	-	16.88
16	23649446	23649446	PALB2	Missense	NM_024675	2	c.A53G	p.K18R	S15-F03-P01	7,6	-	21.8
5	131923673	131923673	RAD50	Missense	NM_005732	7	c.G943T	p.V315L	AOGC-14-0795, S09-F02-P01, S13-F02-P01	12,12; 15,11; 17,10	-	24.4
5	131923740	131923740	RAD50	Missense	NM_005732	7	c.G1010A	p.R337K	AOGC-14-2345	11,11	-	22.3
5	131925413	131925413	RAD50	Missense	NM_005732	9	c.A1336G	p.K446E	S04-F19-P01	6,1	-	28.3
5	131939072	131939072	RAD50	Missense	NM_005732	14	c.G2288A	p.R763H	AOGC-15-0006	15,17	-	23.4
5	131939128	131939128	RAD50	Missense	NM_005732	14	c.G2344C	p.E782Q	S09-F41-P01	20,14	-	27.5
5	131953890	131953890	RAD50	Missense	NM_005732	21	c.G3293A	p.R1098Q	AOGC-14-1520	13,3	-	22.3
5	131953924	131953924	RAD50	Missense	NM_005732	21	c.T3327G	p.I1109M	S11-F05-P01	9,12	-	24.4

5	131972883	131972883	RAD50	Missense	NM_005732	22	c.C3466T	p.R1156C	AOGC-14-3994	7,8	-	17.96
5	131977906	131977906	RAD50	Missense	NM_005732	25	c.G3789C	p.Q1263H	S02-F23-P01	8,9	-	26.7
5	131977952	131977952	RAD50	Missense	NM_005732	25	c.C3835T	p.R1279C	AOGC-02-0045	16,15	-	35
5	131977953	131977953	RAD50	Missense	NM_005732	25	c.G3836A	p.R1279H	AOGC-14-2938	10,14	-	27
5	131978046	131978046	RAD50	Missense	NM_005732	25	c.A3929G	p.N1310S	AOGC-14-3461	15,14	-	11.26
14	68331840	68331840	RAD51B	Missense	NM_133510	5	c.G436A	p.A146T	S09-F59-P01	4,8	-	24.9
14	68352672	68352672	RAD51B	Missense	NM_133510	6	c.A539G	p.Y180C	AOGC-14-0782, AOGC-14-3566, S06-F03-P01, S09-F44-P01	8,6; 7,8; 10,7; 3,2	-	22.5
14	68353925	68353925	RAD51B	Splicing	NM_133509	7	c.756+4A>C	-	S02-F01-P01	7,5	-	15.28
17	56772429	56772429	RAD51C	Missense	NM_002876	2	c.C283T	p.H95Y	AOGC-14-4260	19,15	-	22.2
17	33428058	33428058	RAD51D	Splicing	NM_001142571	11	c.964-3C>T	-	AOGC-14-4782, AOGC-14-5036, S02-F01-P01	12,10; 10,11; 9,4	-	11.11
17	33430511	33430511	RAD51D	Missense	NM_001142571	7	c.C689T	p.A230V	AOGC-14-1548	18,18	-	29.7
11	67161081	67161081	RAD9A	Missense	NM_001243224	1	c.G121A	p.G41R	AOGC-14-4541	22,24	-	21
11	67163780	67163780	RAD9A	Missense	NM_001243224	6	c.G553A	p.V185I	AOGC-14-3952	41,33	-	22.1
11	67165007	67165007	RAD9A	Missense	NM_001243224	8	c.G925A	p.E309K	AOGC-14-5035	28,31	-	22
11	67163284	67163286	RAD9A	Del	NM_001243224	3	c.319_321del	p.107_107del	S07-F10-P01	22,22	-	22.4
18	20548849	20548849	RBBP8	Missense	NM_002894	5	c.G329A	p.R110Q	AOGC-14-4046	12,5	-	24.7
18	20569274	20569274	RBBP8	Missense	NM_002894	9	c.A800G	p.E267G	S09-F25-P01	15,13	-	24.5
18	20572775	20572775	RBBP8	Missense	NM_002894	11	c.G985A	p.V329I	AOGC-03-0132	10,4	-	26.5
18	20573157	20573157	RBBP8	Missense	NM_002894	11	c.A1367G	p.H456R	AOGC-14-0782, AOGC-14-5036, S09-F86-P02	14,15; 11,8; 24,19	-	13.4
16	11444605	11444605	RMI2	Missense	NM_152308	2	c.G402A	p.M134I	AOGC-14-4329	13,13	-	19.97
16	11444630	11444630	RMI2	Missense	NM_152308	2	c.C427T	p.H143Y	S14-F09-P01	11,14	-	22.3
1	28223560	28223560	RPA2	Missense	NM_002946	6	c.G481A	p.E161K	AOGC-14-4129	12,5	-	33
1	28240593	28240593	RPA2	Missense	NM_002946	2	c.C98T	p.S33F	S08-F09-P01	5,3	-	12.53
16	3640271	3640271	SLX4	Missense	NM_032444	12	c.C3368A	p.S1123Y	AOGC-14-3743	30,26	-	25.2
16	3640985	3640985	SLX4	Missense	NM_032444	12	c.C2654T	p.P885L	AOGC-14-3234	42,43	-	22.9

16	3641016	3641016	SLX4	Missense	NM_032444	12	c.G2623A	p.E875K	AOGC-14-3553	18,31	-	22.4
16	3641060	3641060	SLX4	Missense	NM_032444	12	c.C2579T	p.T860I	AOGC-14-4246	18,19	-	27.2
16	3644501	3644501	SLX4	Missense	NM_032444	10	c.G2113A	p.A705T	AOGC-14-3566	12,16	-	26.7
16	3647564	3647564	SLX4	Missense	NM_032444	7	c.C1499T	p.T500M	AOGC-14-3853	19,15	-	24.9
16	3647691	3647691	SLX4	Missense	NM_032444	7	c.A1372G	p.K458E	S09-F32-P01	10,8	-	25.5
16	3656493	3656493	SLX4	Missense	NM_032444	3	c.G742A	p.E248K	S09-F23-P01	2,4	-	22.8
17	18181592	18181592	TOP3A	Missense	NM_004618	18	c.G2224A	p.D742N	AOGC-14-0460, AOGC-14-1548, S04-F26-P01	25,15; 13,5; 11,9	-	23
17	18181601	18181601	TOP3A	Missense	NM_004618	18	c.G2215A	p.G739R	AOGC-14-4677	26,19	-	29.7
17	18198068	18198068	TOP3A	Missense	NM_004618	10	c.G1022A	p.R341K	AOGC-03-0132, S01-F17-P01	15,5; 19,9	-	16.46
17	18211737	18211737	TOP3A	Missense	NM_004618	3	c.A242G	p.N81S	AOGC-14-4454	10,18	-	21.2
12	125397621	125397621	UBC	Missense	NM_021009	2	c.G697A	p.V233M	S02-F13-P01	87,83	-	21.8
12	125397817	125397817	UBC	Splicing	NM_021009	2	c.C501T	p.L167L	AOGC-14-2169	76,72	-	16.53
12	125397982	125397982	UBC	Missense	NM_021009	2	c.T336C	p.I112I	S09-F08-P01	57,77	-	14.73
12	125397997	125397997	UBC	Missense	NM_021009	2	c.G321A	p.Q107Q	AOGC-14-1941	56,8	-	15.82
12	125398015	125398015	UBC	Missense	NM_021009	2	c.C303T	p.N101N	AOGC-14-2169	69,13	-	14.34
12	125398024	125398024	UBC	Missense	NM_021009	2	c.C294T	p.T98T	AOGC-14-1941	45,8	-	22.2
12	125396390	125396393	UBC	FS del	NM_021009	2	c.1925_1928del	p.E642fs	AOGC-14-1270	30,43	-	22.3
1	202304777	202304777	UBE2T	Missense	NM_014176	2	c.G106A	p.A36T	AOGC-14-3002	17,12	-	33
1	62908939	62908939	USP1	Missense	NM_001017416	5	c.A506G	p.K169R	AOGC-14-4616	16,11	-	24.4
1	62910713	62910713	USP1	Missense	NM_001017416	6	c.A862G	p.K288E	AOGC-14-2348	26,24	-	24.6
1	62910957	62910957	USP1	Missense	NM_001017416	6	c.G1106C	p.G369A	S09-F71-P01	41,18	-	16.49
1	62910981	62910981	USP1	Missense	NM_001017416	6	c.A1130T	p.D377V	S04-F18-P01	24,21	-	22.5
1	62916503	62916503	USP1	Missense	NM_001017416	9	c.G2209A	p.V737M	S07-F23-P01	6,10	-	19.53
7	152357788	152357788	XRCC2	Missense	NM_005431	2	c.A119G	p.H40R	AOGC-14-4161	2,3	-	22.3
14	104169623	104169623	XRCC3	Missense	NM_005432	7	c.C448T	p.R150C	AOGC-14-3194	7,7	-	13.56

*Splicing mutations where the exon is not annotated, the genomic coordinate of the variant were listed.

[1]Disease-causing (D-C) mutations reported to cause Fanconi Anaemia are represented as FA, pathogenic mutations reported in HGMD database are represented as DM, and breast cancer mutations are represented as P.