Identification of disease gene variants that can lead to familial myelodysplasia and acute myeloid leukaemia

A thesis submitted for the degree of PhD

Shirleny Romualdo Cardoso

Supervisors: Professor Inderjeet Dokal, and
Professor Tom Vulliamy

Centre for Genomics and Child Health, Blizard Institute
Barts and The London School of Medicine & Dentistry,
Queen Mary University of London

In loving memory of my beloved sister Karla Romualdo Cardoso

I, Shirleny Romualdo Cardoso, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Shirleny Romualdo Cardoso

Abstract

Myelodysplasia (MDS) is characterised by inefficient haematopoiesis with dysplastic features of blood and bone marrow, reduction of mature blood cells and continuous bone marrow failure (BMF). Acute myeloid leukaemia (AML) is characterised by the accumulation of immature myeloid blasts in the bone marrow. MDS and AML are mostly sporadic clonal disorders affecting older patients. Familial occurrence of MDS/AML is rare, and most of these cases occur in the setting of genetic syndromes. However, it has also been reported to be caused by germline heterozygous mutations in genes including *RUNX1*, *CEBPA*, *TERC*, *TERT*, *GATA2*, *SRP72*, *and ANKRD26*.

Our group has collected 115 families that have two or more individuals with BMF with at least one of whom has MDS or AML. The aim of this project was to identify disease causing gene variants that can lead to familial MDS/AML. Identification of predisposing variants to familial MDS/AML is critical for effective management in these families. This will also provide new insights into the biology of MDS/AML in general.

Herein, we have characterised a subset of families with MDS/AML as well as identified candidate disease genes using a range of genetic studies. Specifically, we have: i. Identified new genetic variants in some of the known disease genes such as *RUNX1* and *GATA2*. ii. Our studies have substantiated the discovery of *DDX41* as a disease gene as we have identified several families harbouring novel heterozygous loss of function (LoF) *DDX41* variants. iii. Identified germline heterozygous LoF *RTEL1* variants in a subset of families with myelodysplasia and liver disease. This defines a new disease group in this field, *RTEL1* can now be added to the list of familial MDS/AML disease genes. iv. We have identified nine new candidate disease genes

which are involved in RNA splicing, transcription factor, DNA modification, cell signalling and intracellular transport.

Contents

List of figures

List of tables

List of abbreviations

+8	trisomy 8
µg	microgram
µl	microlitre
µM	micromolar
1000G	1000 Genomes Project
-20q	loss of 20q
5'/3'	5 prime/3 prime
-7q	loss of 7q
AA	aplastic anaemia
ABCA3	ATP-binding cassette subfamily A, member 3
ACD	ACD, shelterin complex subunit and telomerase recruitment factor
ACMG	American College of Medical Genetics and Genomics
AD	autosomal dominant
ALL	T-cell acute lymphoblastic leukaemia
ALT	alternative telomere length
AML	acute myeloid leukaemia
AMP	Molecular Pathology
ANKRD26	ankyrin repeat domain 26
ApE	A plasmid Editor
AR	autosomal recessive
ASXL1	additional sex combs like 1, transcriptional regulator
ATP	adenosine triphosphate
AZA	azacytidine
BAM	binary alignment map
BCOR	BCL6 corepressor
BCORL1	BCL6 corepressor like 1
BLM	Bloom syndrome RecQ like helicase
BMF	bone marrow failure
bp	base pair
BRCA1	BRCA1, DNA repair associated
BRCA2	BRCA2, DNA repair associated
BRIP1	BRCA1 interacting protein C-terminal helicase 1

BSA	bovine serum albumin
BTK	Bruton's tyrosine kinase
Cas-9	CRISPR-associated 9
CADD	Combined Annotation Depletion score
CBF	core binding factor
CBFB	core-binding factor beta subunit
CBL	Cbl proto-oncogene
CCND1	cyclin D1
CD34+	haematopoietic progenitor cell antigen CD34
CD4/CD8	cluster of differentiation 4/8
c-di-AMP	cyclic di-AMP
c-di-GMP	cyclic di-GMP
CDKL1	cyclin-dependent kinase-like 1
CDKN1A	cyclin-dependent kinase inhibitor 1A
cDNA	coding DNA
CEBPA	CCAAT/enhancer binding protein alpha
CEP68	centrosomal protein 68kDa
cGAS	cyclic-GMP-AMP synthetase
CGH-array	Comparative genomic hybridization-array
CHIP	clonal haematopoiesis of indeterminate potential
cm	centimetre
CML	chronic myeloid leukaemia
CMML	chronic myelomonocytic leukaemia
CO_2	carbon dioxide
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTC1	CST telomere replication complex component 1
DAI	DNA-dependent activator of IFN-regulatory factors
DAC	decitabine
DC	dyskeratosis congenita
ddNTPs	dideoxynucleotides triphosphates
DDX1	DEAD-box helicase 1
DDX11	DEAD/H-box helicase 11
DDX17	DEAD-box helicase 17
DDX21	DExD-box helicase 21
DDX3	DEAD-box helicase 3

DDX41	DEAD-box helicase 41 gene
DDX5	DEAD-box helicase 5
DDX58	DExD/H-box helicase 58
DDX60	DExD/H-box helicase 60
DEB	diepoxybutane
DEK	DEK proto-oncogene
del(5q)	loss of 5q
dHPLC	denaturing high-performance liquid chromatography
DHX36	DEAH-box helicase 36
DHX58	DExH-box helicase 58
DHX9	DExH-box helicase 9
DIG	digoxygenin
DKC1	dyskerin
D-loop	displacement-loop structure
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNAJC21	DnaJ heat shock protein family (Hsp40) member C21
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3 alpha
dNTPs	deoxynucleotide triphosphates
DSBs	DNA double-strand breaks
dsDNA	double-stranded DNA
dsRNA	double stranded RNA
DTT	dithiothreitol
E2F	E2F transcription factor 1
EFL1	elongation factor-like GTPase 1
eIF6	eukaryotic initiation factor 6
EPS15	epidermal growth factor receptor pathway substrate 15
ERCC1	ERCC excision repair 1, endonuclease non-catalytic subunit
ERCC4	ERCC excision repair 4, endonuclease catalytic subunit
ESP	Exome Sequencing Project
ETO	eight-21
ETS	E-twenty-six
ETV6	ETS variant 6
EVS	Exome Variant Server

ExAC	Exome Aggregation Consortium
EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit
FA	Fanconi anaemia
FANCA	Fanconi anaemia complementation group A
FANCB	Fanconi anaemia complementation group B
FANCC	Fanconi anaemia complementation group C
FANCD1	BRCA2, DNA repair associated
FANCD2	Fanconi anaemia complementation group D2
FANCE	Fanconi anaemia complementation group E
FANCF	Fanconi anaemia complementation group F
FANCG	Fanconi anaemia complementation group G
FANCI	Fanconi anaemia complementation group I
FANCJ	BRCA1 interacting protein C-terminal helicase 1
FANCL	Fanconi anaemia complementation group L
FANCM	Fanconi anaemia complementation group M
FANCN	partner and localizer of BRCA2
FANCO	RAD51 paralog C
FANCP	SLX4 structure-specific endonuclease subunit
FANCQ	ERCC excision repair 4, endonuclease catalytic subunit
FANCR	RAD51 recombinase
FANCS	BRCA1, DNA repair associated
FANCT	ubiquitin conjugating enzyme E2 T
FANCU	X-ray repair cross complementing 2
FANCV	mitotic arrest deficient 2 like 2
FANCW	ring finger and WD repeat domain 3
FeS	iron-sulphur
FIP	familial interstitial pneumonia
FLI1	Fli-1 proto-oncogene, ETS transcription factor
FLT3	fms related tyrosine kinase 3
FPD	familial platelet disorder
FPD/AML	familial platelet disorder with propensity to myeloid malignancy
G4-DNA	telomeric G-quadruplex
GATA1	GATA binding protein 1
GATA2	GATA binding protein 2
GATA3	GATA binding protein 3

GATA4	GATA binding protein 4
GATA5	GATA binding protein 5
GATA6	GATA binding protein 6
GATK	Genome Analysis Toolkit
GNAS	GNAS complex locus
gVCF	genomic variant call format
HCl	hydrochloric acid
HCV	hepatitis C virus
HH	Hoyeraal Hreidarsson syndrome
HIV	human immunodeficiency virus
HJ	Holliday junction
HR	homologous recombination
IDH1/2	isocitrate dehydrogenase (NADP(+)) 1/2, cytosolic
IFI16	gamma-inducible protein 16
IFIH1	interferon induced with helicase C domain 1
IFN-α	Interferon alpha
IFN-β	Interferon beta
IIP	idiopathic interstitial pneumonia
IKKε	inhibitor of κB kinase ε
indels	Insertion/deletion (s)
IRF3	interferon regulatory factor 3
IRF7	interferon regulatory factor 7
JAK2	Janus kinase 2
JMML	juvenile myelomonocytic leukaemia
KAT3A	CREB binding protein
kb	kilobase
KCl	potassium chloride
kDa	Kilo Dalton
KIT	KIT proto-oncogene receptor tyrosine kinase
KMT2A	lysine methyltransferase 2A
LoF	loss of function
LUC7L2	LUC7 like 2, pre-mRNA splicing factor
M	molar
MAC	Mycobacteria avium complex
MAD2L2	mitotic arrest deficient 2 like 2

MAF	minor allele frequency
MAVS	CARD-domain containing mitochondrial adaptor
MDM2	MDM2 proto-oncogene
MDS	Myelodysplasia syndrome
MDS/AML	myelodysplasia/acute myeloid leukaemia
MECOM	MDS1 and EVI1 complex locus
MES	2-ethanesulfonic acid
$MgCl_2$	magnesium chloride
MGUS	monoclonal gammopathy
ml	mililitre
MLLT3	MLLT3, super elongation complex subunit
mM	milimolar
mm	milimitre
MMqPCR	monochrome multiplex quantitative PCR
MPN	myeloproliferative neoplasms
MUS81	MUS81 structure-specific endonuclease subunit
MYC	MYC proto-oncogene, bHLH transcription factor
MyD88	myeloid differentiation primary response gene 88
MYH11	myosin heavy chain 11
MYM	myeloproliferative and mental retardation
NAF1	nuclear assembly factor 1 ribonucleoprotein
NaOH	sodium hydroxyzine
NCBI36/hg18	Homo sapiens (human) genome assembly
NF1	neurofibromin 1
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
ng	nanogram
NGS	next generation sequencing
NHLBI	National Heart, Lung, and Blood Institute
NHP2	NHP2 ribonucleoprotein
NOP1	fibrillarin
NPH2	inversin
NPM1	nucleophosmin 1
NSD1	nuclear receptor binding SET domain protein 1
NTP	nucleoside triphosphate
NUP214	nucleoporin 214

NUP98	nucleoporin 98
PALB2	partner and localizer of BRCA2
PAMPs	pathogen-associated molecular patterns
PARN	PARN poly(A)-specific ribonuclease
PBS	phosphate buffered saline
PCNA	proliferating cell nuclear antigen interacting protein
PCR	polymerase chain reaction
PHF6	PHD finger protein 6
pLI	probability of LoF intolerance
PML	promyelocytic leukemia
pmol	picomole
POT1	protection of telomeres 1
PPM1D	protein phosphatase, Mg2+/Mn2+ dependent 1D
PRPF8	pre-mRNA processing factor 8
PRRs	pattern recognition receptors
PS-DVB	polystyrene-divinylbenzene
PTPN11	protein tyrosine phosphatase, non-receptor type 11
PVDF	polyvinylidene difluoride
RAD21	RAD21 cohesin complex component
RAD51	RAD51 recombinase
RAEB	refractory anaemia with excess blasts
Rap1	repressor activator protein 1
RARA	retinoic acid receptor alpha
RB	retinoblastoma tumour suppressor
RCA	rolling circle amplification
REV7	mitotic arrest deficient 2 like 2
RFWD3	ring finger and WD repeat domain 3
RLHs	RIG-like helicases
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	pre-ribosome RNA
RTEL1	regulator of telomere elongation helicase 1
RUNX1	runt related transcription factor 1
RUNX1T1	RUNX1 translocation partner 1
RUNX2	runt related transcription factor 2

RUNX3	runt related transcription factor 3
SAMD9	sterile alpha motif domain containing 9
SAMD9L	sterile alpha motif domain containing 9 like
SBDS	SBDS, ribosome maturation factor
SCT	stem-cell transplant
SDS	Shwachman-Diamond syndrome
SDSA	synthesis-dependent strand annealing pathway
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SETD4	SET domain-containing protein 4
SF2	Superfamily 2
SF3B1	splicing factor 3b subunit 1
SF3B2	splicing factor 3b subunit 2
SF3B3	splicing factor 3b subunit 3
SFTPA2	surfactant protein A2
SFTPC	surfactant protein C
SLX1	SLX1 structure-specific endonuclease subunit
SLX4	SLX4 structure-specific endonuclease subunit
SMC1A	structural maintenance of chromosomes 1A
SMC3	structural maintenance of chromosomes 3
SNPs	single nucleotide polymorphisms
SRP72	signal recognition particle 72
SRSF2	serine and arginine rich splicing factor 2
SSC	saline-sodium citrate
ssDNA	single-stranded DNA
STAG1/2	stromal antigen 1/2
STING	stimulator of interferon genes
T/S	telomeric to single-copy gene ratio
TBE	tris-borate-EDTA
TBK1	TANK binding kinase 1
TBS-T	Tween-20
TCFs	ternary complex factors
T-circle	telomeric-circle
TE	tris-EDTA
TERC	telomerase RNA component
TERT	telomerase reverse transcriptase

TET2	tet methylcytosine dioxygenase 2
THC2	thrombocytopenia 2
TINF2	TERF1 interacting nuclear factor 2
TLR	Toll-like receptors
TNRC6C	trinucleotide repeat containing 6C
TP53	tumor protein p53
TRANK1	tetratricopeptide repeat and ankyrin repeat containing 1
TRF1	telomere repeat binding factor 1
TRF2	telomere repeat binding factor 2
TRIF	TIR-domain-adapter-inducing interferon-β
TRIM21	tripartite motif containing 21
TSCA	Truseq Custom Amplicon
TTP1	TINF2-interacting protein 2
TTS	upstream of the transcription start site
U2 snRNP	U2 small nuclear ribonucleoproteins complex
U2AF1	U2 small nuclear RNA auxiliary factor 1
UBE2T	ubiquitin conjugating enzyme E2 T
UTR	untranslated region
V	volt
V/cm	volt/centimetre
VAF	variant allele frequency
vs	versus
VUS	variant of unknown significance
WAS	Wiskott-Aldrich syndrome
WASp	WAS protein
WES	whole exome sequencing
WHO	World Health Organization
WRAP53	WD repeat containing antisense to TP53
WT1	Wilms tumor 1
XL	X-linked
XLN	X-linked neutropenia
XP	xeroderma pigmentosum
XPD	Xeroderma pigmentosum group D
XPF	ERCC excision repair 4, endonuclease catalytic subunit
XRCC2	X-ray repair cross complementing 2

-Y	loss of Y
ZMYM2	Zinc finger MYM-type containing 2

Publications and presentations arising from this thesis

Publications

Tummala H, Walne A, Collopy L, Cardoso S, de la Fuente J, Lawson S, Powell J, Cooper N, Foster A, Mohammed S, Plagnol V, Vulliamy T, Dokal I. Poly(A)-specific ribonuclease deficiency impacts telomere biology and causes dyskeratosis congenita. J Clin Invest. 2015 May;125(5):2151-60.

Collopy LC, Walne AJ, Cardoso S, de la Fuente J, Mohamed M, Toriello H, Tamary H, Ling AJ, Lloyd T, Kassam R, Tummala H, Vulliamy TJ, Dokal I. Triallelic and epigenetic-like inheritance in human disorders of telomerase. Blood. 2015 Jul 9;126(2):176-84.

Tummala H, Walne AJ, Williams M, Bockett N, Collopy L, Cardoso S, Ellison A, Wynn R, Leblanc T, Fitzgibbon J, Kelsell DP, van Heel DA, Payne E, Plagnol V, Dokal I, Vulliamy T. DNAJC21 Mutations Link a Cancer-Prone Bone Marrow Failure Syndrome to Corruption in 60S Ribosome Subunit Maturation. Am J Hum Genet. 2016 Jul 7;99(1):115-24.

Cardoso SR, Ryan G, Walne AJ, Ellison A, Lowe R, Tummala H, Rio-Machin A, Collopy L, Al Seraihi A, Wallis Y, Page P, Akiki S, Fitzgibbon J, Vulliamy T, Dokal I. Germline heterozygous DDX41 variants in a subset of familial myelodysplasia and acute myeloid leukemia. Leukemia. 2016 Oct;30(10):2083-2086.

Walne AJ, Collopy L, Cardoso S, Ellison A, Plagnol V, Albayrak C, Albayrak D, Kilic SS, Patıroglu T, Akar H, Godfrey K, Carter T, Marafie M, Vora A, Sundin M, Vulliamy T, Tummala H, Dokal I. Marked overlap of four genetic syndromes with dyskeratosis congenita confounds clinical diagnosis. Haematologica. 2016 Oct;101(10):1180-1189.

Cardoso SR, Ellison ACM, Walne AJ, Cassiman D, Raghavan M, Kishore B, Ancliff P, Rodríguez-Vigil C, Dobbels B, Rio-Machin A, Al Seraihi AFH, Pontikos N, Tummala H, Vulliamy T, Dokal I. Myelodysplasia and liver disease extend the spectrum of RTEL1 related telomeropathies. Haematologica. 2017 Aug;102(8):e293-e296.

Presentations

Poster presentation at The Blizard Graduate Studies Day (London, UK; April 2015)
Identification of mutations that can lead to familial myelodysplasia and acute myeloid leukaemia

XXXVI World congress - International Society of Haematology hosted by British Society for Haematology (Glasgow, UK; April 2016)
Germline heterozygous DDX41 variants account for a subset of familial myelodysplasia and acute myeloid leukaemia

Oral presentation at Barts and The London School of Medicine and Dentistry Centre for Genomics and Child Health (London, UK; October 2016)
Germline heterozygous DDX41 variants in a subset of familial myelodysplasia and acute myeloid leukemia

Oral presentation at Blizard Institute Post Grads Club (London, UK; February 2017)
Identification and characterisation of variants that can lead to familial myelodysplasia and acute myeloid leukaemia

Oral presentation at Barts and The London School of Medicine and Dentistry Centre for Genomics and Child Health (London, UK; May 2017)
Myelodysplasia and liver disease extend the spectrum of RTEL1 related telomeropathies

Poster presentation at William Harvey Annual Research Day (London, UK; October 2017)
Myelodysplasia and liver disease extend the spectrum of RTEL1 related telomeropathies

Acknowledgments

After an intensive period of four years, today is the day: writing this note of thanks is the finishing touch on my thesis. It has been a period of intense learning for me, not only in the scientific field, but also on a personal level. I would like to reflect on the people who have supported and helped me so much throughout this period.

I would particularly like to thank my supervisors Professors Inderjeet Dokal and Tom Vulliamy for their valuable guidance, motivation and for all of the opportunities I was given to conduct my research. And also, for the hard questions, which stimulated me to widen my research from various perspectives. You definitely provided me with the tools that I needed to choose the right direction and successfully complete my thesis. I thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for funding this work.

I would like to thank my colleagues from Dokal group for their remarkable collaboration. Dr Amanda Walne, Dr Hemanth Tummala, Alicia Ellison and Dr Jasmin Sidhu, I want to thank you for your excellent cooperation and for all of the opportunities I was given to conduct my research. You supported me greatly and were always willing to help me.

I am also grateful to Dr Vicent Plagnol from University College London, who performed the bioinformatics processing of the next generation sequencing data in this thesis and the clinicians who provided the lab with the patient samples for Familial MDS/AML and Dyskeratosis Congenita registries. I also thank the patients and families themselves for their willingness to be included in this research. In addition, I would like to thank María Victoria Niklison-Chirou, Ana Rio-Machain, Ahad Fahad, Jenna Analjar and those also working in Blizard, who have provided advice when needed during these years. I would like to especially thank my lovely friend Dr Chinedu Nwokoro, who helped all this come true. Meeting you was the beginning of my life changing, I will be forever grateful to you.

I am grateful to Julia Moreta Diaz, Jaya Rajamanie, Shirley Dankyi-Larbi for their prompt assistance always when needed. You have made Blizard a much more enjoyable place to work in.

A very special gratitude goes out to my wonderful partner Giovanni Santagiuliana, life is much easier and enjoyable with you beside me. Thank you very much for all your support and patience.

I would also like to thank my mother Tânia Maria and my stepfather Júlio César for their wise counsel and sympathetic ear. I am also grateful to my siblings Roberto, Marcos and Camilla, you are always there for me and I am most proud of you. I will always remember with a warm heart all support I have received from my lovely sister Karla during her short time with us. I am also grateful to my other family members who have supported me along the way.

Finally, there are my friends: Anna Zebrowska, Nisha Liyana, Malihe Moeinian, Alex Wibawa, Minan Al-Ezzi, Karina Salomão, Eduardo Araújo and Michelle Padovan. We were not only able to support each other by deliberating over our problems and findings, but also happily by talking about things other than just our work. I am also extremely grateful to all my good friends from Brazil that despite the geographic distance, they make sure to be constantly present in my life.

This accomplishment would not have been possible without all of you. Thank you.

Chapter 1

Introduction

1.1 General introduction

Myelodysplasia and acute myeloid leukaemia are mostly sporadic haematopoietic malignancies typically affecting older patients, though a small subset has been associated with germline mutations. Familial cases of myelodysplasia/acute myeloid leukaemia are rare, but are extremely relevant for the investigation of the molecular pathogenesis of myelodysplasia and acute myeloid leukaemia in general as many genes associated with these familial myeloid neoplasms are also recurrently mutated in sporadic cases (Liew & Owen, 2011; West *et al*., 2014). Large-scale sequencing of cancer genomes has now been completed for thousands of cancer samples. This initial discovery phase has uncovered many novel genes, pathways, and mutational processes implicated in cancer development (Vogelstein *et al*., 2013). Furthermore, it has been demonstrated that the mutation/deletion status of a set of genes could be used to build a clinically relevant prognostic system as independent variables from clinical parameters. Moving forward, studies are warranted to clarify how to integrate this increased knowledge of gene mutations in our understanding of leukaemogenesis and into clinical practice (Haferlach *et al*., 2014).

1.2 Myelodysplasia syndromes

Myelodysplasia syndromes (MDS) are a heterogeneous group of bone marrow disorders derived from clonal haematopoietic stem cells with increased proliferation or resistance to apoptosis, leading to an over production of abnormal differentiated blood cells. This produces inefficient haematopoiesis with morphological dysplastic features of blood and bone marrow, reduction of blood cells with various degrees of

cytopenia and continuous bone marrow failure (Figure 1.1) (Tefferi & Vardiman, 2009). The incidence rate of sporadic myelodysplasia in 2016 was approximately 3 to 4 cases per 100,000 population per year, with 30 cases per 100,000 population per year in patients >70 years old (Cancer Network - www.cancernetwork.com/cancer-management/mds).

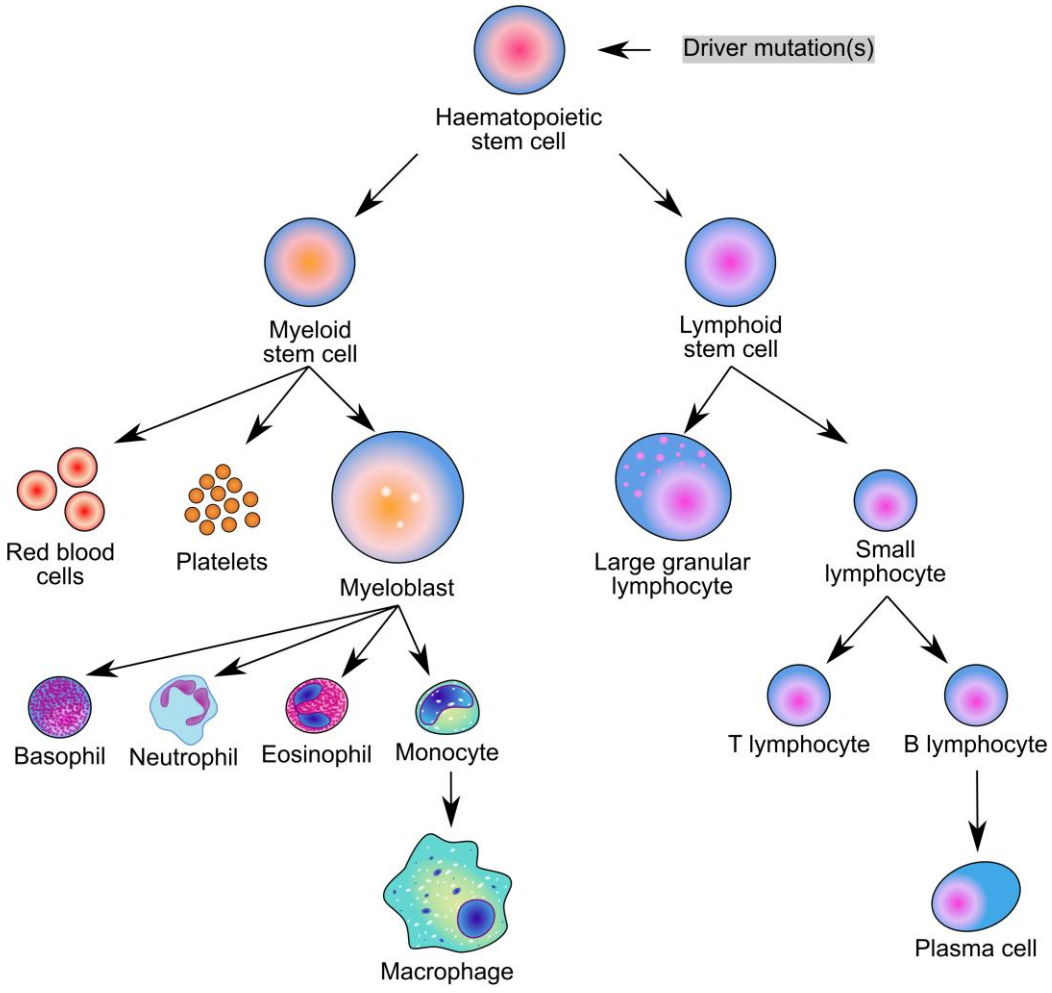


Figure 1.1. Schematic of a normal haematopoietic stem cells differentiation. Driver mutation(s) in the haematopoietic stem cell is necessary for the development of MDS or AML.

There are many MDS subtypes with varying clinical features and pathogenesis depending on which blood cells are affected. Any or all of the blood cells may be affected in MDS and its diagnosis depends mainly on the degree of dysplasia and blast percentages in peripheral blood and bone marrow. MDS should be classified according to the World Health Organization (WHO) criteria, as revised in 2016 (Arber *et al*., 2016). In the revised WHO MDS criteria the terms “refractory anemia” and “refractory cytopenia” were replaced by “myelodysplastic syndrome” followed by the appropriate modifiers: single vs multilineage dysplasia, ring sideroblasts, excess blasts, or the del(5q) cytogenetic abnormality (Table 1.1, adapted from Arber *et al*., 2016).

MDS can evolve into a form of leukaemia, usually acute myeloid leukaemia (AML) in 30% of cases (Mufti *et al*., 2008). The clinical phenotype of patients with MDS are diverse with respect to the number and severity of cytopenias, cellularity and blast count in the bone marrow, rate of progression to AML, overall survival and response to treatment. MDS is an extremely heterogeneous group of disorders, ranging from mild conditions with a near-normal life expectancy to forms approaching AML. Much of this phenotypic heterogeneity is likely due to the variety of genetic alterations that contribute to disease pathogenesis.

Somatically acquired genetic abnormalities, including karyotypic abnormalities, gene alterations and aberrant epigenetic regulation of gene expression lead to the development of MDS (Bejar *et al*., 2011). However, single genetic alterations are unlikely to be the sole disease-causing abnormalities in myeloid neoplasms. Instead, a combination of two or more of these genetic modifications may be needed in cooperation with global changes in epigenetic states and cellular environment (Bejar *et al*., 2011).

Table 1.1. WHO classification of MDS and AML as revised in 2016

WHO myelodysplastic and acute leukaemia classification	
Myelodysplastic syndromes (MDS)	**Acute myeloid leukaemia (AML) and related neoplasms**
MDS with single lineage dysplasia	AML with recurrent genetic abnormalities
MDS with ring sideroblasts (MDS-RS)	AML with t(8;21)(q22;q22.1); *RUNX1-RUNX1T1*
MDS-RS and single lineage dysplasia	AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*
MDS-RS and multilineage dysplasia	AML with *PML-RARA*
MDS with multilineage dysplasia	AML with t(9;11)(q21.3;q23.3); *MLLT3-KMT2A*
MDS with excess blasts	AML with t(6;9)(q23;q34.1); *DEK-NUP214*
MDS with isolated del(5q)	AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2, MECOM*
MDS, unclassifiable	AML (megakaryoblastic) with t(1;22)(p13.3;q13.3); *RBM15-MKL1*
Provisional entity: Refractory cytopenia of childhood	*Provisional entity: AML with BCR-ABL1*
Myeloid neoplasms with germline predisposition	AML with mutated *NPM1*
	AML with biallelic mutations of *CEBPA*
	Provisional entity: AML with mutated RUNX1
	AML with myelodysplasia-related changes
	Therapy-related myeloid neoplasms
	AML, NOS
	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 sarcoma
	Myeloid proliferations related to Down syndrome
	Transient abnormal myelopoiesis (TAM)
	Myeloid leukaemia associated with Down syndrome

Furthermore**,** there is no single factor known to cause these genetic alterations, in many cases it is probably due to a combination of different factors such as smoking, some anti-cancer treatments, exposure to certain chemicals (such as benzene) and exposure to high levels of ionising radiation.

Sporadic MDS, *de novo* MDS and the closely related secondary AML evolving from an antecedent MDS are predominantly sporadic diseases that affect the elderly, with a median age of diagnosis over 70 years, and generally carry a poor prognosis (Sekeres *et al*., 2008; Steensma, 2015b; Khwaja *et al*., 2016). Although, adult patients less than 50 years of age are sporadically affected by MDS. In addition, in children, refractory cytopenia of childhood and juvenile myelomonocytic leukaemia (JMML) are considered distinct entities and are more related to congenital bone marrow failure and familial leukaemia syndromes than to adult MDS (Niemeyer & Baumann, 2011). Furthermore, Hirsch *et al*. (2017) found that high-risk MDS (with excess blasts) was more common among early onset (range, 20-50 years old) MDS patients (35% against 24%, P=0.048) while lower-risk MDS (with single lineage dysplasia, MDS with ring sideroblasts with single and multilineage dysplasia and, MDS with multilineage dysplasia) predominated in MDS patients over 50 years of age (28% against 41%, P=0.042). Additionally, the number of mutations tended to be higher in advanced WHO subtypes and high-risk prognostic groups, which is also associated with increasing intratumoral subpopulations and worse prognosis (Haferlach *et al*., 2014).

Congenital disorders such as Down syndrome, Fanconi anaemia (FA), dyskeratosis congenita (DC) and Bloom syndrome are also associated with MDS. In these syndromes affected individuals usually present with MDS at an earlier age (Germing *et al*., 2008), suggesting a “multiple-hit” mechanism of cancer development with genetic and environmental factors. The genetic variations in haematopoietic cells,

whether inherited or acquired, can affect apoptosis and cell differentiation which are crucial to increase the susceptibility of the affected precursor cell to further DNA damage, contributing to an accumulation of secondary genetic alterations that conclusively results in the development of definite MDS and AML (Tefferi & Vardiman, 2009).

Finally, although MDS is classified as cancer by WHO and is treated by oncologists in many settings, and MDS shares some biological features with leukaemia or other overt neoplasms, there are other features of MDS that are not typical of cancer, such as response to immunosuppressive therapy in some cases and stability for more than a decade in others (Shlush & Minden, 2015).

1.2.1 Patterns of acquired genetic variants in MDS

The spectrum of genetic abnormalities in MDS implicates a wide range of molecular mechanisms in the pathogenesis of this disorder, including RNA splicing, transcription factors, DNA modification, chromatin regulation, cell signalling and epigenetic regulators (Haferlach *et al*., 2014). RNA splicing is the most commonly mutated pathway in MDS and occurs early in disease evolution. These mutations play a major role in determining the clinical features of the disease, with differences in morphological features seen on bone marrow biopsy as they may influence the subsequent genomic evolution of the disease. It happens because the patterns of cooperating mutations are very different between the genes with driver mutations (Papaemmanuil *et al*., 2013). Specific alterations present in individual patients with MDS may explain much of the heterogeneity in clinical phenotype associated with MDS and can predict prognosis and response to therapy (Bejar *et al*., 2011).

Chromosomal abnormalities are present in approximately half of patients with MDS. The most common of these are loss of 5q (-5q), loss of 7 or 7q (-7/7q), trisomy 8 (+8), loss of 20q (-20q), and loss of Y (-Y) (Haase *et al*., 2007; Bejar *et al*., 2011). Del(5q) is the only cytogenetic or molecular genetic abnormality that is considered a specific MDS subtype by WHO criteria. The presence of +8, -Y, or del(20q) is not considered to be MDS-defining in the absence of diagnostic morphologic features of MDS (Arber *et al*., 2016).

Acquired somatic variants have been identified in several genes in MDS (Table 1.2), where the genes involved in chromatin regulation (*ASXL1* and *EZH2*), DNA repair and transcriptor factor (*TP53*), transcriptor factor (*RUNX1*), RNA splicing (*SF3B1*, *SRSF2*, and *U2AF1*), and DNA methylation (*TET2* and *DNMT3A*) are the most frequently mutated (Arber *et al*., 2016). However, deleterious variants in most of these genes can be found in different myeloid neoplasms and are not specific for MDS (Bejar *et al*., 2011).

The number of somatic variants increases linearly with age and on average, patients over 50 years of age have more somatic variants in spliceosomal, epigenetic modifier, and RAS gene families. Furthermore, there are age-related differences in molecular features among elderly patients with MDS, where somatic variants in the genes involved in RNA splicing (*SRSF2*) and DNA methylation (*TET2* and *DNMT3A*) occur with a high frequency. While somatic variants in genes involved in chromatin regulation (*ASXL1*), DNA methylation (*TET2*), DNA repair and transcriptor factor (*TP53* and *RUNX1*) are the most frequently mutated genes in patients with early onset MDS (range, 20-50 years old) (Hirsch *et al*., 2017).

Table 1.2. Frequent genetic abnormalities observed in sporadic MDS

Genes and chromosomal abnormalities grouped by molecular mechanisms			
Chromatin regulation	**Cohesin/CTCF pathway**	**Cell adhesion**	**Other**
ASXL1	*CTCF*	*LAMB4*	*MECOM*
ATRX	*MED12*		*MLL*
BCOR		**Chromosome segregation**	*MLL2*
EZH2	**RNA splicing**	*RAD21*	*OGT*
KDM6A	*DDX41*	*SMC1A*	*PIGA*
NCOR2	*LUC7L2*	*SMC3*	*PPM1D*
PHF6	*PRPF8*	*STAG2*	*PTEN*
IDH1	*PTPN11*		*SETBP1*
IDH2	*SF1*	**Other**	*SH2B3*
	SF3B1	+19	*SIMC1*
DNA repair	*SFRS2*	+8	*SUZ12*
ATM	*SRSF2*	*APC*	
BRCC3	*U2AF1*	*BCORL1*	
DCLRE1C	*U2AF2*	*CALR*	
FANCL	*ZRSR2*	*CDH23*	
TP53		*CDKN2A*	
	DNA methylation	*CREBBP*	
Negative regulation of cellular process	*DNMT3A*	*CSF3R*	
BRAF	*TET2*	*CTNNA1*	
NF1		*CUX1*	
	NOTCH signaling	*DDX54*	
RAS signaling	*FBXW7*	*del(11q)*	
CBL		*del(12q)*	
KRAS	**Cytokine receptor**	*del(17p)*	
NRAS	*FLT3*	*del(20q)*	
	KIT	*del(5q)*	
Transcription factor	*MPL*	*del(7q)*	
CEBPA		*DHX29*	
ETV6	**Receptor/signaling**	*EED*	
GATA2	*GNAS*	*EP300*	
IRF1	*GPRC5A*	*ERBB4*	
NPM1		*GLI1*	
RUNX1	**Kinase signaling**	*GNB1*	
TP53	*JAK2*	*GPR98*	
WT1		*IRF4*	

References: Papaemmanuil *et al*. (2013); Haferlach *et al*. (2014); Lindsley *et al*. (2017).

It is known that there are haematological neoplasm-associated genetic abnormalities in the blood of some healthy people, especially older adults called clonal haematopoiesis of indeterminate potential (CHIP) (Biernaux *et al*., 1995; Steensma *et al*., 2015a). Although, acquired clonal mosaicism predicts an increased risk of development of a neoplasm, indicating that such alterations can represent disease-initiating events in some cases by promoting clonal expansion (Jacobs *et al*., 2012). The most common genes identified with variants in healthy population or in patients with non-haematological malignancies in analysis of a large cohort were *ASXL1*, *BCORL1*, *TP53*, *GNAS*, *SF3B1*, *DNMT3A*, *TET2*, *JAK2* and *PPM1D*. In addition, the frequency of these variants increased with aging (Xie *et al*., 2014; Jaiswal *et al*., 2014; Genovese *et al*., 2014; Kwok *et al*., 2014).

Further studies are required to determine the best management and monitoring of the patients with CHIP, to describe whether they have a higher risk of developing malignancies or any other outcome. The specific genetic variants, number of variants, and variant allele frequency may also influence the risk of progression and could further refine diagnostic criteria (Wong *et al*., 2015).

1.3 Acute myeloid leukaemia

Leukaemia is a cancer of the white blood cells, it is classified according to the type of white blood cell affected and the speed with which the cancer progresses. AML is one type of leukaemia and is an aggressive disorder characterised by a fast clonal proliferation of very large numbers of malignant immature myeloid cells in the bone marrow (Figure 1.2) (Kupsa *et al*., 2012). These cells will never mature into proper blood cells, fundamental to a healthy immune system, and so patients with AML have an increased risk of infection. Normal blood cells are prevented from being made by the clustering of cancer cells in the bone marrow. Most of the complications of leukaemia are caused by the lack of normal cells in the blood (bone marrow failure), rather than the leukaemia cells themselves.

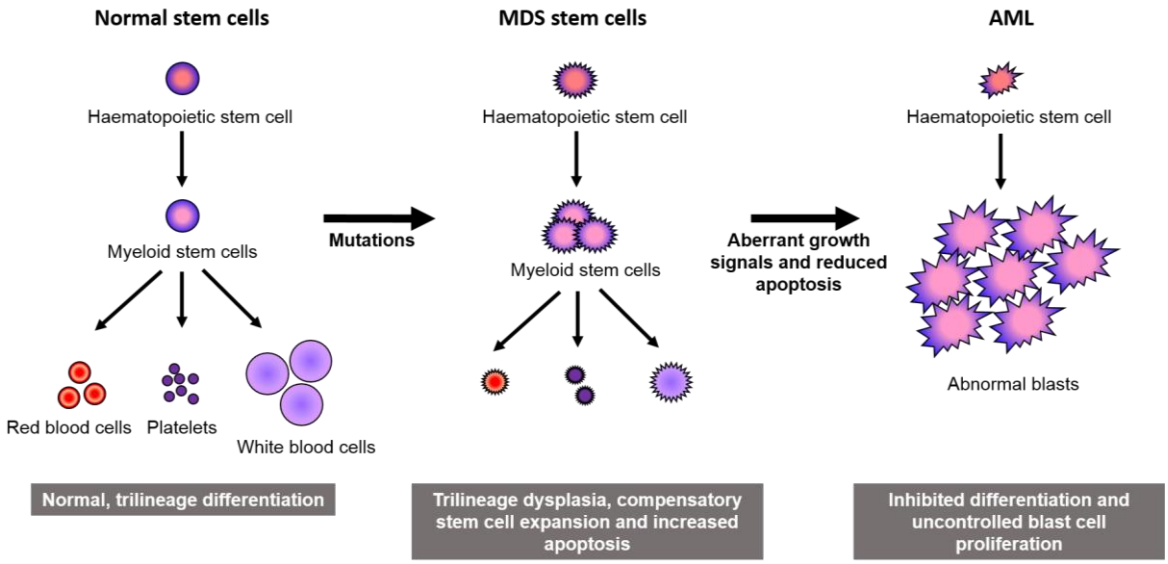


Figure 1.2. Schematic of haematopoietic stem cells differentiation. Differences in cell morphology and proliferation are shown when the haematopoietic stem cells are mutated leading to the development of MDS and AML.

AML accounts for less than 1% of all new cancer cases in 2015 in the UK. There were approximately 3,100 new cases of AML between 2013 and 2015 in the UK, with approximately 8 cases diagnosed every day. Cancer research UK predicts that 1 in 200 men and 1 in 255 women will be diagnosed with AML during their lifetime (http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml). As in MDS, different factors play a role in AML development. Cigarette smoke is a minor risk for AML (Bjork *et al*., 2009) and benzene exposure increases the risk of leukaemia (Vlaanderen *et al*., 2010). Patients who are receiving radiotherapy or chemotherapy to treat some other cancers may go on to develop AML. When this happens, it is called secondary leukaemia or therapy-related leukaemia. The risk of developing AML from other blood disorders is also low (Greaves, 1997).

AML is a heterogeneous genetic disorder characterised by the accumulation of genetic abnormalities. The critical initiation step towards the disease pathogenesis occurs with the generation of chimeric fusion genes from translocation/inversion events or with preleukaemic mutation rising in the haematopoietic stem/progenitor cell (Hou *et al*., 2014). These preleukaemic events precede the development of leukaemia, which happens when further mutations co-occur later in the progenitor cells (Shlush *et al*., 2014). The presence of mutations that precede development of leukaemia likely reflects a cell's cumulative inability to completely repair the multitude of mutations that occur randomly over time (Tomasetti & Vogelstein, 2015).

Studies have shown that translocations/inversions underlie disease pathogenesis in approximately 80% of AML in children and 30% in young adults, whereas only a minority of AMLs in older adults have balanced rearrangements. Moreover, approximately 40% of adult AML with a highly heterogeneous clinical outcome have

normal karyotype but might exhibit molecular abnormalities that retain the normal characteristics of multipotent differentiation (Grimwade & Mrozek, 2011). Chimeric fusion genes are insufficient to induce leukaemic transformation in their own right. However, they may provide a competitive advantage, generating populations of cells in which secondary mutations may arise and be selected for (Miyamoto *et al*., 2000).

Leukaemias often comprise heterogeneous mixtures of subclones (Welch *et al*., 2012). Studies in large cohorts have shown that the development of AML follows specific and ordered evolutionary trajectories (Papaemmamuil *et al*., 2016) and the identification of this stepwise acquisition of mutations during the development of the disease is possible with the use of next generation sequencing (NGS) approaches. The use of NGS with an extensive sequencing depth have allowed the quantification of the differences in the relative proportion of co-occurring mutations within the tumour at the time of diagnosis (Mardis *et al*., 2009).

Variant allele frequency (VAF) is the comparison of the proportion of reads that contain the mutant allele to that of wild-type allele reads, with the obtained relative proportions able to infer clonal architecture (Figure 1.3). Such analysis has demonstrated the emergence of new clones carrying novel mutations at different times during the evolution of the leukaemia (Walter *et al*., 2012) and also those mutations that persist after treatment failure, prioritising obvious critical molecular drivers of leukaemia development and relapse to target therapy (Ding *et al*., 2012). Mutations with a high VAF are predicted to occur early, whereas mutations present only in a minority of cells are likely to be acquired at later stages of leukaemia development. High VAF may also be observed as a result of acquired uniparental disomy, for example in a proportion of AMLs with associated *FLT3*-ITD and *TET2* mutations (Fitzgibbon *et al*., 2005; Mohamedali *et al*., 2009).

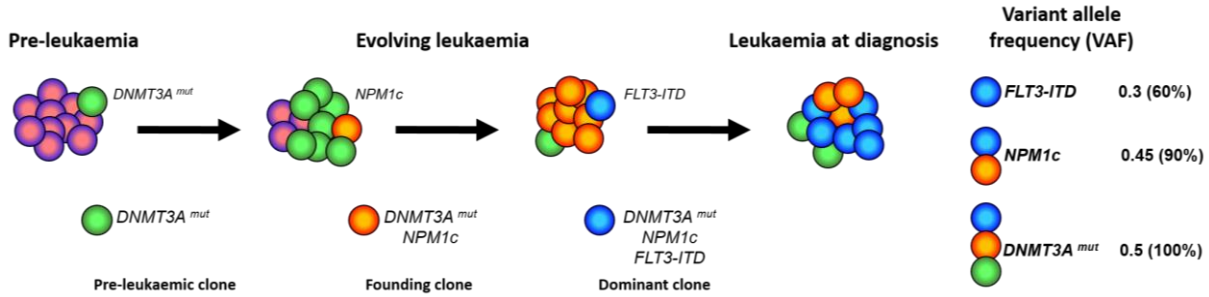


Figure 1.3. Clonal evolution and clonal heterogeneity of AML. Mutation of *DNMT3A* is the earliest event and occurs before the disease development. *NPM1c* occurs as the disease-defining mutation in the founding clone, with further acquisition of a *FLT3-ITD* mutation in a clone during leukaemia expansion, which become dominant at diagnosis. Quantification of VAF of each mutation, allows the demonstration of the evolution of the disease according to the temporal acquisition of mutations (adapted from Grimwade *et al*., 2016).

The outcome of AML patients greatly differs according to their genetic abnormalities and the overall survival of AML patients is correlated with the number of driver mutations independent of the patient's age and white cell count. In addition, the clinical effect of some driver mutations is modified by the wider genomic context in which they occur due to gene-gene interactions, in which the prognostic effect of one gene is significantly altered if another gene is co-mutated (Papaemmanuil *et al*., 2016). Although, studies in healthy individuals have demonstrated that mutation frequency increases as a function of age and clonal expansion in older adults was confirmed by the presence of mutations associated with myeloid malignancies, increasing the risk of blood cancer development (Genovese *et al*., 2014; Jaiswal *et al*., 2014; Xie *et al*., 2014; McKerrell *et al*., 2015).

Undoubtedly, a more complete genetic characterisation of haematological malignancies has great potential to elucidate the molecular basis for the clinical

heterogeneity of these disorders and to identify disease subtypes with shared outcomes and responses to therapy (Bejar *et al*., 2011).

1.3.1 Patterns of acquired genetic variants in AML

The discovery of somatic chromosomal abnormalities including balanced translocations in the leukaemic cells of some patients established that AML is a genetic disease (Rowley, 1973; Rowley *et al*., 1977). This blood disorder is characterised by the accumulation of genetic abnormalities such as the formation of chimeric fusion genes because of gene translocations and inversions listed in Table 1.3. These chromosomal rearrangements occur between genes that encode haematopoietic transcription factors (*RARA*, *RUNX1*, or *CEBPA*), epigenetic regulators (*KMT2A*, *NSD1*, *KAT3A*), and components of the nuclear pore complex (*NUP98*, *NUP214*) (Grimwade *et al*., 2016).

The WHO classifications recognised these recurrent chromosomal rearrangements as separate entities that are sufficient to diagnose AML without evidence of bone marrow blasts percentage ≥20%. In this classification, molecular groups in adult AML include t(15;17), t(9;11), t(8;21), inv(16), t(16;16), t(6;9), inv(3), t(3;3), t(1;22), *MLLT3-KMT2A*, *DEK-NUP214*, *RUNX1-RUNX1T1*, *CBFB-MYH11*, *PML-RARA*, as well as mutations in *CEBPA*, *RUNX1*, *GATA2*, *MECOM* or *NPM1* (Arber *et al*., 2016).

Table 1.3. Frequent genetic abnormalities observed in sporadic AML

Genes and chromosomal abnormalities grouped by molecular mechanism			
Chromatin regulation	**RNA splicing**	**DNA methylation**	**Other**
ASXL1	*DDX41*	*DNMT3A*	inv(3)/t(3;3)
ATRX	*PTPN11*	*TET2*	*KDM5A*
BCOR	*SF1*		*MLL*
EZH2	*SF3A1*	**Other**	*MLL2*
IDH1	*SF3B1*	+11/11q	*MLL3*
IDH2	*SRSF2*	+13	*MLL5*
KDM6A	*U2AF1*	+21	MLL-ELL
PHF6	*U2AF2*	+22	MLL-MLLT1
	ZRSR2	+8/8q	MLL-MLLT10
DNA repair		-12/12p	MLL-MLLT3
TP53	**NOTCH signaling**	-17/17P	MLL-MLLT4
	FBXW7	-18/18q	MLL-PTD
Negative regulation of cellular process		-20/20q	*MYC*
BRAF	**Cytokine receptor**	-4/4q	MYST3-CREBBP
NF1	*FLT3*	-5/5q	NPM1-MLF1
CBLB	*KIT*	-7/7q	NUP98-HOXA9
	MPL	-9q	NUP98-NSD1
RAS signaling		abn3q	PICALM-MLLT10
CBL	**Receptor/signaling**	*ASXL2*	PML-RARA
KRAS	*GNAS*	BCR-ABL	PRPF40B
NRAS		CBFB-MYH11	*PTEN*
	Kinase signaling	*CDKN2A*	*RB1*
Transcription factor	JAK2	*CREBBP*	RUNX1-MECOM
CEBPA		*CUX1*	RUNX1-RUNX1T1
ETV6	**Chromosome segregation**	DEK-NUP214	*SH2B3*
GATA2	*RAD21*	*EP300*	t(11q23;x)
NPM1	*SMC1A*	FLT3-ITD	t(15;17)
RUNX1	*SMC3*	FLT3-TKD	T(6;9)
TP53	*STAG1*	FUS-ERG	t(8;21)
WT1	*STAG2*	*GATA2/MECOM*	-Y
		IKZF1	
		inv(16)	

References: Papaemmanuil *et al*. (2016); Grimwade *et al*. (2016).

The outcome for AML patients differs considerably according to the cytogenetic abnormality and there are variations even in outcomes between patients with the same chromosomal abnormality. In addition, genomic heterogeneity in AML is associated with complex epigenetic heterogeneity (Li *et al*., 2016). However, approximately half of AML patients have normal karyotype and their outcome is also heterogeneous. The advances in technology provided important insights into the molecular abnormalities underlying AML with normal cytogenetics and those with chromosome losses or gains that were previously poor understood (Grimwade *et al*., 2016).

Systematic studies of the genomic landscape of AML have generated an extensive catalogue of leukaemia genes. The AML biology consists of many different genetic alterations in patients that typically have more than one driver mutation. The disease evolves over time, with multiple competing clones coexisting at any time (Walter *et al*., 2012).

The molecular landscape of AML in young adults (<60 years) consists of variants in genes involved in signalling pathway (*FLT3*, *RAS*, *KIT*, *CBL*, *NF1*, and *PTPN11*), epigenetic modifier genes (DNA methylation: *DNMT3A*; DNA demethylation: *TET2*, *IDH*1/2, *WT1*; chromatin regulation: *ASXL*1/2, *PHF6*, *BCOR*, and *EZH2*), splicing factor genes (*SRSF2*, *SF3B1*, *U2AF1*, and *ZRSR2*), cohesion complex members (*RAD21*, *SMC1A*, *SMC3*, and *STAG1/2*), transcription factor genes *CEBPA* and *RUNX1* and, tumour suppressor *NPM1* (Grimwade *et al*., 2016). The same chromatin and splicing factors genes are also frequently mutated in high-risk myeloproliferative neoplasms and in MDS (Haferlach *et al*., 2014; Papaemmanuil *et al*., 2016).

This elucidation of the molecular landscape of AML has already facilitated the

identification and development of novel and targeted therapy agents that includes novel epigenetic therapies, anti-apoptotic agents, and selective inhibitors of nuclear export and immunotherapies that target a number of specific cellular processes such as intracellular signalling, transcriptional control, epigenetic regulation, and mRNA splicing (Chen *et al*., 2013). It has been demonstrated how important it is to target an early and/or driver mutation as it will be present within all clones for the eradication of disease (Fielding *et al*., 2014). On the other hand, recent studies have shown that minimal residual disease assessment using NGS provides a more powerful predictor of disease outcome than mutational profile, defining patients at risk of relapse (Klco *et al*., 2015).

The knowledge of the clonal architecture and the role of clonal haematopoiesis in the development and prognosis of blood cancers are essential for the development of effective therapies. Furthermore, as the complete prognostic significance of additional mutations or combinations of mutations is currently unknown, a key challenge will be to determine which are the most informative molecular markers that most reliably track leukaemic populations irrespective of the recognised clonal heterogeneity, rather than pre-leukemic clones that can persist in patients in long-term remission following chemotherapy. As consequence, it is necessary to refine and optimise combinational knowledge of mutation complement for prognostication along with demographic characteristics and type of AML (*de novo*, secondary or therapy related). Finally, correlation of mutational data with genome-wide scale via epigenetic and proteomic analysis data sets may also further refine our understanding of AML biology, improve outcome prediction and treatment choice (Grimwade *et al*., 2016).

1.4 Familial myelodysplasia and/or acute myeloid leukaemia with germline predisposition syndromes

Family history of myelodysplasia and/or acute myeloid leukaemia (MDS/AML) is here defined as the presence of two or more relatives with bone marrow failure, where at least one of them presents with MDS or AML. Familial occurrence of MDS/AML is rare, less than 300 families were reported to have mutations in one of known disease causing gene (published mutations is listed in Tables A1.1 – A1.9 in Appendix 1). Several of these cases occur in the setting of genetic syndromes associated with increased risk of developing MDS or AML, including inherited bone marrow failure syndromes, such as Fanconi anaemia, severe congenital neutropenia, Shwachman-Diamond syndrome and dyskeratosis congenita. However, rare familial cases of MDS/AML have been reported outside the context of such syndromes who carry germline predisposing mutations (Gao *et al*., 2014).

Genomic investigation of families with MDS/AML has revealed multiple genes that when altered predispose to the disease development. Patients who carry mutation within these genes often have other concomitant characteristics, particularly in those diagnosed in adulthood. Table 1.4 summarises the known genes predispositions to inherited MDS/AML.

WHO classification included familial MDS/AML in the myeloid neoplasms with a predisposition germline mutational group, which comprises MDS, MDS/myeloproliferative neoplasms (MPN), and AML that occur on the background of a predisposing germline mutations with the following classification: (1) myeloid neoplasms with germline predisposition without a pre-existing disorder or organ dysfunction with mutations in *CEPBA*, (2) myeloid neoplasms with germline

predisposition and pre-existing platelet disorders with mutations in *RUNX1*, *ANKRD26* and *ETV6*, (3) myeloid neoplasms with germline predisposition and other organ dysfunction with mutations in *GATA2*, myeloid neoplasms associated with bone marrow failure syndromes, myeloid neoplasms associated with telomere biology disorders, myeloid neoplasms associated with Down syndrome and JMML associated with neurofibromatosis, Noonan syndrome or Noonan syndrome-like disorders (Arber *et al*., 2016). These patients must have a different management and their families should get genetic counselling.

In familial neoplasms, mutations are frequently present in the heterozygous state, most commonly resulting in loss of function alleles and subsequent haploinsufficiency, although gain-of-function mutations have also been reported in *GATA2* (Godley, 2014).

Table 1.4. Genes frequently mutated in familial MDS/AML

Predisposition genes to familial MDS/AML	
Transcription factor	
RUNX1	Song *et al*. (1999)
CEBPA	Smith *et al*. (2004)
GATA2	Hahn *et al*. (2011); Ostergaard *et al*. (2011)
ANKRD26	Noris *et al*. (2013)
ETV6	Zhang *et al*. (2015)
Cell signaling	
SRP72	Kirwan *et al*. (2012)
Telomere maintenance	
TERC	Kirwan *et al*. (2009)
TERT	Kirwan *et al*. (2009)
RNA splicing	
DDX41	Polprasert *et al*. (2015)

1.4.1 Familial platelet disorder with propensity to myeloid malignancy (FPD/AML)

Familial platelet disorder with propensity to myeloid malignancy is an autosomal dominant familial MDS/AML syndrome caused by germline mutations in *RUNX1*, which is located on chromosome 21. Besides reports of missense mutation in the DNA biding domain, the majority of mutations in this gene are frameshift, nonsense, or deletion that result in premature protein truncation, leading to protein loss-of-function or confer dominant-negative effects to the remaining *RUNX1* allele (Sakurai *et al*., 2014). Often, patients with FPD/AML present with life-long mild to moderate thrombocytopenia with platelet defects and it can vary even within affected families from individuals with a normal platelet count, to severe thrombocytopenia or to childhood MDS/AML at the time of diagnosis (Song *et al*., 1999). The haematological malignancies described in FPD/AML patients include MDS, AML, and T-cell acute lymphoblastic leukemia (ALL) with 40% lifetime risk to develop MDS or AML and an average age at diagnosis of 33 years (Owen *et al*., 2008).

1.4.2 Thrombocytopenia 2

Thrombocytopenia 2 is an autosomal dominant disorder caused by germline mutations in the 5' untranslated region (5' UTR) of *ANKRD26* on chromosome 10. These mutations might enhance *ANKRD26* expression and dysregulation of apoptosis might be the pathogenetic mechanism. (Pippucci *et al*., 2011). 5' UTR has been described to be involved in transcription regulation and it is important for the

regulation of translation of transcripts by differing mechanisms (Cenik *et al*., 2011). Hence, *ANKRD26* has a role in transcriptional and translational regulation.

Thrombocytopenia 2 is characterised by moderate thrombocytopenia with or without bleeding propensity, similar to FPD/AML (Pippucci *et al*., 2011). The prevalence of thrombocytopenia 2 is not well described and individuals carrying *ANKRD26* mutations are clinically difficult to distinguish from those with FPD/AML (Noris *et al*., 2013).

1.4.3 Thrombocytopenia 5

Thrombocytopenia 5 is an inherited autosomal dominant MDS/AML predisposition syndrome associated with germline missense mutations in *ETV6* on chromosome 12. These mutations appear to have a dominant negative function, disrupting the nuclear localization of the protein and resulting in reduced expression of platelet-associate genes. Individuals with germline mutations in ETV6 present with variable degree of thrombocytopenia and mild to moderate bleeding tendencies and they are at increased risk for all haematological malignancies, including MDS, AML, chronic myelomonocytic leukaemia (CMML), B-lymphoblastic leukaemia, and plasma cell myeloma (Zhang *et al*., 2015). It is noteworthy mentioning that *ETV6* and *DDX41* were discovered as familial MDS/AML causing genes in 2015 by Zhang *et al*. (2015) and Polprasert *et al*. (2015) respectively during the course of this study. *DDX41* will be discussed in chapter 4.

1.4.4 Familial AML with mutated *CEBPA*

Germline heterozygous mutations with high penetrance in *CEBPA* are the cause of familial AML, an inherited autosomal dominant disorder (Smith *et al*., 2004; Tawana *et al*., 2015). Patients carrying mutations in this gene located on chromosome 19, present long-term survival, with median age of AML onset of 24.5 years. They also appear to have a significant risk of late AML recurrence and this typically represent independent leukaemic episodes, characterised by a unique molecular profile that is distinct from that of the preceding tumour (Tawana *et al*., 2017). Frameshift mutations in *CEBPA* commonly occur in the N-terminal combined with C-terminal in-frame insertions or deletions (disrupting the DNA binding or leucine zipper domains) and, they are frequently accompanied by mutations in *GATA2* or *WT1* in familial AML (Fasan *et al*., 2014).

1.4.5 Familial MDS/AML with mutated *GATA2*

GATA2 deficiency is a clinically heterogeneous predisposition to MDS caused by germline mutations on chromosome 3. Patients carrying *GATA2* mutations are at significantly increased lifetime risk of MDS/AML, with a median age at onset of 29 years and they can present without any haematopoietic or organ system involvement prior to the development of MDS or AML. There are two distinct syndromic presentations caused by *GATA2* mutations, including Emberger syndrome and MonoMac syndrome.

Emberger syndrome is clinically characterised by primary lymphedema, cutaneous/extragenital warts, low CD4/CD8 T-cell ratio, and sensorineural hearing loss with a predisposition to MDS/AML. The MonoMac syndrome is characterised by pulmonary alveolar proteinosis, monocytes, dendritic cells and B/natural killer cell deficiencies, leading to the development of atypical mycobacterial or fungal infections, and MDS/AML predisposition. Both phenotypes can overlap, and these syndromes are considered to be part of the same autosomal dominant genetic disorder with variability because they share the same underlying genetic aetiology (Hahn *et al*., 2011).

1.4.6 Familial aplastic anaemia/MDS with *SRP72*

Only two families with autosomal dominant MDS and aplastic anaemia (AA) in adulthood have been identified to date as a result of germline mutations in the gene *SRP72*. The reported mutations are one missense and one deletion that caused a truncated protein. Little is known regarding the lifetime risk or incidence of this unusual condition given the rarity of these germline mutations (Kirwan *et al*., 2012).

1.4.7 Bone marrow failure syndromes

Typically, bone marrow failures are a group of disorders presenting in young age with characteristic physical features along with bone marrow failure and a predisposition to MDS/AML and other cancers. The majority of individuals with bone marrow failure syndromes will have syndromic phenotypic abnormalities such as pancreatic

dysfunction or multiple congenital anomalies at presentation. Patients with severe congenital neutropenia, Fanconi anaemia, Shwachman-Diamond syndrome and dyskeratosis congenita are at significantly increased risk for treatment-induced malignancies (Dokal & Vulliamy, 2010; Schulz *et al*., 2012).

1.4.7.1 Fanconi anaemia

Fanconi anaemia (FA) is a rare, autosomal recessive or rarely X-linked genetic predisposition to bone marrow failure with congenital limb anomalies including absent thumbs and other radial ray defects. FA cells are characterised by increased chromosomal fragility and breakage when treated with DNA cross-linking agents. Patients with FA present with progressive bone marrow failure with pancytopenia, up to 30% of familial MDS/AML incidence and there is an increased risk of solid tumours, particularly squamous cell carcinomas of the head and neck. Approximately 40% of patients with FA lack physical abnormalities associated with the disease and are also less likely to develop early-onset bone marrow failure (Rosenberg *et al*., 2008). Furthermore, apparently healthy relatives of a FA patient should be referred for genetic counselling and management of solid tumour risks as FA is caused by biallelic mutations in the 21 FA complementation groups, *FANCA*-*FANCW*. Mutations in *FANCA* is responsible for 70% of FA cases, followed by *FANCC* with 14%, while mutation in the remaining genes of this group account for up to 3% of the cases. However, *BRCA1*, *BRCA2*, *PALB2* and *RAD51C* are also part of this group and, when mutated these genes are associated with autosomal dominant predispositions to solid tumour development, mainly ovarian and breast cancer (Bogliolo & Surrallés, 2015; Sawyer *et al*., 2015). FA will be further discussed in chapter 3.

1.4.7.2 Dyskeratosis congenita

Patients with dyskeratosis congenita (DC) are at increased risk for bone marrow failure, MDS, or AML, solid tumours, and pulmonary fibrosis. DC is a telomere biology disorder characterised by very short telomeres and classically by a diagnostic triad of dysplastic nails, lacy reticular skin pigmentation, and oral leukoplakia; the median age at onset is 37 years. However, these features are not present in all individuals with DC and they may or may not develop over time (Dokal, 2011; Tummala *et al*., 2015). Pathogenic germline mutations can be detected in approximately half of DC cases and there are 12 known disease genes: *DKC1*, *TERC*, *TERT*, *NOP10*, *NHP2*, *ACD*, *TINF2*, *WRAP53*, *CTC1*, *RTEL1*, *PARN* and *NAF1* (Walne *et al*., 2013; Walne *et al*., 2016). Additionally, germline mutations in *TERT* and *TERC* have also been associated with idiopathic AA and idiopathic pulmonary fibrosis (Yamaguchi *et al*., 2005).

MDS and AML heterogeneity is not only explained by its morphological diversity, but also by the increasing number of molecular pathways and hallmarks that participate in disease initiation and evolution. The main biological hallmarks in sporadic MDS and AML have been well described and include both genomic and epigenomic alterations in transcription factors, epigenetic modulators, miRNA, microenvironment and innate immunity (Bejar *et al*., 2011). The most common mutations in these sporadic diseases are in epigenetic modifiers (*TET2*, *IDH1/2*, *DNMT3A*, *EZH2* and *ASXL1*) as well as in genes involved in spliceosome machinery (*U2AF1*, *SRSF2* and *SF3B1*) (Shahrabi *et al*., 2016; Cedena *et al*., 2017) Furthermore, it is well known that inducing differentiation and apoptosis of leukaemic blasts by DNA-hypomethylating agents such as azacytidine (AZA) and decitabine (DAC), represent well-tolerated alternative treatment approaches to most patients with MDS and AML as they are older and

exhibit a poor prognosis (Bohl *et al*. 2018). Clinical responses to AZA and DAC support epigenetic alterations as a fundamental pathophysiology in sporadic MDS and subsets of sporadic AML (Fenaux *et al*., 2009; Traina *et al*., 2014). These hypomethylating agents are nucleoside analogs that integrate into DNA and inhibit DNA methyltransferases (Saygin & Carraway, 2017), affecting the proliferation of malignant cells. (Moudra *et al*., 2016).

Although mutations in epigenetic genes were not found to cause familial MDS/AML so far, some known familial MDS/AML causing genes such as *RUNX1*, *CEBPA* and *TERT* have been involved with epigenetic mechanisms in sporadic forms of MDS and AML. Additionally, *ANKRD26* mutations might have a role in transcriptional and translational regulation as it occurs in its 5'UTR.

RUNX1 mutations were observed to be almost mutually exclusive with recurrent genetic abnormalities in sporadic AML, and they frequently co-occurred with mutations in epigenetic modifiers (*ASXL1*, *IDH2*, *KMT2A*, *EZH2*), components of the spliceosome complex (*SRSF2*, *SF3B1*) and *STAG2*, *PHF6*, *BCOR* (Gaidzik *et al*., 2016). *RUNX1* is also subject to translocations that promotes fusion proteins with aberrant transcriptional activities (Loke *et al*., 2018). The RUNX1-ETO t(8;21) is perhaps the most frequent chromosomal translocation associated with AML. The translocation creates a fusion protein that consists of N-terminal portion of RUNX1 on chromosome 21 and full-length eight-21 (ETO) on chromosome 8, producing the chimeric gene RUNX1-ETO. In addition to blocking differentiation, RUNX1-ETO is also shown to induce growth arrest in AML cells and leads to the silencing of myeloid maturation genes. RUNX1-ETO recruits histone deacetylases and DNMT1 to RUNX1 DNA binding sites and acts as a potent negative regulator of transcription of the genes normally controlled by *RUNX1*, such as *CEBPA*. (Buchi *et al*., 2014). RUNX1-ETO

might suppresses endogenous DNA repair in cells to promote mutagenesis, which facilitates acquisition of cooperating secondary events. RUNX1 also induces DNA demethylation by recruiting DNA demethylation machinery to its binding sites, which likely contributes to hematopoietic development (Suzuki *et al*., 2017).

It is well known that mutations in *CEBPA* cause cell differentiation inhibition in AML. However, Sinha *et al.* (2015) found that pathogenic somatic variants in *CEBPA*, *WT1* and *IDH2* were found to be genetic drivers of DNA hypermethylation in sporadic AML using an integrative analysis of The Cancer Genome Atlas data (Cancer Genome Atlas Research Network, 2013) based on Boolean implications.

Genetic and epigenetic regulations of *TERT* seem to play important roles in pathophysiology and clinical outcome in human cancers. The dysregulation of *TERT* in malignant cells can be explained by alteration of the *TERT* through mutations, DNA methylation, histone methylation, histone acetylation, non-coding RNA, and guanine-quadruplexes (GQ) structures (Zhao *et al.*, 2016). Normal human somatic cells have a non-methylated/hypomethylated CpG island within the *TERT* promoter region, while telomerase-positive cells have at least a partially methylated promoter region. Chromatin remodelling changes the state of histones present within the *TERT* promoter by influencing the binding of transcription factors. Non-coding RNAs can target epigenetic-modifying enzymes, as well as transcription factors involved in a post-transcriptional manner by binding to the 3'UTR of TERT mRNA, or by affecting the presence of transcription factors responsible for the transcription or repression of *TERT* (Lewis & Tollefsbol, 2016).

The full epigenetic landscape of the *TERT* promoter region in sporadic AML is characterised by the heterogeneous upstream *TERT* promoter methylation profile

with conservative hypomethylated transcription start sites alleles. This distinct epigenetic change of *TERT* promoter implies alteration of the secondary or tertiary structure of the *TERT* promoter region such as CG structures. These can modify configurational interactions with transcription factors and control *TERT* expression in leukaemia cells. GQ structures within the promoters in cancer related genes, such as *MYC*, *KIT*, *KRAS* and *TERT*, lead to down-regulation of gene expression (Zhao *et al.*, 2016).

Finally, most of germline heterozygous mutations including missense, frameshift, nonsense, or deletion resulting in premature protein truncation, that cause familial MDS/AML were found in transcription factors genes (5) as well as in telomere maintenance (2), cell signalling (1) and RNA splicing (1) genes. Which is in contrast with sporadic form of these diseases where the majority of mutations were found in epigenetic modifiers and RNA splicing genes. Despite the progress made in the comprehension of these haematological malignancies, more study needs to be performed to fully characterise the occurrence of these somatic and germline variation in the biological and clinical setting. Furthermore, genes that are known to cause familial MDS/AML can be associated with complex phenotypes such as *GATA2*, *TERC* and *TERT*, or mainly with familial MDS/AML phenotype such as *RUNX1*, *CEBPA*, *SRP72*, *ANKRD26*, *ETV6* and *DDX41*.

Research on inherited forms of MDS/AML allowed the acknowledgement that these diseases are more common than initially realised and can frequently be present in adulthood, rather than exclusively in childhood. These disorders are heterogeneous regarding their underlying genetic mutations, clinical presentations, and progression to MDS/AML. However, as a group, they all share the unique requirement for a high index of clinical suspicion to allow appropriate genetic counselling, genetic testing,

and mutation-specific clinical management (University of Chicago Hematopoietic Malignancies Cancer Risk Team, 2016).

1.5 Our cohort of families with familial MDS/AML

Over the last 20 years our group has collected a significant number of families (n=115) that have two or more individuals with bone marrow failure (AA, MDS or leukaemia) with at least one individual who has MDS or AML. At the beginning of this project in 2014, 12 MDS/AML families had their predisposing genetic variant to the disease identified using a range of genetic studies (Figure 1.4; Table 1.5).

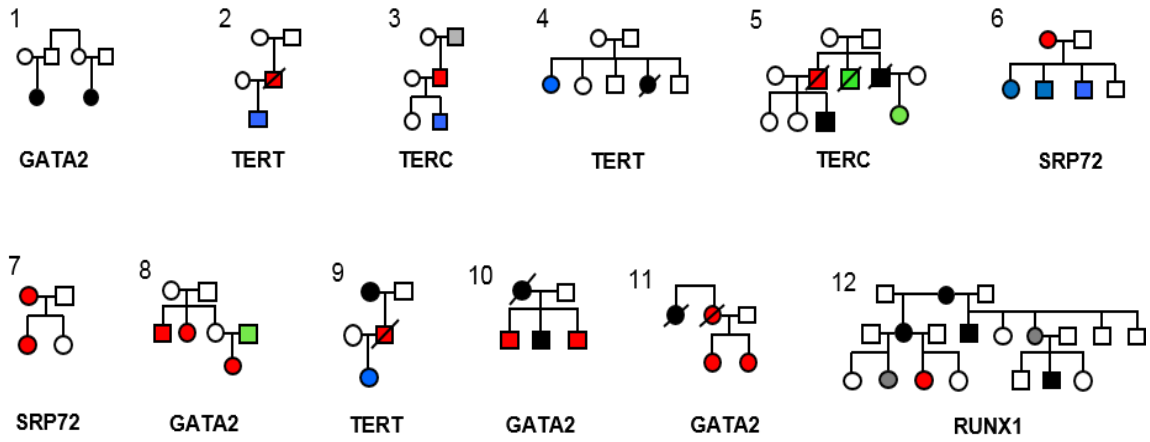


Figure 1.4. Genetically characterised MDS/AML families. Affected individuals are coloured as follows: red: MDS, black: AML, blue: AA, grey: thrombocytopenia, green: other leukaemia/cancer. The predisposing gene to the disease is indicated in each characterised family.

Table 1.5. Characterised MDS/AML families with their respective disease causing variant

Characterised MDS/AML families			
Family	**Gene**	**Variant**	**Amino acid change**
1	*GATA2*	c.310 – 311insCC	p.L105Pfs*15
2	*TERT*	c.1892 G>A	p.R631Q
3	*TERC*	c.212 C>G	N/A
4	*TERT*	c.2354 C>T	p.P785L
5	*TERC*	c.309 G>T	N/A
6	*SRP72*	c.1064 – 1065del	p.T355Kfs*19
7	*SRP72*	c.620 G>A	p.R207H
8	*GATA2*	c.121 C>G	p.P41A
9	*TERT*	c.248 G>C	p.R83P
10	*GATA2*	c.1187 G>A	p.R396E
11	*GATA2*	c.1061 C>T	p.T354M
12	*RUNX1*	c.602 G>A	p.A201Q

N/A: not applicable.

During this PhD studies, an additional nine families from our familial MDS/AML cohort and further 10 families from our DC and idiopathic BMF registries have been characterised using WES and targeting sequencing, respectively. Currently there are 33 MDS/AML families that remain uncharacterised and had whole exome sequencing undertaken (Figure 1.5).

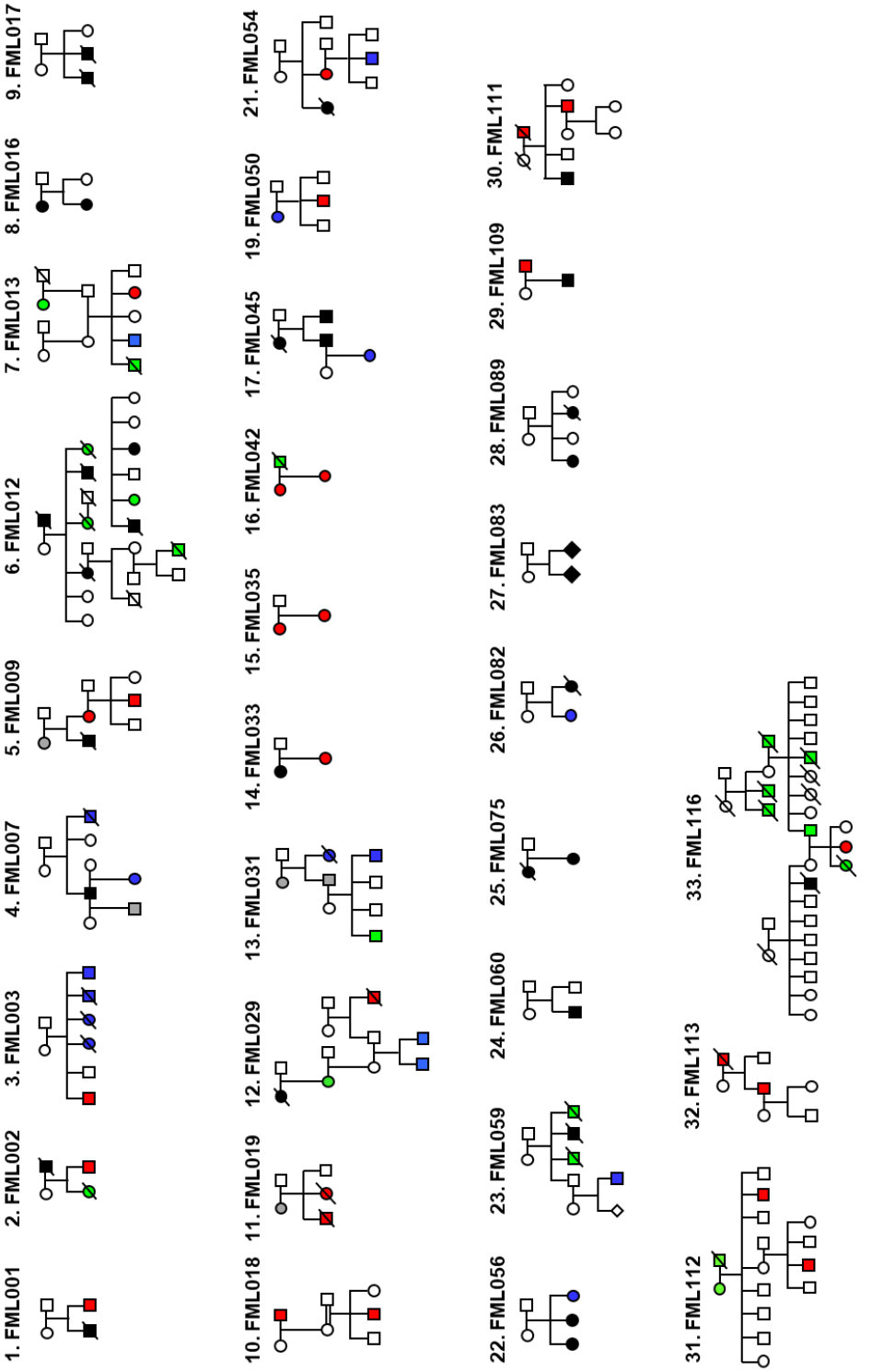


Figure 1.5. Genetically uncharacterised MDS/AML families with WES undertaken. Affected individuals are coloured as follows: red: MDS, black: AML, blue: AA, grey: thrombocytopenia, green: other leukaemia/cancer. Families FML047 and FML051 have no family tree due to lack of information.

1.6 Aims of the project

The overall aim of this project was to identify and characterise disease causing gene variants (functional variants such as missense, nonsense, deletions or insertions) that can lead to familial MDS/AML in the uncharacterised familial MDS/AML cases using whole exome and target sequencing. This includes diagnosis of uncharacterased families and/or identification of candidate genes that could be the genetic cause of the disease. This work will provide important information for the management of these families.

In addition, investigations of these rare families are likely to provide a rare opportunity to understand key pathways underlying the susceptibility and the multistep progression to MDS/AML and allow the possibility of novel strategies for the prevention and treatment of both familial and sporadic forms of MDS/AML (Hahn *et al*., 2011).

Chapter 2

Material and methods

2.1 Introduction

The materials and methods described in this chapter are relevant to the results chapters 3 to 6 presented in this thesis. All techniques are presented together here as some of them were used across all chapters.

2.2 Materials

2.2.1 Patient samples

All samples were part of our international registry of patients with bone marrow failure and related diseases. The samples were obtained with informed consent in accordance with the Declaration of Helsinki. These studies have received approval from the East London and The City Research Ethics Committee. The DNA for each sample was extracted from peripheral blood cells using a Gentra Purgene Blood Kit (Qiagen).

2.2.2 Primers

Primers were designed in the A plasmid Editor (ApE) software (Biology Labs) to have approximately 50% guanine and cytosine content and a melting temperature of 57-60°C for standard PCR. Genomic and coding DNA sequences used were obtained from NCBI - The National Center for Biotechnology Information database (https://www.ncbi.nlm.nih.gov/). All primers were synthesised by Sigma-Aldrich and

dissolved in sterile water to a stock concentration of 100µM unless stated otherwise. Primer sequences can be found in Appendix 2 (Tables A2.1 – A2.7) .

2.2.3 Chemicals and reagents

All chemicals and reagents used were of analytical or molecular biology grade and were obtained from various suppliers, as will be stated throughout the methods section.

2.3 Methods

2.3.1 DNA quantification

A spectrophotometric analysis was done in all selected samples using ND-1000 Nanodrop Spectrophotometer V3.7 (Thermo Scientific, MA, USA) to determine the average concentrations of DNA. In this spectrophometer, the sample is exposed to ultraviolet light at 260nm, and a photo-detector measures the light that passes through the sample. The more light absorbed by the sample, the higher the DNA concentration in the sample. Furthermore, the purity of DNA can be calculated by the ratio of the absorbance at 260 and 280nm ($A_{260/280}$) that is ~1.8 for pure DNA.

When it was required to detect and quantify small amounts of double stranded DNA, Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, USA) was used as in this method there is no interference from contaminants, such as salts, urea, ethanol, chloroform, detergents, proteins, and agarose. In this method, DNA concentration is measured using fluorescent dye that binds to double-stranded DNA. The fluorescence intensity is related to the amount of DNA in the sample. A standard DNA sample at 100µg/ml provided by the supplier, was diluted to 1:10 and then serially diluted to give standard curve samples of 500, 125, 31.25, 7.812 and 1.953ng/µl. PicoGreen master mix consisted of 0.3% PicoGreen dye in tris-EDTA (TE) was added to 100µl of the standard curve samples and the DNA samples being measured in wells of a 96-well plate. FLUOstar Optima microplate reader (BMG Labtech, Germany) was used to determine the fluorescence of the samples. The acquired data was then manipulated in Microsoft Excel to obtain the DNA concentrations.

2.3.2 Polymerase chain reaction – PCR

Polymerase chain reaction (PCR) was performed using DNA Tetrad 2 Peltier Thermal Cycler (Biorad, USA) as follows: 30ng genomic DNA was amplified by PCR in a total volume of 20µl containing 1x Reaction Buffer IV (200mM Tris-HCl pH 8.4, 500mM KCl**,** Thermo Scientific, USA), 2mM $MgCl_2$, 0.5µM forward and reverse primers, 2.5% dimethyl sulfoxide (DMSO), 280mM deoxynucleotide triphosphates (dNTPs) and, 0.02 units/µl of Taq DNA polymerase (Thermo Scientific) were used. A non-template control was included in each instance to exclude the possibility of contamination. The cycling parameters were: initialization step of 5 minutes at 95°C, 35 cycles of denaturation step of 30 seconds at 95°C, annealing step of 30 seconds at 58°C and an extension step of 45 seconds at 72°C and finally, a final elongation of 5 minutes at 72°C.

2.3.3 Agarose gel electrophoresis

PCR products were visualised and separated using an agarose gel electrophoresis. This consisted of 1.5% agarose in 100ml 0.5x TBE (tris-borate-EDTA) buffer and 10mg/ml ethidium bromide. Samples were loaded with a 5x loading buffer (3x TBE, 30% glycerol, 0.3% bromophenol blue) and electrophoresed at 120V in 0.5x TBE buffer for approximately 30 minutes. Gels were visualised by transillumination with ultra-violet light.

2.3.4 DNA restriction digestion

Restriction digestion of DNA fragments were performed for rolling circle amplification reaction in chapter 5. About 0.5-1µg of genomic DNA was double digested with *RsaI* and *HinfI* for detecting telomeric-circles (T-circles) in our patient samples. Enzymes and their corresponding buffers were purchased from New England Biolabs at a concentration of 20,000 units/µl. A stated DNA quantity was digested with 20,000 units of enzyme in 10x reaction buffer made to 100µl with sterilised water. Reactions were incubated at 37°C overnight followed by reaction precipitation. The pellet was precipitated by adding 10µl NaAcetate (3M, pH 5.2), 200µl 96% ethanol to the reaction, they were incubated 30 minutes at -20°C, spun 15 minutes at max speed, the pellet was then washed with 70% ethanol. The pellet was resuspended in 15.5µl distilled water after being air dried.

2.3.5 Rolling circle amplification reaction

A modified rolling circle amplification (RCA) protocol from Zellinger *et al*. (2007) was used in chapter 5 to detect T-circles in DNA from patient samples. It is a molecular technique for *in vitro* rolling nucleic acid synthesis from small single-stranded DNA minicircles using highly processive DNA polymerases. Briefly, DNA was digested by a frequently cutting restriction enzyme which digests the genomic DNA, but not telomeric sequence. T-circles then served as template for RCA using telomere-specific primers and phi29 polymerase. This leads to a synthesis of ~ 100kb long single stranded telomeric molecules. The extension products were separated from DNA by alkaline electrophoresis and detected by Southern hybridisation.

The RCA primer annealing was performed in 0.2ml PCR tubes as follows (this mix is for two parallel reactions: +/- phi29):

Template DNA:	15.5µl
10 µM Thio-$(C3TA_3)_3$:	2,5µl
10 x annealing buffer:	2µl
10 x annealing buffer:	0.2M Tris.HCl pH 7.5
	0.2M KCl
	1mM EDTA

These reactions were mixed well, spun down, and denatured for 5 minutes at 96°C in the PCR machine (DNA Tetrad 2 Peltier Thermal Cycler, Biorad, USA). Afterwards the tubes were again mixed by flicking and incubate at room temperature for 30-60 minutes.

Following primer annealing (primer sequences can be found in Table A2.1, Appendix 2), RCA reaction was carried out by first splitting the 20µl of annealed DNA into 2 tubes for +/- phi DNA polymerase reactions. A premix without enzyme were made, 9µl of the premix were added per each 10µl reaction and then either 1µl (10 units) phi29 DNA polymerase (Thermo Fisher Scientific) or water were added. The reaction tubes were vortexed, spun down and, incubated at 30°C for 16 hours. Following the RCA the tubes were heated at 65°C in a PCR machine for 20 minutes to inactivate the enzyme. The whole reaction was then loaded in a denaturing gel.

Premix:

10x phi29 buffer (provided by supplier):	2µl
2 mM dNTP mix:	2µl
distilled water:	5µl

2.3.6 Alkaline gel electrophoresis

Alkaline gel electrophoresis was carried out according to Sambrook & Russel (2006) in chapter 5 as part of the T-circle assay. A 0.8% agarose gel was made by boiling 1.2g of agarose in 150ml of distilled water. 750µl of 10M NaOH and 300µl of 0.5M EDTA was added to the melted agarose after it was cooled down to 55°C as the NaOH makes the agarose gel fragile at high temperatures. The agarose gel was then mixed well and poured into an electrophoresis tray for 30 minutes before the run. 6x alkaline loading buffer (6x gel loading buffer II from New England BioLabs, NEB) supplemented with NaOH and EDTA (100µl 6 x LB II + 3µl 10M NaOH + 1.2µl 0.5M EDTA) was mixed into the DNA samples and then they were loaded and ran at < 3.5 V/cm in 1L of fresh 1x running buffer (50mM NaOH, 1mM EDTA pH 8). The gel was ran at 25V (2 V/cm) for 16-18 hours.

2.3.7 Southern blotting

Southern blotting was performed in chapter 5 as part of the T-circle assay and it is a modified method from Ming *et al*. (1994). After the alkaline gel electrophoresis, the gel was incubated in 0.25M HCl (10.8ml concentrated HCl in 500ml distilled water) for 10 minutes, then denatured by soaking it in denaturation solution (0.5M NaOH, 1.5M NaCl) for 30 minutes and neutralised in neutralisation solution (0.5M Tris.HCl pH 7.5, 1.5M NaCl) for 30 minutes. A regular Southern blotting of the gel onto neutral nylon transfer membrane (Hybond-N, Amersham Pharmacia Biotech, UK) was carried out.

As seen in Figure 2.1, a sheet of nylon membrane was cut to the size of the gel and a plastic tray was half filled with the transfer buffer, 20x SSC (saline-sodium citrate) buffer. Then, by using a glass plate placed vertically on top of the tray a platform was made and covered with a wick made from three sheets of chromatography 3mm paper saturated in transfer buffer. The treated gel was carefully placed on the wick platform to avoid air bubble formation between the gel and the wick. The gel was surrounded with cling film to prevent the transfer buffer from being absorbed directly into the paper towels. Again, carefully placed the nylon membrane on top of the gel, avoiding air bubbles. Three sheets of chromatography 3mm paper cut to size and saturated in transfer membrane, were placed on top of the membrane, avoiding air bubbles. A stack of absorbent towels with about 10cm high were placed on top of the chromatography 3mm paper. Finally, a glass plate and a weight were placed on top of the paper stack. The transfer was performed overnight, and a mark was made on the membrane to allow identification of the tracks with a cut in the right bottom. The membrane was rinsed with 2x SSC and nucleic acid was fixed to the membrane by baking it at 80°C for 2 hours in the hybridisation oven (Amersham Life Science). The blot was then hybridised for identification of T-circle formation. The transfer pyramid was set up as shown in Figure 2.1.

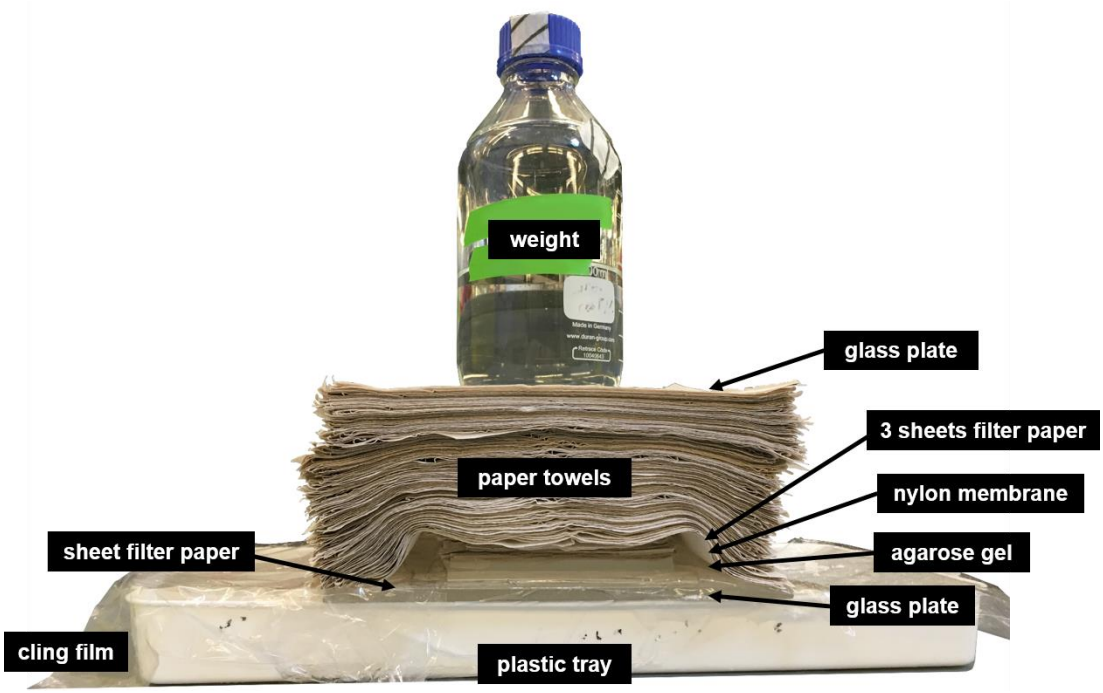


Figure 2.1. Capillary blotting apparatus.

2.3.8 Hybridisation and chemiluminescence detection

Hybridisation of Southern blot membrane was performed using TeloTAGGG Telomere Length Assay version 9.0 (Roche), a non-radioactive chemiluminescent assay to determine telomere length that uses digoxygenin (DIG) label probes. Briefly, the blot was prehybridised with DIG Easy Hyb granules provided by the supplier for 30-60 minutes at 42°C. The prehybridised solution was totally discarded, and the blot was hybridised with telomere probe overnight at 42°C with gentle agitation on Hybridiser incubator HB-1D (Techne). The blot was then washed with stringent buffer, blocked using a 1x blocking solution provided by the supplier for 30 minutes at 15-25°C with gentle agitation. The membrane was incubated with Anti-DIG-AP working solution for 30 minutes at 15-25°C with gentle agitation and then washed with 1x

washing buffer. Then the membrane was incubated with 1x detection buffer for 2-5 minutes at 15-25°C with gentle agitation. Detection buffer was discarded, and excess liquid was removed from the membrane. The wet membrane was immediately placed with DNA side facing up, on an opened hybridisation bag and very quickly ~40 drops of substrate solution were applied to it. The membrane was immediately and carefully covered with the second sheet of the hybridisation bag and it was incubated for 5 minutes at 15-25°C. After sealing the hybridisation bag's edges, the membrane was exposed to X-ray film for 20 minutes at 15-25°C.

2.3.9 Gel extraction and purification

The QIAquick Gel Extraction Kit was used for gel extraction and purification. This kit facilitates removal of nucleotides, enzymes, salts, agarose, ethidium bromide, and other impurities from DNA samples. Briefly, gel bands were removed under UV illumination and dissolved in a 3x volume of Buffer GE (agarose). DNA was precipitated with 100% isopropanol, applied to a Qiagen spin column and washed in Buffer PE and eluted in sterilised water or TE and stored at 2-8°C. Buffers GE and PE were provided by the manufacturer.

2.3.10 Sanger sequencing

Sanger sequencing was carried out by Barts and the London Genome Centre in Chaterhouse Square, London. It was used to validate results from WES and dHPLC results that were of interest. Prior to sequencing, PCR products were cleaned-up by

incubating 5µl of them with 2µl ExoSAP (0.5% exonuclease I, 10% Shrimp alkaline phosphatase) at 37°C for 15 minutes followed by 15 minutes at 80°C to deactivate the enzymes. Exo-SAP uses Exonuclease I to degrade residual single stranded primers and any extraneous single stranded DNA produced by the PCR. Shrimp alkaline phosphatase hydrolyses any remaining dNTPs from the PCR mix, which could interfere with sequencing.

Sequencing was carried out using a BigDye® Terminator Cycle Sequencing V3.1 kit (Invitrogen, USA). This kit utilises chain terminating chemistry. A mix is used which has a DNA polymerase, normal dNTPs and fluorescently labelled dideoxynucleotides (ddNTPs) which lack a 3'-OH group necessary for a phosphodiester bond to form between two nucleotides. As a result, when a ddNTP becomes incorporated into the DNA, the strand is terminated and the ddNTP can emit a fluorescent signal allowing the sequence of nucleotides to be established. The sequencing reaction consists of 1µl purified PCR product, 0.5µl Big Dye Terminator mix, 2µl 5x buffer and 0.05µM primer made to 10µl with water. This reaction mix has 25 cycles of denaturation at 96°C for 60 seconds, primer annealing for 15 seconds at 58°C and extension for 1 minute at 60°C. Following PCR a standard salt-ethanol precipitation is performed by incubating the sample with 30µl 100% ethanol and 125mM EDTA on ice for 30 minutes then spinning at 8°C for 30 minutes at 4000rpm. Samples were dried and washed in 70% ethanol before spinning again at 4000rpm for 10 minutes. Samples were dried before being resuspended in formamide. They were heated to 95°C for 3 minutes to denature the DNA and the presence of formamide helps to maintain this state. Samples were then run on a 3130xl Genetic Analyser (Applied Biosystems, UK).

Occasionally, some PCR products were sent directly to Barts and the London Genome Centre for clean-up and sequencing with 10pmol/ml of an appropriate primer.

2.3.11 Analysis of sequencing traces

Chromas Lite (Technelysium) was used to examine quality and observe variants in the sequence traces (Figure 2.2). The sequence from traces were aligned using the online Basic Local Alignment Research Tool (BLAST, NCBI) to the human genome reference sequence hg19. References sequences are obtained from NCBI database (http://www.ncbi.nlm.nih.gov) and stored in the A plasmid Editor (ApE) software (Biology Labs, USA).

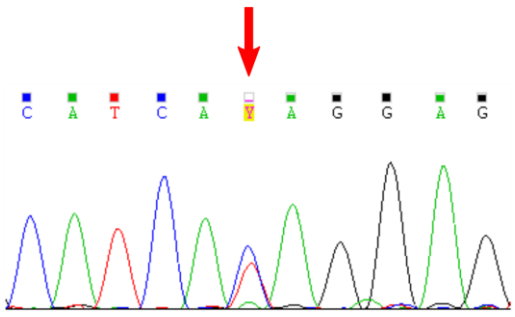


Figure 2.2. Sanger sequencing trace view in Chromas Lite. Variant is highlighted by the red arrow.

2.3.12 RNA extraction

RNA was extracted from primary and WI38-VA13 (foetal lung fibroblasts) cells (chapter 4). Cells were lysed using 1ml TRIzol Reagent (ThermoFisher Scientific). TRIzol™ Reagent is a monophasic solution of phenol, guanidine isothiocyanate, and other proprietary components which facilitate the isolation of RNA. After lysis, 0.2ml chloroform was added and samples centrifuged at 12,000rpm for 15m minutes at 4°C. This separates RNA into an upper aqueous layer, distinct from the red lower organic layer containing proteins and DNA. The clear aqueous phase was transferred to a fresh tube and 0.5ml 100% isopropanol and the RNA was precipitated, incubating at room temperature for 10 minutes then spinning at 4°C for 10 minutes to pellet the RNA. The pellet was washed in 75% ethanol to remove contaminants and air dried before resuspending in water. RNA was stored at 20°C.

2.3.13 Reverse Transcriptase PCR

cDNA was synthesised using reverse transcriptase PCR, where random hexamers and Superscript III reverse transcriptase (Invitrogen) were used according to the manufacturer's protocol. Briefly, up to 5µl RNA was incubated with 50ng of random hexamers (Invitrogen), 1µl 10mM dNTPs and sterile water to 10µl at 65°C for 5 minutes then on ice for 1 minute to denature the RNA. The cDNA was then synthesised using a mix of 10x RT buffer, 25mM $MgCl_2$, 0.1M dithiothreitol (DTT), 40 units/µl RNaseOUT and 200 units/µl SuperScript III Reverse Transcriptase enzyme (Invitrogen). Samples were incubated as follows: 10 minutes at 25°C, 50 minutes at

50°C, 5 minutes at 85°C before chilling on ice. The cDNA was treated with RNase H (Invitrogen) for 20 minutes at 37°C to remove RNA and stored at -20°C.

2.3.14 Monochrome multiplex quantitative PCR

In chapters 4 and 5, patient's telomere length was relatively compared to standard DNA in peripheral blood cells using DNA Monochrome multiplex quantitative PCR (MMqPCR) described by Cawthon (2009). This method compares the ratio of a telomeric PCR product (T) to a single-copy gene (S, T/S ratio). Genomic DNA samples were diluted to 50ng/µl in TE and allowed to equilibrate overnight. Samples were then diluted to 2ng/µl in distilled water. The reference DNA used for the standard curve was diluted to 15ng/µl and then serially diluted to 5.00, 1.67, 0.56 and 1.85ng/µl samples for a five-point standard curve. For each sample the final input of genomic DNA was 10ng and for the standard curve this was 75, 25, 8.33, 2.78 and 0.925ng. 10µl of master mix was added to the DNA, this mix consisted of Roche SYBR Green Master I reaction mix (Roche Applied Science), telomeric primers at 30µlM each and S primers at 6nM each. Therefore, in the same reaction, telomeric primers amplified a 79bp telomeric amplicon and S primers amplified a single copy gene amplicon (human beta-globin, see Table A2.2, Appendix 2). The cycling parameters were: stage 1 consists of an initialization step of 15 minutes at 95°C, in stage 2, two cycles of denaturation step of 15 minutes at 94°C, annealing step of 15 minutes at 49°C and in stage 3, 40 cycles of denaturation step of 15 minutes at 94°C, annealing step of 15 minutes at 62°C and an extension step of 15 minutes at 74°C with fluorescent signal acquisition for conventional qPCR. These are followed by two steps: an incubation at 84°C for 15s and an incubation at 88°C for 15s with a second fluorescent signal

acquisition. The acquired data was then manipulated in Microsoft Excel to obtain the patient's telomere length.

2.3.15 Denaturing high-performance liquid chromatography

Denaturing high-performance liquid chromatography (dHPLC) was used to scan variants in known disease causing genes and candidate genes in chapters 3 and 6. DNA fragments generated by standard PCRs (primer sequences can be seen in Tables A2.3 – A2.6, Appendix 2) were partially denatured at an ideal temperature in the range of 50–70°C, mixed pair wise, and scanned for variants by denaturing high-performance liquid chromatography on a Transgenomic Wave DNA fragment analysis system (Glasgow, UK). Any fragments showing abnormal elution patterns were re-amplified and the variant confirmed by forward and reverse Sanger sequencing.

This technique employs the DNASep® cartridge that contains polymeric sorbent with alkylated nonporous polystyrene-divinylbenzene (PS-DVB) copolymer microspheres for high performance nucleic acid separations. Homoduplexes and heteroduplexes are formed from two amplified DNA fragments with different nucleotide sequences after denaturing and gradual annealing processes. More than one peak are present on the chromatogram because heteroduplexes and homoduplexes have different retention times due to less helical fraction in heteroduplexes (Figure 2.3) (Oefner & Huber, 2002).

dHPLC method is capable of analysing accurately nucleic acid in a high resolution and is sufficiently sensitive for the reliable detection of nearly 100% of DNA sequence variations at optimized partially denaturing temperatures (Xiao & Oefner, 2001).

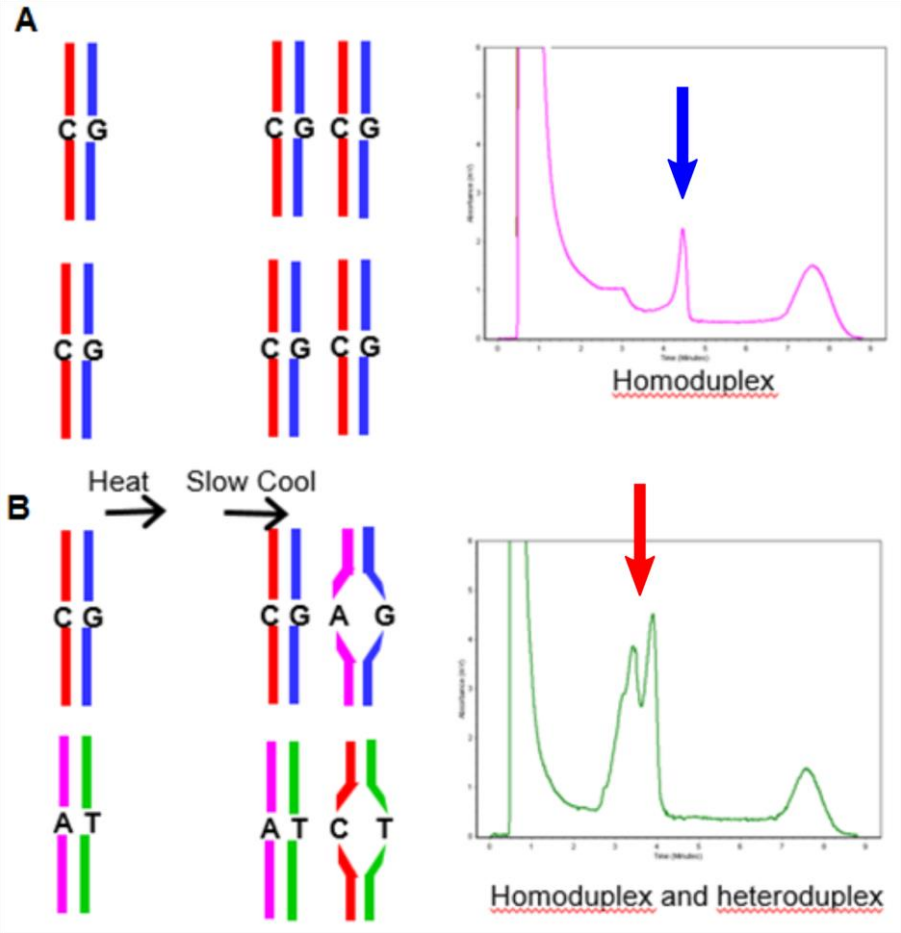


Figure 2.3. Denaturing high pressure liquid chromatography. This method can reveal single nucleotide variation through comparison of two DNA fragments by denaturing and reannealing the samples, then detecting the resulting duplex products. A. Reannealing of two identical DNA fragments forming a homoduplex chromatography elution pattern shown by the blue arrow. B. Reannealing of two different DNA fragments forming a heteroduplex chromatography elution pattern shown by the red arrow.

2.3.16 Whole exome sequencing

Whole exome sequencing (WES) was undertaken on a total of 42 uncharacterised families with familial MDS/AML to identify disease causing genes during the course of this project. Of these, 8 families were processed by myself using Nextera Rapid Capture Exome and Expanded Exome Enrichment kit to prepare the libraries and the remaining samples were done by Dr Amanda Walne (Centre for Genomics and Child Health, Blizard Institute). WES was undertaken on multiple unrelated individuals from these families and when available, distantly affected individuals of the same family were also sequenced. Once a variant in a candidate gene of interest was found, dHPLC was used to screen additional families from our cohort for that specific gene. Abnormal patterns of elution were validated through Sanger sequencing.

2.3.17 Nextera library preparation (enrichment of exonic fragments)

A library of DNA was built and amplified (library enrichment), using the Nextera rapid capture enrichment kit (Illumina, UK). DNA was fragmented and adapter sequences (specific oligonucleotides) were added onto the ends of those fragments. Once constructed, these libraries were clonally amplified in preparation for sequencing (Metzker, 2010; Grada & Weinbrecht, 2013).

The Nextera library preparation kit uses modified transposons for cleaving and adding adapter sequences onto DNA template to generate indexed sequencing libraries that can be carried through enrichment for targeted sequencing applications (Figures 2.4 – 2.8).

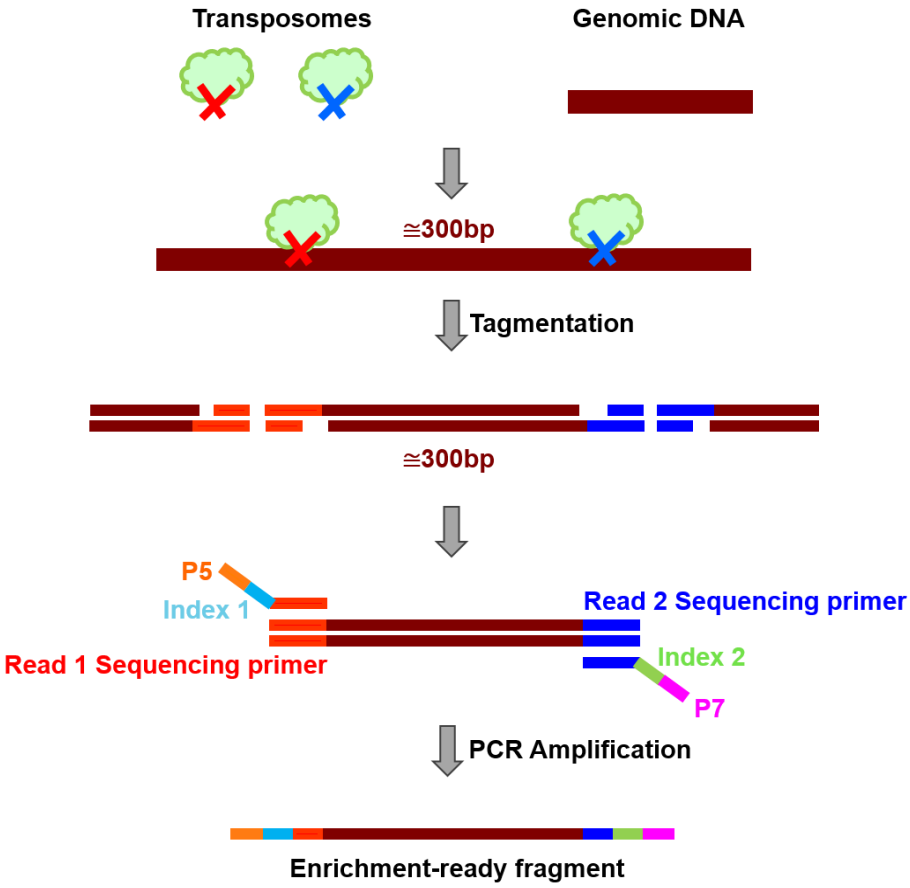


Figure 2.4. Sample preparation. P5 and P7: adaptors for cluster generation generated for sequencing; Index 1 and Index 2: adaptors for identification of the DNA library after sequencing. Schematic illustration of the steps of sample preparation for whole exome sequencing using Illumina Nextera Rapid capture Exome Kit (adapted from http://www.gtbiotech.com.tw).

The success of Nextera enrichment depends on using an accurately quantified amount of input DNA, adding too much of material can lead to undertagmentation and likewise adding to little can lead to overtagmentation. To achieve consistent DNA quantification is important to avoid UV absorbance methods. It is recommended that the Nextera library prep to have an optimal size range within 200bp to 1kb. Larger fragments cluster inefficiently in the flow cell leading to lower than expected output. In

addition to excess DNA input, enzymatic inhibitors such as EDTA, detergents, proteins and phenol can also lead to undertagmentation. When little DNA input is added to the prep reaction, an overtagmentation can occur and it means that the DNA fragment is over cleaved. The clean up step is designed to remove fragments smaller than 200bp, so shorter DNA fragment will be removed resulting in lower unknown yield. Therefore, picogreen fluorescent assay was used for the genomic DNA quantification.

2.3.18 Tagmentation of genomic DNA and first PCR amplification

50ng of genomic DNA with a concentration of 5ng/µl was tagged and fragmented by Nextera transposome. The Nextera transposome simultaneously fragmented the genomic DNA and added adapter sequences to the ends allowing amplification by PCR. Following clean up, the purified tagmented DNA was amplified by a short PCR program of 10 cycles (Figure 2.4). The cycling program was as follows: 72°C for 3 minutes, 98°C for 30s, 10 cycles of denaturation step of 10s at 98°C, annealing step of 30s at 60°C and an extension step of 30s at 72°C, a final elongation of 5 minute at 72°C and finally, held at 10°C.

Indexes needed for sequencing (indexes 1 and 2) as well as common adaptors required for cluster generation (P5 and P7) were added during the tagmented DNA amplification (Figure 2.4).

2.3.19 First hybridisation and first capture

After the PCR clean-up, the DNA library was denatured, and the first hybridisation was performed (Figure 2.5 and 2.6). In this process the denatured DNA library was mixed with capture probes to targeted regions of interest. Following the binding of the probes to their specific region (Figure 2.6), streptavidin beads captured the probes hybridised to the targeted regions of interest (Figure 2.7). Two heated washes removed non-specific binding from the beads. The enriched library was then eluted from the beads (Figure2.8) and prepared for a second round of hybridisation, then second capture and finally second PCR amplification.

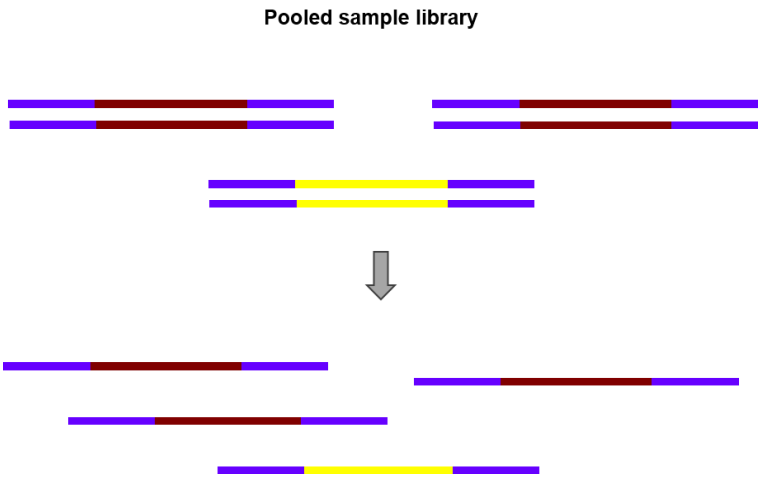


Figure 2.5. Denaturation of double-stranded DNA library. Adaptors and indexes not shown, see Figure 2.4. Schematic illustration of the steps of sample preparation for whole exome sequencing using Illumina Nextera Rapid capture Exome Kit (adapted from http://www.gtbiotech.com.tw).

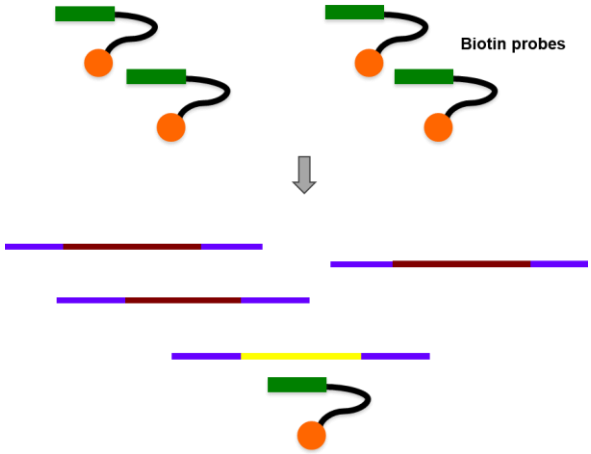


Figure 2.6. Hybridisation of biotinylated probes to targeted regions. Schematic illustration of the steps of sample preparation for whole exome sequencing using Illumina Nextera Rapid Capture Exome Kit (adapted from http://www.gtbiotech.com.tw).

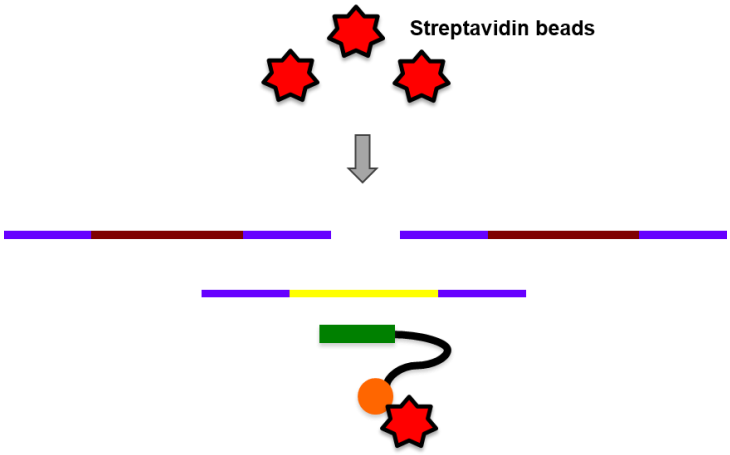


Figure 2.7. Enrichment using streptavidin beads. Schematic illustration of the steps of sample preparation for whole exome sequencing using Illumina Nextera Rapid capture Exome Kit (adapted from http://www.gtbiotech.com.tw).

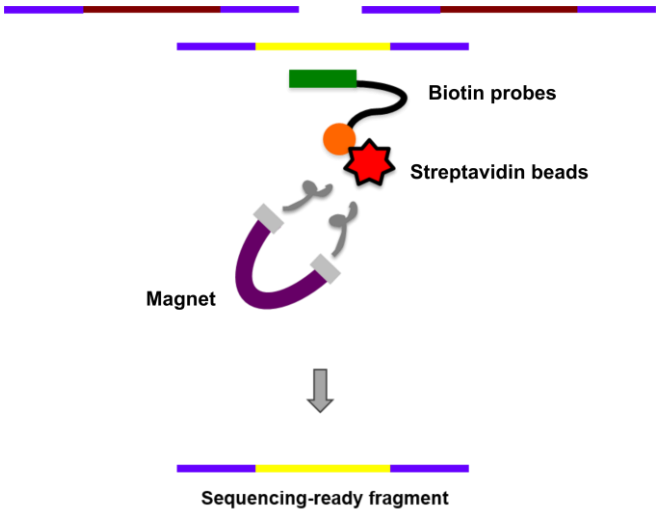


Figure 2.8. Elution of the enriched library from beads. Schematic illustration of the steps of sample preparation for whole exome sequencing using Illumina Nextera Rapid capture Exome Kit (adapted from http://www.gtbiotech.com.tw).

2.3.20 Targeted resequencing – design of probes

The targeted resequencing in chapters 4 and 5 was performed using the Illumina Truseq Custom Amplicon (TSCA) kit by Birmingham Women's NHS Foundation Trust in Birmingham, UK. TSCA permits to focus on genomic regions of interest through sequencing of customised amplicons. The Illumina Design Studio software (https://designstudio.illumina.com) was used to design the TSCA oligonucleotide probes. When the coordinates of the genomic region of interest is logged in this software, it is divided into ~250bp target regions and the probe design is automatically performed by using an algorithm that considers a range of factors, including GC content, specificity, probe interaction, and coverage. Probes are given estimated success scores and it can be filtered with user-defined tags, and then added to, or removed from, the design to improve coverage (the gene panel can be seen in Table

A3.1, Appendix 3).

2.3.21 Targeted resequencing library preparation

Libraries containing the target regions were prepared for sequencing for each DNA sample and subjected to a clean-up and normalisation process following the protocol of the Illumina TSCA kit. The first step for library prep hybridisation of the costumed probes was to add 150ng of genomic DNA input per sample, 5µl of the customised probes and a hybridisation buffer provided by the manufacturer. These samples were incubated at 95°C for 1 minute, cooled slowly for 80 minutes to 40°C to denature double stranded DNA and allow the binding of the probes to their specific DNA fragment (Figure 2.9). The second step is the extension/ligation, in which a DNA polymerase adds bases from probe 1 and then extends all the way down the downstream probe. A ligase comes in and finishes the job by ligating the extended piece to the downstream probe. This fills the gap between the two probes, synthesising a strand complementary to the genomic DNA. Samples were incubated for 45 minutes at 37°C with an extension ligation mix provided by the manufacturer for this extension/ligation to happen. The amplicons are ready after the ligase fills the gap.

The last step is PCR, when the indexes and the oligonucleotides complementary to the flow cell are added and finally the whole fragment is amplified. The probes have complementary sequences to the genomic DNA allowing them to bind either side of their target and also have a primer binding region (Figure 2.9).

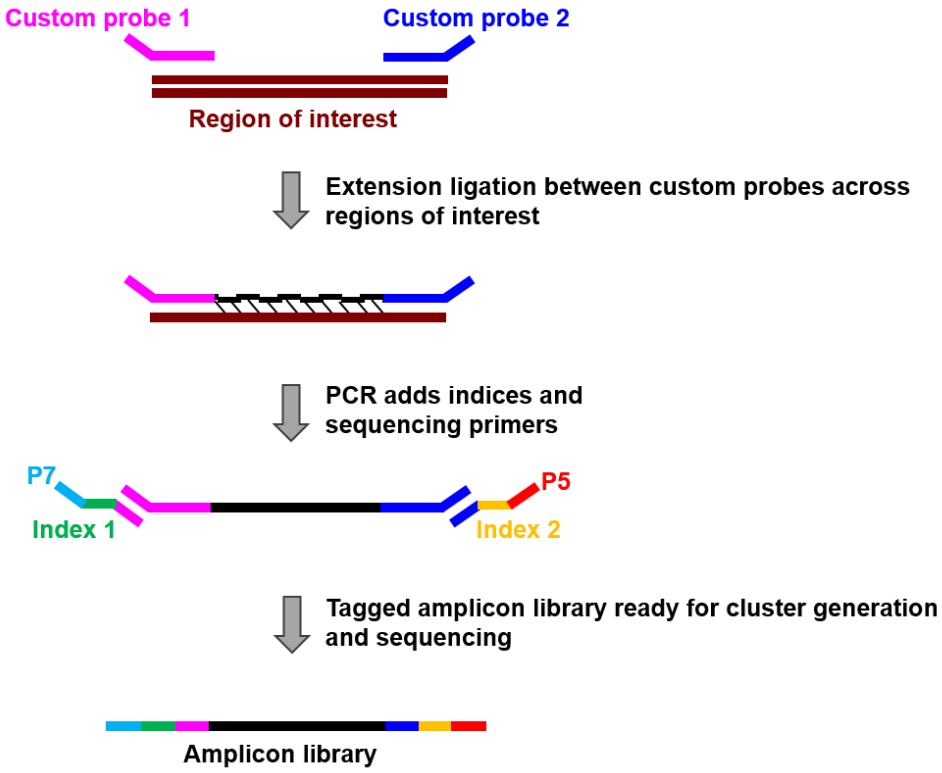


Figure 2.9. Schematic of TSCA library preparation. Adapted from https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_truseq_custom_amplicon.pdf.

A PCR reaction is carried out after the samples are denatured using 50mM NaOH, this reaction amplifies the single stranded products using primers complementary to the probes. Six base pair *index* sequence is used to tag the amplicons of an individual and it is incorporated into the primers which also contain a sequence complementary to an oligonucleotide on the flow cell for sequencing. This is similar to the adapters P5 and P7 seen in Figures 2.4 and 2.9, used in the preparation of samples for exome sequencing, as described in section 2.3.18. Extension/ligation product, PCR master mix, primers and TruSeq DNA polymerase provided with the Illumina kit were used for the PCR reaction. This PCR amplification was carried out under the following

conditions: initialization step of 3 minutes at 95°C, 23 cycles of denaturation step of 30 seconds at 95°C, annealing step of 30 seconds at 66°C and an extension step of 60 seconds at 72°C and finally, a final elongation of 5 minutes at 72°C.

Magnetic AMPure beads were used to enrich the amplicon library. They bind to DNA fragments of a preferred size (≥200bp), this is based on the ratio of DNA to beads (1:1). Samples were cleaned up from small unbound fragments of DNA by placing them on a magnetic stand. The amplicon library was removed from the beads with an elution buffer provided by the manufacturer. Each library has to have an equal quantity of DNA to ensure a good quality of the sequencing, for that libraries for each individual were normalised using a magnetic library beads provided by the manufacturer as they bind to a specific number of DNA. The beads were mixed and added to each DNA library and incubated for 30 minutes on a microplate shaker at 1800rpm. During this process the beads bound to DNA leaving excess in solution. Samples were placed on a magnetic stand and the excess DNA washed away. The library bound to the beads was removed using 0.1N NaOH freshly made. This solution denatures the DNA resulting in single stranded libraries which were mixed with a storage buffer provided by Illumina.

The libraries for each individual were pooled into a single tube. The amplicons of each individual are distinguishable when pooled by the indices that were added to each amplicon in the PCR amplification step. The pooled library was denatured by incubating at 96°C for 2 minutes to break any bonds within and between fragments, removing hairpin structures to ensure the DNA remained single stranded and linearized. It was then immediately placed in an ice-water bath for 5 minutes. The prepared single stranded library was then ready for sequencing.

2.3.22 Sequencing of exome library

Nextera Rapid Capture Kit was used for the preparation of exome libraries that were sequenced on the NextSeq 550 System at University College London while TSCA library was sequenced by Birmingham Women's NHS Foundation Trust in Birmingham, UK.

Sequencing occurs on a flow cell which is a glass slide with lanes and clustering is the first step in which each fragment is isothermally amplified. Each lane has a channel coated with a lawn, composed of two types of oligos (shown as P5 and P7 in Figures 2.4 and 2.9). These are complementary to the adapters ligated to the ends of each DNA fragment during sample preparation (Figure 2.10). Single stranded DNA fragment can therefore hybridise to one of the oligos when the prepared samples are added to the flow cell. A complementary strand is synthesised by a polymerase enzyme and then the double stranded fragment is denatured, and the original template is washed away.

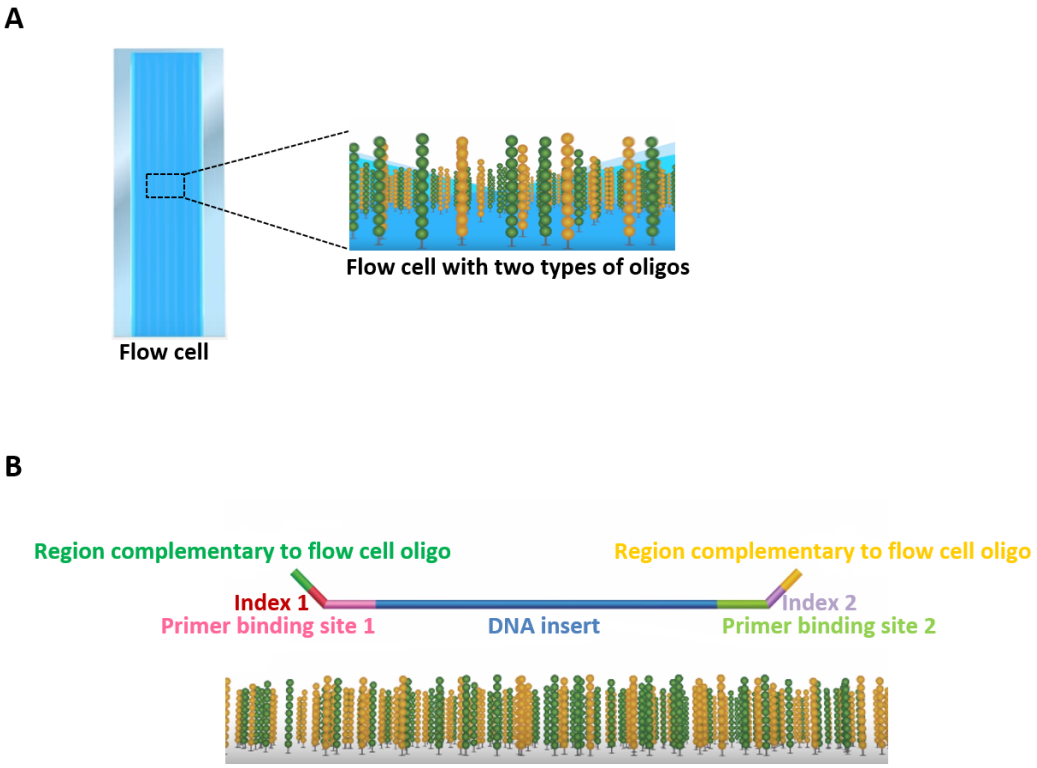


Figure 2.10. Schematic illustration of the cell flow and DNA fragment prepared for sequencing. Adapters such as indices and regions complementary to the cell flow oligos were added to the DNA fragment. Adapted from Illumina Sequencing Technology – https://www.youtube.com/watch?v=womKfikWIxM.

The strands are clonally amplified through bridge amplification. In this process the strand folds over and the adapter region hybridises to the second type of oligo on the flow cell. This forms a single stranded bridge as polymerases generate the compliment strand. This is denatured to give two single stranded fragments which can form separate bridges and undergo bridge amplification for millions of time, resulting in clonal amplification of all the fragments (Figure 2.11). The clusters were sequenced using sequencing-by-synthesis, a method based on reversible dye-terminators that enable the identification of single bases as they are introduced into DNA strands.

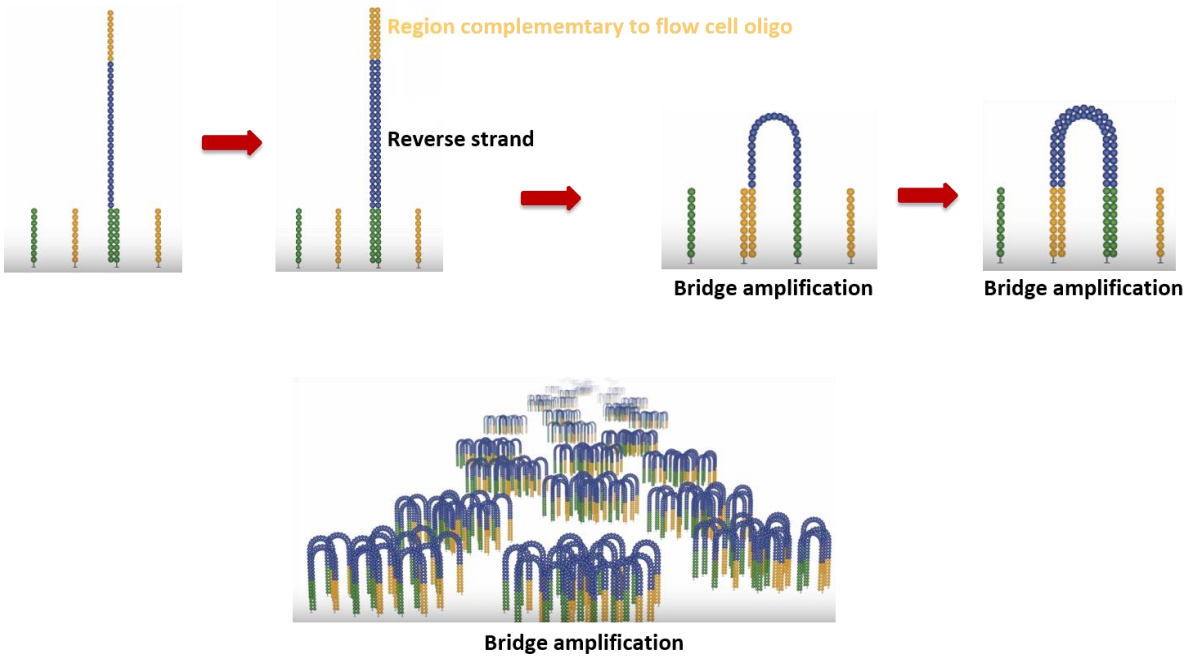


Figure 2.11. Schematic illustration of bridge amplification and generation of clonal clusters of the DNA fragments. Adapted from Illumina Sequencing Technology - https://www.youtube.com/watch?v=womKfikWlxM.

2.3.23 Variant calling

The raw data generated from NGS is processed to remove adapter sequences and low-quality reads and then it is aligned to a reference sequence or assembled *de novo* (Pop & Salzberg, 2008) as described below.

Variant calling of exome data was performed by our collaborator Dr Vincent Plagnol (Genetics Institute, University College London). WES reads were de-multiplexed and Novoalign version 2.08.03 (Novocraft Technologies) was used to align the raw fastq files to the GRCh37 reference genome (Figure 2.12).

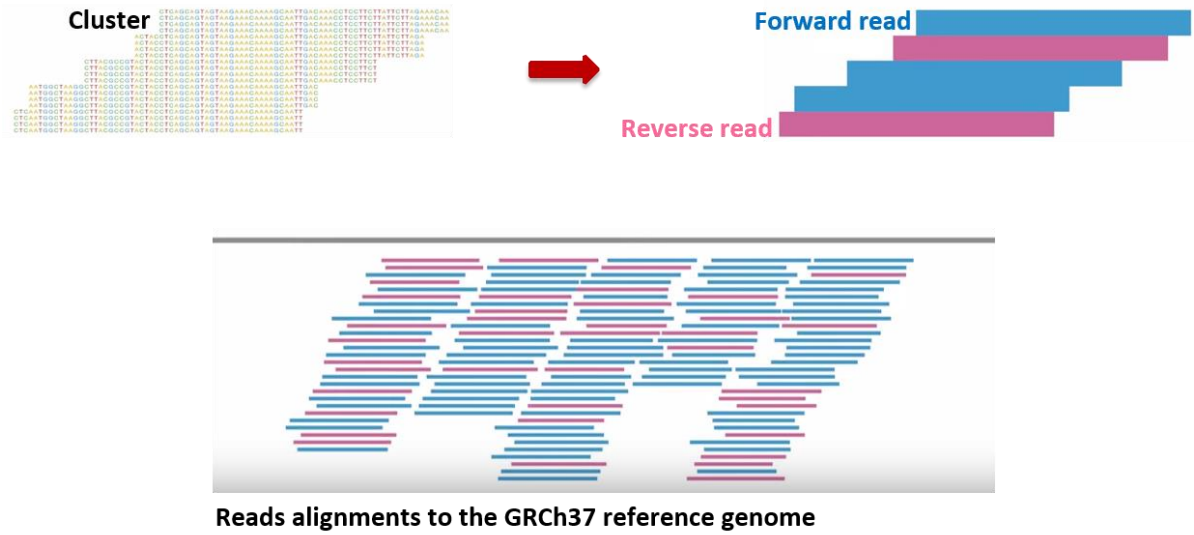


Figure 2.12. Schematic illustration of data analysis. Sequences of pooled sample libraries are separated by the unique indices introduced in the sample preparation. Reads for each sample are locally clustered and forward and reverse reads are paired creating contiguous sequences. These contiguous sequences are aligned to the reference genome for variant identification. This figure has been adapted from Illumina Sequencing Technology (online video https://www.youtube.com/watch?v=womKfikWIxM).

After binary alignment map (BAM) files were generated, duplicate reads were removed using Picard tools MarkDuplicates. Each sample had a genomic variant call format (gVCF) file containing the variant calling that was performed using the haplotype caller module of Genome Analysis Toolkit (GATK- https://www.broadinstitute.org/gatk, version 3.1-1). The individual patient gVCF files, in combination with UCL-exomes consortium that covers 2,500 clinical exomes, were combined into merged VCF files for each chromosome containing on average 100 samples each. The final variant calling was performed using the GATK Genotype GVCFs module jointly for all samples (cases and controls). Variants quality scores were then re-calibrated according to GATK best practices separately for indels and single nucleotide polymorphisms (SNPs). Resulting variants were annotated using

ANNOVAR based on Ensembl data. Candidate variants were filtered based on function (non-synonymous, presumed loss-of-function or splicing) and minor allele frequency of < 0.5% in our internal control group, as well as the National Heart, Lung, and Blood Institute (NHLBI - https://www.nhlbi.nih.gov/) exome sequencing dataset. Splice variants were flagged within 5bp of the exon-intron junction. All relevant variants identified were validated by Sanger sequencing on a 3130xl Genetic Analyzer with a BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems) (Tummala *et al*., 2015).

2.3.24 Mammalian cell culture

Primary and WI38-VA13 cell lines have been used in to make cDNA in chapters 4 and 5. These lines are adherent and were cultured in Dulbecco's Modified Eagle Medium, high glucose ClutaMAX (DMEM, ThermoFisher Scientific) which was supplemented with penicillin (100units/ml) streptomycin (100ug/ml) and 10% foetal calf serum (Life Technologies). Cells were cultured in humidified conditions at 37°C with 5% CO_2. Cells were passaged 1 in 10 when they reached 90% confluence by removing growth media, washing in 1x phosphate buffered saline (PBS) and applying 0.25% Trypsin-EDTA for ~3 minutes. This allows cells to detach from the flask. DMEM was added to neutralise the reaction and the cells were pelleted, washed in PBS and seeded as required. Sterile technique was practised at all times.

2.3.25 Polyacrylamide gel electrophoresis and Western blotting

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting were used to visualise specific proteins from lysates in chapter 5. Lysates were denatured in a sample loading buffer (2% SDS, 5% DTT, 0.002% bromophenol blue, 10% glycerol and 62.5mM Tris-HCl, pH 6.8) at 90°C for 5 minutes and loaded onto a 4-12% bis-tris polyacrylamide gradient gel (Invitrogen) alongside a 10-250kDa protein size marker (Kaleidoscope Pre-stained Protein Standard, BioRad). Samples were run at 150V for ~90 minutes in Invitrogen NuPAGE MES SDS Running buffer (50mM MES - 2-ethanesulfonic acid, 50mM Tris base, 0.1% SDS, 1mM EDTA, pH 7.3) to separate proteins by size. Western blotting was performed using a WesternBreeze Chemiluminescent Kit (Invitrogen). Proteins were electrophoretically transferred onto a polyvinylidene difluoride (PVDF) membrane which was blocked for 30 minutes in 10% bovine serum albumin (BSA) – based in tris-buffered saline supplemented with Tween-20 (TBS-T). The membrane was washed thoroughly before incubating with a primary antibody (against the protein of interest) in 3% BSA in TBS-T overnight at 4°C with gentle shaking. The primary antibody was removed, the membrane was washed in 1x PBS and the protein visualised using a secondary antibody (which targets the primary) conjugated to alkaline phosphatase. The chemiluminescent agent CDP-Star was applied which is dephosphorylated by alkaline phosphatase at the site of the protein on the membrane. This forms a metastable intermediate molecule that emits light as it decomposes. This is detected using X-ray film, thus indicating the presence of the protein under investigation.

2.3.26 *In silico* analyses

Throughout chapters 3 to 5, *in silico* analyses were performed to investigate genetic variants and predict their biological impact. Eight databases were searched for specific information including: National Centre for Biotechnology Information - NCBI (https://www.ncbi.nlm.nih.gov/) and Ensembl (https://www.ensembl.org/) where the reference coding DNA and genomic DNA sequences were obtained from; Genome Bioinformatics Site (https://genome.ucsc.edu/) from University of California Santa Cruz (UCSC) was used to view DNA sequence and exon locations; The Exome Aggregation Consortium - ExAC (http://exac.broadinstitute.org/) was consulted to verify the minor allele frequency of relevant variants; Polyphen-2 (Adzhubei *et al*., 2013; http://genetics.bwh.harvard.edu) and MutationTaster2 (Schwarz *et al*., 2014; http://www.mutationtaster.org/) web-based softwares were used to evaluate the pathogenicity of a variant; ClustalW2 Multiple Sequence Alignment program (https://www.ebi.ac.uk/Tools/msa/clustalw2/) was used to investigate the conservation of amino acid residues and finally, UniProt (http://www.uniprot.org/) was used to view protein domains and motifs.

Chapter 3

Variants identified in known disease genes associated with familial MDS/AML and related disorders

3.1 Introduction

Whole exome sequencing is a diagnostic approach for the identification of molecular defects in patients with suspected inherited disorders and a powerful resource for scientific discoveries. In this chapter we describe the families with MDS/AML from our cohort that were characterised during the course of these studies with variants in the known disease genes associated with familial MDS/AML and related disorders, including Shwachman-Diamond syndrome, Fanconi anaemia and Wiskott-Aldrich syndrome. However, many families in our cohort remain uncharacterised. This highlights the marked genetic heterogeneity of familial MDS/AML and suggests that further unknown genes are involved in the development of MDS/AML. Future studies will be necessary to determine the full spectrum of this genetic heterogeneity, reveal new insights to the biology of these diseases and help guide clinical management.

3.1.1 Genetically uncharacterised familial MDS/AML patients

At the beginning of these studies in 2014, seven genes (*RUNX1*, *CEBPA*, *TERC*, *TERT*, *GATA2*, *SRP72* and *ANKRD26*) had been identified as predisposing to familial MDS/AML. Two other genes (*DDX41* and *ETV6*) also associated with this disease were identified in 2015. Since 2014, nine MDS/AML families were genetically characterised using WES in this study. Of these, seven will be discussed in this chapter and the remaining two families will be discussed in chapter 4.

3.2 Results

Rare germline variants likely to be damaging were identified in the known familial MDS/AML causing genes *RUNX1*, *TERT*, and *GATA2* and also in genes known to cause blood related disorders including *FANCA*, *SBDS*, and *WAS*, in a total of seven families with MDS/AML (Figure 3.1). Segregation analysis were performed when DNA samples of the family members were available. The results are described below.

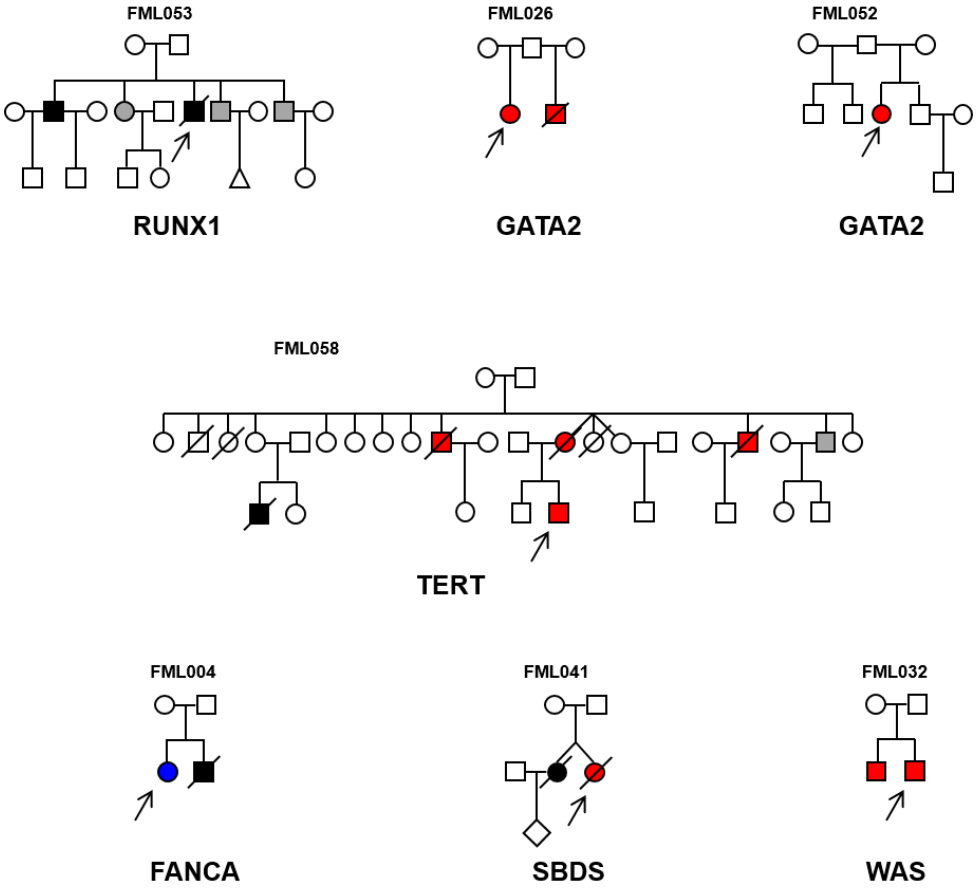


Figure 3.1. Seven characterised families with MDS/AML from our cohort. The altered disease causing gene is indicated below each family. Affected individuals are coloured as following black: AML, red: AML, blue: AA and grey: thrombocytopenia. The arrow highlights the index case.

3.2.1 Variants identified in previously known familial MDS/AML causing genes

3.2.1.1 Variants identified in *RUNX1*

The core binding factor (CBF) is a transcriptional regulator complex composed of α and β subunits. Runt related transcription factor 1 (*RUNX1*), runt related transcription factor 2 (*RUNX2*) and runt related transcription factor 3 (*RUNX3*) code for the α subunits while *CBFβ* codes for the β subunits. The α subunits bind directly to the DNA in the regulatory regions of their targets and the β subunits stabilize the RUNX-DNA complex by interacting with the α subunits (Huang *et al*., 2001). Germline heterozygous variants, including nonsense, missense, deletion, frameshift throughout the gene and gene deletion or duplication in *RUNX1* cause familial platelet disorder with propensity to myeloid malignancies (MDS, AML and T-cell ALL). The proposed mechanisms that underlie progression to leukaemia in *RUNX1* are haploinsufficiency for tumour suppression, dominant-negative effects on normal *RUNX1* function, acquisition of *de novo* variant in the wild type allele, and acquisition of cooperating variants (Jongmans *et al*., 2010). This disorder is an autosomal dominant syndrome characterised by thrombocytopenia and platelet abnormalities that typically can develop into AML (FPD/AML) (Owen *et al*., 2008). In addition, it is also known that chromosomal translocations and somatic variants in *RUNX1* have been associated with several types of sporadic leukaemia and MDS. RUNX1 protein (Figure 3.2A) is a transcription factor that regulates enhancers and promoters of many haematopoietic genes and it is involved in the development of normal haematopoiesis (Hamadou *et al*., 2016).

Here we describe a germline deletion of 162kb in *RUNX1* in the index case (II:3, Figure 3.2B) of family FML053 and his brothers (II:1 and II:5, Figure 3.2B). The index case (II:3, Figure 3.2B) presented with T-cell ALL when he was 10 years old and died at age 22 a year after the progression of the disease into AML at age 21 years. All siblings had altered platelet aggregation. His elder brother (II:1, Figure 3.2B) had AML at age 10 years and his youngest brother (II:5, Figure 3.2B) had mild thrombocytopenia. Comparative genomic hybridization-array (CGH-array) analysis performed by another group from Hospital das Clínicas da Faculdade de Medicina de São Paulo (Brazil) where these samples came from, revealed the deletion in *RUNX1*, chromosome 21q22.12 in Human Genome Issue hg18 (chr21:35,197,917-35,360,669del). This deletion of 162,752bp of chromosome 21 (NCBI36/hg18) affected exons 1 and 2 of *RUNX1*, outside the cluster region of the Runt Homology domain and the transactivation domain (see in Appendix 4, Figure A4.1) and the region of 1,097,908bp just before *RUNX1*. This region included NR_073512 (RUNX1 processed transcript that do not contain an open reading frame – Ensemble - https://www.ensembl.org) and the SET domain-containing protein 4 (SETD4) located in the Down syndrome critical region (Jiang *et al*., 2015). *SETD4* encodes a histone lysine methyltransferase which is involved in breast carcinogenesis. *SETD4* knockdown in breast cancer cell lines significantly suppressed their proliferation and delayed the G1/S cell cycle transition without affecting apoptosis (Faria *et al.*, 2013). To determine the breakpoint location of this deletion in *RUNX1*, several combinations of forward and reverse primers located on either side of the deletion were made with no success (none PCR product was generated).

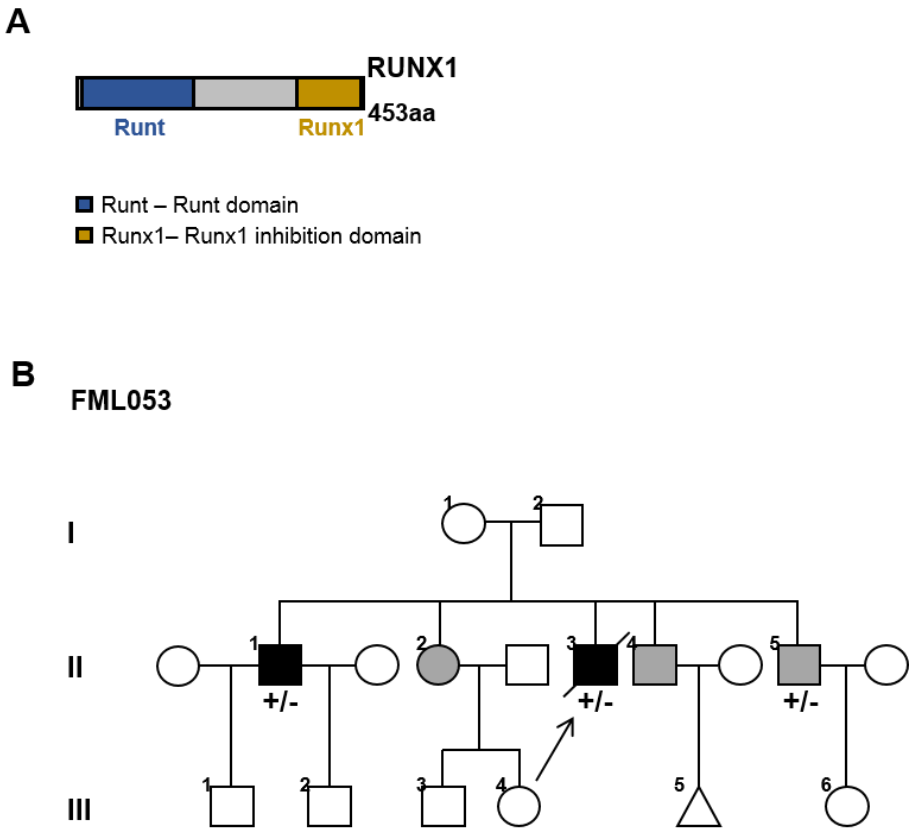


Figure 3.2. Characterisation of the *RUNX1* germline variant in family FML053. A. Schematic of RUNX1 protein (XP_011528068.1). B. Affected individuals are coloured as following black: AML and grey: thrombocytopenia. The arrow highlights the index case. The heterozygous genotyping is denoted as (+/-).

Of additional interest, a frameshift deletion of 21bp in the domain region of *RUNX1* (c.233-254del) was identified in the index case (II:1, Figure 3.3A) of family FML007 (Figure 3.3A).

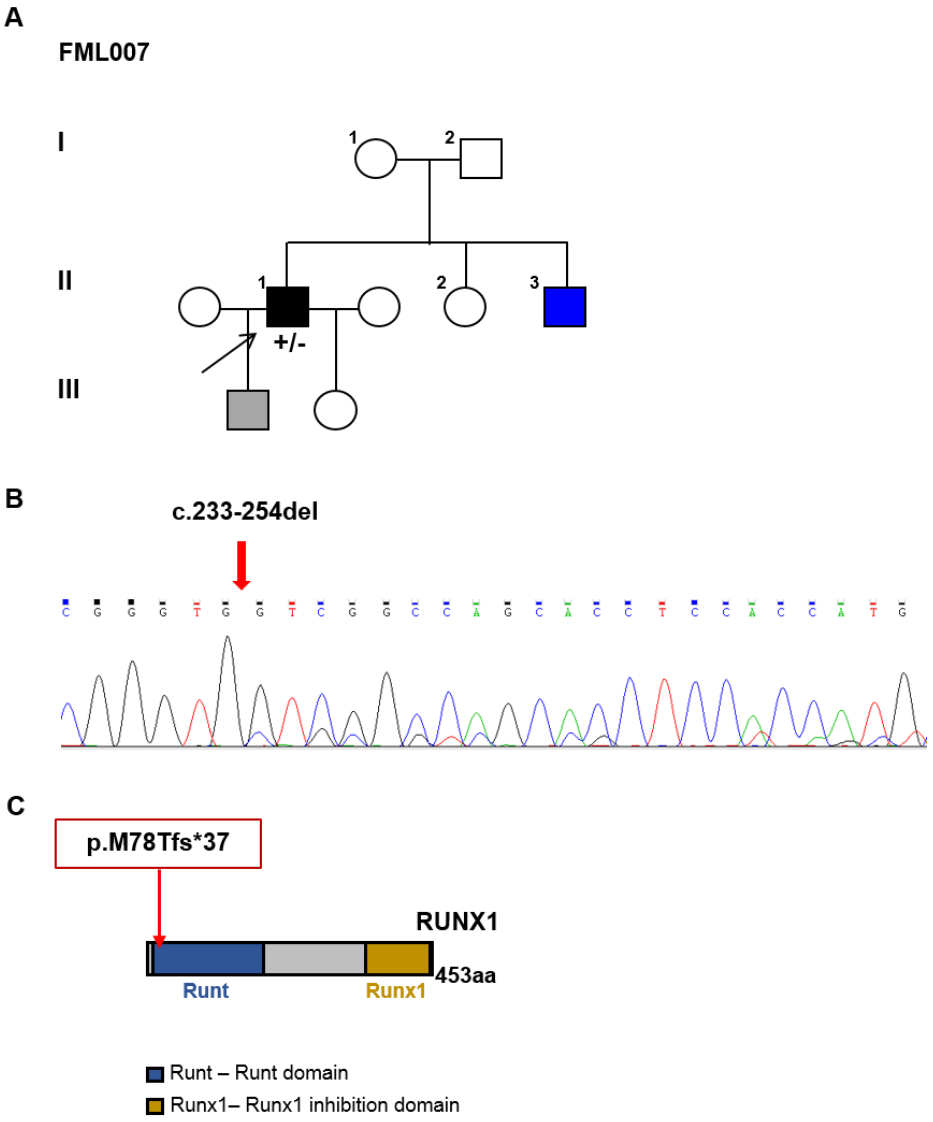


Figure 3.3. Example of a somatic *RUNX1* variant in exon 3. A. Family FML007 with acquired *RUNX1* deletion. Affected individuals are coloured as following black: AML, blue: AA and grey: thrombocytopenia. The black arrow highlights index case. The heterozygous genotyping is denoted as (+/-). B. Sanger sequence chromatogram, red arrow shows the variant in *RUNX1*. C. Schematic showing the position of the variant in RUNX1 protein (XP_011528068.1).

This variant is likely to cause loss of function, however it was found in 20% of the cells according to the peak of the traces from Sanger sequencing compared to the wild type sequence (Figure 3.3B). This observation raises the question as to whether this is a germline variant, since a germline variant would be expected to account for ~50% of the signal. Sanger sequencing done in DNA from fibroblasts confirmed that the variant c.233-254del is acquired and it is notable that acquired variants in *RUNX1* are found in up to 32% of *de novo* cases with AML (Liew & Owen, 2011). This finding highlights the importance of identifying whether a variant is germline or somatic in the first instance. It also shows further studies are needed to characterise this family.

3.2.2 Variants identified in genes associated with complex phenotypes

3.2.2.1 Variants identified in *GATA2*

GATA family consists of six zinc-finger transcription factor proteins that play essential roles in regulating transcription of genes involved in the development of many cell types, including haematopoietic, cardiac, and endodermal lineages. GATA1 (GATA binding protein 1), GATA2 (GATA binding protein 2) and *GATA3* (GATA binding protein 3) are essential for normal haematopoiesis, and alterations in these genes lead to a variety of blood disorders. Germline variants in *GATA1* cause sex-linked recessive forms of inherited thrombocytopenia and dyserythropoietic anaemia (Crispino & Weiss, 2014) and acquired variants in this gene are associated with transient abnormal myelopoiesis and AML associated with Down syndrome (Wechsler *et al*., 2002). Germline variants in *GATA2* are responsible for GATA2

deficiency syndrome, while acquired variants are associated with sporadic forms of MDS and AML (Shiba *et al*., 2014; Collin *et al*., 2015). GATA2 expression is essential for maintenance of haematopoietic stem cells as it activates GATA1, which drives the differentiation of these stem cells into erythroid/megakaryocytic lineage (Grass *et al*., 2003). Germline variants in *GATA3* are responsible for the syndrome of hypoparathyroidism, deafness, and renal anomalities (Van Esch *et al*., 2000) and somatic variants are seen in breast cancer and ALL (Perez-Andreu *et al*., 2013; Ping *et al*., 2016). GATA4, GATA5 and GATA6 are involved in heart formation, and germline variants are associated with congenital heart disease (Peterkin *et al*., 2005).

GATA2 deficiency syndrome is an autosomal dominant bone marrow failure disorder with systemic features caused by germline heterozygous variants in the gene *GATA2* causing a reduction of its expression consistent with this disorder being associated with haploinsufficiency of *GATA2* (Collin *et al*., 2015). This disorder presents variable clinical manifestations, including many other predisposing syndromes to MDS/AML such as familial MDS/AML, Emberger syndrome, MonoMAC syndrome, and a range of bone marrow failures (Hahn *et al*., 2011; Hsu *et al*., 2011; Ostergaard *et al*., 2011; Dickinson *et al*., 2011). Emberger syndrome as described previously, is characterised by primary lymphedema in the lower extremities and genitals, MDS with predisposition to develop into AML, cutaneous warts, and sensorineural deafness. Emberger syndrome can be inherited in an autosomal dominant manner with incomplete penetrance and it can also occur sporadically (Ostergaard *et al*., 2011). The MonoMAC syndrome is characterised by severe monocytopenia and severe infections with nontuberculous *Mycobacteria*, typically M. *avium* complex (MAC) and may present with natural killer cell and B-cell lymphocytopenia, fungal infections, pulmonary alveolar proteinosis, and severely decreases dendritic cells, and predisposition to MDS/AML or chronic myelomonocytic leukaemia (Vinh *et al*., 2010).

GATA2 variants are typically loss of function with no genotype-phenotype correlations (Hyde & Liu, 2011). Several families with *GATA2* variants with highly penetrant autosomal dominant inheritance have been described with no distinguishing phenotypic abnormalities other than early-onset familial MDS/AML (Hahn *et al*., 2011). An acquired variant in *ASXL1* frequently occur and monosomy 7 is the most commonly associated cytogenetic finding although trisomy 8 and trisomy 21 also may emerge (Spinner *et al*., 2014). Furthermore, the prognosis after the development of MDS/AML is usually poor (Churpek *et al*., 2015).

We have identified two loss of function variants in *GATA2* in two families from our cohort. In family FML026 (Figure 3.4), a stop gain variant in *GATA2* (c.1084 C>T; p. R362*) was identified in the index case (II:1, Figure 3.4A) who developed MDS at 36 years of age. She had congenital deafness, trisomy 8, had meningitis at age of 10 years, vulval cancer and fever with no apparent infection. Her brother (II:2, Figure 3.4A) had MDS and died following matched unrelated donor for monosomy 7. This novel nonsense variant is located in the highly conserved zinc finger 2 DNA binding domain (Figure 3.4 B and C) and causes truncation of the protein. As *GATA2* plays an essential role in regulating transcription of genes involved in the development and proliferation of haematopoietic and endocrine cell lineages (Gao *et al*., 2014), it is likely that this *GATA2* nonsense variant is the disease causing in this family.

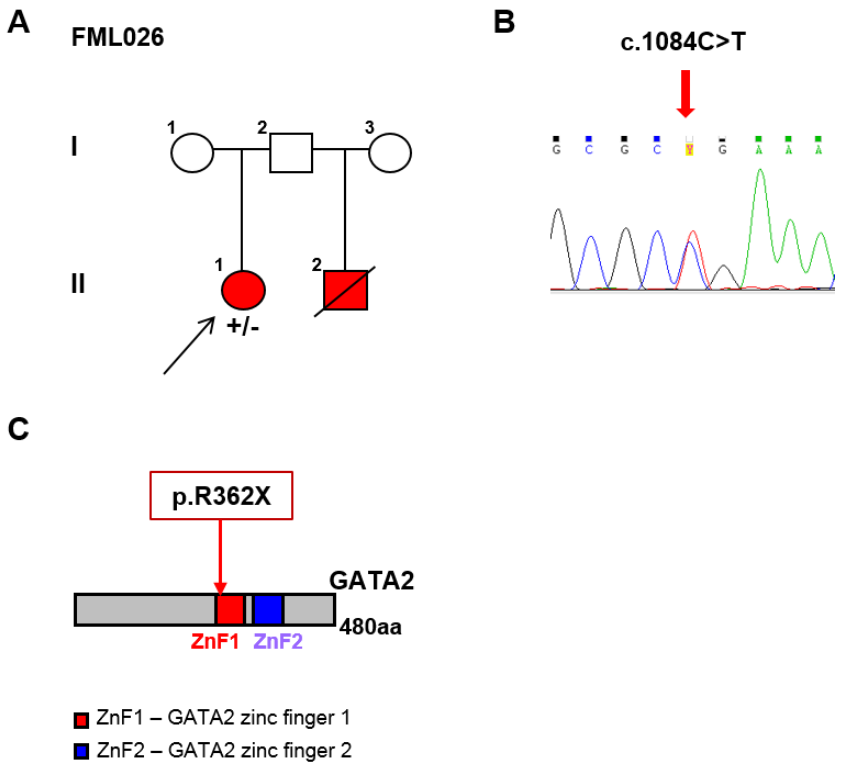


Figure 3.4. Characterisation of *GATA2* variant in exon 6. A. Family FML026 with *GATA2* variant. Affected individuals are coloured in red indicating MDS. The black arrow highlights index case. The heterozygous genotyping is denoted as (+/-). B. Sanger sequence chromatogram, red arrow shows the variant in *GATA2*. C. Schematic showing the position of the variant in GATA2 protein (NP_116027.2).

The second *GATA2* variant (c.630_643del, p.K212Tfs*65) was identified in family FML052 (Figure 3.5). The proband (II:3, Figure 3.5A) had MDS with lymphedema at age 16 years, however there was no family history. This *GATA2* 14bp deletion (Figure 3.5B and C) is likely to be *de novo* (spontaneously arising variants that can be passed to the next generation) and a pathogenic variant as the protein would possibly lose its function. GATA2 deficiency syndrome often explains sporadic cases of bone marrow failure, in the absence of a family history, arising from *de novo* germline *GATA2* variants (Hirabayashi *et al*., 2012).

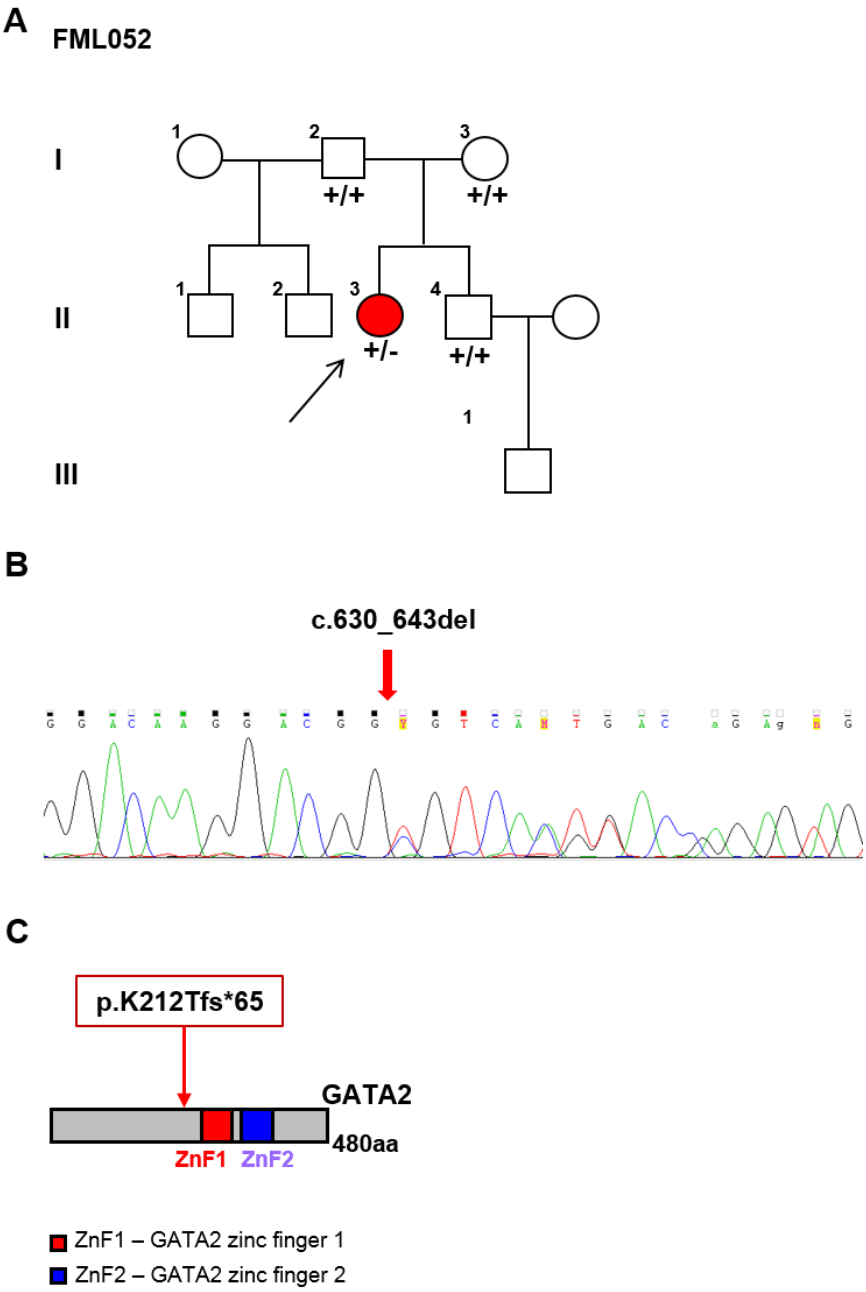


Figure 3.5. Characterisation of *GATA2* variant in exon 4. A. Family FML052 with *GATA2* variant. Affected individuals are coloured in red indicating MDS. The arrow highlights the index case. The genotyping is denoted as follows: wild-type (+/+) or heterozygous (+/-). B. Sanger sequence chromatogram, red arrow shows the variant in *GATA2*. C. Schematic showing the position of the variant in GATA2 protein (NP_116027.2).

3.2.2.2 Variant identified in *TERT*

Telomerase reverse transcriptase (*TERT*) gene is located on chromosome 5 and encode a protein of 1,132 amino acids, which is the catalytic subunit of telomerase. Telomerase is a ribonucleoprotein complex that consists of the protein TERT and its integral RNA subunit telomerase RNA component (TERC). It maintains the telomeres, which are essential structural elements that seal and protect the ends of chromosomes from recombination and end-to-end fusion. In normal somatic cells, telomeres gradually shorten after successive cell division, resulting in senescence (Gomez *et al*., 2012). Regulation of *TERT* impacts on telomerase activity and it is considered to have a critical role in tumour formation (Mosrati *et al*., 2015). Impaired telomerase activity and extremely short telomeres induce chromosomal instability, causing bone marrow failure, fibrosis of the lungs and liver, and tumour formation (Calado *et al*., 2012).

High *TERT* expression is observed in about 90% of cancer cells that typically have high telomerase activity, protecting them from proliferation arrest, senescence and apoptosis (Hanahan & Weinberg, 2011). The dysregulation of *TERT* expression in these cells is a result from alteration of the *TERT* structure by variants, epigenetic modifications or aberrant chromatin environments (Zhao *et al*., 2016) and these are associated with melanoma, gliomas and haematological malignancies. Two somatic variants (228C>T and 250C>T) in the promoter region of *TERT* were reported at positions -124 and -146 base pairs upstream of the *TERT* translation start site in melanomas (Horn *et al*., 2013; Huang *et al*., 2013). Each variant independently generates a novel E-twenty-six (ETS) transcription factor biding site (GGA/T) and has been shown to increase the transcriptional activity of the *TERT* promoter (Killela *et al*., 2013). *TERT* promoter variants occur frequently in several tumours, including

gliomas, liposarcomas, urothelial carcinomas and hepatocellular carcinomas (Huang *et al*., 2015). Haematological malignancies are not reported to carry these variants in *TERT* promoter regions but display enhanced telomerase activity and shortened telomeres (Mosrati *et al*., 2016). However, abnormal hypermethylation of CpG islands in a variety of gene promoters is the hallmark epigenetic changes in both AML and MDS (Cancer Genome Atlas Research Network, 2013). Zhao *et al.* (2016) analysed the methylation status of the *TERT* promoter region and identified a distinct epigenetic landscape of the *TERT* promoter region in patients with AML and MDS/AML. In their studies, haematopoietic cell lines presented with a highly methylated upstream region and a hypomethylated region around the transcription start site. Furthermore, CpG methylation rates were much higher in the cell lines than their corresponding primary leukaemic cells. Primary cells and their cell lines displayed a trend of increasing methylation intensity toward the 5' end while the regions around the transcription start site remained unmethylated in primary cell populations and relatively hypomethylated in the cell lines. These results demonstrated that methylation profiling of leukaemia cell lines does not represent the methylation pattern observed in the original primary leukaemic cells. This finding can be explained by the evolution of specific *TERT* promoter methylation patterns in regions occurring in cancer cell lines during the process of transformation from the original primary leukaemic cell (Smiraglia *et al*., 2001; Ahmed *et al*., 2013; Varley *et al*., 2013).

Germline variants in *TERT* can lead to familial MDS/AML and to telomere biology disorders including DC, HH (West *et al*., 2014). Telomere biology disorders are associated with abnormal telomere maintenance and predisposition to MDS/AML and patients may present initially with bone marrow failure, MDS, or pulmonary fibrosis, without demonstrating mucocutaneous features of DC (Yamaguchi *et al*., 2003). These disorders are caused by a number of genes in three inheritance patterns.

Autosomal recessive DC and HH may be caused by variants in *NOP1*, *TERT*, *NPH2*, *WRAP53*, *CTC1*, *RTEL1* and *PARN* (Dokal, 2011; Keller *et al*., 2012; Walne *et al*., 2013b; Tummala *et al*., 2015). X-linked recessive DC is associated with variants in *DKC1* (Heiss *et al*., 1998). Autosomal dominant DC results from variants in *TERT*, *TERC*, *TINF2*, and *RTEL1* (Ballew *et al*., 2013; Savage *et al*., 2008; Vulliamy *et al*., 2001). Heterozygous variants in *TERT* and *TERC* may result in familial MDS/AML predisposition syndromes (Kirwan *et al*., 2009). In familial MDS/AML, variants in *TERT* have an autosomal dominant form of inheritance with variable clinical manifestations and age at onset, incomplete penetrance and anticipation, in which progressively shorter telomeres passed down through generations (Vulliamy *et al*., 2004; Armanios, 2009).

The proband (III:5, Figure 3.6A) of family FML058 is a 40-year-old male with MDS and a complex karyotype including del(5q) and monosomy 18.

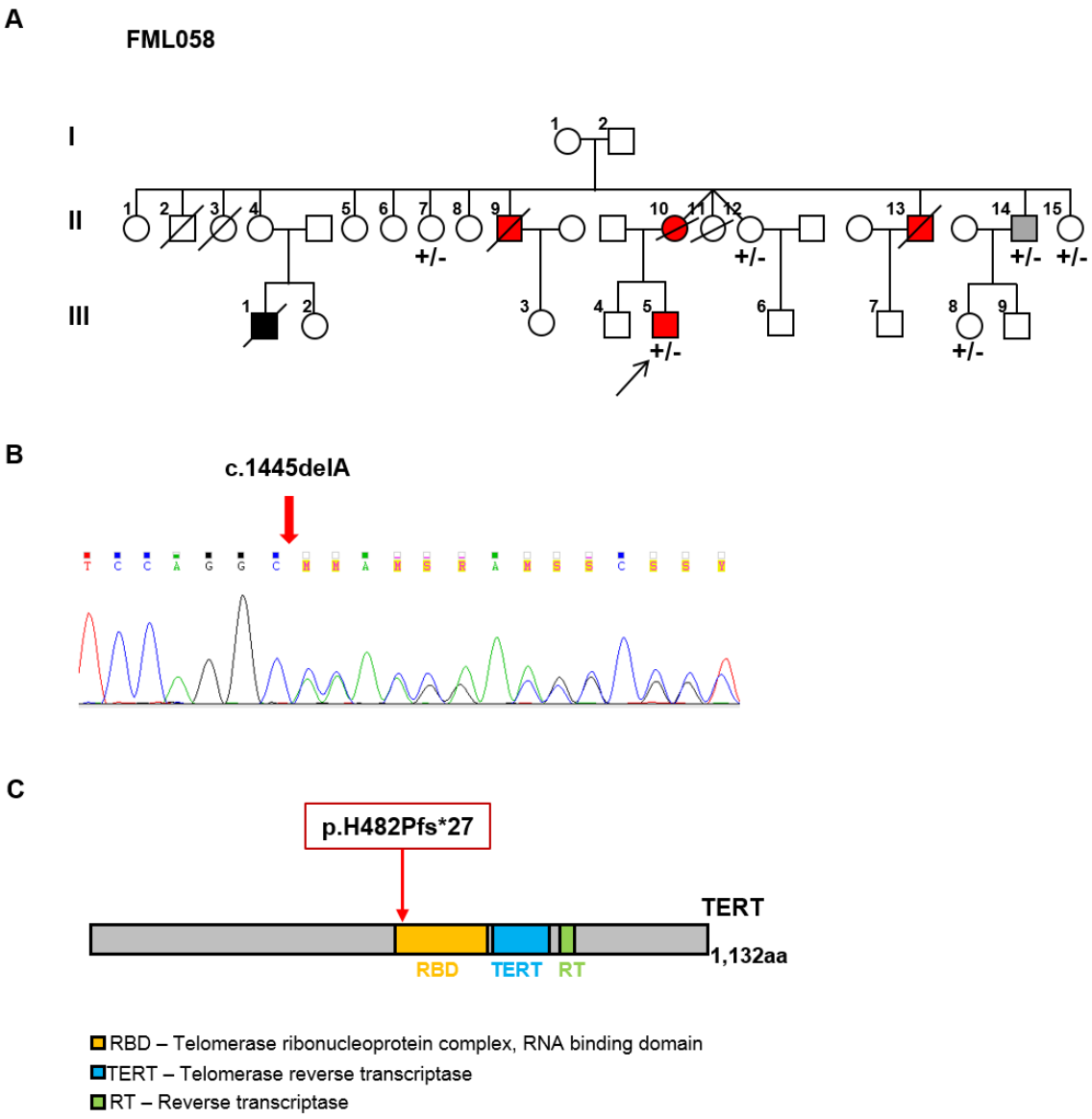


Figure 3.6. Characterisation of *TERT* variant in exon 2. A. Family FML058 with *TERT* variant. Affected individuals are coloured as following red: MDS, black: AML and grey: thrombocytopenia. The arrow highlights index case. The heterozygous genotyping is denoted as (+/-). B. Sanger sequence chromatogram, red arrow shows the variant in *TERT*. C. Schematic showing the position of the variant in TERT protein (NP_937983.2).

His mother (II:10, Figure 3.6A) died aged 60 years of MDS and had pulmonary infections and liver cirrhosis. Two of his uncles died of MDS with infections, one (II:9, Figure 3.6A) at 65 years old; and another (II:13, Figure 3.6A) at 56 years old with an addition of diabetes, arthritis and heart disease. His uncle aged 51 years (II:14, Figure 3.6A) had thrombocytopenia evolving to MDS with lung infections and his cousin (III:1, Figure 3.6A) died of AML. Deletion of a single nucleotide (c.1445delA, p.H482Pfs*27) resulting in a frameshift and truncation of the protein (Figure 3.6B and C) was identified in *TERT* in individuals II:7, II:12, II:14, II:15, III:6 and III:8 seen in Figure 3.6A through direct sequencing. Interestingly, this *TERT* variant has an incomplete penetrance as there are four asymptomatic carriers (II:7, II:12, II:15, III:8, Figure 3.6A) in this family and striking incidence of MDS occurring with pulmonary symptoms, and late onset of disease.

3.2.2.3 *TERT* promoter region and *ANKRD26* 5'UTR screening in our cohort of patients by dHPLC

As described above, two highly recurrent somatic variants in *TERT* promoter region were found in melanoma, bladder and hepatocellular cancer cells: C>T at -124bp (described as 228C>T) and C>T at -146bp (described as 250C>T), these positions are according to Human Genome Issue hg19 (Figure 3.7). In addition, Horn *et al*. (2013) identified one germline variant in a family with melanoma (T>G at -57, shown in light blue in Figure 3.7B) and two tandem variants (CC>TT, shown in pink in Figure 3.7B) occurred at -124/-125bp and at -138/-139bp that also creates binding motifs for the transcription factors ETS and for ternary complex factors (TCFs) in sporadic melanoma.

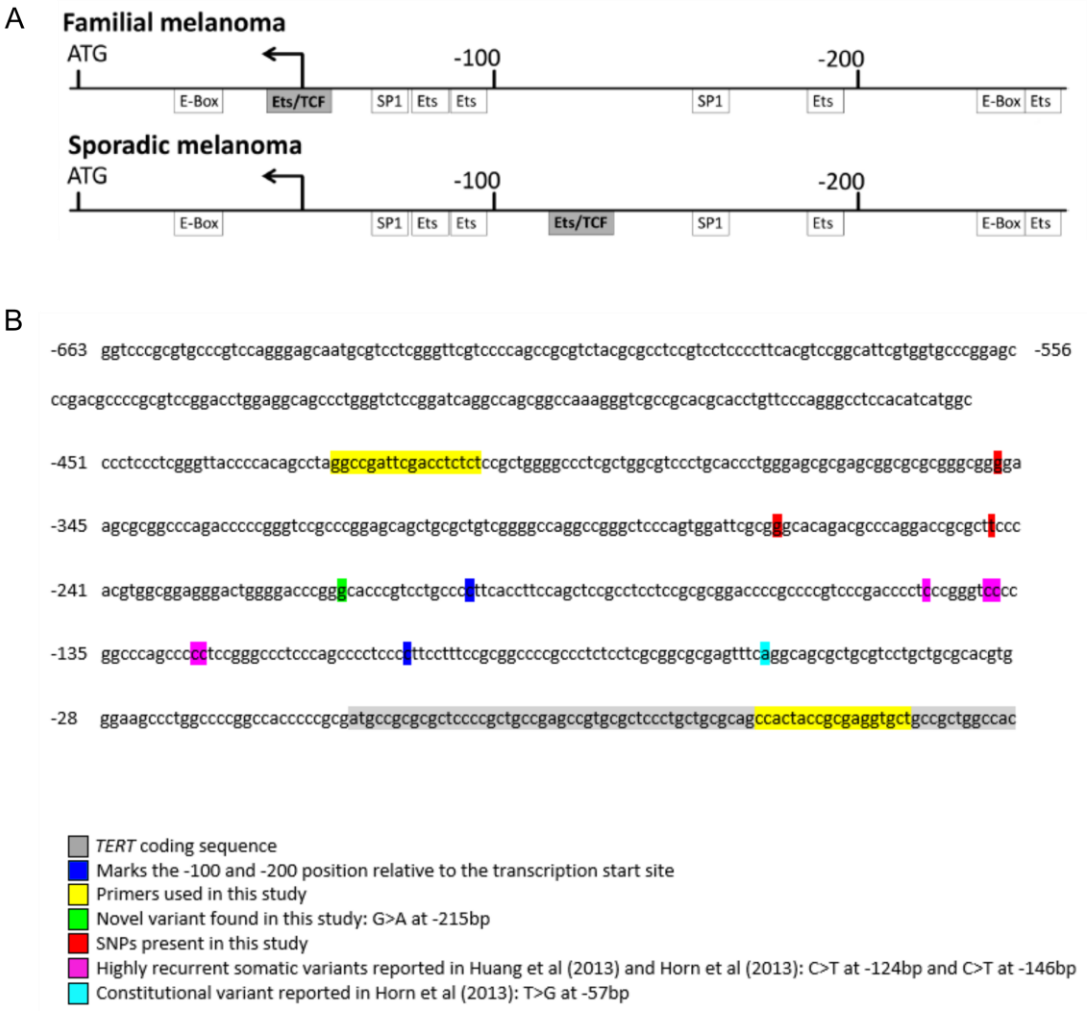


Figure 3.7. *TERT* core promoter. A. Mutations 228C>T and 250C>T created Ets/TCF binding motifs in melanoma (adapted from Horn *et al*., 2013). B. *TERT* promoter region sequence showing the *TERT* coding sequence; the primer position used in this study to cover the full length of *TERT* promoter region; the novel variant and SNPs found in this study; the highly recurrent somatic and germline variants identified in melanoma by Horn *et al*., 2013.

Since germline TERT variants can be found in both sporadic AML (Calado *et al*., 2009) and familial MDS/AML (Kirwan *et al*., 2009), we investigated whether *TERT* promoter variants also occur in MDS/AML families as it had not been described to date. We screened 52 patients from our cohort of familial MDS/AML, 26 constitutional MDS/AML (individuals presenting MDS/AML alongside with one or more somatic

abnormalities), 12 familial MDS with other type of cancer beyond leukaemia and 10 idiopathic cases for variants in the *TERT* full core promoter (489bp from -424 to +65) by dHPLC as the exome sequencing does not cover this region (Figure 3.8).

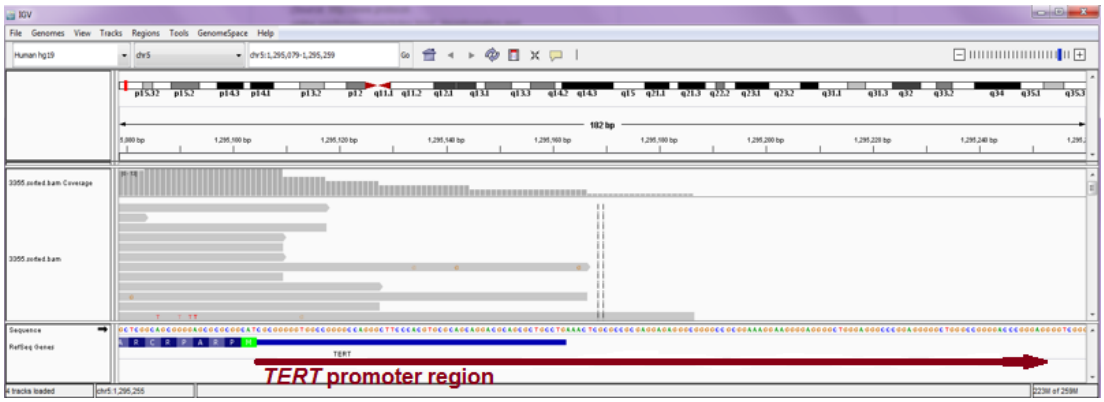


Figure 3.8. BAM file of *TERT* promoter region. The horizontal red arrow shows that *TERT* promoter region is not fully covered by WES.

The SNP rs35226131 (G>A at position -269) was found in four patients and in three of them another SNP rs2853669 (T>C at position -245) was also present (Figure 3.9). In addition, one of these patients have a novel variant G>A at -215bp upstream of the transcription start site (TTS) (Figure 3.9B and C). Overall, this novel variant identified in *TERT* promoter is likely to be of unknown significance as there is no sufficient evidence for pathogenicity. However, considering that there are available samples from the patient harbouring this novel variant, a Luciferase assay could be performed in cell lines established from a patient cell and in the controls. The promoter activity of the constructs containing this novel variant at -215bp of the *TERT* promoter could be compared to the wild type.

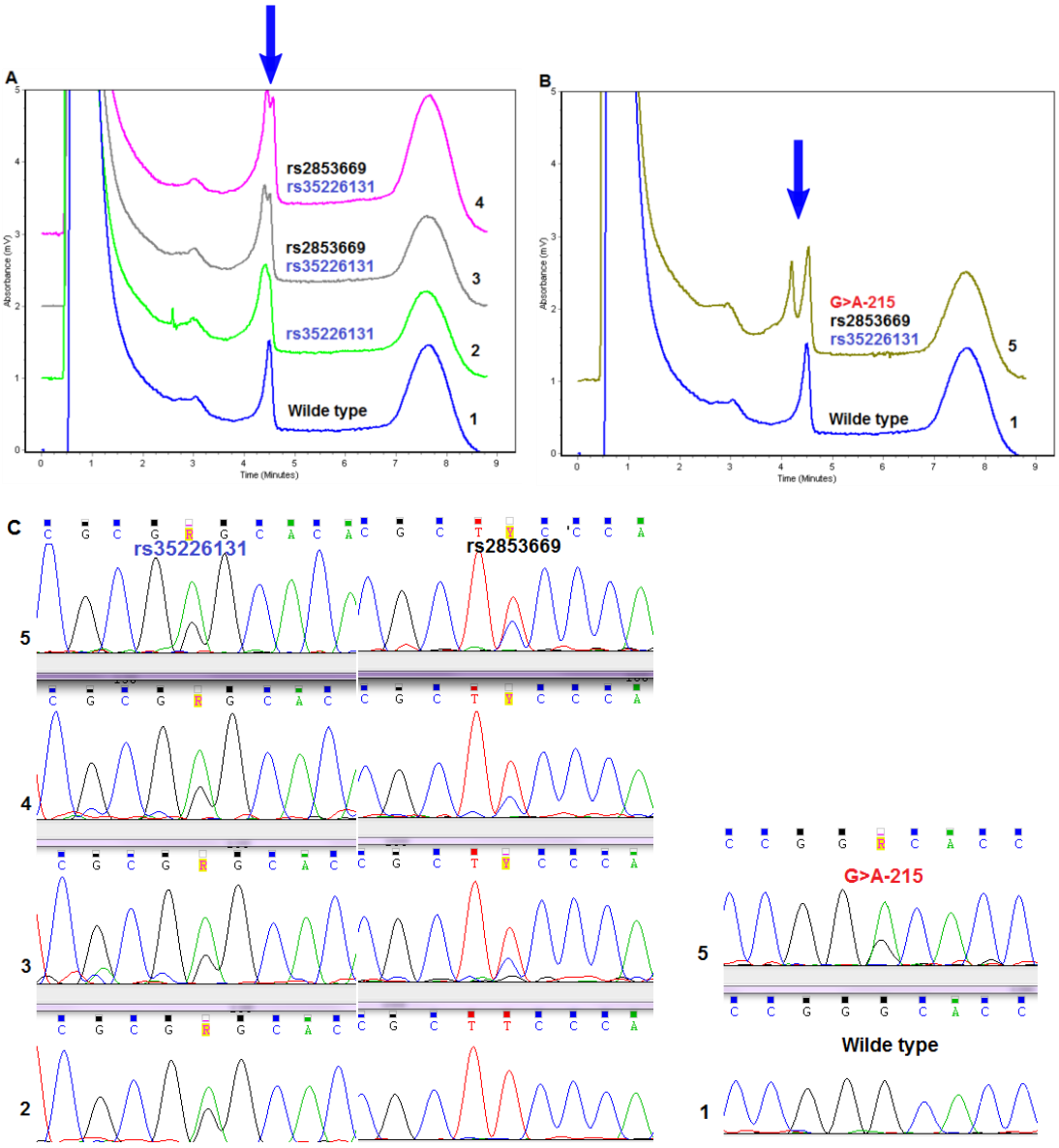


Figure 3.9. Denaturing high performance liquid chromatography (dHPLC) *TERT* promoter results. A and B. Abnormal dHPLC elution patterns in patient samples 2-5 shown by the blue arrows. The abnormal dHPLC elution patterns correspond to SNPs identified in patients 2-5 as indicated by 'rs' numbers and an additional novel variant shown in red was identified in patient 5. The respective Sanger sequencing chromatograms are shown in C.

Furthermore, the same group of patients described above was screened for the 5' untranslated region (UTR) of ankyrin repeat domain 26 gene (*ANKRD26*) by dHPLC

as this region is also not covered by WES (Figure 3.10). *ANKRD26* encodes a protein containing N-terminal ankyrin repeats which function in protein-protein interactions and, variants in *ANKRD26* 5'UTR were identified in families with an autosomal dominant thrombocytopenia 2 (THC2) (Pippucci *et al*., 2011).

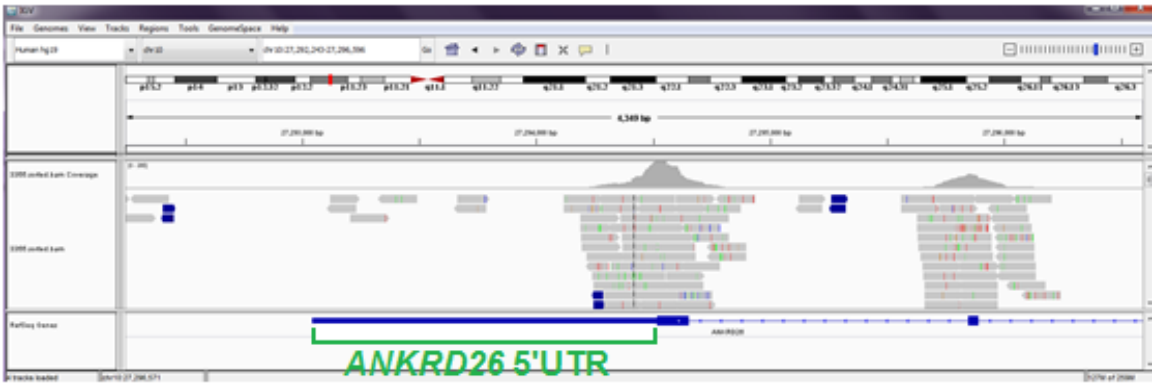


Figure 3.10. BAM file showing that *ANKRD26* 5'UTR (highlighted in green) are not covered by WES.

Bluteau *et al*. (2014) demonstrated that THC2 is caused by the impaired binding of RUNX1 and Fli-1 proto-oncogene, ETS transcription factor (FLI1) to the altered *ANKRD26* which results in the *ANKRD26* down regulation by these same transcription factors and a subsequent *ANKRD26* overexpression in megakaryocytes. Patients with variants in *ANKRD26* are more predisposed to myeloid malignancies, in particular AML and the thrombocytopenia in these patients is characterised by normal platelet size, moderate thrombocytopenia, and absent or mild bleeding tendency (Noris *et al*., 2013; Marconi *et al*., 2017). Through dHPLC screening, few SNPs (c.1-59G>A rs3737056, c.1-140C>G rs41299222 and c.1-229A>C rs7897698) were identified but there were no novel potentially damaging variants (Figure 3.11).

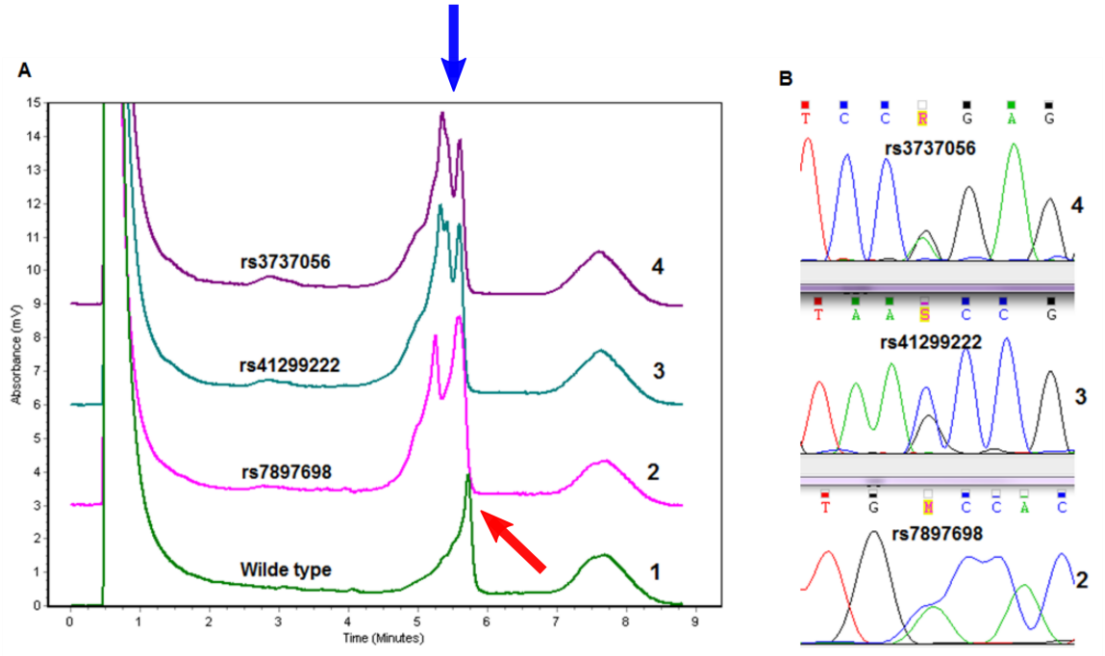


Figure 3.11. Denaturing high performance liquid chromatography (dHPLC) *ANKRD26* 5'UTR results. A. Abnormal dHPLC elution patterns of SNPs identified in patients 2-4, shown in blue arrow compared to the wild type (1) shown in red arrow. B. Sanger sequencing chromatograms.

3.2.3 Variants identified in genes associated with inherited syndromes with predisposition to MDS/AML

3.2.3.1 Shwachman-Diamond syndrome

Ribosomes are essential machines that read genetic code in the cells, translating them into proteins. The ribosomal large 60S and small 40S subunits are pre-assembled in the nucleus and then exported to the cytoplasm to be matured. Genetic variants that occur in genes involved in the ribosome biogenesis cause disorders called ribosomopathies such as Diamond–Blackfan anemia, DC, Shwachman–Diamond syndrome, del(5q) myelodysplastic syndrome, Treacher Collins syndrome,

Cartilage–hair hypoplasia, North American Indian childhood cirrhosis and isolated congenital asplenia (Warren, 2017).

Shwachman-Diamond syndrome (SDS) is an autosomal recessive disorder caused by biallelic variants in *SBDS* and is associated with bone marrow failure and predisposition to MDS/AML. It is characterised by multiple developmental anomalies including skeletal abnormalities, cognitive impairment, exocrine pancreatic and poor growth (Myers *et al*., 2014). *SBDS* is a cofactor for elongation factor-like GTPase 1 (EFL1) to disassociate the eukaryotic initiation factor 6 (eIF6) from the late subunit joining the 60S ribosomal subunit in the cytoplasmic maturation (Finch *et al*., 2011; Weis *et al*., 2015). Recently, biallelic variants in DnaJ heat shock protein family (Hsp40) member C21 (*DNAJC21*) were described by our group (Tummala *et al*., 2016) and another (Dhanraj *et al*., 2017) causing SDS like disease. *DNAJC21* plays a role in cytoplasmic maturation of the 60S ribosomal subunit (Myers *et al*., 2014). However, whilst no homozygotes for a loss of function variant have been reported in *SBDS*, suggesting that these alterations are embryonic lethal, two cases carrying homozygous nonsense variants were reported in *DNAJC21* (Shammas *et al*., 2005; Tummala *et al*., 2016).

Boocock *et al*. (2003) found that more than 90% of patients with SDS carry recurrent biallelic variants in *SBDS* including one of three common pathogenic *SBDS* variants: 183_184TA>CT, 258+2T>C, or the combination of 183_184TA>CT and 258+2T>C. The variant 258+2T>C disrupts the donor splice site of intron 2, while 183_184TA>CT, introduces an in-frame stop codon (K62X). In our cohort, recurrent biallelic variants in *SBDS* (c.184A>T, p.K62*, rs120074160 and c.258+2T>C, rs113993993) were found in the family FML041 (Figure 3.12). This was possible through data analyses from whole exome sequencing performed in all our familial MDS/AML patients that had

good quality of DNA sample. The index case (II:2) and her twin (II:1, Figure 3.12A) did not fulfil the classical clinical criteria of SDS at the time of diagnosis. It is well known that inherited bone marrow failure syndromes are heterogeneous disorders with overlapping features and predisposition to MDS or AML. Patients with SDS often present with a combination of exocrine pancreatic insufficiency with malabsorption and neutropenia in infancy with propensity to develop bone marrow failure malignancies (Lindsley *et al*., 2017). In addition, they may have low birth weight, short stature, metaphyseal dysostosis, neurocognitive deficits, immunodeficiency among other less common findings (Myers *et al*., 2014).

The index case (II:2, Figure 3.12A) of family FML041 was a 23 years old woman; she and her twin (II:1, Figure 3.12A) had longstanding history of MDS since 6 months of age. Both twins had an extensive past medical history including in the index case (II:2, Figure 3.12A) a cleft of her soft palate at birth, speech delay in childhood, hearing loss, a history of congenital hip dislocation, mild short stature, dysmorphic features with round face, small nose, pinched alae nasi, long smooth filtrum, rounded ears and mild joint hyperextensibility. Her twin sister (II:1, Figure 3.12A) was also noted to have recurrent chest infections and asthma. II:1 (Figure 3.12A) died following the transformation of MDS to AML. The index case (II:2, Figure 3.12A) died a year later with refractory AML post allograft. Whole exome sequencing data revealed biallelic variants (c.184A>T, p.K62* rs120074160 and c.258+2T>C, rs113993993) in exon 2 of *SBDS* are likely to be the disease causing in this family (Figure 3.12B and C). These compound heterozygous germline variants are frequently seen in patients with SDS, although *de novo* variants are seen in 10% of SDS cases (Steele *et al*., 2014).

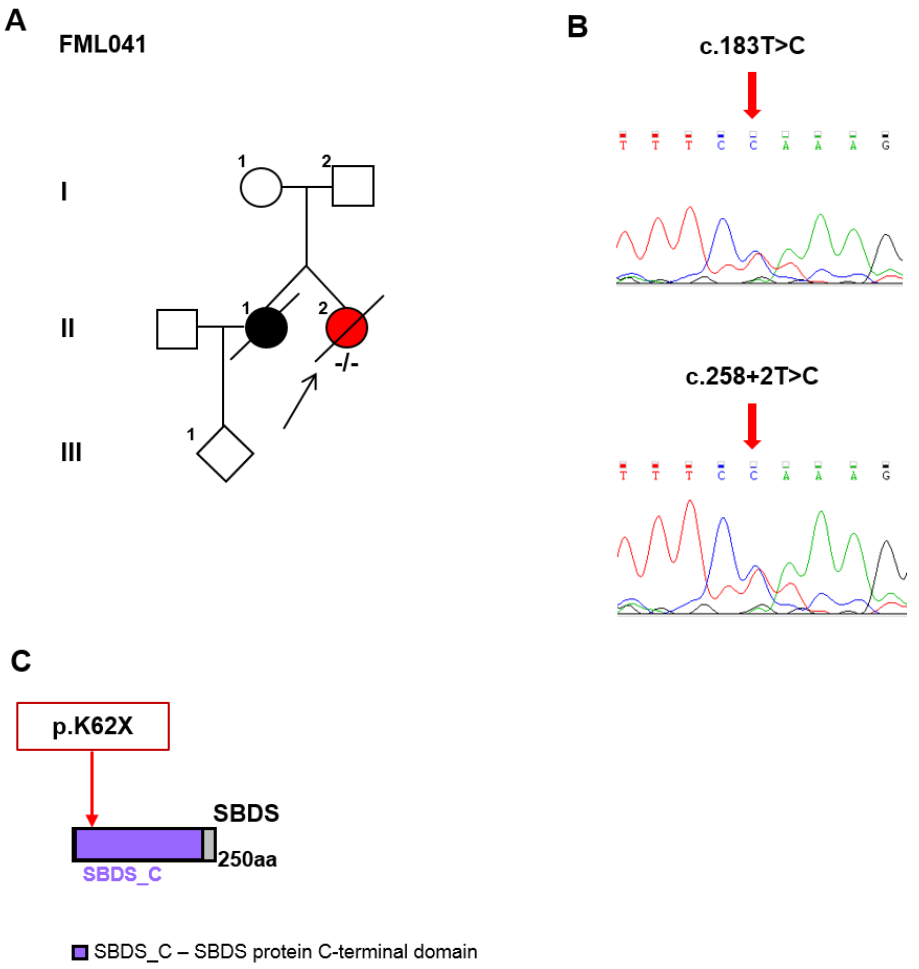


Figure 3.12. Characterisation of *SBDS* variants in exon 2. A. Family FML041 with *SBDS* variants. Affected individuals are coloured as following red: MDS and black: AML. The arrow highlights index case. The compound heterozygous state is denoted as (-/-). B. Sanger sequence chromatogram showing the variant in *SBDS*. C. Schematic showing the position of the variant by the red arrow in SBDS protein (NP_057122.2).

3.2.3.2 Fanconi anaemia

Fanconi anaemia (FA) is associated with congenital abnormalities in multiple organs, haematological manifestations at a young age such as bone marrow failure, MDS, and AML (Alter *et al*., 2010). FA is characterised by increased chromosomal breakage

upon exposure to DNA crosslinking agents, diepoxybutane (DEB), or mytomycin C. It is a rare recessive DNA repair deficiency resulting from biallelic damaging variants in one of at least 20 genes (*FANCA*, *FANCC*, *FANCD1*/*BRCA2*, *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *FANCI*, *FANCJ/BRIP1*, *FANCL*, *FANCM*, *FANCN/PALB2*, *FANCO*, *FANCP*, *FANCQ*/*ERCC4*/*XPF*, *FANCS*/*BRCA1*, *FANCT*/*UBE2T*, *FANCU*/*XRCC2*, *FANCV*/*MAD2L2*/*REV7* and *FANCW*/*RFWD3*). Damaging variants in the X-linked gene *FANCB* and heterozygous damaging variants in *FANCR*/*RAD51* can also cause FA (Auerbach, 2009). These genes are involved in repairing DNA crosslinks associated with the FA/BRCA pathway. Complete molecular diagnosis of the disease causing gene and specific pathogenic variants are required for confirmation of FA and efficient clinical management (Flynn *et al*., 2014). This is due to the fact that different FA complementation groups demonstrate variable outcome. While variants in *FANCA* lead to a mild disease and later onset of bone marrow failure, variants in *FANCG* usually result in severe haematological malignancies (Faivre *et al*., 2000).

Biallelic variants in Fanconi anaemia complementation group A (*FANCA*) (Figure 3.13A) are responsible for the development of the disease in most cases of FA. Large and small deletions, small insertions, and single nucleotide variation are the type of disease causing variants found in *FANCA* (Kimble *et al*., 2017). We have identified biallelic variants in one family of our cohort of familial MDS/AML.

The proband (III:2, Figure 3.13B) of family FML004 was a female with mild short stature associated with non-severe AA at the age of 38 years. She developed AA after receiving treatment for her tongue and throat cancer. Her mother (I:1, Figure 3.13B) had short stature and her brother (III:1, Figure 3.13B) died at age of 19 years with AML. Her maternal aunt had stomach cancer in her forties and a further maternal uncle who died in his sixties.

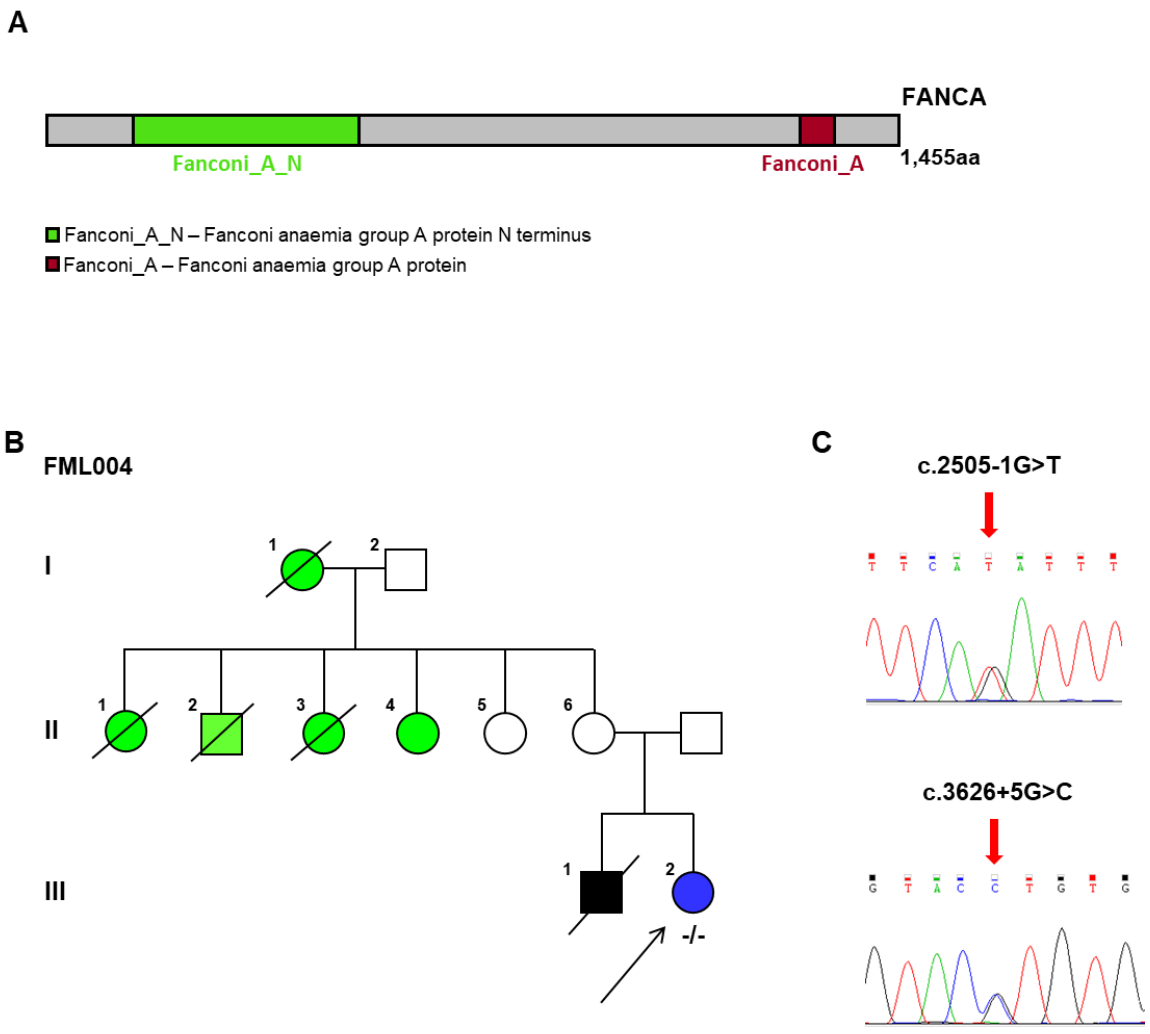


Figure 3.13. Characterisation of *FANCA* variants in exons 27 and 36. A. Schematic of FANCA protein (NP_000126.2). B. Family FML004 with *FANCA* variants. Affected individuals are coloured as following blue: AA and black: AML. The arrow highlights index case. The compound heterozygous state is denoted as (-/-). C. Sanger sequence chromatogram, red arrows show variants in *FANCA*.

Two other maternal aunts have died of breast cancer in their fifties and further aunt had two thumbs. Her maternal grandmother had throat cancer and stomach problems and died in her sixties. The index case had some chromosome instability. This result was not within the normal range but neither was a diagnostic of FA. However, very

rare biallelic variants (c.2505-1G>T and c.3626+5G>C) likely to be pathogenic were identified in *FANCA* by exome sequence performed in the index case and validated by Sanger sequencing (Figure 3.13C). Hence, despite II:1 not presenting with typical congenital abnormalities nor radial-ray, which are features frequently seen in FA, our results support the diagnosis of FA for this patient.

3.2.4 Variants identified in genes associated with other inherited syndromes

3.2.4.1 Wiskott-Aldrich syndrome

Wiskott-Aldrich syndrome (WAS) is a rare X-linked recessive disease characterised by thrombocytopenia, bloody diarrhoea, immunodeficiency, recurrent infections, inflammatory symptoms, and eczema. It was described in 1937 and 1954 by Dr. Alfred Wiskott and Dr. Robert Aldrich respectively. The presentation of WAS is very heterogeneous ranging from mild to severe, in which severe conditions such as the development of autoimmunity and lymphoproliferative disorders and lymphoid malignancies may occur. The presentation of these disorder typically occurs in early life while the diagnosis may sometimes be in adulthood due to its heterogeneity (Ochs *et al*., 2009; Worth & Thrasher, 2015). Damaging variants in *WAS* gene are associated with the development of WAS. These variants are typically missense, splice site, nonsense, insertions and deletions are distributed along *WAS* (Shcherbina *et al*., 2003; Jin *et al*., 2004). Furthermore, the WAS related disorder X-linked neutropenia (XLN) is a very rare type of severe congenital neutropenia and MDS with recurrent infections that develop in an early infancy with a diverse presentation from

that of WAS. Only four *WAS* missense variants have been reported to cause XLN to date, including p.L270P, p.S272P, p.I276S, and p.I290T (Devriendt *et al*., 2001; Ancliff *et al*., 2006; Beel *et al*., 2009; Kobayashi *et al*., 2017). *WAS* is located on the X chromosome and encodes the WAS protein (WASp) which is member of a family of actin nucleation-promoting factors that translate surface signals into actin polymerization through the actin-related protein, Arp2/3 complex. It is important in actin polymerization, cytoskeletal remodelling and it is only expressed in haematopoietic cells (Thrasher & Burns, 2010). As WASp regulates the actin cytoskeleton processes including immune synapse formation, cell signalling, migration and cytokine release in most haematopoietic lineages, it participates in innate and adaptive immunity and is consequently important for normal function of immunological processes (Baptista *et al*., 2016).

Family FML032 (Figure 3.14A) from our cohort was found to have a single nucleotide deletion (c.1336delA, p.K446Rfs*24) causing premature stop codon in the WASp (Figure 3.15B and C). The index case (II:2, Figure 3.14A) was a three months old boy who presented with thrombocytopenia on day 1 and progressed to MDS. His six year old brother (II:1, Figure 3.14A) developed thrombocytopenia and splenomegaly at nine weeks of age and progressed to MDS. The *WAS* variant p.K446Rfs*24 was identified in both brothers. This deletion in *WAS* is likely to be disease causing in both brothers.

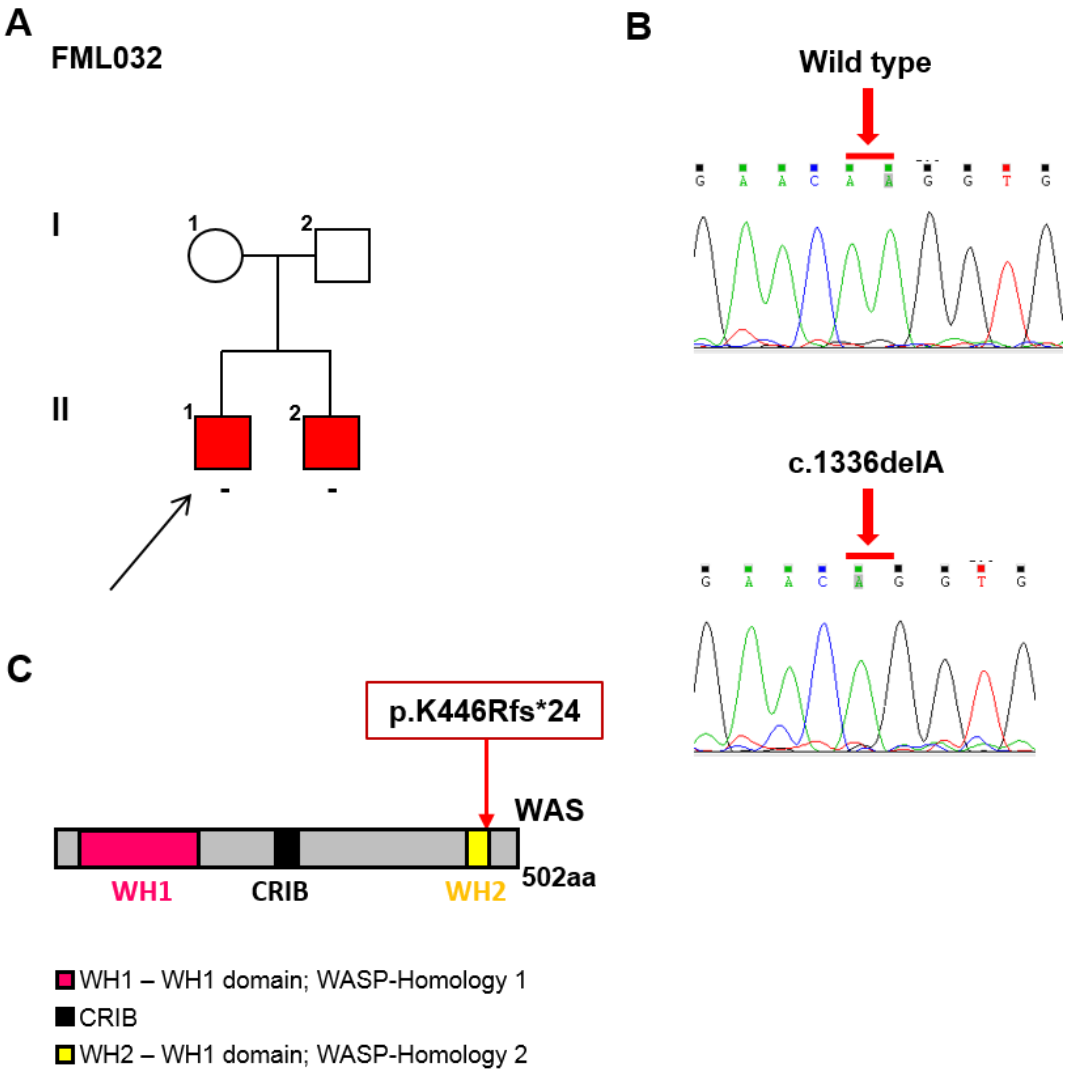


Figure 3.14. Characterisation of *WAS* variant in exon 10. A. Family FML032 with *WAS* variant. Affected individuals are coloured in red indicating MDS. The arrow highlights index case. The hemizygous genotype is denoted as (-). B. Sanger sequence chromatogram, red arrow shows the variant in *WAS*. C. Schematic showing the position of the variant in WAS protein (NP_116575.3).

3.3 Discussion

The recognition of familial MDS/AML is essential for effective clinical management of patients with an inherited susceptibility and their families. Besides documentation of a complete family history, germline genetic testing is also important to help to clarify whether a patient may have inherited predisposition to MDS or AML. Germline genetic testing is becoming more common to aid this recognition. Such officially recognised tests exist for FPD/AML caused by variants in *RUNX1*; familial AML with variants in *CEBPA*; familial MDS/AML with variants in *GATA2*; and for inherited bone marrow failure syndromes including DC and FA (Nickels *et al*., 2013). It is important to note that the segregation of the germline variant with the disease within the family is crucial to assign its pathogenicity. Certain features of the history or laboratory values may increase the likelihood of a particular syndrome over others in individual cases.

Here, we characterised seven families part of our familial MDS/AML cohort. Family FML053 had a large deletion in *RUNX1*, carriers exhibited heterogeneity in their degree of platelet dysfunction (individual II:5 is thrombocytopenic with no myeloid malignancy detected to date) and multiple individuals have developed T-cell ALL and acute myeloid leukaemia. Anticipation is a strong characteristic of FPD/AML and a close clinical follow up for members of the followed generations in the family is critical Disease anticipation leads to occurrence of MDS and AML in younger individuals in subsequent generations (Jongmans *et al.*, 2010; Nickels *et al.*, 2013; Sood *et al.*, 2017). Furthermore, we also identified an example of a deletion of 21bp in *RUNX1* in FML007 that is actually acquired and, therefore it is not likely to be the disease causing variant in this family. This highlights that interpretation of genetic testing in blood can be confounding by the presence of somatic variants found in known disease causing genes. Despite the recommendation of use of a non-blood tissue to validate

a germline variant, we do not have it available in all our families. However, segregation analysis within the family can be useful in distinguishing inherited from somatic variants.

We have not found *CEBPA*, *SRP72*, *TERC* nor *ANKRD26* variants in our uncharacterised familial MDS/AML cohort in these studies. However, *GATA2* variants were found in two families and they are possibly the underlying cause of the disease in these families. The proband of family FML026 harbours a nonsense variant with variable penetrance as the father did not present with the disease and she is also reported to have trisomy 8. Interesting *GATA2* variants are usually highly penetrant with variable cytogenetic abnormalities, including monosomy 7, trisomy 8 and trisomy 21 (Hahn *et al*., 2011). The proband of the second family FML052 was diagnosed with Emberger syndrome and harbours a frameshift deletion of 14bp in *GATA2*. This is likely to be a *de novo* event as her parents and her brother do not carry this variant and it was found in about 50% of the cells according to the peak of the traces from Sanger sequencing compared to the wild type sequence.

The large family FML058 with familial MDS/AML was characterised by a deletion of one nucleotide in *TERT*, which caused frameshift and truncation of the protein. Affected members in this family presented MDS/AML along with lung infections and liver cirrhosis common features seen in patients carrying variants in *TERT* (Dokal, 2011).

Patients with rare inherited bone marrow failure syndromes including FA, DC, Diamond Blackfan anaemia and SDS can be diagnosed during evaluation of haematological manifestations such as MDS or AML as a result of specific clinical phenotypes, screening tests or genomic studies (West & Churpek, 2017). This group

of disorders may present overlapping features and increased rates of malignancies. Despite the common features, these syndromes are very diverse in their manifestations and complications (Dietz *et al*., 2017). Here, we described families from our cohort of familial MDS/AML diagnosed with Fanconi anaemia and SDS as their initial presentation was atypical.

Our case of FA in family FML004 shows the importance of a molecular diagnosis to elucidate a complicated case. The diagnosis of FA of the proband of FML004 was not confirmed until the identification of biallelic variants in *FANCA*. Taking into account that she did not present with classical FA clinical features such as abnormal skin pigmentation, radial ray defects, and organ and skeletal abnormalities and exhibited intermediate chromosome breakage instability for a classical diagnosis of FA. This marked heterogeneity means diagnosis of FA is not always straightforward and it is important to consider this diagnosis in young patients presenting with cancers (Fargo *et al*., 2014).

Family FML041 was found to have recurrent biallelic variants in *SBDS*, however the proband of this family had not been diagnosed with SDS prior to these findings. Regardless of her and her twin sister's longstanding MDS along with many other congenital anomalies, she did not present with the classical exocrine pancreatic insufficiency compatible with a diagnosis of SDS. Again, it was through genetic studies that a diagnosis of SDS was made.

Family FML032 is another interesting case as both young brothers in this family were found to carry a frameshift variant in *WAS*. Although both boys presented with thrombocytopenia at an early age they both progressed to MDS. However, neither of them presented with immunodeficiency, eczema nor repeated respiratory infections

that are typically seen in WAS. This family highlights that early presentation of thrombocytopenia with evolution into MDS, can also arise from variants in WAS in the absence of typical WAS features.

Finally, the incorporation of next-generation sequencing as a standard practice into clinic may change the clinician's diagnostic approach. As a consequence, it is important to consider that diagnosis of an underlying germline variant raises the issue of using an allogeneic stem-cell transplant (STC) for consolidation therapy, since transplant is the only way to eradicate the bone marrow of the underlying predisposing allele. Furthermore, in order to get appropriate counselling and treatment, any new outcome that occurs after a STC in patients with inherited bone marrow failure disorders, needs to be distinguished from an event derived from STC or clinical feature acquired with aging.

Chapter 4

Germline heterozygous *DDX41* variants in a subset of familial myelodysplasia and acute myeloid leukaemia

4.1 Introduction

Considering that germline heterozygous mutations have been identified in DEAD-box helicase 41 gene (*DDX41*) in patients with familial MDS/AML in 2015 by Polprasert *et al*., we have decided to investigate variants in this gene in our cohort using a combination of whole exome and targeted sequencing.

Subsequently, heterozygous germline loss of function (LoF) variants in the gene *DDX41* (canonical transcript, NM_016222.3, NP_057306.2, 622 amino acids) that segregate with disease in four families with MDS/AML were identified and are described in this chapter.

4.1.1 Familial myelodysplasia and acute myeloid leukaemia with germline predisposing variants in *DDX41*

Recently, alterations in *DDX41* were identified to cause familial MDS/AML. Although, the age of onset and potential function of *DDX41* differ this gene from the previous familial MDS/AML predisposing genes (Polprasert *et al*., 2015).

The median age of MDS or AML onset in carriers of germline *DDX41* variants is 62 years, which is similar to the diagnosis age of sporadic MDS and AML. While the other predisposition syndromes occur in an earlier age of onset (Tawana & Fitzgibbon, 2016). Another different aspect is that most of the carriers of variants in *DDX41* have normal blood counts up until the development of MDS or AML, unlike other predisposing genes such as *GATA2* and *RUNX1* that are often characterised by

extended pre-leukemic cytopenic phases (Lewinsohn *et al*., 2016). These features make it harder to diagnose familial MDS/AML in patients carrying germline *DDX41* variants.

Finally, DDX41 has been implicated by functional studies to be involved in innate immune response, mRNA splicing, ribosome biogenesis and post-transcriptional regulation of protein translation in cell growth (Cheah *et al*., 2017). Despite all these function implications, the role of DDX41 in leukemogenesis is yet to be established (Li *et al*., 2016).

4.1.2 DDX41 – DEAD-box helicase 41 structure and function

RNA and DNA helicases are considered to be enzymes that catalyse the separation of double-stranded nucleic acids in an energy-dependent manner. They utilise the energy derived from a nucleoside triphosphate (NTP) hydrolysis to dissociate duplexes or displace bound proteins (Linder & Jankowsky, 2011). However, not all helicases have purely a double-stranded unwinding activity and they might be involved in various functions (Cordin *et al*., 2006).

DExD/H box family of RNA helicases are part of the large SF2 helicase superfamily. There are 59 highly conserved DExD/H helicases in eukaryotes (Zhang *et al*., 2011b) which play important roles in RNA metabolism, including ribosome biogenesis, RNA processing and folding, ribonucleoprotein modelling, RNA nuclear export, in the regulation of RNA translation and transcription, and in nonsense-mediated RNA decay. DExD/H box RNA helicases can have different functions in these processes such as RNA chaperones, ATP-dependent RNA helicases and unwindases, as

RNPases by mediating RNA-protein association and dissociation (Pyle, 2008) or as co-activators and co-repressors of transcription (Wortham *et al*., 2009). In addition, several DExD/H box family members are involved in viral replication. They are captured and regulated by viral proteins (Schröder, 2010), and are involved in viral RNA maturation (Yount *et al*., 2008). They can also mediate antiviral host defence activating the host innate immune response (Rehwinkel & Sousa, 2010).

Within the DExD/H box family, RNA helicases share at least eight conserved motifs (I, Ia, Ib, II, and III in the N terminal domain and motifs IV, V, and VI in the C terminal domain) contained within two RecA-like domains joined by a short flexible link (Cordin *et al*., 2006; Jiang *et al*., 2017; Omura *et al*., 2016). These proteins are further distinguished based on variations within their amino acid sequence of the conserved helicase motif II (DEAD, DEAH, DExH and DExD helicases). Both RecA-like domains contribute to ATP hydrolysis and to the binding site for RNA substrates, facilitating the helicase activity of these proteins such as RNA unwinding or protein displacement. In addition, most DExD/H box contain variable N- and C- terminal regions that confer functional specificity to individual helicases (Fuller-Pace *et al*., 2006).

DEAD box RNA helicases are the largest family from DExD/H box family, comprising more than 500 proteins (Silverman *et al*., 2003) and are characterised by the conserved motif II, Asp-Glu-Ala-Asp (DEAD). The motif DEAD provides the RNA helicase function to this protein family by the activities of ATPase and RNA unwinding (Jankowsky, 2011). Furthermore, members of this family are linked to human disease, including cancer, viral infections and hepatitis C (Schütz *et al*., 2010; Jiang *et al*., 2017b).

DEAD-box helicase 41 (DDX41) contains a disordered N-terminal (amino acids 1-152) which are responsible for its nuclear localisation, a DEAD domain that consists of motifs Q, I (Walker A, P-loop), II (Walker B, dead-BOX), Ia, GG, Ib and III and a helicase domain which consists of motifs IV, V and VI (Schmid & Linder, 1992; Caruthers *et al*., 2000; Cordin *et al*., 2006). Motifs Q, I and II are involved in nucleotide biding; motifs Ia, Ib, IV and V are involved in RNA binding and motifs III and VI are involved in ATP hydrolysis (Jiang *et al*., 2017) (Figure 4.1).

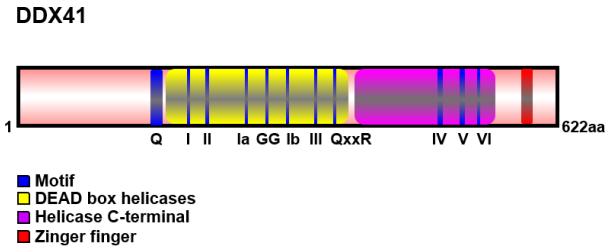


Figure 4.1. Schematic of DDX41 protein.

DDX41 has been shown to participate in the anti-viral innate immunity by functioning as a sensor for cytoplasmic viral DNA in mouse myeloid dendritic cells (Zhang *et al*., 2011b) and also as a germline-encoded pattern recognition receptor when interacting with the bacterial second messengers cyclic di-GMP (c-di-GMP) and cyclic di-AMP (c-di-AMP), which results in the induction of genes involved in the innate immune response in mouse and human cells (Parvatiyar *et al*., 2012). In addition, Polprasert *et al*. (2015) suggested that DDX41 acts as tumour suppressor and demonstrated that alterations in this gene have an impact on mRNA splicing of important downstream genes in leukaemic cells. Furthermore, DDX41 has been implicated in ribosome biogenesis in studies with CD34+ from cord blood cells and leukaemia cell lines

(Kadono *et al*., 2016) and in post-transcriptional regulation of protein translation in studies using colon carcinoma cells (Peters *et al*., 2017).

4.1.3 The role of DExD/H box helicases and DDX41 in the innate immunity response

The recognition of pathogenic DNA from viruses, bacteria, fungi and parasites in the host organism is important to the initiation of innate immune response, which will help the host against the infection of pathogen (Bonjardim *et al*., 2009).

Innate immune response is a conserved, nonspecific first defence of the organism. It includes various defence strategies such as physical barriers (for instance, skin), chemicals in the blood, and immune system cells. In innate cells such as macrophages, innate immunity response is based on the recognition of pathogen-associated molecular patterns (PAMPs) through a set of pattern recognition receptors (PRRs) that stimulate downstream signalling cascades leading to production of pro-inflammatory cytokines and type I interferons (IFN-α and IFN-β), which are cytokines with potent anti-viral activity (Takeuchi & Akira 2010).

The induction of IFN are mediated by activation of the transcription factors interferon regulatory factor 3 (IRF3), interferon regulatory factor 7 (IRF7) or nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), and by signalling pathways downstream of anti-viral PRRs select kinases that phosphorylate and activate IRF3/7 (Sato *et al*., 2000).

Viral PAMPs are mainly genomic RNA, DNA and replication intermediates (Saito & Gale, 2007). The main groups of PRRs sensing PAMPs are endosomal Toll-like

receptors (TLR), the RIG-like helicases (RLHs) which are part of DExD/H-box family, and cytoplasmic DNA receptors (Thompson *et al*., 2011). Most PRRs utilises the kinases TANK binding kinase 1 (TBK1) and inhibitor of κB kinase ε (IKKε) for phosphorylation of IRF3/7 while their upstream signalling is quite diverse. In the upstream signalling, TLR depends on the adaptor molecule TIR-domain-adapter-inducing interferon-β (TRIF) (Yamamoto *et al*., 2002) or on the TIR-domain-adaptor molecule myeloid differentiation primary response gene 88 (MyD88) (Kawai *et al*., 2004), the RLHs utilise a CARD-domain containing mitochondrial adaptor called MAVS (also called IPS-1, Cardif or VISA) (Seth *et al*., 2005), and most cytosol DNA or RNA receptors depend on stimulator of interferon genes (STING), an endoplasmic reticulum-resident adaptor molecule (Ishikawa & Barber, 2008).

Several cytoplasmic DNA or RNA receptors have been identified, including RNA polymerase III, cyclic-GMP-AMP synthetase (cGAS), DNA-dependent activator of IFN-regulatory factors (DAI), gamma-inducible protein 16 (IFI16) and DDX41. Some of these intracellular sensors signal via the adaptor STING (Chiu *et al*., 2009; Sun *et al*., 2013; Takaoka *et al*., 2007; Unterholzner *et al*., 2010).

Hence, DExD/H-box helicases are known to contribute to anti-viral immunity, either by acting as sensors for viral nucleic acids or by facilitating downstream signalling events. The RLH family has the ability to sense double strand RNA (dsRNA) in the cytosol and is constituted by DExD/H-box helicase 58 (DDX58, also known as RIG-I), interferon induced with helicase C domain 1 (IFIH1, also known as MDA5), and DExH-box helicase 58 (DHX58, also known as LGP2) (Schmidt *et al*., 2012). The RLHs recognise their specific RNA ligands through their C-terminal regulatory domain and not the conserved helicase core region (Fullam & Schröder, 2013).

Other DEAD box helicases participate in the innate immune response, such as DEAD-box helicase 3 (DDX3) which acts as a sensor for viral RNA in conjunction with RIG-I and MDA5 (Oshiumi *et al*. 2010). Additionally, DDX3 might function downstream of TBK1 and IKKε either as a signalling adaptor and/or transcriptional regulator (Soto-Rifo *et al*., 2012). DExH-box helicase 9 (DHX9) has been identified as a sensor for dsRNA in myeloid cells (Zhang *et al*., 2011c), and DHX9 and DEAH-box helicase 36 (DHX36) as sensors for CpG oligonucleotides in plasmacytoid dendritic cells (Kim *et al*., 2010). CpG oligonucleotides are known to induce INF-α and pro-inflammatory cytokine production (Krug *et al*., 2001). In myeloid cells, DEAD-box helicase 1 (DDX1) senses dsRNA and then triggers signalling via DExD-box helicase 21 (DDX21) and DDX36 that interact with the downstream protein TRIF to trigger type I interferon responses (Zhang *et al*., 2011a). DExD/H-box helicase 60 (DDX60) is proved to act in conjunction with RIG-I or MDA5 to mediate responses to viral dsRNA (Miyashita *et al*., 2011).

Furthermore, certain bacterial species can release cyclic diguanosine monophosphate (c-di-GMP) or cyclic diadenosine monophosphate (c-di-AMP), which are secondary messengers that regulate the bacteria metabolism, motility and virulence and can also modulate the innate immune response in mammalians cell by acting as PAMPs. c-di-GMP and c-di-AMP can be detected by PRRs and activate the IFN response, however the cytoplasmic RIG-1 (DDX58) nor the Toll-like receptor family of PRRs are involved in the detection of cyclic dinucleotides (Hengge, 2009; McWhirter *et al*., 2009). STING and DDX41 were identified as an innate sensor of cyclic dinucleotides in the cytosol (Burdette *et al*., 2011; Lee *et al*., 2015).

Parvatiyar *et al*. (2012) showed that DDX41 specifically and directly interacts with c-di-GMP via its DEAD domain motif II and motif I, after being phosphorylated by

Bruton's tyrosine kinase (BTK) followed by induction of type I interferon in mice and human cells. Lee *et al*. (2015) using mice and human cells, demonstrated that BTK kinase phosphorylates Tyr414 of DDX41, which is the same site that recognises DNA and bind to STING (Figure 4.2).

DDX41 was also identified as an intracellular DNA sensor in myeloid dendritic cells that depends on STING to sense pathogenic DNA (Zhang *et al*., 2011b). Herein, DDX41 functions through the STING-TBK1-IRF3 pathway, where STING functions as a key scaffolding and adaptor protein to facilitate the signal transduction initiated from upstream cytosolic dsDNA receptors to downstream effectors TBK1, NF-κB and IRF3, leading to the expression of type I interferon (Liu & Wang, 2016) (Figure 4.2).

Uncontrolled sensing of DNA or RNA and excessive production of type I interferon could induce autoimmune diseases, so DDX41 must be degraded or inactivated after immune response. Zhang *et al*. (2013) demonstrated that DDX41 in mice is degraded by the E3 ligase tripartite motif containing 21 (TRIM21), using Lys9 and Lys115 of DDX41 as the ubiquination sites (Figure 4.2).

Interestingly, Lewinsohn *et al*. (2016) and Kadono *et al*. (2016) observed that DDX41 full length protein 70kDa localises in the nucleus, contradicting the DDX41 function as a cytosolic DNA sensor. However, both groups described a short DDX41 isoform of 52kDa translated from the second methionine identified in both nucleus and cytoplasm in human cells. Although several studies showed that DDX41 is important for innate immunity, the exact mechanism of how it functions is still unknown (Jiang *et al*., 2017).

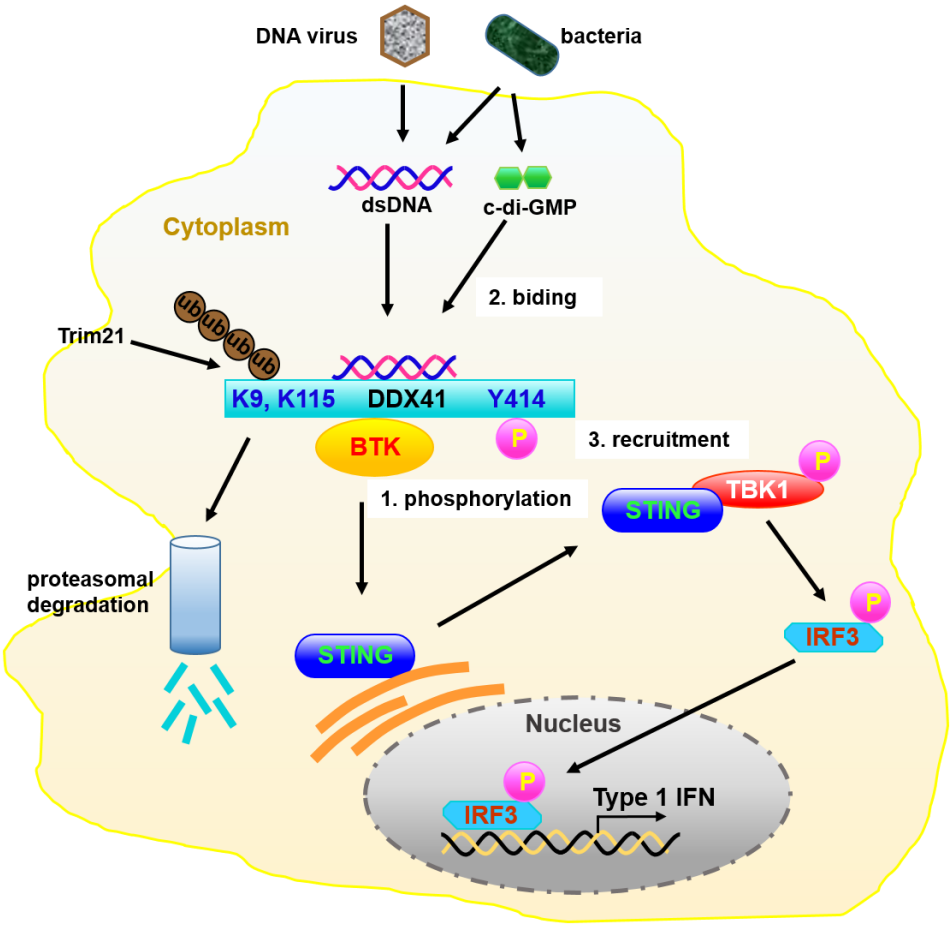


Figure 4.2. DDX41 as an intracellular DNA sensor through the STING-TBK1-IRF3 pathway. Adapted from Lee *et al*. (2015) and Jiang *et al*. (2017).

Finally, in addition to its involvement in viral recognition and anti-viral immunity, DEAD box RNA helicases are also recruited by viruses to facilitate their replication. Some viruses do not encode their own RNA helicases and many cellular helicases have been identified as essential host factors for RNA unwinding in viral replication. The human immunodeficiency virus (HIV) and hepatitis C virus (HCV) are some examples (Lorgeoux *et al*., 2012). DDX1, DDX3 and DEAD-box helicase 5 (DDX5) are required for HCV replication (Owsianka & Patel, 1999; Goh *et al*., 2004; Tingting *et al*., 2006) and DDX1, DHX9, DDX5, DEAD-box helicase 17 (DDX17) and DDX21 have also

been reported to be important for the HIV-1 replication (Fang *et al*., 2004; Naji *et al*., 2012). DHX9 is also required for the replication of the foot-and-mouth-disease virus (Lawrence *et al*., 2009) and the influenza virus (Lin *et al*., 2012).

4.1.4 Association of *DDX41* variants with myeloid neoplasms and defects in mRNA splicing

Besides its roles in innate immunity response, *DDX41* is one of the most frequently mutated familial MDS/AML predisposition gene, with acquired and inherited variants identified (Brown *et al*., 2017). *DDX41* variants profile can be seen in Figure 4.3. In addition to poor survival, normal karyotype, long latency and advanced stage at diagnosis, familial *DDX41* MDS/AML syndrome is also characterised by hypocellular bone marrow, erythroid dysplasia and high risk of MDS and AML (Lewinsohn *et al*., 2016). Furthermore, *DDX41* is located on chromosome 5 and is deleted in approximately 26% of MDS with del(5q) resulting in haploinsufficient expression (Polprasert *et al*., 2015).

Polprasert *et al*. (2015) identified germline and somatic *DDX41* variants in several MDS/AML cases and showed that half of the cases with germline variants also harboured an additional somatic *DDX41* variant in the remaining wild-type allele. The somatic DDX41 R525H was the most frequent acquired variant and it is located in the conserved motif VI at the C-terminal helicase domain where adenosine triphosphate (ATP) hydrolyses occurs, causing a lower ATPase activity in the altered helicase (Kadono *et al*., 2016) (Figure 4.3 and Table 4.1). Biallelic variants were also found with AML progression in MDS/AML with *CEBPA* or *RUNX1* variants (Pabst *et al*., 2008; Preudhomme *et al*., 2009).

In addition, Lewinsohn *et al*. (2016) identified five *DDX41* variants including missense and splicing variants predicted to result in truncated proteins in ten families with MDS/AML (Figure 4.3 and Table 4.1). They identified the recurrent germline D140Gfs*2 variant along with the germline M1I variant, seen subsequently by us and others (Cardoso *et al*., 2016; DiNardo *et al*., 2016; Berger *et al*., 2017).

Further *DDX41* variants were reported by Li *et al*. (2016), they have identified one family with MDS/AML harbouring two consecutive rare heterozygous germline variants (p.L237F and p.P238T) segregating with the disease (Figure 4.3 and Table 4.1). Herein, the disease had a long latency period consistent with previous studies. These variants are located in the DEAD-box domain, affecting a conserved motif that includes the ATP binding site and they are both seen in ExAC database with allele frequency of 1 in 120,916 alleles.

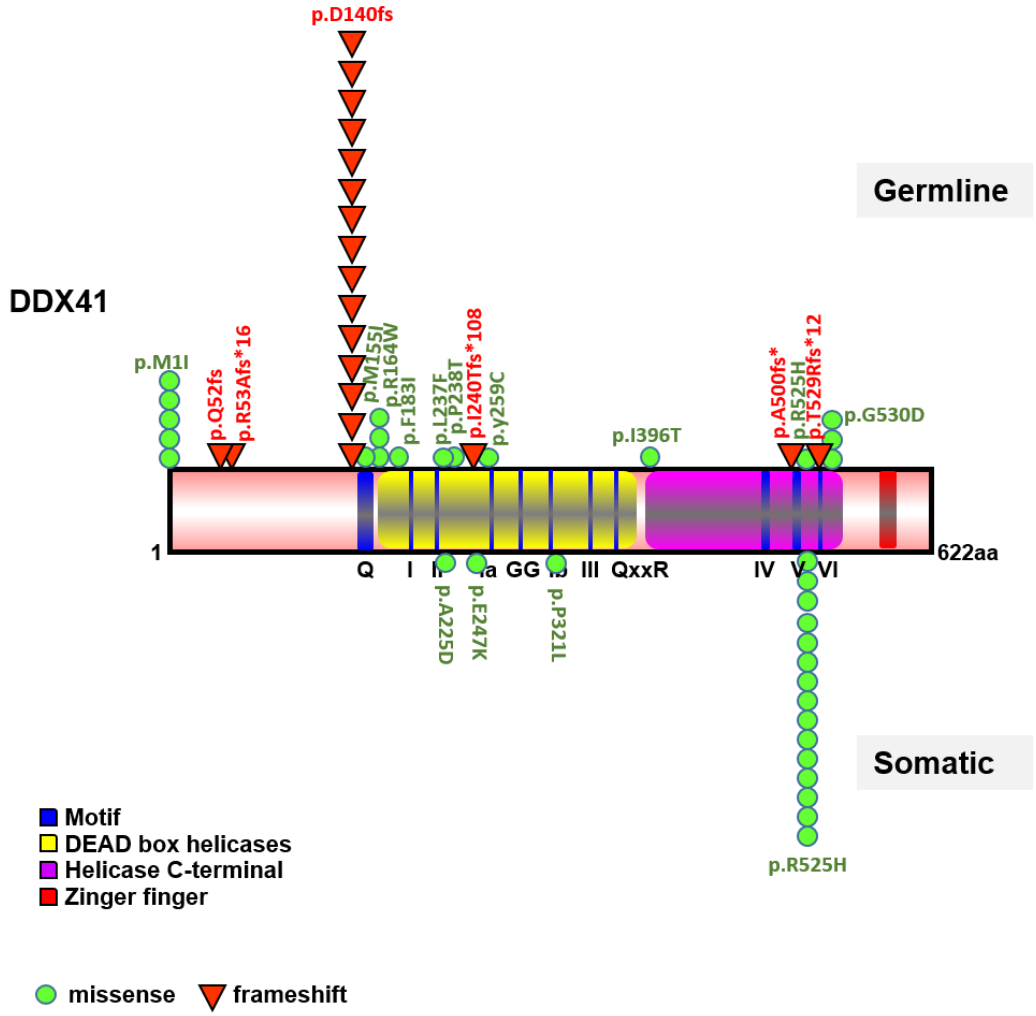


Figure 4.3. Schematic of DDX41 protein with the type of *DDX41* variants identified to date. For further variant details see Table 4.1.

Table 4.1. *DDX41* variants reported to date

Published article	Families	Cases	Age	DDX41 germline variants[#]	ExAC(MAF)	DDX41 somatic variant	ExAC (MAF)	Diagnosis
Polprasert *et al.*, 2015	1	1	70	p.D140fs*2	10 in 121378 (0.00008239)	p.R525H	NR	AML
	1	2	44	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	sAML
	1	3	45	p.D140fs*2	10 in 121378 (0.00008239)	p.R525H	NR	AML
	2	4	64	p.I396T	2 in 121264 (0.00001649)	p.R525H	NR	MDS
	2	5	66	p.I396T	2 in 121264 (0.00001649)	p.R525H	NR	MDS
	3	6	67	p.D140fs*2	10 in 121378 (0.00008239)	p.R525H	NR	MDS
	4	7	73	p.D140fs*2	10 in 121378 (0.00008239)	NA	NA	sAML
	4	8	56	p.D140fs*2	10 in 121378 (0.00008239)	NA	NA	AML
	5	9	72	p.D140fs*2	10 in 121378 (0.00008239)	p.R525H	NR	MDS
	6	10	62	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	MDS
	7	11	65	p.F183I	NR	p.R525H	NR	MDS
		12	85	p.D140fs	10 in 121378 (0.00008239)	p.R525H	NR	sAML
		13	74	p.Q52fs	NR	p.A225D	NR	sAML
		14	58	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	MDS
		15	69	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	CMML
		16	88	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	MDS
		17	71	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	AML
		18	68	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	sAML
		19	78	p.M155I	18 in 121386 (0.0001483)	No	NA	MDS
		20	64	No	NA	p.R525H	NR	5q-syndrome
		21	68	No	NA	p.R525H	NR	MDS
		22	63	No	NA	p.R525H	NR	AML
		23	66	No	NA	p.R525H	NR	MDS
		24	46	No	NA	p.R525H	NR	MDS
		25	78	No	NA	p.P321L	NR	MDS
		26	70	No	NA	p.E247K	NR	5q-syndrome
		27	68	No	NA	splice (e11+1)	NA	AML
Takeda *et al.*, 2015 Abstract 2843	1			p.A500Cfs*	1 in 121292 (0.000008245)	No	NA	
	2			p.Y259C	3 in 119836 (0.00002503)	No	NA	

Table 4.1. Continued

Lewinsohn *et al*. (2016)	1	1	54	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	AML
	2	1	69	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	AML
	3	1	73	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	AML
	4	1	67	c.435-2_435-1delAGinsCA	1 in 121374 (0.000008239)	No	NA	MDS
	5	1	58	p.M1I	18 in 121386 (0.0001483)	No	NA	AML
		2	33	p.M1I	18 in 121386 (0.0001483)	No	NA	CMML
	6	1	72	p.M1I	18 in 121386 (0.0001483)	No	NA	MDS
		2	74	p.M1I	18 in 121386 (0.0001483)	No	NA	AML
	7	1	56	p.R525H	NR	No	NA	MDS
		2	58	p.R525H	NR	No	NA	MDS
		3	55	p.R525H	NR	No	NA	AML
	8	1	53	p.G530D	NR	No	NA	AML
		2	44	p.G530D	NR	No	NA	AML
		3	50	p.G530D	NR	No	NA	AML
	9	1	55	p.R164W	17 in 121384 (0.0001401)	No	NA	Follicular Lymphoma
		2	48	p.R164W	17 in 121384 (0.0001401)	No	NA	Follicular Lymphoma
		3	55	p.R164W	17 in 121384 (0.0001401)	No	NA	Follicular Lymphoma
	10	1	73	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	AML
		2	56	p.D140fs*2	10 in 121378 (0.00008239)	No	NA	AML
Li *et al*. (2016)	1	1	66	p.L237F and p.P238T	1 in 120928 (0.000008269) and 1 in 120916 (0.000008270)	No	NA	AML
		2	59	p.L237F and p.P238T	1 in 120928 (0.000008269) and 1 in 120916 (0.000008270)	No	NA	MDS
		3	70	p.L237F and p.P238T	1 in 120928 (0.000008269) and 1 in 120916 (0.000008270)	No	NA	MDS
Cardoso *et al*., 2016	1	1	77	p.R53Afs*16	NR	No	NA	Asymptomatic
		2	66	p.R53Afs*16	NR	No	NA	MDS
		3	69	p.R53Afs*16	NR	No	NA	MDS
		4	49	p.R53Afs*16	NR	No	NA	MDS
	2	1	60	p.I240Tfs*108	NR	No	NA	AML
	3	1	58	p.T529Rfs*12	NR	No	NA	MDS
		2	NA	p.T529Rfs*12	NR	No	NA	Asymptomatic
	4	1	64	p.M1I and c.-44G>A	18 in 121386 (0.0001483)	No	NA	MDS
		2	41	p.M1I and c.-44G>A	18 in 121386 (0.0001483)	No	NA	MDS

Table 4.1. Continued

Kadono *et al.*, 2016	1	1	64	no	NA	p.R525H	NR	AML
	2	2	72	no	NA	p.R525H	NR	AML
	3	3	81	no	NA	p.R525H	NR	AML
DiNardo *et al.*, 2016	1	1	67	p.M1I	18 in 121386 (0.0001483)	No	NA	AML
		2	64	p.M1I	18 in 121386 (0.0001483)	No	NA	AML
Berger *et al.*, 2017	1	1	58	p.M1I	18 in 121386 (0.0001483)	No	NA	AML
		2	62	p.M1I	18 in 121386 (0.0001483)	No	NA	Asymptomatic
		3	68	p.M1I	18 in 121386 (0.0001483)	p.R525H	NR	AML

[#]: Build GRCh37/hg 19; NR: not registered; NA: not applied; AML: acute myeloid leukaemia; sAML: secondary acute myeloid leukaemia; MDS: myelodysplasia; CMML: chronic myelomonocytic leukemia.

The late onset of familial MDS/AML characteristic of *DDX41* variants make it harder to obtain an accurate diagnosis for this group of patients and misguided selection for an allogenic stem cell transplantation might take place. This is the case reported by Berger *et al*. (2017), where the proband of 58 years presenting AML had a blood stem cell transplantation from his healthy brother of 62 years at the time, followed by the proband achieving full recovered blood cells count. However, the proband had a relapse into MDS four years later to discover that the MDS clone emerged from the donor cells. Meanwhile, another brother carrying the recurrent DDX41 p.R525H variant was diagnosed with AML at the age of 68 years. The recurrent DDX41 p.Met1Ile variant was identified in all three brothers.

Polprasert *et al*. (2015) have demonstrated the tumour suppressor role of DDX41 from their *DDX41* knockdown experiments in K562 (an immortalised human myelogenous leukemia cell line derived from a chronic myelogenous leukemia patient) and $CD34^{+}$ cells results, which displayed an increase of cell growth compared to control. Further experiments in HEK293 (human embryonic kidney cells 293) cells using mass spectrometry identified spliceosomal proteins to associate and interact with DDX41 and this was then confirmed by western blotting of endogenous DDX41 protein from primary and K562 cells.

In addition, data from the mass spectrometry experiment mentioned above showed that the R525H variant reduced the DDX41 interaction with splicing factor 3b subunit 1 (SF3B1) and pre-mRNA processing factor 8 (PRPF8), crucial components in spliceosomes. SF3B1 together with SF3B2 and SF3B3 forms the U2 small nuclear ribonucleoproteins complex (U2 snRNP) and somatic genetic alterations in SF3B1 gene has been linked to MDS and chronic lymphocytic leukaemia patients by induction of aberrant transcription, altered pre-mRNA recognition and alternative

splicing (Cazzola *et al*., 2013; Jenkins & Kielkopf, 2017). While somatic genetic alteration in PRPF8 has been associated with MDS and AML, where decrease of PRPF8 expression is associated with increased exon skipping, probably as a result of splicing proofreading defect. PRPF8 is a scaffold protein component of both U2- and U12-dependent spliceosomes and is required for the assembly of the U4/U6-U5 tri-snRNP complex (Kurtovic-Kozaric *et al*., 2015).

Moreover, Polprasert *et al*. (2015) performed deep whole RNA sequencing in three cases that had deletion (5q-), one case with mutant (p.R525H), and one case that showed low expression of *DDX41*, and 11 wild-type cases. This demonstrated that defects in RNA splicing were associated with the altered DDX41 that resulted in aberrant exon skipping of 61 genes and exon retention in 95 genes. Zinc finger MYM-type containing 2 (*ZMYM2*) was one of the top ten genes that had the most differentially misspliced exons. *ZMYM2* encodes a zinc finger protein member of a family of myeloproliferative and mental retardation (MYM) domain involved in a histone deacetylase complex (Gocke & Yu, 2008). The difference of skipping ratio in *ZMYM2* exon 3 was 13% between mutant and wild-type. The enhanced skipping of this exon, was recapitulated by DDX41 knockdown in K562 and CD34+ cells. In contrast, overexpression of wild-type DDX41 in U937 cells (monocytes isolated from the histiocytic lymphoma of a male patient) led to decreased exon skipping of ZMYM2 in comparison to mock transduction. In addition to the changes in spliced isoform ratios, ZMYM2 mRNA was expressed at significantly lower levels in the cells with low DDX41 expression. Low ZMYM2 mRNA levels were associated with down-modulation of structural maintenance of chromosomes 3 (SMC3), RAD21 cohesin complex component (RAD21), and RUNX1, which were also significantly under-expressed in samples with low DDX41 expression.

Overexpression as well as variants or deletions of other members of DExD/H-box RNA helicase family have been described in several other cancers as well as in myeloid neoplasms, where they can act both as tumour suppressors and oncogenes (Fuller-Pace, 2013).

Finally, DDX41 and other members of the DEAD/H-box family gene variants are identified functionally relevant as a novel family of variants with implications for prognosis and treatment of myeloid malignancies, which may lead to approaches of therapeutic schedule (Antony-Debré & Steidl, 2015). Nonetheless, the role of DDX41 in haematopoiesis and leukemogenesis is not yet clear and further studies are required to better explain its function and molecular pathway (Jiang *et al*., 2017).

4.1.5 The role of DDX41 in ribosome biogenesis and in post-transcriptional regulation of protein translation in cell growth

In addition to mRNA splicing deficiency due to defective DDX41 demonstrated by Polprasert *et al*. (2015), Kadono *et al*. (2016) reported three patients harbouring the somatic DDX41 R525H variant that commonly exhibited sporadic AML with bone marrow cytopenias and low blast counts as also seen in Lewinsohn *et al*. (2016). They suggested that the cell growth arrest seen in cells harbouring this somatic variant in their experiments is due to impaired pre-ribosome RNA (rRNA) processing by the loss of ATPase activity of the protein, suggesting that DDX41 plays a role in ribosome biogenesis in human cells.

To investigate the molecular functions of DDX41 in hematopoietic cells, Kadono *et al*. (2016) firstly observed that ectopically expressed DDX41 in murine fibroblasts and in THP- leukaemia cell line (human monocytic cell line derived from an acute monocytic leukaemia patient) was mostly nuclear regardless of the p.R525H presence. Then, they demonstrated that this acquired variant inhibits haematopoietic cell cycle growth in mutant cord blood-derived human $CD34^+$ cells compared to the wild-type cells control. Subsequently, data from a gene set enrichment analysis in the cultured mutant and wild-type cord blood-derived human $CD34^+$ cells, suggested that a certain ribosomopathy may occur in the cells expressing DDX41 pR525H as there was a suppression of mRNA encoding ribosomal proteins. They have also demonstrated that the altered DDX41 presented a lower ATPase activity by an ATPase assay.

In humans, the processing of pre-rRNA occurs mainly in the nucleolus, where approximately 4,500 putative nucleoclar proteins and small nucleolar RNAs are thought to participate in this process (Ahmad *et al*., 2009). Kadono *et al*. (2016) proposed that the haematopoietic cell growth deficiency in patients with DDX41 R525H variant is caused by the inhibition of E2F transcription factor 1 (E2F) activity, which is caused by an activation of retinoblastoma tumour suppressor (RB). The RB-E2F pathway regulates the cell cycle progression and cell death and consists of inhibitors and activators of cyclin-dependent kinases, the RB, and the E2F-family of transcription factors. According to Kadono *et al*. (2016), the somatic DDX41 R525H variant negatively affects rRNA synthesis, which consecutively releases ribosomal proteins. Released ribosomal proteins eventually bind to MDM2 proto-oncogene (MDM2) and RB pathway is consecutively activated, resulting in cell growth arrest.

On the other hand, data from Peters *et al*. (2017) studies in HCT116 colon carcinoma cells, supports an oncogene role for DDX41. They identified DDX41 as a negative

regulator of cyclin-dependent kinase inhibitor 1A (CDKN1A) or p21 protein translation via association with the 3'UTR of its mRNA in the presence of p53 independently of stress, however this mechanism is not associated with the transcriptional activity of p53. p21 functions as tumour suppressor in several cancer cell lines, including HCT116.

p21 is the most studied cyclin-dependent kinase protein inhibitor. It tightly controls cyclin-dependent kinase proteins, which are serine/threonine kinases that regulates cell cycle. p21 regulates cell growth and apoptosis in the presence of the tumour suppressor protein p53 in response to a variety of stress stimuli. p21 also functions both as tumour suppressor and as oncogene important in stress pathways. Although nuclear p21 functions predominantly as a tumour suppressor by negatively regulating DNA replication and cell proliferation, cytoplasmatic p21 acts in an oncogenic manner by facilitating cell proliferation, inhibiting apoptosis, and regulating migration (Malumbres & Barbacid, 2009).

Interestingly, Peter *et al*. (2017) also found that DDX41 requires its helicase activity to regulate p21 expression at the translational level by generating a DDX41 missense variant (G521S) in the motif VI (which is essential for ATP hydrolysis and, thus, helicase activity of DEAD-box proteins) and comparing luciferase activities in response to either the mutant or the wild-type DDX41 protein in HCT116 cells.

Finally, other DEAD box proteins such as DDX3 and DDX5 have also been demonstrated to control expression of p21. They transcriptionally up-regulate p21 indirectly in a p53-dependent manner (Wu *et al*., 2011; Nicol *et al*., 2013). Furthermore, as observed in DDX41 and p21, DDX3, DDX5 and DDX17 have been described functioning as both tumour suppressor and as oncogene under different

intracellular conditions, depending on the cancer type, treatment modalities, and several co-factors (Chao *et al*., 2006; Yang *et al*., 2007; Botlagunta *et al*., 2008; Germann *et al*., 2012). Nonetheless, further investigation is required to clarify how defected DDX41 in the cells could result into the development of haematological malignancies (Kadono *et al*., 2016).

4.2 Results

We have undertaken a combination of whole exome and targeted sequencing to characterise 55 families from our independent cohorts (DC, idiopathic BMF and familial MDS/AML registries). The targeted sequencing employed a newly designed familial MDS/AML gene panel which included the nine genes, where germline heterozygous variants have been identified in association with development of familial MDS/AML to date. These are *RUNX1*, *CEBPA*, *TERC*, *TERT*, *GATA2*, *SRP72*, *ANKRD26*, *ETV6*, and *DDX41*. This analysis has enabled us to identify four families harbouring heterozygous germline *DDX41* variants (Figure 4.4A-D); three families have novel frameshift variants (c.155dupA, c.1586_1587delCA and c.719delTinsCG) and the fourth family has a recurrent missense variant in the initiation codon (c.3G>A, rs141601766) described previously by Polprasert *et al*. (2015) and Lewinsohn *et al*. (2015). Collectively these four families comprise seven cases of MDS and two cases of AML (age range, 40 to 70 years). These patients did not have any extra-haematopoietic features and therefore represent “pure” MDS/AML (Table 4.2).

Table 4.2. Characteristics and family history of index cases

Family	Case	Age (years)	Diagnosis	Relationship to index	Nucleotide	Amino acid
1	I-1	NA	asymptomatic	grandmother	NA	NA
	I-2	NA	asymptomatic	grandfather	NA	NA
	II-1	77	asymptomatic	maternal aunt	c.155dupA	p.Arg53Alafs*16
	II-2	66	MDS	maternal aunt	c.155dupA	p.Arg53Alafs*16
	II-3	69	MDS	maternal uncle	c.155dupA	p.Arg53Alafs*16
	II-4	NA	asymptomatic	mother	NA	NA
	II-5	NA	CML	father	NA	NA
	III-1	49	MDS	index case	c.155dupA	p.Arg53Alafs*16
2	I-1	NA	AML	mother	NA	NA
	I-2	NA	asymptomatic	father	NA	NA
	II-1	60	AML	index case	c.719delTinsCG	p.Ile240Thrfs*108
3	I-1	NA	MDS	mother	NA	NA
	I-2	NA	stomach cancer	father	NA	NA
	II-1	NA	asymptomatic	husband	NA	NA
	II-2	58	MDS	index case	c.1586-1587delCA	p.Thr529Argfs*12
	II-3	NA	asymptomatic	sister	NV	NV
	II-4	56	tongue cancer	brother	NV	NV
	III-1	NA	asymptomatic	son	NV	NV
	III-2	NA	asymptomatic	daughter	c.1586-1587delCA	p.Thr529Argfs*12
4	I-1	NA	asymptomatic	mother	NA	NA
	I-2	64	MDS	father	c.3G>A	p.Met1Ile
					c.-44G>A Met1Ile	NA
	II-1	41	MDS	index case	c.3G>A	p.Met1Ile
					c.-44G>A Met1Ile	NA

MDS: myelodysplastic syndrome; AML: acute myeloid leukaemia; CML: chronic myeloid leukaemia; NA: not available; NV: does not have the variant.

4.2.1 *DDX41* germline variants identified in our cohort

In Family 1 (Figure 4.4A), a novel heterozygous germline variant c.155dupA (p.Arg53Alafs*16 showed in Figure 4.4E) in *DDX41* was identified in the 49 year-old female index case (III:1, Figure 4.4A) diagnosed with MDS, refractory anaemia with excess blasts (RAEB). Sanger sequencing revealed that her maternal aunt and uncle who both developed RAEB also harbour this frameshift variant (individuals II:3 and II:2, Figure 4.4A respectively). There are two asymptomatic carriers (individuals II:1 and II:4, Figure 4.4A), supporting previous observations that haploinsufficiency for *DDX41* shows variable penetrance (Lewinsohn *et al*., 2015). Further family history included her father who died of chronic myeloid leukaemia (CML), unlikely to be related to the *DDX41* variant.

In Family 2 (Figure 4.4B), the index case is a 60 year-old male (II:1, Figure 4.4B) with AML harbouring a novel heterozygous frameshift variant c.719delTinsCG (p.Ile240Thrfs*108), predicted to cause truncation of the protein and consequent loss of function. His mother died of AML (I:1, Figure 4.4B). Segregation analysis was not possible as there were no family samples available, however the variant allele frequency in the index case is 0.494 indicating heterozygosity. This variant is located in the DEAD box domain of DDX41, in a highly conserved motif that includes the ATP binding site (Figure 4.4E).

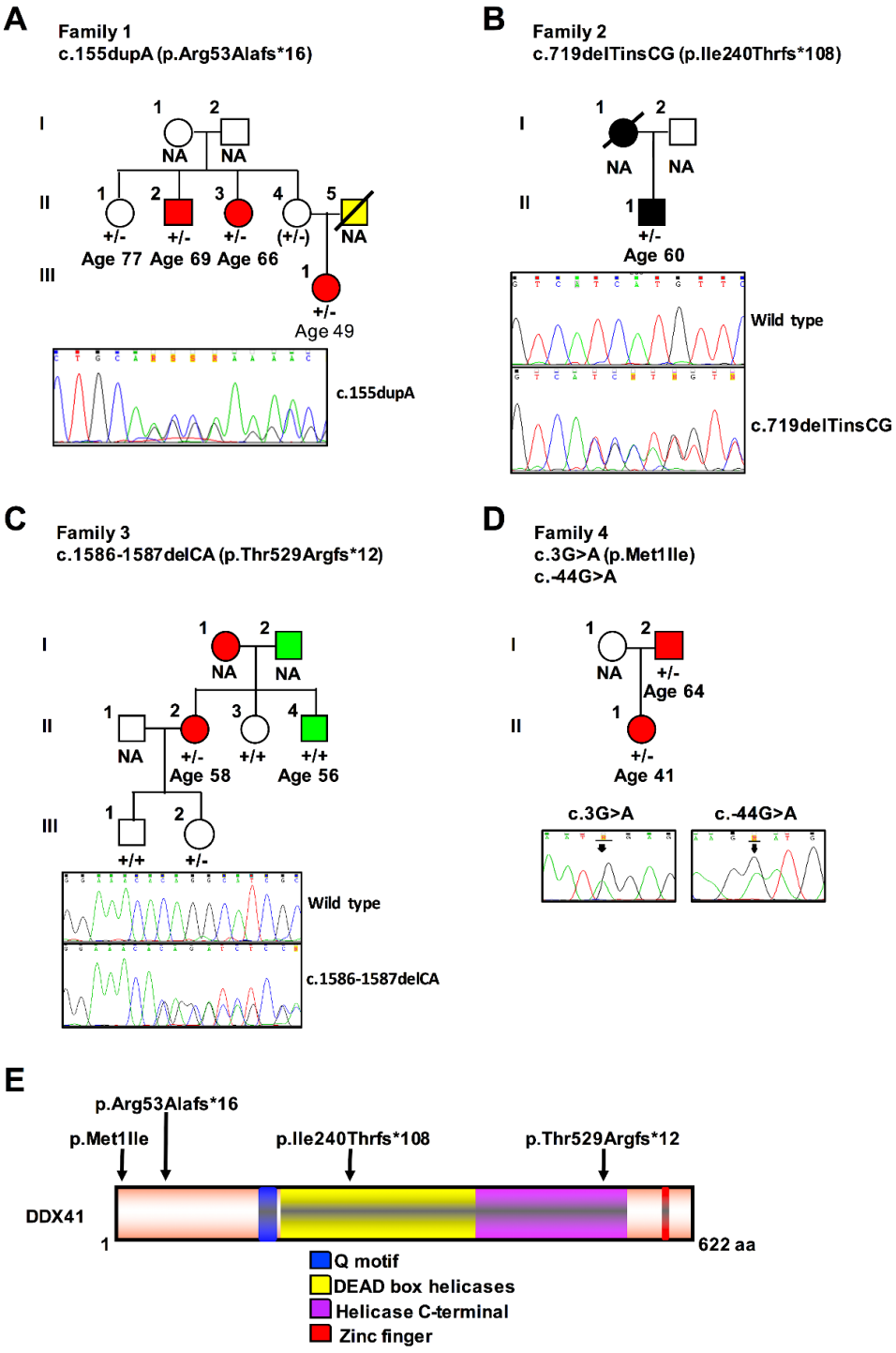


Figure 4.4. Familial MDS/AML caused by LoF *DDX41* variants. A-D. Families with MDS/AML with variants in *DDX41*, their age at diagnosis and their respective Sanger sequencing traces. Affected individuals are coloured as follows: red, MDS; yellow, CML; black, AML; and green, other non-haematological cancer. E. Schematic of DDX41 protein showing the heterozygous variants identified in this study. CML, chronic myeloid leukaemia.

The 58 year-old female index case in Family 3 (II:2, Figure 4.4C) with MDS, has a novel frameshift deletion variant c.1586-1587delCA (p.Thr529Argfs*12) in the helicase domain of DDX41 (Figure 4.4E), which is again predicted to cause truncation of the protein. Her brother had tongue cancer (II:4, Figure 4.4C), her mother had MDS (I:1, Figure 4.4C) and her father had stomach cancer (I:2, Figure 4.4C). In the absence of samples of the index case's parents, Sanger sequencing was undertaken on samples from her siblings and children. The siblings (II:3 and II:4, Figure 4.4C) of the index case do not harbour the variant c.1586-1587delCA, whilst her daughter (III:2, Figure 4.4C) is an asymptomatic carrier. This suggests that the index case and her mother (both with MDS) have disease associated with the *DDX41* variant, while the non-haematological cancers seen in her brother (II:4, Figure 4.4C) and father (I:2, Figure 4.4C) are unrelated to *DDX41*.

The index case of Family 4 (Figure 4.4D) is a 41-year-old female (II:1, Figure 4.4D) diagnosed with MDS/RAEB. Her father (I:2, Figure 4.4D) was also diagnosed with MDS at age 64 years. The heterozygous missense variant c.3G>A (p.Met1Ile – rs141601766, showed in Figure 4.4E) in *DDX41* which segregated with disease in these two individuals has been reported in ExAC database in 6/117,464 alleles (http://exac.broadinstitute.org/, accessed 15th March 2016). Interestingly, both cases with the c.3G>A variant also carried a linked 5'UTR variant (c.-44G>A showed in Figure 4.4D) previously observed by Lewinsohn *et al*. (2015). They also demonstrated that HEK293 cells ectopically expressing the Met1Ile mutant protein used an alternative translation initiation site yielding a smaller cytoplasmatic DDX41 protein when compared to the nuclear full-length of 70kDa. Their experiments suggest that this isoform may occur naturally and has an altered location.

4.2.2 Telomere length analysis in patients harbouring *DDX41* variants

Kirwan *et al*. (2009) demonstrated that *TERC*/*TERT* familial MDS/AML patients have a significant shorter telomere compared to controls. To investigate whether *DDX41* familial MDS/AML families present with short telomeres, we measured the telomere length by monochrome multiplex quantitative PCR method (Cawthon, 2009) in the *DDX41* MDS/AML patients which we had available sample. We found a slightly shorter telomere length in these patients compared to controls ($p<0.05$, Figure 4.5). However, there is no evidence supporting the role of DDX41 in telomere maintenance to date.

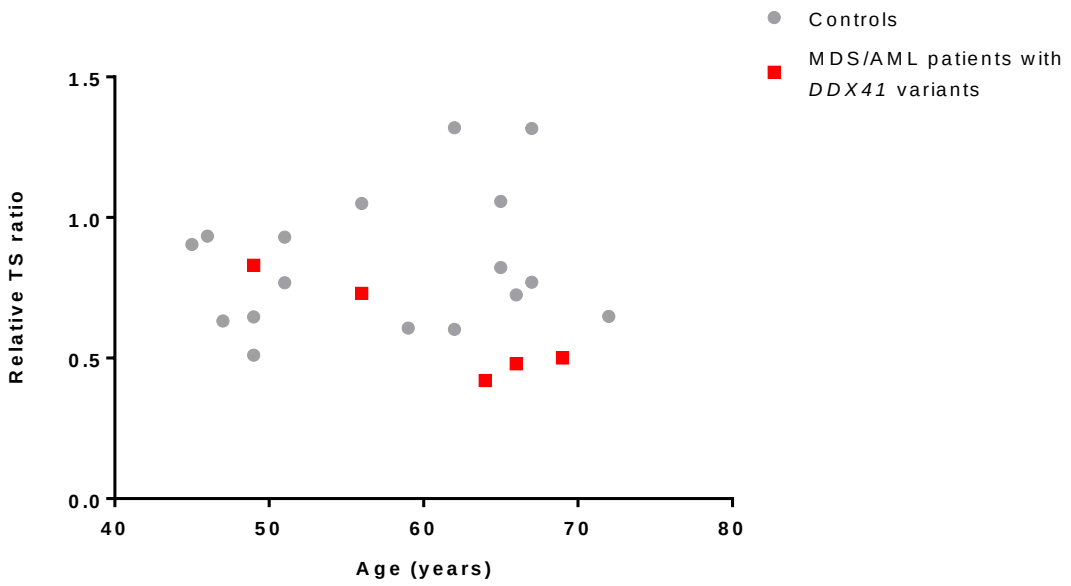


Figure 4.5. Telomere lengths are slightly shorter in affected individuals ($p<0.05$). Telomere lengths were measured by MMqPCR using matched age control.

4.3 Discussion

Here, we reported four new MDS/AML families harbouring LoF variants likely to be disease causing in *DDX41* gene with probands presenting slightly short telomeres when compared to controls. Further functional studies would be necessary to confirm pathogenicity of the variants in these families. Three families presented novel frameshift variant causing premature stop in the protein and one is a recurrent variant missense in the start transcription codon. Most of the variants found in *DDX41* to date are heterozygous LoF that cause truncated protein, suggesting that *DDX41* familial MDS/AML is caused by haploinsufficiency. Additionally, the frequent occurrence of asymptomatic carriers demonstrates the variable penetrance of *DDX41* variants as well as its association with the long latency period of this disease. Indeed, *DDX41* has only recently been associated with the development of familial MDS/AML and the identification of more families with germline *DDX41* variants is important to further clarify the prevalence and penetrance of these variants, as well as the prognosis of individuals that develop the disease (Li *et al*., 2016).

The recurrence of the Met1Ile variant in the ExAC database poses an interesting question as to the causative role of *DDX41* variants in MDS/AML. Excluding any non-canonical and dubious calls in this database, LoF variants (including Met1Ile) are seen to occur at a cumulative frequency of 1 in 1,189 people (46 LoF variants in an average of 109,354 alleles). This is in stark contrast to the few LoF variants reported in *RUNX1* (7)*, CEPBA* (0)*, GATA2* (0) and *ETV6* (4). We also note that in a screen of 1,045 patients with MDS and secondary AML, 16 patients (1 in 65) had germline LoF variants in *DDX41* (Polprasert *et al*., 2015)*.* These data indicate that rather than establishing a causal Mendelian link between germline LoF *DDX41* variants and MDS/AML, it is better to think of them as genetic risk factors. Comparing the frequency

of LoF *DDX41* variants seen in MDS and secondary AML with the frequency seen in ExAC database we obtain an odds ratio of 8.05 ($p = 5.65x10^{-5}$, Fisher's Exact Test). Allowing for a 1/100 probability of getting the disease, this would translate to a relative risk of 7.51. It is inevitable therefore, that MDS/AML driven by *DDX41* LoF variants will sometimes appear as familial.

Patients described in this study presented median age at diagnosis of 58 years corroborating with the observation that *DDX41* families with MDS/AML develop the disease at an age characteristic of de novo disease and this late presentation makes it difficult to distinguish hereditary factors from aging and cumulative environmental exposures (Sekeres, 2010). The late onset of MDS/AML cases suggests that germline variants in *DDX41* alone does not cause MDS/AML. This is supported by the incidence of several other somatic variants in genes known to cause sporadic hematologic malignancies and familial MDS/AML such as *TP53*, *RUNX1*, *LUC7L2*, *DNMT3A* and *ASXL1* alongside to the predisposition for somatic *DDX41* variants in patients carrying germline *DDX41* variants seen in Polprasert *et al*. (2015) and Berger *et al*. (2017). Polymorphisms in *GATA2*, *TERT*, *ANKRD26*, *ASXL1* and *DNMT3A* were also found in our patients carrying *DDX41* germline variants. In addition, population-based studies have revealed that clonal haematopoiesis might occur during aging, whereby *TET2*, *DNMT3A*, *TP53* and *ASXL1* are the most altered genes (Jaiswal *et al*., 2014).

Works on DDX41 function have so far implied the involvement of this RNA helicase in different pathways such as immune response, mRNA splicing, ribosome biogenesis and regulation of protein translation. Intriguingly, none of the reported *DDX41* MDS/AML families presented impaired immune system. As consequence, DDX41 involvement in immune response is yet to be linked to a disease phenotype.

Despite of the genetic underlying cause of familial MDS/AML in patients with germline variants in *DDX41*, its precise mechanism is far to be elucidated. Polprasert *et al*. (2015) using different assays in myeloid and primary cells, demonstrated a tumour suppressor role for DDX41 along with the findings that DDX41 interacts with several spliceosome proteins and that the somatic *DDX41* variant p.R525H altered the native DDX41 interactome especially for major components in U2 and U5 spliceosomes. Furthermore, they demonstrated that DDX41 defects impaired mRNA splicing of several genes in patient-derived cells. Additionally, while investigating the cause of sporadic AML in three patients harbouring the somatic *DDX41* p.R525H variant, Kadono *et al*. (2016) demonstrated as expected that this somatic variant is responsible for the loss of ATP hydrolysis of DDX41 as well as for the cell growth arrest in mutated CD34+ cells from cord blood. This cell cycle inhibition was caused by a disruption of pre-rRNA synthesis. It is not yet clear how this disruption occurs, although their data from a gene set enrichment analysis from mutated CD34+ and patient-derived cells showed that the cell cycle was inhibited by the suppression of E2F activity through the RB-E2F pathway. They concluded that this event coupled with age-dependent epigenetic alterations or additional somatic variants might collaborate with the somatic DDX41 p.R525H to cause AML in the analysed patients.

Conversely, Peters *et al*. (2017) identified DDX41 as negative regulator of p21 mRNA translation p53-dependent and independently of stress suggesting an oncogene role for DDX41 as p21 functions in several cancer cell lines in an antiapoptotic manner. This result is consistent with data from a genome-scale RNA-mediated interference screen in HeLa cells (human epithelial cells) demonstrating reduced cell numbers following knockdown of *DDX41* (Kittler *et al*., 2007). Although being contradictory to Polprasert *et al*. (2015) results, other DExD/H-box RNA helicases were described as oncogene or tumour suppressor depending on the cancer type, treatments modalities,

and several co-factors. Comparatively, DDX5 transcriptionally up-regulates p21 indirectly in a p53-dependent manner (Bates *et al*., 2005; Chao *et al*., 2006) exhibiting tumour suppressor activity (Nicol *et al*., 2013) whilst demonstrating oncogenic functions when up-regulating pro-proliferative genes such as cyclin D1 (*CCND1*) and MYC proto-oncogene, bHLH transcription factor (*MYC*), as well as genes required for DNA replication (Yang *et al*., 2007). Interestingly, Polprasert *et al*. (2015) identified variants in other members of the DExD/H-box RNA helicase family in about 4% of patients in their cohort, while we have identified rare variants in these genes in 18% of ours suggesting that RNA helicase variants represent an entire new family of variants in myeloid neoplasms.

Furthermore, Peters *et al*. (2017) demonstrated that DDX41 requires its helicase activity to regulate p21 expression at the translation level. Thus, the somatic p.R525H variant is possibly hypomorphic as it severely affects the helicase activity of DDX41 based on its location. Therefore, this somatic variant probably disrupts most of the putative DDX41 functions described to date.

In summary, we reported on novel germline heterozygous LoF *DDX41* variants exhibiting variable penetrance in families with MDS/AML and tendency to short telomeres. Our analysis suggests that rather than establishing a causal Mendelian link between *DDX41* germline LoF variants and MDS/AML it is appropriate to consider these as genetic risk factors. Furthermore, additional studies are required in order to clarify DDX41 function and its role in haematopoiesis and leukaemogenesis.

Chapter 5

RTEL1 variants leading to myelodysplasia and liver disease

5.1 Introduction

Due to the fact that there is a considerable clinical heterogeneity and overlapping features seen in patients with dyskeratosis congenita and Hoyeraal Hreidarsson syndrome and that these are bone marrow failure disorders that can develop into myelodysplasia and/or acute myeloid leukaemia, we analysed our cohort of patients with dyskeratosis congenita, Hoyeraal Hreidarsson, aplastic anaemia, myelodysplasia and/or acute myeloid leukaemia families in search of disease causing variants in genes in patients with these phenotypes.

As a result, in this chapter we associate heterozygous loss of function variants in the telomere maintenance gene regulator of telomere elongation helicase 1 (*RTEL1*), with liver disease and myelodysplasia, for the first time. Previously germline biallelic variants in *RTEL1* have been shown to cause dyskeratosis congenita and Hoyeraal Hreidarsson syndrome and germline heterozygous variants in this gene are known to cause pulmonary fibrosis (which is present in approximately 20% of dyskeratosis congenita patients). This study therefore extends the range of phenotypes associated with germline *RTEL1* variants by adding myelodysplasia and liver disease to the clinical spectrum of patients.

5.1.1 RTEL1 protein structure and function

RTEL1 is an iron-sulphur (FeS) ATP-dependent DNA helicase with 1,243 amino acids, classified as a RAD3-related helicase, which belongs to the DEAH subfamily of the Superfamily 2 (SF2) helicases. The Rad3-related DNA helicase is located in

the N-terminal side of the protein and consists of four domains: two Rec-A-like motor domains; helicase domain (HD1 and HD2); a Fe-4S cluster; and an ARCH domain (Figure 5.1). RTEL1 is part of a subclass of FeS cluster-containing DNA helicases known as XPD family, which includes the proteins Xeroderma pigmentosum group D (XPD), DEAD/H-box helicase 11 (DDX11) and Fanconi anemia group J protein (FANCJ) (Rudolf *et al*., 2006).

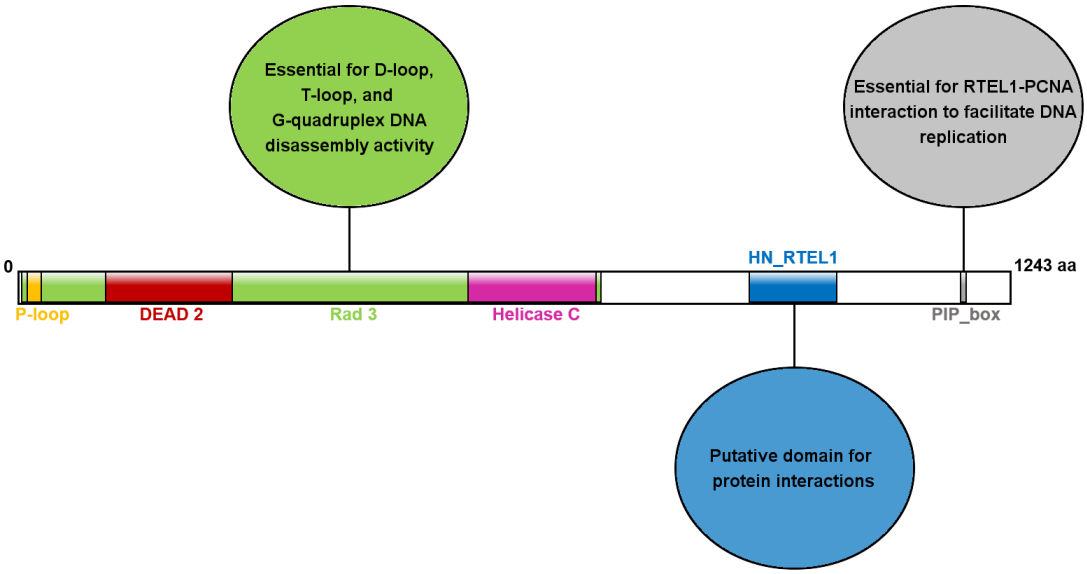


Figure 5.1. Schematic of RTEL1 protein. Putative functions of RAD3, Hamonin-N-like (HN_RTEL1) and PIP domains are shown (NP_116575; ENST00000508582).

The XPD family of helicases are characterised by a FeS cluster composed of four conserved cysteine residues attached to iron ions. Studies in XPD protein revealed that its helicase activity is lost when the FeS domain is removed and that this domain recognises single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) (Pugh *et al*., 2008). Therefore, it is likely that the other FeS cluster containing helicases,

including RTEL1 also recognise ssDNA and dsDNA based on the sequence similarity of the helicase core domain (Uringa *et al*., 2011).

In addition to its helicase motifs and FeS domain, RTEL1 has a conserved eight amino acid PIP box (Warbrick, 1998) (Figure 5.1). PIP box is a proliferating cell nuclear antigen interacting protein (PCNA) interaction motif and PCNA is a highly conserved eukaryotic protein that functions in DNA replication and acts as a cofactor for DNA polymerases.

Variants in *XPD*, *FANCJ* and *DDX11* cause the genetic disorders xeroderma pigmentosum (XP) (Andressoo *et al*., 2005), Fanconi anaemia (Bridge *et al*., 2005) and Warsaw breakage syndrome (van der Lelij *et al*., 2010), respectively. Furthermore, variants in *RTEL1* cause dyskeratosis congenita (Walne *et al*., 2013a), Hoyeraal Hreidarsson syndrome (Ballew *et al*., 2013; Deng *et al*., 2013; Le Guen *et al*., 2013), familial pulmonary fibrosis (Kannengiesser *et al*., 2015; Cogan *et al*., 2015) and myelodysplasia and liver disease (Cardoso *et al*., 2017; work presented in this chapter).

RTEL1 is essential in maintaining genome stability by disassembling DNA secondary structures formed during DNA repair, DNA recombination, and DNA replication and it is also essential in telomere maintenance (Barber *et al*., 2008; Vannier *et al*., 2014)**.** RTEL1 interrupts D-loop formation in homologous recombination upon DNA double-strand breaks (DSBs) formation (Barber *et al*., 2008; Uringa *et al*., 2011; Vannier *et al*., 2012) and it is also crucial for the disruption of G-rich DNA secondary structures and T-loops during DNA replication, thereby protecting telomere length (Ding *et al*., 2004; Vannier *et al*., 2012; Kannengiesser *et al*., 2015).

5.1.2 RTEL1 in homologous recombination

The stability of the genome is critically dependent on the coordinate action of DNA repair pathways during cell cycle (Chapman *et al*., 2012). Homologous recombination (HR) is an essential conserved process for dividing cells. In mitosis, HR is required not only for the accurate repair of DNA DSBs but also for the restart of stalled replication forks. In meiosis, HR is crucial for DSB repair and limitation of excessive crossing over, which is required for accurate chromosome segregation at the first meiotic division (Youds *et al*., 2010).

In HR, the sister chromatid or the homologous chromosome is used as a template for repair through synthesis-dependent strand annealing pathway (SDSA). This involves temporary engagement of a homologous DNA duplex that serves as an information donor by acting as a template for DNA synthesis at the repair site (Pâques & Haber, 1999).

Studies have shown that in HR DNA repair, a DSB is rearranged to produce 3' single-stranded DNA tails that are bound by the DNA strand exchange protein RAD51 to form a nucleoprotein filament (Sung *et al*., 2003). These filaments are the catalyst for strand invasion into homologous duplex DNA, resulting in the formation of a displacement-loop (D-loop) structure (Kasamatsu *et al*., 1971) (Figure 5.2).

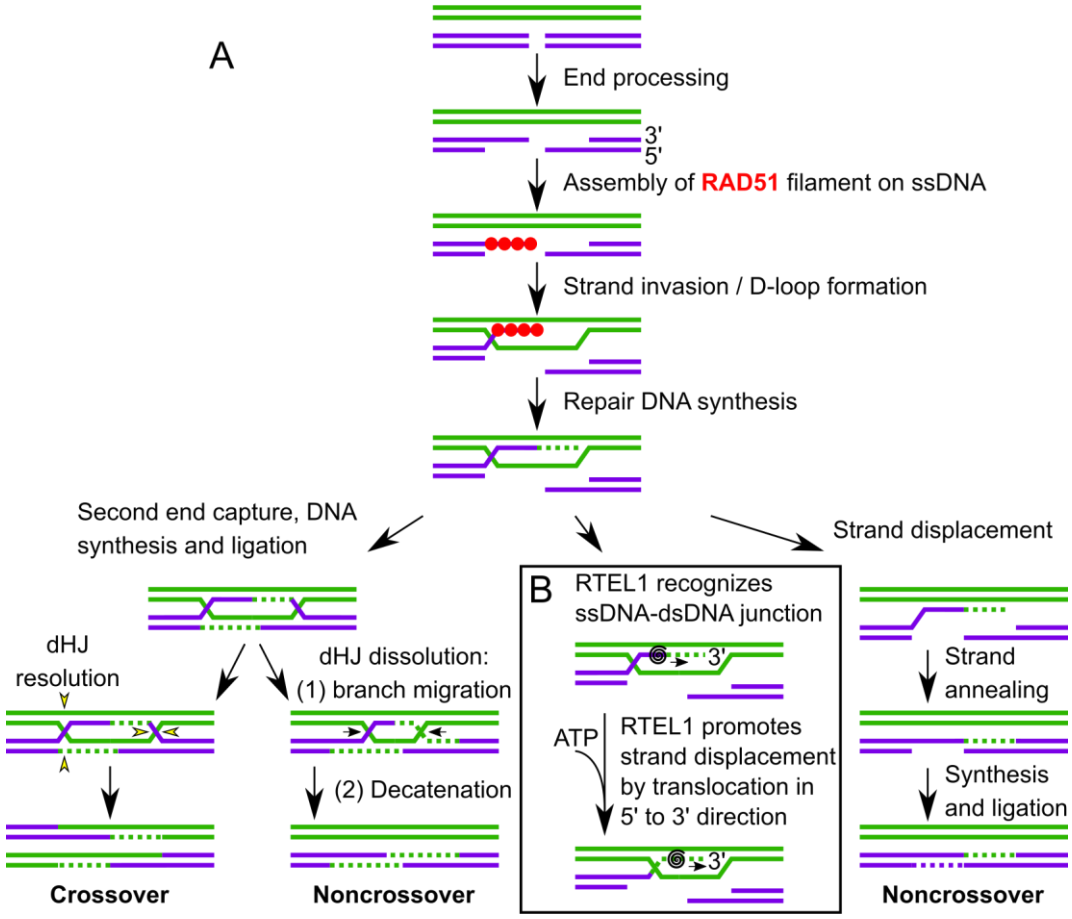


Figure 5.2. RTEL1 role in homologous recombination. A. Homologous recombination pathways of double-strand break repair. B. Model for RTEL1 promoting synthesis-dependent strand annealing producing a non-crossover repair product. Adapted from Villeneuve (2008) and Uringa *et al*. (2010).

The invading 3' end forms a D-loop and provides a primer for DNA synthesis, which can be resolved either through displacement of the invading strand from the D-loop and annealing to the other DSB end (synthesis-dependent strand annealing), resulting in a non-crossover repair product, or by the capture of the other resected end by the altered strand of the D-loop to form a double Holliday junction (dHJ). HR can be completed by endonucleolytic cleavage of the two HJs, which may result in a crossover repair product (Barber *et al*., 2008).

Therefore, HR repairs DSBs through either a crossover or a non-crossover event (Bishop *et al*., 2004). Meiotic DSBs are not randomly distributed along chromosomes but tend to occur in specific regions (Handel & Schimenti, 2010). This ensures that each pair of homologs gets at least one obligate crossover, however how specific DSBs are selected to become crossover is unknown. A mechanism called crossover interference regulates the distribution of crossovers along the chromosome in such way that crossovers tend to occur further apart from each other than expected by chance. Furthermore, crossover homeostasis occurs when the number of meiotic DBSs is reduced, the number of crossover is maintained at the expense of non-crossover (Martini *et al*., 2006).

Barber *et al*. (2008) demonstrated that RTEL1 actively reverts HR at an early stage after strand invasion by D-loop dissociation activity in the presence of calcium, promoting a non-crossover repair product (Figure 5.2) and inhibiting the formation of D-loops *in vitro*. Biochemical studies done by the same group revealed that human RTEL1 disrupts D-loops in HR in both mitotic repair and regulating meiotic recombination. Subsequent studies demonstrated that RTEL1 controls excessive crossover products in meiosis and thus impacts on the outcome of the HR reaction (Sung & Klein, 2006; Youds *et al*., 2010).

Inappropriate HR can give rise to genome instability and cancer as a result of erroneous chromosomal rearrangements and the persistence of intermediate recombination structures such as D-loops that cannot be resolved. Hence, HR must be tightly regulated and temporally coordinated with cell-cycle progression and replication (Barber *et al*., 2008). HR is also important for the formation of T-loop structure at telomeres, which protects the chromosome end from degradation and inappropriate repair (Vannier *et al*., 2014).

Although, Barber *et al*. (2008) were the first to demonstrate that RTEL1 is a conserved anti-recombinase protein, RTEL1 was originally identified through genomic mapping of loci that control telomere length in mice (Zhu, *et al*., 1998). Then the role of RTEL1 in telomere homeostasis was proposed by Ding *et al*. (2004) when *Rtel1* knockout in mice was found to be embryonic lethal and inefficient removal of DNA secondary structures at telomeres could be the reason. This was based on the fact that RTEL1 is most related to human FANCJ and *C. elegans* DOG-1, which unwind DNA secondary structures.

5.1.3 RTEL1 in telomere maintenance

Telomeres are DNA-nucleoprotein complexes that maintain genomic stability by protecting the ends of eukaryotic chromosomes (Uringa *et al*., 2011). In most eukaryotic species, telomeric DNA consists of short G-rich repeat sequences synthesized by telomerase (Greider & Blackburn, 1985; Singer & Gottschiling, 1994). Telomeric DNA in vertebrates consists of a double-strand region composed of TTAGGG repeats associated to proteins, forming the sheltering complex (de Lange, 2005) (Figure 5.3).

An important function of telomeres is to distinguish normal chromosome ends from DSBs, which avoid chromosome end-to-end fusions and inappropriate recombination events (de Lange, 2009). Most proteins that are specific for telomere maintenance (such as telomerase and RTEL1) are recruited to telomeres via the shelterin complex that bind directly or indirectly to telomere repeats (Blackburn, 2001; Vega *et al*., 2003).

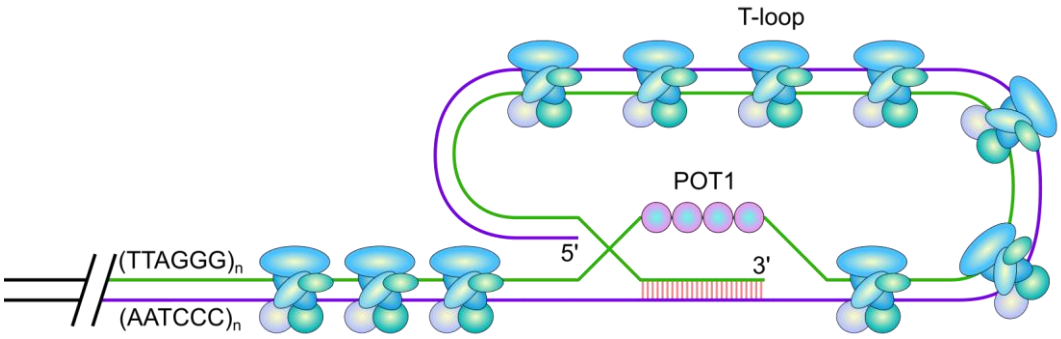


Figure 5.3. Schematic of the human telomere structure and the shelterin complex. Adapted from Titia de Lange (2004).

The shelterin complex includes telomere-specific binding proteins such as telomere repeat binding factor 1 (TRF1), telomere repeat binding factor 2 (TRF2), repressor activator protein 1 (Rap1), TINF2-interacting protein 2 (TTP1), protection of telomeres 1 (POT1) and ACD, shelterin complex subunit and telomerase recruitment factor (ACD) that function to regulate telomerase and protect the telomeres (de Lange, 2005; Bauman & Cech, 2001). In the absence of such proteins, DNA repair-specific proteins are recruited to chromosome ends to avoid chromosome fusions (Mieczkowski *et al*., 2003), chromosome end reallocation (Hackett & Greider, 2003) or other forms of potentially lethal DNA damage (Takai *et al*., 2003; Zhu *et al*., 2003).

Telomeric DNA shortens in each round of DNA replication because of incomplete replication of lagging strand DNA (Lansdorp, 2005) and it could also be lost by C strand degradation (Makarov *et al*., 1997), oxidative stress (von Zglinicki, 2002), and possibly other mechanisms. Telomere length is maintained by telomerase, a reverse transcriptase that utilizes an associated RNA component (TERC) as a template to add *de novo* telomeric sequences to the 3' end of the G-rich strand of the telomere (Greider & Blackburn, 1985; Shippen-Lentz & Blackburn, 1990), or by an alternative

telomere length (ALT) maintenance mechanism that utilises HR in some cancer cells. Cells with short telomeres that lack telomerase activity typically lose the ability to proliferate after a variable number of cell divisions (Blackburn, 2001). It is known that a minimum number of telomere repeats is required at each chromosome end for proper telomere function and to distinguish telomeres from DSBs. Excessive telomere shortening accelerates aging, but telomere elongation may facilitate cancer (Deng *et al*., 2013).

The telomere consists of a 3' single-strand G-rich overhang that invades into duplex telomeric TTAGGG repeats to form a T-loop structure (Griffith *et al*., 1999; Doksani *et al*., 2013). The T-loop structure requires both HR and shelterin complex for its assembly as they catalyse the invasion of the single-stranded telomere end into the subtelomeric region, displacing the identical sequence strand of the duplex telomeric DNA generating a D-loop at the base of the T-loop (Figure 5.4) (de Lange, 2004; Verdun & Karlseder, 2006 and Amiard *et al*., 2007). The D-loop is also an intermediate in the DNA repair pathway via homologous recombination as described previously.

HR has been shown to cause deletion of the protective T-loop to permit telomere replication (Wang *et al*., 2004). RTEL1 unwinds the T-loop structure by disrupting D-loop to allow telomerase access to complete chromosome end replication during cell cycle. Failing to open the T-loop for replication and/or transcription may lead to large telomeric deletions (Barber *et al*., 2008; Vannier *et al*., 2012).

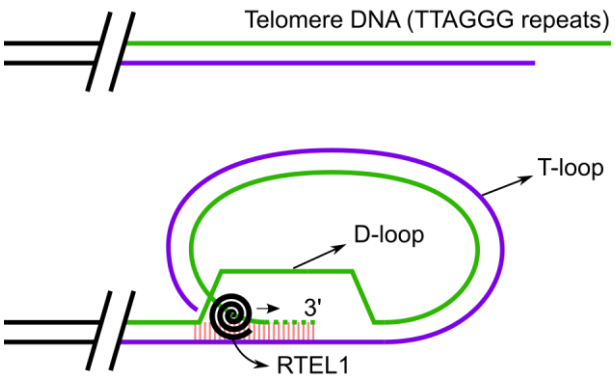


Figure 5.4. T-loop structure. T-loop protects the telomere from being identified as a DSB and it is formed by the telomere end that folds back onto itself. D-loop is created when the 3' overhang invades the double-stranded telomere. Black spiral indicates RTEL1. Adapted from Uringa *et al*. (2010).

In the absence of RTEL1, the T-loops are inappropriately resolved by SLX4 nuclease complex, resulting in loss of the telomere as a circle (T-circle) (Figure 5.5). SLX4 nuclease complex consists of SLX4 structure-specific endonuclease subunit (SLX4) and its associated nucleases MUS81 structure-specific endonuclease subunit (MUS81), ERCC excision repair 1, endonuclease non-catalytic subunit (ERCC1), and SLX1 structure-specific endonuclease subunit (SLX1). Studies have shown that the SLX4 nuclease complex is a Holliday junction-resolving enzyme that colocalises and interacts with the shelterin components TRF2 and Rap1 at telomeres (Muñoz *et al*., 2009; Svendsen *et al*., 2009; Svendsen & Harper, 2010; Wilson *et al*., 2013). It is possible that SLX4 nuclease complex is located at the telomeres to resolve persistent HR intermediates that may arise infrequently in normal cells. In contrast, SLX4 activity to resolve T-loops in RTEL1 deficient cells can also lead to telomere damages (Vannier *et al*., 2012).

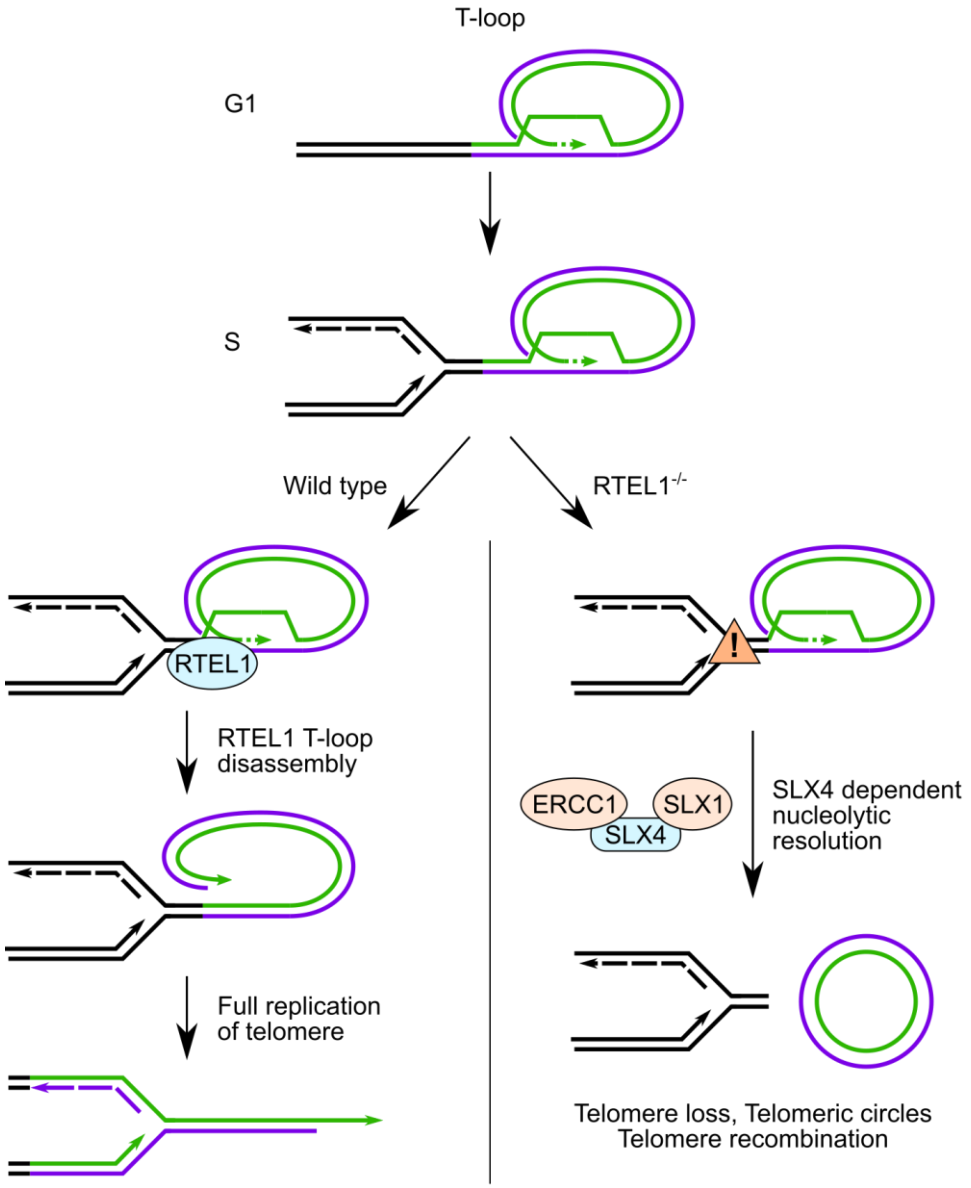


Figure 5.5. Schematic model of the role of RTEL1 in T-circle formation. Adapted from Vannier *et al*. (2012).

Vannier *et al*. (2012) observed that inactivation of RTEL1 in vertebrate cells resulted in a rapid accumulation of T-circles and changes in telomere length and telomere loss. Furthermore, RTEL1 is required to prevent 3' single strand telomere end from invading the telomeres of other chromosomes. This might prevent telomere

recombination events, chromosome entanglements and subsequent breakage when attached chromosomes attempt to segregate during mitosis (Uringa *et al*., 2011).

Mammalian telomeres are difficult regions to replicate and resemble fragile sites, which are hot spots for deletions and other chromosome rearrangements and are associated with an increased frequency of HR (Miller *et al*., 2006). Sfeir *et al*. (2009) showed that TRF1 recruits RTEL1 and Bloom syndrome RecQ like helicase (BLM) to prevent telomere fragility by resolving telomeric G-quadruplex structures (Figure 5.6). Although, Vannier *et al*. (2012) established that RTEL1 and BLM act in different pathways to facilitate telomeric DNA replication.

In addition to the T-loop configuration, the guanine (G)-rich nature of the telomere poses a challenge for telomere maintenance by causing telomere fragility. *In vitro*, single-stranded G-rich telomeric sequences are capable of forming stable secondary structures called G-quadruplex (G4-DNA) DNA that prevent DNA replication (Vannier *et al*., 2012). *In vivo*, G4-DNA might form at telomeres during lagging strand replication, repair and transcription of G-rich telomeric DNA (Ding *et al*., 2004). Vannier *et al*. (2013) found that RTEL1 is able to unwind telomeric G4-DNA structures *in vitro*, suggesting that it also plays a role in suppressing telomere fragility. Thus, RTEL1 performs two functions essential for telomere integrity: it facilitates T-loop disassembly and telomeric G4-DNA unwinding.

Additionally, RTEL1 is essential to facilitate DNA replication as described previously. The presence of RTEL1 within the replication foci is dependent on a PIP box interaction with PCNA, which is a processitivity factor for DNA polymerase and an integral component of the replisome during S-phase. Telomeric G4-DNA unwinding activity of RTEL1 is dependent on the replisome association between RTEL1 and

PCNA. Moreover, the RTEL1-PCNA interaction is also necessary to prevent replication fork stalling, which affects genome-wide replication (Vannier *et al*., 2013).

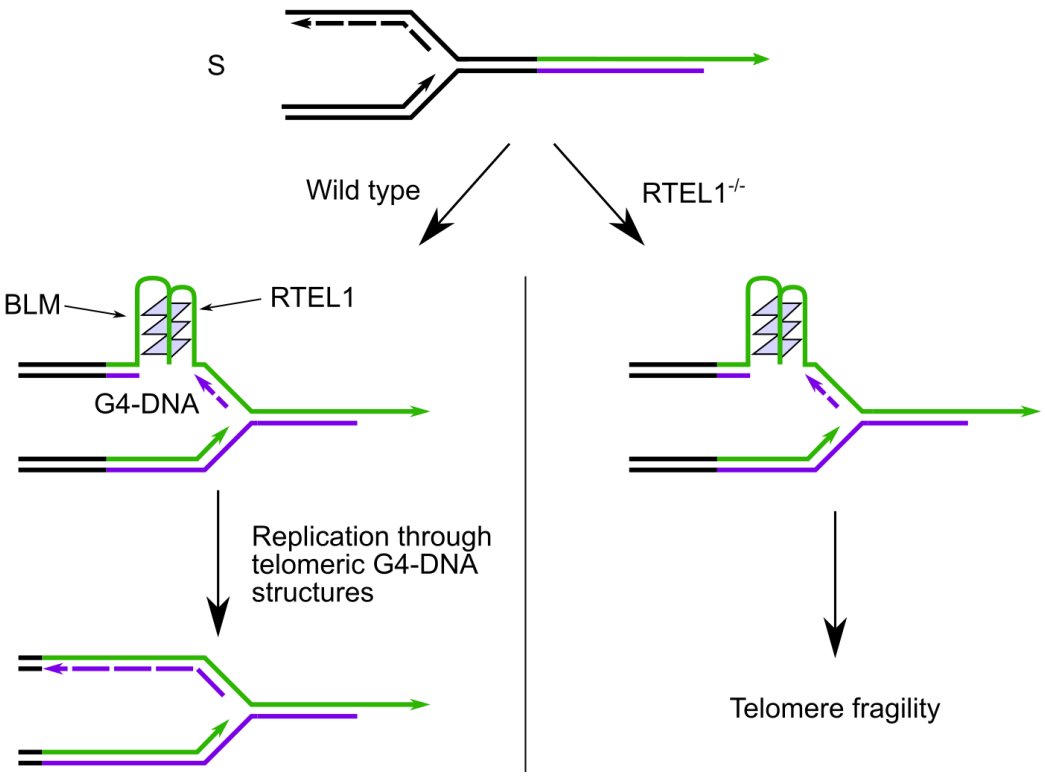


Figure 5.6. Schematic model of the role of RTEL1 in suppressing G4-DNA structure. Adapted from Vannier *et al*. (2012).

5.1.4 RTEL1 in human diseases

5.1.4.1 *RTEL1* germline biallelic variants leading to dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome

Germline biallelic variants in *RTEL1* are associated with dyskeratosis congenita (DC) and its severe clinical variant Hoyeraal Hreidarsson syndrome (HH), which are caused by telomere dysfunction (Walne *et al*., 2013a; Deng *et al*., 2013; Jullien *et al*., 2016).

DC is also known as Zinsser-Engman-Cole syndrome and it is a rare inherited multi-system bone marrow failure syndrome which is characterised by the presentation of abnormal skin pigmentation, nail dystrophy, oral leukoplakia and a variety of other abnormalities including bone marrow failure, pulmonary, gastrointestinal, endocrine, skeletal, urological, immunological and neurological abnormalities (Walne *et al*., 2013a; Walne *et al*., 2016). Its clinical features can manifest at variable ages, even within the same family (Alter *et al*., 2012).

Usually ectodermal dystrophy is the first abnormality to appear in DC patients followed by over 50% of patients developing haematopoietic disorders (Walne & Dokal, 2008). A number of DC cases have been described where the patient has presented with anaemia or other related bone marrow failure disease, pulmonary fibrosis and/or liver disease with high risk of cancer development (Alter *et al*., 2009). Some studies have shown that pancytopenia associated with DC develops more as a result of an intrinsic defect in the haematopoietic stem cells rather than as a result of a defective haematopoietic microenvironment (Marsh *et al*., 1992).

Furthermore, Fanconi anaemia (FA) and DC are related disorders and clinical features overlap between them such as pancytopenia and bone marrow hypoplasia as well as a range of minor abnormalities in addition to mental retardation and generalised growth impairment (Dokal & Vulliamy, 2008). However, there are also clear differences between these two diseases, the main distinguishing feature is the difference in chromosomal breakage in culture when peripheral blood lymphocytes are stressed with clastogenic agents such as diepoxybutane, with breaks commonly appearing in FA. Whereas, in the majority of cases, chromosome breakage analysis of DC is reported as normal (Dokal, 2006). The chromosomal instability seen in DC patient cells usually tends to be chromosomal rearrangements such as end-to-end fusions instead of chromosomal breakage seen in FA patients.

HH is a rare and severe multi-system disorder which is characterised by early onset of bone marrow failure, intrauterine growth retardation, developmental delay, microcephaly, cerebellar hypoplasia and immunodeficiency (Vulliamy *et al*., 2006). There is a high frequency of mortality from cancer and pulmonary fibrosis in patients with DC and HH (Tummala *et al*., 2015), however bone marrow failure is the major cause of death in this group of patients.

DC and HH are clinically and genetically heterogeneous diseases with defective telomere maintenance being their principal pathology and they are associated with short telomeres (Dokal, 2011). Approximately two thirds of patients have germline variants in genes encoding components of telomerase (*TERT*, *TERC*, *DKC1*, *NOP10*, *NHP2)* (Jullien *et al*., 2016; Deng *et al*., 2013; Ballew *et al*., 2013; Tummala *et al*., 2015; Walne *et al*., 2013b), or other factors related to telomere maintenance (*TINF2*, *WRAP53, CTC1*, *RTEL1*, *PARN*, and *ACD*) (Walne *et al*., 2013a; Tummala *et al*., 2015 and Cogan *et al*., 2015). Rare damaging variants in these genes segregate with

disease in autosomal dominant (AD), autosomal recessive (AR) or X-linked (XL) recessive patterns of inheritance. The molecular basis of the various DC and HH pathologies is therefore thought to be an accelerated telomere shortening and a consequent impairment of cell proliferation (Pereboeva *et al*., 2013). Vulliamy *et al*. (2004) observed disease anticipation associated with progressive telomere shortening in the development of AD form of DC in families harbouring variants in *TERC*, which implies that the disease become more severe in successive generations and the telomere lengths were significantly shorter in the affected children compared to their affected parents. Genetic anticipation has been reported also in *TERT* and *TINF2* pedigrees by Armanios *et al*. (2005) and Savage & Bertuch (2010). In addition, haploinsufficiency is the mechanism that leads to the AD-DC phenotype in most cases (Armanios *et al*., 2005).

Walne *et al*. (2013a) and Deng *et al*. (2013), demonstrated that defective human *RTEL1* has a damaging effect on telomere maintenance, suggesting that incorrect resolution of T-loops is a mechanism for telomere shortening in humans.

Overall, about 30% of DC and HH cases remain genetically uncharacterised highlighting the need to search for new disease causing genes (Tummala *et al*., 2015).

5.1.4.2 *RTEL1* germline heterozygous variants leading to pulmonary fibrosis

Up to 20% of cases of idiopathic interstitial pneumonia (IIP) occur in two or more members of the same family, comprising the syndrome of familial interstitial pneumonia (FIP). The most common and severe form of interstitial pneumonia is

pulmonary fibrosis; which is a progressive scarring of the alveolar interstitium, often leading to hypoxemic respiratory insufficiency. It is characterised by an accumulation of extracellular matrix and fibroblasts in the distal lung (Kannengiesser *et al*., 2015). Approximately one in five cases of pulmonary fibrosis run in families (King *et al*., 2011).

Familial and sporadic pulmonary fibrosis diseases are clinically and histologically indistinguishable, but familial pulmonary fibrosis tends to present at an earlier age and might differ slightly in radiological pattern (Spagnolo *et al*., 2014). The age of disease onset is between 50 to 70 years and it can be associated with environmental exposures, such inhalation of fibrogenic dusts or aerosolised organic antigens; drug toxicity; systemic diseases such as connective tissue diseases; by genetic variation; or occur isolated, sporadic disease without extra-pulmonary involvement (as seen in IIP) (Steele, *et al*., 2005).

Rare heterozygous variants causing FIP were identified in genes involved in pulmonary surfactant metabolism such as surfactant protein A2 (*SFTPA2*), surfactant protein C (*SFTPC*) and ATP-binding cassette subfamily A, member 3 (*ABCA3*) (Wang *et al*., 2009; Thomas *et al*., 2002; Nathan *et al*., 2012), and in genes related to telomere biology such as *TERT*, *TERC*, dyskerin (*DKC1*), *TINF2*, PARN poly(A)-specific ribonuclease (*PARN*), *RTEL1* and nuclear assembly factor 1 ribonucleoprotein (*NAF1*) (Armanios *et al*., 2007; Tsakiri, *et al*., 2007; Kropski *et al*., 2014; Fukuhara *et al*., 2013; Stuart *et al*., 2015; Cogan *et al*., 2015; Stanley *et al*., 2016). These variants are reported to segregate with an autosomal dominant pattern of inheritance and are associated with short telomeres and variable penetrance. Short telomeres are commonly identified in blood cells in patients with sporadic and FIP also in the absence of known variants in telomere-related genes.

Interestingly, variants in *TERT*, *TERC*, *DKC1*, *TINF2*, *PARN*, and *RTEL1* were primarily described being responsible for DC and HH and pulmonary disease is found in 20% of DC patients with variants in these genes. However, the mechanism by which telomerase dysfunction and short telomeres lead to lung fibrosis is unknown. It is possible that telomerase pathway variants lead to premature senescence of progenitor cells of the distal lung, resulting in a proliferative defect and failure of repair mechanisms following injury to the alveolar epithelium. Further studies are necessary to clarify the mechanistic role of *RTEL1* and the other genes involved in telomere biology in the pathogenesis of lung fibrosis (Cogan *et al*., 2015; Kannengiesser *et al*., 2015).

5.2 Results

By undertaking whole exome and targeted sequencing in 429 patients with diverse bone marrow failure phenotypes (DC, HH, AA and MDS/AML) we have identified 35 patients with 27 *RTEL1* variants (canonical transcript, NM_032957.4, NP_116575.3, 1,243 amino acids)**.** These are detailed in Table 5.1. Based on the minor allele frequency (MAF) in the population reported on ExAC database, the type of variant (missense, nonsense and indels), the telomere length, the Combined Annotation Depletion (CADD) score (Kircher *et al*., 2014), and their segregation as well as information found in the literature, we classified these variants into four different groups: (1) biallelic variants, (2) heterozygous loss of function (LoF) variants, (3) heterozygous missense variants of unknown significance and (4) heterozygous missense variants likely to be benign.

Table 5.1. *RTEL1* variants identified in 35 index cases

		Index	Diagnosis	Age (years)	Sex	Telomere length T/S ratio (percentile)	DNA change	Protein change	ExAC frequency[#]	CADD PHRED
Biallelic	Likely pathogenic	1	AA	36	F	0.46 (1st)	c.2942G>A (homozygous)	p.R981Q	1 in 119,934	28.4
		2	AA	12	F	0.49 (<10th)	c.3286G>T (homozygous)	p.G1096W	1 in 118,164	26.2
		3	DC	14	M	0.4 (<1st)	c.2300G>A (homozygous)	p.R767Q	NR	27.7
		4	DC	15	F	0.64 (<10th)	c.2785_2787delCAG (heterozygous)	p.Q929del	NR	12.9
							c.2992C>T (heterozygous)	p.R998*	2 in 119,914 (1 in 59,957)	37
	Likely benign	5	HH	2	F	0.45 (<1st)	c.1716C>G (homozygous)	p.I572M	NR	24.9
Heterozygous	LoF Likely pathogenic	6	DC	77	M	0.5 (<10th)	c.3028C>T	p.R1010*	10 in 119,716 (1 in 11,972)	34
		7	MDS	45	F	0.47 (<10th)	c.3028C>T	p.R1010*	10 in 119,716 (1 in 11,972)	34
		8	MDS	55	M	0.73 (<50th)	c.2992C>T	p.R998*	2 in 119,914 (1 in 59,957)	37
		9	MDS	54	M	0.48 (<10th)	c.3012_3028del	p.Q1005Kfs*80	NR	34
	Unknown significance	10	AML	23	F	1.14 (>50th)	c.3464C>T	p.T1155M	5 in 117,408 (1 in 23,482)	12.94
		11	MDS	20	F	0.95 (<50th)	c.2965G>C	p.E989Q	3 in 119,938 (1 in 39,980)	22
		12	DC	10	M	0.39 (<1st)	c.2723C>G	p.P908R	1 in 119,146	0.001
		13	DC	24	M	0.99 (>50th)	c.208C>T	p.R70C	10 in 120,456 (1 in 12,046)	25.3
		14	AA	24	F	0.65 (<10th)	c.2941C>T	p.R981W	6 in 119,930 (1 in 19,988)	33
		15	AA	6	F	1.04 (>50th)	c.2351C>T	p.A784V	6 in 118,274 (1 in 19,712)	2.88
		16	AA	28	F	1.56 (>90th)	c.3595G>A	p.G1199R	4 in 107,372 (1 in 26,843)	5.246
		17	DC	8	F	0.98 (>90th)	c.1603A>G	p.I535V	NR	9.212
		18	DC	18	M	0.95 (<50th)	c.3430G>A	p.V1144M	NR	23.6
		19	AA	28	M	0.56 (<10th)	c.4129A>G†	p.T1377A	NR	1.406
		20	AA	10	F	1.01 (>50th)	c.3608G>A	p.S1203N	NR	23.5
		21	DC	16	F	0.47 (<10th)	c.1991G>T	p.G664V	NR	26.1

Table 5.1. Continued

Heterozygous	**Likely benign**	22	DC	18	M	1.21 (>50th)	c.2618G>A	p.G873D	249 in 19,124 (1 in 77)	10.29
		23	DC	37	M	0.64 (<10th)	c.2516G>T	p.S839I	126 in 71,024 (1 in 564)	17.05
		24	AA	4	F	1 (>50th)	c.3047C>T	p.P1016L	184 in 119,184 (1 in 648)	10.85
		25	DC	50	F	0.94 (<50th)	c.3047C>T	p.P1016L	184 in 119,184 (1 in 648)	10.85
		26	DC	NA	M	NA	c.3047C>T	p.P1016L	184 in 119,184 (1 in 648)	10.85
		27	MDS/AML	61	F	0.63 (<10th)	c.3128A>G	p.Q1043R	151 in 118,626 (1 in 786)	0.276
		28	DC	4	M	0.85 (<50th)	c.3992G>A†	p.R1331Q	120 in 101,400 (1 in 845)	12.7
		29	DC	3	M	1.51 (>90th)	c.3992G>A†	p.R1331Q	120 in 101,400 (1 in 845)	12.7
		30	HH	0	M	0.69 (<50th)	c.2734G>C	p.V912L	85 in 117,986 (1 in 1,388)	6.325
		31	DC	54	M	0.47 (<10th)	c.4159C>T†	p.P1387S	71 in 110,950 (1 in 1,563)	24.7
		32	AA	7	M	1.34 (>50th)	c.4159C>T†	p.P1387S	71 in 110,950 (1 in 1,563)	24.7
		33	DC	31	M	0.64 (<10th)	c.1261C>G	p.Q421E	71 in 120,318 (1 in 1,695)	24.2
		34	AA	34	M	0.6 (<10th)	c.1261C>G	p.Q421E	71 in 120,318 (1 in 1,695)	24.2
		35	DC	3	M	1.44 (>50th)	c.3121G>A	p.D1041N	43 in 118,650 (1 in 2,759)	14.34

NR: not reported; CADD PHRED: combined annotation dependent depletion score; AA: aplastic anemia; AML: acute myeloid leukemia; DC: dyskeratosis congenita; HH: Hoyeraal Hreidarsson syndrome; MDS: myelodysplasia. †variant is not in the canonical transcript ENST00000508582 seen in ExAC, but is found in ENST00000482936. Centiles for T/S ratios, established from a healthy control population (n = 202) are as follows: 99th centile = 1.99, 90th centile = 1.47, 50th centile = 0.96, 10th centile = 0.68, 1st centile = 0.46. Telomeres are considered short if they are at or below the 10th centile, and very short if they are at or below the 1st centile.

$^{\#}$: For each of the rare variants reported (less than 10 heterozygotes), the ethnicity of our patient matched at least one reported on the ExAC.

NB: Six index cases harbour variants in other known disease genes: index cases 11, 15, 28, 31, 33 and 35 harbour variants in *TERT* (heterozygous c.3197C>T; p.P1066L and c.322C>T; p.R108C), *DNAJC21* (homozygous c.793G>T; p.Q265*), *TERT* (heterozygous c.1336_1337insC; p.R446Pfs93* and c.329G>C; p.G110A), *TERT* (homozygous c.3150G>C; p.K1050N) and *TERC* (heterozygous c.205C>T), *TINF2* (heterozygous c.838A>G; p.K280Q), and *DKC1* (hemizygous c.941A>C; p.K314T), respectively.

5.2.1 *RTEL1* germline variants identified in our cohort

5.2.1.1 Patients with biallelic *RTEL1* variants

Four homozygous and one compound heterozygous (Table 5.1 and 5.2) were identified in five patients from unrelated families (Families 1 to 5), two of whom presented with AA, two with DC and one with HH. The six variants identified fall in the C-terminal half of the protein (Figure 5.7).

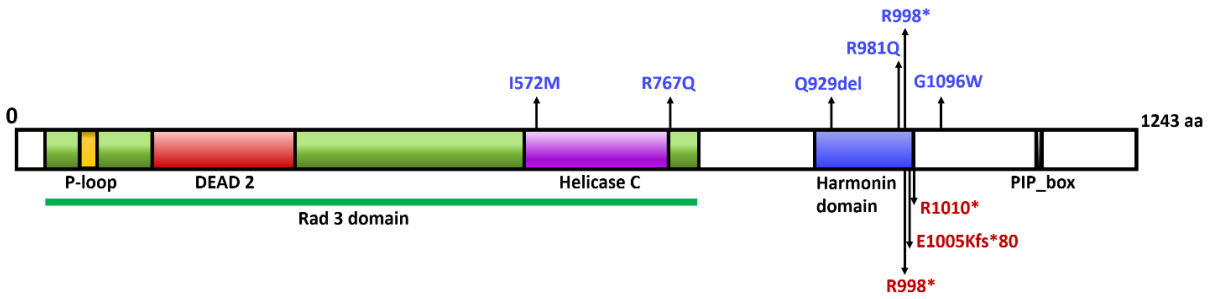


Figure 5.7. RTEL1 variants. Biallelic variants are shown in blue, and heterozygous LoF variants are shown in red. Conserved protein domains (NP_116575.3) include the P-loop NTPase (yellow); the Rad3 domain (green) that includes the DEAD2 domain (red) and the Helicase C-terminal domain (purple); Harmonin N-like domain (blue); PIP-box – the proliferating cell nuclear antigen interacting protein domain (black).

Table 5.2. Characteristics of families with biallelic *RTEL1* variants

Family	Individuals	Age at study (years)	Gender	Clinical status	Nucleotide	Amino acid	Variant status	Clinical features/diagnosis
1	I-1	NA	F	asymptomatic	NA	NA	NA	none
	I-2	NA	M	asymptomatic	NA	NA	NA	none
	II-1	NA	M	asymptomatic	NA	NA	NA	none
	II-2	36	F	affected	c.2942G>A	p.R981Q	homozygous	AA; short stature
2	I-1	70	F	affected	NA	NA	NA	pulmonary fibrosis
	I-2	NA	M	NA	NA	NA	NA	NA
	II-1	40	F	asymptomatic	c.3286G>T	p.G1096W	heterozygous	none
	II-2	43	M	asymptomatic	c.3286G>T	p.G1096W	heterozygous	none
	III-1	16	F	asymptomatic	NA	NA	NA	none
	III-2	15	F	asymptomatic	NA	NA	NA	none
	III-3	12	F	affected	c.3286G>T	p.G1096W	homozygous	AA, pulmonary fibrosis, corrugated tongue
	III-4	3	M	affected	c.3286G>T	p.G1096W	homozygous	AA
3	I-1	40	F	asymptomatic	c.2300G>A	p.R767Q	heterozygous	none
	I-2	37	M	asymptomatic	c.2300G>A	p.R767Q	heterozygous	none
	II-1	15	M	asymptomatic	NA	NA	NA	none
	II-2	14	M	affected	c.2300G>A	p.R767Q	homozygous	DC, blepharitis, conjunctivitis, pancytopenia, atrial septal defect, low birth weight, growth restriction
	II-3	13	M	asymptomatic	NA	NA	NA	none
	II-4	12	F	asymptomatic	NA	NA	NA	none
	II-5	10	M	asymptomatic	NA	NA	NA	none
	II-6	8	M	asymptomatic	c.2300G>A	p.R767Q	heterozygous	none
	II-7	5	M	asymptomatic	NA	NA	NA	none
4	I-1	46	F	asymptomatic	c.2785_2787delCAG	p.Q929del	heterozygous	none
	I-2	50	M	asymptomatic	c.2992C>T	p.R998*	heterozygous	none
	II-1	24	F	asymptomatic	c.2992C>T	p.R998*	heterozygous	none
	II-2	15	F	affected	c.2785_2787delCAG c.2992C>T	p.Q929del p.R998*	compound heterozygous	DC
5	I-1	25	F	asymptomatic	c.1716C>T	p.I572M	homozygous	none
	I-2	31	M	asymptomatic	c.1716C>T	p.I572M	heterozygous	none
	II-1	still birth	M	NA	NA	NA	NA	NA
	II-2	still birth	M	NA	NA	NA	NA	NA
	II-3	6 weeks	NA	affected	NA	NA	NA	Turner syndrome
	II-4	2	F	affected	c.1716C>T	p.I572M	homozygous	HH
	II-5	NA	F	NA	NA	NA	NA	NA

AA: aplastic anemia; DC: dyskeratosis congenita; HH: Hoyeraal Hreidarsson syndrome; NA: not available; F: female; M: male.

The proband of Family 1 (II:2, Figure 5.8) was a 36 year old female presenting with AA and short stature who had the homozygous missense variant, c.2942 G>A, p.R981Q. This *RTEL1* variant is seen in the ExAC database in a heterozygous state with a frequency of 1 in 119,934. Hence, the homozygous occurrence of c.2942 G>A, p.R981Q is predicted to be extremely rare and in the context of very short telomeres and a high CADD score (Table 5.1), we believe this variant is likely to be disease causing in this patient.

In Family 2, the female proband (III:3, Figure 5.8) had AA and short telomeres (Table 5.1 and 5.2). She had a bone marrow transplant at age 12 years from her HLA-matched brother, but developed significant complications following post-transplant. This included bone marrow failure and pulmonary fibrosis and she died aged 16 years. Her donor (brother) went on to develop AA which was responsive to therapy with danazol (III:4, Figure 5.8). The proband's maternal grandmother died of pulmonary fibrosis aged 70 years (I:1, Figure 5.8). Both siblings, offspring of a consanguineous marriage, were homozygous for the missense variant c.3286 G>T, p.G1096W, seen in ExAC in a heterozygous state with the frequency of 1 in 118,164. We believe this variant is pathogenic in this case. The proband's asymptomatic parents in their forties were heterozygous for this variant (II:1 and 2, Figure 5.8).

The proband of Family 3, was a 14-year-old male (II:2, Figure 5.8) from a consanguineous marriage with features of DC as well as very short telomeres (Table 5.1 and 5.2). He was homozygous for the novel missense *RTEL1* variant c.2300 G>A, p.R767Q, which we believe to be disease causing. His asymptomatic parents (I:1 and 2, Figure 5.8) and one of his asymptomatic younger brothers (II:6, Figure 5.7) were all heterozygous for this variant.

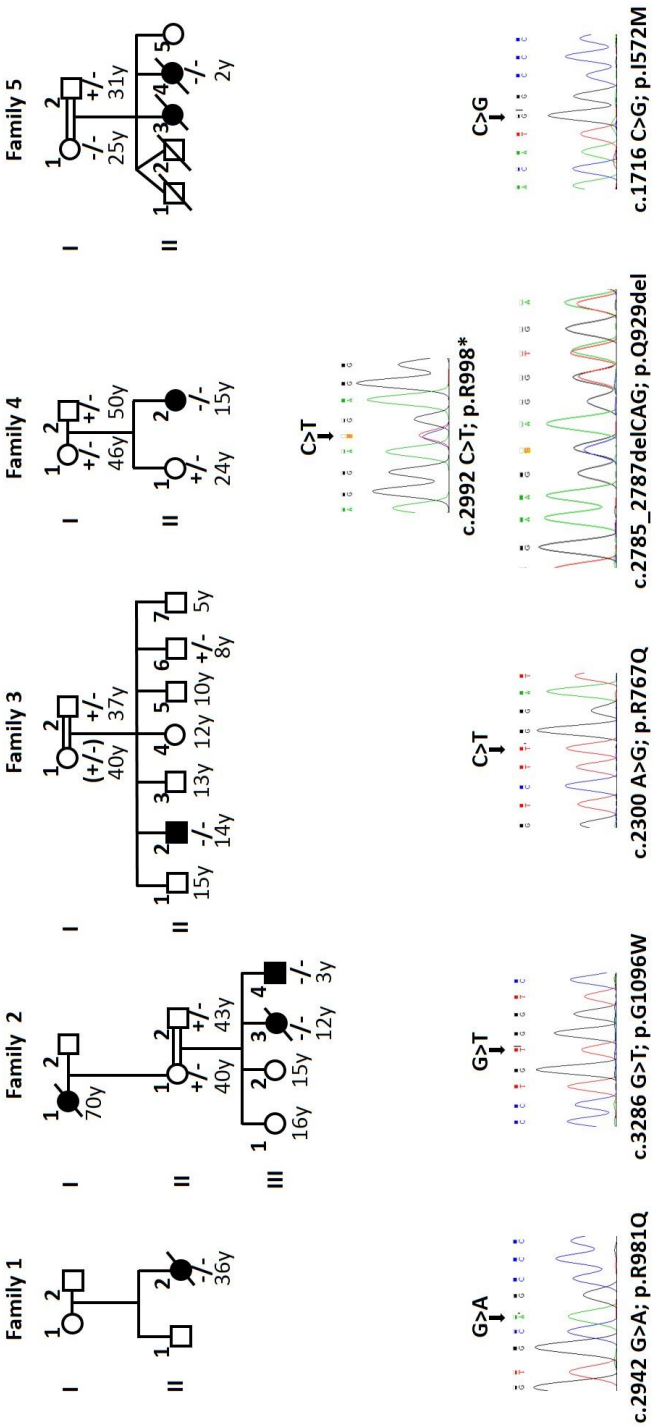


Figure 5.8. Families with biallelic *RTEL1* variants and sequencing traces of index cases. Families 1, 2, 3 and 5 are homozygous; Family 4 is compound heterozygous. The genotyping is denoted as follows: wild-type (+/+), heterozygous (+/-) or biallelic (-/-). The age at study is given in years. Affected individuals are coloured in black. NA: not available.

The proband of Family 4 (II:2, Figure 5.8) was a 15-year-old female with DC with short telomeres (Table 5.1 and 5.2). She was found to harbour a compound heterozygous variant, a novel in-frame deletion c. 2785_2787delCAG, p.Q929del and a recurrent nonsense variant c.2992 C>T, p.R998* (MAF: 2 in 119,914 in ExAC).

The proband (II:4, Figure 5.8) in Family 5 was diagnosed with HH when she was 2 years old. She came from a family with history of multiple consanguineous marriages through generations. She had intrauterine growth restriction, bone marrow failure, low B cell numbers, developmental delay, cerebellar hypoplasia. She also had ataxia and feeding and swallowing difficulties. She was found to be homozygous for a novel missense variant, c.1716 C>G, p.I572M. However, it is very unlikely this variant is causing the severe phenotype observed, as her asymptomatic mother (I:1, Figure 5.8) is also homozygous for this variant. We therefore, classified this variant as a likely benign.

5.2.1.2 Patients with heterozygous loss of function *RTEL1* variants

Three LoF *RTEL1* variants (Table 5.1 and 5.3) were found in patients from four unrelated families (Families 6-9, Figure 5.9), one of whom presented with DC and the others with MDS and/or liver disease. Liver disease and myelodysplasia/BMF were the common clinical feature in families 6, 8 and 9 (Figure 5.9). The clinical features of the four families with LoF variants in *RTEL1* are summarised in Table 5.3. The three LoF variants are seen to cluster at the C terminal end of the harmonin domain of RTEL1 (Figure 5.7).

Table 5.3. Characteristics of families with LoF *RTEL1* variants

Family	Individuals	Age at study (years)	Gender	Clinical status	Nucleotide	Amino acid	Variant status	Clinical features/diagnosis
6	I-1	NA	F	asymptomatic	NA	NA	NA	none
	I-2	NA	M	asymptomatic	NA	NA	NA	none
	II-1	NA	F	asymptomatic	NA	NA	NA	none
	II-2	77	M	affected	c.3028C>T	p.R1010*	heterozygous	DC, lacy skin pigmentation, pancytopenia, pulmonary fibrosis, cirrhosis
	II-3	NA	F	asymptomatic	NA	NA	NA	none
	III-1	NA	M	asymptomatic	NA	NA	NA	none
	III-2	54	F	asymptomatic	c.3028C>T	p.R1010*	heterozygous	none
	III-3	52	F	affected	c.3028C>T	p.R1010*	heterozygous	liver disease (non-specific hepatitic changes)
	III-4	49	F	asymptomatic	wild type	wild type	wild type	none
	III-5	47	M	asymptomatic	c.3028C>T	p.R1010*	heterozygous	none
	III-6	NA	F	asymptomatic	NA	NA	NA	none
	IV-1	20	F	asymptomatic	wild type	wild type	wild type	none
	IV-2	17	F	asymptomatic	wild type	wild type	wild type	none
	IV-3	16	F	asymptomatic	c.3028C>T	p.R1010*	heterozygous	none
	IV-4	23	F	asymptomatic	wild type	wild type	wild type	none
	IV-5	21	F	asymptomatic	wild type	wild type	wild type	none
	IV-6	7	F	asymptomatic	c.3028C>T	p.R1010*	heterozygous	none
	IV-7	4	M	asymptomatic	wild type	wild type	wild type	none
7	I-1	NA	F	NA	NA	NA	NA	NA
	I-2	NA	M	NA	NA	NA	NA	NA
	III-1	45	F	affected	c.3028C>T	p.R1010*	heterozygous	MDS, nail dystrophy
8	I-1	NA	F	asymptomatic	NA	NA	NA	none
	I-2	NA	M	asymptomatic	NA	NA	NA	none
	II-1	55	M	affected	c.2992C>T	p.R998*	heterozygous	MDS (low risk), cirrhosis
	II-2	47	M	affected	c.2992C>T	p.R998*	heterozygous	MDS (low risk), cirrhosis
	II-3	NA	F	asymptomatic	NA	NA	NA	none
9	I-1	46	F	affected	NA	NA	NA	liver and lung disease
	I-2	NA	M	asymptomatic	NA	NA	NA	none
	II-1	NA	M	asymptomatic	NA	NA	NA	none
	II-2	NA	F	asymptomatic	NA	NA	NA	none
	III-1	NA	F	asymptomatic	NA	NA	NA	none
	III-2	54	M	affected	c.3012_3028del	p.E1005Kfs*80	heterozygous	MDS (low risk), interstitial lung disease, cirrhosis, osteoporosis, baldness and psoriatiform skin
	III-3	49	F	asymptomatic	wild type	wild type	wild type	none
	IV-1	32	F	asymptomatic	c.3012_3028del	p.E1005Kfs*80	heterozygous	none
	IV-2	30	F	asymptomatic	wild type	wild type	wild type	none

DC: dyskeratosis congenita; MDS: myelodysplasia; NA: not available; F: female; M: male

The proband of Family 6 (II:2, Figure 5.9) was diagnosed with DC at age 77 years. He progressively developed pulmonary fibrosis, liver disease and bone marrow failure (pancytopenia) with some dysplasia. He died shortly after diagnosis with multi-system complications. He harboured the heterozygous nonsense *RTEL1* variant c.3028 C>T, p.R1010* (MAF: 10 in 119,716 in ExAC), reported previously in patients by Ballew *et al* (2013) and as a compound heterozygote by Moriya *et al* (2016). He had three daughters and one son (age 54, 52, 49 and 47 years, respectively), and all but one of his daughters (III:4, Figure 5.9) harboured this variant. Among his three children harbouring this nonsense variant, one of his daughters had features of liver disease (III:3, Figure 5.9), the other daughter and son were asymptomatic (III:2 and 5, Figure 5.9). The variant was also identified in two of his asymptomatic granddaughters (IV:3 and 6, Figure 5.9).

The proband in Family 7 (II:1, Figure 5.9) was a 45 years old female with MDS, nail dystrophy and short telomeres (Table 5.1 and 5.3). She carried the same heterozygous nonsense variant c.3028 C>T, p.R1010* as seen in Family 6.

In Family 8 (II:1, Figure 5.9), the proband was a 55-year-old male with MDS and liver disease, which progressed to cirrhosis and required liver transplantation. He was heterozygous for the nonsense variant c.2992C>T, p.R998* (MAF: 2 in 119,914 in ExAC) which has been reported as a compound heterozygote causing DC and HH (Walne *et al*., 2013a; Deng *et al*., 2013 and Ballew *et al*., 2013) and as a heterozygote causing familial interstitial pneumonia (Cogan *et al*., 2015). His brother (II:2, Figure 5.9) aged 47 years harboured the same nonsense variant and also had MDS and liver cirrhosis.

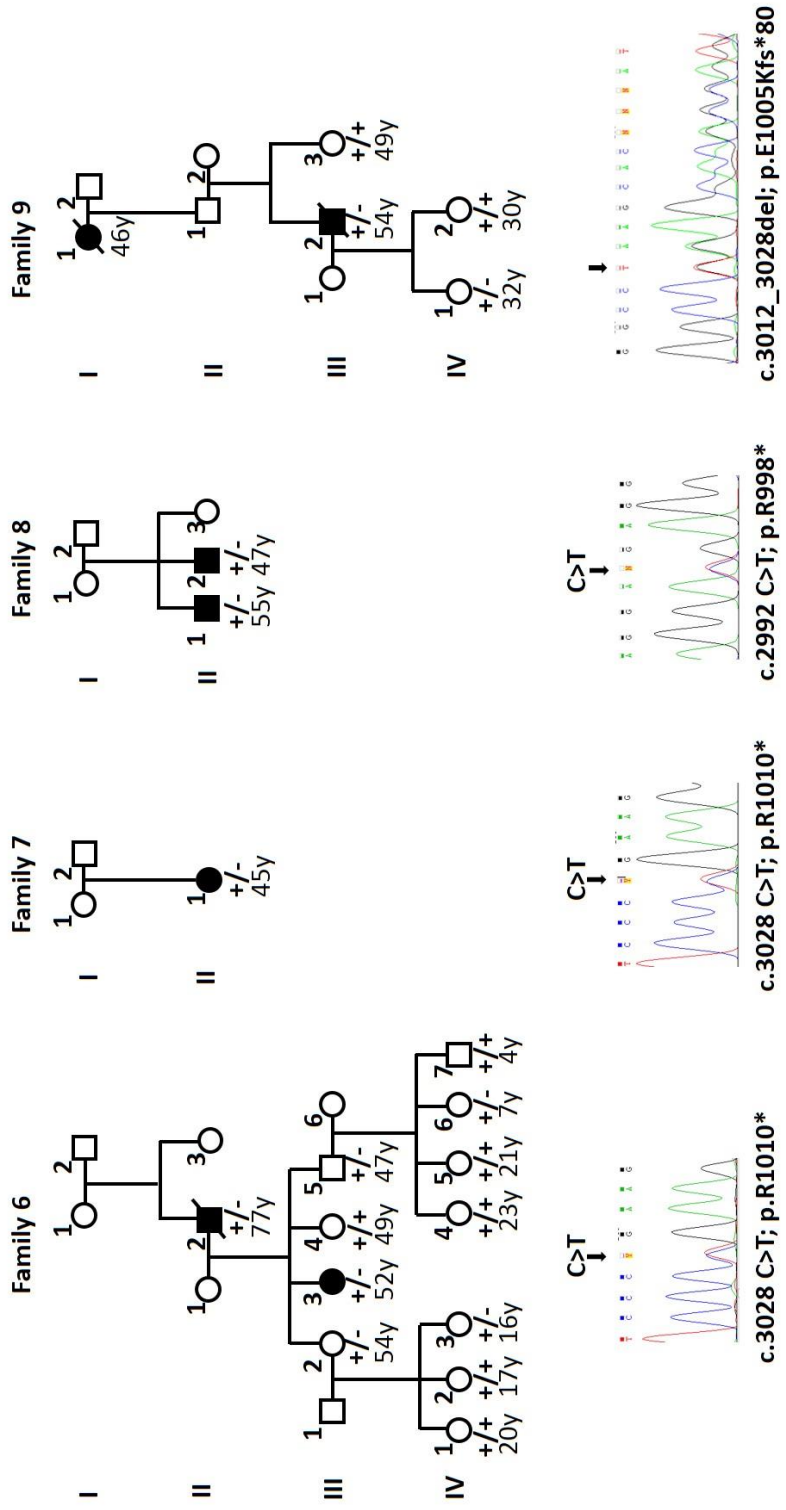


Figure 5.9. Families with heterozygous LoF *RTEL1* variants and sequencing traces of index cases. The genotyping is denoted as follows: wild-type (+/+) and heterozygous (+/-). The age at study is given in years. Affected individuals are coloured in black. NA: not available.

The proband of Family 9 (III:2, Figure 5.9) was diagnosed with MDS. He had liver disease, interstitial lung disease, abnormal skin pigmentation and severe osteoporosis. He died aged 54 years from multi-system complications. He harboured a novel 16bp deletion that caused a frameshift and a premature stop codon (c.3012_3028del, p.E1005Kfs*80). His parents were asymptomatic but his maternal grandmother (I:1, Figure 5.9) died aged 46 years with liver and lung disease. His 32-year-old daughter (IV:1, Figure 5.9) was an asymptomatic carrier of the same variant.

5.2.1.3 Families with heterozygous variants of unknown significance and heterozygous likely benign variants

A variant might be of unknown significance (VUS) when there is no validated association of this variant to a disease risk. In this study, we have identified 12 unrelated patients harbouring heterozygous missense variants of unknown significance (Table 5.1 and 5.4), as these variants are missense that are either not seen in the ExAC population or are present at an allele frequency of less than 1 in 10,000. This makes it difficult to assign a clear status to these and hence we have categorised them as VUS. We notice that there is lower average CADD score for these VUS (average 15.43, range 0.001 – 33), compared to those that we believe to be disease causing (average 30.13, range 12.9 – 37, Table 5.1).

Additionally, 14 unrelated patients were identified with nine heterozygous missense variants that we believe to be likely benign (Table 5.1 and 5.4) due to their occurrence at an allele frequency of less than 1 in 3,000 in the ExAC population.

Table 5.4. Characteristics of index cases with heterozygous VUS and likely benign *RTEL1* variants

	Index	Age at study (years)	Gender	Diagnosis	Additional relevant clinical features
Unknown significance	10	23	F	AML	Short stature
	11	20	F	MDS	Skin pigmentation abnormality and squamous cell carcinoma of oesophagus. This patient harbours variants in *TERT* (heterozygous c.3197C>T; p.P1066L and c.322C>T; p.R108C)
	12	10	M	DC	Developmental delay, short stature, dysmorphic facial features, microcephaly, BMF and pulmonary disease
	13	24	M	DC	Skin pigmentation abnormality, leukoplakia, thin hair and BMF
	14	24	F	AA	
	15	6	F	AA	Short stature and oral ulceration with dysphagia. This patient harbours variant in *DNAJC21* (homozygous c.793G>T; p.Q265*)
	16	28	F	AA	Nail dystrophy and leukoplakia
	17	8	F	DC	Skin pigmentation abnormality, thin hair, extensive dental caries and BMF
	18	18	M	DC	
	19	28	M	AA	
	20	10	F	AA	
	21	16	F	DC	Skin pigmentation abnormality, nail dystrophy, leukoplakia, small teeth, sparse scalp hair, epiphora, microcephaly and BMF
Likely benign	22	18	M	DC	Skin pigmentation abnormality, nail dystrophy, hair loss, extensive dental caries, developmental delay and short stature
	23	37	M	DC	Skin pigmentation abnormality, nail dystrophy, hair loss, frequent otitis, mild hearing loss and extensive caries/ dental loss
	24	4	F	AA	Nail dystrophy, cirrhosis, duodenal ulcers, deafness and developmental delay
	25	50	F	DC	Skin pigmentation abnormality, nail dystrophy, leucoplakia and leukemia
	26	NA	M	DC	
	27	61	F	MDS/AML	
	28	4	M	DC	Skin pigmentation abnormality, nail dystrophy, microcephaly, low birthweight, developmental delay and cerebellar atrophy. This patient harbours variants in *TERT* (heterozygous c.1336_1337insC; p.R446Pfs93* and c.329G>C; p.G110A)
	29	3	M	DC	Skin pigmentation abnormality, nail dystrophy, abnormal facies, microcephaly, ear abnormality and difficulty in swallowing
	30	0	M	HH	Congenital cytomegalovirus infection, microcephaly, generalized seizures, intracranial calcifications, growth restriction, low birth weight and BMF
	31	54	M	DC	Skin pigmentation abnormality, nail dystrophy, hair loss, tooth loss, renal failure and BMF. This patient harbours variants in *TERT* (homozygous c.3150G>C; p.K1050N) and *TERC* (heterozygous c.205C>T)
	32	7	M	AA	
	33	31	M	DC	Skin pigmentation abnormality, leukoplakia, epiphora, duodenal ulcers, cirrhosis, hepato-pulmonary syndrome and BMF. This patient harbours variant in *TINF2* (heterozygous c.838A>G; p.K280Q)
	34	34	M	AA	
	35	3	M	DC	Skin pigmentation abnormality, nail dystrophy, leucoplakia, hair loss, microcephaly, premature birth with intrauterine growth restriction, glaucoma, premature aging, malabsorption, developmental delay and BMF. This patient harbours variant in *DKC1* (hemizygous c.941A>C; p.K314T)

AML: acute myeloid leukemia; MDS: myelodysplasia; DC: dyskeratosis congenita; AA: aplastic anemia; HH: Hoyeraal Hreidarsson syndrome; BMF: bone marrow failure; NA: not available; F: female; M: male.

5.2.2 Short telomeres and T-circles in distinguishing the pathogenic status of *RTEL1* variants

We have measured telomere lengths by MMqPCR (Cawthon, 2009) in peripheral blood DNA acquired from all patients bar one, which had poor DNA quality (Table 5.1, Figure 5.10).

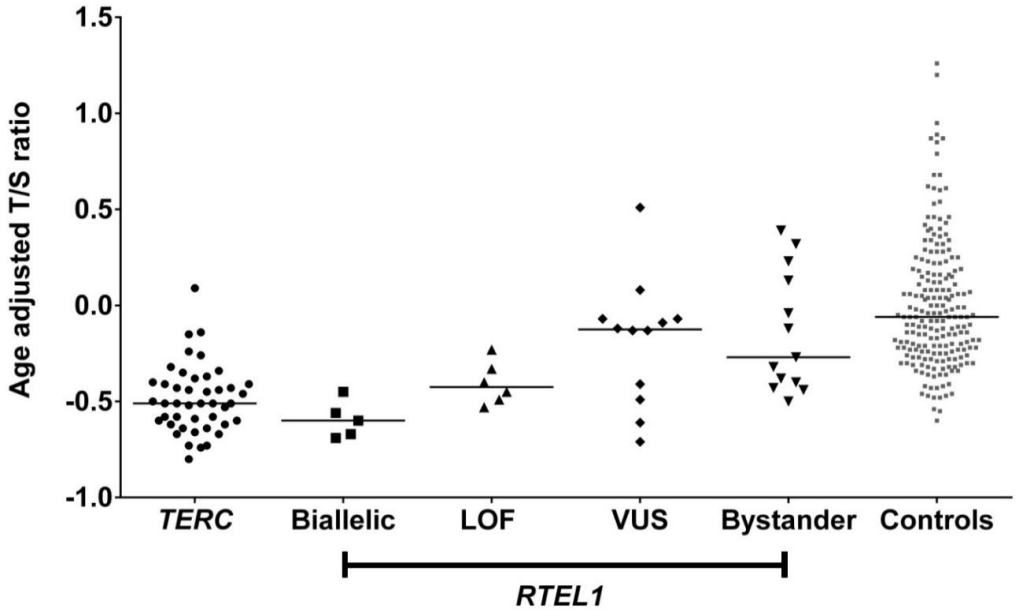


Figure 5.10. Age adjusted telomere length of *RTEL1* patients. Age adjusted telomere length values (delta-tel) were measured by subtracting the observed T/S ratio from the expected T/S ratio, using the equation derived from the line of best fit through the plot of T/S ratios from healthy control samples against age. Patients with TERC variants are included as a group with known short telomeres. Centiles were calculated from the control delta-tel values as follows: 99th centile = 0.95, 90th centile = 0.42, 50th centile = 0.06, 10th centile = -0.34, 1st centile = -0.54. The different genotypes are represented as follows, TERC: circles (n=44); *RTEL1* variants: biallelic: squares (n=5); loss of function (LoF): triangles (n=6); variants of unknown significance (VUS): diamonds (n=12); likely benign: inverted triangles (n=13); controls: grey squares (n=202).

In agreement with previous studies reporting the impact of *RTEL1* variants on telomere length, we observed that patients with biallelic variants and those with heterozygous loss of function variants had significantly shorter telomeres than controls as determined by the age-adjusted T/S ratio ($p = 0.0005$ and $p = 0.003$ respectively, 1 way ANOVA with Dunn's multiple comparison test). The median age adjusted T/S ratio for the biallelic group is below the 1st centile (-0.6 compared with -0.54) and for the LoF group is below the 10th centile (-0.43 compared with -0.34). It is interesting to note that in the VUS group there appears to be two subgroups. The lower four points correspond to the variants p.G664V, p.P908R, p.R981W and p.T1377A. Three of these variants affect key domains within the protein and may impact on the function of RTEL1. These are the helicase C domain (G644V) and the harmonin domain (P908R and R981W).

The T-circle amplification assay (Zellinger *et al*., 2007) was undertaken in patients where good quality DNA was available and revealed a significant increase in intensity of T-circle formation in a patient harbouring RTEL1 LoF variant p.R998* (Family 5.8, Figures 5.9 and 5.11) in comparison to patient with RTEL1 p.R70C, which is a variant of unknown significance (patient 13 in Table 5.1). These studies suggest that LoF status in heterozygous *RTEL1* variants could be established based on the T-circle formation assay and most importantly these LoF variants impact RTEL1 function in telomere maintenance.

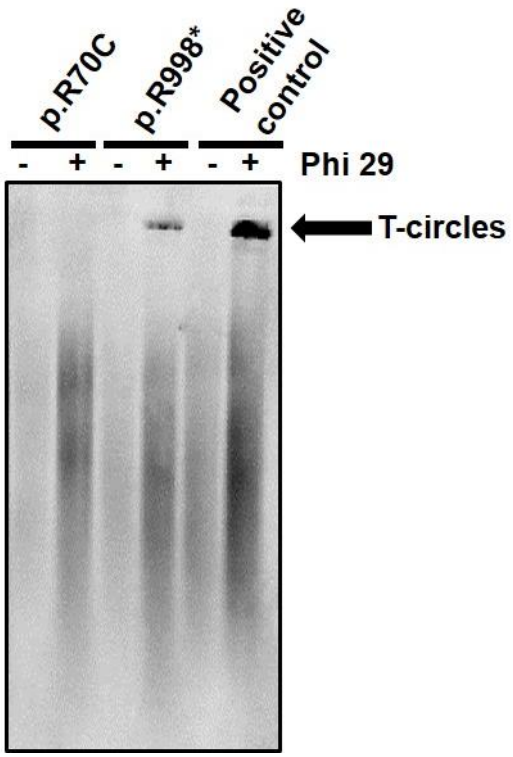


Figure 5.11. T-circle amplification using Phi29 polymerase detected by Southern blot analysis. Samples: p.R70C - patient with sporadic DC carrying this variant of unknown significance (patient 13 in Table 5.1); p.R998* - proband of Family 8 carrying this LoF variant (Table 5.1); positive control - genomic DNA extracted from WI-38 VA-13 cells (human fibroblasts), known to produce T-circle.

5.3 Discussion

In this study, we have undertaken analysis of *RTEL1* in 429 patients with DC, AA, MDS and related phenotypes. We have identified 35 variants, where 5 are biallelic (4 homozygous and 1 compound heterozygous) and 30 are heterozygous. Nine variants are novel. As a result, we have been able to probe further into the relationship between variants in the *RTEL1* gene and this spectrum of disease.

The initial disease association was made when biallelic *RTEL1* variants were shown to cause early onset of a severe form of DC and HH (Walne *et al*., 2013a; Jullien *et al*., 2016; Deng *et al*., 2013; Ballew *et al*., 2013). Herein, we describe five new biallelic families. Interestingly, in one of these the index case presented in adulthood with aplastic anaemia. In another, the index case presented with HH at the age two years, but in this patient, we consider the homozygous *RTEL1* variant to be likely benign because her mother who was asymptomatic was also homozygous for this variant (Family 5.8).

In amongst the earlier papers, there was an exceptional pair of sibs, both heterozygous for an *RTEL1* variant who presented with a severe phenotype (Ballew *et al*., 2013). Since then, several papers have clearly shown that heterozygous variants in *RTEL1* are more often associated with pulmonary fibrosis, occurring later in life, often in large families, but with variable penetrance (Kannengiesser *et al*., 2015; Cogan *et al*., 2015; Stanley *et al*., 2016). The frequency of heterozygous *RTEL1* variants in idiopathic pulmonary fibrosis as a whole is yet to be firmly established.

In this study, we have extended the phenotype associated with heterozygous loss of function *RTEL1* variants to include late onset of MDS and liver disease (Families 6, 8

and 9, Figure 5.9). This triad of late onset pulmonary, haematological and liver disease is very reminiscent of that established for heterozygous variants in another telomere related gene, *TERT,* which can also present with a severe early onset disease when the variants are biallelic.

The families we present clearly illustrate the variable penetrance of heterozygous *RTEL1* variants. This is exemplified by Family 6 where the index case had DC features, which did not become apparent until age 77 years. His daughter had liver disease at age of 52 years, and segregation analysis identified four asymptomatic carriers at age below 50 years old. This family highlights not only variable penetrance of heterozygous LoF variants but also suggests a late onset disease predisposition. The same *RTEL1* variant was identified in Family 7, where it was associated with MDS and nail dystrophy in the 45 years old index case. Interestingly, this is the same variant as reported by Ballew *et al*. (2013) in a heterozygous state as being the cause of HH in two siblings (aged three and one years) with very short telomeres. In that family, the mother also harboured the variant and had short telomeres but was asymptomatic.

Indeed, in most of the families where the index case has disease due to biallelic *RTEL1* variants, both here and in previous reports, the heterozygous parents are generally asymptomatic. However, we must now note that these individuals may nevertheless be predisposed to developing disease in their later years. This is suggested by Family 2 (with p.G1096W) where there is a history of pulmonary disease in the grandmother in her 70s and for the R998* variant which has been seen in both severe recessive (Walne *et al*., 2013a; Deng *et al*., 2013 and Ballew *et al*., 2013) and late onset dominant settings (Family 7 and Cogan *et al*., 2015). Thus, it is important to be careful when counselling families.

Previously we reported the recurrent missense variant c.2941C>T; p.R981W (MAF: 6 in 119,930 in ExAC) as a compound heterozygote in three young probands (under 12 years old) from unrelated families causing HH (Walne *et al*., 2013a). Here, we observed the same variant in a heterozygous state in a 24-year-old patient with AA from a consanguineous family (patient 14 in Table 5.1). In this case, there is no strong evidence that this variant (c.2941C>T; p.R981W) is the cause of AA on account of the relatively high frequency of this variant in the ExAC population. However, we do note the short telomeres in this patient and the very high CADD score of this variant, indicating the possibility that it acts as a risk factor for disease.

A key point arises, therefore, when a patient presents with an *RTEL1* variant, as to whether or not it should be considered pathogenic, as there are a multitude of rare coding *RTEL1* variants in the population at large. Using the ExAC database, the sum of number of very rare heterozygous coding alleles (at a frequency of <0.0001) is 1,195 in an average of approximately 56,700 people. This is significantly lower than the number very rare coding variants that we have identified in our cohort (22 in 429 patients, Fisher's exact test, P = 0.003), but on a case-by-case basis this background poses a problem.

In addition to looking at the ExAC database for population frequency there are several parameters that we have used to assign pathogenic status. The association of the rare variant with the pathology is a given, if the patient under review is presenting with one the *RTEL1* related disease features. Telomere length measurement is now widely used, and our experience here is that the heterozygotes, who are often more elderly, may have telomere lengths that are short, but not necessarily very short. We have also looked at T-circles, and shown that in some cases their presence is clearly increased where there is a LoF variant compared to a common missense variant.

However, this test is not very ‘user-friendly’ and a normal range has not been established. The *in silico* prediction tools are helpful and improving, but remain a guide, and by no means a definitive test.

Finally, the segregation of the variant with disease can be decisive. This is more often the case in exclusion rather than inclusion as we show in Family 5, where the index case presented with HH and a novel homozygous missense variant which was predicted to be damaging. However, this did not segregate with the disease in this family as the asymptomatic mother is also homozygous for this same variant. This suggests that this homozygous *RTEL1* variant is not disease causing in this index case. No other candidate genes were identified in this family by whole exome sequencing.

In summary, this study identifies several important observations. Firstly, heterozygous LoF *RTEL1* variants are associated with myelodysplasia and liver disease in adulthood. Secondly, biallelic *RTEL1* variants can present with just bone marrow failure in adulthood. Thirdly, many heterozygous variants and even some biallelic *RTEL1* variants are likely benign. Therefore, in order to assign an accurate status to each *RTEL1* variant, detailed clinical and laboratory studies are necessary.

Chapter 6

Variants identified in familial MDS/AML candidate genes

6.1 Introduction

Whole exome sequencing (WES) entails the capture, sequencing and analysis of all protein coding genes in the human genome. Regardless of recent improvement in databases and software tools, interpretation of damaging variants and variants of unknown significance (VUS) is one of the major challenges presented by WES data analysis. This technique generates a long list of variants with a large number of them likely to have no known clinical significance. Nevertheless, WES can be used as a diagnostic approach for the identification of genetic alterations in patients with suspected inherited disorders. In this chapter, we present the analysis we have performed on our WES data to search for the gene responsible for the disease in our uncharacterised MDS/AML families.

6.1.1 Genetically uncharacterised familial MDS/AML patients

In our cohort (2018) there are 68 genetically uncharacterised MDS/AML families. Whole exome sequencing was performed on a total of 51 unrelated cases in 42 of these MDS/AML families (Figure 6.1). Initially, WES was performed in 19 MDS/AML families and this number was expanded at different times during the course of this study to include the new MDS/AML families recruited into our cohort. Therefore, data analysis was undertaken at different times under various criteria to try to identify the underlying disease causing gene in these families.

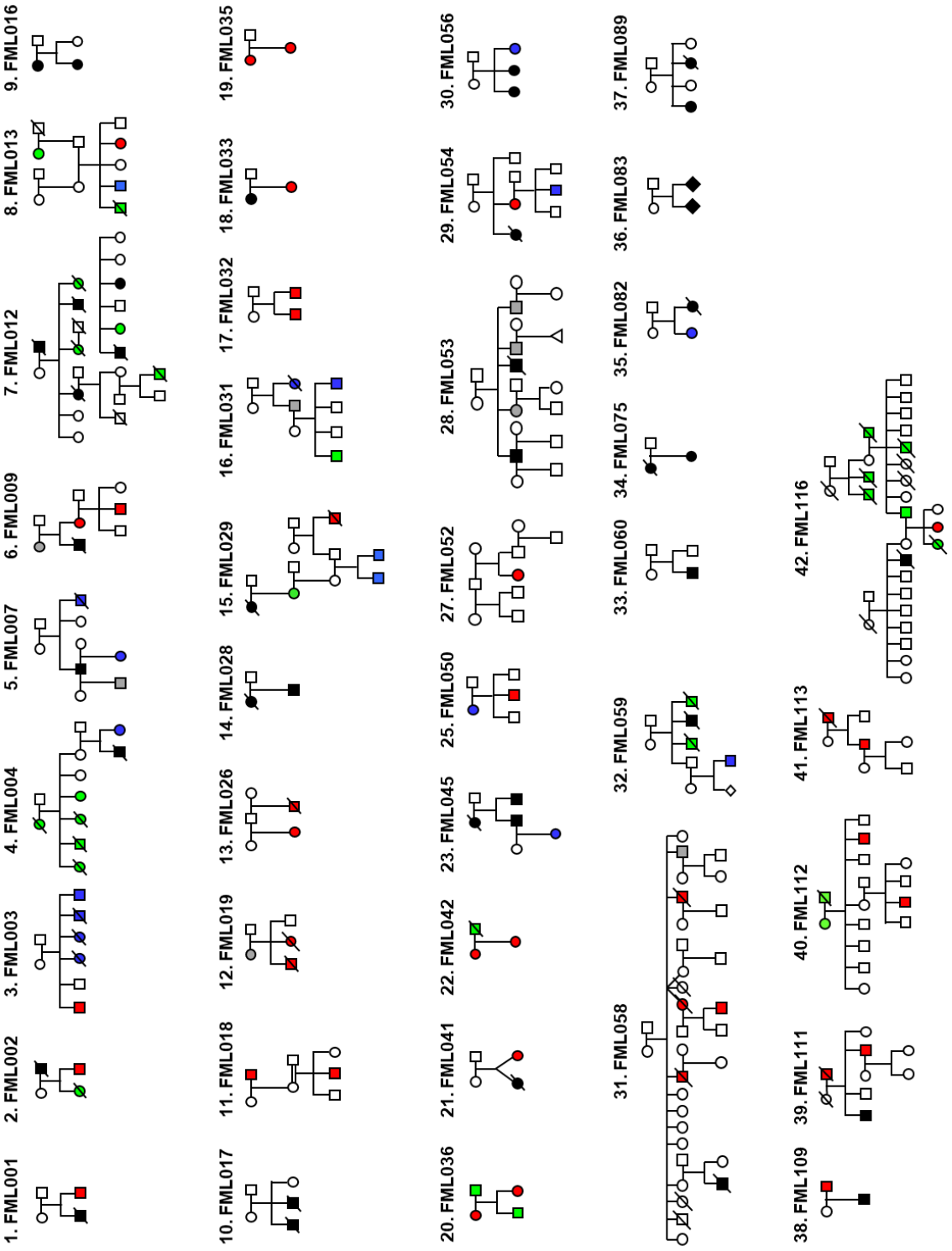


Figure 6.1. All MDS/AML families with WES done. Affected individuals are coloured as follows: red: MDS, black: AML, blue: AA, grey: thrombocytopenia, green: other leukaemia/cancer. Note: families with MDS/AML FML047 and FML051 have no family trees due to lack of sufficient family information.

DNA or blood samples on our familial MDS/AML cohort came from all over the world and it constitutes patients who presented at different ages at the time of diagnosis ranging from newborn to adulthood. Most index cases had undergone prior genetic testing, which could consist of chromosomal analysis, DNA sequencing studies, telomere length assays, chromosomal breakage analysis or a combination of tests. Available samples of parents, siblings, cousins or uncles/aunties were sequenced in approximately 30% of the cohort, in the remaining 70% of the families principally the index case has been analysed.

6.2 Results

6.2.1 Data analysis

Analyses of the generated WES sequence data has been undertaken assuming an autosomal dominant (AD) or an autosomal recessive (AR) pattern of inheritance depending on which group of MDS/AML families were analysed. This is because two families of our cohort are likely to have an AR pattern of inheritance in which there was history of consanguinity (Figure 6.1) and, additionally four other MDS/AML families could be analysed as AD or AR due a lack of family information on them (Figure 6.1). Multistep filtering was used to select potentially damaging variants in canonical transcripts of putative candidate disease genes on the basis of allele frequency, functional consequence and occurrence in multiple unrelated families (Figure 6.2). Depending on the set of criteria selected at this point, different candidate genes can be proposed.

In the AD analysis, WES data was firstly filtered for novel germline heterozygous variants when compared against the Exome Sequencing Project (ESP) and 1000 Genomes Project (1000G) databases, which are publicly available data sets of human DNA sequence variation. Exome Variant Server (EVS) was created as part of the National Heart, Lung, And Blood Institute Exome Sequencing Project and contains frequency information spanning 6,503 exomes (http://evs.gs.washington.edu/EVS/), while 1000G includes individual-level genotype data from whole-genome and exome sequence data for 2,504 individuals (http://www.internationalgenome.org/). Subsequently, allele frequency was verified against ExAC as it is approximately 10 fold bigger than the former databases.

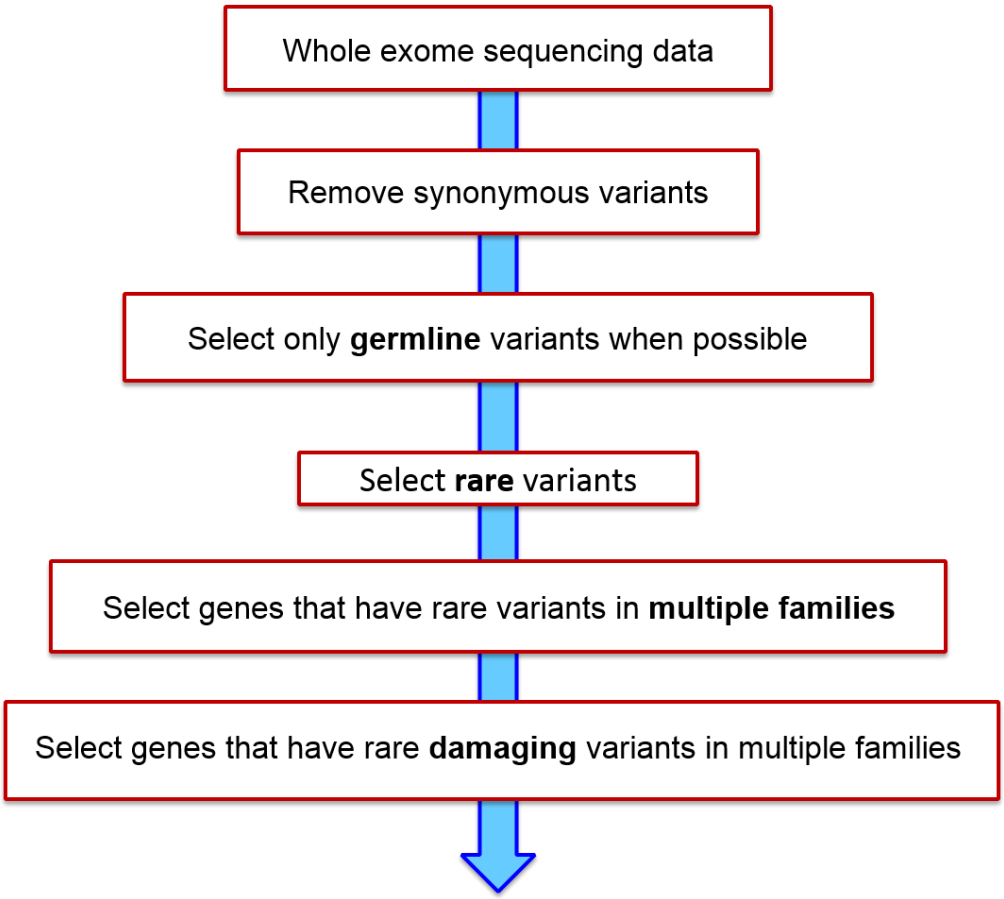


Figure 6.2. Workflow of a typical autosomal dominant analysis of WES data.

It is known that the efficacy of filtering of candidate disease variants by frequency depends on the size of the data and its ancestral diversity data (Lek *et al*., 2016). For this reason, we used the ExAC database to filter rare variants (novel, MAF≤0.001 or MAF≤0.0001) out of our WES data. As described elsewhere ExAC is an aggregation of high quality exome data for 60,706 individuals of diverse ancestries. This catalogue of human genetic diversity contains an average of one variant every eight bases of the exome (http://exac.broadinstitute.org/). For the AR analysis, the variants were filtered for rare germline biallelic variants (MAF≤0.001) when compared against the databases mentioned above.

More than 96% of the variants selected were confirmed by Sanger sequencing. The remaining 4% variants were found to be false positive results and these usually would be calls that had unequal allele fractions, poor mapping scores, or sequence data indicating suboptimal alignment to the reference sequence. Positive cases met each of the diagnostic criteria regarding variant severity, appropriate inheritance patterns (when other family member's data were available), and disease-phenotype concordance. The candidate gene description in this chapter was carried out using NCBI - National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/), UniProt (http://www.uniprot.org/), OMIM - Online Mendelian Inheritance in Man (https://www.omim.org/) and ExAC databases.

6.2.1.1 All MDS/AML families

Analysis of WES data of 30 MDS/AML families was performed assuming autosomal dominant as pattern of inheritance. A total of 34,888 variants somewhat rare (MAF≤0.05) variants were found by comparison with the current reference haploid human genome sequence (human genome 19). Of these, 8,100 were novel non-silent variants compared against ESP and 1000G project databases and further filtering revealed 3,857 variants segregating within families (when samples from more than one member were analysed, which was undertaken in 12 families). These putative germline variants were than filtered by rare allele frequency (MAF≤0.0001) compared against ExAC. Pathogenicity of the selected variants was assessed using predictable damaging tools such as Polyphen2 (http://genetics.bwh.harvard.edu/pph2/) and MutationTaster2 (http://www.mutationtaster.org/). As a result, 12 candidate genes with probably damaging rare heterozygous variants were identified in at least three MDS/AML families (Table 6.1) and 22 candidate genes with probably damaging rare

heterozygous variants were identified in at least two MDS/AML families (Table 6.2). Interestingly, 10 candidate genes harbouring rare putative germline LoF variants (including stopgain, frameshift indel and variants within the consensus splice site dinocleotides) were identified in at least two unrelated MDS/AML families (Table 6.3). Genes known to be associated with another AD disorder were removed. After all this filtering two candidate disease genes harbouring probably damaging rare heterozygous variants were selected (Table 6.4 and Figure 6.3). Detailed description of these candidate genes is shown in Table 6.5. Unfortunately, there is insufficient evidence to date to conclusively say whether any of these can be definitively considered as the disease causing gene in our families. Hence, these probably damaging rare heterozygous variants identified in the selected candidate genes (Table 6.4 and 6.5) are currently considered to be of unknown significance.

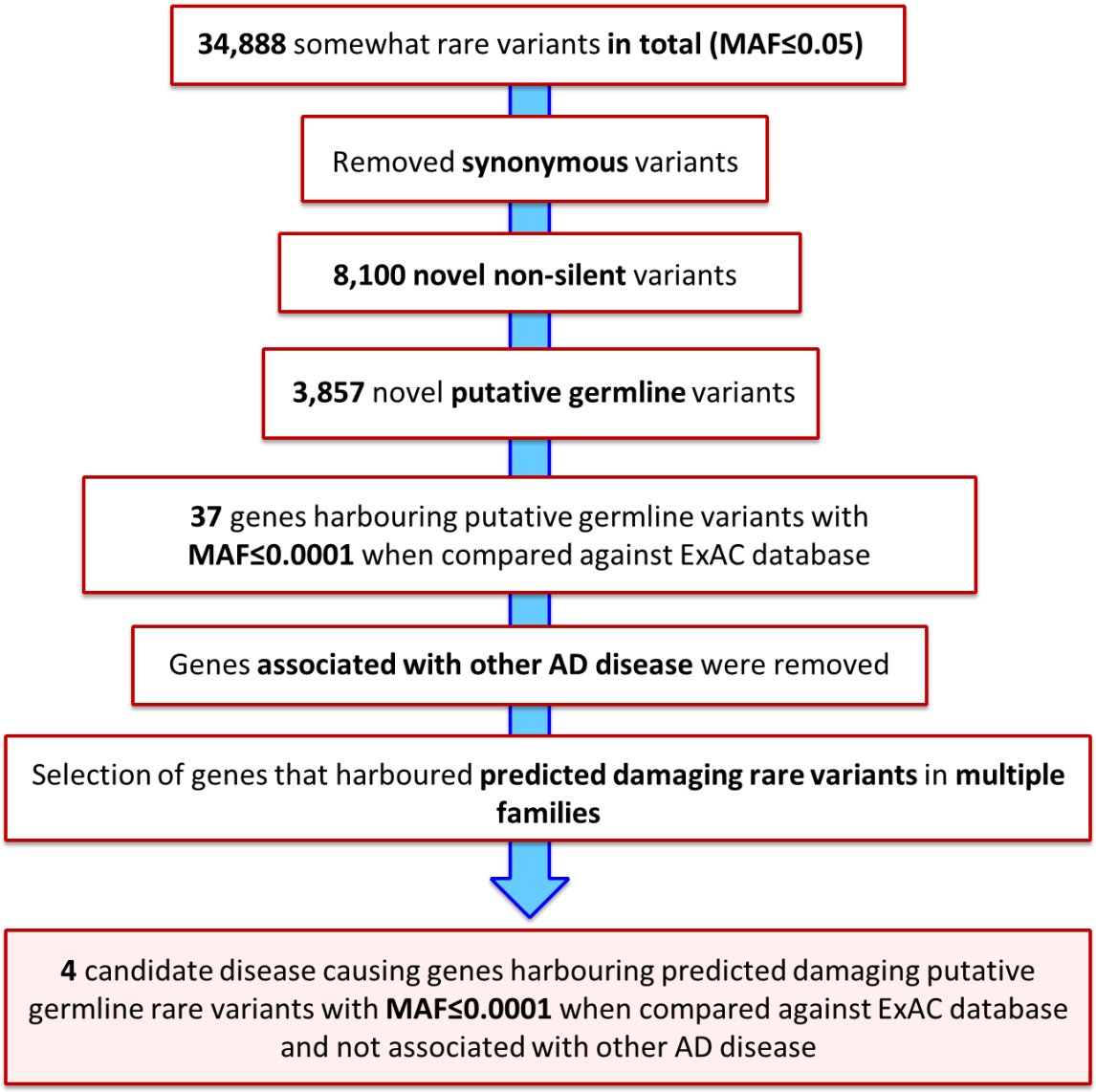


Figure 6.3. Workflow of the autosomal dominant analysis of WES data of 30 uncharacterised MDS/AML families.

Table 6.1. Rare possibly damaging variants (MAF≤0.0001) in the same gene in at least three MDS/AML families

Autosomal dominant analysis MDS/AML families					
	Samples	**Gene**	**AAChange**	**Exac (MAF)**	**PolyPhen2**	**MutationTaster2**
1	FML019	AHNAK2	ENST00000333244:exon7:c.G6634T:p.V2212F	0.00002488	D	B
	FML051	AHNAK2	ENST00000333244:exon7:c.G6408C:p.Q2136H	0.00004585	P	B
	FML056	AHNAK2	ENST00000333244:exon7:c.C16699T:p.Q5567*	NR	NA	D
2	FML048	ATP13A4	ENST00000342695:exon30:c.A3482C:p.D1161A	NR	D	D
	FML050	ATP13A4	ENST00000342695:exon8:c.T752A:p.L251H	NR	D	D
	FML016	ATP13A4	ENST00000342695:exon26:c.G2891A:p.G964E	NR	P	B
3	FML056	CBY3	ENST00000376974:exon2:c.C647T:p.A216V	0.00006702	P	B
	FML059	CBY3	ENST00000376974:exon2:c.C586T:p.R196C	NR	P	D
	FML054	CBY3	ENST00000376974:exon2:c.A329G:p.N110S	NR	P	B
4	FML019	DNAH11	ENST00000328843:exon64:c.C10262T:p.T3421M	0.00005983	P	B
	FML060	DNAH11	ENST00000328843:exon49:c.C7897G:p.P2633A	NR	P	D
	FML056	DNAH11	ENST00000328843:exon60:c.G9734A:p.W3245*	NR	NA	D
5	FML057	DNAH6	ENST00000389394:exon38:c.C6136T:p.R2046W	NR	D	D
	FML017	DNAH6	ENST00000389394:exon7:c.G1091A:p.R364Q	0.00009052	P	D
	FML077	DNAH6	ENST00000389394:exon5:c.A709G:p.S237G	0.0001893	D	D
6	FML033	FAT1	ENST00000441802:exon2:c.T1529C:p.F510S	NR	P	D
	FML042	FAT1	ENST00000441802:exon2:c.T2486C:p.V829A	NR	P	D
	FML002	FAT1	ENST00000441802:exon10:c.A6044G:p.H2015R	NR	B	B
	FML059	FAT1	ENST00000441802:exon3:c.G3487C:p.A1163P	NR	D	D
7	FML043	LRP2	ENST00000263816:exon39:c.A7205C:p.H2402P	NR	P	B
	FML059	LRP2	ENST00000263816:exon39:c.A6982G:p.I2328V	NR	B	B
	FML035	LRP2	ENST00000263816:exon37:c.A6149G:p.K2050R	NR	P	D
	FML018	LRP2	ENST00000263816:exon15:c.G2081A:p.G694D	NR	D	D
8	FML001	LTK	ENST00000263800:exon10:c.C1257A:p.H419Q	0.000009514	P	B
	FML056	LTK	ENST00000263800:exon16:c.G1964T:p.G655V	NR	D	D
	FML056	LTK	ENST00000263800:exon17:c.2086delC:p.L696fs	NR	NA	D
	FML109	LTK	ENST00000263800:exon11:c.1357_1359del:p.453_453del	NR	NA	NA
9	FML049	RYR1	ENST00000359596:exon34:c.C5216T:p.P1739L	NR	B	B
	FML059	RYR1	ENST00000359596:exon71:c.G10478A:p.R3493H	0.00000827	D	B
	FML110	RYR1	ENST00000359596:exon74:c.G10883A:p.R3628H	NR	D	B
	FML075	RYR1	ENST00000359596:exon34:c.C5213A:p.T1738K	NR	D	B
10	FML033	SZT2	ENST00000562955:exon33:c.T4832A:p.L1611Q	0.00004119	P	D
	FML048	SZT2	ENST00000562955:exon64:c.G8870A:p.R2957Q	0.00004119	D	B
	FML017	SZT2	ENST00000562955:exon17:c.A2549G:p.Q850R	0.00006598	D	NA
	FML017	SZT2	ENST00000562955:exon3:c.A158G:p.Q53R	0.00001649	B	B
11	FML054	TNRC6C	ENST00000335749:exon13:c.C3646T:p.P1216S	NR	P	D
	FML109	TNRC6C	ENST00000335749:exon12:c.G3469T:p.A1157S	NR	NA	D
	FML018	TNRC6C	ENST00000335749:exon3:c.A1475C:p.K492T	NR	D	D
12	FML042	ZNF845	ENST00000458035:exon4:c.2214_2215insA:p.L738fs	NR	NA	D
	FML060	ZNF845	ENST00000458035:exon4:c.G1619A:p.R540H	NR	D	B
	FML109	ZNF845	ENST00000458035:exon4:c.C860T:p.T287I	0.00002481	D	B

MAF: minor allele frequency; NR: not registered; P: possibly damaging; B: benign; D: damaging. Predicted damaging variants are highlighted in yellow.

Table 6.2. Rare possibly damaging rare heterozygous variants (MAF≤0.0001) in the same gene in at least two MDS/AML families

Autosomal dominant analysis MDS/AML families						
	Samples	**Gene**	**AAChange**	**Exac (MAF)**	**PolyPhen2**	**MutationTaster**
1	FML049	ABCA13	ENST00000435803:exon17:c.G3907A:p.D1303N	NR	B	B
	FML110	ABCA13	ENST00000435803:exon17:c.A2127C:p.L709F	NR	B	B
	FML013	ABCA13	ENST00000435803:exon31:c.G10594A:p.E3532K	NR	D	D
	FML035	ABCA13	ENST00000435803:exon7:c.C722T:p.S241L	NR	D	B
2	FML019	ARHGAP23	ENST00000431231:exon7:c.C793T:p.R265W	0.0001392	P	B
	FML002	ARHGAP23	ENST00000431231:exon12:c.C2410G:p.R804G	NR	P	B
	FML031	ARHGAP23	ENST00000431231:exon24:c.G3548C:p.R1183P	NR	B	B
3	FML043	BIN1	ENST00000316724:exon1:c.G52A:p.V18M	0.00003711	B	D
	FML059	BIN1	ENST00000316724:exon10:c.A791G:p.N264S	0.00005182	B	D
4	FML019	COL27A1	ENST00000356083:exon14:c.C2461T:p.L821F	0.00009077	B	D
	FML060	COL27A1	ENST00000356083:exon37:c.G3649A:p.E1217K	0.00005202	NA	D
5	FML048	CPEB2	ENST00000538197:exon1:c.C1049T:p.P350L	NR	NA	D
	FML050	CPEB2	ENST00000538197:exon1:c.1448_1450delGCG:p.483_484del	NR	NA	B
	FML077	CPEB2	ENSG00000137449:ENST00000538197:exon1:c.A427T:p.S143C	NR	NA	B
6	FML049	DUOX1	ENST00000321429:exon19:c.G2255C:p.R752T	0.00001652	B	D
	FML109	DUOX1	ENST00000321429:exon30:c.G3829A:p.V1277M	0.0001072	D	D
7	FML060	FAM107B	ENST00000181796:exon4:c.A707G:p.K236R	0.00004119	D	D
	FML054	FAM107B	ENST00000181796:exon1:c.C286T:p.R96C	0.000008237	P	B
8	FML056	FANCA	ENST00000389301:exon6:c.C553A:p.L185I	0.000008253	D	D
	FML075	FANCA	ENST00000389301:exon20:c.C1805G:p.A602G	NR	P	B
9	FML019	GRIK1	ENST00000399907:exon17:c.C2609A:p.A870D	0.000008361	B	B
	FML056	GRIK1	ENST00000399907:exon2:c.C256A:p.L86I	NR	B	D
	FML013	GRIK1	ENST00000399907:exon12:c.A1690T:p.N564Y	NR	D	D
10	FML049	HERC1	ENST00000443617:exon78:c.C14462G:p.S4821C	NR	D	D
	FML059	HERC1	ENST00000443617:exon10:c.A2080G:p.M694V	NR	D	D
11	FML043	IFT172	ENST00000260570:exon13:c.A1306G:p.M436V	0.000008239	B	D
	FML045	IFT172	ENST00000260570:exon35:c.C3907T:p.R1303*	NR	NA	D
	FML077	IFT172	ENST00000260570:exon16:c.A1585G:p.M529V	0.00003295	B	B
12	FML002	KAT6A	ENST00000396930:exon18:c.4989_4991del:p.1663_1664del		NA	NA
	FML077	KAT6A	ENST00000396930:exon18:c.A5740G:p.M1914V	0.0001648	D	B
13	FML007	KMT2D	ENST00000301067:exon33:c.G8291A:p.G2764E	NR	D	NA
	FML043	KMT2D	ENST00000301067:exon39:c.10821_10823del:p.3607_3608del	NR	NA	NA
	FML002	KMT2D	ENST00000301067:exon8:c.G1076A:p.R359H	NR	B	NA
	FML029	KMT2D	ENST00000301067:exon10:c.C2233T:p.P745S	NR	B	NA
	FML035	KMT2D	ENST00000301067:exon48:c.CFML0073G:p.T5018R	NR	D	NA

Table 6.2. Continued

14	FML019	KIAA1549L	ENST00000321505:exon1:c.G607A:p.D203N	0.00003316	B	B
	FML060	KIAA1549L	ENST00000321505:exon6:c.G3046A:p.V1016I	0.00005797	P	B
	FML029	KIAA1549L	ENST00000321505:exon1:c.C404T:p.T135I	NR	B	B
	FML031	KIAA1549L	ENST00000321505:exon20:c.C5456T:p.T1819I	NR	P	D
15	FML045	MYCBP2	ENST00000544440:exon71:c.T12119C:p.I4040T	NR	B	D
	FML016	MYCBP2	ENST00000544440:exon22:c.G3106C:p.V1036L	0.00004126	P	D
16	FML012	PCSK5	ENST00000545128:exon30:c.G3871A:p.A1291T	0.00001747	D	D
	FML077	PCSK5	ENST00000545128:exon37:c.A5239G:p.I1747V	0.00003477	B	B
	FML033	PCSK5	ENST00000545128:exon34:c.4668-2A≥G	NR	NA	NA
17	FML043	PIEZO1	ENST00000301015:exon37:c.G5140A:p.V1714I	0.00005702	P	D
	FML056	PIEZO1	ENST00000301015:exon33:c.A4481C:p.E1494A	NR	B	B
	FML013	PIEZO1	ENST00000301015:exon36:c.G4856T:p.S1619I	NR	D	B
	FML017	PIEZO1	ENST00000301015:exon50:c.G7195A:p.G2399S	0.0001547	B	B
18	FML007	PIK3C2B	ENST00000367187:exon3:c.G751T:p.D251Y	NR	P	D
	FML019	PIK3C2B	ENST00000367187:exon25:c.A3536T:p.Y1179F	0.00005107	P	B
19	FML033	PLEKHG4	ENST00000360461:exon19:c.A3352G:p.R1118G	NR	B	B
	FML059	PLEKHG4	ENST00000360461:exon12:c.C1832T:p.A611V	NR	P	B
	FML109	PLEKHG4	ENST00000360461:exon6:c.C1000T:p.R334*	NR	NA	D
20	FML007	SPTBN1	ENST00000356805:exon20:c.A4201T:p.S1401C	NR	P	D
	FML060	SPTBN1	ENST00000356805:exon26:c.A5276G:p.N1759S	0.00002472	NA	D
21	FML048	TLDC1	ENST00000343629:exon5:c.G574T:p.D192Y	NR	D	D
	FML057	TLDC1	ENST00000343629:exon6:c.G995C:p.S332T	0.00000826	B	B
	FML016	TLDC1	ENST00000343629:exon8:c.G1363A:p.D455N	0.000008776	NA	D
22	FML001	VPS13C	ENST00000261517:exon15:c.G1261C:p.V421L	NR	B	B
	FML056	VPS13C	ENST00000261517:exon59:c.A7694T:p.K2565M	NR	D	B
	FML075	VPS13C	ENST00000261517:exon49:c.T5846G:p.L1949R	0.000008258	D	D
	FML017	VPS13C	ENST00000261517:exon17:c.1291-4A≥G	0.0004363	NA	NA

MAF: minor allele frequency; NR: not registered; P: possibly damaging; B: benign; D: damaging. Predicted damaging variants are highlighted in yellow.

Table 6.3. Rare heterozygous loss of function variants (MAF≤0.0001) in the same gene in at least two families

	LoF analysis MDS/AML families				
	Samples	**ExonicFunc**	**Gene**	**AAChange**	**ExAC (MAF)**
1	FML056	nonframeshift deletion	AGAP2	ENST00000547588:exon1:c.715_726del:p.239_242del	NR
	FML059	nonframeshift deletion	AGAP2	ENST00000547588:exon1:c.873_875del:p.291_292del	NR
2	FML019	frameshift insertion	ADA	ENST00000372874:exon4:c.259_260insT:p.V87fs	NR
	FML056	stopgain SNV	ADA	ENST00000372874:exon1:c.C7T:p.Q3*	NR
3	FML049	stopgain SNV	AMIGO3	ENST00000535833:exon10:c.C1203G:p.Y401*	NR
	FML059	stopgain SNV	AMIGO3	ENST00000535833:exon10:c.C669A:p.C223*	0.0001019
4	FML001	frameshift deletion	FAM200B	ENST00000422728:exon2:c.24_25del:p.8_9del	NR
	FML077	stopgain SNV	FAM200B	ENST00000422728:exon2:c.G1164A:p.W388*	NR
5	FML042	stopgain SNV	LOXL2	ENST00000389131:exon5:c.C938G:p.S313*	NR
	FML033	unknown	LOXL2	ENST00000389131:exon7:c.967-2A≥G	NR
6	FML056	frameshift deletion	LTK	ENST00000263800:exon17:c.2086delC:p.L696fs	NR
	FML109	nonframeshift deletion	LTK	ENST00000263800:exon11:c.1357_1359del:p.453_453del	NR
7	FML110	frameshift insertion	NPM1	ENST00000296930:exon11:c.859_860insTCTG:p.L287fs	NR
	FML050	nonframeshift deletion	NPM1	ENST00000296930:exon7:c.525_527del:p.175_176del	NR
8	FML056	nonframeshift deletion	RP11-723O4.6	ENST00000508239:exon6:c.1413_1415del:p.471_472del	NR
	FML059	nonframeshift deletion	RP11-723O4.6	ENST00000508239:exon6:c.1396_1398del:p.466_466del	NR
9	FML049	stopgain SNV	SI	ENST00000264382:exon9:c.C919T:p.Q307*	NR
	FML056	splicing	SI	ENST00000264382:exon46:c.5109-3T≥C	0.00001653
10	FML049;FML043	nonframeshift deletion	SRRM2	ENST00000301740:exon11:c.7654_7656del:p.2552_2552del	NR

Genes MUC, TTN and splicing variants on position 4 or over were removed. MAF: minor allele frequency; NR: not registered.

Table 6.4. Candidate genes with possibly damaging rare heterozygous variants (MAF<0.0001) in 3 or more families and not associated with other AD disorders

	Gene	Families	Segregating within a family	Variant	AA Change	ExAC (MAF)	PolyPhen2	MutationTaster2	Features of the affected families
1	*FAT1*	FML033	N/A	c.1529 T>C	p.Phe510Ser	NR	P	D	AD; MDS,AML
		FML002	N/A	c.6044 A>G	p.His2015Arg	NR	B	N	AD; AML, AA, other cancer
		FML042	N/A	c.2486 T>C	p.Val829Ala	NR	P	D	AD; MDS
		FML059	N/A	c.3487 G>C	p.Ala1163Pro	NR	D	D	AD; AA, other ca, leukaemia
2	*TNRC6C*	FML054	yes	c.3646 C>T	p.Pro1216GSer	NR	P	D	AD; AA, acute leukaemia
		FML109	yes	c.3469 G>T	p.Ala1157Ser	NR	NA	D	AD; AML
		FML018	N/A	c.1475 A>C	p.Lys492Thr	NR	D	D	AD; MDS

MAF: minor allele frequency; P: possibly damaging; B: benign; D: damaging; AD: Autosomal dominant; AR: Autosomal recessive; N/A: not available.

Table 6.5. Description of the candidate genes with possibly damaging rare heterozygous variants (MAF≤0.0001) in 3 or more families

	Gene	Additional information on the gene	Disease associated
1	*FAT1* (FAT atypical cadherin 1) novel variants	FAT1 is a receptor for a signalling pathway that regulates growth, gene expression, and cell polarity. FAT1 is upregulated in 11% of AML, 29% of preB acute lymphoblastic leukaemia (ALL) and 63% of T-cell acute lymphoblastic leukaemia (T-ALL). Morris *et al.* (2013) reported that FAT1 encodes a cadherin-like protein, which was able to potently suppress cancer cell growth *in vitro* and *in vivo* by binding beta-catenin and antagonizing its nuclear localization; suggesting FAT1 as a tumor suppressor gene. ENST00000441802:NM_005245.3 → NP_005236.2, 4588aa	none
2	*TNRC6C* (trinucleotide repeat containing 6C)	Plays a role in RNA-mediated gene silencing by micro-RNAs (miRNAs). Required for miRNA-dependent translational repression of complementary mRNAs by argonaute family proteins. As scaffolding protein associates with argonaute proteins bound to partially complementary mRNAs and simultaneously can recruit CCR4-NOT and PAN deadenylase complexes (Zipprich *et al.*, 2009; Braun *et al.*, 2011; Fabian *et al.*, 2011 and Chekulaeva *et al.*, 2011). ENST00000335749:NM_001142640.1 → NP_001136112.1, 1726aa	none

AD: Autosomal dominant; AR: Autosomal recessive. Information associated with leukaemia is highlighted in red.

6.2.1.2 Family FML012

Thirteen MDS/AML families had WES performed in more than one family member (FML009, FML012, FML029, FML031, FML035/FML073, FML036, FML054, FML061, FML075, FML081, FML109 and, FML112 seen in Figure 6.1). Here, we describe the autosomal dominant analysis of WES data of family FML012, specifically from individuals III:7 and the index case IV:2 (Figure 6.4). This is a large family with MDS/AML history along with various solid cancers. The index case (IV:2, Figure 6.4) was a male who died of ALL at age 36 years. His grandmother (II:3, Figure 6.4) died of AML at the aged 46 years and his great grandfather (I:2, Figure 6.4) also died of AML at age 50 years. One of his grandaunts (II:5, Figure 6.4) had stomach cancer and another grandaunt had pancreatic cancer (II:8, Figure 6.4). His granduncle (II:7, Figure 6.4) had AML. Furthermore, the index case had two cousins with AML (III:4 and III:7, Figure 6.4) and another with cervical cancer (III:5, Figure 6.4). Conveniently, the number of germline variants shared between these two cousins are diminished as they are distantly related.

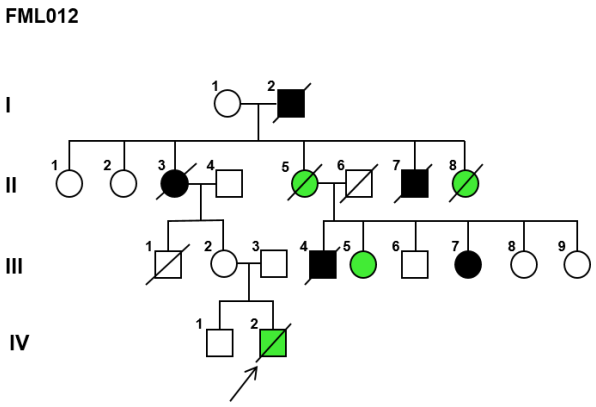


Figure 6.4. MDS/AML family tree of FML012. Affected individuals are coloured as following, black: AML and green: other cancer. The arrow highlights index case.

WES of the proband (IV:2, Figure 6.4) and his cousin (III:7, Figure 6.4) revealed an overall of 461 non-silent rare heterozygous variants with MAF≤0.0001, where 446 were novel when compared to ESP and 1000G and 32 were loss of function variants (including stopgain, frameshift indel and variants within the consensus splice site dinocleotides). Strikingly, ten germline variants were found in total in family FML012 (Table 6.6) with just two rare germline heterozygous variants in the canonical transcripts with MAF≤0.0001 in genes that are not associated with other autosomal dominant disease (*CDKL1* and *CEP68*) (Table 6.7). Consequently, cyclin-dependent kinase-like 1 (*CDKL1*) and centrosomal protein 68kDa (*CEP68*) are the only candidate genes found in FML012 in this analysis.

Additionally, dHPLC was performed on 30 constitutional MDS/AML cases as described in section 3.2.2.3 in chapter 3 for both genes. No further rare heterozygous variants in neither of the genes *CDKL1* nor *CEP68* were found.

Unfortunately, neither *CDKL1* nor *CEP68* presented enough evidence to be considered the underlying cause of disease in this family. For instance, Sanger sequencing of the rare heterozygous variant (c.416C>T; p.S139F) located in CDKL1 domain performed in two additional members (III:8 and III:9, Figure 6.5 A) of FML012 revealed the presence of this variant in one of them (III:8) (Figure 6.5 A, B and C). III:8 (Figure 6.5 A) was diagnosed with arthritis and monoclonal gammopathy of unknown significance (MGUS). The amino acid serine in the position 139 in CDKL1 is well conserved among mammals (Figure 6.5 D) and this amino acid substitution to F is considered to be probably disease causing by MutationTaster2 but not by Polyphen2, where this variant is predicted to be benign (Table 6.8). Furthermore, no additional family from our cohort was found to have a germline variant in *CDKL1* to strengthen its case to be a disease gene.

Table 6.6. All germline heterozygous variants in FML012

	Gene	freq.controls	AAChange	ESP	X1000project	dbSNP	PolyPhen2	MutationTaster
1	CDKL1	0	c.C416T:p.S139F	NA	NA	NA	NA	D
2	CEP68	0	c.A1373G:p.Q458R	NA	NA	NA	D	B
3	GNLY	0.002340094	c.125delC:p.S42fs	0.004073	NA	NA	NA	D
4	HK2	0.000489237	c.G994A:p.G332S	0.001	NA	rs35677390	D	D
5	HTT	0	c.C55A:p.Q19K	NA	NA	NA	NA	B
	HTT	0.015060241	c.G57T:p.Q19H	NA	NA	NA	NA	B
6	IQSEC2	0	c.G1592C:p.R531P	0.000284	0.0012	rs149027201	D	D
7	MINK1	0.0033867	c.T893G:p.V298G	NA	NA	NA	D	D
8	PSD4	0	c.G1094A:p.C365Y	0.000923	0	rs45626940	D	B
9	TMC3	0.000280741	c.C2206T:p.R736*	0.000401	NA	rs201527196	NA	D
10	TMEM178B	0.004232804	c.G133A:p.A45T	NA	0.0046	rs139501428	NA	B

Freq.controls: allele frequency in the internal controls; ESP: Exome Variant Server available by National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP); X1000 project: 1000 Genomes Project; dbSNP: The Single Nucleotide Polymorphism Database from the National Center for Biotechnology Information (NCBI) in collaboration with the National Human Genome Research Institute (NHGRI).

Table 6.7. Rare germline heterozygous variants with MAF ≤0.0001 shared in both affected members (III:7 and IV:2) of FML012

	Gene	AAChange	ExAC (MAF)	PolyPhen2	MutationTaster
1	CEP68	c.A1373G:p.Q458R	NR	D	B
2	CDKL1	c.C416T:p.S139F	0.00001648	NA	D
3	HTT	c.C55A:p.Q19K	NR	NA	B
	HTT	c.G57T:p.Q19H	NR	NA	B

MAF: minor allele frequency; D: damaging; B: benign; NA: not available.

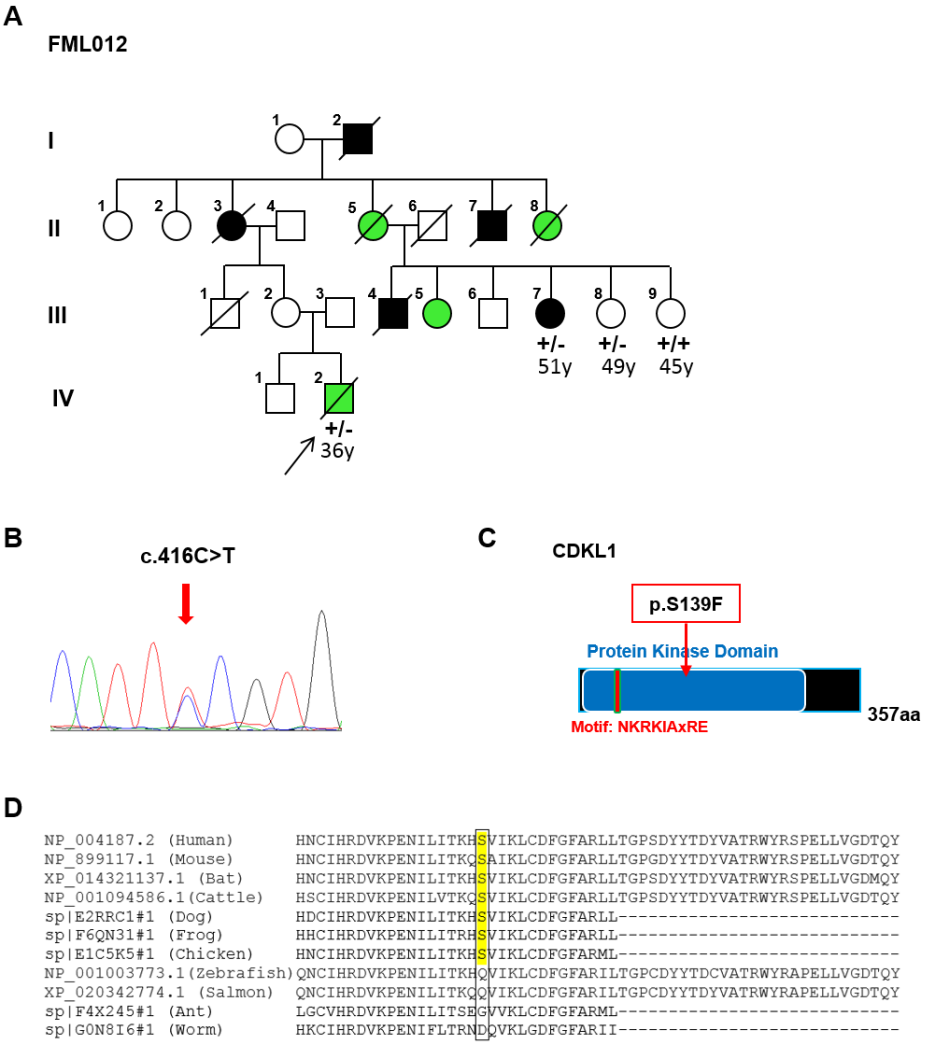


Figure 6.5. *CDKL1* variant analysis. A. MDS/AML family tree of FML012. Affected individuals are coloured as following; black: AML and green: other cancer. The arrow highlights index case. B. Sanger sequencing trace representing the nucleotide substitution in *CDKL1*. C. Schematic CDKL1 protein showing the location of the amino acid change. D. Multiple CDKL1 protein amino acid alignment in the position 139.

Table 6.8. Description of candidate genes identified in FML012

	CDKL1 - Cyclin-dependent kinase-like 1, chromosome 14; 68001bp; 14 exons; 358 aa	***CEP68*** - Centrosomal protein 68kDa, chromosome 2; 30648 bp; 8 exons; 757aa
OMIM	603441	616889
Protein	Cyclin-dependent kinase-like 1 is a member of a large family of CDC2-related serine/threonine protein kinases (Meyerson *et al.*, 1992). Protein serine/threonine kinase activity: Catalysis of the reactions: ATP + protein serine = ADP + protein serine phosphate, and ATP + protein threonine = ADP + protein threonine phosphate. Kinase activity: Catalysis of the transfer of a phosphate group, usually from ATP, to a substrate molecule. Transferase activity: Catalysis of the transfer of a group, e.g. a methyl group, glycosyl group, acyl group, phosphorus-containing, or other groups, from one compound (generally regarded as the donor) to another compound (generally regarded as the acceptor) (Taglienti *et al.*, 1996).	Centrosomal protein 68kDa is highly conserved and required for centrosome cohesion during interphase in mammalian cell cycle process (Graser *et al.*, 2007; Fang *et al.*, 2014). Although CEP68 variants have not been described in haematological cancer, aberrant centrosomes (structural and numeric aberrations) in a cell have been associated with various types of human cancer due to chromosomal instability and aneuploidy (Chan, 2011). Furthermore deregulation of any oncogenic and tumor suppressor proteins that are located in the centrosomes might induce centrosome abnormalities (Fukasawa, 2007).
Protein Domains	1 Domain: Pkinase; Protein kinase domain	1 Domain: PHA03247; large tegument protein UL36
Variant	c.416 C≥T; p.S139F ENST00000395834.5:NM_004196.4 → NP_004187.2	c.1373 A≥G; Q458R ENST00000377990.6:NM_015147.2 → NP_055962.2
ExAC (MAF)	2 in 121332; 0.00001648	Not registered
Polyphen2 score	0.003, benign	0.112, benign
Mutationtaster2 score	Disease causing	Benign

MAF: minor allele frequency.

Sanger sequencing of the novel heterozygous *CEP68* variant (c.1373A≥G; p.Q458R) located immediately after CEP68 domain, was performed on the additional other two members of FML012 (III:8 and III:9, Figure 6.6 A). The *CEP68* variant was identified in III:9 who is reported to be asymptomatic to date (Figure 6.6 A, B and C). The amino acid Q is conserved among mammals (Figure 6.6 D) and this variant is predicted to be benign by both MutationTaster2 and Polyphen 2 tools (Table 6.8). No additional family with similar phenotype was identified to harbor a rare heterozygous variant in *CEP68*. Further information on *CDKL1* and *CEP68* can be seen in Table 6.8.

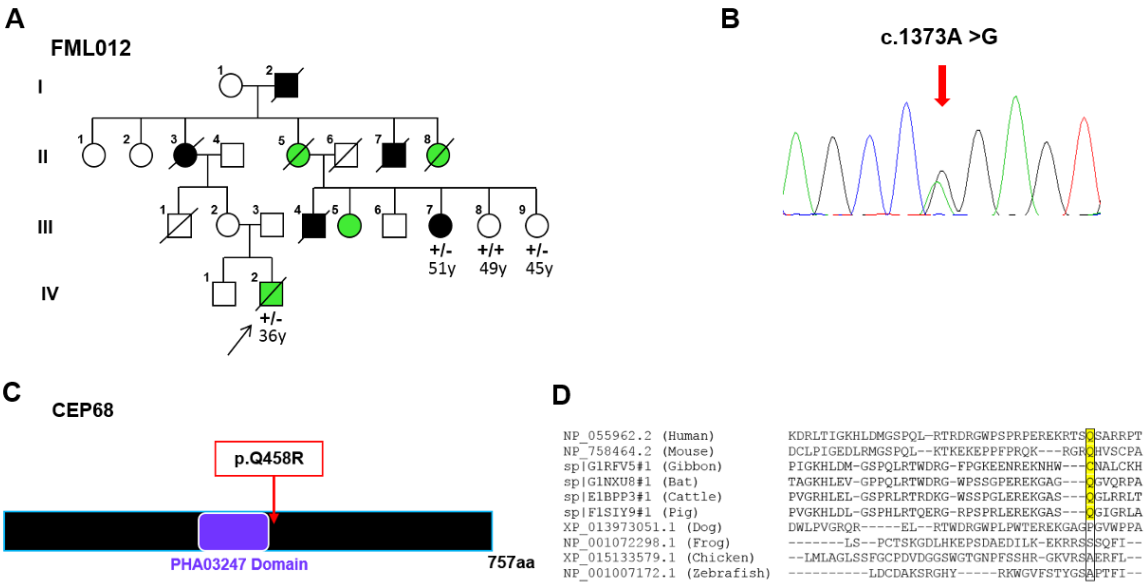


Figure 6.6. *CEP68* variant analysis. A. MDS/AML family tree of FML012. Affected individuals are coloured as following, black: AML and green: other cancer. The arrow highlights index case. B. Sanger sequencing trace representing the nucleotide substitution in *CEP68*. C. Schematic CEP68 protein showing the location of the amino acid change. D. Multiple CDKL1 protein amino acid alignment in the position 458.

6.2.1.3 Family FML003

MDS/AML families FML003 (Figure 6.7) was analysed assuming autosomal recessive as pattern of inheritance. Here, we describe the AR analysis of WES data of one member (II:6, Figure 6.7) of family FML003. It was a consanguineous marriage in which both parents were asymptomatic and five out of six children were affected. The oldest child had MDS (II:1, Figure 6.7), another three died with AA at age 10 years old (II:3, II:4 and II:5, Figure 6.7) and, the proband (II:6, Figure 6.7) presented with monosomy 7, MDS and died at age 14 years old after an autologous bone marrow transplant.

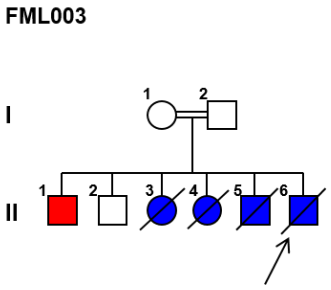


Figure 6.7. MDS/AML family tree of FML003. Affected individuals are coloured as following, red: MDS and blue: AA. The arrow highlights index case.

WES of the proband (II:6, Figure 6.7) revealed 769 non-silent variants, in which 281 were novel when compared to ESP and 1000G databases and 33 loss of function variants. 23 biallelic variants were found in FML003 with MAF≤0.001, 16 were homozygous and 7 were compound heterozygous. From these, five genes harbouring rare homozygous missense variants were identified and four that carried predicted damaging variants were selected to be the candidate genes in this family as they were not associated with any autosomal recessive disease (Table 6.9 and Table 6.10).

Segregation analysis in these five candidate genes were carried out as DNA samples from both parents were available (Figure 6.8) in an attempt to exclude any of the identified candidate genes as disease causing. Nonetheless, the variants found in the selected candidate genes were also present in a heterozygous estate in both parents making it a challenge to identify the best disease causing gene in this family.

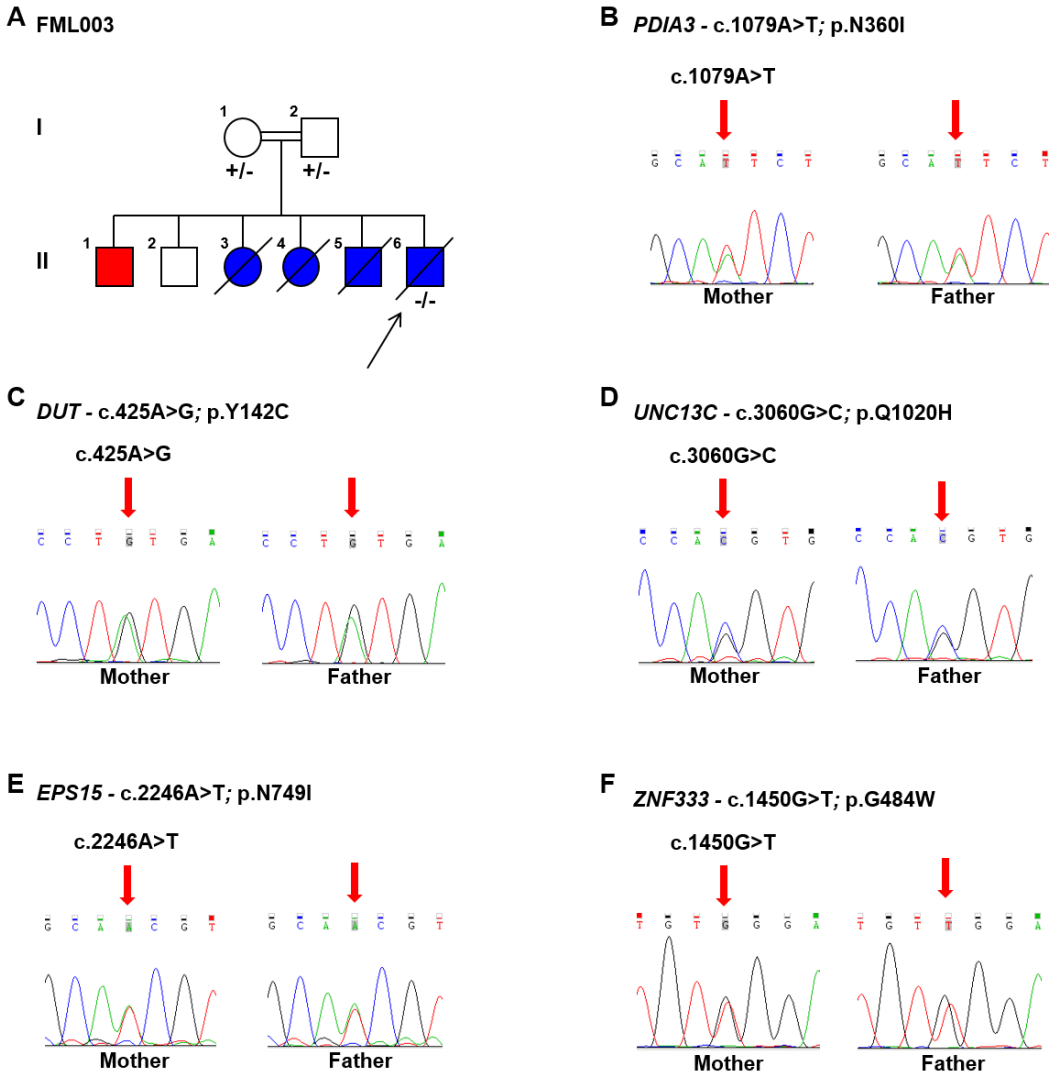


Figure 6.8. Segregation analysis of candidate genes in FML003. A. MDS/AML family tree of FML003. Affected individuals are coloured as following, red: MDS and blue: AA. The arrow highlights index case. B-F. Sanger sequencing showing the amino acid changes present in both parents of the index case.

Table 6.9. Rare homozygous variants with MAF≤0.001 in FML003

Autosomal recessive analysis in FML003						
	Gene	**Description**	**Variant**	**Exac (MAF)**	**PolyPhen2**	**MutationTaster2**
1	DUT	deoxyuridine triphosphatase	c.A425G:p.Y142C	0.000066	P	D
2	EPS15	epidermal growth factor receptor pathway substrate 15	c.A2246T:p.N749I	0.000033	B	B
3	PDIA3	protein disulfide isomerase family A, member 3	c.A1079T:p.N360I	0.000033	P	D
4	UNC13C	unc-13 homolog C (C. elegans)	c.G3060C:p.Q1020H	0.000034	D	D
5	ZNF333	zinc finger protein 333	c.G1450T:p.G484W	0.00026	D	D

MUC4 and AHNAK2 were removed. MAF: minor allele frequency; P: possibly damaging; B: benign; D: damaging.

Table 6.10. Candidate genes harbouring rare homozygous variants (MAF≤0.001) identified in FML003

	Gene	Additional information on the gene	Disease associated
1	*DUT* - Deoxiuridine triphosphatase	This gene encodes an enzyme of nucleotide metabolism. DUT hydrolyzes dUTP to dUMP and pyrophosphate. dUMP is used for the synthesis of thymine nucleotides needed for DNA replication. In addition, DUT also limits intracellular concentration of dUTP, which elevated levels lead to increased incorporation of uracil into DNA. Uracil induces extensive excision repair mediated by uracil glycosylase. This repair process, resulting in the removal and reincorporation of dUTP, is self-defeating and leads to DNA fragmentation and cell death. The more abundant form (DUT-N) localizes to nuclei while the higher molecular weight form (DUT-M) is associated with mitochondria (Mol *et al.*, 1996). ENST00000331200: NM_001025248.1 → NP_001020419.1, 252aa	None
2	*PDIA3* - Protein disulfide isomerase family A member 3	This gene encodes a protein of the endoplasmic reticulum that interacts with lectin chaperones calreticulin and calnexin to modulate folding of newly synthesized glycoproteins as it catalyzes the rearrangement of -S-S- bonds in proteins. PDIA3 has protein disulfide isomerase activity and it is thought that complexes of lectins and PDIA3 mediate protein folding by promoting formation of disulfide bonds in their glycoprotein substrates (Bourdi *et al.*, 1995). ENST00000300289: NM_005313.4 → NP_005304.3, 505aa	None
3	*UNC13C* - Unc-13 homolog C	UNC13 proteins, such as UNC13C share homology with *C. elegans* Unc13 and are predominantly expressed in brain. UNC13 proteins contain a phorbol ester-binding C1 domain and 2 C2 domains (Augustin *et al.*, 2001). UNC13C might play a role in vesicle maturation during exocytosis as a target of the diacylglycerol second messenger pathway and it might be involved in the regulation of synaptic transmission at parallel fibre - Purkinje cell synapses (By similarity). ENST00000260323: NM_001080534.2 → NP_001074003.1, 2214aa	None
4	ZNF333 - Zinc finger protein 333	Jing *et al.* (2004) showed that ZNF333 repressed transcription of promoters containing the ATAAT sequence in a luciferase reporter assay. It may be involved in transcriptional regulation. ENST00000292530: NM_032433.3 → NP_115809.1, 665aa	None

6.2.1.4 All MDS/AML families – gene-level and variant-level metrics combined to assess potential pathogenicity of a variant

From WES of 51 unrelated cases in 42 uncharacterized MDS/AML families, approximately 70,000 somewhat rare (MAF≤0.05) variants were found by comparison with the current reference haploid human genome sequence (human genome 19) and, 26,435 of these were non-silent variants (Figure 6.9). Further filtering resulted in 19,801 variants that segregated within the families (when samples from more than one member were analysed, which was undertaken in 12 families). Therefore, we have analysed 19,801 possibly germline somewhat rare considering autosomal dominant as pattern of inheritance. A further filtering using ExAC retained 4,676 possibly germline variants with MAF≤0.0001 in a total of 1,976 genes (Table 6.11 and Figure 6.9).

Table 6.11. Genes with variants MAF≤0.0001 identified in multiple MDS/AML families

Number of families	Altered genes		
12	*FLG*	521 genes	1976 genes
11	*RYR1*		
10	*DNAH14, MUC16, AHNAK2*		
9	*MACF1, MUC12, OBSCN*		
7	*LRP2, MYH13*		
6	8 genes		
5	16 genes		
4	83 genes		
3	404 genes		
2	1,455 genes		

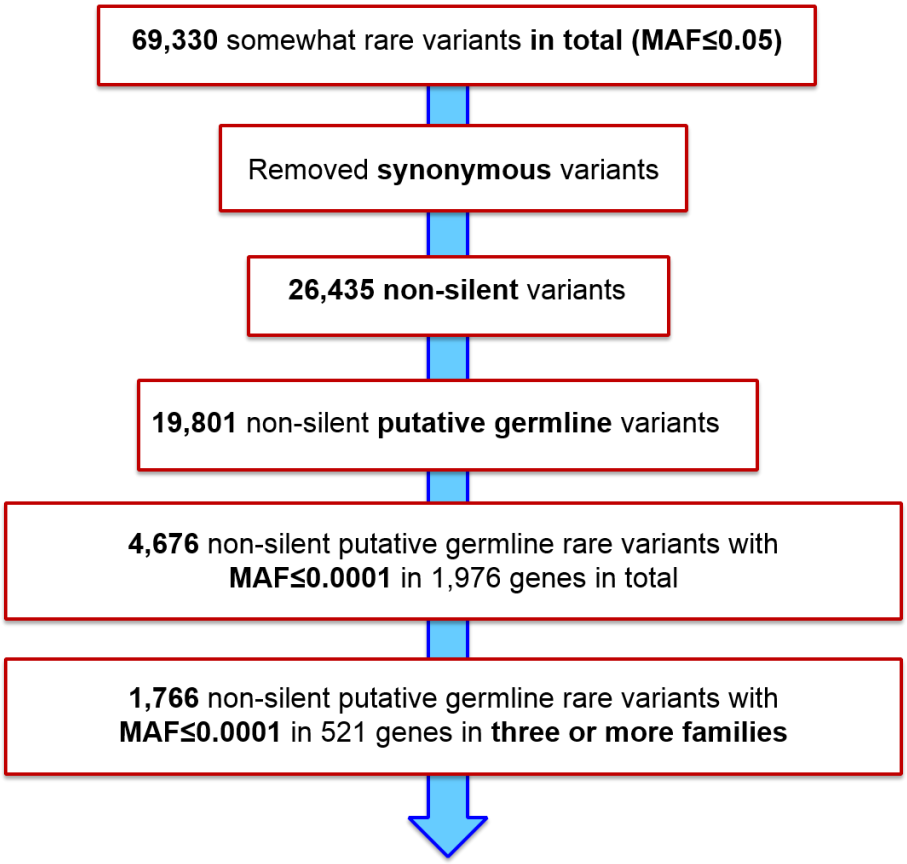


Figure 6.9. Workflow of the autosomal dominant analysis of WES data of 42 uncharacterised MDS/AML families.

Analysing manually all 4,676 rare variants on basis of their predictability effect to the protein using Polyphen2 and MutationTaster2 tools would be a laborious task. Instead, we worked alongside with a new pipeline based on the assumption that some genes in the human genome are sensitive to alterations and they would be the most likely to contribute to disease when mutated. We used two metrics including missense Z and loss of function pLI (probability of LoF intolerance) scores according to Samocha *et al*., 2014 and Lek *et al*., 2016 to evaluate the proportion of common functional variation in each gene, thereby identifying genes that appeared to be intolerant of variants. The knowledge that damaging variants are expected to have

lower allele frequencies than neutral ones and the deep discovery of rare variation in ExAC allows inference of the degree of selection against specific functional classes of variation (for instance, missense or LoF) on a gene basis. It is done by examining the proportion of variation that is missing compared to expectations under random variants (Samocha *et al*., 2014; Lek *et al*., 2016).

Gene-level measures of constraint such as missense Z and loss of function pLI scores offer additional information to variant-level metrics (such as Polyphen 2 and MutationTaster2) in assessing potential pathogenicity. As rare variants found in genes that are altered in multiple unrelated families with similar phenotypes are more likely to be pathogenic, we prioritised the analysis on rare variants in genes that were altered in 3 or more families. Hence, we further filtered 1,766 rare variants MAF≤0.0001 identified in 521 genes by selecting genes with missense Z score≥2.9 or pLI≥0.9 present in 3 or more families. We have set the threshold missense Z score≥2.9 as Samocha *et al*. (2014) found that missense Z score≥3.09 corresponded to excessively constrained genes in their data analyses. This anlaysis resulted in 83 genes with Z≥2.9 or pLI≥0.9 harbouring rare germline heterozygous variants (MAF≤0.0001). From these, 30 genes were removed as they were observed to be associated with other autosomal dominant diseases. This left us with 179 rare germline heterozygous variants in 53 candidate genes. A further filtering based on variants with MAF≤0.00001 and, genes with Z≥2.9 and pLI≥0.9 identified approximately 80 germline heterozygous rare variants in 46 genes present in 3 or more unrelated families. As result, eight selected candidate disease genes presented damaging rare variants predicted by the Polyphen 2 and MutationTaster2 tools in 3 or 4 families (Table 6.12). The analysis workflow is described in Figure 6.10. Description of these candidate disease genes can be seen in Table 6.13.

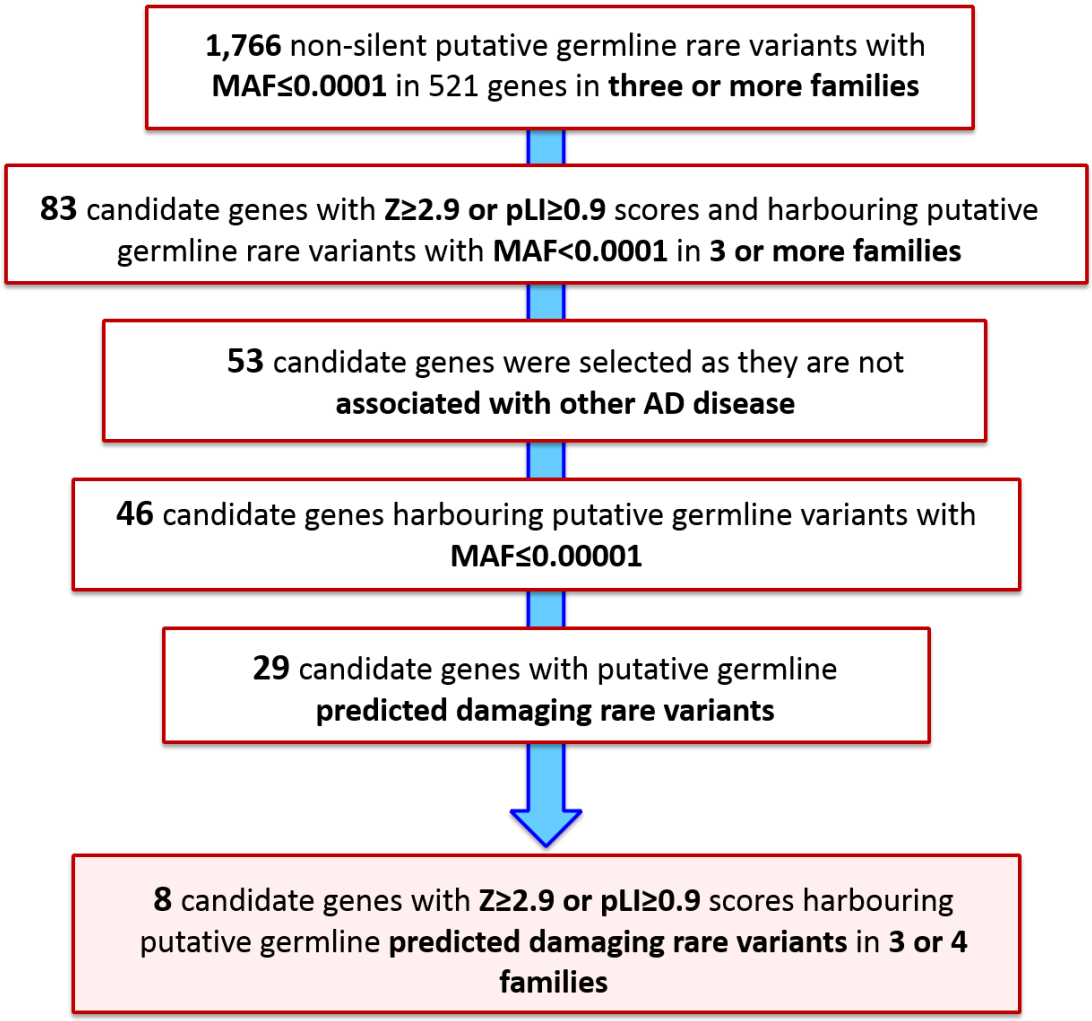


Figure 6.10. Workflow of the AD analysis of WES data of 42 uncharacterised MDS/AML families using gene-level and variant-level metrics combined to assess potential pathogenicity of a variant and select candidate genes.

Table 6.12. Candidate genes based on possibly damaging rare heterozygous variants with MAF≤0.00001, Z≤2.9 and/or pLI≥0.9 scores in 3 or more families

Autosomal dominant analysis - MDS/AML families						
	Samples	**Gene**	**AAChange**	**ExAC (MAF)**	**z**	**pLI**
1	FML116	ARNTL	ENST00000389707:exon15:c.C1246T:p.R416W	1 in 116060 (0.000008616)	3.08	0.99
	FML113	ARNTL	ENST00000389707:exon19:c.A1630G:p.T544A	1 in 120858 (0.000008274)	3.08	0.99
	FML018	ARNTL **(3fam)**	ENST00000389707:exon16:c.C1348A:p.P450T	1 in 121074 (0.000008259)	3.08	0.99
2	FML019	BRD1	ENST00000216267:exon6:c.C2342T:p.P781L	2 in 118952 (0.00001681)	4.29	0.99
	FML116	BRD1	ENST00000216267:exon6:c.A2127C:p.R709S	NR	4.29	0.99
	FML009	BRD1 **(3fam)**	ENST00000216267:exon12:c.2839-4G≥A	NR	4.29	0.99
3	FML045	KIF13A	ENST00000259711:exon39:c.G4900A:p.E1634K	2 in 120346 (0.00001662)	0.91	1
	FML033	KIF13A	ENST00000259711:exon4:c.A203G:p.N68S	NR	0.91	1
	FML112	KIF13A **(3fam)**	ENST00000259711:exon16:c.1764_1766delATG:p.588_589del	NR	0.91	1
4	FML002	NCOR2	ENST00000405201:exon23:c.3179_3180insC:p.R1060fs	NR	2.09	1
	FML018	NCOR2	ENST00000405201:exon14:c.1523_1525delGCT:p.508_509del	NR	2.09	1
	FML017	NCOR2	ENST00000405201:exon9:c.C971T:p.A324V	1 in 120610 (0.000008291)	2.09	1
	FML089	NCOR2 **(4fam)**	ENST00000405201:exon20:c.G2364C:p.E788D	1 in 83880 (0.00001192)	2.09	1
5	FML047	PTPN4	ENST00000263708:exon23:c.T2212C:p.W738R	NR	2.16	1
	FML060	PTPN4	ENST00000263708:exon14:c.C1093T:p.R365W	NR	2.16	1
	FML116	PTPN4 **(3fam)**	ENST00000263708:exon21:c.C1984A:p.L662M	NR	2.16	1
6	FML002	SCAF1	ENST00000360565:exon7:c.1727_1728insCCGCTCCCGCTC:p.S576delinsSRSRS	NR	4.32	1
	FML056	SCAF1	ENST00000360565:exon7:c.A683G:p.Y228C	NR	4.32	1
	FML111	SCAF1 **(3fam)**	ENST00000360565:exon7:c.C2485T:p.R829W	2 in 118840 (0.00001683)	4.32	1
7	FML054	TNRC6C	ENST00000335749:exon13:c.C3646T:p.P1216S	NR	1.34	1
	FML109	TNRC6C	ENST00000335749:exon12:c.G3469T:p.A1157S	NR	1.34	1
	FML018	TNRC6C **(3fam)**	ENST00000335749:exon3:c.A1475C:p.K492T	NR	1.34	1
8	FML075	ZZEF1	ENST00000381638:exon43:c.C6944G:p.S2315C	2 in 121230 (0.00001650)	0.55	1
	FML082	ZZEF1	ENST00000381638:exon7:c.C1346T:p.S449F	NR	0.55	1
	FML083	ZZEF1 **(3fam)**	ENST00000381638:exon24:c.G3598T:p.V1200L	NR	0.55	1

Z: Zscore; pLI: probability of loss of function score; MAF: minor allele frequency; 3fam: three families; NR: not registered.

Table 6.13. Description of candidate disease genes identified by using a combination of gene-level and variant-level metrics of constraint

Autosomal dominant analysis – MDS/AML families				
	Gene	**Description**	**Disease associated**	**Inheritance**
1	ARNTL - aryl hydrocarbon receptor nuclear translocator like 625aa; 23 exons z:3.08; pLI:0.99 ENST00000389707: NM_001178.5 → NP_001169.3	Transcriptional activator which forms a core component of the circadian clock. The circadian clock, an internal time-keeping system, regulates various physiological processes through the generation of approximately 24 hour circadian rhythms in gene expression, which are translated into rhythms in metabolism and behaviour. Disruptions in the circadian rhythms contribute to the pathology of cardiovascular diseases, cancer, metabolic syndromes and aging. Alterations in this gene have been linked to infertility, problems with gluconeogenesis and lipogenesis, and altered sleep patterns (Wang *et al.*, 2013).	None	NA
2	BRD1 - bromodomain containing 1 1189 aa; 20 exons z:4.29; pLI:0.99 ENST00000216267: NM_001304808.2 → NP_001291737.1	This gene encodes a bromodomain-containing protein that localizes to the nucleus and can interact with DNA and histone tails. The encoded protein is a component of the MOZ/MORF acetyltransferase complex and can stimulate acetylation of histones H3 and H4, thereby potentially playing a role in gene activation. It is related to the AF10 leukaemia gene (involved with translocations) through the presence of a cysteine-rich region and a leucine zipper. McCullagh *et al.* (1999) suggested that the identification of other proteins with these structures might aid understanding of their role in normal and leukemic cells.	Schizophrenia and bipolar disorder (from population studies: SNP rs138880 in the promoter region of BRD1, C is the risk allele).	NA
3	KIF13A - kinesin family member 13A 1805 aa; 46 exons z:0.91; pLI:1 ENST00000259711: NM_022113.5 → NP_071396.4	Plus end-directed microtubule-dependent motor protein involved in intracellular transport and regulating various processes such as mannose-6-phosphate receptor (M6PR) transport to the plasma membrane, endosomal sorting during melanosome biogenesis and cytokinesis. Mediates the transport of M6PR-containing vesicles from trans-Golgi network to the plasma membrane via direct interaction with the AP-1 complex. During melanosome maturation, required for delivering melanogenic enzymes from recycling endosomes to nascent melanosomes by creating peripheral recycling endosomal subdomains in melanocytes. Also required for the abscission step in cytokinesis: mediates translocation of ZFYVE26, and possibly TTC19, to the mid body during cytokinesis (Sagona *et al.*, 2010).	None	NA

Table 6.13. Continued

4	NCOR2 - nuclear receptor corepressor 2 2514 aa; 49 exons z:2.09; pLI:1 ENST00000405201: NM_006312.5 → NP_006303.4	This gene encodes a nuclear receptor co-repressor that mediates transcriptional silencing of certain target genes. The encoded protein is a member of a family of thyroid hormone- and retinoic acid receptor-associated co-repressors. This protein acts as part of a multisubunit complex which includes histone deacetylases to modify chromatin structure that prevents basal transcriptional activity of target genes. Aberrant expression of this gene is associated with certain cancers (Mendez *et al*., 2008).	None	NA
5	PTPN4 - protein tyrosine phosphatase, non-receptor type 4 926 aa; 29 exons z:2.16; pLI:1 ENST00000263708: NM_002830.3 → NP_002821.1	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signalling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. PTPN4 prevents cell death (Caillet-Saguy *et al*., 2017).	None	NA
6	SCAF1 - SR-related CTD-associated factor 1 1312 aa; 13 exons z:4.32; pLI:1 ENST00000360565: NM_021228.2 → NP_067051.2	SCAF1, interact with the C-terminal domain (CTD) of the large subunit of RNA polymerase II and participate in pre-mRNA splicing (Scorilas *et al.*, 2001).	None	NA

Table 6.13. Continued

7	TNRC6C - trinucleotide repeat containing 6C 1726 aa; 24 exons z:1.34; pLI:1 ENST00000335749: NM_001142640.1 → NP_001136112.1	Plays a role in RNA-mediated gene silencing by micro-RNAs (miRNAs). Required for miRNA-dependent translational repression of complementary mRNAs by argonaute family proteins. As scaffolding protein associates with argonaute proteins bound to partially complementary mRNAs and simultaneously can recruit CCR4-NOT and PAN deadenylase complexes (Fabian *et al.*, 2011).	None	NA
8	ZZEF1 - zinc finger ZZ-type and EF-hand domain containing 1 2961 aa; 55 exons z:0.55; pLI:1 ENST00000381638: NM_015113.3 → NP_055928.3	No description of its function on NCBI nor Uniprot. No OMIM.	None	NA

NA: not applicable. Information associated with leukaemia is highlighted in red.

6.3 Discussion

Searching for the underlying disease gene in families with MDS/AML is a massive challenge due to the vast clinical and genetic heterogeneity of this rare disorder. We have undertaken WES in 42 uncharacterised families with MDS/AML and have not identified the disease causing gene in half of them. Generally in this study, the selection of candidate genes were performed by selecting rare predicted damaging variants in genes not previously associated with other disease and that occurred in multiple families. This generated a number of candidate genes that are listed fully in Tables 6.12, 6.4, 6.7, and 6.9 in chapter 6. However, with the data to hand it is not possible to say, which if any, will turn out to be definitive disease causing genes.

One of the initial issues when analysing sequencing data is defining the threshold at which a variant should be considered rare or a polymorphism. Traditional methods for setting allele frequency thresholds for variant classification are based on the expected incidence of disease. However, accurate incidence and penetrance numbers of very rare diseases such as familial MDS/AML are not available (Shearer *et al*., 2014; Kobayashi *et al*., 2017). The frequency of a variant in the general population is crucial as the rarity of a variant is a prerequisite for pathogenicity. However, population allele frequency is a potential empirical data for improving variant interpretation as an allele frequency greater than expected for disorder is strong evidence for a benign classification. Without careful consideration, aggressive allele frequency thresholds may increase the risk of incorrectly classifying pathogenic variants with elevated allele frequencies as benign. If done carefully, a very low allele frequency threshold (including novel, MAF≤0.0001 or MAF≤0.00001) can be adopted to more accurately interpret sequence variants (Richards *et al*., 2015).

The approach of using a small cohort of presumed healthy individuals to distinguish benign polymorphisms from potentially pathogenic variants, as any variants observed in unaffected individuals are unlikely to be disease causing, is effective for early-onset dominant disorders with high penetrance. It is essential to consider that the likelihood of the presence of disease causing variants in population databases such as ExAC and ESP is high as they include non-healthy individuals. This likelihood increases even further when analysing recessive disorders or disorders with low-penetrance or late-onset in which unaffected carriers are expected to be present. However, the size of the ExAC population database is a comprehensive representation of very rare variants that allows for more accurate minor allele frequency calculations. Furthermore, scientific literature can also be used in agreement with population data for variant classification as rare pathogenic variants are frequently published. Overall, this is a scalable approach that allows for rapid adoption of new datasets and refinements to MAF thresholds as larger and higher-quality datasets are published (Walsh *et al*., 2016).

On the other hand, variant classification is a complex process involving the evaluation and interpretation of multiple pieces of evidence, which in turn requires considerable knowledge and expertise. A variant's absence or presence in ExAC at very low frequency is clearly not sufficient to indicate that the variant is pathogenic. Many variants are private, novel, or rare, and the vast majority of these are also not pathogenic (Richards *et al*., 2015). However, finding rare predicted damaging variant in the same gene in multiple families with similar phenotype and where alterations in the gene in question is not linked to other disease are some strong lines of evidence to identify a good candidate disease causing gene.

Furthermore, missense variants are a common finding in our cohort and the assessment of their protein functional impact is demanding without adequate functional studies. Therefore, the use of an intolerance ranking system can facilitate identification of high impact variants through the gene in which they occur. Gene-level score can be integrated with well-established variant-level scores to highlight candidate casual variants (Petrovski *et al*., 2013).

Genes intolerant to genetic variation in the human population are more likely to cause some disorders. The most highly constrained missense (top 25% missense Z score) and LoF (pLI≥0.9) genes show higher expression levels and broader tissue expression than the least constrained genes and are involved in core biological processes (spliceosome, ribosome, and proteasome components). Genes intolerant of LoF variation would be expected to be dosage-sensitive, as in such genes natural selection does not tolerate a 50% deficit in expression due to the loss of single allele. Therefore, smaller changes in the expression of these genes are more likely to contribute to medically relevant phenotypes. Although, this extreme constraint does not necessarily reflect a lethal disease or status as a disease gene, but probably points to genes in which heterozygous LoF confers some non-trivial survival or reproductive disadvantage (Table 6.14). Disease genes that act after post-reproductive age – do not necessarily have a high pLI values (e.g. *BRCA1*).

LoF intolerant genes include virtually all known severe haploinsufficient human disease genes and yet there are 3,230 genes in ExAC with near complete depletion of predicted protein-truncating variants, with 72% of these genes having no currently established human disease phenotype in the OMIM (https://www.omim.org/) or ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) databases (Lek *et al*., 2016).

Table 6.14. Z and pLI scores of genes associated with familial MDS/AML

Gene	Z score	pLI score
RUNX1	2.8	0.45
CEBPA	NA	NA
TERC	NA	NA
TERT	6.3	0.87
GATA2	2.88	0.98
SRP72	0.45	0.01
ANKRD26	-2.05	0
ETV6	2.2	1
DDX41	2.97	0

NA: not available.

Finally, it might be necessary to use different approaches combined with WES such as other types of next generation sequencing methods that can identify variants in non-coding regions, large insertions/deletions and copy numbers variants to search for the underlying gene disease in the uncharacterised MDS/AML families. This combined use of various techniques would increase genome coverage and decrease the chances of missing a causal genetic alteration. Additionally, new WES data from further registered MDS/AML families will supplement this dataset and increase germline allelic series and therefore it may facilitate disease gene identification.

Chapter 7

Discussion

7.1 Variants identified in known disease genes associated with familial MDS/AML and related disorders

In chapter 3, seven MDS/AML families that were genetically characterised during this study have been discussed. Interestingly, in addition to the identification of known disease causing genes associated with familial MDS/AML, three MDS/AML families harboured variants in genes that are associated with other bone marrow failure syndromes including *FANCA*, *SBDS* and *WAS*. This highlights the marked clinical and genetic heterogeneity of these disorders. Bone marrow failure syndromes are a heterogeneous group of diseases that includes familial MDS/AML, DC, HH, FA, Diamond–Blackfan anaemia, Shwachman–Diamond syndrome among others. They present with cytopenias in at least one haematopoietic cell lineage that may develop into pancytopenia and with an increased risk of haematological or solid cancers (Alter, 2017). Although the classical expression of these syndromes is very distinct, it is well known that they have overlapping clinical features. Furthermore, phenotypic variability even among patients sharing the same variant in a disease gene is recognised in these group of disorders (Dokal *et al*., 2015; Bertuch, 2016). The atypical cases of bone marrow failure syndromes may initially present with isolated MDS or AML and, the classic features may manifest over time or alternatively they may present with reminiscent characteristics but not entirely coherent with a known inherited disorder as observed in our families FML004, FML032 and FML041. Indeed, Rochowski *et al*. (2012) reported that approximately 1% of patients recovering from leukaemia after chemotherapy had undiagnosed FA. As also exemplified by our results, genomic evaluation using next generation sequencing is efficient to identify the pathogenic variant in genes known to cause the disease and to define a complex diagnosis (Zhang *et al*., 2015; Ghemlas *et al*., 2015). Use of next generation sequencing in these atypical patients and in healthy individuals will help further understanding the biology

of bone marrow failure diseases. An additional important consideration when working up a challenging diagnosis is to be aware of revertant somatic mosaicism. This refers to a rare genetic event when a pathogenic disease variant spontaneously reverts into a benign variant. Revertant mosaicism is likely to have taken place in cases where children are affected, and the asymptomatic parents do not harbour the pathogenic variant. This event has been observed in FA and in patients with DC harbouring *TERT* variants (Soulier *et al*., 2005; Jongmas *et al*., 2012).

The specific pathway disrupted by a variant in a known disease causing gene also contributes to the progress and specific disease outcome. Each of these diseases causing genes are affected by their unique variant types (missense, LoF, indels or nonsense) and location (throughout the protein or in specific regions) and, by the mechanisms that causes the disease such as haploinsufficiency or dominant-negative effects, although, autosomal dominant pattern of inheritance and haploinsufficiency seem to be a general characteristic of familial MDS/AML. Germline heterozygous variants, including nonsense, missense, deletion, frameshift throughout the gene and gene deletion or duplication in *RUNX1* cause familial platelet disorder with propensity to myeloid malignancies (MDS, AML and T-cell ALL). *GATA2* variants are typically loss of function with no genotype-phenotype correlations. Impaired telomerase activity (due to germline variants in *TERC* and *TERT*) and extremely short telomeres induce chromosomal instability, causing bone marrow failure, fibrosis of the lungs and liver, and tumour formation. Somatic variants in the promoter region of *TERT* have been reported in melanomas, gliomas, liposarcomas, urothelial carcinomas and hepatocellular carcinomas. Biallelic germline variants in *TERT* can lead to telomere biology disorders including DC and HH and *TERT* heterozygous variants cause familial MDS/AML.

In our study, rare variants were classified as likely pathogenic, VUS and likely benign according to the joint consensus recommendation for the interpretation of sequence variants by the American College of Medical Genetics and Genomics (ACMG) and the association for Molecular Pathology (AMP). Likely to be pathogenic and VUS were further analysed in order to assign a variant as causal. This involved strict criteria for determining pathogenic variants: a) families that carried proven variants in known disease causing genes were expected to have similar disease presentation in accordance with the literature, b) recurrent variants were favourable to assign pathogenicity to a variant, c) protein impact was assessed for all proven variants using prediction software as a guide, d) family segregation studies were carried out when possible and, e) investigation of all likely pathogenic variants were performed in the literature and databases.

The index case in FML007 who harbours a large *RUNX1* deletion (described in section 3.2.1.1 in chapter 3) highlights that interpretation of genetic testing in blood can be confounding by the presence of somatic variants found in known disease causing genes. Despite the recommendation of use of a non-blood tissue to validate a germline variant, we do not have it available for all our families. However, segregation analysis within families can be useful in distinguishing inherited from somatic variants. Furthermore, it is also possible to identify if a variant is germline or somatic through Sanger sequencing trace analysis when the variant is a heterozygous indel. This is carried out by calculating an average of the ratio of the chromatogram peak heights of the variant and the sum of the chromatogram peak heights of wild type plus variant, the variant in question is likely germline when the results are near 50%. For obvious reasons, this method could not be performed on missense nor nonsense variants.

In conclusion, the observations in these known disease genes serve to highlight the complexity in assigning pathogenic status to a newly identified variant. They also show the value of next generation sequencing in picking up pathogenic variants in cases which have an atypical clinical presentation.

7.2 Germline heterozygous LoF *DDX41* variants in a subset of familial myelodysplasia and acute myeloid leukaemia

In chapter 4 we have described four families with MDS/AML from our cohort harbouring heterozygous LoF variants in *DDX41*, a relatively new gene associated with familial MDS/AML. Variants in *DDX41* were reported in various families with MDS/AML and it is characterised by late onset, advanced state of the disease at diagnosis, normal karyotype and poor overall survival. This makes *DDX41* different from the other known familial MDS/AML disease genes identified to date along with its unclear function in leukaemogenesis. *DDX41* is shown to be involved in spliceosomes defects when altered in patients with familial MDS/AML, however the exact mechanism leading to disease is not yet established. Nonetheless, *DDX41* is known to participate in innate immunity response by acting as an intracellular DNA sensor and direct interactions with bacteria pathogenic particles stimulating the immune response. Interestingly, Shwartz *et al.* (2017) reported mutations in *SAMD9/SAMD9L* to be the cause of familial MDS/AML in children. And these genes present some similarities with *DDX41* such as a) monosomy of chromosome 7 causes haploinsufficiency in familial MDS/AML indicating a loss of tumor suppressor role as it happens with monosomy of chromosome 5, b) there are cases in paediatric familial MDS/AML patients with germline mutations in *SAMD9/SAMD9L* harbouring somatic mutations in the other allele, and c) *SAMD9L* is also involved in immune response (Nagamachi *et al.*, 2013).

There is data suggesting association of *DDX41* variants with post-transcriptional regulation of protein translation in cell growth in Peters *et al*. (2017) studies. This group suggested an oncogenic role for *DDX41*. In addition, *DDX41* deletions and frameshift variants as well as its reduced expression demonstrate a tumour

suppressor function in MDS/AML reported by Polprasert *et al*. (2015). Similarly, variants in other members of the DEAD/H-box RNA helicase family can act as tumour suppressors and oncogenes in cancer (Fuller-Pace, 2013). In addition, Polprasert *et al*. (2015) demonstrated the interaction of *DDX41* with spliceosome proteins and that altered *DDX41* impaired this interaction. Spliceosome defects caused by mutated genes can result in alterations in the balance of isoforms and in the inactivation of tumour suppressor genes, promoting cell proliferation. Further studies are needed in order to clarify the precise role of *DDX41* variants in haematopoiesis and leukaemia.

Patients harbouring germline variants in *DDX41* are phenotypically very similar to sporadic AML cases. In many cases the disease develops with an acquisition of a second *DDX41* somatic variant in the other allele, however total inactivation of *DDX41* appears to be cell-lethal. *DDX41* is located on chromosome 5 and is deleted in some sporadic cases of MDS with del (5q) resulting in haploinsufficient expression. However, frameshift germline variants have not been found to be accompanied by deletions of the DDX41 locus on 5q (Maciejewski *et al*., 2017).

Virtually all families with MDS/AML carrying *DDX41* germline LoF variants reported in our study (chapter 4) had a late onset of the disease, except for family FML037 (Figure 4.5, chapter 4) in which the index case was diagnosed aged 41 years. In addition, the recurrent *DDX41* variant (c.3G>A, p.M1I) found in family FML037 is seen in ExAC with frequency of 1 in 1,189 people, raises a query about the pathogenicity of *DDX41* germline LoF variants. Deleterious variants are seen in *DDX41* with a high frequency in the population compared to variants in other known familial MDS/AML disease genes. The frequency of a variant in the general population is important as rarity of a variant is essential for pathogenicity. According to the ACMG and AMP guidelines, an “allele frequency greater than expected for disorder” is strong evidence for a benign

classification (Richards *et al*., 2015). In that context, *DDX41* germline variants can easily be mistaken and excluded as polymorphisms or interpreted falsely as pathogenic from some applied analysis strategies in many studies. Thus, it is necessary to carefully use a combination of evidence that rigorously assess the pathogenicity of *DDX41* variants.

Finally, the late onset of familial MDS/AML characteristic of *DDX41* variants makes it harder to obtain an accurate diagnosis for this group of patients and inappropriate selection of asymptomatic family donors for allogenic stem cell transplantation might take place. Clinically, recognition of *DDX41* mutated cases may have implications for surveillance, assessment of prognosis, and, perhaps, for design of targeted therapies. Further functional studies are necessary to confirm the pathogenicity of the variants in these families.

7.3 *RTEL1* LoF variants leading to myelodysplasia and liver disease

RTEL1 variants in our cohort of patients with bone marrow failure disorders (DC, HH, AA and, familial MDS/AML) are very interesting starting with the amount of variants found in this gene. We have found 29 *RTEL1* variants in 35 patients. The next step was to interpret these variants in order to identify the disease causing variant by filtering, annotating the variants (population variant allele frequencies in the canonical transcript) and using *in silico* variant function impact prediction. Most of these *RTEL1* variants were classified as heterozygous missense likely benign (9 variants in 14 patients), many were heterozygous missense of unknown significance (12), few were biallelic likely pathogenic (4) bar one was biallelic likely benign variant and, some were identified as LoF variants likely pathogenic (three in four unrelated families) as described in chapter 5. In this section of our study, we were able to use the prediction tool Combined Annotation Depletion score (Kircher *et al*., 2014) as variant call format files were available on this data.

CADD is a method that measures deleterious variants systematically across the genome. It combines many different annotations including allelic diversity, annotations of functionality, pathogenicity, disease severity, experimentally measured regulatory effects and complex trait associations into the C score, a specific measure for each variant. This highly ranks known pathogenic variants within individual genomes (Kircher *et al*., 2014). However, Polyphen2 and MutationTaster2 were the prediction tools used in all other chapters as they are also well-established annotations important in the identification of disease causing genes (Kircher *et al*., 2014).

It is largely known that biallelic variants in *RTEL1* are associated with DC and HH and *RTEL1* heterozygous variants have been associated with idiopathic and familial cases of interstitial pulmonary fibrosis. In this study we have identified five families carrying *RTEL1* biallelic variants and four families carrying heterozygous *RTEL1* LoF variants. However, the clinical phenotypes of these families vary from what is already established in the literature and consequently it expands the disease spectrum of *RTEL1* (Cardoso *et al*., 2017). *RTEL1* biallelic variants were found in two families with DC, however *RTEL1* biallelic variants also were identified in family 5 (Figure 5.8 seen in chapter 5) and it is unlikely to be the disease causing in this family as the asymptomatic mother harbours the same biallelic variants. Furthermore, other two families harbouring *RTEL1* biallelic variants presented with AA in adulthood again expanding what is reported in the literature to date. Marsh *et al*. (2018) have just reported similar data whilst this thesis was being prepared. Finally, *RTEL1* heterozygous LoF variants were identified in four families with MDS and liver disease in adults, expanding the spectrum of RTEL1 associated diseases.

It is noteworthy that patients with complex diseases do not always present with a clear classical phenotype at the time of presentation. From the families described in chapter 5 that present with MDS and liver disease with variants in *RTEL1*, only family 8 is classified as familial MDS/AML (defined by us as being when there are two or more members with bone marrow failure in the family, where at least one of them presents with MDS or AML). This family prompted us to search for further *RTEL1* variants in independent cohorts from our lab (dyskeratosis congenita, idiopathic bone marrow failure).

RTEL1 is part of the telomere biology- associated genes (*DKC1*, *TERC*, *TERT*, *USB1*, *CTC1*, *NHP2*, *NOP10*, *WRAP53*, *TINF2*, *PARN* and *ACD*) that cause telomeropathies

where germline variants result in short telomeres and bone marrow, lung, liver, and skin can be affected. In agreement with the literature on patients carrying biallelic or LoF variants presented short telomeres compared to age-matched controls. Furthermore, RTEL1 defect is associated with production and accumulation of T-circles in cells and we have seen that one case carrying a LoF variant presented with increased T-circles when compared to a case carrying a missense variant. Further T-circle tests should be undertaken in a larger group of samples in order to establish a normal range of T-circle production in *RTEL1* patients. Furthermore, if enough samples of affected patients and resources were available, the following functional experiments could be performed to verify the biological impact of the germline heterozygous LoF *RTEL1* variants identified in this study: a) demonstrate the expected lower expression of RTEL1 in the patients with heterozygous germline LoF variants compared to a control by Western blotting; b) mass spectrometry to identify the proteins associated with RTEL1 in the patients and in a control along with RNA seq and gene set enrichment analysis to recognise the possible affected biological pathways; c) analyse all results and verify whether there is a plausible explanation for the MDS and liver disease phenotypes present in these patients.

In conclusion, the data from this section clearly extends the spectrum of RTEL1 related telomeropathies. It also shows that *RTEL1* can be added to the list of disease genes associated with familial MDS/AML.

7.4 Variants identified in familial MDS/AML candidate genes

We have been analysing WES data from MDS/AML families from the beginning of this project, working on different numbers of families, using adequate criteria for searching the genetic disease cause and generating different lists of candidate genes at each time. However, we have not identified the definite disease causing gene in a significant number of these uncharacterised MDS/AML families. Certainly, we first searched for predicted damaging variants in the known familial MDS/AML disease genes and only few of these families were characterised by predicted pathogenic variants in these known genes. This left us with many uncharacterised families with possibly variants in new disease causing genes. Finding these new disease causing genes is a very challenging task for many reasons including quality of the acquired information on the family history and the type and quality of collected sample(s). Consequently, new challenges arise from the chosen approaches as not only one technique would be enough to identifying these new causal genes. In addition, there are some technical problems that might occur from sample collection to preparation, storage and interpretation of data. Not to mention the technical challenges relating to genome coverage.

Considering these previously mentioned challenges, the first encountered issue was the selection of an adequate variant allele frequency that could be considered rare enough to cause familial MDS/AML. We have first chosen to analyse novel variants as the majority of published variants in the known familial MDS/AML disease genes are novel. No plausible candidate gene was identified using this criteria. We have then analysed predicted to be damaging variants with MAF≤0.0001 as damaging variants in *DDX41* were reported within this allele frequency. Again, no evident candidate gene was found. The second challenge in the analysis was the variant classification into

likely to be damaging and variant of unknown significance, mainly because missense variants are found in abundance in our WES data. Although there are many prediction tools to evaluate the pathogenicity of a variant, they are not enough to define its pathogenicity. Finally, the decision of a specific candidate gene being the possible underlying cause of the disease in a family is not straightforward even after the identification of a handful number of candidate genes that are not previously associated with other AD disease and that harbour predicted damaging rare germline variants which segregate with the disease in multiple families. This is because a) the results we obtained are not conclusive as about half of the families harbour variants in more than one identified candidate gene, b) the majority of the identified variants are missense making it harder to predict their pathogenicity and, c) lastly, it would help if any identified candidate gene was functionally linked to known familial MDS/AML disease genes to explain the disease development.

AD analysis of WES data from 30 MDS/AML families was performed based on predicted damaging rare germline variants that segregate with the disease in three or more families generated two candidate genes seen in Table 6.4 in section 6.2.1.1 in chapter 6 and Figure 7.1. In addition, AD analysis of WES data from 42 MDS/AML families was performed based on predicted damaging rare germline variants that segregate with the disease in three or more families generated eight candidate genes that presented $Z \leq 2.9$ or $pLI \geq 0.9$ seen in Table 6.13 in section 6.2.1.4 in chapter 6 and Figure 7.1. These analyses identified only one candidate gene in common, which is trinucleotide repeat containing 6C (*TNRC6C*). Hence, the first interpretation would be that *TNRC6C* might be one of the genes we were searching for. However, family FML018 has a consanguineous marriage and AR is likely to be the inheritance mode in this family even though there is a possibility of it being AD due to family history.

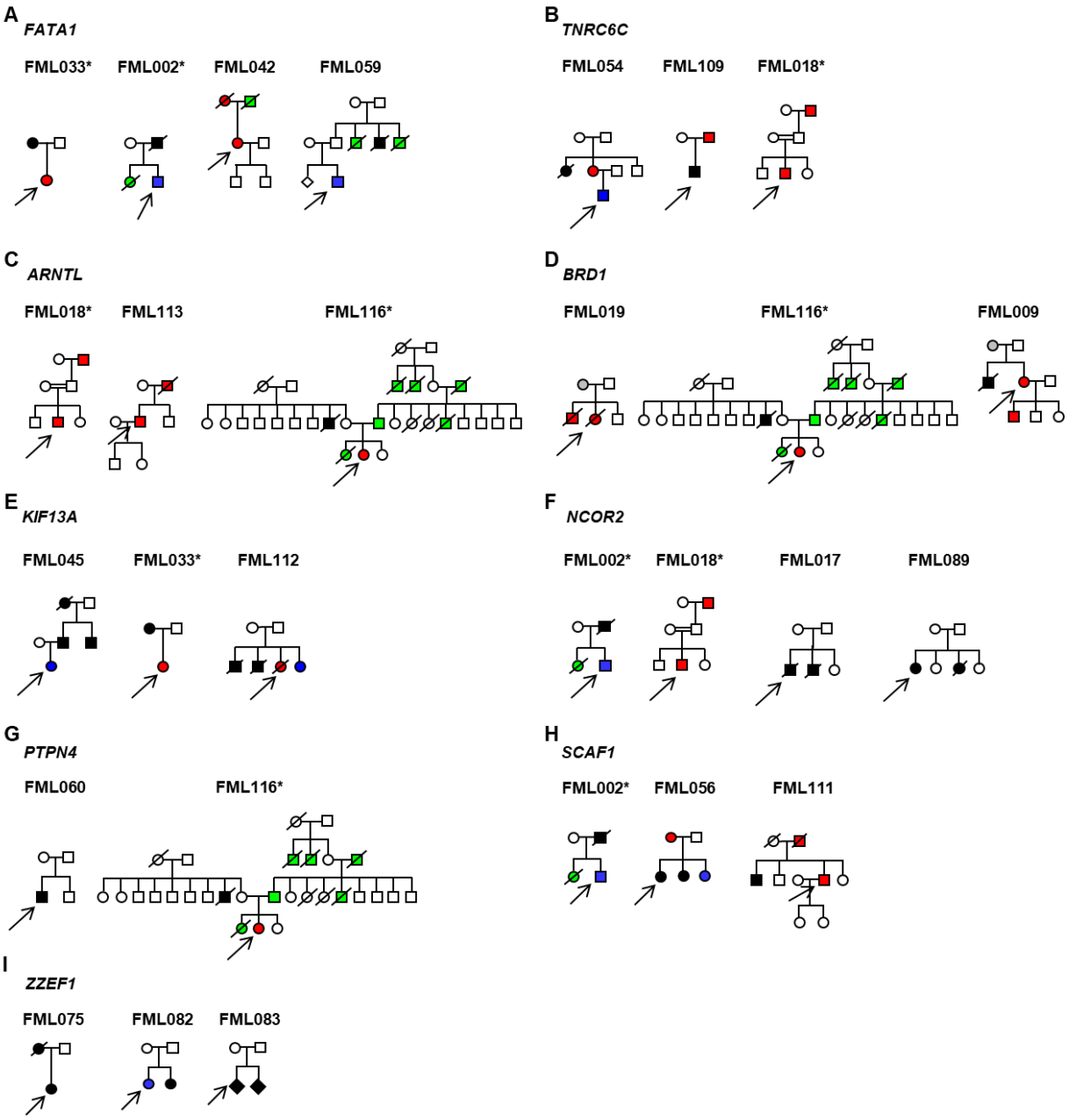


Figure 7.1. Uncharacterised families with MDS/AML that harbour predicted damaging heterozygous rare variants in the candidate genes seen in Table 7.1. Affected individuals are coloured as following black: AML, red: AML, blue: AA, grey: thrombocytopenia and green: solid cancer. The arrow highlights the index case. *: families that harbour variants in at least two of the candidate genes here presented. Family FML047 presents history of leukaemia, but there is no sufficient information for designing a family tree.

Segregation analysis is important to clarify the inheritance pattern in family FML018, however there are no DNA samples available from other members of this family. Furthermore, it was observed that families had variants in more than one candidate genes, which made it harder to distinguish which variant could be pathogenic (Table 7.1).

Table 7.1. Candidate genes selected from AD analysis of WES data from 30 and 42 MDS/AML families

1	FAT1	FML033* FML002* FML042 FML059	FAT1 is a receptor for a signalling pathway that regulates growth, gene expression, and cell polarity. FAT1 is up regulated in 11% of AML, 29% of preB acute lymphoblastic leukaemia (ALL) and 63% of T-cell acute lymphoblastic leukaemia (T-ALL). 4588aa
2	TNRC6C	FML054 FML109 FML018*	Plays a role in RNA-mediated gene silencing by micro-RNAs (miRNAs). 1726aa
3	ARNTL	FML116* FML113 FML018*	Transcriptional activator which forms a core component of the circadian clock. Alterations in this gene have been linked to infertility, problems with gluconeogenesis and lipogenesis, and altered sleep patterns. 625aa
4	BRD1	FML019 FML116* FML009	The encoded protein is a component of the MOZ/MORF acetyltransferase complex and can stimulate acetylation of histones H3 and H4, thereby potentially playing a role in gene activation. It is related to the AF10 leukaemia gene (involved with translocations). 1189aa
5	KIF13A	FML045 FML033* FML112	Involved in intracellular transport and regulating various processes such as mannose-6-phosphate receptor (M6PR) transport to the plasma membrane, endosomal sorting during melanosome biogenesis and cytokinesis. 1805aa
6	NCOR2	FML002* FML018* FML017 FML089	Mediates transcriptional silencing of certain target genes. This protein acts as part of a multisubunit complex which includes histone deacetylases to modify chromatin structure that prevents basal transcriptional activity of target genes. Aberrant expression of this gene is associated with certain cancers. 2514aa
7	PTPN4	FML047 FML060 FML116*	It is a tyrosine phosphatase (PTP) protein. PTPs are known to be signalling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. PTPN4 prevents cell death. 926aa
8	SCAF1	FML002* FML056 FML111	SCAF1, interact with the C-terminal domain (CTD) of the large subunit of RNA polymerase II and participate in pre-mRNA splicing. 1312aa
9	ZZEF1	FML075 FML082 FML083	No description of its function on NCBI nor Uniprot. No OMIM. 2961aa

*: families that harbour variants in at least two of the candidate genes here presented.
Information associated with leukaemia is highlighted in red.

Further analysis using The human protein atlas (https://www.proteinatlas.org/) allowed a refined selection of more suitable candidate genes from the list presented in Table 7.1. This showed that *FAT1*, *BRD1*, *PTPN4*, and *ZZEF1* are unlikely to be good candidate genes as they are not expressed in haematopoietic cells. In addition, *KIF13A* and *SCAF1* present low protein expression in haematopoietic cells. These results leave us with *ARNTL*, *NCOR2*, and *TNRC6C* as good candidate genes. However, one of the families that harbour germline damaging heterozygous variant in *ARNTL* and *NCOR2* is FML018, the same family that was discussed previously for also carrying a germline damaging heterozygous variant in *TNRC6C*. Hence, without FML018, there are two other families (FML113 and FML116) harbouring variants in *ARNTL*. Finally, *NCOR2* seems to be the reasonable candidate gene to have its pathogenicity further validated at this moment from results of this analysis as still there are three other families (FML002, FML017 and FML089 see Figure 7.2) that carry predicted damaging variants in this gene. One novel predicted to be damaging in frame deletion (c.1523_1525delGCT; p.508_509del) and two rare variants predicted to be damaging (MAF: 0.0000647, c.3179_3180insC, p.R1060fs and, MAF: 0.00001192, c.2364G>C, p.E7888D) were identified in *NCOR2* in these families.

NCOR2 (2,514aa, Z:2.09 and pLI:1) is a well known transcriptional co-repressor that mediates transcriptional silencing of target genes and is part of NCOR complex, which includes histone deacetylases to modify chromatin structures that prevents basal transcriptional activity of target genes. *NCOR2* plays key roles in the cell cycle, apoptosis and proliferation of many cancer cells, including head and neck squamous cell carcinoma and breast cancer. Furthermore, cellular differentiation of human *in vitro* generated monocytes is regulated by NCOR2 and time-dependent Interleukin-4 signalling (Sander *et al*., 2017).

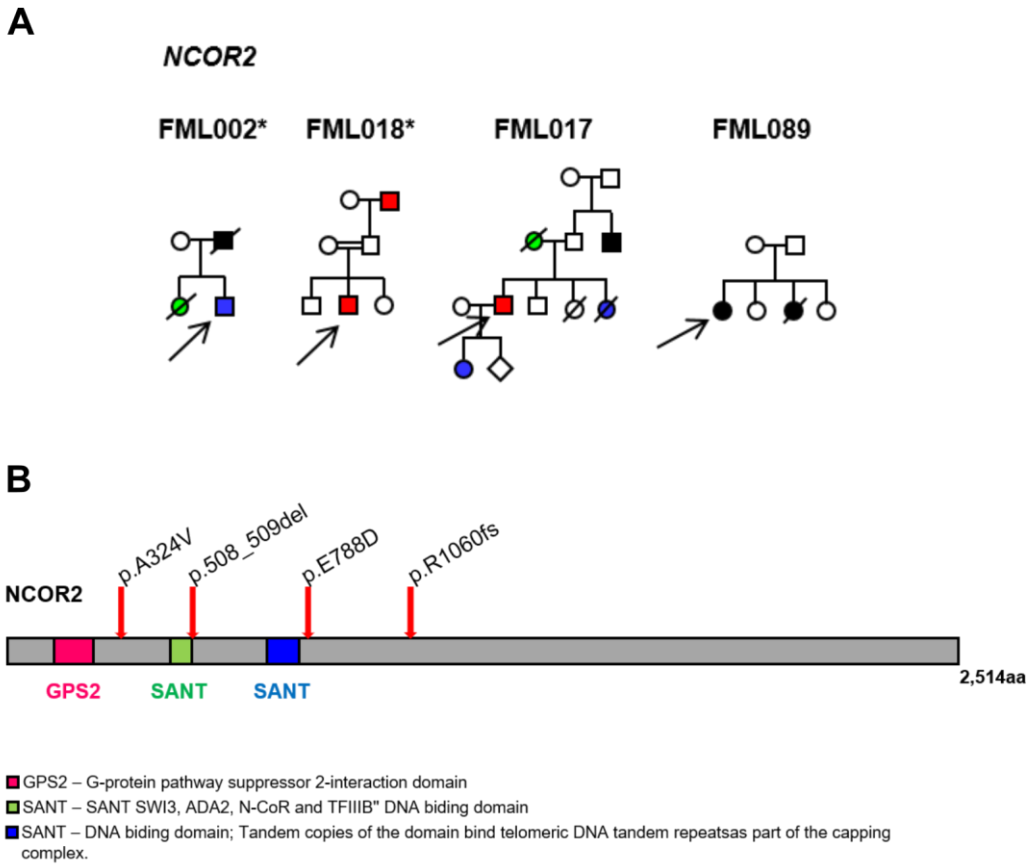


Figure 7.2. Familial MDS/AML candidate gene *NCOR2*. A. Uncharacterised families with MDS/AML that harbour predicted damaging heterozygous rare variants in *NCOR2*. Affected individuals are coloured as following black: AML, red: AML, blue: AA, grey: thrombocytopenia and green: solid cancer. The arrow highlights the index case. *: families that harbour variants in at least two of the candidate genes presented in Table 7.1. B. Schematic of RUNX1 protein (NP_006303.4).

Although the role of NCOR2 in the FLT3/p65 signalling pathway and aberrant cell growth is unclear, NCOR2 nuclear protein levels are significantly increased when *FLT3* and *p65* are downregulated *in vitro* and *in vivo* (Wang *et al*., 2012). *FLT3* is an independent poor prognostic marker of sporadic AML, it has an important role in the survival, proliferation and differentiation of haematopoietic cells and AML blasts. Overexpression of *FLT3* is present in 70-90% of sporadic AML. Activating mutations

of *FLT3* may disrupt transcriptional repressor function resulting in aberrant gene regulation and abnormal leukaemia cell growth. In addition, between 20% and 30% of AML patients have either an internal tandem duplication region or a point mutation in the FLT3 receptor leading to the constitutive activation of downstream signalling pathways and aberrant cell growth. *FLT3* internal tandem duplication induced translocation of NCOR2 protein from the nucleus to cytoplasm as well as it inhibited the function of transcriptional repressors by blocking protein interactions with NCOR2 (Takahashi *et al*., 2004). A gene fusion NCO2/SCARB1 and a missense variant of unknown significance were identified in AML cases in the TCGA published in The Cancer Genome Atlas (Cancer Genome Atlas Research Network, 2013).

It would be interestingly to see the differences in NCOR2 protein expression between the affected individuals compared to a control, considering available samples of these patients by Western blotting. If there were significant differences, it would be important to discover the pathway that might have been disrupted to give rise to MDS/AML phenotype by designing new experiments to be performed. Furthermore, CRISPR-Cas9 system could be used to introduce the *NCOR2* variants identified in this study into haematological cells and a cell culture of these modified cells and a wild type *NCOR2* cell would be useful to verify the differences in cell growth as a result of specific *NCOR2* variants.

We have analysed MDS/AML families where we had undertaken WES in more than one member within the family. In chapter 6, the AD analysis of WES data from family FML012 (Figure 6.4, section 6.2.1.2) was described. In this family we were left with only two germline variants in the genes *CDKL1* and *CEP68* (Table 6.9, section 6.2.1.2). As occurred previously, there is not sufficient evidence on either of these candidate genes to be considered disease causing in this family. Reasons for these

conclusion are: a) both variants are missense not predicted to be damaging by both prediction tools used in this study (Polyphen2 and MutationTaster2), b) both were found in another asymptomatic family member, c) no additional families from our cohort were found to harbour damaging variants in neither of these genes. Despite all discussed above, it is important to note *CDKL1* might be the best candidate gene in FML012 as the individual III:8 (Figure 6.5 A) harboured a germline heterozygous *CDKL1* variant (c.416C>T; p.S139F). She was at the time of this study asymptomatic, although described to have monoclonal gammopathy (MGUS). People with MGUS usually do not present with symptoms, but it can develop into myeloma or a related blood disorder (Umemura *et al.*, 2018). To investigate the pathogenicity of this *CDKL1* missense variant, it would be reasonable to a) verify the expression of *CDKL1* in the patient's blood cells compared to a control by Western blotting and qPCR as *CDKL1* is known to be overexpressed in malignant tumors such as melanoma, breast cancer, and gastric cancer (Qin *et al.*, 2017); b) the next decision should be made upon obtained results from the CDKL1 protein levels present in the affected individuals.

Family FML003 (Figure 6.7, section 6.2.1.3 in chapter 6) is another inconclusive case as AR analysis generated five genes that harboured rare homozygous variants (Table 6.10, section 6.2.1.3 in chapter 6). However all of them were found in a heterozygous state in both parents of the index case. This finding left us five genes as candidates. From these genes only epidermal growth factor receptor pathway substrate 15 (*EPS15*) variant is predicted to be benign by Polyphen2 and MutationTaster2 and therefore this is a weak candidate gene; this still leaves four candidate genes in FML003 (Table 6.10, 6.2.1.3 in chapter 6). Thus, it would be necessary to have DNA sample from one of the affected siblings (which was not possible) in order to identify the likely disease causing variant in this family as well as the identification of likely damaging variants in the same candidate gene in other affected MDS/AML families.

Finally, after such an extensive search for the disease gene (s) in these uncharacterised families with no conclusive disease gene to date, there are some points that need to be considered: a) the pathogenic variant might be in a region not covered by WES, and b) a technical or procedure issue that need to be identified and resolved.

7.5 Future work

Firstly, collection of ideal germline tissue (blood and another non-blood samples) and complete documentation of family history when possible are essential in order to distinguish germline from somatic variants when segregation analysis is not possible. Furthermore, as studies have suggested that AML occurs after a process of genetic and epigenetic changes that are found in preleukaemic clone before the cells are transformed into AML by an acquisition of additional alterations (Pandolfi *et al*., 2013), it would be interesting to perform clone evolution studies of patients with familial MDS/AML. Thus, collection of appropriate samples is important. The initial germline variant in patients with familial MDS/AML could be considered a preleukaemic event, required but sometimes not sufficient for cancer development as in principle, both germline and somatic variants of the same gene lead to AML progression. Additional somatic variants are reported in *CEBPA*, *RUNX1* and *DDX41* patients and clonal architecture of leukaemic clones revealed *DDX41* variants could be a founder event in some patients (Polprasert *et al*., 2015).

The whole exome sequencing technology is proven to be of great value for identification of pathogenic variants in inherited disorders, although many of the families with MDS/AML from our cohort remain uncharacterised. Performing a combination of techniques such as array comparative genomic hybridisation (aCGH) and RNA sequencing (mRNA expression analysis) when feasible, in order to cover all DNA regions and assemble a more complete variant information may provide additional help in search for the genetic cause of these families. A comprehensive genomic landscape of our familial MDS/AML cohort could be described using WES along with RNA sequencing and computational analysis. In addition, copy number variation could be analysed from our WES data.

Following up of potential disease causing variants by screening additional family samples and by performing segregation studies are also very good future approaches to strengthen the genetic evidence for the pathogenic role of identified variants (for instance, for the four candidate genes identified in family FML003 and for the candidates listed in Table 7.1).

In conclusion, the recruitment of new families (to increase the allelic series of subtypes of patients) together with incorporation of new techniques (such as RNA sequencing) as well as functional analysis of specific candidate genes should enable the identification of new disease causing genes in familial MDS/AML.

7.6 Concluding remarks

Familial MDS/AML occurs in both children as well as in adults and studies of this disease is providing better understanding of the biology of both sporadic and familial forms. Furthermore, understanding when to suspect germline predisposition in a patient is essential as early and accurate diagnosis influence clinical care of these patients and their families. Hence, identification of causal genetic variants is critical for this diagnosis. Knowledge of which genes contribute to familial MDS/AML predisposition and attention to the acquired variants in genes associated with this inherited disease is very important. This is crucial for ensuring that appropriate donors are selected for allogeneic transplantation. Although complete human genomes alteration are not obtained from any currently single sequencing technology, exome sequencing is an effective disease causing variant detection method because it targets protein-change and amino acid substitutions which currently account for approximately half of the known gene alterations responsible for human inherited disease (Ng *et al*., 2010). However, WES as well as other next generation sequencing bring some technical, data management and interpretation issues. Many guidelines are being proposed in order to help overcome these challenges.

Variant interpretation is one of the most encountered challenge in WES data analysis as it is a collaborative work of different highly trained specialists that require a long and complex research process. Certainly, the creation of large scale databases by sharing sequencing results from many different studies will help to exclude variants with higher minor allele frequency as causal. Computational prediction of variants on protein impact is a good guide on the clinical interpretation of variants of uncertain significance and on genome studies of diseases. However, it is important to be clear that these well-established *in silico* prediction tools are indicators and for confirming

a variant pathogenicity it is essential to perform a combination of approaches such as functional analysis, use of population database as ExAC to verify the population variant allele frequency and family investigation. Overall, there is a need to develop more efficient and standardised bioinformatics tools to filter, analyse and interpret WES variants.

Our results also reflected the extensive genetic heterogeneity and phenotypic complexity of bone marrow failure diseases including familial MDS/AML. These disorders present significant overlapping features and variable expressivity and penetrance within syndromes. These along with absent, subtle, or previously unreported clinical findings add to the challenges of a timely and precise diagnosis.

Finally, it is essential that clinicians, genetic counsellors, basic scientists and patients work together for establishing a better understanding and new therapies to improve patient outcomes.

Chapter 8

Appendices

Appendix 1 – Known familial MDS/AML causing genes – published mutations to date

Table A1.1. Published *RUNX1* mutations in familial MDS/AML cases to date

Paper	Families	Variant
Song *et al.* 1999 Nature Genetics	1	intragenic deletion
	2	cryptic splice acceptor to stop
	3	nonsense
	4	nonsense
	5	c.R201Q
	6	c.R166Q
Owen *et al*., 2008 Blood	1	c.1007_1013del;p.G336fsX563
	2	c.83insG;p.A28fsX109
	3	cG286C; p.D96H
	4	c.+3G>A; p.K90fs
	5	cC877T; p.R292X
Ripperger *et al*., 2009 Leukemia	1	c.C520T; p.R174X
Jongmans *et al*., 2010 Leukemia	1	162 kb duplication in chromosome 21
	2	105bp deletion
Langabeer *et al*., 2010 European Journal of Haematology	1	c.507delA; p.W169fsX182
Schmit *et al*., 2015 Leukemia Research Reports	1	c.G837A; p.W279X c.422_423insAAGGC; p.S141_A142insRP
about 30 affected families have been reported since 1999 (Yoshimi *et al.*, 2014 Nature communications)		

Table A1.1. Continued

Antony-Debré *et al*., 2016 Leukemia less than 45 affected families have been reported since 1999	1	p.R177Q
	2	p.Q308RfsX259
	3	Complete deletion of RUNX1
	4	p.R139X
	5	p.P218S
	6	p.G108V
	7	p.D305TfsX262
	8	p.H377PfsX191
	9	p.G108V
	10	p.G143RfsX43
	11	p.T169R
	12	Complete deletion of RUNX1
	13	p.T219RfsX8
	14	p.T121HfsX9
	15	p.A129E
Haslam *et al*., 2016 British Journal of Haematology	1	c.496C>T; p.Arg166X
Sakurai *et al*., 2016 Blood Cancer Journal	1	285 kb heterozygous deletion including the promoter and the 5'-half of RUNX1
	2	2 Mb heterozygous deletion in chromosome 21 encompassing the entire RUNX1 gene and a large genomic region of 5'-RUNX1
Yoshimi *et al*., 2016 Annals of Oncology	1	p.Thr233fs
	2	p.Phe303fs
	3	p.Arg174
	4	p.Gly262fs
	5	p.Ser140Asn
	6	p. Gly172Glu
	7	p.Asn438Lys
	8	p.Leu445Pro
Hamadou *et al*., 2016 Annals of Hematology	1	p.L56S
Latger-Cannard *et al*., 2016 Orphanet Journal of Rare Diseases	1	c.320G > A; p.Arg107His
	2	c.467C > A; p.Ala156Glu
	3	c.602G > A; p.Arg201Gln
	4	c.611G > A; p.Arg204Gln
	5	c.587C > G; P.Thr196Arg
	6	c. 999_1003dup; p.Gln335Argfs261
	7	c. 1092del; p.Ile364Metfs230
	8	c. 442_449del; p.Thr148Hisfs9
	9	c. 496C > T; p.Arg166X

Table A1.1. Continued

DiNardo *et al*., 2016 Clinical Lymphoma myelomaLeukemia	1 2 3	c.582A > C; p.K194N c.610C > T; p.Arg201X c. 1098_1103dupCGGCAT; p.I366_G367dup
Badin *et al*., 2017 Haemophilia	1	c.583dupA; p.Ile195AsnfsX18 c. G>A; p.T246M
Tawana *et al.*, 2017 European Journal of human genetics	1	c.601C>T; p.Arg201*
Kanagal-Shamanna *et al*., 2017 Haematologica	1 2 3 4 5 6 7	c.582A>C p.K194N c.719delC p.Pro240Hisfs and c.167T>T p.Leu56Ser Partial gene deletion (at least exons 1-6) c.836G>A p.W279* c.496C>T p.R166* c.308dup p.T104fs c.1098_1103dupCGGCAT p.I366_G367dup
de Andrade Silva *et al*., 2018 Cancer genetics	1	Deletion of exon 1 and 2 of RUNX1

Table A1.2. Published *CEBPA* mutations in familial MDS/AML cases to date

Paper	**Families**	**Variant**
Smith *et al. 2004* The new England Journal of Medicine	1	c.212delC c.1050_1085dup; p.302_313KAKQRNVETQQK
Sellick *et al*., 2005 Leukemia	1	c.217insC;p.fsX106 c.1071delGAGACGCinsCTGGAGGCCA; p.E308_Q310delinsLEAK c.107ins GAC; p.E308dup
Nanri *et al*., 2006 Blood	1	c.350_351insCTAC; p.I68fsX107 c.1063ins18bp c.1079ins3bp c.1083ins3bp

Table A1.2. Continued

Pabst et al., 2008 Journal of Clinical Oncology	1	c.744-745GC>TT;p.A199L
		c.1167G>A;p.G340S
	2	c.563-564insCG;p.Y138fsX160
		c.1094-1095insCTG;p.L315-316ins
	3	c327-328insC.;p.E59fsX107
		c.1098-1099insGTC;p.V316-317ins
	4	c.551G>A;p.A134A
		c.742-743insGCCGCCCC;p.P199fsX318
	5	c236-237insGC.;p.A29fsX160
	6	c.395del;p.F82fsX159
		c.1076-1077insAAG;p.K309-310ins
	7	c.213delC;p.S21Q
		c.1088-1089insTCT;p.P22fsX159
		c.212C>A;p.S313-314ins
	8	c.1083C>T;p.Q312X
	9	c.672C>G;p.L175V
		c.676C>T;p.A176V
		c.678-679GG>TT;p.G177F
		c.683C>T;p.L178L
		c.688C>A;p.P180H
		c.692C>G;p.Y181X
	10	c.1079-1080insTCT;p.S310-311ins
	11	c.420-421insT;p.A91fsX107
	12	c.327-328insC;p.E59fsX107
		c.1098-1099insGTC;p.V316-317ins
	13	c.465insT;p.D106fsX107
		c.1089insAAG;p.314insK
		c.1207G>C
		c.1210A>C
		c.1089insAAG
	14	c.291delC;p.A47fsX159
		c.1086insCAG;p.313insQ
		c.1086insCAG
	15	c.245delG;p.G32fsX159
	16	c.216-217insCG;p.P23fsX160
		c.1165G>C;p.R339P
	17	c.286-287insTC;p.P46fsX160
		c.1076-1077insCCG;p.K309-310ins
	18	c.1094-1095insCTG;p.L315-316ins
		c.687T>C;p.P180P
Renneville *et al*., 2009	1	c.217–218insC
		c.1083–1085delAAG
		c.1065–1066insGGG

Table A1.2. Continued

Nanri *et al*., 2010 Genes, Chromosomes & Cancer	1	c.350_351insCTAC; p.I68fsX107 c.1063–1064ins18-bp c.1079–1080insCAG c.1085–1086insAAG
Stelljes *et al*., 2011 Leukemia	1	c. 338delC; p. c.1085insGAA; p. c. 1072_1083dup; p.
Green *et al*., 2013 British Journal of Haematology	1	c.68delC, p.P23fs c.K302_K313dup
Tawana *et al*., 2015 Blood	1	c.218delC; p.P23RfsX137 c.1054_1089dup; p.K302_K313dup c.1087_1089dup; p.K313dup
	2	c.218_219insC; p.H24AfsX84 c.991_992insGA; p.N281RfsX38 c.1067_1068insGCG; p.R306dup
	3	c.218_219insC; p.H24AfsX84 c.1047_1088dup; p. R300_K313dup
	4	c.297_315del; p.E50AfsX104 c.1087_1089dup, c.1061_1210del; p.K313dup, p.K304_A353del c.1087_1089dup; p.K313dup
	5	c.351_352 ins CTAC; p.I68LfsX41 c.1067_1068insGGCCCTCGCCCCCCCGCG; p.R306_N307insALAPPR c.1087_1089dup; p.K313dup
	6	c.218_219insC; p.H24AfsX84 c.1075_1081delinsCTGGAGGCCA; p.E309_Q311 delins LEAK c.1075_1077dup; p.E309dup
	7	c.339delC; p.D63EfsX97 c.1087_1089dup; p.K313dup c.1076_1087dup; p.E309-Q312dup
	8	c.308delG; p.G53AfsX107 c.1126_1127ins1079_1227; p.K326 insT310_X359
	9	c.291delC; p.A48PfsX112 c.1085_1087dup; p.Q312dup
	10	c.464_465insT; p.F106LfsX2 c.G1207C; c.A1210C; p.A353P; p.M354L c.1087_1089dup; p.K313dup
Pathak *et al*., 2016 Haematologica	1	c.932A>C; p.Q311P
Ram *et al*., 2017 Blood Advances	1	c.68dupC; p.H24fs*84 c.442G.T; p.Glu148*

Table A1.3. Published *TERC* mutations in familial MDS/AML cases to date

Paper	Families	Variant
Kirwan *et al. 2009* Human mutation and in Holme *et al*., 2012 British Journal of Haematology	1 2	c.212C>G c.309G>T

Table A1.4. Published *TERT* mutations in familial MDS/AML cases to date

Paper	Families	Variant
Kirwan *et al. 2009* Human mutation and in *Holme et al*., 2012 British Journal of Haematology	1 2 1	c.1892G>A; p.Arg631Gln c.2354C>T; p.Pro785Leu c.248G>C; p.Arg83Pro

Table A1.5. Published *GATA2* mutations in familial MDS/AML cases to date

Paper	Families	Variant
Hahn *et al. 2011* Nature Genetics	1	c.1007_1013del;p.G336fsX563
	2	c.83insG;p.A28fsX109
	3	cG286C; p.D96H
	4	c.+3G>A; p.K90fs
Hsu *et al*., 2011 Blood	1	c.1192C>T; p. R398W
	2	c.1192C>T; p. R398W
	3	c.1192C>T; p. R398W
	4	c.1061>T; p.T354M
	5	c.243_244delAinsGC; p.G81fs
	6	c.1192C>T; p. R398W
	7	c.1113 C>G; p.N371K
	8	c.1083_1094del 12 bp; p.R361delRNAN
	9	c.1–200_871; 527del 2033 bp M1del290
	10	c.1186 C>T; p.R396W
	11	c.1061 C>T; p.T354M
	12	c.1187 G>A; p.R396Q
	13	c.1061 C>T; p.T354M
	14	c.778_779ins 10 bp D259fs
	15	c.1192C>T; p. R398W
	16	c. 951_952ins 11 bp N317fs
	17	c. 751 C>T; p. P254L
	18	c. 1018–1 G>A; p.D340–381

Table A1.5. Continued

Ostergaard *et al*., 2011 Nature Genetics	1 2	c.310_311insCC; p.Leu105ProfsX15 c.230-1_230insC; p.Arg78ProfsX107
Dickinson *et al*., 2011 Blood	1 2	c.1192C>T; p. R398W c.1018-1 G>T; del 340-381
Kazenwadel *et al*., 2012 Blood	1 2	c.; p.Thr354Me c.;p.Met1del290
Holme *et al*., 2012 British Journal of Haematology	1 2 3 4	c.310_311insCC; p.Leu105ProfsX15 c.121C>G; p.Pro41Ala c.1187G>A, p.Arg396Glu c.1061C>T, p.Thr354Met
Bodor *et al*., 2012 Haematologica Secondary mutations in ASXL1	1	c.; p.Thr354Met
Pasquet *et al*., 2013 Blood	1 2 3 4 5 6 7	c.1187G>A; p.R396Q c.1114G>A,; p.A372T c.1162A>G; p.M388V c.c.988C>T; p.R330X c.610C>T,; p.R204X c.670G>T; p.E224X deletion of 61 kb
Green *et al*., 2013 British Journal of Haematology	1	c.310_311insCC; p.L321F c.310_311insCC; p.R330Q
Gao *et al*., 2014 Journal of Haematology and Oncology	1 2	c.; p.p.Thr358Asn c.; p.Leu359Val
Churpek *et al*., 2015 Blood	1 2 3	c.10171+572 C>T c.1192C>T; p.R398W c.1061C>T; p.T354M
Malhi *et al*., 2016 Pediatric Transplantation	1 2	c.917G>A; p.Trp306* c.1009 C>T; p.Arg337X
Fisher *et al*., 2017 Blood Advances	1 2 3 4 5	c.1018-1G>A c.1018-2A>C c.1144-1G>C c.599delG, p.G200VfsX18 3.1-3.3 Mb het del encompassing *GATA2*

Table A1.6. Published *SRP72* mutations in familial MDS/AML cases to date

Paper	**Families**	**Variant**
Kirwan *et al. 2009* The American Journal of Human Genetics	1 2	c.1064_1065del; p.Thr355Lysfs*19 c.620G>A; p.Arg207His

Table A1.7. Published *ANKRD26* 5'UTR mutations in familial MDS/AML cases to date

Paper	Families	Variant
Pippucci *et al. 2011* American Journal Human Genetics	1	c.-118C>T
	2	c.-127A>T
	3	c.-128G>A
	4	c.-134G>A
	5	c.-127A>T
	6	c.-128G>A
	7	c.-125T>G
	8	c.-116C>T
Noris *et al*., 2011 Blood	1	c.-113A>C
	2	c.-118C >T
	3	c.-118C>A
	4	c.-119C>A
	5	c.-121A>C
	6	c.-125T>G
	7	c.-126T>G
	8	c.-127A>T
	9	c.-127A>G
	10	c.-128G>A
	11	c.-128G> A
	12	c.-134G>A
Noris *et al*., 2013 Blood	1	c.-116C.G*
	2	c.-118C>A
	3	c.-118C>T
	4	c.-119C>A
	5	c.-126T>G
	6	c.-127A>G
	7	c.-127A>G
	8	c.-127A>T
	9	c.-127A>T
	10	c.-127delAT*
	11	c.-128G>A
	12	c.-128G>A
	13	c.-128G>A
	14	c.-128G>A
	15	c.-128G>A
	16	c.-128G>A
	17	c.-128G>A
	18	c.-128G>A
	19	c.-128G>A
	20	c.-134G>A
	21	c.-134G>A
	22	c.-134G>A
	23	c.-134G>A

Table A1.7. Continued

Perez Botero *et al*., 2015 Blood Cancer	1	c.-116 C4T ASXL1 c.2290delC
Ouchi-Uchiyama *et al*., 2016 Pediatric Blood Cancer	1 2 3 4	c.-118C >T c.-118C >T c.-134G>A c.-134G>A
Yoshimi *et al*., 2016 Annals of Oncology	1	c.-134G > A
Vincenot *et al*., 2016 Annales de Biologie Clinique	1	c.-127A>T
Tsang *et al*., 2017 Modern pathology	1 2	c.-134G > A
Marconi *et al*., 2017 Journal of Haematology and Oncology	1	c.-125T>G
Averina *et al*., 2017 Thrombosis Research	1	c.-128G>T
Ferrari *et al*., 2017 Platelets	1 2 3	c.-128G>A c.-134G > A c.-140C > G
Guison *et al*., 2017 Mediterranean Journal of Hematology and Infectious Diseases	1	c.-127C>A
Zaninetti *et al*., 2017 Journal of Thrombosis and Haemostasis	1 2	c.-128G>A c.-128G>A

Table A1.8. Published *ETV6* mutations in familial MDS/AML cases to date

Paper	**Families**	**Variant**
Zhang *et al. 2015* Nature Genetics	1 2 3	c.1106G>A ; p.Arg369Gln c.1195C>T; p.Arg399Cys c.641C>T; p.Pro214Leu
Noetzli *et al*., 2015 Nature Genetics	1 2 3	c.641C>T; p.Pro214Leu c.641C>T; p.Pro214Leu c.1252A>G; p.Arg418Gly
Topka *et al*., 201 Plos Genetics	1 2	c.T1046C; p. L349P c.1153-5_1153_1delAACAG; p. N385fs
Poggi *et al*., 2017 Haematologica	1 2 3 4 5 6	p.A377T p.Y401H c.641C>T; p.Pro214Leu p.Y401N p.I358M p.R396G

Table A1.8. Continued

Melazzini *et al*., 2016 Haematologica	1	c.641C>T; p.P214L
	2	c.1252A>G; p.R418G c.1153-1_1165del; p.N385Vfs*7
	3	c.1138T>A; p.W380R
	4	c.1105C>T; p.R369W
	5	c.641C>T; p.P214L
	6	c.1153-1_1165del; p.N385Vfs*7
Duployez *et al*., 2017 European Journal of Haematology	1	R378X

Table A1.9. Published *DDX41* mutations in familial MDS/AML cases to date

See **Table 4.1** in chapter 4, section 4.14, page 144.

Appendix 2 – Primer sequences

Table A2.1. Primers used for rolling circle amplification

	Primer sequence
Arqbidopsis	Thio-$(C_3TA_3)_3$ 5'-CCCTAAACCCTAAACCCTaaa-3'
Homo sapiens	hC21Thio 5'-CCCTAACCCTAACCCTAAccc-3'

Table A2.2. Primers used for monochrome multiplex qPCR

	Primer sequence
telg	ACACTAAGGTTTGGGTTTGGGTTTGGGTTTGGGTTAGTGT
telc	TGTTAGGTATCCCTATCCCTATCCCTATCCCTATCCCTAACA
hbgu	CGGCGGCGGGCGGCGCGGGCTGGGCGGCTTCATCCACGTTCACCTTG
hbgd	GCCCGGCCCGCCGCGCCCGTCCCGCCGGAGGAGAAGTCTGCCGTT

Table A2.3. Primers used for dHPLC – *TERT* promoter

	Forward primer	**Reverse primer**	**Oligo length (bp)**
***TERT* promoter**	GGCCGATTCGACCTCTCT	AGCACCTCGCGGTAGTGG	453

Table A2.4. Primers used for dHPLC – *ANKRD26* 5'UTR

	Forward primer	**Reverse primer**	**Oligo length (bp)**
***ANKRD26* 5'UTR**	CCAGTCGCCGAGATTTGC	CTTGCCGAGATCTCGGTC	396

Table A2.5. Primers used for dHPLC - *CDKL1*

***CDKL1* exons**	**Forward primer**	**Reverse primer**	**Oligo length (bp)**
1	TCTCCTGAGTTGCTGGGAC	GAGGCAGAAGCATGGCTT G	362
2	GGACAAACAACTTACTTTTTATTGG	GCCTTACCTAGTCTTAAAAAGAG	315
3	CTAAGATTTTTCACTACTAAAGCC	GTGTGACATAAACTTTGTACAAG	402
4	TAGTCATGTGTTTCATTTATCCC	TAGAGCATTTAGATCCTTTGTTG	358
5	CCACAAAATTGGGCACAGG	TGAGACTGAGAACTTGGCTC	391
6	CAGCATTGATGGAAGAAAACC	GGCTCAGGAGGTGAATAAC	432
7	GGACGTGAGGAAGGTGC	TTTCATTGCATGGATTGACATAAG	321
8	GACTGGTGTGTTTTCCTGC	CTCAAGTCTAGATTCCAACTC	384
9	CTCTTTTGTGATATATTCTAAATAATG	GGCCTCCCAGTTTCTTGC	359

Table A2.6. Primers used for dHPLC – *CEP68*

***CEP68* exons**	**Forward primer**	**Reverse primer**	**Oligo length (bp)**
2A	GCAAGACTTCCTGAACACAG	GGTCAGTCCCAATCCAGC	324
2B	TCCTATGGGAGAGGGAGC	GAAAGGATGCAAGATGGTGAG	386
3A	CCACAAGGAAACTCCTGGG	CCTGTGGAGGAAGCAGAG	341
3B	TGAGCCTTCCCAGAACAAC	TGCCTAGCCCAGAAGCATC	387
3C	GCATCTCTGCTTCCTCCAC	ACGGCTATCAAGGTGCTTTG	377
3D	CTCTCCTTCCAGGCTGAG	CCATGCCACCCTGTTTCTG	370
3E	TGATAGCCGTGTGCCAGC	GGGAGAGCCCATATCAAGG	352
3F	TAAGCAGTGGCCCTCCAG	GAAACCTGTGTCAGCCGAG	354
3G	GCAAGCACCTTGATATGGG	AGCTGGCTTTGGCTGGAG	334
3H	GGA AGTGGAAAGTGATGACG	CAGCAGTCCAGAGGAGAC	340
3I	CAGACAGTGATGGGCCAG	GGGGGAAGAGACAAGACAC	375
4	AGTCTGTCT CCAAGTCCTC	GCGTTCCGCTTTAAGAATGTG	382
5	ACACATTAGAGGGGA ATAAGG	TACCAGGGATAGTGCGGT C	285
6	TGCCAGGTCCTACTTTGTG	TGTTACAGCAGGGTTGGTG	343

Table A2.7. Primers used for Sanger sequencing validation

	Forward primer	**Reverse primer**	**Oligo length (bp)**
***RUNX1* exon 3**	CTACACAAATGCCCTAAAAGTG	ACCGAGTTTCTAGGGATTCC	330
***GATA2* exon 4**	CCACCCAAAGAAGTGTCTCCTGA	GCCGGCACATAGGAGGGGTAG	392
***GATA2* exon 6**	GTCAGGGAGGGGGGTC	GCCCTTCTGGCGCTCAC	345
***TERT* exon 2**	GTTTCTGGAGCTGCTTGGGA	AGCCCCTACTGCATTCAGCT	455
***SBDS* exon 2**	GGCTGAGGTTACAGTGACC	TGCTTGGTTAGTCTTTCCTCC	478
***FANCA*exon 27**	TGCTCAGGCCATCCAGTTC	CCTGAGATGGGCACAAAGC	322
***FANCA*exon 36**	GTA GTG GCC TGT AGG AGC	CCACCACCACGAGAACTC	368
***WAS* exon 10**	ACTGGACGTTCTGGACCAC	CCAACCTTTCAACCCTATCAC	353
***DDX41* exon 1**	CTCCGAGGTCGTTCCTAC	GTCCTCGTCGTCCTCATC	200
***DDX41* exon 3**	GACCGACGGCTTGATCTG	GACTCTTTGCGCGCTGAG	402
***DDX41* exon 8A**	CAACACCCATTCAGATCCAG	GTCTCCATCTGCTCTTTCAC	510
***DDX41* exon 15**	AGAACTATGGTAAGAGCCTGG	GGTCCATCAGCACTGACTC	319
***RTEL1* exon 20**	AGCACTGAGGCCTGAGGTC	AAGCTGTGAGAGGCAGGGG	423
***RTEL1* exon 25**	GCAGCAGATGAGGGTCTTC	CAAAGCCAGGTGAGTCGC	362
***RTEL1* exon 29**	TTTCTCAGGCAGCAGCCC	AGAGAACAGAGACCACCTTG	351
***RTEL1* exon 30A**	CCAAGGTGGTCTCTGTTCTC	CCACGCAGGAGTCTGAGG	326
***RTEL1* exon 30B**	ACTACAAGGGTTCCGATGAC	GGTCGTCGTCTTGCTTATAG	513
***RTEL1* exon 32**	GCAGTTGTCCTGAGCAGC	TCAATCAGACCCGGCACAG	415
***DUT* Exon 3**	GGTAATTCATCATAGCAAGGTTG	GGTGCTTCTTTTAGGACACAG	365
***EPS15* Exon 23**	TGAGCTGTTTTTGGTTTGATCT	AGAATGAATGACAGCAGCAGG	549
***PDIA3* Exon 9**	TCTGCATATTGAGAGATGAGAG	AACGAAGTCTTCATTTAGACCC	339
***UNC13C* Exon 3**	GGAGCTGACCCTGCTTAG	CAGCATCTTGCACTCAAAGC	294
***ZNF333* Exon 12**	TTGCACCAGAGAAACCACAC	TTCGCATGTGACTCTTCAGG	349

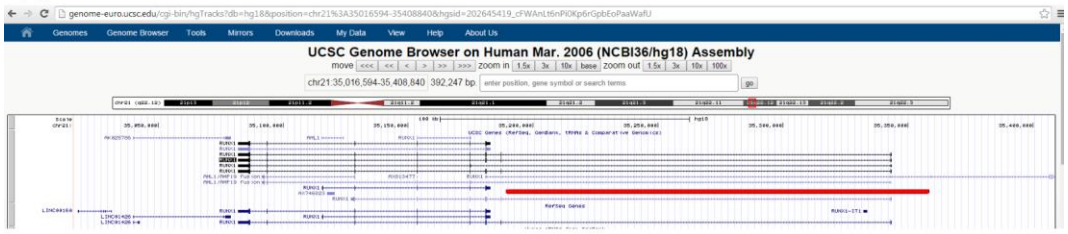
Appendix 3 – TSCA studies

Table A3.1. TSCA gene panel performed in some DNA samples described in chapters 4 and 5

Genes	
ACD	*NOP10*
ANKRD26	*PARN*
CEBPA	*PAX5*
CTC1	*RECQL4*
DDX41	*RMRP*
DKC1	*RTEL1*
DNAJC21	*RUNX1*
ERCC6L2	*SLX4*
ETV6	*SRP72*
FANCA	*TERC*
FANCC	*TERT*
FANCD2	*TINF2*
FANCG	*TP53*
GATA2	USB1
GRHL2	*WAS*
LIG4	*WRAP53*
NHP2	

Appendix 4 – Deletion of *RUNX1* in family FML053

A



B

Figure A4.1. Deletion of *RUNX1* in family FML053. An UCSC Genome Browser (https://genome.ucsc.edu/) window showing the position of the 162kb *RUNX1* deletion in the chromosome 21. The horizontal red bar indicates the deletion. B.BAM file showing the *RUNX1* position in the chromosome 21. The horizontal red bar indicates the region that is included in the 162kb *RUNX1* deletion (NR_073512 and *SETD4*).

Chapter 9

References

Abdelhaleem M. Do human RNA helicases have a role in cancer? Biochimica et Biophysica Acta 2004; 1704:37-46.

Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013 Jan; Chapter 7:Unit7.20.

Ahmad Y, Boisvert FM, Gregor P, Cobley A, Lamond AI. NOPdb: Nucleolar Proteome Database--2008 update. Nucleic Acids Res. 2009 Jan; 37:D181-184.

Ahmed D, Eide PW, Eilertsen IA, Danielsen SA, Eknæs M, Hektoen M, Lind GE, Lothe RA. Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis. 2013 Sep 16;2:e71.

Alter BP, Giri N, Savage SA, Rosenberg PS. Cancer in dyskeratosis congenita. Blood. 2009 Jun 25;113(26):6549-6557.

Alter BP, Giri N, Savage SA, Peters JA, Loud JT, Leathwood L, Carr AG, Greene MH, Rosenberg PS. Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. Br J Haematol. 2010 Jul;150(2):179-188.

Alter BP, Rosenberg PS, Giri N, Baerlocher GM, Lansdorp PM, Savage SA. Telomere length is associated with disease severity and declines with age in dyskeratosis congenita. Haematologica. 2012 Mar;97(3):353-359.

Alter BP. Inherited bone marrow failure syndromes: considerations pre- and posttransplant. Hematology Am Soc Hematol Educ Program. 2017 Dec 8;2017(1):88-95.

Amiard S, Doudeau M, Pinte S, Poulet A, Lenain C, Faivre-Moskalenko C, Angelov D, Hug N, Vindigni A, Bouvet P, Paoletti J, Gilson E, Giraud-Panis MJ. A topological mechanism for TRF2-enhanced strand invasion. Nat Struct Mol Biol. 2007 Feb;14(2):147-154.

Ancliff PJ, Blundell MP, Cory GO, Calle Y, Worth A, Kempski H, Burns S, Jones GE, Sinclair J, Kinnon C, Hann IM, Gale RE, Linch DC, Thrasher AJ. Two novel activating mutations in the Wiskott-Aldrich syndrome protein result in congenital neutropenia. Blood. 2006 Oct 1;108(7):2182-2189.

Andressoo JO, Hoeijmakers JH. Transcription-coupled repair and premature ageing. Mutat Res. 2005 Sep 4;577(1-2):179-194.

Antony-Debré I, Steidl U. Functionally relevant RNA helicase mutations in familial and sporadic myeloid malignancies. Cancer Cell. 2015 May 11;27(5):609-611.

Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016 May 19;127(20):2391-2405.

Armanios M, Chen JL, Chang YP, Brodsky RA, Hawkins A, Griffin CA, Eshleman JR, Cohen AR, Chakravarti A, Hamosh A, Greider CW. Haploinsufficiency of

telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. Proc Natl Acad Sci U S A. 2005 Nov 1;102(44):15960-15964.

Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, Lawson WE, Xie M, Vulto I, Phillips JA 3rd, Lansdorp PM, Greider CW, Loyd JE. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med. 2007 Mar 29;356(13):1317-1326.

Armanios M. Syndromes of telomere shortening. Annu Rev Genomics Hum Genet. 2009; 10:45-61.

Auerbach AD. Fanconi anemia and its diagnosis. Mutat Res. 2009 Jul 31;668(1-2):4-10.

Augustin I, Korte S, Rickmann M, Kretzschmar HA, Südhof TC, Herms JW, Brose N. The cerebellum-specific Munc13 isoform Munc13-3 regulates cerebellar synaptic transmission and motor learning in mice. J Neurosci. 2001 Jan 1;21(1):10-17.

Averina M, Jensvoll H, Strand H, Sovershaev M. A novel ANKRD26 gene variant causing inherited thrombocytopenia in a family of Finnish origin: Another brick in the wall? Thromb Res. 2017 Mar;151:41-43.

Badin MS, Iyer JK, Chong M, Graf L, Rivard GE, Waye JS, Paterson AD, Pare G, Hayward CPM. Molecular phenotype and bleeding risks of an inherited platelet disorder in a family with a RUNX1 frameshift mutation. Haemophilia. 2017 May;23(3):e204-e213.

Ballew BJ, Yeager M, Jacobs K, Giri N, Boland J, Burdett L, Alter BP, Savage SA. Germline mutations of regulator of telomere elongation helicase 1, RTEL1, in Dyskeratosis congenita. Hum Genet. 2013 Apr;132(4):473-480.

Baptista MA, Keszei M, Oliveira M, Sunahara KK, Andersson J, Dahlberg CI, Worth AJ, Liedén A, Kuo IC, Wallin RP, Snapper SB, Eidsmo L, Scheynius A, Karlsson MC, Bouma G, Burns SO, Forsell MN, Thrasher AJ, Nylén S, Westerberg LS. Deletion of Wiskott-Aldrich syndrome protein triggers Rac2 activity and increased cross-presentation by dendritic cells. Nat Commun. 2016 Jul 18;7:12175.

Barber LJ, Youds JL, Ward JD, McIlwraith MJ, O'Neil NJ, Petalcorin MI, Martin JS, Collis SJ, Cantor SB, Auclair M, Tissenbaum H, West SC, Rose AM, Boulton SJ. RTEL1 maintains genomic stability by suppressing homologous recombination. Cell. 2008 Oct 17;135(2):261-271.

Bates GJ, Nicol SM, Wilson BJ, Jacobs AM, Bourdon JC, Wardrop J, Gregory DJ, Lane DP, Perkins ND, Fuller-Pace FV. The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. EMBO J. 2005 Feb 9;24(3):543-553.

Baumann P, Cech TR. Pot1, the putative telomere end-binding protein in fission yeast and humans. Science. 2001 May 11;292(5519):1171-5. Erratum in: Science 2001 Jul 13;293(5528):214.

Beel K, Cotter MM, Blatny J, Bond J, Lucas G, Green F, Vanduppen V, Leung DW, Rooney S, Smith OP, Rosen MK, Vandenberghe P. A large kindred with X-linked neutropenia with an I294T mutation of the Wiskott-Aldrich syndrome gene. Br J Haematol. 2009 Jan;144(1):120-126.

Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, Kantarjian H, Raza A, Levine RL, Neuberg D, Ebert BL. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med. 2011 Jun 30;364(26):2496-2506.

Berger G, van den Berg E, Sikkema-Raddatz B, Abbott KM, Sinke RJ, Bungener LB, Mulder AB, Vellenga E. Re-emergence of acute myeloid leukemia in donor cells following allogeneic transplantation in a family with a germline DDX41 mutation. Leukemia. 2017 Feb;31(2):520-522.

Bertuch AA. The molecular genetics of the telomere biology disorders. RNA Biol. 2016 Aug 2;13(8):696-706.

Biernaux C, Loos M, Sels A, Huez G, Stryckmans P. Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals. Blood. 1995 Oct 15;86(8):3118-3122.

Bishop DK, Zickler D. Early decision; meiotic crossover interference prior to stable strand exchange and synapsis. Cell. 2004 Apr 2;117(1):9-15.

Bjork J, Johansson B, Broberg K & Albin M. Smoking as a risk factor for myelodysplastic syndromes and acute myeloid leukaemia and its relation to cytogenetic findings: a case-control study. Leukaemia Research 2009; 33,788–791.

Blackburn EH. Switching and signaling at the telomere. Cell. 2001 Sep 21;106(6):661-73.

Bluteau D, Balduini A, Balayn N, Currao M, Nurden P, Deswarte C, Leverger G, Noris P, Perrotta S, Solary E, Vainchenker W, Debili N, Favier R, Raslova H. Thrombocytopenia-associated mutations in the ANKRD26 regulatory region induce MAPK hyperactivation. J Clin Invest. 2014 Feb;124(2):580-591.

Bödör C, Renneville A, Smith M, Charazac A, Iqbal S, Etancelin P, Cavenagh J, Barnett MJ, Kramarzová K, Krishnan B, Matolcsy A, Preudhomme C, Fitzgibbon J, Owen C. Germ-line GATA2 p.THR354MET mutation in familial myelodysplastic syndrome with acquired monosomy 7 and ASXL1 mutation demonstrating rapid onset and poor survival. Haematologica. 2012 Jun;97(6):890-894.

Bogliolo M, Surrallés J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. Curr Opin Genet Dev. 2015 Aug; 33:32-40.

Bohl SR, Bullinger L, Rücker FG. Epigenetic therapy: azacytidine and decitabine in acute myeloid leukemia. Expert Rev Hematol. 2018 May;11(5):361-371.

Bonjardim CA, Ferreira PC, Kroon EG. Interferons: signaling, antiviral and viral evasion. Immunol Lett. 2009 Jan 29;122(1):1-11.

Boocock GR, Morrison JA, Popovic M, Richards N, Ellis L, Durie PR, Rommens JM. Mutations in SBDS are associated with Shwachman-Diamond syndrome. Nat Genet. 2003 Jan;33(1):97-101.

Borgdorff V, Lleonart ME, Bishop CL, Fessart D, Bergin AH, Overhoff MG, Beach DH. Multiple microRNAs rescue from Ras-induced senescence by inhibiting p21(Waf1/Cip1). Oncogene. 2010 Apr 15;29(15):2262-2271.

Botlagunta M, Vesuna F, Mironchik Y, Raman A, Lisok A, Winnard P Jr, Mukadam S, Van Diest P, Chen JH, Farabaugh P, Patel AH, Raman V. Oncogenic role of DDX3 in breast cancer biogenesis. Oncogene. 2008 Jun 26;27(28):3912-3922.

Bourdi M, Demady D, Martin JL, Jabbour SK, Martin BM, George JW, Pohl LR. cDNA cloning and baculovirus expression of the human liver endoplasmic reticulum P58: characterization as a protein disulfide isomerase isoform, but not as a protease or a carnitine acyltransferase. Arch Biochem Biophys. 1995 Nov 10;323(2):397-403.

Braun JE, Huntzinger E, Fauser M, Izaurralde E. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell. 2011 Oct 7;44(1):120-133.

Bridge WL, Vandenberg CJ, Franklin RJ, Hiom K. The BRIP1 helicase functions independently of BRCA1 in the Fanconi anemia pathway for DNA crosslink repair. Nat Genet. 2005 Sep;37(9):953-957.

Brown AL, Churpek JE, Malcovati L, Döhner H, Godley LA. Recognition of familial myeloid neoplasia in adults. Semin Hematol. 2017 Apr;54(2):60-68.

Buchi F, Masala E, Rossi A, Valencia A, Spinelli E, Sanna A, Gozzini A, Santini V. Redistribution of H3K27me3 and acetylated histone H4 upon exposure to azacitidine and decitabine results in de-repression of the AML1/ETO target gene IL3. Epigenetics. 2014 Mar;9(3):387-395.

Burdette DL, Monroe KM, Sotelo-Troha K, Iwig JS, Eckert B, Hyodo M, Hayakawa Y, Vance RE. STING is a direct innate immune sensor of cyclic di-GMP. Nature. 2011 Sep 25;478(7370):515-518.

Caillet-Saguy C, Toto A, Guerois R, Maisonneuve P, di Silvio E, Sawyer K, Gianni S, Wolff N. Regulation of the Human Phosphatase PTPN4 by the inter-domain linker connecting the PDZ and the phosphatase domains. Sci Rep. 2017 Aug 11;7(1):7875.

Calado RT, Regal JA, Hills M, Yewdell WT, Dalmazzo LF, Zago MA, Lansdorp PM, Hogge D, Chanock SJ, Estey EH, Falcão RP, Young NS. Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. Proc Natl Acad Sci U S A. 2009 Jan 27;106(4):1187-1192.

Calado RT, Cooper JN, Padilla-Nash HM, Sloand EM, Wu CO, Scheinberg P, Ried T, Young NS. Short telomeres result in chromosomal instability in hematopoietic cells and precede malignant evolution in human aplastic anemia. Leukemia. 2012 Apr;26(4):700-707.

Cancer Genome Atlas Research Network, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ Jr, Laird PW, Baty JD, Fulton LL, Fulton R, Heath SE, Kalicki-Veizer J, Kandoth C, Klco JM, Koboldt DC, Kanchi KL, Kulkarni S, Lamprecht TL, Larson DE, Lin L, Lu C, McLellan MD, McMichael JF, Payton J, Schmidt H, Spencer DH, Tomasson MH, Wallis JW, Wartman LD, Watson MA, Welch J, Wendl MC, Ally A, Balasundaram M, Birol I, Butterfield Y, Chiu R, Chu A, Chuah E, Chun HJ, Corbett R, Dhalla N, Guin R, He A, Hirst C, Hirst M, Holt RA, Jones S, Karsan A, Lee D, Li HI, Marra MA, Mayo M, Moore RA, Mungall K, Parker J, Pleasance E, Plettner P, Schein J, Stoll D, Swanson L, Tam A, Thiessen N, Varhol R, Wye N, Zhao Y, Gabriel S, Getz G, Sougnez C, Zou L, Leiserson MD, Vandin F, Wu HT, Applebaum F, Baylin SB, Akbani R, Broom BM, Chen K, Motter TC, Nguyen K, Weinstein JN, Zhang N, Ferguson ML, Adams C, Black A, Bowen J, Gastier-Foster J, Grossman T, Lichtenberg T, Wise L, Davidsen T, Demchok JA, Shaw KR, Sheth M, Sofia HJ, Yang L, Downing JR, Eley G. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013 May 30;368(22):2059-74. Erratum in: N Engl J Med. 2013 Jul 4;369(1):98.

Cardoso SR, Ryan G, Walne AJ, Ellison A, Lowe R, Tummala H, Rio-Machin A, Collopy L, Al Seraihi A, Wallis Y, Page P, Akiki S, Fitzgibbon J, Vulliamy T, Dokal I. Germline heterozygous DDX41 variants in a subset of familial myelodysplasia and acute myeloid leukemia. Leukemia. 2016 Oct;30(10):2083-2086.

Cardoso SR, Ellison ACM, Walne AJ, Cassiman D, Raghavan M, Kishore B, Ancliff P, Rodríguez-Vigil C, Dobbels B, Rio-Machin A, Al Seraihi AFH, Pontikos N, Tummala H, Vulliamy T, Dokal I. Myelodysplasia and liver disease extend the spectrum of RTEL1 related telomeropathies. Haematologica. 2017 Aug;102(8):e293-e296.

Caruthers JM, Johnson ER, McKay DB. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. Proc Natl Acad Sci U S A. 2000 Nov 21;97(24):13080-13085.

Cavalcante de Andrade Silva M, Krepischi ACV, Kulikowski LD, Zanardo EA, Nardinelli L, Leal AM, Costa SS, Muto NH, Rocha V, Velloso EDRP. Deletion of RUNX1 exons 1 and 2 associated with familial platelet disorder with propensity to acute myeloid leukemia. Cancer Genet. 2018 Apr;222-223:32-37.

Cawthon RM. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic Acids Res. 2009; 37(3):e21.

Cazzola M, Rossi M, Malcovati L; Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie Mieloproliferative. Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. Blood. 2013 Jan 10;121(2):260-269.

Cedena MT, Rapado I, Santos-Lozano A, Ayala R, Onecha E, Abaigar M, Such E, Ramos F, Cervera J, Díez-Campelo M, Sanz G, Rivas JH, Lucía A, Martínez-López J. Mutations in the DNA methylation pathway and number of driver mutations predict response to azacitidine in myelodysplastic syndromes. Oncotarget. 2017 Oct 27;8(63):106948-106961.

Cenik C, Chua HN, Zhang H, Tarnawsky SP, Akef A, Derti A, Tasan M, Moore MJ, Palazzo AF, Roth FP. Genome analysis reveals interplay between 5'UTR introns

and nuclear mRNA export for secretory and mitochondrial genes. PLoS Genet. 2011 Apr;7(4):e1001366.

Chao CH, Chen CM, Cheng PL, Shih JW, Tsou AP, Lee YH. DDX3, a DEAD box RNA helicase with tumor growth-suppressive property and transcriptional regulation activity of the p21waf1/cip1 promoter, is a candidate tumor suppressor. Cancer Res. 2006 Jul 1;66(13):6579-6588.

Chapman JR, Taylor MR, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. Mol Cell. 2012 Aug 24;47(4):497-510.

Cheah JJC, Hahn CN, Hiwase DK, Scott HS, Brown AL. Myeloid neoplasms with germline DDX41 mutation. Int J Hematol. 2017 Aug;106(2):163-174.

Chekulaeva M, Mathys H, Zipprich JT, Attig J, Colic M, Parker R, Filipowicz W. miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol. 2011 Oct 7;18(11):1218-1226.

Chen SJ, Shen Y, Chen Z. A panoramic view of acute myeloid leukemia. Nat Genet. 2013 Jun;45(6):586-587.

Child ES, Mann DJ. The intricacies of p21 phosphorylation: protein/protein interactions, subcellular localization and stability. Cell Cycle. 2006 Jun;5(12):1313-1319.

Chiu YH, Macmillan JB, Chen ZJ. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. Cell. 2009 Aug 7;138(3):576-591.

Churpek JE, Pyrtel K, Kanchi KL, Shao J, Koboldt D, Miller CA, Shen D, Fulton R, O'Laughlin M, Fronick C, Pusic I, Uy GL, Braunstein EM, Levis M, Ross J, Elliott K, Heath S, Jiang A, Westervelt P, DiPersio JF, Link DC, Walter MJ, Welch J, Wilson R, Ley TJ, Godley LA, Graubert TA. Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. Blood. 2015 Nov 26;126(22):2484-2490.

Cogan JD, Kropski JA, Zhao M, Mitchell DB, Rives L, Markin C, Garnett ET, Montgomery KH, Mason WR, McKean DF, Powers J, Murphy E, Olson LM, Choi L, Cheng DS, Blue EM, Young LR, Lancaster LH, Steele MP, Brown KK, Schwarz MI, Fingerlin TE, Schwartz DA, Lawson WE, Loyd JE, Zhao Z, Phillips JA 3rd, Blackwell TS. Rare variants in RTEL1 are associated with familial interstitial pneumonia. Am J Respir Crit Care Med. 2015 Mar 15;191(6):646-655.

Collin M, Dickinson R, Bigley V. Haematopoietic and immune defects associated with GATA2 mutation. Br J Haematol. 2015 Apr;169(2):173-187.

Cordin O, Banroques J, Tanner NK, Linder P. The DEAD-box protein family of RNA helicases. Gene. 2006 Feb 15; 367:17-37.

Crispino JD, Weiss MJ. Erythro-megakaryocytic transcription factors associated with hereditary anemia. Blood. 2014 May 15;123(20):3080-3088.

de Lange T. T-loops and the origin of telomeres. Nat Rev Mol Cell Biol. 2004 Apr;5(4):323-9. Review. Erratum in: Nat Rev Mol Cell Biol. 2004 Jun;5(6):492.

de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 2005 Sep 15;19(18):2100-2110.

de Lange T. How telomeres solve the end-protection problem. Science. 2009 Nov 13;326(5955):948-952.

Deng Z, Glousker G, Molczan A, Fox AJ, Lamm N, Dheekollu J, Weizman OE, Schertzer M, Wang Z, Vladimirova O, Schug J, Aker M, Londoño-Vallejo A, Kaestner KH, Lieberman PM, Tzfati Y. Inherited mutations in the helicase RTEL1 cause telomere dysfunction and Hoyeraal-Hreidarsson syndrome. Proc Natl Acad Sci U S A. 2013 Sep 3;110(36):E3408-3416.

Devriendt K, Kim AS, Mathijs G, Frints SG, Schwartz M, Van Den Oord JJ, Verhoef GE, Boogaerts MA, Fryns JP, You D, Rosen MK, Vandenberghe P. Constitutively activating mutation in WASP causes X-linked severe congenital neutropenia. Nat Genet. 2001 Mar;27(3):313-317.

Dhanraj S, Matveev A, Li H, Lauhasurayotin S, Jardine L, Cada M, Zlateska B, Tailor CS, Zhou J, Mendoza-Londono R, Vincent A, Durie PR, Scherer SW, Rommens JM, Heon E, Dror Y. Biallelic mutations in DNAJC21 cause Shwachman-Diamond syndrome. Blood. 2017 Mar 16;129(11):1557-1562.

Dickinson RE, Griffin H, Bigley V, Reynard LN, Hussain R, Haniffa M, Lakey JH, Rahman T, Wang XN, McGovern N, Pagan S, Cookson S, McDonald D, Chua I, Wallis J, Cant A, Wright M, Keavney B, Chinnery PF, Loughlin J, Hambleton S, Santibanez-Koref M, Collin M. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. Blood. 2011 Sep 8;118(10):2656-2658.

Dietz AC, Mehta PA, Vlachos A, Savage SA, Bresters D, Tolar J, Boulad F, Dalle JH, Bonfim C, de la Fuente J, Duncan CN, Baker KS, Pulsipher MA, Lipton JM, Wagner JE, Alter BP. Current Knowledge and Priorities for Future Research in Late Effects after Hematopoietic Cell Transplantation for Inherited Bone Marrow Failure Syndromes: Consensus Statement from the Second Pediatric Blood and Marrow Transplant Consortium International Conference on Late Effects after Pediatric Hematopoietic Cell Transplantation. Biol Blood Marrow Transplant. 2017 May;23(5):726-735.

DiNardo CD, Bannon SA, Routbort M, Franklin A, Mork M, Armanios M, Mace EM, Orange JS, Jeff-Eke M, Churpek JE, Takahashi K, Jorgensen JL, Garcia-Manero G, Kornblau S, Bertuch A, Cheung H, Bhalla K, Futreal A, Godley LA, Patel KP. Evaluation of Patients and Families With Concern for Predispositions to Hematologic Malignancies Within the Hereditary Hematologic Malignancy Clinic (HHMC). Clin Lymphoma Myeloma Leuk. 2016 Jul;16(7):417-428.

Ding H, Schertzer M, Wu X, Gertsenstein M, Selig S, Kammori M, Pourvali R, Poon S, Vulto I, Chavez E, Tam PP, Nagy A, Lansdorp PM. Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. Cell. 2004 Jun 25;117(7):873-886.

Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, Shen D, Harris CC, Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, McGrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, DiPersio JF. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genomesequencing. Nature 2012; 481:506-510.

Dokal I. Fanconi's anaemia and related bone marrow failure syndromes. Br Med Bull. 2006;77-78:37-53.

Dokal I, Vulliamy T. Inherited aplastic anaemias/bone marrow failure syndromes. Blood Rev. 2008 May;22(3):141-153.

Dokal I, Vulliamy T. Inherited bone marrow failure syndromes. Haematologica. 2010 Aug;95(8):1236-1240.

Dokal I. Dyskeratosis congenita. Hematology Am Soc Hematol Educ Program. 2011; 2011:480-486.

Dokal I, Vulliamy T, Mason P, Bessler M. Clinical utility gene card for: Dyskeratosis congenita - update 2015. Eur J Hum Genet. 2015 Apr;23(4).

Doksani Y, Wu JY, de Lange T, Zhuang X. Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. Cell. 2013 Oct 10;155(2):345-356.

Duployez N, Abou Chahla W, Lejeune S, Marceau-Renaut A, Letizia G, Boyer T, Geffroy S, Peyrouze P, Grardel N, Nelken B, Michel G, Bertrand Y, Preudhomme C. Detection of a new heterozygous germline ETV6 mutation in a case with hyperdiploid acute lymphoblastic leukemia. Eur J Haematol. 2018 Jan;100(1):104-107.

Fabian MR, Cieplak MK, Frank F, Morita M, Green J, Srikumar T, Nagar B, Yamamoto T, Raught B, Duchaine TF, Sonenberg N. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nat Struct Mol Biol. 2011 Oct 7;18(11):1211-1217.

Faivre L, Guardiola P, Lewis C, Dokal I, Ebell W, Zatterale A, Altay C, Poole J, Stones D, Kwee ML, van Weel-Sipman M, Havenga C, Morgan N, de Winter J, Digweed M, Savoia A, Pronk J, de Ravel T, Jansen S, Joenje H, Gluckman E, Mathew CG. Association of complementation group and mutation type with clinical outcome in fanconi anemia. European Fanconi Anemia Research Group. Blood. 2000 Dec 15;96(13):4064-4070.

Fang G, Zhang D, Yin H, Zheng L, Bi X, Yuan L. Centlein mediates an interaction between C-Nap1 and Cep68 to maintain centrosome cohesion. J Cell Sci. 2014 Apr 15;127(Pt 8):1631-1639.

Fang J, Kubota S, Yang B, Zhou N, Zhang H, Godbout R, Pomerantz RJ. A DEAD box protein facilitates HIV-1 replication as a cellular co-factor of Rev. Virology. 2004 Dec 20;330(2):471-480.

Fargo JH, Rochowski A, Giri N, Savage SA, Olson SB, Alter BP. Comparison of chromosome breakage in non-mosaic and mosaic patients with Fanconi anemia, relatives, and patients with other inherited bone marrow failure syndromes. Cytogenet Genome Res. 2014;144(1):15-27.

Faria JA, Corrêa NC, de Andrade C, de Angelis Campos AC, Dos Santos Samuel de Almeida R, Rodrigues TS, de Goes AM, Gomes DA, Silva FP. SET domain-containing Protein 4 (SETD4) is a Newly Identified Cytosolic and Nuclear Lysine Methyltransferase involved in Breast Cancer Cell Proliferation. J Cancer Sci Ther. 2013 Jan 21;5(2):58-65.

Fasan A, Haferlach C, Alpermann T, Jeromin S, Grossmann V, Eder C, Weissmann S, Dicker F, Kohlmann A, Schindela S, Kern W, Haferlach T, Schnittger S. The role of different genetic subtypes of CEBPA mutated AML. Leukemia. 2014 Apr;28(4):794-803.

Fenaux P, Mufti GJ, Hellstrom-Lindberg E, Santini V, Finelli C, Giagounidis A, Schoch R, Gattermann N, Sanz G, List A, Gore SD, Seymour JF, Bennett JM, Byrd J, Backstrom J, Zimmerman L, McKenzie D, Beach C, Silverman LR; International Vidaza High-Risk MDS Survival Study Group. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. Lancet Oncol. 2009 Mar;10(3):223-232.

Ferrari S, Lombardi AM, Putti MC, Bertomoro A, Cortella I, Barzon I, Girolami A, Fabris F. Spectrum of 5'UTR mutations in ANKRD26 gene in patients with inherited thrombocytopenia: c.-140C>G mutation is more frequent than expected. Platelets. 2017 Sep;28(6):621-624.

Fielding AK, Rowe JM, Buck G, Foroni L, Gerrard G, Litzow MR, Lazarus H, Luger SM, Marks DI, McMillan AK, Moorman AV, Patel B, Paietta E, Tallman MS, Goldstone AH. UKALLXII/ECOG2993: addition of imatinib to a standard treatment regimen enhances long-term outcomes in Philadelphia positive acute lymphoblastic leukemia. Blood. 2014 Feb 6;123(6):843-850.

Finch AJ, Hilcenko C, Basse N, Drynan LF, Goyenechea B, Menne TF, González Fernández A, Simpson P, D'Santos CS, Arends MJ, Donadieu J, Bellanné-Chantelot C, Costanzo M, Boone C, McKenzie AN, Freund SM, Warren AJ. Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes Shwachman-Diamond syndrome. Genes Dev. 2011 May 1;25(9):917-929.

Fisher KE, Hsu AP, Williams CL, Sayeed H, Merritt BY, Elghetany MT, Holland SM, Bertuch AA, Gramatges MM. Somatic mutations in children with GATA2-associated myelodysplastic syndrome who lack other features of GATA2 deficiency. Blood Adv. 2017 Feb 27;1(7):443-448.

Fitzgibbon J, Smith LL, Raghavan M, Smith ML, Debernardi S, Skoulakis S, Lillington D, Lister TA, Young BD. Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. Cancer Res. 2005 Oct 15;65(20):9152-9154.

Flynn EK, Kamat A, Lach FP, Donovan FX, Kimble DC, Narisu N, Sanborn E, Boulad F, Davies SM, Gillio AP 3rd, Harris RE, MacMillan ML, Wagner JE,

Smogorzewska A, Auerbach AD, Ostrander EA, Chandrasekharappa SC. Comprehensive analysis of pathogenic deletion variants in Fanconi anemia genes. Hum Mutat. 2014 Nov;35(11):1342-1353.

Fukuhara A, Tanino Y, Ishii T, Inokoshi Y, Saito K, Fukuhara N, Sato S, Saito J, Ishida T, Yamaguchi H, Munakata M. Pulmonary fibrosis in dyskeratosis congenita with TINF2 gene mutation. Eur Respir J. 2013 Dec;42(6):1757-1759.

Fullam A, Schröder M. DExD/H-box RNA helicases as mediators of anti-viral innate immunity and essential host factors for viral replication. Biochim Biophys Acta. 2013 Aug;1829(8):854-865.

Fuller-Pace FV. DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic Acids Res. 2006;34(15):4206-15. Epub 2006 Aug 25.

Fuller-Pace FV. DEAD box RNA helicase functions in cancer. RNA Biol. 2013 Jan;10(1):121-132.

Gaidzik VI, Teleanu V, Papaemmanuil E, Weber D, Paschka P, Hahn J, Wallrabenstein T, Kolbinger B, Köhne CH, Horst HA, Brossart P, Held G, Kündgen A, Ringhoffer M, Götze K, Rummel M, Gerstung M, Campbell P, Kraus JM, Kestler HA, Thol F, Heuser M, Schlegelberger B, Ganser A, Bullinger L, Schlenk RF, Döhner K, Döhner H. RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features. Leukemia. 2016 Nov;30(11):2160-2168.

Gao J, Gentzler RD, Andrew ET, Horwitz M S, Frankfurt O, Altman JK & Peterson LC. Heritable GATA2 mutations associated with familial AML-MDS: a case report and review of literature. Journal of Hematology & Oncology 2014; 7, 1-7.

Gao J, Gentzler RD, Timms AE, Horwitz MS, Frankfurt O, Altman JK, Peterson LC.Heritable GATA2 mutations associated with familial AML-MDS: a case report and review of literature. J Hematol Oncol. 2014 Apr 22; 7:36. Erratum in: J Hematol Oncol. 2015; 8:131.

Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, Purcell SM, Svantesson O, Landén M, Höglund M, Lehmann S, Gabriel SB, Moran JL, Lander ES, Sullivan PF, Sklar P, Grönberg H, Hultman CM, McCarroll SA. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014 Dec 25;371(26):2477-2487.

Germann S, Gratadou L, Zonta E, Dardenne E, Gaudineau B, Fougère M, Samaan S, Dutertre M, Jauliac S, Auboeuf D. Dual role of the ddx5/ddx17 RNA helicases in the control of the pro-migratory NFAT5 transcription factor. Oncogene. 2012 Oct 18;31(42):4536-4549.

Germing U, Aul C, Niemeyer CM, Haas R, Bennett JM. Epidemiology, classification and prognosis of adults and children with myelodysplastic syndromes. Annals of Hematology 2008; 87, 691–699.

Ghemlas I, Li H, Zlateska B, Klaassen R, Fernandez CV, Yanofsky RA, Wu J, Pastore Y, Silva M, Lipton JH, Brossard J, Michon B, Abish S, Steele M, Sinha R, Belletrutti M, Breakey VR, Jardine L, Goodyear L, Sung L, Dhanraj S, Reble E, Wagner A, Beyene J, Ray P, Meyn S, Cada M, Dror Y. Improving diagnostic precision, care and syndrome definitions using comprehensive next-generation sequencing for the inherited bone marrow failure syndromes. J Med Genet. 2015 Sep;52(9):575-584.

Gocke CB, Yu H. ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. PLoS One. 2008 Sep 22;3(9):e3255.

Godley LA. Inherited predisposition to acute myeloid leukemia. Semin Hematol. 2014 Oct;51(4):306-321.

Gogol-Döring A & Chen W. An overview of the analysis of next generation sequencing data. Methods in Mol Biol 2012; 802, 249–257.

Goh PY, Tan YJ, Lim SP, Tan YH, Lim SG, Fuller-Pace F, Hong W. Cellular RNA helicase p68 relocalization and interaction with the hepatitis C virus (HCV) NS5B protein and the potential role of p68 in HCV RNA replication. J Virol. 2004 May;78(10):5288-5298.

Gomez DE, Armando RG, Farina HG, Menna PL, Cerrudo CS, Ghiringhelli PD, Alonso DF. Telomere structure and telomerase in health and disease (review). Int J Oncol. 2012 Nov;41(5):1561-1569.

Grada A & Weinbrecht K. Next-generation sequencing: methodology and application. J Invest Dermatol. 2013; 133**,** e11.

Graser S, Stierhof YD, Nigg EA. Cep68 and Cep215 (Cdk5rap2) are required for centrosome cohesion. J Cell Sci. 2007 Dec 15;120(Pt 24):4321-4331. Epub 2007 Nov 27.

Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. Proc Natl Acad Sci U S A. 2003 Jul 22;100(15):8811-8816.

Greaves MF. Aetiology of acute leukaemia. The Lancet 1997; 349, 344–349.

Green CL, Tawana K, Hills RK, Bödör C, Fitzgibbon J, Inglott S, Ancliff P, Burnett AK, Linch DC, Gale RE. GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations. Br J Haematol. 2013 Jun;161(5):701-705.

Greider CW and Blackburn EH. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell. 1985 Dec;43(2 Pt 1):405-413.

Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, de Lange T. Mammalian telomeres end in a large duplex loop. Cell. 1999 May 14;97(4):503-514.

Grimwade D, Mrózek K. Diagnostic and prognostic value of cytogenetics in acute myeloid leukemia. Hematol Oncol Clin North Am. 2011 Dec;25(6):1135-1161.

Grimwade D, Ivey A, Huntly BJ. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. Blood. 2016 Jan 7;127(1):29-41.

Guison J, Blaison G, Stoica O, Hurstel R, Favier M, Favier R. Idiopathic Pulmonary Embolism in a case of Severe Family ANKRD26 Thrombocytopenia. Mediterr J Hematol Infect Dis. 2017 Jun 16;9(1):e2017038.

Haase D, Germing U, Schanz J, Pfeilstöcker M, Nösslinger T, Hildebrandt B, Kundgen A, Lübbert M, Kunzmann R, Giagounidis AA, Aul C, Trümper L, Krieger O, Stauder R, Müller TH, Wimazal F, Valent P, Fonatsch C, Steidl C. New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. Blood. 2007 Dec 15;110(13):4385-4395.

Hackett JA, Greider CW. End resection initiates genomic instability in the absence of telomerase. Mol Cell Biol. 2003 Dec;23(23):8450-8461.

Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, Schnittger S, Sanada M, Kon A, Alpermann T, Yoshida K, Roller A, Nadarajah N, Shiraishi Y, Shiozawa Y, Chiba K, Tanaka H, Koeffler HP, Klein HU, Dugas M, Aburatani H, Kohlmann A, Miyano S, Haferlach C, Kern W, Ogawa S. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014 Feb;28(2):241-247.

Hahn C N, Chong C-E, Carmichael CL, Wilkins EJ, Brautigan PJ, Li X-C, Babic M, Lin M, Carmagnac A, Lee YK, Kok CH, Gagliardi L, Friend KL, Ekert PG, Butcher CM, Brown AL, Lewis ID, To LB, Timms AE, Storek J, Moore S, Altree M, Escher R, Bardy PG, Suthers G K, D'Andrea RJ, Horwitz MS & Scott HS. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukaemia. Nature Genetics 2011; 43, 1012-1019.

Hamadou WS, Bourdon V, Gaildrat P, Besbes S, Fabre A, Youssef YB, Regaieg H, Laatiri MA, Eisinger F, Mari V, Gesta P, Dreyfus H, Bonadona V, Dugast C, Zattara H, Faivre L, Jemni SY, Noguchi T, Khélif A, Sobol H, Soua Z. Mutational analysis of JAK2, CBL, RUNX1, and NPM1 genes in familial aggregation of haematological malignancies. Ann Hematol. 2016 Jun;95(7):1043-1050.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar 4;144(5):646-674.

Handel MA, Schimenti JC. Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. Nat Rev Genet. 2010 Feb;11(2):124-136.

Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A, Dokal I. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat Genet. 1998 May;19(1):32-38.

Hengge R. Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol. 2009 Apr;7(4):263-273.

Hirabayashi S, Wlodarski MW, Kozyra E, Niemeyer CM. Heterogeneity of GATA2-related myeloid neoplasms. Int J Hematol. 2017 Aug;106(2):175-182.

Hirsch CM, Przychodzen BP, Radivoyevitch T, Patel B, Thota S, Clemente MJ, Nagata Y, LaFramboise T, Carraway HE, Nazha A, Sekeres MA, Makishima H, Maciejewski JP. Molecular features of early onset adult myelodysplastic syndrome. Haematologica. 2017 Jun;102(6):1028-1034.

Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, Kadel S, Moll I, Nagore E, Hemminki K, Schadendorf D, Kumar R. TERT promoter mutations in familial and sporadic melanoma. Science. 2013 Feb 22;339(6122):959-961.

Hou HA, Lin CC, Chou WC, Liu CY, Chen CY, Tang JL, Lai YJ, Tseng MH, Huang CF, Chiang YC, Lee FY, Kuo YY, Lee MC, Liu MC, Liu CW, Lin LI, Yao M, Huang SY, Ko BS, Hsu SC, Wu SJ, Tsay W, Chen YC, Tien HF. Integration of cytogenetic and molecular alterations in risk stratification of 318 patients with de novo non-M3 acute myeloid leukemia. Leukemia. 2014 Jan;28(1):50-58.

Hsu AP, Sampaio EP, Khan J, Calvo KR, Lemieux JE, Patel SY, Frucht DM, Vinh DC, Auth RD, Freeman AF, Olivier KN, Uzel G, Zerbe CS, Spalding C, Pittaluga S, Raffeld M, Kuhns DB, Ding L, Paulson ML, Marciano BE, Gea-Banacloche JC, Orange JS, Cuellar-Rodriguez J, Hickstein DD, Holland SM. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. Blood. 2011 Sep 8;118(10):2653-2655.

Huang DS, Wang Z, He XJ, Diplas BH, Yang R, Killela PJ, Meng Q, Ye ZY, Wang W, Jiang XT, Xu L, He XL, Zhao ZS, Xu WJ, Wang HJ, Ma YY, Xia YJ, Li L, Zhang RX, Jin T, Zhao ZK, Xu J, Yu S, Wu F, Liang J, Wang S, Jiao Y, Yan H, Tao HQ. Recurrent TERT promoter mutations identified in a large-scale study of multiple tumour types are associated with increased TERT expression and telomerase activation. Eur J Cancer. 2015 May;51(8):969-976.

Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. Science. 2013 Feb 22;339(6122):957-959.

Huang G, Shigesada K, Ito K, Wee HJ, Yokomizo T, Ito Y. Dimerization with PEBP2beta protects RUNX1/AML1 from ubiquitin-proteasome-mediated degradation. EMBO J. 2001 Feb 15;20(4):723-733.

Hyde RK, Liu PP. GATA2 mutations lead to MDS and AML. Nat Genet. 2011 Sep 28;43(10):926-927.

Ishikawa H, Barber GN. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. Nature. 2008 Oct 2;455(7213):674-8. Erratum in: Nature. 2008 Nov 13;456(7219):274.

Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, Hutchinson A, Deng X, Liu C, Horner MJ, Cullen M, Epstein CG, Burdett L, Dean MC, Chatterjee N, Sampson J, Chung CC, Kovaks J, Gapstur SM, Stevens VL, Teras LT, Gaudet MM, Albanes D, Weinstein SJ, Virtamo J, Taylor PR, Freedman ND, Abnet CC, Goldstein AM, Hu N, Yu K, Yuan JM, Liao L, Ding T, Qiao YL, Gao YT, Koh WP, Xiang YB, Tang ZZ, Fan JH, Aldrich MC, Amos C, Blot WJ, Bock CH, Gillanders EM, Harris CC, Haiman CA, Henderson BE, Kolonel LN, Le Marchand L, McNeill LH, Rybicki BA, Schwartz AG, Signorello LB, Spitz MR, Wiencke JK,

Wrensch M, Wu X, Zanetti KA, Ziegler RG, Figueroa JD, Garcia-Closas M, Malats N, Marenne G, Prokunina-Olsson L, Baris D, Schwenn M, Johnson A, Landi MT, Goldin L, Consonni D, Bertazzi PA, Rotunno M, Rajaraman P, Andersson U, Beane Freeman LE, Berg CD, Buring JE, Butler MA, Carreon T, Feychting M, Ahlbom A, Gaziano JM, Giles GG, Hallmans G, Hankinson SE, Hartge P, Henriksson R, Inskip PD, Johansen C, Landgren A, McKean-Cowdin R, Michaud DS, Melin BS, Peters U, Ruder AM, Sesso HD, Severi G, Shu XO, Visvanathan K, White E, Wolk A, Zeleniuch-Jacquotte A, Zheng W, Silverman DT, Kogevinas M, Gonzalez JR, Villa O, Li D, Duell EJ, Risch HA, Olson SH, Kooperberg C, Wolpin BM, Jiao L, Hassan M, Wheeler W, Arslan AA, Bueno-de-Mesquita HB, Fuchs CS, Gallinger S, Gross MD, Holly EA, Klein AP, LaCroix A, Mandelson MT, Petersen G, Boutron-Ruault MC, Bracci PM, Canzian F, Chang K, Cotterchio M, Giovannucci EL, Goggins M, Hoffman Bolton JA, Jenab M, Khaw KT, Krogh V, Kurtz RC, McWilliams RR, Mendelsohn JB, Rabe KG, Riboli E, Tjønneland A, Tobias GS, Trichopoulos D, Elena JW, Yu H, Amundadottir L, Stolzenberg-Solomon RZ, Kraft P, Schumacher F, Stram D, Savage SA, Mirabello L, Andrulis IL, Wunder JS, Patiño García A, Sierrasesúmaga L, Barkauskas DA, Gorlick RG, Purdue M, Chow WH, Moore LE, Schwartz KL, Davis FG, Hsing AW, Berndt SI, Black A, Wentzensen N, Brinton LA, Lissowska J, Peplonska B, McGlynn KA, Cook MB, Graubard BI, Kratz CP, Greene MH, Erickson RL, Hunter DJ, Thomas G, Hoover RN, Real FX, Fraumeni JF Jr, Caporaso NE, Tucker M, Rothman N, Pérez-Jurado LA, Chanock SJ. Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. 2012 May 6;44(6):651-658.

Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burtt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen HA, Ladenvall C, Getz G, Correa A, Banahan BF, Gabriel S, Kathiresan S, Stringham HM, McCarthy MI, Boehnke M, Tuomilehto J, Haiman C, Groop L, Atzmon G, Wilson JG, Neuberg D, Altshuler D, Ebert BL. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014 Dec 25;371(26):2488-2498.

Jankowsky E. RNA helicases at work: binding and rearranging. Trends in Biochemical Sciences 2011; 36:19-29.

Jenkins JL, Kielkopf CL. Splicing Factor Mutations in Myelodysplasias: Insights from Spliceosome Structures. Trends Genet. 2017 May;33(5):336-348.

Jiang X, Liu C, Yu T, Zhang L, Meng K, Xing Z, Belichenko PV, Kleschevnikov AM, Pao A, Peresie J, Wie S, Mobley WC, Yu YE. Genetic dissection of the Down syndrome critical region. Hum Mol Genet. 2015 Nov 15;24(22):6540-6551.

Jiang Y, Zhu Y, Liu ZJ, Ouyang S. The emerging roles of the DDX41 protein in immunity and diseases. Protein Cell. 2017 Feb;8(2):83-89.

Jin Y, Mazza C, Christie JR, Giliani S, Fiorini M, Mella P, Gandellini F, Stewart DM, Zhu Q, Nelson DL, Notarangelo LD, Ochs HD. Mutations of the Wiskott-Aldrich Syndrome Protein (WASP): hotspots, effect on transcription, and translation and phenotype/genotype correlation. Blood. 2004 Dec 15;104(13):4010-4019.

Jing Z, Liu Y, Dong M, Hu S, Huang S. Identification of the DNA binding element of the human ZNF333 protein. J Biochem Mol Biol. 2004 Nov 30;37(6):663-670.

Jongmans MC, Kuiper RP, Carmichael CL, Wilkins EJ, Dors N, Carmagnac A, Schouten-van Meeteren AY, Li X, Stankovic M, Kamping E, Bengtsson H, Schoenmakers EF, van Kessel AG, Hoogerbrugge PM, Hahn CN, Brons PP, Scott HS, Hoogerbrugge N. Novel RUNX1 mutations in familial platelet disorder with enhanced risk for acute myeloid leukemia: clues for improved identification of the FPD/AML syndrome. Leukemia. 2010 Jan;24(1):242-246.

Jongmans MC, Verwiel ET, Heijdra Y, Vulliamy T, Kamping EJ, Hehir-Kwa JY, Bongers EM, Pfundt R, van Emst L, van Leeuwen FN, van Gassen KL, Geurts van Kessel A, Dokal I, Hoogerbrugge N, Ligtenberg MJ, Kuiper RP. Revertant somatic mosaicism by mitotic recombination in dyskeratosis congenita. Am J Hum Genet. 2012 Mar 9;90(3):426-433.

Jullien L, Kannengiesser C, Kermasson L, Cormier-Daire V, Leblanc T, Soulier J, Londono-Vallejo A, de Villartay JP, Callebaut I, Revy P. Mutations of the RTEL1 Helicase in a Hoyeraal-Hreidarsson Syndrome Patient Highlight the Importance of the ARCH Domain. Hum Mutat. 2016 May;37(5):469-472.

Jung YS, Qian Y, Chen X. Examination of the expanding pathways for the regulation of p21 expression and activity. Cell Signal. 2010 Jul;22(7):1003-1012.

Kadono M, Kanai A, Nagamachi A, Shinriki S, Kawata J, Iwato K, Kyo T, Oshima K, Yokoyama A, Kawamura T, Nagase R, Inoue D, Kitamura T, Inaba T, Ichinohe T, Matsui H. Biological implications of somatic DDX41 p.R525H mutation in acute myeloid leukemia. Exp Hematol. 2016 Aug;44(8):745-54.e4.

Kanagal-Shamanna R, Loghavi S, DiNardo CD, Medeiros LJ, Garcia-Manero G, Jabbour E, Routbort MJ, Luthra R, Bueso-Ramos CE, Khoury JD. Bone marrow pathologic abnormalities in familial platelet disorder with propensity for myeloid malignancy and germline RUNX1 mutation. Haematologica. 2017 Oct;102(10):1661-1670.

Kannengiesser C, Borie R, Ménard C, Réocreux M, Nitschké P, Gazal S, Mal H, Taillé C, Cadranel J, Nunes H, Valeyre D, Cordier JF, Callebaut I, Boileau C, Cottin V, Grandchamp B, Revy P, Crestani B. Heterozygous RTEL1 mutations are associated with familial pulmonary fibrosis. Eur Respir J. 2015 Aug;46(2):474-485.

Kasamatsu H, Robberson DL, Vinograd J. A novel closed-circular mitochondrial DNA with properties of a replicating intermediate. Proc Natl Acad Sci U S A. 1971 Sep;68(9):2252-2257.

Kawai T, Sato S, Ishii KJ, Coban C, Hemmi H, Yamamoto M, Terai K, Matsuda M, Inoue J, Uematsu S, Takeuchi O, Akira S. Interferon-alpha induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. Nat Immunol. 2004 Oct;5(10):1061-1068.

Kazenwadel J, Secker GA, Liu YJ, Rosenfeld JA, Wildin RS, Cuellar-Rodriguez J, Hsu AP, Dyack S, Fernandez CV, Chong CE, Babic M, Bardy PG, Shimamura A, Zhang MY, Walsh T, Holland SM, Hickstein DD, Horwitz MS, Hahn CN, Scott HS, Harvey NL. Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature. Blood. 2012 Feb 2;119(5):1283-1291.

Keller RB, Gagne KE, Usmani GN, Asdourian GK, Williams DA, Hofmann I, Agarwal S. CTC1 Mutations in a patient with dyskeratosis congenita. Pediatr Blood Cancer. 2012 Aug;59(2):311-314.

Khwaja A, Bjorkholm M, Gale RE, Levine RL, Jordan CT, Ehninger G, Bloomfield CD, Estey E, Burnett A, Cornelissen JJ, Scheinberg DA, Bouscary D, Linch DC. Acute myeloid leukaemia. Nat Rev Dis Primers. 2016 Mar 10;2:16010.

Killela PJ, Reitman ZJ, Jiao Y, Bettegowda C, Agrawal N, Diaz LA Jr, Friedman AH, Friedman H, Gallia GL, Giovanella BC, Grollman AP, He TC, He Y, Hruban RH, Jallo GI, Mandahl N, Meeker AK, Mertens F, Netto GJ, Rasheed BA, Riggins GJ, Rosenquist TA, Schiffman M, Shih IeM, Theodorescu D, Torbenson MS, Velculescu VE, Wang TL, Wentzensen N, Wood LD, Zhang M, McLendon RE, Bigner DD, Kinzler KW, Vogelstein B, Papadopoulos N, Yan H. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. Proc Natl Acad Sci U S A. 2013 Apr 9;110(15):6021-6026.

Kim T, Pazhoor S, Bao M, Zhang Z, Hanabuchi S, Facchinetti V, Bover L, Plumas J, Chaperot L, Qin J, Liu YJ. Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells. Proc Natl Acad Sci U S A. 2010 Aug 24;107(34):15181-15186.

Kimble DC, Lach FP, Gregg SQ, Donovan FX, Flynn EK, Kamat A, Young A, Vemulapalli M, Thomas JW, Mullikin JC, Auerbach AD, Smogorzewska A, Chandrasekharappa SC. A comprehensive approach to identification of pathogenic FANCA variants in Fanconi anemia patients and their families. Hum Mutat. 2018 Feb;39(2):237-254.

King TE Jr, Pardo A, Selman M. Idiopathic pulmonary fibrosis. Lancet. 2011 Dec 3;378(9807):1949-1961.

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.

Kirstetter P, Schuster MB, Bereshchenko O, Moore S, Dvinge H, Kurz E, Theilgaard-Mönch K, Månsson R, Pedersen TA, Pabst T, Schrock E, Porse BT, Jacobsen SE, Bertone P, Tenen DG, Nerlov C. Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. Cancer Cell. 2008 Apr;13(4):299-310.

Kirwan M, Vulliamy T, Marrone A, Walne A J, Beswick R, Hillmen P, Kelly R, Stewart A, Bowen D, Schonland SO, Whittle AM, McVerry A, Gilleece M & Dokal I. Defining the pathogenic role of telomerase mutations in myelodysplastic syndrome and acute myeloid leukaemia. Human Mutation 2009; 30, 1567–1573.

Kirwan M, Walne A J, Plagnol V, Velangi M, Ho A, Hossain U, Vulliamy T & Dokal I. Exome sequencing identifies autosomal-dominant SRP72 mutations associated with familial aplasia and myelodysplasia. The American Journal of Human Genetics 2012; 90, 888–892.

Kittler R, Pelletier L, Heninger AK, Slabicki M, Theis M, Miroslaw L, Poser I, Lawo S, Grabner H, Kozak K, Wagner J, Surendranath V, Richter C, Bowen W, Jackson AL, Habermann B, Hyman AA, Buchholz F. Genome-scale RNAi profiling of cell division in human tissue culture cells. Nat Cell Biol. 2007 Dec;9(12):1401-1412.

Klco JM, Miller CA, Griffith M, Petti A, Spencer DH, Ketkar-Kulkarni S, Wartman LD, Christopher M, Lamprecht TL, Helton NM, Duncavage EJ, Payton JE, Baty J, Heath SE, Griffith OL, Shen D, Hundal J, Chang GS, Fulton R, O'Laughlin M, Fronick C, Magrini V, Demeter RT, Larson DE, Kulkarni S, Ozenberger BA, Welch JS, Walter MJ, Graubert TA, Westervelt P, Radich JP, Link DC, Mardis ER, DiPersio JF, Wilson RK, Ley TJ. Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. JAMA. 2015 Aug 25;314(8):811-822.

Kobayashi M, Yokoyama K, Shimizu E, Yusa N, Ito M, Yamaguchi R, Imoto S, Miyano S, Tojo A. Phenotype-based gene analysis allowed successful diagnosis of X-linked neutropenia associated with a novel WASp mutation. Ann Hematol. 2018 Feb;97(2):367-369.

Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. Genome Med. 2017 Feb 6;9(1):13.

Kropski JA, Mitchell DB, Markin C, Polosukhin VV, Choi L, Johnson JE, Lawson WE, Phillips JA 3rd, Cogan JD, Blackwell TS, Loyd JE. A novel dyskerin (DKC1) mutation is associated with familial interstitial pneumonia. Chest. 2014 Jul;146(1):e1-e7.

Krug A, Rothenfusser S, Hornung V, Jahrsdörfer B, Blackwell S, Ballas ZK, Endres S, Krieg AM, Hartmann G. Identification of CpG oligonucleotide sequences with high induction of IFN-alpha/beta in plasmacytoid dendritic cells. Eur J Immunol. 2001 Jul;31(7):2154-2163.

Kupsa T, Horacek JM, Jebavy L. The role of cytokines in acute myeloid leukaemia: a systematic review. Biomedical Papers of the Medical Faculty of the University Palacky Olomouc Czech Repub. 2012; 156, 291–301.

Kurtovic-Kozaric A, Przychodzen B, Singh J, Konarska MM, Clemente MJ, Otrock ZK, Nakashima M, Hsi ED, Yoshida K, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Boultwood J, Makishima H, Maciejewski JP, Padgett RA. PRPF8 defects cause missplicing in myeloid malignancies. Leukemia. 2015 Jan;29(1):126-136.

Kwok B, Reddy P, Lin K, Flamholz R, Yung A, Dabbas B, McGinniss M, Nahas S, Kines J, Xu Y. Next-generation sequencing (NGS)-based profiling of idiopathic cytopenia of undetermined significance (ICUS) identifies a subset of patients with genomic similarities to lower-risk myelodysplastic syndrome (MDS) [abstract]. Blood. 2014;124(21). Abstract 166.

Lai-Cheong JE, McGrath JA. Next-generation diagnostics for inherited skin disorders. J Invest Dermatol. 2011; 131, 971–973.

Langabeer SE, Owen CJ, McCarron SL, Fitzgibbon J, Smith OP, O'Marcaigh A, Browne P. A novel RUNX1 mutation in a kindred with familial platelet disorder with

propensity to acute myeloid leukaemia: male predominance of affected individuals. Eur J Haematol. 2010 Dec;85(6):552-553.

Lansdorp PM. Major cutbacks at chromosome ends. Trends Biochem Sci. 2005 Jul;30(7):388-395.

Latger-Cannard V, Philippe C, Bouquet A, Baccini V, Alessi MC, Ankri A, Bauters A, Bayart S, Cornillet-Lefebvre P, Daliphard S, Mozziconacci MJ, Renneville A, Ballerini P, Leverger G, Sobol H, Jonveaux P, Preudhomme C, Nurden P, Lecompte T, Favier R. Haematological spectrum and genotype-phenotype correlations in nine unrelated families with RUNX1 mutations from the French network on inherited platelet disorders. Orphanet J Rare Dis. 2016 Apr 26;11:49.

Lawrence P, Rieder E. Identification of RNA helicase A as a new host factor in the replication cycle of foot-and-mouth disease virus. J Virol. 2009 Nov;83(21):11356-11366.

Le Guen T, Jullien L, Touzot F, Schertzer M, Gaillard L, Perderiset M, Carpentier W, Nitschke P, Picard C, Couillault G, Soulier J, Fischer A, Callebaut I, Jabado N, Londono-Vallejo A, de Villartay JP, Revy P. Human RTEL1 deficiency causes Hoyeraal-Hreidarsson syndrome with short telomeres and genome instability. Hum Mol Genet. 2013 Aug 15;22(16):3239-3249.

Lee KG, Kim SS, Kui L, Voon DC, Mauduit M, Bist P, Bi X, Pereira NA, Liu C, Sukumaran B, Rénia L, Ito Y, Lam KP. Bruton's tyrosine kinase phosphorylates DDX41 and activates its binding of dsDNA and STING to initiate type 1 interferon response. Cell Rep. 2015 Feb 24;10(7):1055-1065.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 Aug 18;536(7616):285-291.

Lesage S, Drouet V, Majounie E, Deramecourt V, Jacoupy M, Nicolas A, Cormier-Dequaire F, Hassoun SM, Pujol C, Ciura S, Erpapazoglou Z, Usenko T, Maurage CA, Sahbatou M, Liebau S, Ding J, Bilgic B, Emre M, Erginel-Unaltuna N, Guven G, Tison F, Tranchant C, Vidailhet M, Corvol JC, Krack P, Leutenegger AL, Nalls MA, Hernandez DG, Heutink P, Gibbs JR, Hardy J, Wood NW, Gasser T, Durr A, Deleuze JF, Tazir M, Destée A, Lohmann E, Kabashi E, Singleton A, Corti O, Brice A; French Parkinson's Disease Genetics Study (PDG); International Parkinson's Disease Genomics Consortium (IPDGC). Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. Am J Hum Genet. 2016 Mar 3;98(3):500-513.

Lewinsohn M, Brown AL, Weinel LM, Phung C, Rafidi G, Lee MK, Schreiber AW, Feng J, Babic M, Chong CE, Lee Y, Yong A, Suthers GK, Poplawski N, Altree M, Phillips K, Jaensch L, Fine M, D'Andrea RJ, Lewis ID, Medeiros BC, Pollyea DA, King MC, Walsh T, Keel S, Shimamura A, Godley LA, Hahn CN, Churpek JE, Scott HS. Novel germ line DDX41 mutations define families with a lower age of MDS/AML onset and lymphoid malignancies. Blood. 2016 Feb 25;127(8):1017-1023.

Li J, Tang H, Mullen TM, Westberg C, Reddy TR, Rose DW, Wong-Staal F. A role for RNA helicase A in post-transcriptional regulation of HIV type 1. Proc Natl Acad Sci U S A. 1999 Jan 19;96(2):709-714.

Li R, Sobreira N, Witmer PD, Pratz KW, Braunstein EM. Two novel germline DDX41 mutations in a family with inherited myelodysplasia/acute myeloid leukemia. Haematologica. 2016 Jun;101(6):e228-31.

Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, Patel J, Dillon R, Vijay P, Brown AL, Perl AE, Cannon J, Bullinger L, Luger S, Becker M, Lewis ID, To LB, Delwel R, Löwenberg B, Döhner H, Döhner K, Guzman ML, Hassane DC, Roboz GJ, Grimwade D, Valk PJ, D'Andrea RJ, Carroll M, Park CY, Neuberg D, Levine R, Melnick AM, Mason CE. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. Nat Med. 2016 Jul;22(7):792-799.

Liew E & Owen C. Familial myelodysplastic syndromes: a review of the literature. Haematologica 2011; 96, 1536–1542.

Lin L, Li Y, Pyo HM, Lu X, Raman SN, Liu Q, Brown EG, Zhou Y. Identification of RNA helicase A as a cellular factor that interacts with influenza A virus NS1 protein and its role in the virus life cycle. J Virol. 2012 Feb;86(4):1942-1954.

Linder P, Jankowsky E. From unwinding to clamping - the DEAD box RNA helicase family. Nat Rev Mol Cell Biol. 2011 Jul 22;12(8):505-516.

Lindsley RC, Saber W, Mar BG, Redd R, Wang T, Haagenson MD, Grauman PV, Hu ZH, Spellman SR, Lee SJ, Verneris MR, Hsu K, Fleischhauer K, Cutler C, Antin JH, Neuberg D, Ebert BL. Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. N Engl J Med. 2017 Feb 9;376(6):536-547.

Liu X, Wang C. The emerging roles of the STING adaptor protein in immunity and diseases. Immunology. 2016 Mar;147(3):285-291.

Loke J, Chin PS, Keane P, Pickin A, Assi SA, Ptasinska A, Imperato MR, Cockerill PN, Bonifer C. C/EBPα overrides epigenetic reprogramming by oncogenic transcription factors in acute myeloid leukemia. Blood Adv. 2018 Feb 13;2(3):271-284.

Lorgeoux RP, Guo F, Liang C. From promoting to inhibiting: diverse roles of helicases in HIV-1 Replication. Retrovirology. 2012 Sep 28; 9:79.

Maciejewski JP, Padgett RA, Brown AL, Müller-Tidow C. DDX41-related myeloid neoplasia. Semin Hematol. 2017 Apr;54(2):94-97.

Makarov VL, Hirose Y, Langmore JP. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. Cell. 1997 Mar 7;88(5):657-666.

Malcovati L, Hellström-Lindberg E, Bowen D, Adès L, Cermak J, Del Cañizo C, Della Porta MG, Fenaux P, Gattermann N, Germing U, Jansen JH, Mittelman M, Mufti G, Platzbecker U, Sanz GF, Selleslag D, Skov-Holm M, Stauder R, Symeonidis A, van de Loosdrecht AA, de Witte T, Cazzola M; European Leukemia Net. Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. Blood. 2013 Oct 24;122(17):2943-2964.

Mallhi K, Dix DB, Niederhoffer KY, Armstrong L, Rozmus J. Successful umbilical cord blood hematopoietic stem cell transplantation in pediatric patients with MDS/AML associated with underlying GATA2 mutations: two case reports and review of literature. Pediatr Transplant. 2016 Nov;20(7):1004-1007.

Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer. 2009 Mar;9(3):153-166.

Marconi C, Canobbio I, Bozzi V, Pippucci T, Simonetti G, Melazzini F, Angori S, Martinelli G, Saglio G, Torti M, Pastan I, Seri M, Pecci A. 5'UTR point substitutions and N-terminal truncating mutations of ANKRD26 in acute myeloid leukemia. J Hematol Oncol. 2017 Jan 18;10(1):18.

Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott RM, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, McMichael JF, Meyer RJ, Schindler JK, Pohl CS, Wallis JW, Shi X, Lin L, Schmidt H, Tang Y, Haipek C, Wiechert ME, Ivy JV, Kalicki J, Elliott G, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson MA, Baty J, Heath S, Shannon WD, Nagarajan R, Link DC, Walter MJ, Graubert TA, DiPersio JF, Wilson RK, Ley TJ. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med. 2009 Sep 10;361(11):1058-1066.

Marsh JC, Will AJ, Hows JM, Sartori P, Darbyshire PJ, Williamson PJ, Oscier DG, Dexter TM, Testa NG. "Stem cell" origin of the hematopoietic defect in dyskeratosis congenita. Blood. 1992 Jun 15;79(12):3138-3144.

Marsh JCW, Gutierrez-Rodrigues F, Cooper J, Jiang J, Gandhi S, Kajigaya S, Feng X, Ibanez MDPF, Donaires FS, Lopes da Silva JP, Li Z, Das S, Ibanez M, Smith AE, Lea N, Best S, Ireland R, Kulasekararaj AG, McLornan DP, Pagliuca A, Callebaut I, Young NS, Calado RT, Townsley DM, Mufti GJ. Heterozygous RTEL1 variants in bone marrow failure and myeloid neoplasms. Blood Adv. 2018 Jan 4;2(1):36-48.

Martini E, Diaz RL, Hunter N, Keeney S. Crossover homeostasis in yeast meiosis. Cell. 2006 Jul 28;126(2):285-295.

McCullagh P, Chaplin T, Meerabux J, Grenzelias D, Lillington D, Poulsom R, Gregorini A, Saha V, Young BD. The cloning, mapping and expression of a novel

gene, BRL, related to the AF10 leukaemia gene. Oncogene. 1999 Dec 9;18(52):7442-7452.

McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J; Understanding Society Scientific Group, Crawley C, Craig J, Scott MA, Hodkinson C, Baxter J, Rad R, Forsyth DR, Quail MA, Zeggini E, Ouwehand W, Varela I, Vassiliou GS. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Rep. 2015 Mar 3;10(8):1239-1245.

McWhirter SM, Barbalat R, Monroe KM, Fontana MF, Hyodo M, Joncker NT, Ishii KJ, Akira S, Colonna M, Chen ZJ, Fitzgerald KA, Hayakawa Y, Vance RE. A host type I interferon response is induced by cytosolic sensing of the bacterial second messenger cyclic-di-GMP. J Exp Med. 2009 Aug 31;206(9):1899-1911.

Mendez LM, Polo JM, Yu JJ, Krupski M, Ding BB, Melnick A, Ye BH. CtBP is an essential corepressor for BCL6 autoregulation. Mol Cell Biol. 2008 Apr;28(7):2175-2186.

Metzker ML. Sequencing technologies — the next generation. Nature Rev Genet 2010; 11, 31-46.

Meyerson M, Enders GH, Wu CL, Su LK, Gorka C, Nelson C, Harlow E, Tsai LH. A family of human cdc2-related protein kinases. EMBO J. 1992 Aug;11(8):2909-2917.

Mieczkowski PA, Mieczkowska JO, Dominska M, Petes TD. Genetic regulation of telomere-telomere fusions in the yeast Saccharomyces cerevisae. Proc Natl Acad Sci U S A. 2003 Sep 16;100(19):10854-10859.

Miller KM, Rog O, Cooper JP. Semi-conservative DNA replication through telomeres requires Taz1. Nature. 2006 Apr 6;440(7085):824-828.

Ming YZ, Di X, Gomez-Sanchez EP, Gomez-Sanchez CE. Improved downward capillary transfer for blotting of DNA and RNA. Biotechniques. 1994 Jan;16(1):58-59.

Miyamoto T, Weissman IL, Akashi K. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. Proc Natl Acad Sci U S A. 2000 Jun 20;97(13):7521-7526.

Miyashita M, Oshiumi H, Matsumoto M, Seya T. DDX60, a DEXD/H box helicase, is a novel antiviral factor promoting RIG-I-like receptor-mediated signaling. Mol Cell Biol. 2011 Sep;31(18):3802-3819.

Mohamedali AM, Smith AE, Gaken J, Lea NC, Mian SA, Westwood NB, Strupp C, Gattermann N, Germing U, Mufti GJ. Novel TET2 mutations associated with UPD4q24 in myelodysplastic syndrome. J Clin Oncol. 2009 Aug 20;27(24):4002-4006.

Mol CD, Harris JM, McIntosh EM, Tainer JA. Human dUTP pyrophosphatase: uracil recognition by a beta hairpin and active sites formed by three separate subunits. Structure. 1996 Sep 15;4(9):1077-1092.

Moriya K, Niizuma H, Rikiishi T, Yamaguchi H, Sasahara Y, Kure S. Novel compound heterozygous RTEL1 gene mutations in a patient with Hoyeraal-Hreidarsson syndrome. Pediatr Blood Cancer. 2016; 63(9):1683-1684.

Morris LG, Kaufman AM, Gong Y, Ramaswami D, Walsh LA, Turcan Ş, Eng S, Kannan K, Zou Y, Peng L, Banuchi VE, Paty P, Zeng Z, Vakiani E, Solit D, Singh B, Ganly I, Liau L, Cloughesy TC, Mischel PS, Mellinghoff IK, Chan TA. Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. Nat Genet. 2013 Mar;45(3):253-261.

Mosrati MA, Willander K, Falk IJ, Hermanson M, Höglund M, Stockelberg D, Wei Y, Lotfi K, Söderkvist P. Association between TERT promoter polymorphisms and acute myeloid leukemia risk and prognosis. Oncotarget. 2015 Sep 22;6(28):25109-25120.

Moudra A, Hubackova S, Machalova V, Vancurova M, Bartek J, Reinis M, Hodny Z, Jonasova A. Dynamic alterations of bone marrow cytokine landscape of myelodysplastic syndromes patients treated with 5-azacytidine. Oncoimmunology. 2016 May 13;5(10):e1183860.

Mufti GJ, Bennett JM, Goasguen J, Bain BJ, Baumann I, Brunning R, Cazzola M, Fenaux P, Germing U, Hellström-Lindberg E, Jinnai I, Manabe A, Matsuda A, Niemeyer CM, Sanz G, Tomonaga M, Vallespi T, Yoshimi A; International Working Group on Morphology of Myelodysplastic Syndrome. Diagnosis and classification of myelodysplastic syndrome: International Working Group on Morphology of myelodysplastic syndrome (IWGM-MDS) consensus proposals for the definition and enumeration of myeloblasts and ring sideroblasts. Haematologica 2008; 93, 1712-1717.

Muñoz IM, Hain K, Déclais AC, Gardiner M, Toh GW, Sanchez-Pulido L, Heuckmann JM, Toth R, Macartney T, Eppink B, Kanaar R, Ponting CP, Lilley DM, Rouse J. Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. Mol Cell. 2009 Jul 10;35(1):116-127.

Myers KC, Bolyard AA, Otto B, Wong TE, Jones AT, Harris RE, Davies SM, Dale DC, Shimamura A. Variable clinical presentation of Shwachman-Diamond syndrome: update from the North American Shwachman-Diamond Syndrome Registry. J Pediatr. 2014 Apr;164(4):866-870.

Naji S, Ambrus G, Cimermančič P, Reyes JR, Johnson JR, Filbrandt R, Huber MD, Vesely P, Krogan NJ, Yates JR 3rd, Saphire AC, Gerace L. Host cell interactome of HIV-1 Rev includes RNA helicases involved in multiple facets of virus production. Mol Cell Proteomics. 2012 Apr;11(4):M111.015313.

Nathan N, Taam RA, Epaud R, Delacourt C, Deschildre A, Reix P, Chiron R, de Pontbriand U, Brouard J, Fayon M, Dubus JC, Giovannini-Chami L, Bremont F, Bessaci K, Schweitzer C, Dalphin ML, Marguet C, Houdouin V, Troussier F, Sardet A, Hullo E, Gibertini I, Mahloul M, Michon D, Priouzeau A, Galeron L, Vibert JF, Thouvenin G, Corvol H, Deblic J, Clement A; French RespiRare® Group. A national internet-linked based database for pediatric interstitial lung diseases: the French network. Orphanet J Rare Dis. 2012 Jun 15;7:40.

Nanri T, Uike N, Kawakita T, Iwanaga E, Mitsuya H, Asou N. A family harboring a germ-line N-terminal C/EBPalpha mutation and development of acute myeloid leukemia with an additional somatic C-terminal C/EBPalpha mutation. Genes Chromosomes Cancer. 2010 Mar;49(3):237-241.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010 Sep;42(9):790-793.

Nickels EM, Soodalter J, Churpek JE, Godley LA. Recognizing familial myeloid leukemia in adults. Ther Adv Hematol. 2013 Aug;4(4):254-269.

Nicol SM, Bray SE, Black HD, Lorimore SA, Wright EG, Lane DP, Meek DW, Coates PJ, Fuller-Pace FV. The RNA helicase p68 (DDX5) is selectively required for the induction of p53-dependent p21 expression and cell-cycle arrest after DNA damage. Oncogene. 2013 Jul 18;32(29):3461-3469.

Niemeyer CM, Baumann I. Classification of childhood aplastic anemia and myelodysplastic syndrome. Hematology Am Soc Hematol Educ Program. 2011;2011:84-89.

Noetzli L, Lo RW, Lee-Sherick AB, Callaghan M, Noris P, Savoia A, Rajpurkar M, Jones K, Gowan K, Balduini C, Pecci A, Gnan C, De Rocco D, Doubek M, Li L, Lu L, Leung R, Landolt-Marticorena C, Hunger S, Heller P, Gutierrez-Hartmann A, Xiayuan L, Pluthero FG, Rowley JW, Weyrich AS, Kahr WHA, Porter CC, Di Paola J. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. Nat Genet. 2015 May;47(5):535-538.

Noris P, Perrotta S, Seri M, Pecci A, Gnan C, Loffredo G, Pujol-Moix N, Zecca M, Scognamiglio F, De Rocco D, Punzo F, Melazzini F, Sciarretta S, Casale M, Marconi C, Pippucci T, Amendola G, Notarangelo LD, Klersy C, Civaschi E, Balduini CL, Savoia A. Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families. Blood. 2011 Jun 16;117(24):6673-6680.

Noris P, Favier R, Alessi MC, Geddis AE, Kunishima S, Heller PG, Giordano P, Niederhoffer KY, Bussel JB, Podda GM, Vianelli N, Kersseboom R, Pecci A, Gnan C, Marconi C, Auvrignon A, Cohen W, Yu JC, Iguchi A, Miller Imahiyerobo A, Boehlen F, Ghalloussi D, De Rocco D, Magini P, Civaschi E, Biino G, Seri M, Savoia A, Balduini CL. ANKRD26-related thrombocytopenia and myeloid malignancies. Blood 2013; 122, 1987-1989.

Ochs HD, Filipovich AH, Veys P, Cowan MJ, Kapoor N. Wiskott-Aldrich syndrome: diagnosis, clinical and laboratory manifestations, and treatment. Biol Blood Marrow Transplant. 2009 Jan;15(1 Suppl):84-90.

Oefner PJ & Huber CG. A decade of high-resolution liquid chromatography of nucleic acids on styrene-divinylbenzene copolymers. J Chromatogr B 2002; 782, 27–55.

Omura H, Oikawa D, Nakane T, Kato M, Ishii R, Ishitani R, Tokunaga F, Nureki O. Structural and Functional Analysis of DDX41: a bispecific immune receptor for DNA and cyclic dinucleotide. Sci Rep. 2016 Oct 10;6:34756.

Oshiumi H, Sakai K, Matsumoto M, Seya T. DEAD/H BOX 3 (DDX3) helicase binds the RIG-I adaptor IPS-1 to up-regulate IFN-beta-inducing potential. Eur J Immunol. 2010 Apr;40(4):940-948.

Ostergaard P, Simpson MA, Connell FC, Steward CG, Brice G, Woollard WJ, Dafou D, Kilo T, Smithson S, Lunt P, Murday VA, Hodgson S, Keenan R, Pilz DT, Martinez-Corral I, Makinen T, Mortimer PS, Jeffery S, Trembath RC, Mansour S. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukaemia (Emberger syndrome). Nature Genetics 2011; 43, 929-931.

Owen CJ, Toze CL, Koochin A, Forrest DL, Smith CA, Stevens JM, Jackson SC, Poon MC, Sinclair GD, Leber B, Johnson PRE, Macheta A, Yin JAL, Barnett MJ, Lister AT & Fitzgibbon F. Five new pedi- grees with inherited RUNX1 mutations causing familial platelet disorder with propensity to myeloid malignancy. Blood 2008; 112, 4639-4645.

Owsianka AM, Patel AH. Hepatitis C virus core protein interacts with a human DEAD box protein DDX3. Virology. 1999 May 10;257(2):330-40.

Pabst T, Eyholzer M, Haefliger S, Schardt J, Mueller BU. Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia. Journal of Clinical Oncology 2008; 26:5088-93.

Pandolfi A, Barreyro L, Steidl U. Concise review: preleukemic stem cells: molecular biology and clinical implications of the precursors to leukemia stem cells. Stem Cells Transl Med. 2013 Feb;2(2):143-150.

Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, Yoon CJ, Ellis P, Wedge DC, Pellagatti A, Shlien A, Groves MJ, Forbes SA, Raine K, Hinton J, Mudie LJ, McLaren S, Hardy C, Latimer C, Della Porta MG, O'Meara S, Ambaglio I, Galli A, Butler AP, Walldin G, Teague JW, Quek L, Sternberg A, Gambacorti-Passerini C, Cross NC, Green AR, Boultwood J, Vyas P, Hellstrom-Lindberg E, Bowen D, Cazzola M, Stratton MR, Campbell PJ; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013 Nov 21;122(22):3616-3627.

Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, 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 DR, Teague JW, Butler AP, Greaves MF, Ganser A, Döhner K, Schlenk RF, Döhner H, Campbell PJ. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016 Jun 9;374(23):2209-2221.

Pâques F, Haber JE. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev. 1999 Jun;63(2):349-404.

Parvatiyar K, Zhang Z, Teles RM, Ouyang S, Jiang Y, Iyer SS, Zaver SA, Schenk M, Zeng S, Zhong W, Liu ZJ, Modlin RL, Liu YJ, Cheng G. The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. Nat Immunol. 2012 Dec;13(12):1155-1161.

Pasquet M, Bellanné-Chantelot C, Tavitian S, Prade N, Beaupain B, Larochelle O, Petit A, Rohrlich P, Ferrand C, Van Den Neste E, Poirel HA, Lamy T, Ouachée-Chardin M, Mansat-De Mas V, Corre J, Récher C, Plat G, Bachelerie F, Donadieu J, Delabesse E. High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. Blood. 2013 Jan 31;121(5):822-829.

Pereboeva L, Westin E, Patel T, Flaniken I, Lamb L, Klingelhutz A, Goldman F. DNA damage responses and oxidative stress in dyskeratosis congenita. PLoS One 2013;8(10):e76473.

Perez-Andreu V, Roberts KG, Harvey RC, Yang W, Cheng C, Pei D, Xu H, Gastier-Foster J, E S, Lim JY, Chen IM, Fan Y, Devidas M, Borowitz MJ, Smith C, Neale G, Burchard EG, Torgerson DG, Klussmann FA, Villagran CR, Winick NJ, Camitta BM, Raetz E, Wood B, Yue F, Carroll WL, Larsen E, Bowman WP, Loh ML, Dean M, Bhojwani D, Pui CH, Evans WE, Relling MV, Hunger SP, Willman CL, Mullighan CG, Yang JJ. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. Nat Genet. 2013 Dec;45(12):1494-1498.

Perez Botero J, Oliveira JL, Chen D, Reichard KK, Viswanatha DS, Nguyen PL, Pruthi RK, Majerus J, Gada P, Gangat N, Tefferi A, Patnaik MM. ASXL1 mutated chronic myelomonocytic leukemia in a patient with familial thrombocytopenia secondary to germline mutation in ANKRD26. Blood Cancer J. 2015 May 22;5:e315.

Peterkin T, Gibson A, Loose M, Patient R. The roles of GATA-4, -5 and -6 in vertebrate heart development. Semin Cell Dev Biol. 2005 Feb;16(1):83-94.

Peters D, Radine C, Reese A, Budach W, Sohn D, Jänicke RU. The DEAD-box RNA helicase DDX41 is a novel repressor of p21(WAF1/CIP1) mRNA translation. J Biol Chem. 2017 May 19;292(20):8331-8341.

Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013;9(8):e1003709.

Ping Z, Xia Y, Shen T, Parekh V, Siegal GP, Eltoum IE, He J, Chen D, Deng M, Xi R, Shen D. A microscopic landscape of the invasive breast cancer genome. Sci Rep. 2016 Jun 10;6:27545.

Pippucci T, Savoia A, Perrotta S, Pujol-Moix N, Noris P, Castegnaro G, Pecci A, Gnan C, Punzo F, Marconi C, Gherardi S, Loffredo G, De Rocco D, Scianguetta S, Barozzi S, Magini P, Bozzi V, Dezzani L, Di Stazio M, Ferraro M, Perini G, Seri M, Balduini CL. Mutations in the 5' UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. Am J Hum Genet. 2011 Jan 7;88(1):115-120.

Poggi M, Canault M, Favier M, Turro E, Saultier P, Ghalloussi D, Baccini V, Vidal L, Mezzapesa A, Chelghoum N, Mohand-Oumoussa B, Falaise C, Favier R, Ouwehand WH, Fiore M, Peiretti F, Morange PE, Saut N, Bernot D, Greinacher A, BioResource N, Nurden AT, Nurden P, Freson K, Trégouët DA, Raslova H, Alessi MC. Germline variants in ETV6 underlie reduced platelet formation, platelet dysfunction and increased levels of circulating CD34+ progenitors. Haematologica. 2017 Feb;102(2):282-294.

Polprasert C, Schulze I, Sekeres MA, Makishima H, Przychodzen B, Hosono N, Singh J, Padgett RA, Gu X, Phillips JG, Clemente M, Parker Y, Lindner D, Dienes B, Jankowsky E, Sauntharajah Y, Du Y, Oakley K, Nguyen N, Mukherjee S, Pabst C, Godley LA, Churpek JE, Pollyea DA, Krug U, Berdel WE, Klein HU, Dugas M, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Yoshida K, Ogawa S, Müller-Tidow C, Maciejewski JP. Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. Cancer Cell 2015; 27, 658-670.

Pop M & Salzberg SL. Bioinformatics challenges of new sequencing technology. Trends Genet 2008; 24, 142–149.

Preudhomme C, Renneville A, Bourdon V, Philippe N, Roche-Lestienne C, Boissel N, Dhedin N, André JM, Cornillet-Lefebvre P, Baruchel A, Mozziconacci MJ, Sobol H. High frequency of RUNX1 biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder. Blood 2009; 113:5583-7.

Pugh RA, Honda M, Leesley H, Thomas A, Lin Y, Nilges MJ, Cann IK, Spies M. The iron-containing domain is essential in Rad3 helicases for coupling of ATPhydrolysis to DNA translocation and for targeting the helicase to the single-stranded DNA-double-stranded DNA junction. J Biol Chem. 2008 Jan 18;283(3):1732-1743.

Pyle AM. Translocation and unwinding mechanisms of RNA and DNA helicases. Annu Rev Biophys. 2008;37:317-336.

Qin C, Ren L, Ji M, Lv S, Wei Y, Zhu D, Lin Q, Xu P, Chang W, Xu J. CDKL1 promotes tumor proliferation and invasion in colorectal cancer. Onco Targets Ther. 2017 Mar 16;10:1613-1624.

Rassa JC, Ross SR. Viruses and Toll-like receptors. Microbes Infect. 2003 Sep;5(11):961-968.

Rehwinkel J, Reis e Sousa C. RIGorous detection: exposing virus through RNA sensing. Science. 2010 Jan 15;327(5963):284-286.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405-424.

Rochowski A, Olson SB, Alonzo TA, Gerbing RB, Lange BJ, Alter BP. Patients with Fanconi anemia and AML have different cytogenetic clones than de novo cases of AML. Pediatr Blood Cancer. 2012 Nov;59(5):922-924.

Rosenberg PS, Alter BP, Ebell W. Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. Haematologica. 2008 Apr;93(4):511-517.

Rowley JD. Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. Ann Genet. 1973 Jun;16(2):109-112.

Rowley JD, Golomb HM, Dougherty C. 15/17 translocation, a consistent chromosomal change in acute promyelocytic leukaemia. Lancet. 1977 Mar 5;1(8010):549-550.

Rudolf J, Makrantoni V, Ingledew WJ, Stark MJ, White MF. The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. Mol Cell. 2006 Sep 15;23(6):801-808.

Sagona AP, Nezis IP, Pedersen NM, Liestøl K, Poulton J, Rusten TE, Skotheim RI, Raiborg C, Stenmark H. PtdIns(3)P controls cytokinesis through KIF13A-mediated recruitment of FYVE-CENT to the midbody. Nat Cell Biol. 2010 Apr;12(4):362-371.

Saito T, Gale M Jr. Principles of intracellular viral recognition. Curr Opin Immunol. 2007 Feb;19(1):17-23.

Sakurai M, Kunimoto H, Watanabe N, Fukuchi Y, Yuasa S, Yamazaki S, Nishimura T, Sadahira K, Fukuda K, Okano H, Nakauchi H, Morita Y, Matsumura I, Kudo K, Ito E, Ebihara Y, Tsuji K, Harada Y, Harada H, Okamoto S, Nakajima H. Impaired hematopoietic differentiation of RUNX1-mutated induced pluripotent stem cells derived from FPD/AML patients. Leukemia. 2014 Dec;28(12):2344-2354.

Sakurai M, Kasahara H, Yoshida K, Yoshimi A, Kunimoto H, Watanabe N, Shiraishi Y, Chiba K, Tanaka H, Harada Y, Harada H, Kawakita T, Kurokawa M, Miyano S, Takahashi S, Ogawa S, Okamoto S, Nakajima H. Genetic basis of myeloid transformation in familial platelet disorder/acute myeloid leukemia patients with haploinsufficient RUNX1 allele. Blood Cancer J. 2016 Feb 5;6:e392.

Sambrook J & Russell D. Protocol Alkaline Agarose Gel Electrophoresis. Cold Spring Harb Protoc 2006; doi:10.1101/pdb.prot4027.

Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnström K, Mallick S, Kirby A, Wall DP, MacArthur DG, Gabriel SB, DePristo M, Purcell SM, Palotie A, Boerwinkle E, Buxbaum JD, Cook EH Jr, Gibbs RA, Schellenberg GD, Sutcliffe JS, Devlin B, Roeder K, Neale BM, Daly MJ. A framework for the interpretation of de novo mutation in human disease. Nat Genet. 2014 Sep;46(9):944-950.

Sander J, Schmidt SV, Cirovic B, McGovern N, Papantonopoulou O, Hardt AL, Aschenbrenner AC, Kreer C, Quast T, Xu AM, Schmidleithner LM, Theis H, Thi Huong LD, Sumatoh HRB, Lauterbach MAR, Schulte-Schrepping J, Günther P, Xue J, Baßler K, Ulas T, Klee K, Katzmarski N, Herresthal S, Krebs W, Martin B, Latz E, Händler K, Kraut M, Kolanus W, Beyer M, Falk CS, Wiegmann B, Burgdorf S, Melosh NA, Newell EW, Ginhoux F, Schlitzer A, Schultze JL. Cellular Differentiation of Human Monocytes Is Regulated by Time-Dependent Interleukin-4 Signaling and theTranscriptional Regulator NCOR2. Immunity. 2017 Dec 19;47(6):1051-1066.e12.

Sato M, Suemori H, Hata N, Asagiri M, Ogasawara K, Nakao K, Nakaya T, Katsuki M, Noguchi S, Tanaka N, Taniguchi T. Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN-alpha/beta gene induction. Immunity. 2000 Oct;13(4):539-548.

Savage SA, Giri N, Baerlocher GM, Orr N, Lansdorp PM, Alter BP. TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita. Am J Hum Genet. 2008 Feb;82(2):501-509.

Savage SA, Bertuch AA. The genetics and clinical manifestations of telomere biology disorders. Genet Med. 2010 Dec;12(12):753-764.

Sawyer SL, Tian L, Kähkönen M, Schwartzentruber J, Kircher M; University of Washington Centre for Mendelian Genomics; FORGE Canada Consortium, Majewski J, Dyment DA, Innes AM, Boycott KM, Moreau LA, Moilanen JS, Greenberg RA. Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. Cancer Discov. 2015 Feb;5(2):135-142.

Saygin C, Carraway HE. Emerging therapies for acute myeloid leukemia. J Hematol Oncol. 2017 Apr 18;10(1):93.

Schmid SR, Linder P. D-E-A-D protein family of putative RNA helicases. Mol Microbiol. 1992 Feb;6(3):283-291.

Schmidt A, Rothenfusser S, Hopfner KP. Sensing of viral nucleic acids by RIG-I: from translocation to translation. Eur J Cell Biol. 2012 Jan;91(1):78-85.

Schmit JM, Turner DJ, Hromas RA, Wingard JR, Brown RA, Li Y, Li MM, Slayton WB, Cogle CR. Two novel RUNX1 mutations in a patient with congenital thrombocytopenia that evolved into a high grade myelodysplastic syndrome. Leuk Res Rep. 2015 Apr 3;4(1):24-27.

Schröder M. Human DEAD-box protein 3 has multiple functions in gene regulation and cell cycle control and is a prime target for viral manipulation. Biochem Pharmacol. 2010 Feb 1;79(3):297-306.

Schulz E, Valentin A, Ulz P, Beham-Schmid C, Lind K, Rupp V, Lackner H, Wölfler A, Zebisch A, Olipitz W, Geigl J, Berghold A, Speicher MR, Sill H. Germline mutations in the DNA damage response genes BRCA1, BRCA2, BARD1 and TP53 in patients with therapy related myeloid neoplasms. J Med Genet. 2012 Jul;49(7):422-428.

Schütz P, Karlberg T, van den Berg S, Collins R, Lehtiö L, Högbom M, Holmberg-Schiavone L, Tempel W, Park HW, Hammarström M, Moche M, Thorsell AG, Schüler H. Comparative structural analysis of human DEAD-box RNA helicases. PLoS One. 2010 Sep 30;5(9).

Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014 Apr;11(4):361-362.

Scorilas A, Kyriakopoulou L, Katsaros D, Diamandis EP. Cloning of a gene (SR-A1), encoding for a new member of the human Ser/Arg-rich family of pre-mRNA

splicing factors: overexpression in aggressive ovarian cancer. Br J Cancer. 2001 Jul 20;85(2):190-198.

Sekeres MA, Schoonen WM, Kantarjian H, List A, Fryzek J, Paquette R, Maciejewski JP. Characteristics of US patients with myelodysplastic syndromes: results of six cross-sectional physician surveys. J Natl Cancer Inst. 2008 Nov 5;100(21):1542-1551.

Sekeres MA. The epidemiology of myelodysplastic syndromes. Hematol Oncol Clin North Am. 2010 Apr;24(2):287-294.

Sellick GS, Spendlove HE, Catovsky D, Pritchard-Jones K, Houlston RS. Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia. Leukemia. 2005 Jul;19(7):1276-1278.

Seth RB, Sun L, Ea CK, Chen ZJ. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. Cell. 2005 Sep 9;122(5):669-682.

Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, de Lange T. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. Cell. 2009 Jul 10;138(1):90-103.

Shammas C, Menne TF, Hilcenko C, Michell SR, Goyenechea B, Boocock GR, Durie PR, Rommens JM, Warren AJ. Structural and mutational analysis of the SBDS protein family. Insight into the leukemia-associated Shwachman-Diamond Syndrome. J Biol Chem. 2005 May 13;280(19):19221-19229.

Shahrabi S, Khosravi A, Shahjahani M, Rahim F, Saki N. Genetics and Epigenetics of Myelodysplastic Syndromes and Response to Drug Therapy: New Insights. Oncol Rev. 2016 Dec 14;10(2):311.

Shcherbina A, Candotti F, Rosen FS, Remold-O'Donnell E. High incidence of lymphomas in a subgroup of Wiskott-Aldrich syndrome patients. Br J Haematol. 2003 May;121(3):529-530.

Shearer AE, Eppsteiner RW, Booth KT, Ephraim SS, Gurrola J 2nd, Simpson A, Black-Ziegelbein EA, Joshi S, Ravi H, Giuffre AC, Happe S, Hildebrand MS, Azaiez H, Bayazit YA, Erdal ME, Lopez-Escamez JA, Gazquez I, Tamayo ML, Gelvez NY, Leal GL, Jalas C, Ekstein J, Yang T, Usami S, Kahrizi K, Bazazzadegan N, Najmabadi H, Scheetz TE, Braun TA, Casavant TL, LeProust EM, Smith RJ. Utilizing ethnic-specific differences in minor allele frequency to recategorize reported pathogenic deafness variants. Am J Hum Genet. 2014 Oct 2;95(4):445-453.

Shiba N, Funato M, Ohki K, Park MJ, Mizushima Y, Adachi S, Kobayashi M, Kinoshita A, Sotomatsu M, Arakawa H, Tawa A, Horibe K, Tsukimoto I, Hayashi Y. Mutations of the GATA2 and CEBPA genes in paediatric acute myeloid leukaemia. Br J Haematol. 2014 Jan;164(1):142-145.

Shih LY, Liang DC, Huang CF, Wu JH, Lin TL, Wang PN, Dunn P, Kuo MC, Tang TC. AML patients with CEBPalpha mutations mostly retain identical mutant patterns

but frequently change in allelic distribution at relapse: a comparative analysis on paired diagnosis and relapse samples. Leukemia 2006; 20, 604-609.

Shippen-Lentz D, Blackburn EH. Functional evidence for an RNA template in telomerase. Science. 1990 Feb 2;247(4942):546-552.

Shlush LI, Minden MD. Preleukemia: the normal side of cancer. Curr Opin Hematol. 2015 Mar;22(2):77-84.

Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, Kennedy JA, Schimmer AD, Schuh AC, Yee KW, McLeod JL, Doedens M, Medeiros JJ, Marke R, Kim HJ, Lee K, McPherson JD, Hudson TJ; HALT Pan-Leukemia Gene Panel Consortium, Brown AM, Yousif F, Trinh QM, Stein LD, Minden MD, Wang JC, Dick JE. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature. 2014 Feb 20;506(7488):328-333. Erratum in: Nature. 2014 Apr 17;508(7496):420.

Silverman E, Edwalds-Gilbert G, Lin RJ. DExD/H-box proteins and their partners: helping RNA helicases unwind. Gene. 2003 Jul 17;312:1-16.

Singer MS, Gottschling DE. TLC1: template RNA component of Saccharomyces cerevisiae telomerase. Science. 1994 Oct 21;266(5184):404-409.

Sinha S, Thomas D, Yu L, Gentles AJ, Jung N, Corces-Zimmerman MR, Chan SM, Reinisch A, Feinberg AP, Dill DL, Majeti R. Mutant WT1 is associated with DNA hypermethylation of PRC2 targets in AML and responds to EZH2 inhibition. Blood. 2015 Jan 8;125(2):316-326.

Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J. Mutation of CEBPA in familial acute myeloid leukaemia. The New England Journal of Medicine 2004; 351, 2403-2407.

Smiraglia DJ, Rush LJ, Frühwald MC, Dai Z, Held WA, Costello JF, Lang JC, Eng C, Li B, Wright FA, Caligiuri MA, Plass C. Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. Hum Mol Genet. 2001 Jun15;10(13):1413-1419.

Sohn D, Peters D, Piekorz RP, Budach W, Jänicke RU. miR-30e controls DNA damage-induced stress responses by modulating expression of the CDK inhibitor p21WAF1/CIP1 and caspase-3. Oncotarget. 2016 Mar 29;7(13):15915-15929.

Song W-J, Sullivan MG, Legare RD, Hutchings S, Tan X, Kufrin D, Ratajczak J, Resende IC, Haworth C, Hock R, Loh M, Felix C, Roy DC, Busque L, Kurnit D, Willman C, Gewirtz AM, Speck NA, Bushweller JH, Li FP, Gardiner K, Poncz M, Maris JM, Gilliland DG. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. Nature Genetics 1999; 23,166- 175.

Sood R, Kamikubo Y, Liu P. Role of RUNX1 in hematological malignancies. Blood. 2017 Apr 13;129(15):2070-2082.

Soto-Rifo R, Rubilar PS, Limousin T, de Breyne S, Décimo D, Ohlmann T. DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. EMBO J. 2012 Sep 12;31(18):3745-3756.

Soulier J, Leblanc T, Larghero J, Dastot H, Shimamura A, Guardiola P, Esperou H, Ferry C, Jubert C, Feugeas JP, Henri A, Toubert A, Socié G, Baruchel A, Sigaux F, D'Andrea AD, Gluckman E. Detection of somatic mosaicism and classification of Fanconi anemia patients by analysis of the FA/BRCA pathway. Blood. 2005 Feb 1;105(3):1329-1336.

Spagnolo P, Grunewald J, du Bois RM. Genetic determinants of pulmonary fibrosis: evolving concepts. Lancet Respir Med. 2014 May;2(5):416-428.

Spinner MA, Sanchez LA, Hsu AP, Shaw PA, Zerbe CS, Calvo KR, Arthur DC, Gu W, Gould CM, Brewer CC, Cowen EW, Freeman AF, Olivier KN, Uzel G, Zelazny AM, Daub JR, Spalding CD, Claypool RJ, Giri NK, Alter BP, Mace EM, Orange JS, Cuellar-Rodriguez J, Hickstein DD, Holland SM. GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. Blood. 2014 Feb 6;123(6):809-821.

Stanley SE, Gable DL, Wagner CL, Carlile TM, Hanumanthu VS, Podlevsky JD, Khalil SE, DeZern AE, Rojas-Duran MF, Applegate CD, Alder JK, Parry EM, Gilbert WV, Armanios M. Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. Sci Transl Med. 2016 Aug 10;8(351):351ra107.

Steele L, Rommens JM, Stockley T, Baskin B, Ray PN. De Novo Mutations Causing Shwachman-Diamond Syndrome and a Founder Mutation in SBDS in the French-Canadian Population. J Investig Genomics 2014 July:1(2): 00008.

Steele MP, Speer MC, Loyd JE, Brown KK, Herron A, Slifer SH, Burch LH, Wahidi MM, Phillips JA 3rd, Sporn TA, McAdams HP, Schwarz MI, Schwartz DA. Clinical and pathologic features of familial interstitial pneumonia. Am J Respir Crit Care Med. 2005 Nov 1;172(9):1146-1152.

Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, Ebert BL. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood. 2015a Jul 2;126(1):9-16.

Steensma DP. Myelodysplastic Syndromes: Diagnosis and Treatment. Mayo Clin Proc. 2015b Jul;90(7):969-983.

Stuart BD, Choi J, Zaidi S, Xing C, Holohan B, Chen R, Choi M, Dharwadkar P, Torres F, Girod CE, Weissler J, Fitzgerald J, Kershaw C, Klesney-Tait J, Mageto Y, Shay JW, Ji W, Bilguvar K, Mane S, Lifton RP, Garcia CK. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. Nat Genet. 2015 May;47(5):512-517.

Sun L, Wu J, Du F, Chen X, Chen ZJ. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. Science. 2013 Feb 15;339(6121):786-791.

Sung P, Krejci L, Van Komen S, Sehorn MG. Rad51 recombinase and recombination mediators. J Biol Chem. 2003 Oct 31;278(44):42729-42732.

Sung P, Klein H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nat Rev Mol Cell Biol. 2006 Oct;7(10):739-750.

Suzuki T, Maeda S, Furuhata E, Shimizu Y, Nishimura H, Kishima M, Suzuki H. A screening system to identify transcription factors that induce binding site-directed DNA demethylation. Epigenetics Chromatin. 2017 Dec 8;10(1):60.

Svendsen JM, Smogorzewska A, Sowa ME, O'Connell BC, Gygi SP, Elledge SJ, Harper JW. Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. Cell. 2009 Jul 10;138(1):63-77.

Svendsen JM, Harper JW. GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution. Genes Dev. 2010 Mar 15;24(6):521-536.

Taglienti CA, Wysk M, Davis RJ. Molecular cloning of the epidermal growth factor-stimulated protein kinase p56 KKIAMRE. Oncogene. 1996 Dec 19;13(12):2563-2574.

Takahashi S, McConnell MJ, Harigae H, Kaku M, Sasaki T, Melnick AM, Licht JD. The Flt3 internal tandem duplication mutant inhibits the function of transcriptional repressors by blocking interactions with SMRT. Blood. 2004 Jun 15;103(12):4650-4658.

Takai H, Smogorzewska A, de Lange T. DNA damage foci at dysfunctional telomeres. Curr Biol. 2003 Sep 2;13(17):1549-1556.

Takaoka A, Wang Z, Choi MK, Yanai H, Negishi H, Ban T, Lu Y, Miyagishi M, Kodama T, Honda K, Ohba Y, Taniguchi T. DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. Nature. 2007 Jul 26;448(7152):501-505.

Takeuchi O, Akira S. Pattern recognition receptors and inflammation. Cell. 2010 Mar 19;140(6):805-820.

Tawana K, Wang J, Renneville A, Bödör C, Hills R, Loveday C, Savic A, Van Delft FW, Treleaven J, Georgiades P, Uglow E, Asou N, Uike N, Debeljak M, Jazbec J, Ancliff P, Gale R, Thomas X, Mialou V, Döhner K, Bullinger L, Mueller B, Pabst T, Stelljes M, Schlegelberger B, Wozniak E, Iqbal S, Okosun J, Araf S, Frank AK, Lauridsen FB, Porse B, Nerlov C, Owen C, Dokal I, Gribben J, Smith M, Preudhomme C, Chelala C, Cavenagh J, Fitzgibbon J. Disease evolution and outcomes in familial AML with germline CEBPA mutations. Blood 2015; 126, 1214-1223.

Tawana K, Fitzgibbon J. Inherited DDX41 mutations: 11 genes and counting. Blood. 2016 Feb 25;127(8):960-961.

Tawana K, Rio-Machin A, Preudhomme C, Fitzgibbon J. Familial CEBPA-mutated acute myeloid leukemia. Semin Hematol. 2017 Apr;54(2):87-93.

Tefferi A & Vardiman JW. Myelodysplastic syndromes. The New England Journal of Medicine 2009; 361, 1872-1885.

Thomas AQ, Lane K, Phillips J 3rd, Prince M, Markin C, Speer M, Schwartz DA, Gaddipati R, Marney A, Johnson J, Roberts R, Haines J, Stahlman M, Loyd JE. Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspecific interstitial pneumonitis in one kindred. Am J Respir Crit Care Med. 2002 May 1;165(9):1322-1328.

Thompson MR, Kaminski JJ, Kurt-Jones EA, Fitzgerald KA. Pattern recognition receptors and the innate immune response to viral infection. Viruses. 2011 Jun;3(6):920-940.

Thrasher AJ, Burns SO. WASP: a key immunological multitasker. Nat Rev Immunol. 2010 Mar;10(3):182-192.

Tingting P, Caiyun F, Zhigang Y, Pengyuan Y, Zhenghong Y. Subproteomic analysis of the cellular proteins associated with the 3' untranslated region of the hepatitis C virus genome in human liver cells. Biochem Biophys Res Commun. 2006 Sep 1;347(3):683-691.

Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015 Jan 2;347(6217):78-81.

Topka S, Vijai J, Walsh MF, Jacobs L, Maria A, Villano D, Gaddam P, Wu G, McGee RB, Quinn E, Inaba H, Hartford C, Pui CH, Pappo A, Edmonson M, Zhang MY, Stepensky P, Steinherz P, Schrader K, Lincoln A, Bussel J, Lipkin SM, Goldgur Y, Harit M, Stadler ZK, Mullighan C, Weintraub M, Shimamura A, Zhang J, Downing JR, Nichols KE, Offit K. Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. PLoS Genet. 2015 Jun 23;11(6):e1005262.

Traina F, Visconte V, Elson P, Tabarroki A, Jankowska AM, Hasrouni E, Sugimoto Y, Szpurka H, Makishima H, O'Keefe CL, Sekeres MA, Advani AS, Kalaycio M, Copelan EA, Sauntharajah Y, Olalla Saad ST, Maciejewski JP, Tiu RV. Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms. Leukemia. 2014 Jan;28(1):78-87.

Trapnell C & Salzberg SL. How to map billions of short reads onto genomes. Nat Biotechnol 2009; 27, 455–457.

Tsang HC, Bussel JB, Mathew S, Liu YC, Imahiyerobo AA, Orazi A, Geyer JT. Bone marrow morphology and disease progression in congenital thrombocytopenia: a detailed clinicopathologic and genetic study of eight cases. Mod Pathol. 2017 Apr;30(4):486-498.

Tsakiri KD, Cronkhite JT, Kuan PJ, Xing C, Raghu G, Weissler JC, Rosenblatt RL, Shay JW, Garcia CK. Adult-onset pulmonary fibrosis caused by mutations in telomerase. Proc Natl Acad Sci U S A. 2007 May 1;104(18):7552-7557.

Tummala H, Walne A, Collopy L, Cardoso S, de la Fuente J, Lawson S, Powell J, Cooper N, Foster A, Mohammed S, Plagnol V, Vulliamy T, Dokal I. Poly(A)-specific

ribonuclease deficiency impacts telomere biology and causes dyskeratosis congenita. J Clin Invest. 2015 May;125(5):2151-2160.

Tummala H, Walne AJ, Williams M, Bockett N, Collopy L, Cardoso S, Ellison A, Wynn R, Leblanc T, Fitzgibbon J, Kelsell DP, van Heel DA, Payne E, Plagnol V, Dokal I, Vulliamy T. DNAJC21 Mutations Link a Cancer-Prone Bone Marrow Failure Syndrome to Corruption in 60S Ribosome Subunit Maturation. Am J Hum Genet. 2016 Jul 7;99(1):115-124.

Umemura H, Yamasaki O, Iwatsuki K. Leukocytoclastic vasculitis associated with immunoglobulin A lambda monoclonal gammopathy of undetermined significance: A case report and review of previously reported cases. J Dermatol. 2018 May 15. doi: 10.1111/1346-8138.14466. [Epub ahead of print].

University of Chicago Hematopoietic Malignancies Cancer Risk Team. How I diagnose and manage individuals at risk for inherited myeloid malignancies. Blood. 2016 Oct 6;128(14):1800-1813.

Unterholzner L, Keating SE, Baran M, Horan KA, Jensen SB, Sharma S, Sirois CM, Jin T, Latz E, Xiao TS, Fitzgerald KA, Paludan SR, Bowie AG. IFI16 is an innate immune sensor for intracellular DNA. Nat Immunol. 2010 Nov;11(11):997-1004.

Uringa EJ, Youds JL, Lisaingo K, Lansdorp PM, Boulton SJ. RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination. Nucleic Acids Res. 2011; 39 (5):1647-1655.

van der Lelij P, Chrzanowska KH, Godthelp BC, Rooimans MA, Oostra AB, Stumm M, Zdzienicka MZ, Joenje H, de Winter JP. Warsaw breakage syndrome, a cohesinopathy associated with mutations in the XPD helicase family member DDX11/ChlR1. Am J Hum Genet. 2010 Feb 12;86(2):262-266.

Van Esch H, Groenen P, Nesbit MA, Schuffenhauer S, Lichtner P, Vanderlinden G, Harding B, Beetz R, Bilous RW, Holdaway I, Shaw NJ, Fryns JP, Van de Ven W, Thakker RV, Devriendt K. GATA3 haplo-insufficiency causes human HDR syndrome. Nature. 2000 Jul 27;406(6794):419-422.

Vannier JB, Pavicic-Kaltenbrunner V, Petalcorin MI, Ding H, Boulton SJ. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. Cell. 2012; 149(4):795-806.

Vannier JB, Sandhu S, Petalcorin MI, Wu X, Nabi Z, Ding H, Boulton SJ. RTEL1 is a replisome-associated helicase that promotes telomere and genome-wide replication. Science. 2013 Oct 11;342(6155):239-242.

Vannier JB, Sarek G, Boulton SJ. RTEL1: functions of a disease-associated helicase. Trends Cell Biol. 2014 Jul;24(7):416-425.

Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, Crawford GE, Absher DM, Wold BJ, Myers RM. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 2013 Mar;23(3):555-567.

Vega LR, Mateyak MK, Zakian VA. Getting to the end: telomerase access in yeast and humans. Nat Rev Mol Cell Biol. 2003 Dec;4(12):948-959.

Verdun RE, Karlseder J. The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres. Cell. 2006 Nov 17;127(4):709-720.

Vinh DC, Patel SY, Uzel G, Anderson VL, Freeman AF, Olivier KN, Spalding C, Hughes S, Pittaluga S, Raffeld M, Sorbara LR, Elloumi HZ, Kuhns DB, Turner ML, Cowen EW, Fink D, Long-Priel D, Hsu AP, Ding L, Paulson ML, Whitney AR, Sampaio EP, Frucht DM, DeLeo FR, Holland SM. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. Blood. 2010 Feb 25;115(8):1519-1529.

Vincenot A, Hurtaud-Roux MF, René O, Binard S, Fenneteau O, Schlegel N. ANKRD26 normocytic thrombocytopenia: a family report. Ann Biol Clin (Paris). 2016 Jun 1;74(3):317-322.

Vlaanderen J, Portengen L, Rothman N, Lan Q, Kromhout H, Vermeulen R. Flexible meta-regression to assess the shape of the benzene-leukemia exposure-response curve. Environmental Health Perspectives 2010; 118: 526–532.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013 Mar 29;339(6127):1546-1558.

von Zglinicki T. Oxidative stress shortens telomeres. Trends Biochem Sci. 2002 Jul;27(7):339-344.

Vulliamy T, Marrone A, Goldman F, Dearlove A, Bessler M, Mason PJ, Goldman F, Dearlove A & Dokal I. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature 2001; 413, 432-435.

Vulliamy T, Marrone A, Szydlo R, Walne A, Mason PJ, Dokal I. Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. Nat Genet. 2004 May;36(5):447-449.

Vulliamy TJ, Marrone A, Knight SW, Walne A, Mason PJ, Dokal I. Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation. Blood. 2006 Apr 1;107(7):2680-2685.

Walne AJ, Dokal I. Dyskeratosis Congenita: a historical perspective. Mech Ageing Dev. 2008 Jan-Feb;129(1-2):48-59.

Walne AJ, Vulliamy T, Kirwan M, Plagnol V, Dokal I. Constitutional mutations in RTEL1 cause severe dyskeratosis congenita. Am J Hum Genet. 2013a; 92(3):448-453.

Walne AJ, Bhagat T, Kirwan M, Gitiaux C, Desguerre I, Leonard N, Nogales E, Vulliamy T, Dokal IS. Mutations in the telomere capping complex in bone marrow failure and related syndromes. Haematologica. 2013b Mar;98(3):334-338.

Walne AJ, Collopy L, Cardoso S, Ellison A, Plagnol V, Albayrak C, Albayrak D, Kilic SS, Patıroglu T, Akar H, Godfrey K, Carter T, Marafie M, Vora A, Sundin M, Vulliamy

T, Tummala H, Dokal I. Marked overlap of four genetic syndromes with dyskeratosis congenita confounds clinical diagnosis. Haematologica. 2016 Oct;101(10):1180-1189.

Walsh R, Thomson KL, Ware JS, Funke BH, Woodley J, McGuire KJ, Mazzarotto F, Blair E, Seller A, Taylor JC, Minikel EV, Exome Aggregation Consortium, MacArthur DG, Farrall M, Cook SA, Watkins H. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. Genet Med. 2017 Feb;19(2):192-203.

Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, Larson DE, McLellan MD, Dooling D, Abbott R, Fulton R, Magrini V, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Fan X, Grillot M, Witowski S, Heath S, Frater JL, Eades W, Tomasson M, Westervelt P, DiPersio JF, Link DC, Mardis ER, Ley TJ, Wilson RK, Graubert TA. Clonal architecture of secondary acute myeloid leukemia. N Engl J Med. 2012 Mar 22;366(12):1090-1098.

Wang RC, Smogorzewska A, de Lange T. Homologous recombination generates T-loop-sized deletions at human telomeres. Cell. 2004 Oct 29;119(3):355-368.

Wang Y, Kuan PJ, Xing C, Cronkhite JT, Torres F, Rosenblatt RL, DiMaio JM, Kinch LN, Grishin NV, Garcia CK. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. Am J Hum Genet. 2009 Jan;84(1):52-59.

Wang C, Lu J, Wang Y, Bai S, Wang Y, Wang L, Sheng G. Combined effects of FLT3 and NF-κB selective inhibitors on acute myeloid leukemia in vivo. J Biochem Mol Toxicol. 2012 Jan;26(1):35-43.

Wang Z, Wu Y, Li L, Su XD. Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA. Cell Res. 2013 Feb;23(2):213-224.

Warbrick E. PCNA binding through a conserved motif. Bioessays. 1998 Mar;20(3):195-199.

Warren AJ. Molecular basis of the human ribosomopathy Shwachman-Diamond syndrome. Adv Biol Regul. 2017 Sep 6. pii: S2212-4926(17)30153-30157.

Wechsler J, Greene M, McDevitt MA, Anastasi J, Karp JE, Le Beau MM, Crispino JD. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nat Genet. 2002 Sep;32(1):148-152.

Weis F, Giudice E, Churcher M, Jin L, Hilcenko C, Wong CC, Traynor D, Kay RR, Warren AJ. Mechanism of eIF6 release from the nascent 60S ribosomal subunit. Nat Struct Mol Biol. 2015 Nov;22(11):914-99.

Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J, Kandoth C, Fulton RS, McLellan MD, Dooling DJ, Wallis JW, Chen K, Harris CC, Schmidt HK, Kalicki-Veizer JM, Lu C, Zhang Q, Lin L, O'Laughlin MD, McMichael JF, Delehaunty KD, Fulton LA, Magrini VJ, McGrath SD, Demeter RT, Vickery TL, Hundal J, Cook LL, Swift GW, Reed JP, Alldredge PA, Wylie TN, Walker JR, Watson MA, Heath SE, Shannon WD, Varghese N, Nagarajan R, Payton JE, Baty JD, Kulkarni S, Klco JM, Tomasson MH, Westervelt

P, Walter MJ, Graubert TA, DiPersio JF, Ding L, Mardis ER, Wilson RK. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012 Jul 20;150(2):264-278.

West AH, Churpek JE. Old and new tools in the clinical diagnosis of inherited bone marrow failure syndromes. Hematology Am Soc Hematol Educ Program. 2017 Dec 8;2017(1):79-87.

West AH, Godley LA, Churpek JE. Familial myelodysplastic syndrome/acute leukemia syndromes: a review and utility for translational investigations. Ann N Y Acad Sci. 2014 Mar;1310:111-118.

Wilson JS, Tejera AM, Castor D, Toth R, Blasco MA, Rouse J. Localization-dependent and -independent roles of SLX4 in regulating telomeres. Cell Rep. 2013 Sep 12;4(5):853-860.

Wong TN, Ramsingh G, Young AL, Miller CA, Touma W, Welch JS, Lamprecht TL, Shen D, Hundal J, Fulton RS, Heath S, Baty JD, Klco JM, Ding L, Mardis ER, Westervelt P, DiPersio JF, Walter MJ, Graubert TA, Ley TJ, Druley T, Link DC, Wilson RK. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. Nature. 2015 Feb 26;518(7540):552-555.

Worth AJ, Thrasher AJ. Current and emerging treatment options for Wiskott-Aldrich syndrome. Expert Rev Clin Immunol. 2015;11(9):1015-1032.

Wortham NC, Ahamed E, Nicol SM, Thomas RS, Periyasamy M, Jiang J, Ochocka AM, Shousha S, Huson L, Bray SE, Coombes RC, Ali S, Fuller-Pace FV. The DEAD-box protein p72 regulates ERalpha-/oestrogen-dependent transcription and cell growth, and is associated with improved survival in ERalpha-positive breast cancer. Oncogene. 2009 Nov 19;28(46):4053-4064.

Wu DW, Liu WS, Wang J, Chen CY, Cheng YW, Lee H. Reduced p21(WAF1/CIP1) via alteration of p53-DDX3 pathway is associated with poor relapse-free survival in early-stage human papillomavirus-associated lung cancer. Clin Cancer Res. 2011 Apr 1;17(7):1895-1905.

Xiao W & Oefner PJ. Denaturing high-performance liquid chromatography: a review. Hum Mutat 2001; 17, 439-474.

Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, McMichael JF, Schmidt HK, Yellapantula V, Miller CA, Ozenberger BA, Welch JS, Link DC, Walter MJ, Mardis ER, Dipersio JF, Chen F, Wilson RK, Ley TJ, Ding L. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med. 2014 Dec;20(12):1472-1478.

Yamaguchi H, Baerlocher GM, Lansdorp PM, Chanock SJ, Nunez O, Sloand E & Young SN. Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome. Blood 2003; 102, 916-918.

Yamaguchi H, Calado RT, Ly H, Kajigaya S, Baerlocher GM, Chanock SJ, Lansdorp PM, Young NS. Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. N Engl J Med. 2005 Apr 7;352(14):1413-1424.

Yamamoto M, Sato S, Mori K, Hoshino K, Takeuchi O, Takeda K, Akira S. Cutting edge: a novel Toll/IL-1 receptor domain-containing adapter that preferentially activates the IFN-beta promoter in the Toll-like receptor signaling. J Immunol. 2002 Dec 15;169(12):6668-6672.

Yang L, Lin C, Zhao S, Wang H, Liu ZR. Phosphorylation of p68 RNA helicase plays a role in platelet-derived growth factor-induced cell proliferation by up-regulating cyclin D1 and c-Myc expression. J Biol Chem. 2007 Jun 8;282(23):16811-16819.

Yoshimi A, Toya T, Nannya Y, Takaoka K, Kirito K, Ito E, Nakajima H, Hayashi Y, Takahashi T, Moriya-Saito A, Suzuki K, Harada H, Komatsu N, Usuki K, Ichikawa M, Kurokawa M. Spectrum of clinical and genetic features of patients with inherited platelet disorder with suspected predisposition to hematological malignancies: a nationwide survey in Japan. Ann Oncol. 2016 May;27(5):887-895.

Youds JL, Mets DG, McIlwraith MJ, Martin JS, Ward JD, ONeil NJ, Rose AM, West SC, Meyer BJ, Boulton SJ. RTEL-1 enforces meiotic crossover interference and homeostasis. Science. 2010 Mar 5;327(5970):1254-1258.

Yount JS, Gitlin L, Moran TM, López CB. MDA5 participates in the detection of paramyxovirus infection and is essential for the early activation of dendritic cells in response to Sendai Virus defective interfering particles. J Immunol.2008 Apr 1;180(7):4910-4918.

Zaninetti C, Santini V, Tiniakou M, Barozzi S, Savoia A, Pecci A. Inherited thrombocytopenia caused by ANKRD26 mutations misdiagnosed and treated as myelodysplastic syndrome: report on two cases. J Thromb Haemost. 2017 Dec;15(12):2388-2392.

Zellinger B, Akimcheva S, Puizina J, Schirato M, Riha K. Ku suppresses formation of telomeric circles and alternative telomere lengthening in Arabidopsis. Mol Cell. 2007 Jul 6;27(1):163-169.

Zhang J, Chen X. Posttranscriptional regulation of p53 and its targets by RNA-binding proteins. Curr Mol Med. 2008 Dec;8(8):845-849.

Zhang MY, Churpek JE, Keel SB, Walsh T, Lee MK, Loeb KR, Gulsuner S, Pritchard CC, Sanchez-Bonilla M, Delrow JJ, Basom RS, Forouhar M, Gyurkocza B, Schwartz BS, Neistadt B, Marquez R, Mariani CJ, Coats SA, Hofmann I, Lindsley RC, Williams DA, Abkowitz JL, Horwitz MS, King MC, Godley LA, Shimamura A. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. Nature Genetics 2015; 47:180-185.

Zhang X, Lancet JE, Zhang L. Molecular pathology of myelodysplastic syndromes: new developments and implications for diagnosis and treatment. Leuk Lymphoma. 2015;56(11):3022-3030.

Zhang Z, Kim T, Bao M, Facchinetti V, Jung SY, Ghaffari AA, Qin J, Cheng G, Liu YJ. DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. Immunity. 2011(a) Jun 24;34(6):866-878.

Zhang Z, Yuan B, Bao M, Lu N, Kim T, Liu YJ. The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. Nat Immunol. 2011(b) Sep 4;12(10):959-965.

Zhang Z, Yuan B, Lu N, Facchinetti V, Liu YJ. DHX9 pairs with IPS-1 to sense double-stranded RNA in myeloid dendritic cells. J Immunol. 2011(c) Nov 1;187(9):4501-4508.

Zhang Z, Bao M, Lu N, Weng L, Yuan B, Liu YJ. The E3 ubiquitin ligase TRIM21 negatively regulates the innate immune response to intracellular double-stranded DNA. Nat Immunol. 2013 Feb;14(2):172-178.

Zhao X, Tian X, Kajigaya S, Cantilena CR, Strickland S, Savani BN, Mohan S, Feng X, Keyvanfar K, Dunavin N, Townsley DM, Dumitriu B, Battiwalla M, Rezvani K, Young NS, Barrett AJ, Ito S. Epigenetic landscape of the TERT promoter: a potential biomarker for high risk AML/MDS. Br J Haematol. 2016 Nov;175(3):427-439.

Zhu L, Hathcock KS, Hande P, Lansdorp PM, Seldin MF, Hodes RJ. Telomere length regulation in mice is linked to a novel chromosome locus. Proc Natl Acad Sci U S A. 1998 Jul 21;95(15):8648-8653.

Zhu XD, Niedernhofer L, Kuster B, Mann M, Hoeijmakers JH, de Lange T. ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. Mol Cell. 2003 Dec;12(6):1489-1498.

Zipprich JT, Bhattacharyya S, Mathys H, Filipowicz W. Importance of the C-terminal domain of the human GW182 protein TNRC6C for translational repression. RNA. 2009 May;15(5):781-793.