The Molecular Investigation of Familial Leukaemia

Kiran Tawana

Queen Mary, University of London

Submitted in fulfilment of the requirements of the degree of Doctor of Philosophy

December 2015

Statement of Originality

I, Kiran Tawana, 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. 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.

Details of Collaboration:

- Dr Jun Wang, Centre for Molecular Oncology, Barts Cancer Institute performed bioinformatic analysis of whole exome and deep sequencing data
- Dr Robert Hills, Cardiff University, performed the statistical analysis comparing sporadic and familial *CEBPA*-mutated cases, detailed in Chapter 3.
- Array CGH experiments were performed by Marcel Tauscher at The Institute of Human Genetics, Hannover University Medical School, Germany. I observed and assisted with these experiments and the subsequent analysis.
- Sanger sequencing experiments detailed in Chapter 4 were performed in collaboration with Dr Csaba Bödör and Péter Király, Semmelweis University, Budapest
- Sanger sequencing verification of variants in Chapters 3 and 5 were assisted by Panayiotis Alex Georgiades, as part of his MSc work, and Dr Elizabeth Uglow, an academic clinical trainee in Haemato-Oncology.
- Novel germline lesions detected in patients from the UK were verified in accredited laboratories at Gt. Ormond St. Hospital (Dr Alison Male) and the West Midlands Regional Genetics Laboratory (Dr Susanna Akiki and Nicola Trim).

Details of Publications:

- Tawana K, Wang J, Renneville A, *et al*. Disease evolution and outcomes in familial AML with germline CEBPA mutations. Blood. 2015;126(10):1214-1223.
- Green CL, Tawana K, Hills RK, *et al*. GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations. Br J Haematol. 2013;161(5):701-705
- Mutsaers PG, van de Loosdrecht AA, Tawana K *et al*. Highly variable clinical manifestations in a large family with a novel GATA2 mutation. Leukemia. 2013;27(11):2247-2248.
- Tawana K, Bödör C, Cavenagh J *et al*. The Molecular Pathogenesis of Acute Myeloid Leukaemia. CML – Leukemia & Lymphoma 2013;21(3):67–77

Signature: 17th December 2015

Abstract

Acute myeloid leukaemia (AML) and myelodysplastic syndrome (MDS) are clonal disorders of haematopoiesis which culminate in bone marrow failure. In rare cases, families demonstrate an autosomal dominant predisposition to these disorders, associated with germline mutations in the haematopoietic transcription factors *RUNX1*, *CEBPA*, *GATA2* and *ETV6*. This thesis investigates the molecular profiling of tumours from multiple families with different germline mutations, examining longitudinal patterns of disease evolution and identifying novel candidate germline lesions.

Studies initially focused on the comprehensive characterisation of AML associated with germline *CEBPA* mutations. In-depth molecular profiling was performed to detail both the genetic events initiating leukaemia and the clinical progression of disease across multiple families. Integrated whole-exome (WES) and deep sequencing analysis revealed the presentation of entirely new leukaemic episodes initiating disease recurrence. This represents a unique model of disease progression in AML and clinical data support this hypothesis by demonstrating continued sensitivity to secondary therapies, with some patients experiencing several episodes of AML recurrence over decades.

Convergence of tumour profiles was demonstrated across relatives from two families with different germline *CEBPA* mutations. To further explore this phenomenon, tumours from multiple siblings within a *RUNX1*-mutated family were investigated with WES. Three of the four siblings presented in an identical manner, all developing high risk MDS or secondary AML at 5 years of age. The molecular disease profiles were remarkable, as these 3 siblings all demonstrated somatic mutations causing upregulation of JAK-STAT signalling. This pattern of intra-familial disease convergence has not previously been reported in haematopoietic or solid familial tumours and highlights the importance of host genetic factors in governing somatic mutation acquisition.

The final chapter of this thesis describes the investigation of novel germline lesions including partial allelic deletions of *RUNX1* and atypical *CEBPA* and *GATA2* mutations, with clinical manifestations varying significantly within these families. Novel genetic

candidates are also discussed, highlighting the need for further work to demonstrate recurrence of these lesions.

To conclude, this work reveals novel insights into the heterogeneous entity of familial leukaemia. By revealing unique patterns of disease evolution and intra-familial tumour convergence, these studies highlight the distinct properties of germline mutations in governing the biology of familial leukaemia. Further work is required to characterise the somatic and constitutional molecular landscape and determine the precise mechanisms governing disease latency, penetrance and phenotype.

Dedication

To all the patients and families who have contributed to this project and the many others who may be helped by ongoing work in this field.

Acknowledgements

I would like to thank everyone who has helped make this project possible over the years. First and foremost, I am very grateful to Cancer Research UK who funded my three-year Clinical fellowship and Children with Cancer and the Elimination of Leukaemia Fund for their additional funding of this project. The rarity of familial leukaemia provided the opportunity to collaborate with many groups who share this common interest and I am immensely grateful to all for their help, special thanks go to Prof. Brigitte Schlegelberger, Dr Doris Steinemann and Marcel Tauscher for their kindness and generosity during my time in Hannover, Germany. Closer to home, special thanks go to Prof Jamie Cavenagh and Dr Matt Smith at St. Bartholomew's Hospital for their invaluable collaboration and support, without their initial discovery the project would not have been possible! I would also like to thank Dr Jun Wang and Claude Chelala for their expertise with the bioinformatic analysis and Graham Clark and the sequencing team at the LRI for their tremendous help with Sanger sequencing and MLPA. Further thanks go to Dr Robert Hills for the statistical comparison of sporadic and familial *CEBPA*-mutated AML outcome data, described in Chapter 3. Above all, I would like to thank the many patients and families who have kindly donated samples for research, often in the most difficult times. It was seeing the overwhelming impact of this syndrome upon families that first prompted me to take up this project and has provided a constant source of motivation ever since.

A huge thank you goes to my supervisor Prof Jude Fitzgibbon for his guidance, encouragement and friendship and also to my second supervisor Prof John Gribben for his direction and guidance throughout my time at Barts. It has been a pleasure to work with everyone in the team: Csaba Bödör, Jessica Okosun, Shamzah Araf, Lola Koniali, Nik Popov, James Heward and Chey Loveday. Extended thanks go to Ana Rio-Machin and Ahad Al Seraihi who have recently joined the group and have been exceptionally supportive during the final stages of my lab work and write-up. I would also like to thank everyone in the Centre for Haemato-Oncology, in particular the Tissue Bank staff under the management of Sameena Iqbal and Dave Williamson, for his unrivalled knowledge of the lab.

Finally, I would like to thank my friends and family for their unwavering kindness, consideration and patience. Special thanks go to Dr Áine McCarthy and to Jean-Baptise Blanc for making my time in London so enjoyable. Mum and Dad, thank you for everything.

Table of Contents

Statement of Originality ---- 2

Abstract ---- 4

Dedication ---- 6

Acknowledgements ---- 7

Table of Contents ---- 8

List of Figures ---- 14

List of Tables ---- 18

Abbreviations ---- 18

Chapter 1 Introduction ---- 21

1 Introduction ---- 22

1.1 Acute myeloid leukaemia ---- 23

1.1.1 Incidence and clinical presentation of AML ---- 23

1.1.2 Classification of AML ---- 23

1.1.3 Cytogenetic risk groups in AML ---- 26

1.1.4 Outline of current treatment approaches in AML ---- 27

1.1.5 Definition of remission in AML ---- 29

1.1.6 Determinants of clinical outcomes in AML ---- 30

1.1.7 Outcomes of AML post relapse ---- 30

1.2 Myelodysplastic syndrome ---- 32

1.2.1 Classification and prognostic markers in MDS ---- 32

1.2.2 Treatment strategies in MDS ---- 35

1.3 The evolving molecular landscape of MDS and AML ---- 36

1.3.1 FMS-like tyrosine kinase ---- 36

1.3.2 Nucleophosmin 1 ---- 37

1.3.3 CCAAT/enhancer-binding protein α ---- 38

1.3.4 Expansion of the molecular landscape in MDS and AML ---- 38

1.4 Clonal architecture and evolution determined by whole genome/exome sequencing ---------- 47

1.5 The cell of origin in AML ---------- 49

1.5.1 Overview of haematopoiesis 49

1.5.2 Characterisation of leukaemia initiating cells 51

1.5.3 Pre-leukaemic mutations in LICs frequently target epigenetic modifiers 51

1.6 Leukaemia predisposition syndromes ---------- 54

1.6.1 The molecular basis of familial MDS/AML 54

1.6.2 Runt-related transcription factor 1 54

1.6.3 Alternative genetic lesions associated with thrombocytopenia and predisposition to malignancy 59

1.6.4 Familial AML with *CEBPA* mutations 61

1.6.5 Familial syndromes associated with *GATA2* mutations 63

1.6.6 Germline *DDX41* mutations 66

1.6.7 Inherited bone marrow failure syndromes 67

1.7 The introduction of next generation sequencing in cancer genomics ---------- 69

1.7.1 Novel technologies for discovery genomics 69

1.7.2 WES in the investigation of familial leukaemia 70

1.8 Aims and Objectives ---------- 71

Chapter 2 Materials and Methods ---------- 72

2.1 Patients and sample collection ---------- 73

2.2 DNA extraction ---------- 75

2.2.1 Genomic DNA extraction 75

2.2.2 DNA extraction from BM aspirate slides 75

2.2.3 Salivary DNA extraction 75

2.3 Nucleic acid quantification ---------- 76

2.4 Screening of established familial MDS/AML loci ---------- 77

2.4.1 Polymerase chain reaction 77

2.5 Sanger sequencing ---------- 82

2.6 Identification of mutations and validation ---------- 82

2.7 Cloning of PCR products into competent *E.coli* ---------- 82

2.8 Familial MDS/AML customised aCGH protocol ---------- 85

2.9 Multiplex ligation-dependent probe amplification ---------- 87

2.9.1 Principles of MLPA 87

2.9.2 MLPA Experimental method 90

2.9.3 MLPA data analysis and interpretation 91

2.10 Whole-exome sequencing ---------- 93

2.10.1 Bioinformatic analysis of WES data 93

2.11 Filtering and prioritisation of WES germline variants ---------- 97

2.11.1 The Exome Aggregation Consortium Browser 97

2.11.2 Mutation prediction algorithms 98

2.12 Targeted deep sequencing of sequential *CEBPA*-mutated tumour samples -------- 99

Chapter 3 Clinical and Molecular Profiling of Families with Germline *CEBPA* Mutations ---------- 103

3.1 Introduction ---------- 104

3.2 Aims and Objectives ---------- 105

3.3 Methods ---------- 106

3.3.1 Collection of familial AML samples 106

3.3.2 Sanger sequencing of *CEBPA* mutations 108

3.3.3 Whole-exome sequencing of familial AML samples 108

3.3.4 Mutation analysis of WES reads 109

3.3.5 Exclusion of pre-leukaemic mutations in tumour and remission DNA 109

3.3.6 Tumour subpopulation identification and clonality analysis 110

3.3.7 Integrated whole-exome and deep sequencing analysis of clonal evolution 110

3.3.8 Clinical profiling of an extended familial AML cohort 110

3.4 Results ---------- **113**

3.4.1 Clinical feature of familial AML cohort 113

3.4.2 Whole-exome profiling of familial AML tumours 116

3.4.3 WES reveals recurrent intra-familial somatic mutations 120

3.4.4 Instability of somatic *CEBPA* mutations at disease recurrence 122

3.4.5 Deep sequencing confirms the tumour specificity of somatic *CEBPA* mutations 125

3.4.6 Molecular profiling reveals two distinct models of AML recurrence 125

3.5 Limitations of Study ---------- **132**

3.6 Discussion ---------- **134**

3.6.1 A new model of disease progression in familial AML 134

3.6.2 Convergence of disease evolution in family members 135

3.6.3 Conclusion 136

Chapter 4 ---------- **137**

Convergence of somatic JAK-STAT mutations in a novel *RUNX1*-mutated pedigree -- **137**

4.1 Introduction ---------- **138**

4.2 Aims and Objectives ---------- **139**

4.3 Materials and Methods ---------- **140**

4.3.1 DNA extraction and Sanger sequencing 140

4.3.2 In-depth genetic profiling of tumour samples 140

4.4 Results ---------- **143**

4.4.1 Identification of the germline *RUNX1* R201X mutation 143

4.4.2 Clinical profiles of siblings with MDS/AML 143

4.4.3 WES metrics for pedigree K 146

4.4.4 Molecular variants detected by WES 147

4.4.5 Chromosomal LOH within tumour samples 149

4.4.6 Copy-neutral LOH increases mutation dosage 151

4.4.7 Deciphering the sequence of genetic events in K.II.1 and K.II.2 155

4.4.8 Clonal evolution at relapse in K.II.4 157

4.4.9 Clinical outcomes in family K 159

4.4.10 *JAK2* haplotype analysis 160

4.4.11 Sporadic AML with *RUNX1* and *JAK2* mutations 162

4.5 Limitations of study 163

4.6 Discussion 164

4.6.1 JAK-STAT signalling activation within Pedigree K 164

4.6.2 JAK-STAT mutations may co-operate with RUNX1 dysfunction 166

4.6.3 Non-random acquisition of mutations in family K 167

4.6.4 Mutation dosage is increased by chromosomal LOH 168

4.6.5 Clonal evolution in sibling K.II.4 168

4.6.7 Conclusion 169

Chapter 5 Novel and Atypical Germline Mutations in Familial AML 170

5.1 Introduction 171

5.2 Aims and Objectives 172

5.3 Materials and Methods 173

5.3.1 DNA Extraction and Sanger Sequencing 173

5.3.2 Array CGH 173

5.3.3 MLPA 173

5.3.4 SNP array 174

5.3.5 WES to detect novel germline variants 174

5.3.5 Families investigated for novel and atypical variants 174

5.4 Results 176

5.4.1 Families with germline *RUNX1* deletions 176

5.4.2 Detection of *RUNX*1 deletions using array CGH 188

5.4.3 Clinical phenotypes associated with *RUNX1* deletions 183

5.4.4 Atypical germline mutations in *GATA2* and *CEBPA* 183

5.4.5 Families investigated for novel germline mutations 188

5.4.6 Molecular variants prioritised across Pedigrees P-T 193

5.5 Limitations of this study ---------- **205**

5.6 Discussion ---------- **206**

5.6.1 Familial platelet disorder and myeloid malignancy with *RUNX1* deletions..... 206

5.6.2 Novel disease insights with atypical lesions in *GATA2* and *CEBPA*..... 207

5.6.3 The challenges of investigating unknown germline mutations..... 208

5.6.4 Conclusion..... 209

Chapter 6 Discussion ---------- **210**

6 Discussion ---------- **211**

6.1 Models of disease progression in sporadic and familial AML ---------- **211**

6.2 A new model of disease progression in familial AML ---------- **212**

6.3 The molecular convergence of tumours within families ---------- **214**

6.4 Challenges in the clinical recognition of leukaemia predisposition syndromes ---- **215**

6.5 Suggestions for future work ---------- **217**

6.6 Conclusion ---------- **217**

References ---------- **218**

Appendix 1 ---------- **256**

Appendix 2 ---------- **258**

Appendix 3 ---------- **262**

Appendix 4 ---------- **264**

List of Figures

Chapter 1 Introduction

Figure.1.1 MRC trial data showing the impact of cytogenetic and molecular characterisation on survival outcomes for young adults with AML. 28
Figure 1.2 Diagrammatic representation of components of the spliceosome 43
Figure 1.3 Current models of clonal evolution in AML 48
Figure 1.4 Diagram of haematopoiesis showing cell-markers at HSC and progenitor stages. 50
Figure 1.5 Current theory of leukaemia initiation 53
Figure 1.6 Distribution of germline *RUNX1* mutations in families with leukaemia predisposition and familial platelet disorder. 58
Figure 1.7 Distribution of *GATA2* mutations associated with leukaemia predisposition 65

Chapter 2 Materials and Methods

Figure 2.1 Overview of TA-cloning procedure 84
Figure 2.2 Overview of array CGH procedure. 86
Figure 2.3 Overview of MLPA experimental procedure 92

Chapter 3 Clinical and Molecular Profiling of families with germline *CEBPA* mutations

Figure 3.1 Genealogical trees of families with germline *CEBPA* mutations..... 108
Figure 3.2 Familial AML pedigrees and timeline of clinical events. 116
Figure 3.3 Molecular profiling of familial AML with germline *CEBPA* mutations 119
Figure 3.4 Recurrent genetic lesions acquired within family members with germline *CEBPA* mutations. 121
Figure 3.5 Sanger sequencing reveals variation of somatic *CEBPA* mutations at recurrence. 124
Figure 3.6 Molecular and copy number profiling of sequential tumour episodes 127
Figure 3.7 Deep sequencing of 28 variants to define pattern of clonal evolution in B.I.1. . 129
Figure 3.8 Deep sequencing of sequential disease episodes in B.I.1..... 130
Figure 3.9 Model of disease progression in B.I.1 131

Chapter 4 Convergence of somatic JAK-STAT mutations in a novel *RUNX1*-mutated pedigree

Figure 4.1 Characterisation of a novel *RUNX1*-mutated pedigree. 144
Figure 4.2 Somatic mutations and LOH in tumour samples from family K. 152
Figure 4.3 Somatic CNAs and aUPD detected in twin-sibling tumours. 154
Figure 4.4 Sequential acquisition of CNAs and aUPD during leukaemic evolution. 156
Figure 4.5 Corrected VAFs for *U2AF2* and *JAK2* mutations at diagnosis (T1) and relapse (T2) in K.II.4. 158
Figure 4.6 Sanger sequencing of *JAK2* 46/1 tagged SNP rs12340895 in pedigree K. 161

Chapter 5 Novel and Atypical Germline Mutations in Familial AML

Figure 5.1 Integrated genetic investigation to identify novel germline lesions. 175
Figure 5.2 Pedigrees and disease phenotypes of cases demonstrating *RUNX1* deletions. . 176
Figure 5.3 *RUNX1* deletions detected in pedigrees L,M and N. 180
Figure 5.4 Verification of *RUNX1* deletions and alignment to reference genome. 181
Figure 5.5 Compound germline *GATA2* mutation in a family with varying clinical phenotypes 185
Figure 5.6 Detection of a germline *CEBPA* C-terminal mutation. 187
Figure 5.7 Families investigated for novel germline mutations using WES. 188
Figure 5.8 Segregation of germline variants across affected individuals in pedigree P 195
Figure 5.9 Schematic representation of the NuRD complex. 198
Figure 5.10 Somatic mutations and timeline of disease events in Pedigree P. 200

List of Tables

Chapter 1 Introduction

Table 1.1 French-American-British Classification of AML 25
Table 1.2 WHO Classification of AML and related neoplasms 25
Table 1.3 WHO Classification of MDS and MDS/MPN 34
Table 1.4 IPSS-R Cytogenetic classification of MDS 34
Table 1.5.Recurrent mutations in NK-AML and MDS with predicted impact on clinical outcome. 46

Chapter 2 Materials and Methods

Table 2.1. Summary of MDS/AML pedigrees and cases investigated 74
Table 2.2 Oligonucleotide Primers for PCR of *RUNX1*, *GATA2*, *CEBPA* and *ETV6* 79
Table 2.3 PCR thermal cycling parameters for *RUNX1* exons 3-7 80
Table 2.4 PCR thermal cycling parameters for *RUNX1* exon 8 81
Table 2.5 PCR thermal cycling parameters for *CEBPA* Fragments 1-3 81
Table 2.6 SALSA MLPA probemix X060-X2 MDS-AML (MRC Holland) 88
Table 2.7 SALSA MLPA probemix P437-A1 Familial MDS-AML (MRC Holland) 89
Table 2.8 WES variants tested with ultra-deep sequencing across 3 sequential tumours in patient B.I.1. 100
Table 2.9 Touchdown PCR cycling parameters 101

Chapter 3 Clinical and Molecular Profiling of families with germline *CEBPA* mutations

Table 3.1 Summary of patient characteristics for familial and sporadic CEBPA-mutated AML cohorts 112
Table 3.2 Disease presentation profiles in familial AML 114
Table 3.3 WES metrics of familial AML and remission samples 116
Table 3.4 Discordance of somatic *CEBPA* mutations at diagnosis and relapse 123

Chapter 4 Convergence of somatic JAK-STAT mutations in a novel *RUNX1*-mutated pedigree

Table 4.1 WES coverage for pedigree K samples 146
Table 4.2WES VAF of germline *RUNX1* mutation across family members 146

Chapter 5 Novel and Atypical Germline Mutations in Familial AML

Table 5.1 WES metrics for samples from pedigrees P-T 191
Table 5.2 Prioritised germline variants following WES of pedigrees P-T 194

Appendix 2

Table 1 Total number of somatic variants detected in *CEBPA*-mutated AML samples using WES* 258
Table 2 Somatic coding variants for pedigrees A, C and D * 259
Table 3 Somatic coding variants for B.I.1 T1-T3 and B.II.2 260
Table 4 Somatic WES variants verified with Sanger and/or deep sequencing 261

Appendix 3

Table 1 Somatic mutations identified in sibling tumour samples using WES* 262
Table 2 Somatic WES variants verified in pedigree K with Sanger sequencing 263

Appendix 4

Table 1 Germline variants verified by Sanger sequencing in peedigrees P-T 264
Table 2 Somatic coding variants detected in P.III.2 265

Abbreviations

AA	Amino acid
AC	Asymptomatic carrier
AID	Acquired isodisomy
AIE	Cytarabine, idarubicin, etoposide
AL	Acute leukaemia
ALL	Acute lymphoblastic leukaemia
Allo-SCT	Allogeneic haematopoietic stem cell transplantation
AML	Acute myeloid leukaemia
APL	Acute promyelocytic leukaemia
aUPD	Acquired uniparental disomy
Auto-SCT	Autologous haematopoietic stem cell transplantation
BAF	B-allele frequency
BM	Bone marrow
BMT	Bone marrow transplant
BWA	Burrows-Wheeler Aligner
CNA	Copy number aberration
CBF	Core binding factor
CBS	Circular binary segmentation
CCF	Cancer cell fraction
CGH	Comparative genomic hybridisation
CIR	Cumulative incidence of relapse
CLP	Common lymphoid progenitor
CMML	Chronic myelomonocytic leukaemia
CMP	Common myeloid progenitor
COSMIC	Catalogue of Somatic Mutations in Cancer
CpG	Cytosine-phosphate-guanine
CR	Complete remission
CR1	First complete remission
C-terminal	Carboxyl(COOH)-terminal
DA	Daunorubicin and cytarabine
DBD	DNA-binding domain
DC	Dyskeratosis Congenita
Del	Deletion
DFS	Disease free survival
DNA	Deoxyribonucleic acid
DQ	Dosage quotient
Dup	Duplication
ECOG	Eastern Co-operative Oncology Group
ESA	Erythroid stimulating agents
FA	Fanconi Anaemia
FAB	French-American-British
FISH	Fluorescence in situ hybridisation
FLAG-IDA	Fludarabine, cytarabine, G-CSF, idarubicin

FPD/MM	Familial platelet disorder/myeloid malignancy
GATK	Genome analysis toolkit
GMPP	Granulocyte macrophage progenitor population
GvHD	Graft versus host disease
GvL	Graft versus leukaemia
HAT	Histone acetyl transferase
Hb	Haemoglobin
Hg19	Human genome issue hg 19
HSC	Haematopoietic stem cell
HSPC	Haematopoietic stem and progenitor cells
HSCT	Haematopoietic stem cell transplantation
IBMF	Inherited bone marrow failure
IMF	Idiopathic myelofibrosis
Indel	Insertion/deletion
ICGC	International Cancer Genome Consortium
ITD	Internal tandem duplication
LAIP	Leukaemia-associated immunophenotype
LD	Linkage disequilibrium
LIC	Leukaemia initiating cell
LOH	Loss of heterozygosity
LRR	Log R ratio
LZD	Leucine zipper domain
MAF	Minor allele frequency
MDS	Myelodysplastic syndrome
MFC	Multiparameter flow cytometry
MLPA	Multiplex ligation dependent probe amplification
MPN	Myeloproliferative neoplasm
MRC	Medical Research Council
MRD	Minimal residual disease
NCRI	National Cancer Research Institute
NGS	Next generation sequencing
NHL	Non-Hodgkin lymphoma
NHLBI	National Heart Lung and Blood Institute
NICE	National Institute for Health and Care Excellence
NK	Normal karyotype
NMD	Nonsense-mediated decay
NTC	Non-template control
N-terminal	Amino (NH_2) terminal
OMIM	Online Mendelian Inheritance in Man
OS	Overall survival
PB	Peripheral blood
PCR	Polymerase chain reaction
Plt	Platelets
PTD	Partial tandem duplication
RA	Refractory anaemia
RAEB	Refractory anaemia with excess blasts

RARS	Refractory anaemia with ring sideroblasts
RCMD	Refractory cytopenia with multilineage dysplasia
RFI	Relapse free interval
RHD	Runt homology domain
RIC	Reduced intensity conditioning
RNA	Riboxynucleic acid
RQ-PCR	Real-time quantitative polymerase chain reaction
SD	Standard deviation
sDNA	Salivary DNA
SNP	Single nucleotide polymoprhism
sn-RNP	Small nuclear ribo-nuclear protein
SNV	Single nucleotide variant
SP	Subpopulation
SS	Splice site
TAD	Transactivation domain
TBD	Telomere breakage disorder
TCGARN	The Cancer Genome Atlas Research Network
TGP	Thousand Genomes Project
TKD	Tyrosine kinase domain
TRM	Treatment related mortality
TSS	Transcription start site
VAF	Variant allele frequency
VUD	Volunteer unrelated donor
WCC	White blood cell count
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WHO	World Health Organisation
WT	Wild-type

Chapter 1

Introduction

1 Introduction

Acute myeloid leukaemia (AML) and myelodysplastic syndrome (MDS) are heterogenous, clonal disorders of myelopoiesis that culminate in bone marrow (BM) failure. The majority of cases occur *de novo* or as a consequence of exposure to cytotoxic agents. In rare cases, families demonstrate an autosomal dominant predisposition to acute leukaemia, this is predominantly AML (often preceded by MDS). These familial cases represent a unique subset of patients requiring tailored follow up and treatment strategies to achieve cancer risk reduction, prevention and best management.

To date, leukaemia predisposition syndromes are typically associated with germline mutations in the haematopoietic transcription factors runt related transcription factor 1 (*RUNX1*), CCAAT/enhancer-binding protein alpha (*CEBPA*) and GATA binding protein 2 (*GATA2*). The disease manifestations associated with each genetic subgroup are highly heterogenous and vary from asymptomatic mild thrombocytopenia (*RUNX1* mutations) to complex immune defects and marked lymphoedema (*GATA2* mutations). Furthermore, the penetrance of mutations varies significantly, with marked differences in disease latency and many individuals remaining healthy carriers. The clinical recognition of leukaemia-predisposition syndromes is further complicated by an absence of standardised diagnostic guidelines and availability of screening assays. Ultimately, this leads to deficiencies in patient care, most notably with the ill-judged consideration of related donors for allogeneic haematopoietic stem cell transplantation (HSCT).

In the era of next generation sequencing (NGS), we are now beginning to elucidate the genomic landscape of many diseases. Familial leukaemia provides an unrivalled model with which to investigate both the stepwise evolution of leukaemia and to characterise, as yet, unknown disease causing alleles. Unbiased investigation of the entire genome (whole genome sequencing, WGS) or exome (whole exome sequencing, WES) can now be performed with relative efficiency and accuracy, whilst targeted deep sequencing offers rapid analysis of multiple known genetic loci sequenced simultaneously across numerous samples with increased sensitivity for the detection of subclonal variants. With these technologies, we have observed a paradigm shift in the molecular investigation of leukaemia from that of a single gene approach, towards the detailed characterisation of tumour architecture, co-operating lesions and clonal evolution. This

thesis describes the in-depth molecular profiling of familial leukaemia, focusing on the acquired somatic lesions initiating disease across multiple families, patterns of disease evolution and the identification of novel germline mutations to build upon our knowledge of the current germline repertoire.

1.1 Acute myeloid leukaemia

1.1.1 Incidence and clinical presentation of AML

AML is the most common form of aggressive leukaemia in adults. In the United Kingdom there are approximately 3000 new cases per year, with the majority occurring in adults over 60 years of age (Cancer Research UK http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml/incidence, accessed July 2015). It is a complex and heterogeneous disease which may be secondary to the transformation of an antecedent myeloid malignancy, such as a chronic myeloproliferative neoplasm (MPN) or MDS, following exposure to DNA damaging agents such as chemotherapy for unrelated conditions or, more commonly, as an isolated sporadic neoplasm. AML is characterised by the accumulation of immature blast cells within the bone marrow, with variable overspill into the peripheral blood (PB). Patients predominantly present with symptoms of bone marrow (BM) failure, most frequently sepsis (due to neutropenia) or bleeding (due to thrombocytopenia). Without the prompt treatment and supportive care, the disease is universally fatal.

1.1.2 Classification of AML

The diagnosis of AML is primarily based on morphological examination of bone marrow using light microscopy. The French-American-British (FAB) co-operative group were the first to describe 8 distinct subtypes of AML (designated M0-M7) based on morphological and cytochemical evaluation, with =/>30% BM blasts required for a diagnosis of AML, Table 1.1 (Bennett *et al*., 1976). This classification revealed that specific subtypes of AML demonstrate clinical and phenotypic similarities, largely due to recurrent chromosomal translocations involving the core binding factor complex (FAB M2 and M4 AML) or the PML-RARA fusion gene associated with acute promyelocytic leukaemia (APL, FAB M3).

In 2001, the World Health Organisation (WHO), in collaboration with the Society for Hematopathology and the European Association of Hematopathology, published a

'Classification of Tumours of the Hematopoietic and Lymphoid Tissues' within the WHO Classification of Tumors (Jaffe *et al*., 2001). The WHO 2001 classification superseded the FAB system, with several important differences; the BM blast threshold for AML was redefined as 20% (with abrogation of this criteria in the presence of recurrent chromosomal translocations), AML with recurrent cytogenetic abnormalities was classified as a separate entity, multilineage dysplasia was recognised, in addition to new morphological subtypes of AML. Further understanding of molecular aberrations in AML prompted the revised WHO classification in 2008 (Swerdlow *et al*., 2008). This integrated approach encompassed chromosomal, molecular and morphological findings and forms an important foundation for therapeutic decision making (Table 1.2)

Table 1.1 French-American-British Classification of AML

French-American-British classification of AML	
M0	Undifferentiated acute myeloblastic leukaemia
M1	Acute myeloblastic leukaemia with minimal maturation
M2	Acute myeloblastic leukaemia with maturation
M3	Acute promyelocytic leukaemia
M4(eo)	Acute myelomonocytic leukaemia (with eosinophilia)
M5a	Acute monoblastic leukaemia
M5b	Acute monocytic leukaemia
M6	Acute erythroid leukaemia
M7	Acute megakaryoblastic leukaemia

Table 1.2 WHO Classification of AML and related neoplasms

AML and related neoplasms	
AML with recurrent genetic abnormalities	AML wth t(8;21)(q22;q22); *RUNX1-RUNX1T1* AML with inv(16)(p13.1q22) or t(16; 16)(p13.1; q22); *CBFB-MYH11* APL with t(15;17)(q22;q12); *PML-RARA* AML with t(9;11)(p22;q23); *MLLT3-MLL* AML with t(6;9)(p23;q34); *DEK-NUP214* AML with inv(3)(q21q26.2)or t(3;3)(q21;q26.2); *RPN-EVI1* AML (megakaryoblastic) with t(1;22)(p13;q13); *RBM15-MKL1* Provisional entity: AML with mutated *NPM1* Provisional entity: AML with mutated *CEBPA*
AML with MDS-related changes	
Therapy-related myeloid neoplasms	
AML, not otherwise specified	AML with minimal differentiation AML without maturation AML wth maturation Acute myelomonocytic leukaemia Acute monoblastic/monocytic leukaemia Acute erythroid leukaemia i)Pure eythroid leukaemia ii) Erythroleukaemia Acute megakaryoblastic leukaemia Acute basophilic leukaemia Acute panmyelosis with myelofibrosis
Myeloid Sarcoma	
Myeloid proliferations related to Down syndrome	Transient abnormal myelopoiesis Myeloid leukaemia associated with Down Syndrome
Blastic/plasmacytoid dendritic cell neoplasm	

1.1.3 Cytogenetic risk groups in AML

Over the last 20 years, large leukaemia research groups have assessed the importance of AML karyotype in uniformly treated cohorts. Chromosomal abnormalities are now the most established prognostic markers in AML, segregating disease according to favourable, intermediate and poor risk outcomes (Grimwade *et al*., 1998). The favourable risk group includes core binding factor complex translocations and the PML-RARA fusion gene, whilst the poor risk group includes monosomies of chromosomes 5 and 7, chromosome 3 abnormalities or complex karyotypes (>3 abnormalities). The intermediate risk category represents 70% of AML cases, including all normal karyotype AML (NK-AML) and those not otherwise classified as favourable or poor risk. Clinical outcomes in intermediate risk patients are heterogeneous with long term survival estimated at 35-40% in young adults, this has prompted the search for additional molecular markers to help discriminate patients who would benefit from more intensive treatment strategies. The importance of cytogenetic status is illustrated in Figure 1.1 A which reveals survival outcomes according to AML karyotype from the 3 distinct Medical Research Council (MRC) AML 10, 12 and 15 trials (Grimwade *et al*., 2010a). Whilst cytogenetic risk groups do not influence initial AML therapy, they are highly informative in determining appropriate post-remission strategies to minimise the risk of relapse.

1.1.4 Outline of current treatment approaches in AML

The basis of current AML treatment regimens was first described over 40 years ago, where Yates *et al*. reported the findings of a pilot trial of cytarabine (also known as cytosine arabinoside or ara-C) with 45 mg/m^2 of daunorubicin, in a regimen called "7&3 DNR" (Yates *et al*.; 1973). In 2009, the Eastern Cooperative Oncology Group (ECOG), reported that dose intensification of daunorubicin to 90mg/m^2 resulted in higher rates of remission and improved overall survival (OS) in younger patients with standard risk AML (Fernandez *et al*, 2009). This is currently under further evaluation in the UK National Cancer Research Institute (NCRI) AML 17 and AML 19 trials.

In general, the standard approach to AML therapy comprises two cycles of induction chemotherapy, most frequently daunorubicin and cytarabine (DA), followed by one to two cycles of consolidation therapy. In the UK, the majority of patients are enrolled in national clinical trials, which may include randomisation to receive investigational therapies considered to be of potential benefit from earlier phase I and II trials. Patients deemed to be at high risk of relapse may proceed to allogeneic HSCT following completion of induction therapy, this remains one of the most effective therapies in the armamentarium of AML treatment. Advances in donor selection techniques (high resolution human leucocyte antigen [HLA] typing), combined with improvements in supportive care and the management of graft vs. host disease (GvHD) have led to reductions in transplant related mortality (TRM) (Horan *et al*., 2011). Furthermore, non-myeloablative conditioning schedules have extended the range of individuals for whom transplantation is applicable, but patients still need to be physiologically fit to withstand the possible complications of this procedure, including sepsis and GvHD.

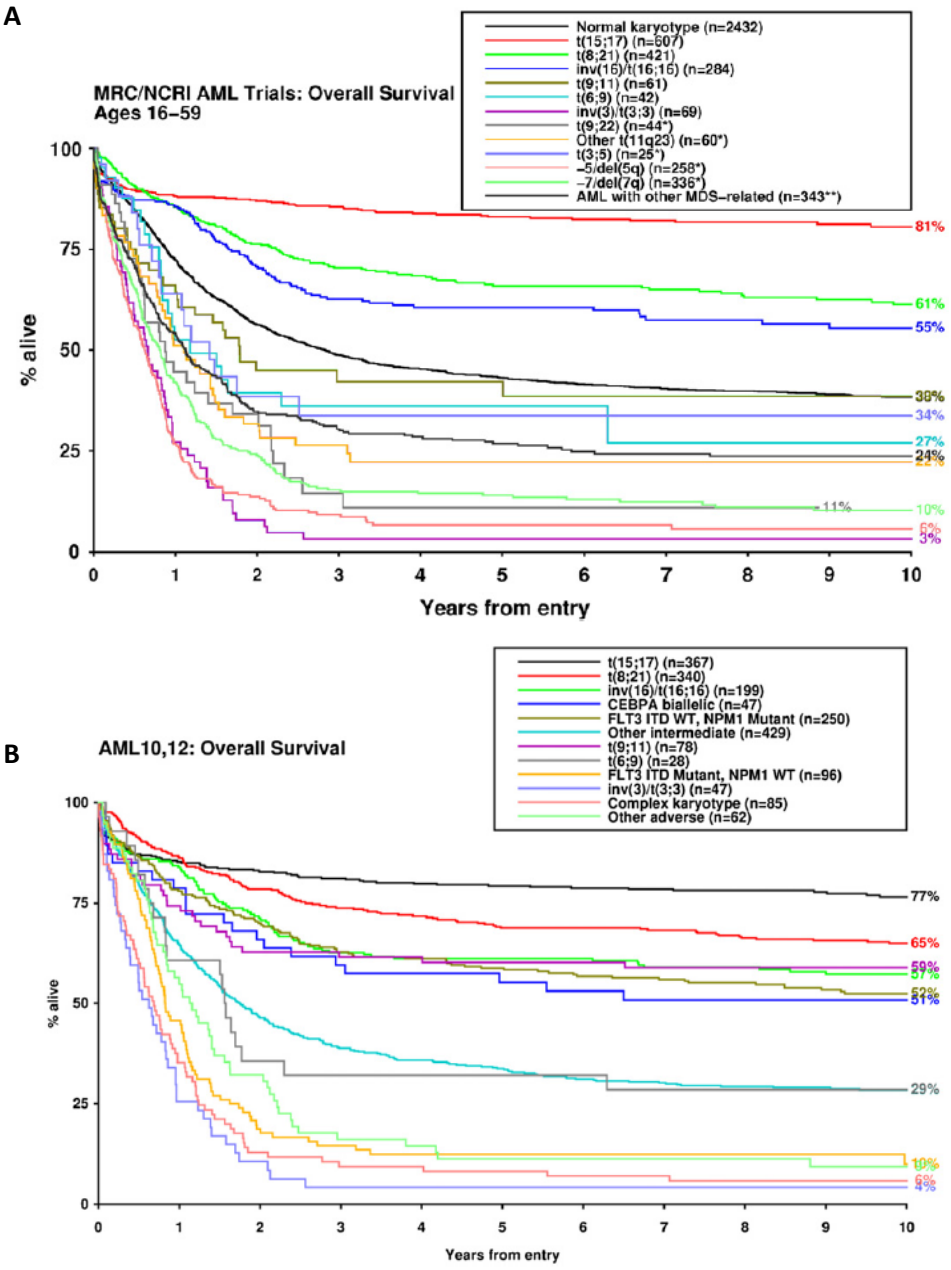


Figure 1.1 MRC trial data showing the impact of cytogenetic and molecular characterisation on survival outcomes for young adults with AML

(A) Overall survival according to recurrent cytogenetic entities described in the WHO Classification, 2008 (figure from Grimwade et al., 2010a) (B) Further characterisation of survival outcome using *FLT3*-ITD, *NPM1,* and bi-allelic (double mutated) *CEBPA* mutations (figure from Grimwade, 2012).

It is often challenging to manage patients who are not eligible for intensive therapy, with limited treatment options available, including low dose cytarabine, the hypomethylating agents azacytidine and decitabine and supportive care (with or without hydroxycarbamide for leucocytosis). Use of the hypomethylating agent, 5-azacytidine, is currently approved by the National Institute for Health and Care Excellence (NICE) for patients with high-risk MDS, chronic myelomonocytic leukaemia (CMML), or AML with 20-30% blasts and evidence of multilineage dysplasia (NICE, 2011). Fenaux *et al* initially demonstrated superior disease control in older patients with high-risk MDS (Fenaux *et al*., 2009) or AML with 20-30% blasts (Fenaux *et al*., 2010). The latter analysis demonstrated a CR rate of 18% and a survival benefit of 24.5 vs 16 months (p=.005), following azacytidine therapy. Notably, a higher 2-year survival (38% vs 0%, p=0.01) was observed in patients with adverse cytogenetics. These findings suggest hypomethylating agents may offer significant responses in patients with limited therapeutic alternatives. With this in mind, the management of elderly AML remains a critical area for future research, where the implementation of less toxic and targeted therapeutic approaches will be highly beneficial for this patient population.

1.1.5 Definition of remission in AML

The conventional definition of remission post therapy involves the achievement of specific haematological parameters (indicating restoration of normal haematopoiesis) and the identification of <5% BM blasts. In addition to morphological assessment, minimal residual disease (MRD), which defines the persistence of leukaemic cells at levels below the sensitivity of detection by morphology, may also be assessed. Assays for detection of MRD include real-time quantitative polymerase chain reaction (RQ-PCR, detecting fusion gene transcripts and molecular variants) or multiparameter flow cytometry (MFC, for the detection of leukaemia-associated immunophenotypes, [LAIPs]).

Currently, the clinical implementation of MRD analysis is largely limited to acute promyelocytic leukaemia and acute lymphoblastic leukaemia (ALL). In non-APL AML, MRD positivity of molecular abnormalities or LAIPs has been associated with relapse and adverse outcomes (Grimwade *et al*., 2010b and Schnittger *et al*., 2009), suggesting its potential utility as a prognostic marker. Discussions are often polarised regarding the

advantages of MRD analysis and are currently the subject of investigation in the UK NCRI AML17/AML19 trials. It seems likely that this information will be a useful adjunct to help select patients who will benefit from allogeneic HSCT in first remission (CR1) in the absence of targeted therapies (Grimwade *et al*., 2014).

1.1.6 Determinants of clinical outcomes in AML

Despite advances in therapeutics and supportive care, the majority of adults die from their disease (with the important exception of APL, comprising ~10% of patients with a cure rate of >75%) (Roboz, 2011). Since the majority of patients achieve morphological CR following chemotherapy, the main determinants of favourable long term outcomes are factors that reduce relapse and treatment related mortality. Older patients with AML (age > 60 years) consistently belong to an adverse prognostic group, with long term survival estimated at <10% (Büchner *et al*.; 2009). The reasons for this are multifactorial and include a higher frequency of sAML, adverse cytogenetic profiles, drug resistance and reduced physiological reserve (performance status) to withstand the side effects of chemotherapy. The MRC identified AML karyotype, white cell count at diagnosis, age and de novo or sAML as the salient prognostic factors in over 2000 older patients with AML, although less objective factors such as patient "fitness" or performance status were also a key factor in treatment decisions for this elderly population (Wheatley *et al*.; 2009). In contrast, younger patients (less than 60 years of age) enter complete remission in over 70% of cases but over half will go on to relapse (often with chemo-refractory disease) with OS plateauing at approximately 40% (Büchner *et al*., 2009).

1.1.7 Outcomes of AML post relapse

As mentioned above, relapse is frequent and occurs in 50-70% of patients with intermediate risk AML, increasing to 80-90% in those with adverse-risk disease (reviewed in Ofran and Rowe, 2012). Following analysis of over 1000 patients treated with chemotherapy alone between 1991 and 2003, annual failure rates of 69% were reported in year 1, 27.7% in year 2 and 17% in year 3, falling to <8% in years 4 and 5. Very late relapses (>5 years) were reported in only 1% of cases (Verma *et al*., 2010). Data from 667 patients aged 15-60 years post first relapse and identified that AML karyotype at diagnosis, relapse free interval, age at relapse and previous HSCT were all predictive of outcome. Notably relapse free intervals (RFI) of > 18 months were associated with

significantly superior outcomes (26% 5 year OS), compared to earlier relapses at occurring at < 6 months (5% 5 year OS). Combining all four variables, five year OS rates varied from 46% in those with a favourable risk calculation, compared to 12% and 6% in the intermediate and poor risk groups respectively (Breems *et al*., 2005). To investigate the survival outcomes following very late relapses, Medeiros *et al*. (2007) described 15 cases with a median age of 48 years, in whom relapse occurred after 5 or more years (median time to relapse 8.6 years). All cases had an intermediate risk karyotype which remained identical at relapse (in 6 patients with available information). Notably, over 86% achieved CR and 51% were disease free at 5 years, these limited data suggest that very late relapses are associated with higher chemotherapy sensitivity.

1.2 Myelodysplastic syndrome

The term MDS represents a heterogenous collection of clonal myeloid neoplasms arising from a common HSC or myeloid progenitor. This syndrome encompasses multiple disease entities affecting one or more myeloid lineages, with variable clinical features and outcomes. Assessment of BM and PB with light microscopy reveals characteristic morphological appearances. Furthermore, evaluation of PB cytopenias and BM blast percentage are also essential for the classification and risk prognostication in MDS, as described below. The incidence of MDS is approximately 4/100,000 population/year, affecting predominantly older individuals, where the incidence increases to 30/100, 000/year over 70 years (Killick *et al*., 2014). Many patients present with the incidental discovery of low blood counts, symptoms related to anaemia (dyserythropoiesis), an increased frequency of bacterial infections (reduced or dysfunctional neutrophils) or a bleeding tendency (reduced or dysfunctional platelets).

1.2.1 Classification and prognostic markers in MDS

A patient's age and perceived risk of transformation to AML are paramount in guiding the clinical management of MDS, the latter being determined by disease classification and prognostic scoring systems. MDS represents a constellation of diagnoses and morphological findings are essential for distinguishing specific disease entities. An updated system for defining these distinct disease subgroups was proposed by the WHO (Swerdlow *et al*., 2008) and details the classification of MDS and MDS/MPN crossover syndromes (Table 1.3).

The International Prognostic Scoring System (IPSS) is most frequently used for risk stratification of patients newly diagnosed with MDS. The overall score is calculated from several variables including BM blast percentage, disease karyotype and the number of PB cytopenias, all of which are used to estimate the median OS and time to leukaemic transformation (Greenberg *et al*., 1997). A revision of this prognostic index was recently published with the inclusion of 2 further cytogenetic subgroups (detailed in Table 1.4) and distinction regarding the level of cytopenias and BM blast percentage (Greenberg *et al*., 2012). As the molecular landscape of MDS becomes increasingly well defined,

mutation profiles are likely to be incorporated into novel scoring systems, although detailed prognostic evaluation is required for many novel lesions.

Chromosomal abnormalities are observed in over 50% of MDS cases, with common aberrations including deletion of chromosomes 5q and 20q, trisomy 8, and monosomy 7 or deletion of 7q. The occurrence of isolated deletion of chromosome 5q, defines a unique disease entity, first described by van den Berghe *et al*. (1974). This syndrome is characterised by haploinsufficiency of *RPS14*, encoding the 40s ribosomal subunit, and located within the deleted chromosomal region. Patients typically present with macrocytic anaemia and thrombocytosis, with BM morphology revealing a low blast count and small hypolobated megakaryocytes (Boultwood *et al*., 1994). Notably, this disease phenotype is typically associated with a low risk of leukaemic transformation, although anaemia may deteriorate with time. Therapy with lenalidomide was found to be effective at inducing cytogenetic remission and abolishing transfusion requirements in 67% of patients (List *et al*., 2005), further underlining the importance of clinical recognition of this syndrome.

Table 1.3 WHO Classification of MDS and MDS/MPN crossover syndromes

Classification of MDS and MDS/MPN	
Myelodysplastic syndrome (MDS)	
Refractory cytopenia with unilineage dysplasia (RCUD)	Refractory anaemia Refractory neutropenia Refractory thrombocytopenia
Refractory anaemia with ring sideroblasts (RARS)	
Refractory cytopenia with multilineage dysplasia (RCMD)	
Refractory anaemia wih excess blasts (RAEB)	RAEB-1 (5-9% BM blasts) RAEB-2 (10-19% BM blasts)
Myelodysplastic syndrome with isolated del (5q)	
Myelodysplastic syndrome, unclassifiable (MDS-U)	
Childhood myelodysplastic syndrome	
Myelodysplastic/myeloproliferative neoplasms (MDS/MPN)	
Chronic myelomonocytic leukaemia (CMML)	
Atypical chronic myeloid leukaemia (atypical chronic myeloid leukaemia (*BCR-ABL1* negative)	
Juvenile myelomonocytic leukaemia (JMML)	
MDS/MPN, unclassifiable	
Provisional entity: refractory anaemia with ring sideroblasts and thrombocytosis	

Table 1.4 IPSS-R Cytogenetic classification of MDS

MDS prognostic subgroup	Cytogenetic abnormalities
Very good	-Y, del(11q)
Good	Normal, del(5q), del(12p), del(20q), double including del(5q)
Intermediate	del(7q), +8, +19, i(17q), any other single or double independent clones
Poor	-7, inv(3)/t(3q)/del(3q), double clone including -7/del(7q), Complex: 3 abnormalities
Very poor	Complex: >3 abnormalities

1.2.2 Treatment strategies in MDS

The heterogeneity of MDS is mirrored by life expectancies ranging from almost a decade for very low risk disease, to several months in high-risk patients. The clinical management therefore varies according to a patient's age, risk of leukaemic transformation and transfusion requirements. Treatment of low-risk disease aims to reduce disease related complications and minimise transfusion requirements. Current therapies available for low-risk MDS include growth factor support, in particular erythroid stimulating agents (ESA), such erythropoietin and granulocyte-colony stimulating factor (G-CSF). Disease modulating agents include lenalidomide which, as discussed above, is efficacious in treating anaemia associated with isolated deletion of chromosome 5q (List *et al*., 2005).

In high-risk MDS, the main objective of current treatment approaches is to prolong survival, with potentially curative management, including allogeneic HSCT. Hypomethylating agents have shown clinical utility in high-risk patients who are not eligible for allogeneic HSCT or as a bridge to transplant (Bejar *et al*., 2014). Allogeneic HSCT is restricted to patients with availability of an appropriate stem cell donor and good performance status. In the interim period, patients may receive induction chemotherapy, a hypomethylating agent or be enrolled in clinical trials, although the optimal pre-transplantation therapy remains unclear.

1.3 The evolving molecular landscape of MDS and AML

As detailed earlier in this chapter, treatment decisions in MDS and AML are predominantly governed by a patient's performance status and predicted prognosis, as determined by clinical and pathological features, including the cytogenetic profile. Over the last two decades, research efforts have focused on deciphering the molecular biology of myeloid malignancies, particularly NK-AML (Schlenk *et al*., 2008). Collectively, these studies have identified recurrent molecular lesions involving FMS-like tyrosine kinase (*FLT3*), nucleophosmin 1 (*NPM1*) and CCAAT/enhancer-binding protein α (*CEBPA*), all of which are now recognised as an integral component in the clinical risk stratification of AML, leading to their incorporation in the revised WHO classification as provisional disease entities (Vardiman *et al*, 2009). Figure 1.1 B shows the distinction in survival outcomes associated with recurrent cytogenetic and molecular lesions in AML (Grimwade, 2012).

1.3.1 FMS-like tyrosine kinase

FLT3 encodes a class III receptor tyrosine kinase which is recurrently mutated in AML. Internal tandem duplication mutations (*FLT3*-ITD) were first described by Nakao *et al*. (1996) and are detected in approximately 30% of NK-AML, targeting the juxtamembrane domain (Kottaridis *et al.*, 2001; Thiede *et al.*, 2002). These in-frame mutations disrupt autoregulation of the receptor's kinase activity and can also increase sensitivity to FLT3 ligand signalling (Zheng *et al*., 2011), causing constitutive activation and an ongoing proliferative signal within leukemic cells. A second type of *FLT3* mutation targets the activation loop of the tyrosine kinase domain (*FLT3*-TKD) and occurs in approximately 10% of NK-AML patients, usually as missense mutations at the aspartic acid residue D835 (Yamamoto *et al*., 2001; Thiede *et al*., 2002; Frohling *et al*., 2002). *FLT3*-TKD mutations lead to constitutive activation of the receptor, although downstream signaling pathways appear to differ from those of *FLT3*-ITD mutations (Grundler *et al*., 2005; Leischner *et al*., 2012). Both *FLT3*-ITD- and *FLT3*-TKD-positive cases are associated with high peripheral white cell counts at presentation, although inferior outcomes are predominantly associated with *FLT3*-ITD mutations, which predict for a significantly increased risk of relapse and reduced OS (Kottaridis *et al*., 2001; Thiede *et al*., 2002; Frohling *et al*., 2002). These adverse effects appear to correlate with increased mutation dosage (or *FLT3*-

ITD:WT allelic ratio), which is frequently caused by loss of heterozygosity (LOH) arising from aUPD of chromosome 13 (Griffiths *et al*., 2005). Although thresholds vary across studies, 20%-40% of *FLT3*-ITD positive cases demonstrate a high ITD allelic ratio at diagnosis (Whitman *et al*., 2001; Kottaridis *et al.* ,2001; Thiede *et al.*, 2002; Meshinchi *et al.*, 2006; Gale *et al.*, 2008).

FLT3-ITD mutations are often associated with favourable-risk founding lesions, including the t(15:17) translocation in APML (Kottaridis *et al*., 2001; Gale *et al*., 2005) and *NPM1* mutations (Gale *et al*., 2008). In APML, *FLT3*-ITD mutations are associated with higher presenting white cell counts although, in the era of targeted therapy and MRD monitoring, adverse impacts upon relapse risk and OS are not observed (Gale *et al*., 2005). In contrast, co-occurrence of *FLT3*-ITD and *NPM1* mutations was associated with inferior OS and disease-free survival (DFS) with clinical outcomes mirroring intermediate risk (*NPM1* WT/*FLT3* WT) disease (Thiede *et al*., 2006). Notably, this effect was augmented with increasing *FLT3*-ITD mutation burden, which led to significant worsening of relapse risk and OS (Gale *et al*., 2008).

1.3.2 Nucleophosmin 1

Mutations in nucleophosmin 1 (*NPM1*) were first described in 2005 and are present in 50–60% of patients with NK-AML (Falini *et al.*, 2005). NPM1 is a nuclear–cytoplasmic shuttling protein, which is ubiquitously expressed and has important functions in the regulation of cell growth, cell proliferation, and pro-apoptotic pathways (Federici *et al.*, 2013). The mutant protein aberrantly localises to the cytoplasm (NPM1c) and secondary mutations are thought to be required for the initiation of disease, as indicated by the long latency or lack of AML demonstrated in murine models (Vassiliou *et al.*, 2011; Chou *et al.*, 2012; Falini *et al.*, 2008). Greater than 80% of mutations occur in exon 12 and are commonly associated with *de novo* AML of myelomonocytic or monocytic morphology, often with high white cell counts and extramedullary disease. In the absence of *FLT3*-ITD mutations, the 5-year OS rate in *NPM1*-mutated cases is estimated at 60% (Schlenk *et al.*, 2008).

1.3.3 CCAAT/enhancer-binding protein α

The transcription factor CEBPA coordinates myeloid differentiation in addition to cellular growth arrest (reviewed by Nerlov, 2007). This intronless gene encodes two isoforms of 42kDa and 30kDa (p42 and p30, respectively), which are regulated by alternative translation initiation, with the shorter p30 isoform lacking the first transactivation domain (TAD1).

CEBPA mutations occur in 10–15% of sporadic NK-AML cases, with approximately equal numbers harbouring single (*CEBPA*sm) or double mutations (*CEBPA*dm) (Dufour *et al*., 2010; Green *et al*., 2010). In *CEBPA*sm AML, frameshift mutations are most frequently detected and are distributed throughout the gene, while in-frame mutations specifically target the carboxyl- (C-) terminal. In *CEBPA*dm AML, mutations arise on different alleles with a distinctive profile characterised by amino- (N-) terminal frameshift mutations (promoting translation of the shorter p30 isoform) with C-terminal in-frame insertions or deletions (disrupting the DNA-binding [DBD] or leucine-zipper [LZD] domains) (Green *et al*., 2010; Fasan *et al*., 2014). The latter subgroup is of particular clinical relevance, predicting favourable therapeutic responses and prognosis, with long term OS approaching 60% (Dufour *et al*., 2010; Green *et al*., 2010).

1.3.4 Expansion of the molecular landscape in MDS and AML

A useful model describing the molecular pathogenesis of AML suggests that two fundamental effects occur: a proliferative advantage (caused by activating [Class I] mutations in signalling transduction pathways), accompanied by a block in differentiation (Class II mutations, often involving transcription factors or fusion oncogenes) (Gilliland, 2002). It has since been noted that up to 50% of AML cases demonstrate lesions within novel molecular subgroups. These findings have revealed increased complexity in the genetic profile of AML, although the precise functional sequelae for many novel mutations are yet to be defined.

1.3.4.1 Epigenetic mutations in MDS/AML

In recent years we have witnessed significant changes in the mutational landscape of myeloid malignancies with attention shifting to epigenetic lesions which regulate DNA methylation and post translational histone modifications. Prior to the introduction of NGS technologies, the sole recurrent epigenetic lesions described in AML were partial tandem duplications (PTDs) of the mixed lineage leukaemia (*MLL*) gene on chromosome 11q23. *MLL*-PTDs were identified in 5% of NK-AML and led to inferior event free (EFS) and OS (Basecke *et al*., 2006). These aberrations characteristically retained the C-terminal SET domain, associated with the transcription-activating histone 3 lysine 4 (H3K4) methyltransferase activity of *MLL* (Nakamura *et al*., 2002; Milne *et al*., 2002). *MLL* is also subject to multiple translocations which, in contrast to PTDs, cause loss of the SET domain, but continue to confer an adverse risk profile. Following elucidation of the methylation signature associated with the *MLL-AF9* fusion protein, Bernt *et al* (2011) recently identified the histone methyltransferase DOT1L as the critical enzymatic mediator of the transcriptional signature associated with the fusion protein. This finding generated immense interest and paved the way for the development of DOT1L inhibitors aimed at treating *MLL*-rearranged leukaemia, highlighting the importance and translational potential of *in vivo* molecular modelling.

In 2009, the sequencing of a NK-AML genome identified recurrent mutations affecting an evolutionarily conserved arginine residue (R132) of the cytoplasmic metabolic enzyme isocitrate dehydrogenase 1 (*IDH1*) (Mardis *et al.*, 2009). Additional work revealed recurrent mutations targeting codons R140 and R172 of isocitrate dehydrogenase 2 (*IDH2*), the mitochondrial homologue of *IDH1* (Ward *et al.,* 2009). Cumulatively, *IDH1* and *IDH2* mutations affect 30% of all AML cases, particularly those with a normal karyotype (Abdel-Wahab *et al.*, 2013). Mutations of *IDH1/2* were associated with increased enzymatic conversion of alpha-ketoglutarate (α-KG) to 2-hydroxyglutarate (2-HG), causing a reduction of histone demethylation (Lu *et al*., 2012). The production of 2-HG by mutant forms of IDH1 and IDH2 has also been recognized as a potential target for novel therapeutic design. A small molecule (AGI-6780) that selectively inhibits the IDH2 R140Q mutant protein was recently shown to induce differentiation of erythroleukemic cell lines and primary human AML cells *in vitro* (Wang

et al., 2013), this again highlighted the potential for future targeted therapeutic development.

Delhommeau *et al*. expanded the epigenetic repertoire further, following their report of recurrent deletions or UPD in the chromosomal region 4q24. This corresponded to the genetic locus of tet methylcytosine dioxygenase 2 (*TET2)*, encoding a member of the TET family of proteins, associated with DNA hydroxymethylation. Loss-of-function mutations were identified in approximately 20% of patients with MDS or secondary AML (Delhommeau *et al.*, 2009), with a similar mutation frequency later identified in NK-AML patients where they were associated with reduced OS in the European Leukemia Net favourable-risk subgroup (*FLT3*-ITD-negative patients with *CEBPA* and/or *NPM1* mutations) (Metzeler *et al.*, 2011). *TET2* and *IDH1/2* mutations are mutually exclusive in AML and share similar methylation profiles, characterised by global promoter hypermethylation. Biochemical analysis revealed that TET proteins and IDH enzymes cooperate through the production of α-KG which acts as a co-factor for TET-mediated DNA hydroxymethylation, this effect is inhibited by 2-HG produced by mutant IDH1/2 (Xu *et al*., 2011; Figueroa *et al*., 2010).

In 2010, Ley *et al*. reported mutations of the DNA methyltransferase *DNMT3A* in 22% of an intermediate-risk AML cohort (Ley *et al*., 2010). These mutations are frequently associated with *FLT3*, *IDH1*, and *NPM1* mutations and most commonly occur in FAB M4 and M5 subtypes, where they independently predict for reduced OS (Thol *et al.*, 2011; Yan *et al.*, 2011; Ribeiro *et al.*, 2012; Patel *et al*., 2012). *DNMT3A* mutations are invariably heterozygous and are predicted to disrupt the catalytic activity of the enzyme, which normally mediates methylation of cytosine residues at the 5' position of cytosine-phosphate-guanine (CpG) dinucleotide islands, forming 5 methylcytosine (5mC) and silencing gene expression (reviewed in Rodriguez-Paredes, 2011). Notably, the effect of recurrent R882 missense mutations is unclear and mechanisms by which *DNMT3A* mutations initiate AML remain elusive.

The polycomb repressive complex 2 (PRC2) is associated with methylation of histone H3 lysine 27 (H3K27) and is made up of four components: enhancer of zeste homologue 2 (EZH2), suppressor of zeste homologue 12 (SUZ12), jumonji, AT rich interactive domain

2 (JARID2) and embryonic ectoderm development (EED). EZH2, the enzymatic component of the PRC2, is a methyltransferase which controls cell fate decisions by orchestrating gene expression to regulate self-renewal and differentiation (Bracken *et al*., 2006). *EZH2* mutations are frequently identified in germinal centre (GC) lymphomas, occurring in 25% of follicular lymphoma (Bödör *et al*., 2013) and 21% of GC-derived diffuse large B-cell lymphoma (Morin *et al*., 2010). These mutations are predominantly missense lesions which target the SET domain causing gain of function with enhanced tri-methylation of H3K27. *EZH2* mutations occur in approximately 2% of *de novo* AML, but are more frequently observed in MDS, MPNs and MDS/MPN crossover syndromes, where they occur in 5-15% of cases and are associated with adverse outcomes (Bejar *et al*., 2011; Grossman *et al*., 2011; Guglielmelli *et al*., 2011). In contrast to mutations in lymphoid malignancies, myeloid neoplasms harbour missense, nonsense and frameshift mutations distributed throughout the coding region of *EZH2*. Emerging data have revealed that *EZH2* mutations in myeloid malignancy are frequently associated with loss of heterozygosity (LOH) arising from acquired uniparental disomy (aUPD) or deletion of chromosome 7q (Jerez *et al*., 2012; Wang *et al*., 2013), suggesting bi-allelic disruption or loss is important in the pathogenesis of these disorders.

Additional sex combs like transcriptional regulator 1 (*ASXL1*) is considered to associate with the PRC2 complex, promoting its recruitment to target loci (Abdel-Wahab *et al*., 2012). *ASXL1* mutations cluster in exon 12 and are predominantly frameshift and nonsense variants leading to loss of the C-terminal planthomeodomain (PHD) finger. In a large cohort of 740 AML cases, Schnittger *et al*. (2013) identified *ASXL1* mutations in 17.6% of AML with an intermediate karyotype. A higher frequency was observed in AML with trisomy 8, where more than 50% of cases had an associated *ASXL1* mutation. Importantly, *ASXL1* mutations were an independent risk factor for OS in both univariate and multivariate analyses; median OS was 11 months in the mutated group vs. 62 months in the wild-type (WT) group.

The identification and functional characterisation of lesions in epigenetic regulators has led to new appreciation of the co-ordinated disruption of transcriptional activation and repression mediated by affecting chromatin structure and non-covalent interactions between nucleosomes. Further work involving *in vitro* and *in vivo* modelling of these

lesions aims to determine the precise mechanisms by which leukaemia is initiated and this information will prove critically important for the development of targeted therapies where, as discussed for *IDH1/2* and *MLL*-rearranged leukaemias, novel agents are now emerging.

1.3.4.2 NGS identifies novel genetic lesions in myeloid malignancies

Expanding the mutational repertoire of MDS and AML further, WES studies led to the characterisation of cohesin complex mutations in myeloid malignancies. Cohesin is a multimeric protein complex which is composed of four subunits: SMC1, SMC3, RAD21 and STAG. The ring-like structure has been shown to govern the cohesion of sister chromatids during mitosis, transcriptional regulation and post replicative DNA repair, emphasising its essential role in maintaining genomic stability (reviewed in Wu and Yu, 2012). Kon *et al*. (2013) detected mutations and deletions involving multiple components of the cohesin complex (*STAG2, RAD21, SMC1A*, and *SMC3*) in 8% of patients with MDS, 10% with CMML, and 12% with AML. Dysregulation of cohesin function appears a significant step in the pathogenesis of myeloid malignancy, although functional consequences and prognostic associations of mutations are yet to be defined.

WES also led to the identification of novel mutations affecting multiple components of the RNA splicing machinery in patients with MDS. Several components were recurrently mutated, including *U2AF35* (also known as *U2AF1*), *ZRSR2*, *SRSF2* and *SF3B1* (Yoshida *et al*., 2011). The spliceosome is responsible for splicing of pre-messenger RNA (pre-mRNA) and is a macromolecule composed of small nuclear RNAs associated with proteins termed small nuclear ribo-nuclear proteins (snRNP). Whilst mutations of the spliceosome are relatively infrequent in AML (occurring in 5-10% of cases), up to 50% of MDS cases harbour lesions within this pathway, highlighting the importance of RNA splicing in the pathogenesis of MDS (Papaemmanuil *et al*., 2013; Cazzola *et al*., 2013; Visconte *et al*., 2012). Notably, *SF3B1* mutations were highly enriched in RARS, occurring in over 65% of cases and suggesting a direct association with this disease phenotype (Papaemmanuil *et al*., 2011). Virtually all spliceosome mutations are mutually exclusive missense substitutions, indicating a dominant negative action of the mutant snRNP. While the molecular consequences have not been fully elucidated, Figure 1.2 details the

major spliceosome components and physiological variations in RNA splicing that may arise.

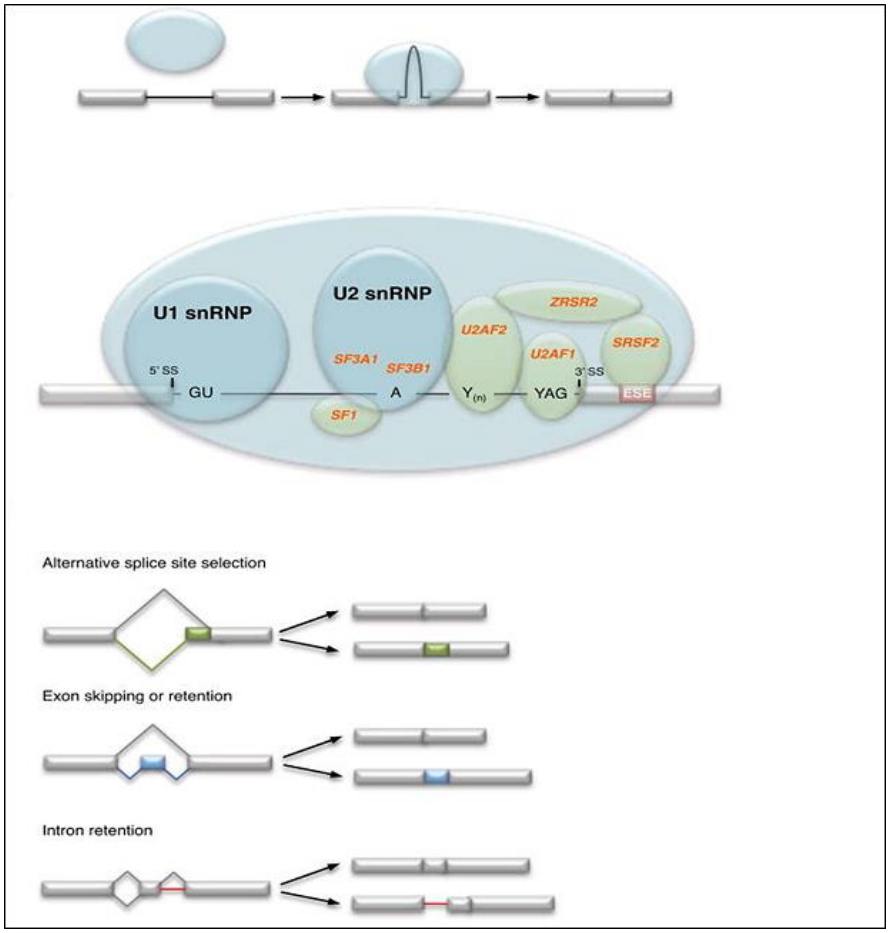


Figure 1.2 Diagrammatic representation of components of the spliceosome, figure from Lindsley and Ebert (2013).

The multi-component spliceosome processes precursor messenger-RNA to excise the intervening intron and ligate adjacent exons. The U1 sn-RNP binds to the 5' splice site (SS) and the U2 sn-RNP to the adenine-containing branch point (A) which is upstream of the polypyrimidine tract (Yn, a pyrimidine rich region of messenger RNA near the 3' end of the intron). The U2 auxiliary factor (composed of subunits U2AF1 and U2AF2) binds to the 3' SS and the Yn. The serine- and arginine-rich SRSF2 binds to the exonic splice enhancer (ESE). Several variations of splicing are observed in normal physiology, leading to alternative protein isoforms.

Importantly novel genetic lesions are helping to characterise and distinguish distinct disease entities within the spectrum of MDS/MPN. SET binding protein 1 (*SETBP1*) mutations were recently detected in 25% of atypical CML cases, 17% of sAML and 15% of CMML (Makishima *et al*., 2013; Piazza *et al*., 2013). The function of SETBP1 is incompletely characterised, but mutations are believed to inhibit activity of the putative tumour suppressor, protein phosphatase type 2A (PP2A), enhancing proliferative capacity and potentially leading to leukaemic transformation of MDS. In other subtypes of MDS/MPN, activating mutations of *CSF3R,* the G-CSF receptor, were recently identified in 59% of patients with chronic neutrophilic leukaemia or atypical CML, leading to upregulation of kinase signaling through the SRC family or JAK-STAT pathway (Maxson *et al.*, 2013; Pardanani *et al*., 2013).

1.3.4.3 Emerging Molecular Associations in MDS and AML

Overall, it is estimated that 90% of patients with MDS have one or more pathogenic mutations and approximately 60% of these occur in the context of a normal karyotype (Cazzola *et al*., 2013). The majority of mutations occur in molecular pathways previously described in AML: RNA splicing, DNA methylation, chromatin modification, transcription regulation, DNA repair, signal transduction and the cohesin complex. Importantly, molecular profiles differ from *de novo* AML varies, with RNA splicing mutations dominating the molecular landscape of MDS. It is important to note that mutations are not isolated events and extensive analysis of over 700 MDS patients identified complex pairwise molecular associations, typically affecting the components of the spliceosome and epigenetic modifiers, suggesting that disease trajectories were pre-defined and that founder mutations may dictate clinical and morphological phenotypes of disease (Papaemmanuil *et al*., 2013).

Thus, emerging data appear to suggest that myeloid neoplasia is initiated by the non-random sequential acquisition and selection of mutations within primitive cell populations. In their extensive analysis of 200 cases of *de novo* AML, the TCGA reported that mutations in *DNMT3A*, *IDH1/2*, *RUNX1*, *CEBPA* and *NPM1* were mutually exclusive with transcription factor fusion genes, suggesting the former were all putative initiating lesions (TCGA, 2013). In contrast, mutational co-operation is frequently reported between, *FLT3*-ITD and *NPM1* (Gale *et al*., 2008) *or* CBF translocations and *c-KIT*

mutations (Paschka *et al*., 2006; Cairoli *et al*., 2006), alluding to Gilliland's two-step model of leukaemogenesis combining proliferative signalling with differentiation blockade (Gilliland, 2002). With the multitude of novel genetic lesions now emerging in MDS/AML, it seems we have a more comprehensive knowledge of the individual components and pathways involved in myeloid malignancy. The challenge ahead lies in determining how to piece this information together, to help decipher the sequence of events governing leukaemia initiation and evolution. A summary of the key molecular lesions in MDS and AML is provided in Table 1.5 (TCGA, 2013; Cazzola *et al.*, 2013; Bejar *et al.*, 2011; Marcucci *et al.*, 2011; Visconte *et al.*, 2013; Thol *et al*., 2013).

Table 1.5 Recurrent mutations in NK-AML and MDS and predicted impacts on clinical outcome.

Class of gene	Gene	Frequency in AML	Frequency in MDS	Impact on survival
Transcription factors	*CEBPA*	10% NK AML	rare	DM -favourable
	RUNX1	10% NK AML	15-20%	Adverse
	WT1	10-13% NK AML	Very rare	Conflicting data
Nuclear regulation	*NPM1*	25-35% all AML 50-60% (NK AML)	Rare	*NPM1* without *FLT3*-ITD: favourable
Tyrosine kinases	*FLT3-ITD*	20% all AML (30% NK AML)	Very rare	Adverse
	C-KIT	25-30% in CBF AML	Very rare	Adverse in CBF AML
Epigenetic modifiers	*DNMT3A*	18-22% (all AML) 35% (NK AML)	10%	Adverse
	TET2	7-23% all AML 18-23% (NK AML)	20-25%	Adverse
	IDH1/2	5-30% all AML 20% (NK AML)	<5%	Conflicting data
	MLL-PTD	10% NK-AML	Very rare	Adverse
	ASXL1	10-18% all AML	15-25%	Adverse
	EZH2	<1% all AML	6%	Adverse in MDS
Cohesin complex	*SMC1A*	5-10% all AML	Rare	Unknown
	SMC3	5-10% all AML	Rare	Unknown
Spliceosome complex	*SF3B1*	2-5% all AML	15-30% (70% RARS)	Favourable in MDS
	SRSF2	1-5% all AML	10% (28% CMML)	Adverse in MDS
	U2AF1 (U2AF35)	5-10% all AML	~10%	Adverse in MDS

1.4 Clonal architecture and evolution determined by whole genome/exome sequencing

NGS datasets have also led to the development of novel algorithms with which to study clonal architecture and evolution in MDS and AML. Since the pioneering study of the first AML genome (Ley *et al*., 2008), we are now beginning to decipher the clonal composition of tumours and study their evolution in disease relapse or transformation to AML. This work has been demonstrated in several concept-defining studies which have revealed that *de novo* AML is an oligoclonal entity with a limited number of mutations (<50 variants) observed within tumours. Furthermore, the majority of mutations represent passenger or bystander variants in aging HSCs (Welch *et al*., 2012). Two distinct patterns of clonal evolution may precipitate relapse: firstly, a linear pattern (whereby relapse is driven by evolution of the founding clone) or secondly, branching evolution, whereby subclones persist and expand following chemotherapy, Figure 1.3 (Ding *et al*., 2012). Further work will undoubtedly focus on characterising tumour architecture and evolution in greater detail, most likely employing xenograft models of purified HSC/HPC subsets and techniques such as single cell sequencing. This work will help unravel the clonal complexity of tumours, representing a significant advance in determining the early molecular lesions governing chemotherapy resistance and disease relapse. Such findings will ultimately guide the future design and implementation of targeted therapies to minimise the risk of relapse, the leading cause of mortality in AML.

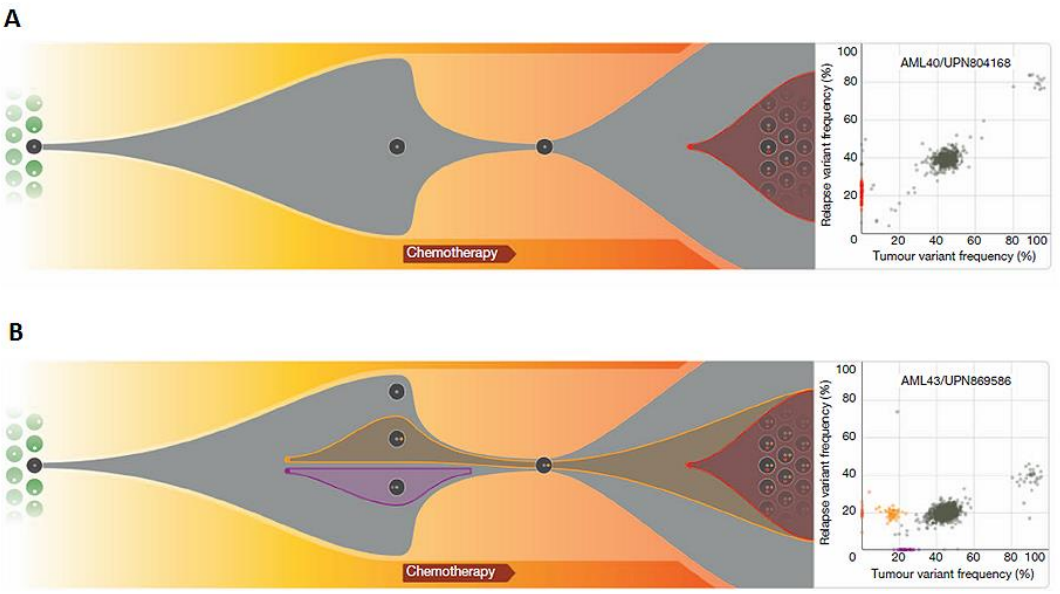


Figure 1.3 Current models of clonal evolution in AML, figure from Ding *et al.*, (2012)

(A) A linear pattern is observed when evolution of the founding clone initiates relapse. (B) Branching evolution occurs with expansion of a minor subclone from the diagnostic tumour. In both scenarios novel genetic lesions are acquired and promote clonal proliferation and selection. Scatter plots show variant allele frequencies (VAFs) within clonal subpopulations for paired diagnostic and relapse tumours.

1.5 The cell of origin in AML

Equally important to understanding the molecular pathogenesis of AML and MDS, has been elucidation the process of haematopoiesis, leading to the production of mature blood cells. The Russian Biologist, Maximow, first described haematopoiesis as a cellular hierarchy, arising from a common precursor, the haematopoietic stem cell (HSC) (Maximow, 1909). Several decades later, Lorenz *et al*. (1951) reported the existence of self-renewing HSCs, by showing that irradiated mice could be rescued from BM failure by injections of BM from non-irradiated donors.

1.5.1 Overview of haematopoiesis

HSCs represent a small group of pluripotent cells which are capable of rapid cell division followed by long term quiescence. It has been shown that foetal HSCs undergo rapid self-renewal, leading to a dramatic increase in cell number to meet the demands of adult haematopoiesis. This rapid proliferation is thought to continue until 3 weeks after birth, after which HSCs undergo a rapid change in their cycling and become quiescent (Bowie *et al*., 2006), this state is predominantly maintained throughout adult life, although they retain the capacity for self-renewal (clonal expansion) and differentiating cell division (clonal extinction) ensuring the steady maintenance of HSC numbers (Cheshier *et al*., 1999; Passegue *et al*., 2005; Wilson *et al*., 2008, Pina and Enver, 2007). During normal haematopoiesis, it is estimated that approximately 1000 HSCs contribute to the development of circulating differentiated cells (Catlin *et al*., 2011).

In the process of HSC differentiation, myeloid and lymphoid lineage specification leads to the formation of progenitor cell intermediates. The common myeloid progenitor (CMP) generates erythroid, megakaryocytic and subsequently granulocytic cells, whilst the common lymphoid progenitor (CLP) yields B-, T-, and NK- cell populations (Figure 1.4). While it is generally considered that CMP and CLP populations harbour limited self-renewal capacity, disruption of this process can lead to haematological malignancy, with dysregulated proliferation.

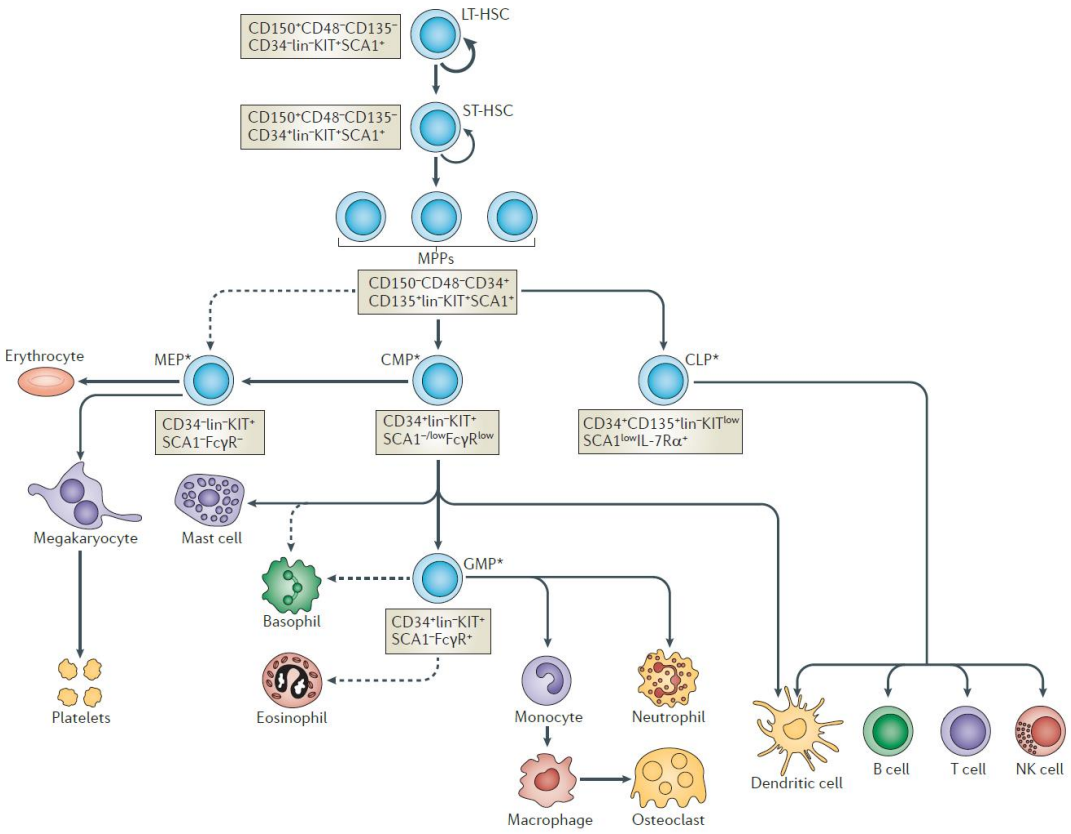


Figure 1.4 Diagram of haematopoiesis showing cell-markers at HSC and progenitor stages, figure from Wang and Wagers (2011).

All haematopoietic cells ultimately derive from a small population of HSCs comprising either long term (LT-HSCs) or short term (ST-HSCs) repopulating cells. The former subset maintain multilineage potential and reconstitution throughout life, whilst ST-HSCs have more limited self-renewal. Multipotent (MPPs) and oligopotent progenitors arise from ST-HSCs and these immature cells progressively lose differentiation potential as they produce terminally differentiated mature blood cells (shown at the bottom). There are multiple subsets of MPPs but these are simplified to show one population in this diagram. The common lymphoid progenitor (CLP), common myeloid progenitor (CMP), megakaryocyte-erythrocyte progenitor (MEP) and granulocyte-monocyte progenitor (GMP) populations are oligopotent lineage-committed cells which derive from MPPs. Commonly denoted flow cytometric profiles associated with HSC and progenitor populations are shown. Dotted arrows indicated putative lineage connections. CD135 is also known as FLT3; IL-7R, interleukin-7 receptor; lin, lineage markers; NK, natural killer; SCA1, surface cell antigen 1.

1.5.2 Characterisation of leukaemia initiating cells

AML is thought to arise from a hierarchic progression of precursor cells, with leukaemia initiating cells (LICs) at the apex and xenotransplantation assays with severe combined immunodeficient (SCID) mouse models, have enabled investigators to characterise these cells in detail. In 1994, Lapidot *et al*. identified that the CD34+CD38- subset of cells were capable of recapitulating the leukemic population of blasts. Subsequent studies revealed that CD34- cells were also capable of leukaemia initiation, specifically in *NPM1*-mutated tumours, in keeping with low CD34 expression within this genetic subtype of AML (Taussig *et al*., 2010). Further heterogeneity of AML precursors was described by Goardon *et al*. (2011), who identified a primitive CD38-CD45+ (LMP) population and a more mature CD38+CD45+ granulocyte-macrophage progenitor (GMP)-like population, giving rise to more-differentiated AML subtypes.

LIC subsets have significant clinical relevance, as these predominantly quiescent cells, are often resistant to chemotherapy and may form a residual reservoir of disease which is capable of initiating relapse (Ishikawa *et al*., 2007; Konopleva *et al*., 2011). Work is now focusing on characterising the molecular aberrations in LICs, in the hope that this may provide new therapeutic opportunities to eradicate residual reservoirs of disease.

1.5.3 Pre-leukaemic mutations in LICs frequently target epigenetic modifiers

Recent studies have shown that LICs are capable of continuing normal haematopoietic reconstitution with their differentiated daughter cells harbouring identical genetic lesions (Corces-Zimmerman *et al*., 2014; Shlush *et al*., 2014). Essentially, these are now considered to be 'pre-leukaemic' clones which persist in a latent state, pending the acquisition of 'driver' mutations to initiate clonal proliferation and expansion. Corces-Zimmerman *et al* (2014) reported genes affecting DNA methylation, histone modification and chromatin looping were significantly over-represented with the majority of pre-leukaemic mutations involving *DNMT3A*, *IDH1/2*, *ASXL1*, *IKZF1* and the fusion gene *CBFB-MYH11*. Shlush *et al*. (2014), elegantly demonstrated that *DNMT3A* mutations preceded the acquisition of other founding mutations such as *NPM1* and were present in differentiated cell subsets and during remission.

These early lesions in LICs are thought to promote clonal expansion, which predisposes to the acquisition of secondary events initiating leukaemic transformation, as shown in Figure 1.5. Awareness of pre-leukaemic mutations is likely to herald the emergence of a paradigm shift in current clinical practice. With further prospective evaluation of their associated relapse risks, detection of residual mutations may provide a strong rationale to escalate consolidation therapy. Ultimately, these therapeutic approaches are governed by weighing the risks and benefits of treatment, further highlighting the need for less toxic, targeted therapies for the treatment of latent clonal expansions.

Pre-leukaemic mutations have also been identified in asymptomatic individuals who have normal haematopoietic development and maturation. Busque *et al* (2013) identified somatic *TET2* mutations in healthy elderly women with clonal haematopoiesis demonstrated by skewing of X-inactivation in peripheral blood polymorphonuclear leucocytes. After follow up of seven individuals with somatic *TET2* mutations for 5 years, one developed a *JAK2*-V617F mutated myeloproliferative disorder. In an extensive analysis of over 12,000 individuals unselected for cancer or haematological disease, Genovese *et al*. (2014) examined PB WES data revealing clonotypic expansion in 10% of persons over 65 years of age, with somatic mutations most frequently affecting *DNMT3A*, *TET2* and *ASXL1.*

Thus, multiple studies have now confirmed that somatic genetic mutations (particularly affecting epigenetic modifiers) accumulate with age and lead to subclinical clonal haematopoiesis. In 1976, Nowell proposed that cancer is caused by multiple mutations which promote the expansion and selection of specific 'sublines' (Nowell, 1976). Consistent with the concept of pre-leukaemic mutations, Nowell had postulated that initial mutations in some way promote genetic instability and that environmental pressures allow selection of mutant cells. Nearly 40 years on, we are yet to determine the nature of the host 'factors' governing selection of tumour clones. With approximately 25,000 genes in the human genome, clear patterns of molecular co-operation and/or exclusivity, a random process of mutation acquisition and selection seems increasingly unlikely. The study of leukaemia evolution in family members provides a valuable opportunity to evaluate the impact of inherited factors upon mutation acquisition and disease progression, as discussed in Chapter 4.

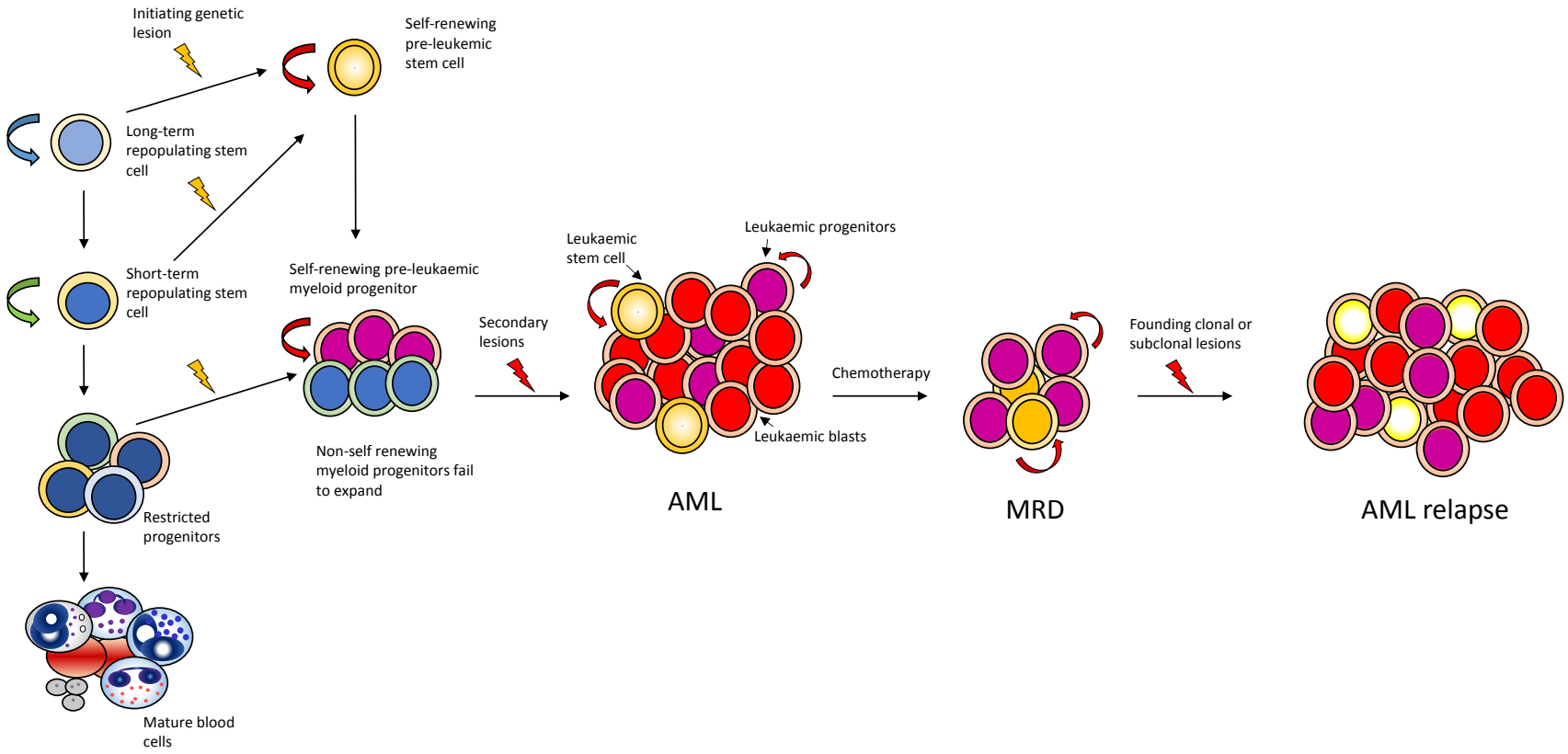


Figure 1.5 Current theory of leukaemia initiation (figure adapted from Roboz, 2011)

Pre-leukaemic stem cells and myeloid progenitors are LICs which harbour initiating genetic lesions (yellow bolt). These cells are capable of haematopoietic reconstitution however clonotypic expansion occurs, predisposing to the acquisition of secondary genetic lesions (red bolt), which initiate AML. Following chemotherapy, residual LICs persist and following the acquisition of further lesions, undergo clonal evolution leading to overt relapse.

1.6 Leukaemia predisposition syndromes

1.6.1 The molecular basis of familial leukaemia

Our current understanding of MDS/AML demonstrates that several genetic events are required to initiate and 'accelerate' the onset of overt disease. This is perhaps best demonstrated in monogenic leukaemia predisposition syndromes, where individuals carrying single germline mutations typically develop disease in an autosomal dominant pattern, with variable penetrance and latency. Familial leukaemia can be defined by two or more directly related individuals presenting with haematological maligancy. Although MDS and AML are the predominant manifestations, ALL and biphenotypic leukaemia may also occur, suggesting evolution of disease from more primitive progenitors. Germline mutations are confirmed in constitutional DNA (often from saliva, buccal epithelial cells, skin biopsy or cultured fibroblasts). While inherited mutations have been reported in at least 10 genetic loci, the most established leukaemia predisposition syndromes are caused by lesions in *RUNX1*, *CEBPA* or *GATA2*.

1.6.2 Runt-related transcription factor 1

Germline mutations in *RUNX1* are associated with a propensity to develop familial platelet disorder (FPD) and MDS/AML. FPD/AML (Online Mendelian Inheritance in Man, OMIM 601399) is a heterogeneous syndrome characterised by defects in platelet function and number, often progressing to haematological malignancy, most commonly MDS and AML. The role of RUNX1 as a critical regulator of haematopoiesis, underpins its importance as a molecular target in familial and sporadic haematological malignancy, as discussed in detail below.

1.6.2.1 The role of *RUNX1* isoforms in haematopoietic regulation

RUNX1 encodes the α-subunit of the core-binding factor complex and its expression and activity of are tightly regulated by multiple mechanisms, including post-translational modifications (reviewed by Goyama *et al*., 2014) and alternative transcription by promoters distal (P1) and proximal (P2) to the transcription start site (TSS) (reviewed in Levanon and Groner, 2004) . Three isoforms of *RUNX1* are described (*RUNX1a-c*) with the most common forms, *RUNX1b* and *RUNX1c*, both possessing the DNA binding and transcription regulatory regions located within the N- and C-termini respectively

(Miyoshi *et al*., 1995; Ghozi *et al*., 1996; Levanon *et al*., 1996). Transcription of *RUNX1b* (comprising exons 3-8) is initiated by the P2 promoter, whilst the P1 promoter directs transcription of *RUNX1c*, translated from exons 1 and 2 which encode an additional 27 amino acids. As with *RUNX1b*, *RUNX1a* expression is also regulated by the P2 promoter, although this isoform lacks the transcription-regulating C-terminal. Mouse studies have shown that both promoters are temporally modulated, with P2 expressed in early embryonic haematopoiesis, whilst P1 is predominantly expressed in haematopoietic progenitors (Sroczynska *et al*., Blood 2009).

The essential role of Runx1 in embryonic haematopoesis was demonstrated by the observation of mid-gestation lethality in knockout mice (Wang et al., 1996; Okuda et al., 1996). By conditional deletion in mouse models, Chen et al. (2009) identified that Runx1 mediates the transition from haemogenic endothelium to HSCs. In adult mice with conditional deletion of Runx1, an expansion of myeloid progenitors and lineage negative (LSK) stem cells was observed (Growney et al., 2005). In addition, multi-lineage blocks in B- and T- lymphoid development and megakaryocyte maturation, rendered mice lymphopenic and thrombocytopenic (Ichikawa et al., 2004). These findings suggest that adult haematopoiesis is not precluded by loss of Runx1, but rather that negative regulation of HSC activity is reduced (Ichikawa et al., 2008).

This co-ordinated regulation of haematopoiesis, lineage specification and differentiation is mediated by the interaction of RUNX1 with multiple transcription factors. A critical downstream target is PU.1, a member of the E-twenty-six (ETS) family of transcription factors and itself a key regulator of haematopoiesis (Scott *et al*., 1994). Furthermore, RUNX1 interacts with epigenetic or chromatin modifiers such as *MLL*, *MYST3* and *P300/CBP* (Yamagata *et al.*, 2005, Zhao *et al.*, 2008, Reed-Inderbitzin *et al.*, 2006) and is downregulated in AML with MLL fusion proteins, suggesting MLL may in turn regulate the expression of *RUNX1* (Zhao *et al*., Blood 2014). The latter association was further illustrated by Huang *et al* (2011), who demonstrated that *RUNX1* mutations disrupted binding with MLL, interfering with the downstream regulation of PU.1. These observations provide a valuable insight into the intricate regulatory network mediated by RUNX1 and importantly reveal clues to the specific effects caused by mutations or translocation of the individual components within this network.

1.6.2.2 Chromosomal rearrangement and mutation of *RUNX1* in sporadic leukaemia

The function of RUNX1 is dependent upon two key regions. The highly conserved RHD, located within the N-terminal, is associated with DNA binding and heterodimerisation with core binding factor β (CBFβ), the latter serves to stabilise the protein, enhancing binding to DNA and preventing degradation (Tahirov *et al*., 2001; Yan *et al*., 2004). The C-terminal transactivation domain (TAD) mediates transcriptional activation and interaction with downstream regulatory targets.

Recurrent translocations involving *RUNX1* have been reported in AML and ALL. The recurrent t(8;21) translocation between *RUNX1* and *ETO* (also known as *RUNX1T1*) represents one of the most frequent chromosomal lesions in AML (predominantly FAB M2 subtype), as described earlier in this chapter. The fusion gene is created by juxtaposition of the N-terminal of *RUNX1* (with an intact RHD) to *ETO* on chromosome 8, permitting DNA binding but disrupting RUNX1-dependent transactivation (reviewed in Peterson and Zhang, 2004). A second translocation, t(12:21) which occurs in 25% of paediatric B-cell precursor ALL (common ALL), involves fusion of the N-terminal region of *ETV6* to an intact *RUNX1* locus (Romana *et al*., 1995; Golub *et al*., 1995) and is associated with superior EFS (Borkhardt *et al*., 1997; Maloney *et al*., 1999).

Point mutations in *RUNX1* occur in approximately 10% of sporadic AML (most frequently FAB M0 subtypes) and in 15% of MDS cases. The majority of mutations arise in AML with intermediate risk cytogenetics (predominantly NK-AML) and are often insertions, deletions or nonsense mutations leading to truncation of the RUNX1 protein, with biallelic or double mutations observed in approximately 10% of mutated cases. Mendler *et al*. (2012) reported a higher frequency of mutations in older patients (>60 years), where mutations appeared to cluster in the RHD. In a study of patients aged 18-60 years, Gaidzik *et al*. reported over 50% of *RUNX1* mutations were located in the C-terminal of the gene (exons 6-8) affecting the TAD, again both RHD and TAD mutations were associated with protein truncation. Co-operating mutations varied across the age groups; whilst *MLL*-PTD, *IDH1* and *IDH2* mutations were significantly associated with *RUNX1* mutations in younger patients (Gaidzik *et al*., 2011), *ASXL1* mutations were more

frequently detected in older patients (Mendler *et al*., 2012). Both studies reported a negative association with *NPM1* mutations. Clinical outcomes were poor across both age groups with higher rates of chemo-refractory disease, reduced relapse-free and overall survival.

1.6.2.3 Germline *RUNX1* mutations

The initial discovery of germline *RUNX1* mutations followed linkage analysis in an extensive FPD/AML family (Dowton *et al*., 1985), which identified the chromosomal region of interest at 21q22.1 (Ho *et al*., 1996). Mutations in *RUNX1* were subsequently described in 6 families, all exhibiting the characteristic FPD/AML phenotype (Song *et al*., 1999). The first pedigree was found to transmit an intragenic deletion of *RUNX1*, while the others harboured frameshift or nonsense mutations, suggesting that haploinsufficiency was an important factor in mediating disease.

Since the initial report by Song *et al* (1999), over 30 families have now been described, harbouring a spectrum of lesions including point mutations (missense or nonsense) within both the RHD and TAD. Whole or partial allelic *RUNX1* deletions are less frequently observed (Nickels *et al*., 2013). Investigators have attempted to distinguish the functional consequences of different germline *RUNX1* mutations, although this has been hampered by the limited numbers of patients. Michaud *et al* (2002) reported that lesions in the the RHD significantly reduced or abolished DNA binding, while lesions in the C-terminal were more likely to demonstrate a dominant-negative effect by retaining the ability to bind to DNA and form heterodimers with CBFβ, consequently antagonising WT RUNX1 function. To date, there appears to be no clear distinction in the clinical phenotypes associated with dominant-negative mutations or those predicted to cause haploinsufficiency, although some authors suggest a higher incidence of MDS/AML with the former (Liew and Owen, 2011). Figure 1.6 shows the distribution of germline *RUNX1* mutations reported to date.

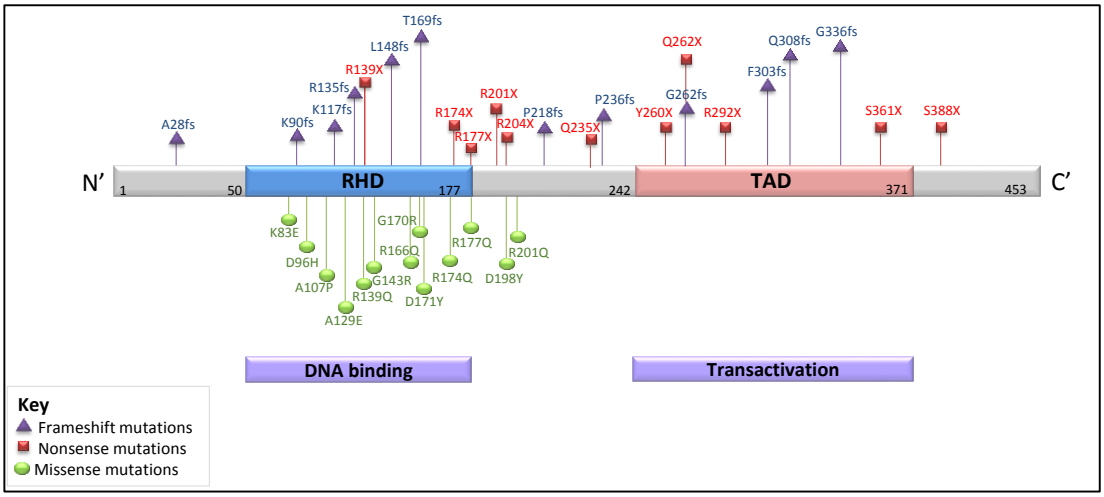


Figure 1.6 Distribution of germline *RUNX1* mutations in familial platelet disorder and myeloid malignancy, figure adapted from Nickels *et al*. (2013).

The majority of mutations localise to the highly conserved RHD associated with DNA binding and include a combination of nonsense, missense and frameshift lesions. In contrast, the TAD domain appears to harbour truncating mutations alone.

1.6.2.4 Penetrance, latency and clinical phenotypes observed in *RUNX1* pedigrees

Germline *RUNX1* mutations lead to haematopoietic malignancy in 40-60% of carriers, with MDS and AML most frequently reported. T-cell ALL and biphenotypic leukaemia have also been described, emphasising the clinical heterogeneity of disease and indicating that tumours may arise from primitive non-lineage committed progenitor cells (Owen *et al*., 2008). It is likely that secondary molecular events are therefore important determinants of the disease phenotype. Preudhomme *et al* (2009) described secondary lesions involving the WT *RUNX1* allele or duplication of the germline mutated *RUNX1* allele in 6 unrelated cases, all of whom developed AML, suggesting that RUNX1 may act as either an oncogene or tumour suppressor in FPD/AML.

The manifestations of familial platelet disorder are also highly variable, often making this syndrome difficult to recognise. Carriers of the same *RUNX1* mutation may display heterogeneity in their degree of platelet dysfunction, with some family members having moderate thrombocytopenia and/or mucocutaneous bleeding, while others are asymptomatic. Platelet size and morphology are usually normal although platelet aggregation is often reduced, particularly in response to agonists such as collagen and

adrenaline, with features of storage pool deficiency and impaired GpIIb/IIIa activation (Liew and Owen, 2011).

1.6.3 Alternative genetic lesions associated with thrombocytopenia and predisposition to malignancy

1.6.3.1 *ANKRD26*

Thrombocytopenia 2 (THC2 OMIM 188000) is an autosomal dominant disorder, characterised by mild to severe abnormalities in platelet number and a mild bleeding diathesis. In 2011, Pippucci *et al*. reported 9 families with THC2, all of whom harboured base substitutions within the 5'UTR of *ANKRD26*. Further screening efforts identified a total of 21 pedigrees with THC2 and mutations in the 5'UTR of *Ankyrin repeat domain-containing protein 26* (*ANKRD26*) (Noris *et al*., 2011). In the majority of these cases, platelets had defective expression of glycoprotein Ia (GPIa) and reduced alpha granules, with dysmegakaryopoiesis observed in the bone marrow. An increased frequency of haematologic malignancy has also been observed in THC2 families, with approximately 10% developing acute leukaemia (predominantly AML), MDS or CML.

1.6.3.2 *ETV6*

ETV6 encodes the ETS family transcriptional repressor *ETS variant 6* (also known as TEL), which is recurrently translocated with *RUNX1* in paediatric B-cell ALL. Germline *ETV6* mutations were recently identified in 6 families with thrombocytopenia and a predisposition to haematologic malignancy, most notably ALL (Zhang *et al*., 2015; Noetzli *et al*., 2015). In the pedigrees reported by Zhang *et al*., two mutations were localised to the highly conserved C-terminal ETS domain (mediating DNA binding), while the third mutation targeted the central regulatory domain, which regulates DNA binding and also interacts with transcriptional corepressor complexes (Chakrabarti *et al*., 1999). The clinical phenotype of affected individuals was variable with haematological malignancies including MDS, mixed-phenotype acute leukaemia, ALL and myeloma. In addition, epithelial tumours such as skin cancer and colon cancer were also reported. Noetzli *et al*. (2015) described three pedigrees with a phenotype characterised by thrombocytopenia, red cell macrocytosis and B-ALL. As previously described, the

mutations localised to the central regulatory and ETS domains, with recurrence of the p.P214L mutation reported by Zhang *et al*.

In both studies, germline mutation constructs displayed reduced transcriptional repression. Importantly, homodimerisation of ETV6 is required to exert its transcriptional repressor effect. All mutant isoforms were capable of dimerising with WT ETV6, but this led to aberrant cytoplasmic localisation, rather than the normal nuclear distribution, which may explain the dominant negative effect observed in these pedigrees (Zhang *et al*., 2015; Noetzli *et al*., 2015). In light of these studies, further assessment is needed to fully characterise the genetic distribution, clinical phenotypes and disease course associated with germline *ETV6* mutations.

1.6.3.3 *FLI1*

Friend leukemia integration 1 (*FLI1)* is another member of the ETS family of transcription factors, which is expressed primarily in haematopoietic cells and regulates genes expressed in megakaryopoiesis. Mutations within the *FLI1* DNA binding domain were recently identified in three families with thrombocytopenia, excessive bleeding and platelet granule secretion defects similar to those observed with *RUNX1* mutations (Stockley *et al*., 2013). Whilst the bleeding diatheses were similar, it is important to note that no cases of haematological malignancy were reported with *FLI1* mutations. Of note, ETV6 modulates the activity of FLI1 (Kwiatkowski *et al*., 1998) and recent evidence suggests that this subsequently co-operates with RUNX1 to govern the transcriptional regulation of *ANKRD26,* the latter function disrupted by germline mutations within the 5'UTR (Bluteau *et al*., 2014). These findings are highly significant as they signify an underlying molecular network with mutations in several key components associated with familial haematological disease/malignancy. It is therefore highly likely that additional target genes may also reside within this pathway.

1.6.4 Familial AML with *CEBPA* mutations

Germline mutations in the transcription factor *CEBPA* were first reported by Smith *et al* (2004). This discovery followed the presentation of two siblings with *de novo* AML, both harbouring an identical frameshift mutation (p.P23fs, 212delC) in the N-terminal of the gene. An acquired C-terminal *CEBPA* mutation was also detected in leukaemic DNA from each sibling with reversion to a WT sequence during remission (Smith *et al.*, 2004). Since then a further 10 pedigrees have been reported (Sellick *et al.*, 2005; Pabst *et al.*, 2008; Renneville *et al*., 2009; Nanri *et al*., 2010, Taskesen *et al.*, 2011; Stelljes *et al.*, 2011; Xiao *et al.*, 2011; Debeljak *et al.*, 2013) although it is estimated that a further 5-10% of sporadic *CEBPA*dm AML patients may harbour an underlying germline mutation, in studies where comparison of matched leukaemic and germline DNA has been possible (Taskesen *et al*., 2011; Pabst *et al*., 2008).

1.6.4.1 Current investigation of *CEBPA*-mutated leukaemogenesis

CEBPA represents another important link in the regulatory network of genes associated with familial leukaemia. In mouse studies, Runx1 binds to Cebpa via a conserved site within the promoter and 4 distal enhancer sites located in a 450-bp region at +37Kb; furthermore, deletion of *Runx1* was associated with reduced transcription of *Cebpa* leading to a reduction of granulopoiesis (Guo *et al*., 2012). This study highlighted the co-operation between these key transcription factors and further alluded to loss of CEBPA function as a significant mediator of *RUNX1*-mutated leukaemia. The group later reported that the +37Kb murine enhancer also harboured conserved bindings sites for CEBPA, GATA2, SCL, PU.1 and FLI1 and additional ETS factors that were conserved in the human +42Kb promoter, further broadening the potential network of haematopoietic regulators interacting with CEBPA (Cooper *et al*., 2015).

CEBPA is capable of inducing terminal differentiation in multiple cell types through its ability to induce specific gene expression profiles and cell cycle arrest (Nerlov, 2007). Zhang *et al*., (1997) showed that disruption of CEBPA impedes the transition of CMP to GMP, preventing granulocytic differentiation. It was later shown that loss of *Cebpa* in foetal HSCs caused increased repopulation of cells suggesting that CEBPA negatively regulated self-renewal (Zhang *et al*., 2004). In a recent study Ye *et al.* (2013) used an inducible knockout model to show that loss of Cebpa in adult HSCs led to a gene

expression signature resembling that of foetal liver HSCs with increased proliferation partially caused by reduced transcriptional repression of *N-Myc*.

It is unclear how specific *CEBPA* mutations disrupt its function and initiate AML. The distinct pattern of N- and C-terminal mutations in *CEBPA*dm AML, suggests that these lesions have distinct cooperating roles and knock in mutagenesis in mice provided a model with which to observe the haematopoietic abnormalities arising from these mutations. Bereshchenko *et al*. (2009) reported that mice harbouring the recurrent C-terminal p.K313dup mutation demonstrated an expansion of pre-malignant HSCs in all phenotypic subsets, although homozygous expression of this mutation blocked lineage commitment. In contrast, the p30 CEBPA isoform (representative of N-terminal frameshift mutations) had minimal impact on HSC numbers in the homozygous state and allowed the formation of myeloid progenitors (Kirstetter *et al*., 2008). Mice harbouring both the N- and C-terminal mutations developed AML more rapidly, suggesting molecular synergy through the promotion of HSC expansion and myeloid commitment. Notably, the observed latency of disease suggested that secondary lesions were required for the development of leukaemia (Bereshchenko *et al*., 2009).

As mentioned earlier in this chapter, sporadic cases of AML can harbour single or double *CEBPA* mutations. Lone *CEBPA* mutations are predominantly frameshift or nonsense mutations distributed throughout the gene, however *CEBPA*dm tumours have a characteristic combination of N-terminal frameshift and C-terminal in-frame mutations, the latter targeting the DNA-binding or leucine zipper domains (Green *et al*., 2010; Fasan *et al*., 2013). *CEBPA*dm tumours also have a limited profile of secondary mutations suggesting that fewer secondary hits are required for leukaemia initiation compared with *CEBPA*sm-AML. In a comprehensive genetic analysis of over 200 cases with sporadic *CEBPA*-mutated AML, Fasan *et al*. reported multiple co-operating lesions in the *CEBPA*sm cohort, including *FLT3*-ITD, *NPM1*, *ASXL1*, *IDH1/2* and *TET2* mutations. In contrast, *CEBPA*dm tumours harboured a limited repertoire of mutations in *GATA2*, *ASXL1* and *TET2*, the majority of which were mutually exclusive. The reduced genetic heterogeneity in the latter may contribute to the homogenous presentation of *CEBPA*dm tumours and their favourable clinical outcomes. While these studies have led to significant advances in our understanding of *CEBPA* mutations in sporadic AML, familial subtypes of AML have

received significantly less scrutiny due to the rarity of patients. A detailed analysis of germline *CEBPA*-mutated AML will be described in Chapter 3.

1.6.5 Familial syndromes associated with *GATA2* mutations

In 2011, germline mutations were initially reported in the haematopoietic transcription factor, *GATA2*. Hahn *et al* (2011) described four families with a 'pure' MDS/AML phenotype, all harboured lesions within the highly conserved second zinc finger (ZF2) domain. Notably, the missense mutation, p.T354M was identified in 3 out 4 families, whilst a 3bp deletion of the adjacent codon (p.T355del) was detected in the fourth family. In the same year, Bödör *et al* (2011) reported two cousins who presented to St. Bartholomew's Hospital with high grade MDS transforming to AML. Although earlier pedigrees had noted little in the way of acquired genetic lesions, both affected cousins in this family possessed an identical somatic *ASXL1* mutation (p.G646WfsX12) in addition to monosomy 7. These aberrations are associated with adverse outcomes in sporadic MDS/AML and both patients died shortly after allogeneic HSCT, one from sepsis and the other from disease relapse. Whilst clinical outcomes have not collectively been reported for familial *GATA2*-mutated patients, many cases now appear to harbour adverse prognostic features, such as trisomy 8 or monosomy 7, and leukaemic transformation of MDS.

Germline *GATA2* mutations can give rise to a several disease phenotypes, often with significant overlap in their clinical presentation. Hsu *et al*.(2011) and Dickinson *et al*. (2011) independently reported a pattern of immunodeficiency (Immunodeficiency 21, OMIM 614172) associated with *GATA2* mutations. The syndrome was termed MonoMac (monoctopenia and mycobacterium avium complex) and DCML (Dendritic cell, monocyte, B- and NK-lymphoid) deficiency by the respective groups, with both representing subtle phenotypic variations of the same clinical entity. As the latter term suggests, this syndrome is characterised by monocytopenia, B- and NK-cell lymphopenia, dendritic cell deficiency and an increased risk of mycobacterial, viral and fungal infections, in addition to pulmonary alveolar proteinosis and a predisposition to MDS/AML (Hsu *et al*., 2011; Dickinson *et al*., 2011).

The mutations described in these families are distributed throughout *GATA2* and this may contribute to the observed clinical heterogeneity, which contrasts with initial reports of 'pure' MDS/AML (Hahn *et al*. 2011; Bödör *et al*. 2011). Furthermore, the disruption of GATA2 structure and function may differ significantly, with a combination missense mutations, insertions and deletions all causing premature termination of the protein or a dominant negative effect on the WT isoform (Hsu *et al.*, 2011; Dickinson *et al.*, 2011; Vinh *et al.,* 2010). In addition to coding mutations, heterozygous deletion of a conserved intronic composite element consisting of an E box, a spacer and a *GATA* motif was recently reported in a patient with MDS and mycobacterial infection. Murine modelling of this deletion showed that homozygous mice demonstrated embryonic lethality with haematopoietic stem/progenitor cell (HSPC) depletion and impaired vascular integrity. In contrast, heterozygous mice were viable but deficient in foetal liver HSPC numbers and colony forming activity, with reduced expression of *GATA2* and its target genes (Johnson *et al*., 2012).

Emberger syndrome (OMIM 614038), characterised by lower limb lymphoedema, sensorineural hearing loss and a propensity to myeloid malignancy represents an additional clinical entity associated with *GATA2* mutations. Truncating mutations in the N-terminal of the gene have been described in addition to missense mutations in the ZF2 domain (Ostergaard *et al.*, 2011). *GATA2* mutations may also have a more insidious presentation characterised by chronic mild neutropenia (often associated with mild-moderate monocytopenia, B- and NK-cell deficiencies). Pasquet *et al*. identified 7 families harbouring germline *GATA2* mutations and presenting with a mild chronic neutropenia which, upon detailed investigation, was often associated with monocytopenia and/or additional features associated with the *GATA2*-mutated immunodeficiency states. These findings emphasise the importance of comprehensive analysis of all haematopoietic cell subsets in individuals with *GATA2* mutations.

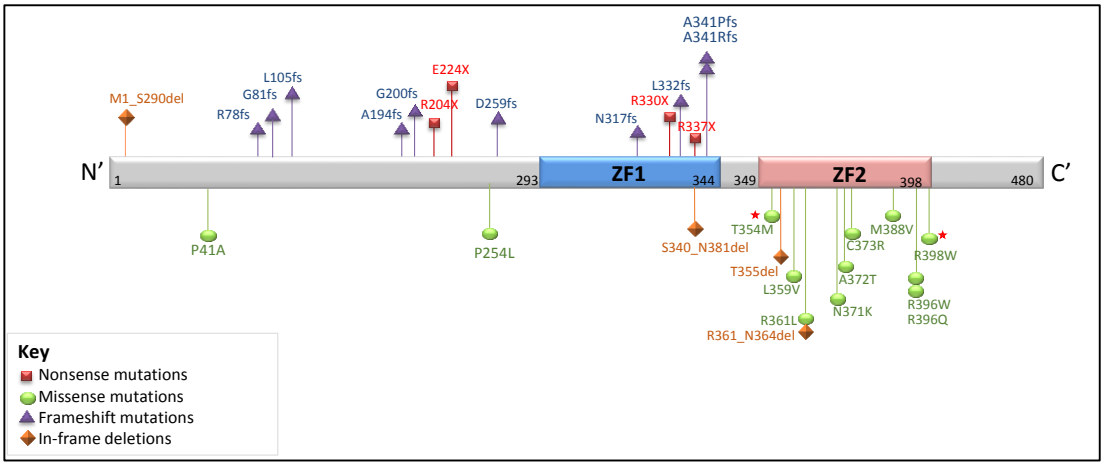


Figure 1.7 Distribution of *GATA2* mutations associated with leukaemia predisposition, adapted from Nickels *et al*. (2013) and Hyde *et al*. (2011).

Recurrent missense mutations at T354 and R398 (red star) are associated with MDS/AML and MonoMAC phenotypes, respectively. Nonsense and frameshift mutations occurring in the N-terminal are more frequently associated with MonoMAC and Emberger phenotypes.

1.6.5.1 The role of GATA2 in the regulation of haematopoiesis

GATA2 is expressed in HSPCs and in differentiated cells such as megakaryocytes, eosinophils and mast cells. Its key function in the haematopoietic system is to regulate the proliferation and survival of HSCs and to orchestrate cell fate decisions, with a critical role in the transition from haemogenic endothelium to HSCs, with homozygous knockout causing lethality due to a failure of definitive haematopoiesis (Tsai *et al.*, 1994; Tsai *et al.*, 1997). GATA2 is predominantly expressed in mature megakaryocytes, mast cells and monocytes with cell fate determined by interactions with CEBPA, GATA1 and PU.1. In particular, GATA1 mediated displacement of GATA2 from chromatin (a process known as the GATA-switch) critically determines the survival and differentiation of erythroid and megakaryocytic lineages (Vicente *et al.*, 2011 and Collin *et al.*, 2015).

GATA2 forms a heptad regulatory unit with six other transcription factors (SCL, LYL1, LMO2, ERG, FLI1 and RUNX1) with direct protein-protein interactions between GATA2, SCL, RUNX1 and ERG to stabilise binding of the complex to DNA (Wilson *et al.*, 2010). Specific interactions with FOG-1 and PU.1 are mediated via the two highly conserved zinc finger domains, ZF1 and ZF2 respectively, suggesting differential regulation of haematopoiesis through these domains (Vicente *et al.*, 2011). The level of GATA2 expression also impacts on haematopoietic regulation with heterozygous loss causing a

reduction in the granulocyte-macrophage progenitor (GMP) population (Rodrigues *et al*., 2008), while enforced expression impairs haematopoiesis by increased quiescence and reduced proliferation of HSCs (Tipping *et al*., 2009). Notably, Wilson *et al* (2010) reported a significant interplay between Runx1 and Gata2 expression in the development of primitive haematopoiesis, demonstrating non-viability of Runx1 +/-:Gata2 +/- compound heterozygous mice, with severe embryonic foetal liver haematopoietic defects demonstrated at mid gestation. Thus we can infer that complex qualitative and quantitative interactions between key haematopoietic regulators are essential in governing the development of early haematopoiesis and this further highlights how missense or truncating germline mutations within these key factors may predispose to the development of haematopoietic malignancy.

1.6.6 Germline *DDX41* mutations

Mutations in the RNA helicase DEAD box polypeptide 41 (*DDX41*) were most recently associated with an inherited predisposition to MDS/AML. *DDX41* is located on chromosome 5q35 and mediates RNA splicing, representing a novel genetic subgroup within the field of familial MDS/AML. Polprasert et al (2015) described a distinctive clinical phenotype, predominantly characterised by late onset MDS in individuals > 60 years. Notably, germline *DDX41* mutations were reported in seven families with a history of MDS/AML, five of these harboured the recurrent p.D140fs lesion, which also occured at a frequency of 0.00008% in the National Heart, Lung and Blood Institute (NHLBI) GO exome sequencing project (ESP) database. This suggests asymptomatic carriage of the germline variant within a control population and may reflect the prolonged latency of MDS/AML associated with this genetic subtype. In 50% of affected family members with *DDX41* mutations, disease progression was associated with the acquisition of somatic mutations or deletions of the WT *DDX41* allele, as previously described with germline *CEBPA* (Smith *et al*. ,2004) and *RUNX1* mutations (Renneville *et al*., 2009).

1.6.7 Inherited bone marrow failure syndromes

The inherited bone marrow failure (IBMF) syndromes are a heterogeneous group of disorders with an inherited predisposition to aplastic anaemia, haematopoietic and solid organ malignancies and characteristic physical features. Individuals usually present in childhood, however the variable disease phenotypes and penetrance of predisposing mutations lead to a significant number of individuals being diagnosed later in life, often with bone marrow failure or MDS/AML as the presenting feature. As IBMF disorders often confer hypersensitivity to chemotherapy, their clinical recognition is essential for appropriate modification of transplant conditioning regimens, avoiding morbidity and mortality from excessive toxicity.

Fanconi Anaemia (FA), the most common IBMF syndrome, is an autosomal recessive disorder characterised by multiple physical features including short stature, thumb or radial ray anomalies, microcephaly, micropthalmia, café au lait spots and genitourinary malformations. Bone marrow failure, of varying degrees of severity, usually presents in the first or second decade of life and patients are also predisposed to the development of haematological and solid organ malignancy. Prompt recognition and diagnosis are essential for appropriate clinical investigation, management and familial screening; this requires the demonstration of chromosomal instability/breakage in lymphocytes exposed to mutagens such as diepoxybutane (DEB) or mitomycin C (MMC) (Sasaki *et al*., 1973).

Seventeen genes have been associated with the Fanconi phenotype since 1992 and these are involved in a common DNA repair signalling pathway, the FA/BRCA pathway, which closely cooperates with other DNA repair proteins (Moldovan *et al*., 2009). Two thirds of cases are due to *FANCA* mutations, whilst *FANCG* and *FANCC* are responsible for a further 25% and *FANCE* and *FANCF* for 8%. *FANCD1*, also known as *BRCA2*, is associated with a strong cancer predisposing FA phenotype (including medulloblastoma and Wilm's tumour) often preceding the onset of aplastic anaemia and/or leukaemia (Reid *et al*., 2005). Biallelic mutations of *PALB2*, the partner of *BRCA2*, are also associated with FA, conferring a similar phenotype to that of *BRCA2/FANCD1* mutations (Reid *et al*., 2007). It is important to note that, in contrast to other leukaemia predisposition syndromes, heterozygous carriage of Fanconi complex mutations is not associated with an increased

risk of acute leukaemia, although solid tumours such as breast cancer may occur with *BRCA2* and *PALB2* mutations (Antoniou *et al*., 2014).

Telomere breakage disorders (TBDs) are caused by mutations affecting genes in the telomerase complex, which lead to abnormal telomere maintenance. *DKC1* is situated on the X chromosome and was first identified in individuals with the X-linked recessive disorder, Dyskeratosis Congenita (DC, Heiss *et al*., 1998). This syndrome presents with a diagnostic triad of nail dystrophy, abnormal skin pigmentation and oral leukoplakia with bone marrow failure and a predisposition to malignancy. Mutations in two further components of the telomerase complex, *TERT* (telomerase reverse transcriptase, Yamaguchi *et al*., 2005) and *TERC* (telomerase RNA component, Vulliamy *et al*., 2001) lack the mucocutaneous manifestations seen in DC and affected individuals present with BM aplasia or MDS/AML, inherited in an autosomal dominant fashion with variable penetrance. The genetic and phenotypic complexity of telomeropathies has been further elucidated by the discovery of mutations in alternative genes including *NOP10* (Walne *et al*., 2007), *NHP2* (Vulliamy *et al*., 2008), *TINF2* (Savage *et al*, 2008), *C16ORF57* (Walne *et al*., 2010), *TCAB1* (Zhong *et al*., 2011) and *RTEL1* (Walne *et al*., 2013).

1.7 The introduction of next generation sequencing in cancer genomics

1.7.1 Novel technologies for discovery genomics

The emergence of NGS technologies has provided the opportunity to perform both comprehensive and unbiased investigation of the genome. This has led to the discovery of multiple novel genetic lesions and pathways associated with inherited and sporadic disorders and, consequently, our understanding of the molecular basis of disease is rapidly advancing. To date, thousands of cancer genomes and exomes have been sequenced and analysed worldwide, with large consortia now forming public databases enabling the advancement of further research. In particular, the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas Research Network (TCGARN), now provide data for the coding variants in >12, 000 cancers with the Catalogue of Somatic Mutations in Cancer database (COSMIC) being regularly updated with the novel variants identified from these extensive sequencing projects. To analyse this exponential growth of NGS data, multiple bioinformatic algorithms have been developed for the assembly, annotation and comparison of genomes and exomes to comprehensively detail genetic variants, copy number aberrations (CNA) and LOH. The future interpretation of this data will involve in-depth functional and clinical-pathological correlation and perhaps most importantly, the selection of molecular targets for novel therapeutic development.

WES is a powerful and cost effective tool for deciphering the molecular basis of disease. In contrast to WGS, WES covers only 1% of the genome allowing for rapid sequencing of protein-coding regions. Whilst non-coding regions of the genome may reveal important insights into leukaemogenesis, the interpretation of data is less clear at the present time and will undoubtedly require functional evaluation. A number of different platforms are commercially available for WES, although the underlying concept remains similar to all; DNA is digested into small fragments, coding regions are then targeted by RNA baits, captured and sequenced repeatedly generating millions of paired end reads. A greater depth of sequencing improves sensitivity for low frequency variant detection permitting more comprehensive analysis of tumour clonality and architecture.

1.7.2 WES in the investigation of familial leukaemia

Within recent years, WES has led to the identification of causal alleles for multiple Mendelian disorders. The propensity to familial leukaemia also reflects a Mendelian trend of autosomal dominant inheritance with variable penetrance. Notably, over 40% of cases investigated to date fail to demonstrate a germline mutation in the known genes. This suggests that novel, unique lesions underlie the inherited predisposition, and efforts to identify these variants are essential for the provision of accurate genetic counselling, screening and management of patients and their families.

The use of WES to investigate causal alleles in several families with MDS/AML will be described in Chapter 5. WES identifies over 20,000 variants in any one individual and >95% of these are recognised single nucleotide polymorphisms (SNPs), occurring in cohorts of presumed healthy individuals and are therefore not considered pathogenic (Bamshad *et al*., 2011). Whilst discrete filtering with SNP databases can reduce the number of variants significantly (only 2% of SNVs in an individual exome are considered novel), further narrowing of the search is required. One approach is to sequence multiple unrelated individuals with a similar phenotype and to identify the novel alleles that segregate with disease, this methodical approach was elegantly demonstrated by Ng *et al* who first described the association of *MLL2* mutations with Kabuki syndrome (Ng *et al*., 2010). For Mendelian disorders, such as familial MDS/AML, pedigree information is also helpful and sequencing of the most distantly related affected individuals (sharing limited inherited variants) may substantially restrict the genomic search.

An additional question remains regarding the acquired (somatic) lesions initiating disease within family members. The investigation of familial leukaemogenesis provides an unrivalled model for understanding the stepwise initiation of leukaemia in humans. While the majority of current research focuses on sporadic disease in unrelated individuals (or our attempts to recapitulate this with genetically engineered mouse models), the investigation of disease evolution across family members sharing a common genotype enables us to determine whether inherited factors may govern the mutations acquired. This question has far reaching implications and challenges the prevailing

dogma that tumourigenesis occurs through the random acquisition and selection of mutated clones.

1.8 Aims and Objectives

The aims of this thesis are to employ in-depth genetic profiling techniques to investigate the somatic mutations and models of leukaemic evolution within families with an inherited propensity to MDS/AML. In particular, this work aims to identify features which distinguish familial disease evolution from that of sporadic AML and may confer unique biological properties specific to inherited forms of disease. By studying somatic mutation profiles within and between families, we hope to study trends of molecular co-operation and determine whether the acquisition of somatic mutations is, in part, governed or 'predetermined' by the host's genotype, challenging the current stochastic view of tumour evolution. As a significant number of familial cases harbour unknown genetic lesions, we aim to characterise novel or atypical germline lesions, expanding the current repertoire of genes and mutations in familial leukaemia. It is hoped that this multifaceted approach will help improve our understanding of familial leukaemia evolution and may also provide valuable insights for sporadic leukaemogenesis.

Chapter 2

Materials and Methods

2.1 Patients and sample collection

Samples were collected via national and international referrals from clinicians managing patients suspected to have an inherited leukaemia predisposition syndrome. Informed, written consent was obtained under the application 06/Q0401/31, while samples obtained from collaborating centres had local ethical approval. Tumour DNA samples were extracted from bone marrow (BM) and peripheral blood (PB) at diagnosis and relapse of MDS/AML. Constitutional or germline variants were assessed using PB or BM DNA sampled during confirmed remission, where possible salivary or skin DNA were also tested. PB and/or salivary DNA were tested in asymptomatic individuals. All pedigrees investigated in this study are described in detail within the relevant chapters of this thesis although an overview of cases is provided in Table 2.1.

Table 2.1 Summary of pedigrees and cases investigated

Family	Country	Relevant clinical history	Chapter	Germline mutation
A	UK	A.I.1, male, AML 10y. Son (A.II.1) AML 30y and daughter (A.II.5) AML 18y. Son of A.II.5 (A.III.2) AML 2 y	3	*CEBPA*
B	France	B.I.1, female, AML 23y. Son (B.II.2) AML 5y.	3	*CEBPA*
C	Serbia	Sisters: C.II.1 AML 32y and C.II.2 AML 3y. Daugther of C.II.1 (C.III.1)AML at 18y	3	*CEBPA*
D	Slovenia	Two female identical twins: D.II.1 AML 18 months, D.II.2 AML 15 y	3	*CEBPA*
E	Japan	E.I.1, male, AML 39y, one affected son (E.II.1) AML 26y and another (E.II.2) asymptomatic carrier (24y)	3	*CEBPA*
F	UK	F.III.2, male, AML 34 y, two affected sons: F.IV.2 AML 25y and F.IV.4, AML 24y. Son of F.IV.2 (F.V.1) AML 4y	3	*CEBPA*
G	Germany	G.II.2, female, AML at 28y. Son (G.III.1) AML 2y.	3	*CEBPA*
H	Germany	H.II.2, female, AML. Daughter (H.III.2) AML 25y	3	*CEBPA*
I	Switzerland	I.I.2 AML 46y. Daughter (I.II.2) AML 40y. Son of I.II.2 (I.III.3) asymptomatic carrier 19y.	3	*CEBPA*
J	Switzerland	J.I.2, male, AML 42y. Daughter (J.II.1) AML 27y.	3	*CEBPA*
K	Hungary	K.I.1 female asymptomatic carrier. Four children: K.II.1 and K.II.2 (non-identical twins) sAML 5y, K.II.3 MDS 14y, K.II.4 sAML 5y.	4	*RUNX1*
L	UK	L.I.2, male, AML, died 67y. Son (L.II.2) thrombocytopenia in early 50's and daughter (L.II.4) AML died 7y. Daughter of L.II.2 AML died 15y	5	*RUNX1*
M	Spain	M.I.2, female, CMML transformed to AML. Son (M.II.3) childhood thrombocytopenia then AML, daughter (M.II.5) thrombocytopenia then AML 38y, son (M.II.7) thrombocytopenia. Older daughter of M.II.3 (M.III.1) thrombocytopenia, younger daughter (M.III.2) thrombocytopenia 8y then AML 14y. Both sons of M.II.5 (M.III.3 and M.III.4) have thrombocytopenia.	5	*RUNX1*
N	UK	N.I.1, male, FPD/mucosal bleeding disorder (normal plt count). Son (N.II.1) AML 13y, T-ALL 14y.	5	*RUNX1*
O	UK	O.II.1 (single case), male 3y, global developmental delay, hypospadias, thrombocytopenia.	5	*RUNX1*
P	Norway	P.I.2 , male,' leukaemia' 50y, daughter (P.II.1) MDS/AML 37y. Son of P.II.1 (P.III.2) sAML/RAEB2, 42y. Sons of P.III.2; P.IV.1, thrombocytopenia 18y and P.IV.2, CMML, 8y	5	Unknown
Q	France	Q.I.1 AML, 53y with relapse after 8 months; daughter (Q.II.1) AML 36y	5	Unknown
R	Canada	R.I.1, male, MDS/AML. Three sons with MDS/AML R.II.1, R.II.4 and R.II.6. The sons of R.II.1 (R.III.2) and R.II.6 (R.III.4) both developed MDS	5	Unknown
S	France	Mother and daughter S.I.1 and S.II.1 with macrocytic anaemia and thrombocytopenia	5	Unknown
T	Belgium	Two brothers: T.II.1 AML 16y and T.II.2 MDS at 11 months	5	Unknown
U	France	Two brothers: U.II.1, AML 77y and U.II.2, MDS (RCMD) 67y	5	Unknown
V	Belgium	Two brothers: V.II.1 aplastic anaemia 13y, V.II.2 sAML 6y	5	Unknown
W	Belgium	Two siblings: W.II.1 B-ALL 3y, W.II.2 Burkitt's leukaemia/lymphoma 3y	5	Unknown
X	Canada	Two brothers: X.II.1 AML 62y (HSCT from healthy sister X.II.3), X.II.13 MDS (RAEB2)- 5 years post sibling allo-HSCT from healthy sister X.II.2. Paternal uncle X.I.1 NHL, 58y	5	Unknown
Y	UK	Mother (Y.I.1) AML 56y. Daughter (Y.II.1) AML 42y (died 43y, due to intracranial haemorrhage post VUD-HSCT)	5	Unknown
Z	UK	Two sisters: Z.II.1 AML 60y, Z.II.2 hypoplastic MDS, 61y	5	Unknown
Additional families				
1	Netherlands	Male (II.7) MDS/AML and EBV-associated T-NHL. Daughter (III.6) Sino-pulmonary infections, veruccae, MDS and niece (III.10) lymphoedema, warts and MDS	5	*GATA2*
2	UK	Female (IV.1) AML 5y , asymptomatic mother (III.2) and maternal grandfather (II.2), maternal great uncle 'acute leukaemia', 53 y	5	*CEBPA*

2.2 DNA extraction

2.2.1 Genomic DNA extraction

Genomic DNA was extracted from PB and BM samples using the DNeasy Blood & Tissue Kit (Qiagen) as per the manufacturer's instructions. Samples had initially been enriched for mononuclear cells by density centrifugation, prior to cryopreservation and storage in liquid nitrogen. After thawing, the mononuclear cell sample was centrifuged and re-suspended in 200µl phosphate buffered saline (PBS, pH7.2 50mM potassium phosphate, 150mM NaCl). Proteinase K and lysis buffer (Buffer AL) were added to the cell suspension prior to vortexing and incubation at 56°C for 10 minutes. The addition of 100% ethanol allowed DNA to precipitate and, after sequential washes with ethanol containing buffers (AW1 and AW2), the purified DNA sample was eluted in 50-100ul of elution buffer (or Tris-EDTA buffer, TE, pH8) prior to long-term storage at -20^{o}C.

2.2.2 DNA extraction from BM aspirate slides

A modification of the above technique was used to extract DNA from archived unstained BM aspirate slides. In this case, 20ul of TE buffer was applied directly to the slide and the sample was scraped into a 1.5ml microcentrifuge tube with a sterile scalpel. Following the addition of 180ul tissue lysis buffer (Buffer ATL, Qiagen) and 20µl proteinase K solution (Qiagen), the sample was incubated overnight at 56^{o}C. Lysis buffer (Buffer AL) was then added and the sample incubated at 70^{o}C for 10 minutes. DNA precipitates formed with the addition of 100% ethanol and the sample was washed as per the manufacturer's protocol for DNA purification from dried blood (Guthrie) spots using the QIAamp DNA mini Kit (Qiagen). To maximise elution efficiency, the elution buffer was warmed to 55^{o}C and incubated with the sample for 1 minute, prior to the final centrifugation step.

2.2.3 Salivary DNA extraction

Salivary DNA was extracted from the Oragene OG-500™ collection kit using PT-L2P solution™ (DNA Genotek). Samples were transferred from the collection kit to sterile 15ml conical centrifuge tubes and incubated in a water bath at 50^{o}C for a minimum of 1 hour. A 500µl aliquot of each sample was transferred to a sterile 2ml microcentrifuge tube and 20µl of the PT-L2P reagent was added. Sample mixtures were incubated on ice

for 10 minutes to precipitate impurities, followed by centrifugation at 13000 rpm for 5 minutes. The DNA-rich supernatant was transferred to a clean microcentrifuge tube and 600ul of 100% ethanol was added for DNA precipitation. The sample was centrifuged at 13000 rpm for 1 minute and the supernatant discarded, without disrupting the pellet. DNA pellets were washed with 250µl 70% ethanol and then eluted in TE buffer.

2.3 Nucleic acid quantification

Nucleic acid quality or purity was assessed using a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, USA) to determine 260nm/280nm and 260nm/230nm absorption ratios. A 260/280 ratio of ~1.8 or higher was suggestive of 'pure' DNA, while lower ratios indicated contamination by proteins or phenols. 260/230 ratios were secondary measures of nucleic acid purity, with optimal values ranging from 1.8-2.2, ratios considerably lower than these reflected co-purified contaminants.

2.4 Screening of established familial MDS/AML loci

2.4.1 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify specific regions of genomic DNA (ranging between 200-700bp in length) flanked by forward and reverse oligonucleotide primers. Primers were designed using the Primer3Plus software (http://www.primer3plus.com/) and were subjected to In-Silico PCR analysis (UCSC Genome bioinformatics) to confirm their specificity for the genomic region of interest. The primers used for Sanger sequencing of *RUNX1*, *CEBPA*, *GATA2 TERC* and *ETV6* are shown in Table 2.2. In addition to amplifying coding regions of *GATA2*, primers were also designed to amplify the regulatory cis element located at + 9.5kb in IVS4 (intron 4), following recent reports of a heterozygous 28bp deletion of this region in a patient with MonoMac syndrome (Johnson *et al*, 2012).

The majority of reactions were performed using 1.1x ReddyMix™ (Life Technologies) which included ThermoPrime Taq DNA Polymerase, dNTPs, PCR reaction buffer, magnesium chloride (1.5mM) and an inert red loading dye for visualisation of PCR products in gel electrophoresis. A 10µl reaction was prepared with 8 µl ReddyMix, 0.5µl of forward and reverse primers (final concentration 0.5µM per primer) and 1µl template DNA (at 25-50ng/µl). A non-template control (NTC) was included in all experiments to exclude reagent contamination.

Standard PCR cycling conditions consisted of an initial denaturation step at 95oC for 15 minutes, followed by 35 cycles of DNA denaturation at 95^{o}C for 30 seconds, primer annealing at 58^{o}C for 30 seconds and extension at 72^{o}C for 30 seconds, with a final extension step of 72^{o}C for 5 minutes. These conditions were successfully used for Sanger sequencing of the majority of variants within this study, including *TERC*, *GATA2* and *ETV6*. Important modifications were required for PCR amplification of *RUNX1* and *CEBPA*, as described below.

2.4.1.1 Reagents and cycling parameters for *RUNX1* and *CEBPA* PCR amplification

RUNX1 exons 3-8 (ENST00000344691) were amplified using Megamix™ (Microzone) which contained recombinant Taq polymerase in a 1.1x reaction buffer with 2.75mM MgCl2 and 250µM dNTPS. The total reaction volume was 20µl (with 1µl forward and reverse primers (10µM of each primer, final concentration 0.5µM) and 1µl template DNA (at 25-50ng/µl). Cycling parameters for exons 3-7 required a gradient based reduction in annealing temperature, while exon 8 was amplified with a fixed annealing temperature of 58°C (Table 2.3 and 2.4, respectively). Due to the high GC content in *CEBPA*, PCR amplification was performed using the AmpliTaq Gold™ 360 mastermix (Applied Biosystems) with the addition of GC-enhancer (5µl), nuclease free water (4.5µl), forward and reverse primers (1µl of each, final concentration 0.4µM in a total reaction volume of 25µl). The thermal cycling conditions for *CEBPA* are shown in Table 2.5.

Table 2.2 Oligonucleotide Primers for PCR of *RUNX1*, *GATA2*, *CEBPA*, *TERC* and *ETV6*

Gene (Ensembl transcript)	Exon	Forwards primer sequence (5'to 3')	Reverse primer sequence (5' to 3')
RUNX1 ENST00000344691	3	GAGCTGCTTGCTGAAGATCC	CTTTCCAGAATCCGGCCC
	4	CATTGCTATTCCTCTGCAACC	AAGACAGACCGAGTTTCTAGGG
	5	TGAAGCAAAGCCAAAATTCC	TCTGAGACATGGTCCCTGAG
	6	AAGAAAAGCCCCAGTTTTAGG	AGTTGGTCTGGGAAGGTGTG
	7	GAACAAGGGCCACTCATTTC	AAGAAAATCAGTGCATGGGC
	8	TCCGTTCTCTTGCCCGC	GGCCTGGCGCCTCAGTA
GATA2 ENST00000341105	2	ACACCTCGTGGTGGGACTT	CGCCTGGGTTCTCATCAC
	3	CTGGTTCTGGGAGTCGTGAT	ATCTGGGAAACCAACACTGC
	4	GACTCCCTCCCGAGAACTTG	GCGTCTGCATTTGAAGGAGT
	5	TTAGCCCTCCTTGACTGAGC	AGCCAAGCTGGATATTGTGG
	6	GTTGCTGGAGGAAGGAACTG	AACTGTCCATGCAGGAAACC
	IVS-5	ACTCACACCGGCCACTAAAA	AGCGAGAGGCAGGACTGAG
CEBPA ENST00000498907	F 1	TCGCCATGCCGGGAGAACTCTAAC	AGCTGCTTGGCTTCATCCTCCT
	F2	GCTGGTGATCAAGCAGGAGC	CCGCCACTCGCGCGGAGGTCG
	F3	GGCAGCGCGCTCAAGGGGCTG	CACGGCTCGGGCAAGCCTCGAGAT
TERC ENST00000602385	Linc RNA	GCTCCCTTTATAAGCCGACTC	AGGTTTGGGGGTTCACAAG
ETV6 ENST00000396373	1	GGGAGAGGAAAGGAAAGTGG	AACAGAGCAACTGCAACAGC
	2	CCATTCCAAGCTTTCATTGTC	AAAACCACCTCAGAGCTTGC
	3	TTCCAGCTGTCTAACTGACAAGA	GGGGTGTTAAAGACCAACCA
	4	TCCCTTTCCTTTTTCTTTCCA	CACGAAGAAGACCAGCTTATCA
	5.1	GTCTTTCCCTCTGCTCCACA	GATCAGAGGGTGCATGATGG
	5.2	ACCAGGAGAACAACCACCAG	CTGGCTGCAAAGATCACAGG
	6	GCAGAAGCAGTTGCTGGATT	GGGCAGACGTAAAAAGCAGA
	7	GCAGTGCCTTTTCTGAGGTT	GGGTATCCGATAGGCAGGTA
	8	TGGAATCTCTTACCTCCTCCA	CCTTTTCCACTCTCCTCAGC

2.4.1.2 Agarose gel analysis and purification of PCR products

Agarose gels were used to confirm the presence and appropriate size of amplified PCR products, 2% gels were prepared using 4g of agarose with 200ml of 1x Tris-Borate-EDTA (TBE) buffer (10x TBE: 108.9g Tris base, 55.7g boric acid and 7.4g EDTA in 1 litre of ddH2O). The agarose was dissolved by microwave heating for approximately 4 minutes and, on partial cooling, 20µl of 1 x GelRed™ (Biotium, Hayward, CA) was added. The gel was then poured into a mould with appropriate combs and allowed to set. Once set, 3-5ul volumes of PCR products and a 1kB DNA ladder (exACTGene™, Fisher Scientific) were loaded into individual wells and electrophoresis was performed in a 1 x TBE running buffer at 120V. Visualisation of products was performed under ultra violet (UV) light transillumination and digital and/or Polaroid images were collected.

Table 2.3 PCR thermal cycling parameters for *RUNX1* exons 3-7

Step	Time	Temperature	Cycles
Initial heat activation	5 mins	95°C	
Denaturation	60 secs	95°C	PART A consisting of **20 cycles**: (decreasing by 0.5°C each cycle)
Annealing	60 secs	65°C	
Extension	60 secs	72°C	
Denaturation	60 secs	95°C	PART B consisting of **10 cycles**
Annealing	60 secs	55°C	
Extension	60 secs	72°C	
Final Extension	10 mins	72°C	

Table 2.4 PCR thermal cycling parameters for *RUNX1* exon 8

Step	Time	Temperature	Cycles
Initial heat activation	5 mins	95°C	
Denaturation	60 secs	95°C	**35 cycles**:
Annealing	60 secs	62°C	
Extension	60 secs	72°C	
Final Extension	10 mins	72°C	

Table 2.5 PCR thermal cycling parameters for *CEBPA* fragments 1-3

Step	Time	Temperature	Cycles
Initial heat activation	10 mins	95°C	
Denaturation	30 secs	95°C	**35 cycles**:
Annealing	30 secs	62°C	
Extension	30 secs	72°C	
Final Extension	7 mins	72°C	

2.4.1.3 Enzymatic clean-up of PCR products

Clean-up of PCR products was performed prior to DNA sequencing to remove unconsumed dNTPs and primers. An Exo-SAP stock solution was made by adding 1µl of recombinant Exonuclease I (*E. coli*) (New England Biolabs) and 20µl Shrimp Alkaline Phophatase (SAP, Affymetrix, Cleveland, U.S.A.) to 179µl of nuclease-free water. This mix underwent thermal cycling at 37°C for 15 minutes, 80°C for 15 minutes before storage at 4°C or proceeding straight to the sequencing reaction.

2.5 Sanger sequencing

Purified PCR products were sequenced in both forward and reverse directions using the BigDye™ Terminator Cycle Sequencing kit version 3.1 (Applied Biosystems) on an ABI PRISM 3730 Genetic Analyser (Applied Biosystems). The 10µl sequencing reaction mix was prepared by the addition of 0.25µl Big Dye Terminator mix™, 1.875µl 5x Buffer, 0.5µl of primer (final concentration 0.5µM) and 1µl of the purified PCR product to 6.375µl RN-ase free water. Thermal cycling of this reaction involved an initial denaturation step at 96°C for 1 minute, followed by 26 cycles of denaturation at 96°C (1 minute), annealing at 50°C (15 seconds) and extension at 60°C (4 minutes) before storage at 10°C. Samples were then transferred to CR-UK sequencing services at the London Research Institute for clean-up and sequencing on the ABI3730.

2.6 Identification of mutations and validation

Sequencing traces were analysed and interpreted using the Lasergene SeqMan software package (DNAStar Inc., USA), in combination with BioEdit software (available at http://www.mbio.ncsu.edu/bioedit/bioedit.html). Visual comparison of sequences was performed against reference sequences obtained from the recent version of the human genome 19 (hg19) on the Ensembl genome browser (available at http://www.ensembl.org). In each case, variants were confirmed by a minimum of two independent sequencing reactions. All variants in this thesis were annotated from the first nucleotide of a consensus coding sequence (CCDS) approved cDNA transcript, noting the corresponding protein alteration. This cDNA nomenclature provided consistency with published germline *CEBPA* mutations and was maintained in subsequent chapters.

2.7 Cloning of PCR products into competent *E.coli*

To determine the allelic distribution of mutations and/or characterise the sequence of large insertions or deletions (>10bp), PCR products were cloned into a plasmid vector which was used to transform chemically competent *E.coli*, allowing Sanger sequencing of individual colonies. This was performed using the Original-TA cloning kit (Invitrogen) and the procedure is summarised in Figure 2.1.

PCR amplicons were first generated non-proofreading Taq DNA polymerases (including those in ReddyMix [Life Technologies], AmpliTaq [Applied Biosystems] and HotStarTaq Plus [Qiagen]). The single 3' deoxyadenosine (A)-overhangs on PCR amplicons could adhere to 3' deoxythymidine (T)-overhangs on the pCR2.1 vector, enabling efficient ligation. The ligation reaction required pCR2.1 vector (2µl, 50ng), 1µl T4 DNA ligase, 1µl ligation buffer (10x), 5ul nuclease-free water and 1µl of PCR product (10ng DNA) and was incubated at 14^{o}C overnight on a thermal cycler. A later modification of the protocol included the ExpressLink T4 DNA ligase, requiring a shorter incubation of 1 hour at room temperature.

The cloned vector was then used to transform chemically competent *E.Coli* (OneShot TOP10™ cells, Invitrogen). To each thawed vial of *E.Coli* cells, 2µl of ligation reaction was added and the mixture was incubated on ice for 30 minutes. Cells were then heat-shocked for 30 seconds at 42^{o}C (to promote incorporation of the vector), 250µl SOC medium (2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl2, 10mM MgSO4, 20mM glucose) was added to each vial, prior to incubation at 37^{o}C for 1 hour on a shaking platform at 225rpm. Between 40-80µl of the transformed cells were then spread onto Luria Bertani (LB) agar containing ampicillin (100µg/ml) and coated with Xgal solution (20mg/ml) in dimethylformamide, to enable blue/white screening of colonies. In colonies without inserts, the empty vector enabled production of B-galactosidase within cells, which was subsequently hydrolysed by X-gal, producing an insoluble blue pigment. The appearance of white colonies indicated the vector and insert were both present within the bacterial clones. Following overnight incubation at 37^{o}C, a minimum of 20 white colonies were selected for PCR amplification of the plasmid-DNA construct followed by Sanger sequencing.

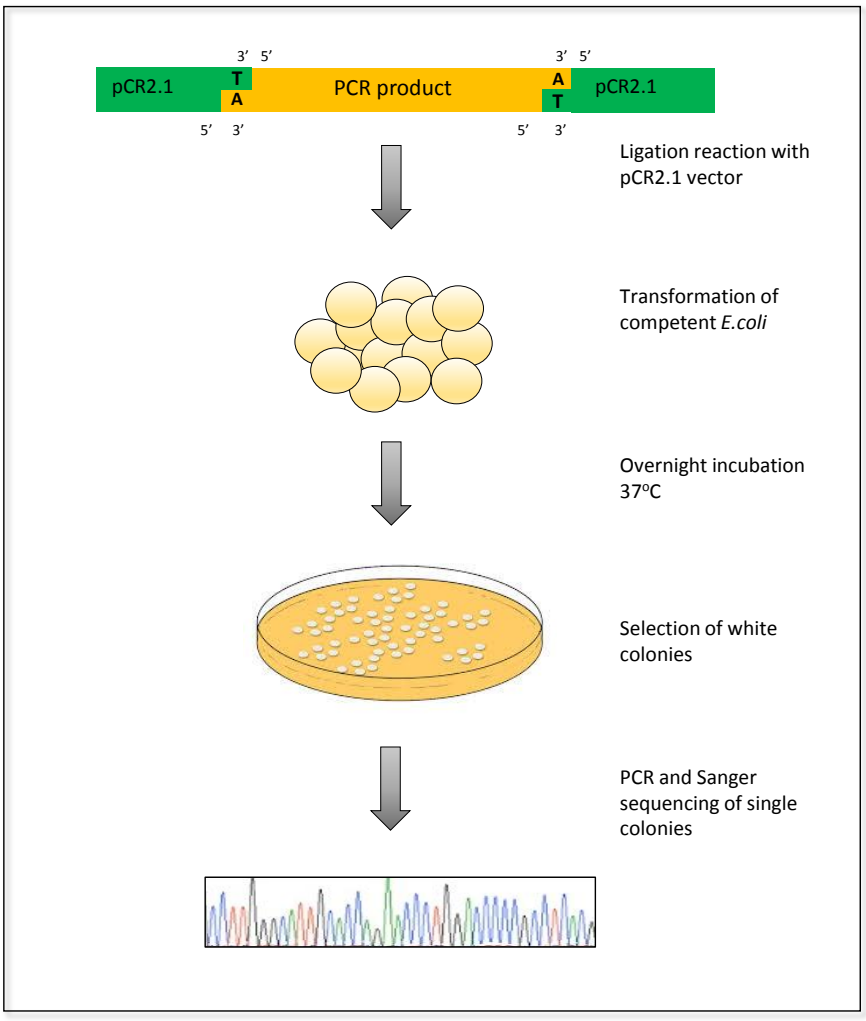


Figure 2.1 Overview of TA-cloning procedure

The cloned pCR2.1 vector containing the PCR amplicon, is used to transform chemically competent E.Coli. Following overnight incubation, bacterial colonies which contain the insert can be distinguished by a white colour, compared with blue colonies containing an empty vector.

2.8 Familial MDS/AML customised aCGH protocol

Array-based comparative genomic hybridisation (aCGH) employs a whole genome oligonucleotide hybridisation technique to identify chromosomal losses and gains. In collaboration with Prof. Brigitte Schlegelberger's group (Institute for Human Genetics, Hannover, Germany) we assessed 10 families with no detectable mutations in *RUNX1*, *CEBPA* and *GATA2* using a customised 2x400k oligo-array (Agilent). This array incorporated a high density of probes within genetic loci associated with leukaemia predisposition syndromes (listed in Appendix 1) whilst standard probes were distributed at 5Kb intervals. Prior to commencing, an equal number of gender matched reference and test DNA samples were assessed by spectrophotometry to ensure a concentration >20ng/ul, 260/280 ratio ≥ 1.8 and 260/230 ratio ≥1.9. A total DNA quantity of 1-1.5µg DNA was considered optimal for this technique.

Random fluorescent labelling of gDNA was performed using cyanine 3-dUTP (green fluorescence-test samples) and cyanine 5-dUTP (red fluorescence –reference DNA samples) with the genomic DNA labelling kit 5190-0449 (Agilent). Care was taken to ensure minimal light exposure to prevent degradation of the photosensitive labels. The labelled DNA samples were stored at -20^{o}C overnight (in foil) and transferred to a purification column, 430 µl 1xTE (pH 8) was added followed by centrifugation for 10mins at 14000g. The wash through was discarded and the step repeated with the addition of 480µl of 1xTE (pH 8). Purified samples were removed from filters by further centrifugation using a clean collection tube. The yield and degree of labelling and specific activity were assessed using Nanodrop spectrophotometry with absorbance criteria specific to each fluorescent label. Matched Cy-3(reference) and Cy-5 (test) labelled DNA samples were then combined to a total volume of 79µl. Following preparation of labelled DNA for hybridisation, 245µl of each hybridisation mix was slowly dispensed onto a clean gasket slide and the microarray slide was placed directly on top and the chamber tightened. Slide chambers were placed in a rotator rack in a hybridisation oven set to 65^{o}C with a rotor speed of 20 rpm for hybridisation over 40 hours.

Following hybridisation, microarray array washes were performed in an ozone free environment with wash buffer 1 used at room temperature whilst wash buffer two was used at 37^{o}C. Microarray slides were then analysed using the SureScan Microarray Scanner (Agilent) and data was subsequently interpreted using the Genomic Workbench program (Agilent). An overview of this process is shown in Figure 2.2 (Agilent, Array based CGH for genomic DNA analysis v7.2).

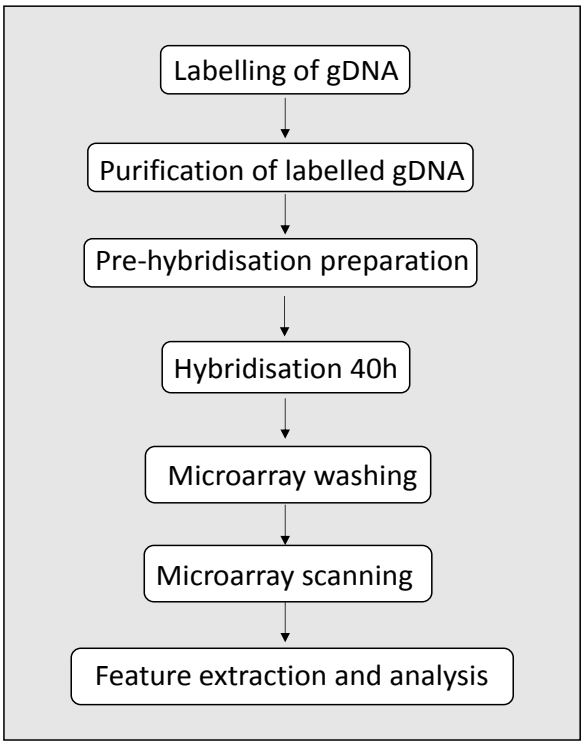


Figure 2.2 Overview of array CGH procedure.

Patient and gender matched control DNA samples are labelled with fluorescent dyes and hybridised to a customised oligonucleotide array for the identification of copy number gains and losses in genes associated with MDS/AML predisposition syndromes.

2.9 Multiplex ligation-dependent probe amplification

Array-based approaches (such as CGH and SNP array) permit high-resolution scanning of the entire genome for CNVs within a single hybridisation reaction, although their use in the clinical setting is limited by equipment and array costs, large DNA requirements and labour-intensive protocols. With the improved characterisation of CNAs in multiple malignancies, an increasing number of customised multiplex ligation-dependent probe amplification (MLPA) kits now offer disease-specific investigation of recurrent chromosomal and genetic CNAs, providing time and cost-effective benefits.

2.9.1 Principles of MLPA

MLPA is used to determine the copy number of up to 60 DNA sequences in a single multiple PCR-based reaction. Each MLPA probe consists of two oligonucleotides that bind to directly adjacent target DNA sequences in a region spanning 55-80 nucleotides (the left and right hybridisation sequences, LHS and RHS respectively). Flanking each hybridisation sequence, there is a primer binding region which is complementary to a universal primer pair capable of amplifying all MLPA probes within the assay. A 'stuffer sequence' may be incorporated between the hybridisation sequences and their respective primer binding regions, to modify the lengths of probes. This allows appropriate separation and identification of amplification products, whilst the use of FAM-fluorescently labelled primers allows their visualisation and quantification with capillary electrophoresis (Schouten *et al*., 2002).

Two different probemixes were used in this study; SALSA MLPA X060-X2 was used to profile somatic CNAs in familial tumour samples, whilst the SALSA MLPA P437-A1 probemix assessed CNAs within known leukaemia predisposition genes. Tables 2.6 and 2.7 show the genomic loci and recurrent point mutations assessed in the probemixes X060-X2 and P437-A1, respectively. In addition, each probemix harboured multiple reference probes, sited at autosomal chromosome locations which are considered relatively stable in MDS and AML.

Table 2.6 SALSA MLPA probemix X060-X2 MDS-AML (MRC Holland)

Length (nt)	SALSA MLPA probe	Chromosomal position			MV location (HG18)
		reference	mutation-specific	other	
64-70-76-82	Q-fragments: DNA quantity; only visible with less than 100 ng sample DNA				
88-92-96	D-fragments: Low signal of 88 or 96 nt fragment indicates incomplete denaturation				
100	X-fragment: Specific for the X chromosome				
105	Y-fragment: Specific for the Y chromosome				
126	Reference probe S0959-L26042	9q21.13			09-078.132782
134 Δ	Reference probe 13867-L23116	16p13.2			16-008.7654561
138 ±	**TP53 probe** 08304-L21074			17p13.1	17-007.517784
142 §	**FLT3 probe** 18510-L24287		**D835Y** (c.2503G>T)		13-027.490602
148	Reference probe 17858-L23867	2p21			02-044.400839
154	**NF1 probe** 12018-L12866			17q11.2	17-026.708124
160 § ⁂ ‡	**ASXL1 probe** 18261-SP0848-L26260		**G646fs*12** (c.1934dupG)		20-030.486043
166 «	**ASXL1 probe** 18512-L23803			20q11.21	20-030.409847
173	**JARID2 probe** 18513-L24441			6p22.3	06-015.604650
179 «	**SUZ12 probe** 18514-L23805			17q11.2	17-027.347941
185	**KMT2A probe** 09500-L09757			11q23.3	11-117.897898
193 § ⁂	**IDH2 probe** 16925-SP0403-L20941		**R140Q** (c.419G>A)		15-088.432856
197	**KMT2A probe** 17085-L23868			11q23.3	11-117.852750
203 §	**IDH1 probe** 17381-L26266		**R132H** (c.395G>A)		02-208.821304
208	Reference probe 13384-L25019	6q12			06-064.998570
214 § ⁂	**NPM1 probe** 18516-SP0880-L26748		**W288fs*12** (c.863_864insTCTG)		05-170.770055
220 «	**U2AF1 probe** 18517-L23808			21q22.3	21-043.386385
227 §	**IDH1 probe** 14787-L23869		**R132C** (c.394C>T)		02-208.821324
234 § ⁂	**SF3B1 probe** 17802-SP0549-L21900		**K700E** (c.2098A>G)		02-197.975049
240	**CUX1 probe** 19886-L26753			7q22.1	07-101.527414
246	**EZH2 probe** 18267-L26041			7q36.1	07-148.160619
251 § ⁂	**DNMT3A probe** 19243-SP0777-L25339		**R882H** (c.2645G>A)		02-025.310683
256	**TP53 probe** 02376-L25000			17p13.1	17-007.519217
264	**IKZF1 probe** 13873-L15917			7p12.2	07-050.411797
269 «	**CUX1 probe** 09472-L23871			7q22.1	07-101.664162
274	Reference probe 09792-L23872	15q21.1			15-042.654415
280	**KMT2A probe** 19132-L25682			11q23.3	11-117.853966
285 «	**NPM1 probe** 18521-L25083			5q35.1	05-170.752385
290	**ATM probe** 08445-L23873			11q22.3	11-107.741279
296	**RUNX1 probe** 17084-L17531			21q22.12	21-035.343030
301 «	**ETV6 probe** 14058-L15656			12p13.2	12-011.694211
308	**EZH2 probe** 18272-L23874			7q36.1	07-148.142933
315	Reference probe 14763-L24999	1q23.3			01-159.404486
320	**AEBP2 probe** 18882-L23816			12p12.3	12-019.506843
328	**CTNNA1 probe** 18523-L23814			5q31.2	05-138.188350
335	Reference probe 11444-L26026	1q41			01-214.662432
342	**ATM probe** 02663-L24998			11q22.3	11-107.648710
349	**ETV6 probe** 18883-L24997			12p13.2	12-011.938395
356	**JARID2 probe** 18800-L27166			6p22.3	06-015.628957
364 «	**NPM1 probe** 18527-L23818			5q35.1	05-170.749614
371	**ATM probe** 08420-L22087			11q22.3	11-107.629847
378	Reference probe 08141-L24356	14q11.2			14-020.862831
384	**MLL5 probe** 18528-L24355			7q22.3	07-104.505044
391	**AEBP2 probe** 18529-L24996			12p12.3	12-019.564543
399	**TP53 probe** 17423-L24353			17p13.1	17-007.517580
404	**TET2 probe** 16297-L24352			4q24	04-106.415798
412	**NF1 probe** 18544-L24995			17q11.2	17-026.611533
418	**MLL5 probe** 18796-L24994			7q22.3	07-104.489851
427	**TET2 probe** 16299-L18591			4q24	04-106.374651
436	**RUNX1 probe** 02838-L02269			21q22.12	21-035.153664
442	**KMT2A probe** 19133-L25082			11q23.3	11-117.848390
452	Reference probe 16286-L24993	13q14.3			13-050.428652
459 ⁂ «	**CTNNA1 probe** 18534-SP0665-L24992			5q31.2	05-138.146832
466	**ASXL1 probe** 18535-L25524			20q11.21	20-030.481425
472	**IKZF1 probe** 14061-L24990			7p12.2	07-050.422502
481 «	**U2AF1 probe** 18537-L23828			21q22.3	21-043.397447
489	Reference probe 17939-L15290	3q25.31			03-157.715622
498	**SUZ12 probe** 18539-L23830			17q11.2	17-027.339557
505	Reference probe 15203-L23403	3p12.2			03-081.774613

§ Mutation-specific probe. This probe will only generate a signal when the specific mutation is present. It has been tested on artificial test DNA **but not on positive human samples!**

Table 2.7 SALSA MLPA probemix P437-A1 Familial MDS-AML (MRC Holland)

Length (nt)	SALSA MLPA probe	Chromosomal position					
		reference	GATA2	TERC	TERT	CEBPA	RUNX1
64-70-76-82	Q-fragments: DNA quantity; only visible with less than 100 ng sample DNA						
88-92-96	D-fragments: Low signal of 88 or 96 nt fragment indicates incomplete denaturation						
100	X-fragment: Specific for the X chromosome						
105	Y-fragment: Specific for the Y chromosome						
124	Reference probe 18709-L21056	5q31.1					
130	**TERC probe** 08640-L08656			exon 1			
136 «	**CEBPA probe** 19049-L24936					exon 1	
142 «	**GATA2 probe** 19050-L24937		exon 6				
149	**TERT probe** 03078-L25347				exon 15		
156	**RUNX1 probe** 19051-L25507						exon 8
160	**TERT probe** 10346-L25509				exon 2		
168 § Ж	**GATA2 probe** 19052-SP0847-L24939	**1192C>T; R398W**					
172	**RUNX1 probe** 02835-L25510						exon 2
180	**TERT probe** 15272-L18513				exon 1		
185	Reference probe 08731-L25567	9q21.13					
190 § Ж	**GATA2 probe** 19053-SP0738-L25512	**1061C>T; T354M**					
195	**TERT probe** 08647-L25704				exon 7		
201	Reference probe 15519-L26387	16q13					
206	**TERT probe** 19055-L25515				exon 6		
213	**RUNX1 probe** 19057-L25344						exon 1
219 «	**GATA2 probe** 19058-L25516		exon 3				
224	**TERT probe** 19059-L25517				exon 9		
229	**TERT probe** 08652-L25343				exon 12		
233	Reference probe 15154-L25342	18q21.32					
239	**RUNX1 probe** 19014-L26113						exon 4b
244 «	**GATA2 probe** 19060-L24947		exon 8				
251	**TERT probe** 19061-L24948				exon 4		
256	Reference probe 10702-L11284	6p12.2					
265 «	**CEBPA probe** 19062-L24949					exon 1	
270	**TERC probe** 19299-L24943			exon 1			
278	**GATA2 probe** 19063-L25518		exon 7a				
283	Reference probe 08834-L08894	2p13.2					
289 Ж «	**GATA2 probe** 19064-SP0740-L25940		exon 4				
296	**TERT probe** 19065-L26112				exon 3		
301 «	**RUNX1 probe** 02840-L25520						exon 9
309 «	**CEBPA probe** 19066-L26430					exon 1	
318 «	**GATA2 probe** 19067-L26432		exon 1				
324 «	**RUNX1 probe** 19017-L25523						exon 10
331	**TERT probe** 08651-L26433				exon 11		
338	Reference probe 16276-L26434	13q14.3					
346 «	**GATA2 probe** 19588-L26183		exon 2				
355	**RUNX1 probe** 19069-L24956						exon 3
364	**TERT probe** 08653-L23271				exon 13		
372	**TERC probe** 19070-L24957			exon 1			
381	Reference probe 01136-L00694	4q35.2					
388	**RUNX1 probe** 19019-L26111						exon 7
395	**TERT probe** 08656-L25442				exon 16		
402	Reference probe 01237-L25675	10p14					
409 «	**GATA2 probe** 19071-L24958		exon 5				
418	**TERT probe** 19072-L24959				exon 10		
425	**RUNX1 probe** 19237-L25317						exon 5
432	Reference probe 15541-L25346	2q22.3					
440	**RUNX1 probe** 02838-L25676						exon 6
447 Ø «	**GATA2 probe** 19298-L25506		intron 6				
454	**TERT probe** 08645-L25504				exon 5		
460	Reference probe 16287-L25505	20q11.23					
468	**TERT probe** 19073-L25677				exon 14		
472	**TERT probe** 19074-L25703				exon 8		
481	Reference probe 09772-L10187	15q21.1					
489	**TERC probe** 19295-L25478			exon 1			
503	Reference probe 06676-L23439	11p15.4					

2.9.2 MLPA Experimental method

Briefly, the MLPA technique consists of 5 steps: 1) DNA denaturation; 2) Probe hybridisation; 3) Enzymatic ligation of probes; 4) PCR amplification and; 5) Fragment separation and sizing by capillary electrophoresis

2.9.2.1 DNA denaturation

50-100ng of genomic DNA was diluted in $TE_{0.1}$ (10 mM Tris-HCl pH 8.2 + 0.1 mM EDTA) to a final volume of 5µl and placed into individual wells within a 96-well plate. A minimum of 3 different reference PB or BM DNA samples from healthy individuals were included in each run. For experiments with over 21 test samples, an additional reference DNA was included per 7 additional samples. Identical reference DNA samples were used in each MLPA experiment to ensure consistency of data interpretation. An assay-specific DNA sample was also included in each run, this included human DNA mixed with plasmid and synthetic DNA to incorporate the specific point mutations in each probemix (providing a positive control for these variants). We also included at least one DNA sample with a known copy gain or loss per experiment, these lesions were identified by previous array CGH or WES. All DNA samples were then denatured in a thermal cycler at 98^{o}C for 5 minutes to generate single stranded DNA.

2.9.2.2 Probe Hybridisation

A hybridisation master mix was prepared with 1.5µl MLPA buffer and 1.5µl MLPA probemix per DNA sample. This was added to each denatured DNA sample and incubated overnight at 60°C for 16 hours to allow hybridisation of probes to the target DNA regions.

2.9.2.3 Ligation

Ligation of hybridisation sequences and target DNA strands was performed by adding a mixture containing 3µl of Ligase Buffer A and B, 1µl Salsa Ligase-65 and 25µl nuclease-free water to each well, whilst samples remained in the thermal cycler at 54^{o}C. Samples were then incubated for a further 15 minutes at 54°C to allow completion of ligation followed by inactivation of the ligase enzyme at 98^{o}C for 5 minutes.

2.9.2.4 PCR reaction

A PCR master mix was prepared using 2µl FAM-fluorescently labelled primers, 0.5µl SALSA-DNA polymerase and 7.5µl nuclease-free water per reaction and this was added to the ligated samples, for PCR amplification with 35 cycles of 95°C (30 seconds), 60°C (30 seconds), 72°C (60 seconds) and final extension of 72°C for 20 minutes and then cooled to 15°C.

2.9.2.5 Fragment separation

Amplified MLPA products were size separated by capillary electrophoresis at the Genomics/Equipment Park at the London Research Institute (now part of the Francis Crick Institute) using an ABI 3130XL Genetic Analyzer instrument (Applied Biosystems, CA) according to the exact electrophoresis specifications detailed by the manufacturer.

2.9.3 MLPA data analysis and interpretation

Raw data obtained from each sample were analysed using Peak Scanner software v1.0 (Applied Biosystems, CA) with the size standard LIZ3130. Detailed quality control (QC) checks were performed in accordance with the provided manufacturer's instructions and only data passing these QC were used for subsequent analysis. Data from the Peak Scanner software were first exported into a customised Excel™ file for data normalisation and to determine the dosage quotient (DQ) of each probe (also referred to as peak ratios). The absolute fluorescence of each probe (peak height) was normalised prior to calculation of the DQ, as described below:

i) Intra-sample normalisation

This step normalised peak heights of target probes to reference probes within the same DNA sample. It was calculated by dividing the peak height of each probe by the total of the reference peaks.

ii) Inter- sample normalisation

The final DQ (or peak ratio) for each probe was determined by dividing the intra-sample normalised peak height by the mean normalised peak height across the reference DNA samples (including the kit specific sample DNA).

Peak ratios or DQs within the range 0.8–1.2 were considered normal, and heterozygous constitutional deletions typically displayed a DQ between 0.4-0.65, according to the manufacturer's protocol (MRC Holland). The DQ of somatic CNAs was dependent on the tumour purity of the sample, for this study values <0.8 were considered to represent heterozygous deletions, while values >1.2 were indicative of heterozygous amplification/duplication. In theory, these thresholds aimed to identify heterozygous CNAs in samples with a tumour purity of > 40% . Each positive result was confirmed in a second independent experiment.

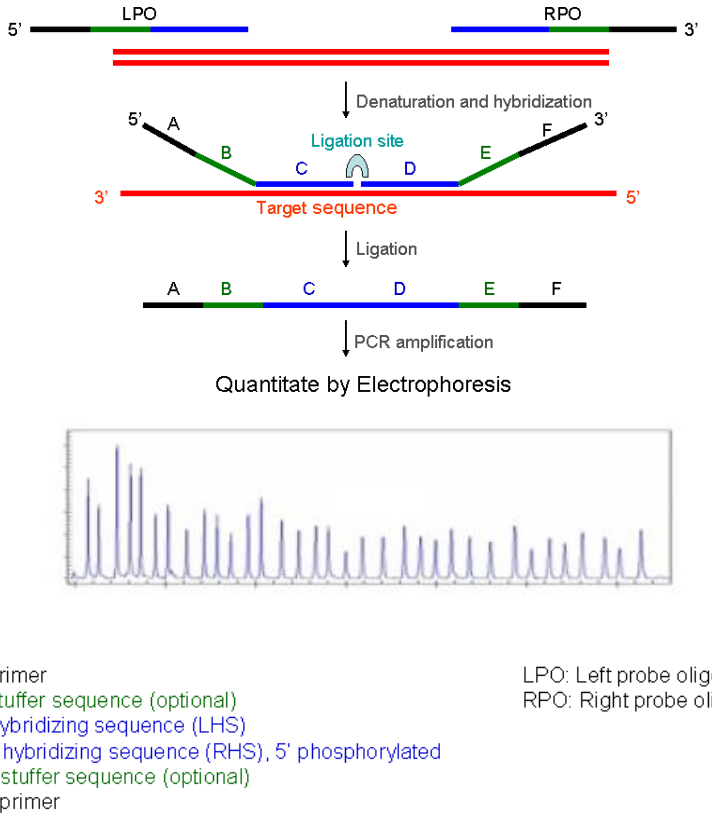


Figure 2.3 Overview of MLPA experimental procedure, figure from Zhi and Hatchwell (2008).

Probes ligate to a target sequence on denatured DNA, following PCR amplification, probe signals are then quantified by capillary electrophoresis. Following comparison with reference peaks, the DQ for each probe was calculated to determine the copy number. Key: LPO, left probe oligonucleotide; RPO, right probe oligonucleotide.

2.10 Whole-exome sequencing

WES was performed in tumour DNA (extracted from PB or BM sampled at disease presentation) and germline DNA (extracted from PB, BM or skin biopsy during remission). Whole exome capture libraries were prepared at The Genome Centre, Barts Cancer Institute or externally by Beckman Coulter Genomics™ (Denver, US) using 1.5-3µg DNA with a 260/280 nanodrop ratio of >1.7. Following sonication of genomic DNA, libraries were prepared by hybridising DNA fragments with biotinylated RNA 'baits' using the Agilent SureSelect XT Human All Exon V5 kit. Hybridised DNA fragments were then amplified by PCR, 'captured' using streptavidin coated magnetic beads. These captured libraries underwent further PCR to amplify the amount of captured DNA and to add sample specific indexes to each library. Capture libraries were then assessed for quality, quantity and size distribution using the DNA High Sensitivity Chip on the Bioanalyzer 2100 (Agilent), prior to pooling or multiplexing in groups of 3. Equimolar concentrations of each pooled product were then loaded into a single lane on an Illumina HiSeq 2000 and sequenced using 100bp paired-end reads to a mean depth of 80-100x per exome.

2.10.1 Bioinformatic analysis of WES data

2.10.1.1 Alignment of WES reads

The primary bioinformatic analysis of WES data was performed by Dr Jun Wang (Centre for Molecular Oncology, BCI) and the methods employed are summarised below. Sequencing reads were initially aligned to the human reference genome (hg19) using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). Picard tools software (version 1.86) was used to generate BAM files and remove PCR duplicates with local local realignment of discordantly mapped or unmapped read pairs around indels and base quality score recalibration performed using the Genome Analysis Toolkit (GATK) version 2.3.9 (DePristo *et al*., 2011). Manual inspection of read alignments were visualised with the Integrative Genomics Viewer (IGV).

2.10.1.2 Somatic variant calling

Somatic variants were detected by comparison of paired tumour and germline exomes. Somatic single nucleotide variant (SNV) calling was performed using the MuTect algorithm (Cibulskis *et al*., 2013), selected for its enhanced sensitivity in detection of subclonal or low allelic fraction variants. Short indels (<10 bp) were identified using the GATK Somatic Indel Detector module, with a minimum coverage of 8 reads at the identified indel sites in both tumour and remission samples. Long indels, were detected using Pindel (Ye *et al*., 2009), with a minimum of 8 support reads in the tumour sample. Only the somatic calls with homozygous reference alleles in the germline sample were considered and annotated.

2.10.1.3 Germline variant calling

Our analysis of germline variants focussed on novel non-synonymous (protein altering) lesions that were absent from polymorphism databases such as HAPMAP, the Thousand Genome Project (TGP) and dbSNP137. For the majority of families, exomes from two affected family members were compared, with unique mutations that were shared between both individuals selected for further prioritisation. In one family (Pedigree P) novel germline variants were detected in the remission sample of one affected member (P.III.2). After filtering strategies to exclude polymorphisms and rank candidate variants, the prioritised mutations were then assessed by Sanger sequencing of his two affected sons (P.IV.1 and P.IV.2), to identify lesions segregating with disease.

In brief, the GATK ReduceReads algorithm was first applied to minimize WES file sizes and improve calling performance for novel germline variants. The GATK Unified Genotyper algorithm was then used to generate raw SNV and indel calls, which were subsequently subjected to filtering with GATK VQSR (Variant Quality Score Recalibration) and ApplyRecalibration methods to identify polymorphisms archived in HAPMAP, TGP and dbSNP137 datasets. After filtering of known polymorphisms, shared unique non-synonymous variants were identified between two affected family members using the GATK SelectVariants module, with a minimum variant confidence score of 10.

2.10.1.4 Annotation of variants

All somatic SNVs, indels and germline variants were annotated using the SNPnexus tool (Dayem Ullah *et al*., 2012). Here, genomic coordinates and variants were used as input and this yielded functional annotation including the location of the variants within the genes (for example, 5'UTR, 3'UTR or coding sequence) and the expected result of the alteration (for example, amino-acid change, frameshift and stop codon gain). The Ensembl database (v63) was used for gene/transcript identification and detection of amino acid changes.

2.10.1.5 Calculation of variant allele frequency (VAF)

NGS methodologies such as WES provide a quantitative analysis of sequencing variants, calculated by their variant allele frequencies (VAFs). For a specific variant, the VAF is calculated as shown below:

VAF = no. of variant allele reads/total (variant + reference allele) reads

The figure may be provided as a decimal value or as a percentage with 1.0 or 100% representing a homozygous variant in all cells. For germline heterozygous mutations, a VAF of 50% or 0.5 is expected (i.e. the mutation occurs in all cells on one of the two alleles). This rationale provided a useful strategy for prioritising germline variants with allelic frequencies ranging from 35-65% and excluding those with VAFs <10%.

By plotting VAFs within tumour samples, VAF clusters emerge (groups of mutations with similar VAFs determined by statistical approaches) and these help to define tumour purity, clonal architecture and provide an insight into the clonal dynamics underlying disease evolution. These concepts are implemented in Chapter 3 to decipher the pattern of disease progression in three sequential leukaemic episodes, and in Chapter 4, where VAF clusters within regions of LOH enabled us to calculate the cancer cell fraction for each event, elucidating the genetic timeline of events within the tumour.

2.10.1.6 Validation of exome data

To validate both acquired and germline variant datasets, variant-spanning primers were designed using Primer3plus™ for PCR amplification and Sanger sequencing. A random selection of variants was initially chosen to ensure data validity, followed by sequencing of significant clonal mutations. As Sanger sequencing may reliably detect variants with allelic frequencies =/>15%, verification focussed on variants with VAFs =/> 15%. Deep sequencing was later performed for a selection of clonal and subclonal variants across the three consecutive tumours of B.I.1 and verified the presence of subclonal mutations that were undetectable by Sanger sequencing, this technique is described in detail in section 2.12.

2.10.1.7 Copy number and LOH analysis

In addition to sequencing variants, WES reads with a minimum coverage of 20 and mapping and base quality scores ≥ 20 were also used to generate reliable copy number estimates using VarScan2 (Koboldt *et al*., 2012). Raw copy number calls were adjusted for GC content and re-centered to 0 based on the modal LogR value determined by kernel density estimates, using the VarScan2 copyCaller module. The DNAcopy R Bioconductor package (Venkatramen *et al*., 2007) identified joint segments of LogR values using the circular binary segmentation (CBS), allowing local segments to be merged where LogR values were within 3 standard deviations (s.d.). Absolute copy numbers, tumour purity and ploidy estimates were then generated from DNA copy segments and SNVs from each tumour-normal pair using absCN-seq R (Bao *et al*., 2014).

2.11 Filtering and prioritisation of WES germline variants

2.11.1 The Exome Aggregation Consortium Browser

The exome aggregation consortium (ExAC) (http://exac.broadinstitute.org) is a coalition of investigators seeking to unify WES data from a wide variety of large-scale sequencing projects, in order for data to become available to the wider scientific community. The multiple consortia contributing to this project are listed in Appendix 1 and the dataset includes over 60,000 unrelated individuals who were sequenced as part of various disease-specific and population genetic studies. Individuals with severe paediatric disease were excluded from this data to serve as a useful reference set of allele frequencies for severe disease studies. Furthermore, all raw data were reprocessed through the same pipeline, to increase data consistency across projects.

The ExAC browser is publically accessible and by inputting the gene of interest, all variants within this locus are listed by their hg19 genomic coordinates and protein consequences. The total number of alleles across the dataset are shown, in addition to the allelic frequency corresponding to each variant and the number of homozygotes. If any homozygotes are observed, it is highly probable that the variant will represent a non-disease causing allele; similarly if the allelic frequency exceeds 0.01 the probability it represents a highly penetrant disease related allele is considerably reduced.

ExAC provides a useful guide for filtering variants, however the information in this (and other SNP databases) must be considered in detail before excluding variants of interest. As described in Chapter 1, we are now increasingly aware of a relatively high frequency of mutations, particularly in *DNMT3A* and *TET2*, which may occur in healthy, older individuals, potentially leading to their inaccurate designation as SNPs. For example, the recurrent p.R882H *DNMT3A* mutation is listed as a polymorphism (rs147001633) and occurs in 66 alleles in ExAC with an allelic frequency of 0.0005. This scenario is reminiscent of asymptomatic carriers harbouring germline mutations and may potentially lead to the false exclusion of candidate genes without due vigilance.

2.11.2 Mutation prediction algorithms

The vast number of variants generated from NGS data often pose significant difficulties in distinguishing between pathogenic and bystander mutations. Although functional assays are required for a definitive answer, an increasing number of computational algorithms can help to discern somatic mutations with deleterious effects. The majority of algorithms evaluate the level of evolutionary conservation, combined with the predicted disruption to protein structure and functional consequences, to differentiate pathogenic mutations from benign variants.

In a comprehensive comparison of 15 mutation prediction algorithms, Martelotto *et al* revealed highly variable accuracy in the classification of 989 mutations selected from 15 cancer-associated genes (Martelotto *et al*., 2014). MutationTaster (Schwarz *et al*., 2010) achieved the highest composite score for accuracy (84.5%), sensitivity (84.0%), specificity (85.3%), positive and negative predictive values (90.8% and 75.6% respectively). More recently the algorithm has been updated to include all publicly available SNPs and indels from the TGP (The Thousand Genomes Project, http://www.1000genomes.org/) as well as known disease variants from ClinVar (Landrum *et al*., 2014) and HGMD Public (Stenson *et al*., 2014), which have improved the accuracy, sensitivity and negative predictive values to approaching 90% (Schwarz *et al*., 2014). This software was therefore used to assess the potential pathogenicity of novel germline variants identified by WES within this study.

2.12 Targeted deep sequencing of sequential *CEBPA*-mutated tumour samples

In addition to WES, targeted deep sequencing was also performed using a tagged amplicon approach to detect known mutations at minor subclonal frequencies (VAF < 5%). This analysis primarily aimed to decipher the pattern of disease evolution across three sequential tumours (T1-T3) in patient B.I.1. A combination of 28 WES variants, were sequenced across each tumour sample, these are listed in Table 2.8. Two control DNA samples from healthy volunteers were also tested to exclude contamination and sequencing artefacts.

Targeted deep sequencing primers were designed according to the Fluidigm Access Array™ System guidelines for Illumina Sequencing Systems, G1. The mean amplicon length was 197bp (range 157-241bp) with the majority of amplicons within 20% of the mean length to ensure uniform sequencing coverage. Universal primer sequences CS1: 5'ACACTGACGACATGGTTCTACA and CS2: 5'TACGGTAGCAGAGACTTGGTCT, were tagged to the 5' end of all variant spanning primers, for use in conjunction with the Access Array Barcode Library for Illumina Sequencers (Fluidigm). All primers were manufactured by Sigma-Aldrich (UK) with polyacrylamide gel electrophoresis (PAGE) purification and were diluted to 100uM stock concentration with nuclease-free water. All primer pairs were initially optimised by singleplex PCR and Sanger sequencing using control DNA samples. Once validated in singleplex, combinations of 3-6 primer pairs were pooled for multiplex PCR reactions to achieve a final primer concentration of 1µM. For example, in a 5-plex primer pool (5 primer pairs), 1µl of each 100µM stock primer was diluted with 90µl nuclease-free water, to give a final volume of 100µl and a final primer concentration of 1µM. A 10µl multiplex PCR reaction mixture was set up using 5µl HotStar Taq Plus DNA polymerase master mix (Qiagen), 1µl pooled primers (1µM pool, final concentration 0.1µM), 2µl nuclease-free water, 2µl control DNA template (100ng) and Touchdown (TD) cycling parameters, Table 2.9.

Table 2.8 WES variants tested with ultra-deep sequencing across 3 sequential tumours in patient B.I.1

Tumour	Chr	Genome no	Ref	Alt	Gene	Transcript	cDNA	CDS	Codon	AA	WES VAF	Deep sequencing primer-F	Deep sequencing primer-R	Amplicon size, bp
T1	chr7	148523708	C	T	*EZH2*	ENST00000320356	867	745	249	E>K	0.1485	ACACTGACGACATGGTTCTACACATGCACAATATTTAGTTGGCTCT	TACGGTAGCAGAGACTTGGTCTTAGCATTTGGTCCATCTATGTTG	241
T1	chr5	148742349	G	A	*PCYOX1L*	ENST00000274569	300	238	80	G>R	0.2	ACACTGACGACATGGTTCTACAGCAGATCGACGTGTACGAGA	TACGGTAGCAGAGACTTGGTCTAAGTCTGAGAAGGGCCCAGT	211
T1	chr2	184025818	A	G	*NUP35*	ENST00000295119	1043	940	314	S>G	0.0857	ACACTGACGACATGGTTCTACATTGCCTATGACACTGAAGTAACAAT	TACGGTAGCAGAGACTTGGTCTCTAACCGAAGGAACCAGCAG	208
T1	chr18	47109974	C	T	*LIPG*	ENST00000261292	1458	1206	402	T>T	0.1667	ACACTGACGACATGGTTCTACAGGGTTGTTACTGCCACTGTGT	TACGGTAGCAGAGACTTGGTCTCCCAGGTGAGCTGGATCTT	180
T1	chr10	31809520	G	A	*ZEB1*	ENST00000361642	1323	1260	420	A>A	0.0714	ACACTGACGACATGGTTCTACAAGGCAACCAGTTCTCCTCAG	TACGGTAGCAGAGACTTGGTCTATGGCCACCTTGTTGTATGG	218
T1	chr20	44167975	G	A	*WFDC6*	ENST00000372665	72	72	24	H>H	0.058	ACACTGACGACATGGTTCTACAATTCCAGGCAGAGGTGGAG	TACGGTAGCAGAGACTTGGTCTCTCAGGACTTCTGCCAATCC	195
T1	chr3	64142965	G	A	*PRICKLE2*	ENST00000295902	1059	473	158	P>L	0.1111	ACACTGACGACATGGTTCTACACTTCAGGCACTCAGCATGG	TACGGTAGCAGAGACTTGGTCTTTCTGTTTGCCTCACAGTGC	197
T1,T3	chr5	148730509	G	A	*GRPEL2*	ENST00000329271	452	342	114	V>V	0.3148	ACACTGACGACATGGTTCTACACCGTCTATCTAGGCCTTACCC	TACGGTAGCAGAGACTTGGTCTCTCAGGCTCCGATTCTTCAG	175
T1,T3	chr10	112343991	G	A	*SMC3*	ENST00000361804	1268	1142	381	R>Q	0.2936	ACACTGACGACATGGTTCTACATTTCATCCCATTTGCTTCTATTT	TACGGTAGCAGAGACTTGGTCTGCTTCAGTGTCTTCCAAATCCT	213
T1,T3	chr7	47880145	G	A	*PKD1L1*	ENST00000289672	5517	5466	1822	G>G	0.3506	ACACTGACGACATGGTTCTACAGGGAGAAAGTGCGAGGAAG	TACGGTAGCAGAGACTTGGTCTTCAGGATAAAGGTGTGTCTGGA	194
T1,T3	chr20	61513343	AAG	A	*DIDO1*	ENST00000266070	4291	3965	1322	AP>AL|AK	0.33	ACACTGACGACATGGTTCTACAAGCTTCCTCCACCGCTTC	TACGGTAGCAGAGACTTGGTCTGCTGGACGATCGGATCTAAC	200
T1,T3	chr19	33792479	T	TTC	*CEBPA*	ENST00000498907	992	842	281	N>R	0.271	ACACTGACGACATGGTTCTACAGGGCAAGGCCAAGAAGTC	TACGGTAGCAGAGACTTGGTCTGCGGTCATTGTCACTGGTC	157
T1,T3	chr6	21065380	TC	T	*CDKAL1*	ENST00000274695	1324	1157	386	F>F	0.476	ACACTGACGACATGGTTCTACATTCCTGGAGAAACAGATCAGG	TACGGTAGCAGAGACTTGGTCTTGTGTGAAGGTAGGCAACACA	217
T2	chrX	62885780	C/G	A/T	*ARHGEF9*	ENST00000253401	1843	1042	348	V>F	0.3333	ACACTGACGACATGGTTCTACACTGGATCTACCAGCCCTACG	TACGGTAGCAGAGACTTGGTCTATTGCTGCCAATCTCCTGTC	215
T2	chr12	125441648	C	T	*DHX37*	ENST00000308736	2290	2191	731	V>M	0.3091	ACACTGACGACATGGTTCTACATCTTGGCATTCACCCTGTC	TACGGTAGCAGAGACTTGGTCTGGAGAGAGGGCCCTGAAC	205
T2	chr3	183904005	T	C	*ABCF3*	ENST00000292808	44	10	4	C>R	0.2833	ACACTGACGACATGGTTCTACACTGCGTCACCGCTACACAT	TACGGTAGCAGAGACTTGGTCTGGCCGATTCAGTTCTGCTTA	211
T2	chr3	56044587	G	A	*ERC2*	ENST00000288221	2066	1810	604	R>*	0.2381	ACACTGACGACATGGTTCTACAACACTGATGGGTTCTCACCTCT	TACGGTAGCAGAGACTTGGTCTTTTCTCAGTCAGTTCAGCCTGT	184
T2	chr3	97806611	C	T	*OR5AC2*	ENST00000358642	595	595	199	L>L	0.2738	ACACTGACGACATGGTTCTACATGCGACTAACTTTCTGCAGGT	TACGGTAGCAGAGACTTGGTCTTCAGAATATCAAAGAGCACACGA	189
T2	chr19	33792403	G	GCGC	*CEBPA*	ENST00000498907	1068	918	306	RN>RR	0.357	ACACTGACGACATGGTTCTACAGGGCAAGGCCAAGAAGTC	TACGGTAGCAGAGACTTGGTCTGCGGTCATTGTCACTGGTC	157
T3	chr7	148513776	C	T	*EZH2*	ENST00000320356	1627	1505	502	R>Q	0.5	ACACTGACGACATGGTTCTACACATGCACAATATTTAGTTGGCTCT	TACGGTAGCAGAGACTTGGTCTTAGCATTTGGTCCATCTATGTTG	195
T3	chr4	41621279	C	G	*LIMCH1*	ENST00000313860	811	757	253	R>G	0.4066	ACACTGACGACATGGTTCTACAGGGAGAGGAAGCGACTCTG	TACGGTAGCAGAGACTTGGTCTCGGTATTCCTCTCTCTCTGCTT	203
T3	chr17	72346692	C	G	*KIF19*	ENST00000389916	1504	1366	456	R>G	0.42	ACACTGACGACATGGTTCTACATTCCAGGAGCAGATGGATGT	TACGGTAGCAGAGACTTGGTCTGCACACTGTACAGCCACAGG	182
T3	chr4	106156747	C	T	*TET2*	ENST00000513237	2508	1711	571	R>*	0.3094	ACACTGACGACATGGTTCTACACCTCAAGCATAACCCACCAA	TACGGTAGCAGAGACTTGGTCTGGATTCCGCTTGGTGAAA	199
T3	chr3	128202758	A	G	*GATA2*	ENST00000341105	1294	962	321	L>P	0.25	ACACTGACGACATGGTTCTACACCGGGAGTGTGTCAACTGT	TACGGTAGCAGAGACTTGGTCTAATTAACCGCCAGCTCCTG	200
T3	chr1	36933434	G	A	*CSF3R*	ENST00000331941	1853	1853	618	T>I	0.1429	ACACTGACGACATGGTTCTACAGCCTATGCTGACCGTACACC	TACGGTAGCAGAGACTTGGTCTTACCCTCCAAACAGCCATCT	214
T3	chr1	19209636	C	T	*ALDH4A1*	ENST00000290597	690	660	220	A>A	0.3333	ACACTGACGACATGGTTCTACACCCATGACTAGGCTGAGACC	TACGGTAGCAGAGACTTGGTCTATCTCACCGCTGCCTCCT	205
T3	chr1	11190823	C	T	*MTOR*	ENST00000361445	5453	5376	1792	A>A	0.4531	ACACTGACGACATGGTTCTACAGGCATGACGCAGTTTCTTCT	TACGGTAGCAGAGACTTGGTCTGACACACAGAACAGGCAGGA	207
T3	chr12	53880222	C	T	*MAP3K12*	ENST00000267079	NA	NA	NA	Splice site	0.3684	ACACTGACGACATGGTTCTACAAGGTGCGAGACCTCAAAGAA	TACGGTAGCAGAGACTTGGTCTATCTCCTCTGGGGAAGGATG	171

Table 2.9 Touchdown PCR cycling parameters

Step	Time	Temperature	Comments
Initial heat activation	5 mins	95°C	
Denaturation	30 secs	94°C	PART A consisting of **18 cycles**: (decreasing by 1°C each cycle)
Annealing	45 secs	68°C	
Extension	60 secs	72°C	
Denaturation	30 secs	94°C	PART B consisting of **17 cycles**
Annealing	45 secs	50°C	
Extension	60 secs	72°C	
Final Extension	10 mins	72°C	

Following optimisation, the 28 tagged amplicons were generated in 3-6plex PCR reactions for the three tumour and two control DNA samples, limiting the number of cycles in part B of the TD program to 5-7 (depending on Bioanalyzer quantitation), this aimed to minimise the occurrence of sequencing artefacts. All amplified products were then quantified using the 2100 Bioanalyzer (Agilent) and pooled in equimolar ratios, resulting in a single pooled library for each DNA sample. The libraries were then transferred to the Genome Centre (Charterhouse Square) where they underwent a second round of PCR to attach barcode primers and adaptor sequences for Illumina cluster generation using the FastStart High Fidelity PCR kit (Roche). The average size of each indexed library was determined using the 2200 Tapestation (Agilent) and each was quantified using the Qubit 2.0 Fluorometer with the dsDNA High Sensitivity assay kit (Invitrogen). The final libraries were pooled at equal concentration, and purified using Agencourt AMPure XP Beads (Beckman Coulter) before being deep sequenced on an Illumina MiSeq run, according to the standard protocol for 150bp paired end reads.

Analysis of deep sequencing data was performed by Dr Jun Wang and is summarised below. Raw reads were initially aligned to the reference genome hg19 using Bowtie2 (Langmead *et al*., 2012). Following BAM conversion, indel realignment and base quality recalibration, VarScan2 was used for variant calling, selecting reads with a mapping

quality > 20 and base quality score > 15. A minimum coverage of 100 reads and 10 variant supporting reads were required for the variant sites, with strand filtering applied. A minimum variant allele frequency (VAF) threshold of 0.01% was applied to run VarScan2. This threshold enabled sufficient detection of subclonal and/or minimal residual disease (MRD) mutations within the sequential samples, but also allowed measurement of the background sequencing artefact or 'noise' rates across each amplicon.

To accurately detect low frequency mutations across sequential samples, the VAF of each target variant was compared with the two control DNA samples, followed by the remaining tumour sample/s. The number of reads supporting the reference allele and the number of reads supporting the alternative allele were compared between any two samples using Fisher's exact test. Secondly, an intra-amplicon analysis was performed, whereby the VAF of targeted variants/sites were compared with the highest frequency variants within surrounding base pairs (30bp upstream and downstream of targeted sites). This analysis helped to exclude false positive calls due to background noise and also allowed calculation of the rate of background noise within each amplicon. Targeted variants detected with a probability $<1^{-5}$ using a two-tailed Fisher's exact test in all comparisons were considered to be true positive variants.

Chapter 3

Clinical and Molecular Profiling of Families with Germline *CEBPA* Mutations

3.1 Introduction

Following the advent of NGS technologies, large research initiatives such as those of the Cancer Genome Research Atlas Network (TCGARN) have sought to comprehensively define the molecular basis of sporadic MDS and AML. Their studies have revealed novel functional classes of mutations and provided a valuable insight into the molecular profiles of myeloid malignancy, in addition to detailing the clonal architecture and evolution of tumours. In contrast, similar investigation of familial leukaemia has been significantly limited by the rarity of cases and material available. This has led to a general lack of scientific and clinical awareness regarding the patterns of disease presentation and progression within this unique patient population.

The transcription factor *CEBPA* is located on chromosome 19q13.1 and has an essential role in mediating granulocytic differentiation and cellular growth arrest (Nerlov, 2007). Heterozygous mutations within this gene were first reported in sporadic AML by Pabst *et al.* (2001). Although all 10 cases in the original series harboured single mutations (*CEBPA*sm), subsequent reports revealed that *CEBPA* was mutated in 10-15% of NK-AML and that 50%-60% of these harboured double mutations (*CEBPA*dm), now considered to represent a distinct molecular entity associated with favourable clinical outcomes. *CEBPA*dm tumours typically display N-terminal frameshift mutations (enforcing translation of the shorter p30 isoform) combined with C-terminal in-frame insertions or deletions (disrupting the DNA-binding or Leucine zipper domain). Mutations within the transcription factors, *GATA2* and *WT1* are also enriched within this subtype of AML, further defining its unique molecular profile (Grief *et al.,* 2012; Green *et al.,* 2013; Fasan *et al.,* 2014).

Germline *CEBPA* mutations were first reported in 2004 (Smith *et al.,* 2004), following the presentation of *de novo* AML in two siblings (aged 30 and 18 years) within a period of two weeks. Since then, over ten pedigrees have been reported worldwide and while cases are rare, published figures are likely to represent an underestimate due to shortcomings in the prospective detection and reporting of germline mutations. To perform the current study, we established an international, multi-centre collaboration to analyse multiple families with germline *CEBPA* mutations, collecting both long-term clinical follow-up information and DNA samples for in-depth genetic profiling.

3.2 Aims and Objectives

This study aimed to integrate molecular and clinical profiling to investigate the somatic mutations and patterns of disease progression in patients with familial *CEBPA*-mutated leukaemia. These data could subsequently be compared with sporadic *CEBPA*-mutated AML to determine important similarities and differences between both patient populations. The main objective was for this information to provide a much needed clinical resource for the genetic counselling, investigation and management of these rare families. We also hypothesised that this comprehensive analysis may provide novel insights into the biology of familial leukaemia, establishing a foundation for the future investigation of this unique subset of patients.

3.3 Methods

3.3.1 Collection of familial AML samples

The initial focus of this work was to establish an international collaboration to identify multiple families for detailed clinical and molecular investigation. Research groups who had previously reported families were contacted to participate in the study and 10 pedigrees were identified for analysis (labelled A-J), including 24 affected individuals, designated numerically according to the genealogical tree, Figure 3.1. Disease episodes within each individual were distinguished in numerical order as follows: primary/diagnostic tumours (T1), first relapse (T2), second relapse (T3) etc. While nine pedigrees had been previously reported, Pedigree C represented a novel family referred to our group from Serbia.

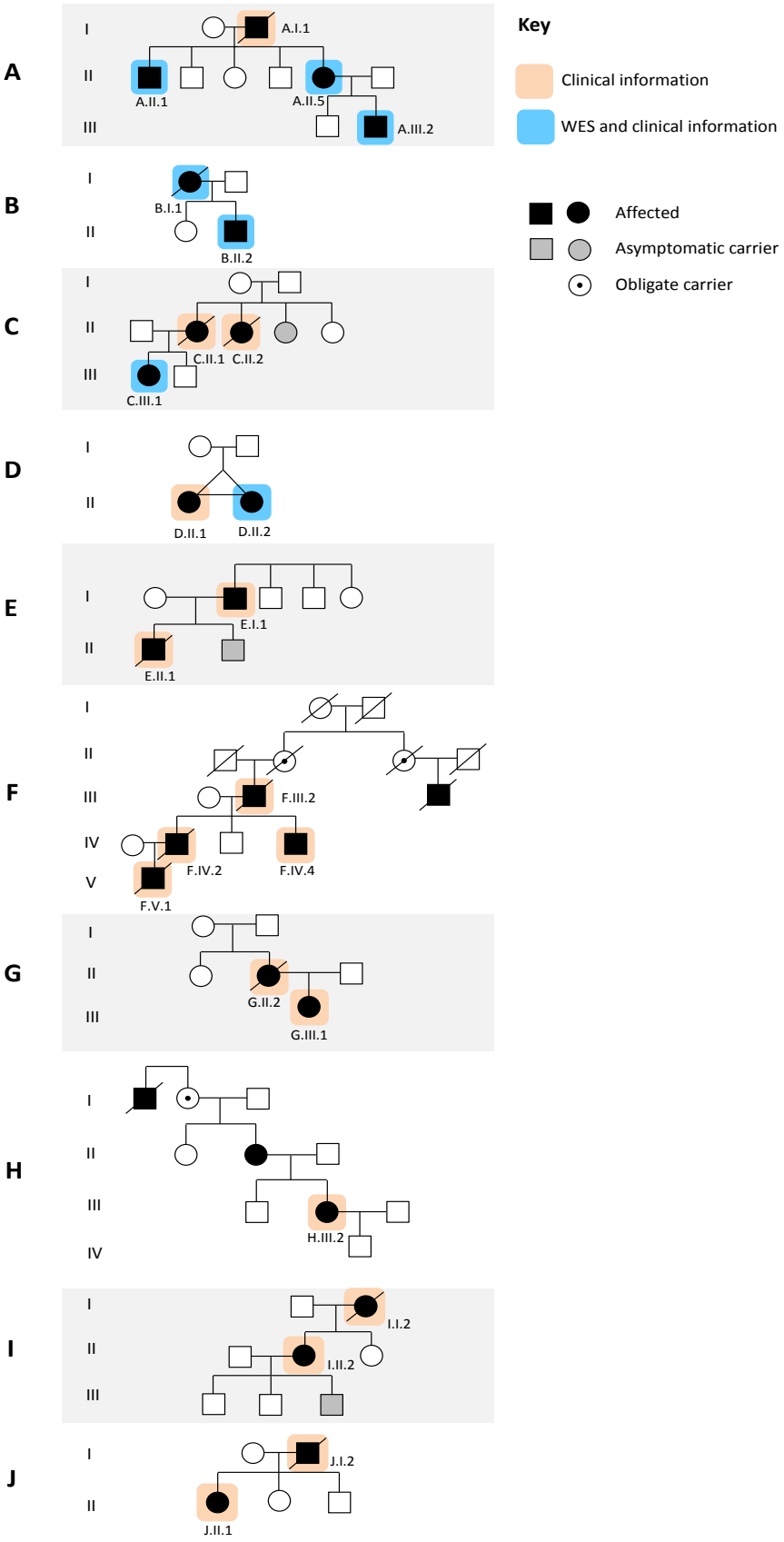

A
I
II
III
A.I.1
A.II.1
A.II.5
A.III.2
B
I
II
B.I.1
B.II.2
C
I
II
III
C.II.1
C.II.2
C.III.1
D
I
II
D.II.1
D.II.2
E
I
II
E.I.1
E.II.1
F
I
II
III
IV
V
F.III.2
F.IV.2
F.IV.4
F.V.1
G
I
II
III
G.II.2
G.III.1
H
I
II
III
IV
H.III.2
I
I
II
III
I.I.2
I.II.2
J
I
II
J.I.2
J.II.1
Key
Clinical information
WES and clinical information
Affected
Asymptomatic carrier
Obligate carrier

Figure 3.1 Genealogical trees of families with germline *CEBPA* mutations.

Ten pedigrees were included for clinical and molecular analysis: family A (Smith *et al.,* 2004), B (Renneville *et al.,* 2009), C (unpublished), D (Debeljak *et al.,* 2013), E (Nanri *et al.,* 2010), F (Sellick *et al.,* 2005), G (Taskesen *et al.,* 2011, Stelljes *et al.,* 2011), H (Taskesen *et al*., 2011), I and J (Pabst *et al.,* 2009). A total of 24 patients were studied with clinical follow-up alone (n=17, highlighted orange) or in combination with WES of tumour DNA (n=7, highlighted blue).

3.3.2 Sanger sequencing of *CEBPA* mutations

DNA extraction from BM, PB, saliva and archival BM smears was performed as described in Chapter 2. Germline *CEBPA* mutations were verified in genomic DNA from affected individuals from pedigrees A-E, using bidirectional sequencing as previously described (Burnett *et al.,* 2010). The sources of germline (constitutional) DNA included peripheral blood (PB) or bone marrow (BM) during remission (A.II.1, A.II.5, A.III.2, B.II.2, C.III.1, D.II.2, E.I.1), saliva (C.III.1, D.II.1) or skin biopsy (B.I.1). Somatic *CEBPA* mutations were detected by testing of tumour samples in these patients and were predominantly in-frame insertions or deletions within the C-terminal. These mutations were annotated by cloning of PCR amplicons, followed by sequencing of individual colonies, in A.II.1, A.II.5, B.II.2 (T2), C.III.1 and F.IV.2 (T3).

3.3.3 Whole-exome sequencing of familial AML samples

Whole-exome profiling of somatic mutations was performed in AML samples from 4 pedigrees with germline *CEBPA* mutations (A-D). Primary (T1) tumours were investigated in 7 patients (A.II.1, A.II.5, A.III.2, B.I.1, B.II.2, C.III.1 and D.II.2) with 2 sequential episodes of disease recurrence in patient B.I.1 (T2 and T3), Figure 3.1. Matched remission DNA was used as a germline reference for primary leukemic episodes, whilst skin DNA was employed as a germline reference for the 3 sequential tumours (T1-T3) in B.I.1. Capture libraries were constructed from 1.5-3µg of DNA using the Agilent SureSelect XT Human All Exon V5 kit. Enriched exome libraries were multiplexed and sequenced on the Illumina 2500 platform to generate 100-bp paired-end reads.

3.3.4 Mutation analysis of WES reads

WES data were analysed by Dr Jun Wang (Centre for Molecular Oncology, Barts Cancer Institute) and a summary of the analytical pipeline is described. Following demultiplexing of reads and alignment to hg19, somatic variant calling was performed by comparing paired tumour and remission samples, with MuTect (Cibulskis *et al*., 2013) for SNV detection, GATK (DePristo *et al.,* 2011) for short indels and Pindel (Ye *et al.,* 2009) for targeted analysis of longer indels (> 10bp), such as *CEBPA* C-terminal insertions. Further genetic profiling of tumours involved the detection of copy number alterations (CNA) and loss of heterozygosity (LOH) from WES reads, with copy-neutral LOH (also known as acquired uniparental disomy [aUPD] or acquired isodisomy [AID]) revealed by intersecting these data.

3.3.5 Exclusion of pre-leukaemic mutations in tumour and remission DNA

Recent studies have revealed the persistence of nascent (pre-leukemic) clones during remission from sporadic AML (Corces-Zimmerman *et al.,* 2014; Xie *et al.,* 2014; Klco *et al.,* 2014). These clones predominantly harbour mutations in epigenetic modifiers, most commonly *DNMT3A* (Shlush *et al.,* 2014; Pløen *et al.,* 2014). To ensure that our analytical pipeline had not led to the false exclusion of these mutations, all variants in candidate pre-leukemic genes (including *DNMT3A*, *TET2*, *ASXL1*, *IDH1*, *IDH2*, *SMC1A*, *SMC3*, *STAG2* and *RAD21*) were identified in tumour and remission samples using VarScan2 (Koboldt *et al.,* 2012) and hg19 as the reference genome. With the exception of common annotated single nucleotide polymorphisms (SNPs), no identical variants were detected in paired samples, at a mean sequencing depth of 91x. These data confirmed that common pre-leukemic mutations were not detected within our WES dataset.

3.3.6 Tumour subpopulation identification and clonality analysis

To analyse the clonal architecture and population dynamics in tumour samples, all somatic SNVs, indels, LOH variants and CNAs were combined to estimate clonal expansions and the cellular frequency of each subpopulation (SP) using EXPANDS (Andor *et al.,* 2014). Somatic variants were assigned to nested populations, with the largest SP forming the dominant clone and any SPs with lower cellular frequencies regarded as subclonal populations.

3.3.7 Integrated whole-exome and deep sequencing analysis of clonal evolution

Somatic *CEBPA* mutations were tested at diagnosis and relapse in 6 patients (B.I.1, B.II.2, E.I.1, E.II.1, F.IV.2 and G.II.2) using Sanger sequencing. Targeted deep sequencing was subsequently performed in tumour pairs from B.I.1 and E.I.1, to determine whether relapse-associated *CEBPA* mutations pre-existed at diagnosis or represented a *de novo* event. To comprehensively characterize disease progression, we performed deep sequencing of 28 selected variants across 3 sequential tumours (T1-T3) in patient B.I.1, enabling us to decipher the pattern of clonal evolution over a disease course of 17 years, a detailed description of the deep sequencing is provided in Chapter 2.

3.3.8 Clinical profiling of an extended familial AML cohort

Clinical demographic and survival data were collected for the 24 affected individuals across all 10 pedigrees (highlighted in Figure 3.1). The median duration of follow up in surviving patients was 7.5 years (range 2-46 years), 4 patients died within a year of diagnosis and one (F.III.2) was lost to follow up. Overall, 96% of patients (n=23) were treated with induction chemotherapy comprising an anthracycline and/or cytarabine regimen. The exception to standard therapy was patient A.I.1, who presented in 1963 and achieved a 2 year remission following prednisolone and mercaptopurine therapy. He experienced two episodes of disease recurrence, the first was treated with prednisolone and methotrexate and the second with cyclophosphamide, resulting in a long term remission of 45 years (Smith *et al.,* 2004).

The majority of familial AML patients (n=11) received chemotherapy alone for consolidation of remission. A third of patients received more intensive therapies such as autologous (n=5), or allogeneic HSCT (sibling donor, n=1; VUD n=1; syngeneic twin, n=1). Following treatment related complications, 2 patients did not receive consolidation therapy post remission (A.I.1 and I.I.2), whilst J.I.2 and C.II.1 died from haemorrhage and chemotherapy-resistant disease, respectively. Consolidation strategies across this cohort are summarised in Table 3.1.

Clinical endpoint definitions followed the revised International Working Group Criteria (Cheson *et al.,* 2003), except that count recovery was not required for complete remission (CR). Survival percentages were calculated using Kaplan-Meier methodology and cumulative incidence of relapse using death as a competing risk. Univariate analyses were performed using the logrank test, comparing outcomes for familial and sporadic *CEBPA*-mutated AML, with the latter categorised into somatic single (n=48, *CEBPA*sm) and double (n=59, *CEBPA*dm) mutated subgroups. This comparative analysis was performed by Dr Robert Hills, Cardiff University Medical School, using data from presumed sporadic *CEBPA*-mutated cases within the UK MRC AML 10 and 12 Trials, as previously described by Green *et al*. (2010). Table 3.1 provides a summary of disease presentation details and consolidation therapy in familial and sporadic single (*CEBPA*sm) and double (*CEBPA*dm) mutated cases.

Table 3.1 Summary of patient characteristics for familial and sporadic *CEBPA*-mutated AML cohorts

	Familial		*CEBPA* sm		*CEBPA* dm	
	No.	%	No.	%	No.	%
No of patients	24		48		59	
Age range, years	1.8-46		16-62		16-67	
Median age*	24.5		46.5		35	
Female	11	46	23	48	30	51
Male	13	54	25	52	29	49
Cytogenetic analysis†						
Favorable karyotype	0		0		0	
Normal karyotype	17	81	26	70	34	72
Other intermediate	2	9.5	7	19	13	28
Adverse	2	9.5	4	11	0	
Unknown	3		11		12	
FAB type						
M0	0	0	2	4	1	2
M1	13	59	15	33	20	34
M2	8	36	16	36	27	47
M4	1	5	7	16	7	12
M5	0		3	7	2	3
M6	0		1	2	0	
M7	0		0		0	
RAEB-t	0		1	2	1	2
Unknown	2		3		1	
Remission rate	22	92	37	77	53	90
Consolidation therapy						
Chemotherapy	11	48	38	79	46	78
Autologous SCT	5	22	5	10	8	14
Allogeneic-VUD SCT	1	4	0		0	
Allogeneic-syngeneic/sib SCT	2	9	5	10	5	8
No consolidation	4	17	0		0	
Unknown	1					

* *CEBPA* sm <45 years, median age 36 years; *CEBPA* dm <45 years, median age 29 years

† Cytogenetic analysis using G-banding.

Abbreviations: SCT- stem cell transplantation, VUD- volunteer unrelated donor, sib- sibling

3.4 Results

3.4.1 Clinical features of familial AML cohort

Familial AML presented early, at a median age of 24.5 years (range 1.75-46 years). Three asymptomatic carriers were identified (C.II.3, E.II.2 and I.III.3) aged 41, 24 and 19 years respectively, they continue to receive annual follow up with normal PB counts and BM morphology (Table 3.2). Figure 3.2 summarises the clinical timeline of events and the consolidation therapy received for each disease episode within the familial AML cohort.

All primary episodes presented as *de novo* AML, predominantly with a normal karyotype, FAB M1 or M2 morphology and aberrant expression of the T-cell antigen, CD7. Germline *CEBPA* mutations were universally located in the N-terminal, each causing a frameshift preceding the internal p30 start codon. Somatic C-terminal *CEBPA* mutations were detected in all diagnostic tumours tested (n=18), with the majority representing in-frame insertions or deletions within the DNA-binding and Leucine zipper domains (Figure 3.3 A)

Table 3.2 Disease presentation profiles in familial AML

Pedigree	Germline mutation		Patient	Gender	Age at diagnosis, years	Diagnostic FAB subtype	CD7 expression	Diagnostic Cytogenetics	Somatic *CEBPA* mutation at diagnosis	
	amino acid*	cDNA mutation**							amino acid	cDNA mutation**
A	P23RfsX137	218delC	I.1	M	10	M1	N/A	NK	N/A	N/A
			II.1	M	30	M2	yes	NK	K302_K313dup	1054_1089dup
			II.5	F	18	M2	yes	NK	Q305_K313dup	1063_1089dup
			III.2	M	2	M1	yes	NK	K313dup	1087_1089dup
B	H24AfsX84	218_219insC	I.1	F	23	M1	yes	NK	N281RfsX38	991_992insGA
			II.2	M	5	M1	yes	NK	R306dup	1067_1068insGCG
C	H24AfsX84	218_219insC	II.1	F	32	M1	N/A	NK	N/A	N/A
			II.2	F	3	M2	N/A	NK	N/A	N/A
			II.3	F	41-AC	_	_	_	_	_
			III.1	F	18	M4	yes	failed	R300_K313dup	1047_1088dup
D	E50AfsX104	297_315del	II.1	F	1.75	M1	N/A	NK	K313dup, K304_A353del	1087_1089dup, c.1061_1210del
			II.2	F	15	M1	N/A	Trisomy 12	K313dup	1087_1089dup
E	I68LfsX41	351_352 ins CTAC	I.1	M	39	M2Eo	yes	NK	R306_N307insALAPPR	1067_1068insGGCCCTCGCCCCCCCGCG
			II.1	M	26	M2Eo	N/A	failed	K313dup	1087_1089dup
			II.2	M	24-AC	_	_	_	_	_
F	H24AfsX84	218_219insC	III.2	M	34	N/A	N/A	N/A	N/A	N/A
			IV.2	M	25	M4	yes	Del6q	E309_Q311 delins LEAK	1075_1081delinsCTGGAGGCCA
			IV.4	M	24	M1	yes	NK	E309dup	1075_1077dup
			V.1	M	4	M1	N/A	NK†	N/A	N/A
G	D63EfsX97	339delC	II.2	F	28	M2	N/A	NK	K313dup	1087_1089dup
			III.1	F	2	M1	N/A	NK	E309-Q312dup	1076_1087dup
H	G53AfsX107	308delG	III.2	F	25	M1	N/A	NK	K326 insT310_X359	1126_1127ins1079_1227
I	A48PfsX112	291delC	I.2	F	46	M2	N/A	Mono 7	N/A	N/A
			II.2	F	40	M1	yes	NK	Q312dup	1085_1087dup
			III.3	M	19-AC	_	_	_	_	_
J	F106LfsX2	464_465insT	I.2	M	42	M1	N/A	Mono 7	A353P; M354L	G1207C; A1210C
			II.1	F	27	M2	yes	NK	K313dup	1087_1089dup

Abbreviations: NK: normal karyotype; N/A: information not available; AC- Asymptomatic carrier; Del-deletion; Mono-monosomy.

Key: * frameshift mutations showing number of amino acids to creation of stop codon
** *CEBPA* transcript: ENST00000498907 (hg19)
† trisomy 8 at 1st relapse and trisomy 21 at 2nd relapse

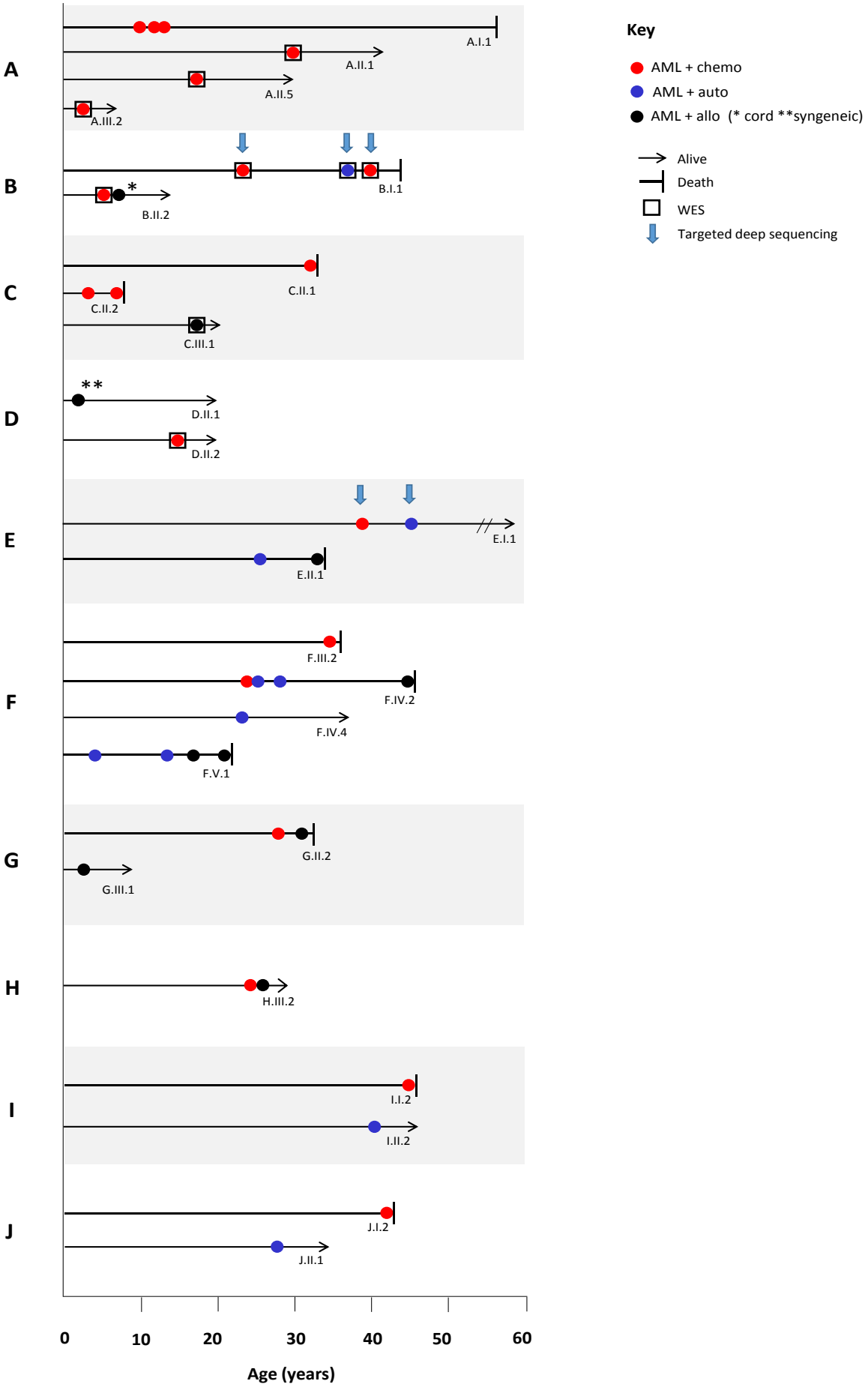

A
A.I.1
A.II.1
A.II.5
A.III.2
B
B.I.1
*
B.II.2
C
C.II.1
C.II.2
C.III.1
D
**
D.II.1
D.II.2
E
E.I.1
E.II.1
F
F.III.2
F.IV.2
F.IV.4
F.V.1
G
G.II.2
G.III.1
H
H.III.2
I
I.I.2
I.II.2
J
J.I.2
J.II.1
0
10
20
30
40
50
60
Age (years)
Key
AML + chemo
AML + auto
AML + allo (* cord **syngeneic)
Alive
Death
WES
Targeted deep sequencing

Figure 3.2 Familial AML pedigrees and timeline of clinical events.

The clinical timeline identifies disease events in each individual, with the main modality of treatment received: conventional chemotherapy (red), autologous transplant (blue) and allogeneic transplant (black). Over 50% of patients experienced recurrent disease, often with prolonged intervening periods of remission (>5 years), as seen in B.I.1, E.I.1, E.II.1, F.IV.2, F.V.1. Blue arrows denote the disease episodes in which deep sequencing was performed to trace patterns of clonal evolution.

3.4.2 Whole-exome profiling of familial AML tumours

WES of 9 familial tumours provided a mean exonic coverage of 89x with 94% of bases covered >10-fold (Table 3.3). In total, 300 somatic mutations were detected across 252 genes, comprising 160 non-synonymous/splice site mutations (protein-altering), 57 synonymous and 83 untranslated region variants (Appendix 2, Tables 1-3). Sanger and deep sequencing approaches were used to verify 107 out of 118 variants, achieving concordance of 83% for indels and 93% for SNVs (Appendix2, Table 4), confirming the accuracy of our analytical pipeline.

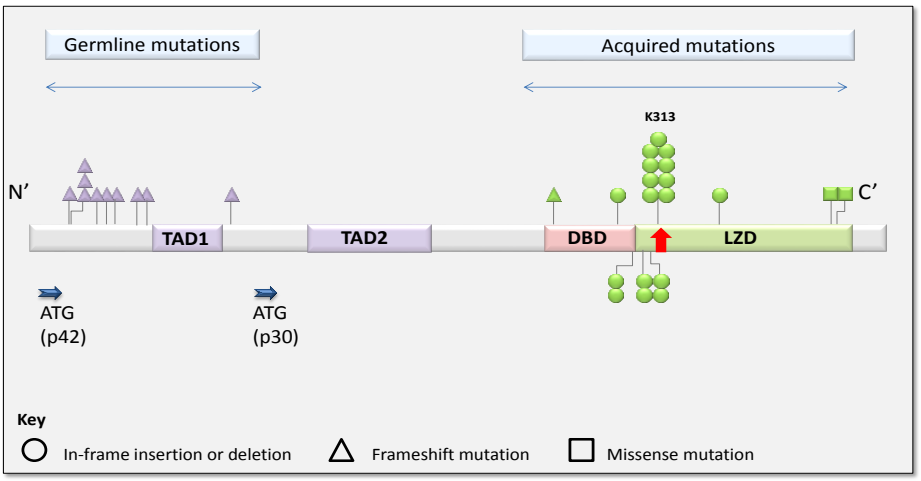
Table 3.3 WES metrics of familial AML and remission samples

Sample	Tissue source	Sequenced nucleotides (Gb)	Mean exome coverage	% bases ≥ 10X coverage
A.II.1-T1	Tumor-diagnosis (BM)	4.29	84.78	98.6
A.II.1-CR	CR1 (PB)	4.92	97.14	98.7
A.II.5-T1	Tumor-diagnosis (BM)	4.12	81.38	92.6
A.II.5-CR	CR1 (PB)	3.23	63.71	94.4
A.III.2-T1	Tumor-diagnosis (BM)	3.84	75.9	98.1
A.III.2-CR	CR1 (BM)	4.68	92.36	98.2
B.I.1-T1	Tumor-diagnosis (PB)	4.69	92.7	96.5
B.I.1-T2	Tumor-diagnosis (BM)	5.31	104.98	97.9
B.I.1-T3	Tumor-diagnosis (BM)	3.09	61.11	90.4
B.I.1-sk	skin biopsy (CR2)	3.6	71.28	94.7
B.II.2-T1	Tumor-diagnosis (BM)	3.78	74.61	95.1
B.II.2-CR	CR1 (BM)	4.93	97.47	95.4
C.III.1-T1	Tumor-diagnosis (BM)	3.93	77.63	64.5
C.III.1-CR	CR1 (BM)	4.12	81.38	92.6
D.II.2-T1	Tumor-diagnosis (BM)	7.11	140.44	99.6
D.II.2-CR	CR1 (BM)	6.48	128.05	99
	Mean	**4.51**	**89.06**	**94.14**

Analysis of familial tumours focused on somatic non-synonymous mutations which were compared with recurrently mutated genes detected in the CGARN series describing WES/WGS of 200 *de novo* AML cases (2013). Specifically, C-terminal *CEBPA* mutations were acquired in all 9 familial tumours, with further mutations observed in *GATA2* (n=5), *WT1* (n=3), *EZH2* (n=2), *SMC3* (n=1), *TET2* (n=1), *NRAS* (n=1), *DDX41* (n=1) and *CSF3R* (n=1). These genes were also found to be mutated in sporadic *CEBPA*-mutated tumours, with the exception of *EZH2* and *DDX41*, Figure 3.3 B.

As described in Chapter 1, *GATA2* mutations are frequent co-operating lesions in sporadic *CEBPA*dm AML, predominantly localising to the zinc finger 1 (ZF1) domain. A similar pattern was observed in familial AML, where clonal (A.II.1 and A.III.2) and subclonal (A.II.5 and B.I.1-T3) *GATA2* mutations universally targeted the ZF1 domain. *WT1* mutations are also a recurrent event in sporadic *CEBPA*dm AML, although their distribution has not been clearly defined in this subgroup of patients. *WT1* mutations occur in 10% of sporadic NK-AML and over 85% of these lesions are found in the C-terminal exons 7 and 9 (Gaidzik *et al*., 2009). In contrast, in familial AML all *WT1* mutations occurred in the transcription-regulatory N-terminal of the gene, specifically exons 1 (C.III.1), 2 (B.I.1-T2) and 6 (D.II.1), suggesting a unique association in this subgroup of patients.

A


Germline mutations
Acquired mutations
K313
N'
C'
TAD1
TAD2
DBD
LZD
ATG
(p42)
ATG
(p30)
Key
In-frame insertion or deletion
Frameshift mutation
Missense mutation

B

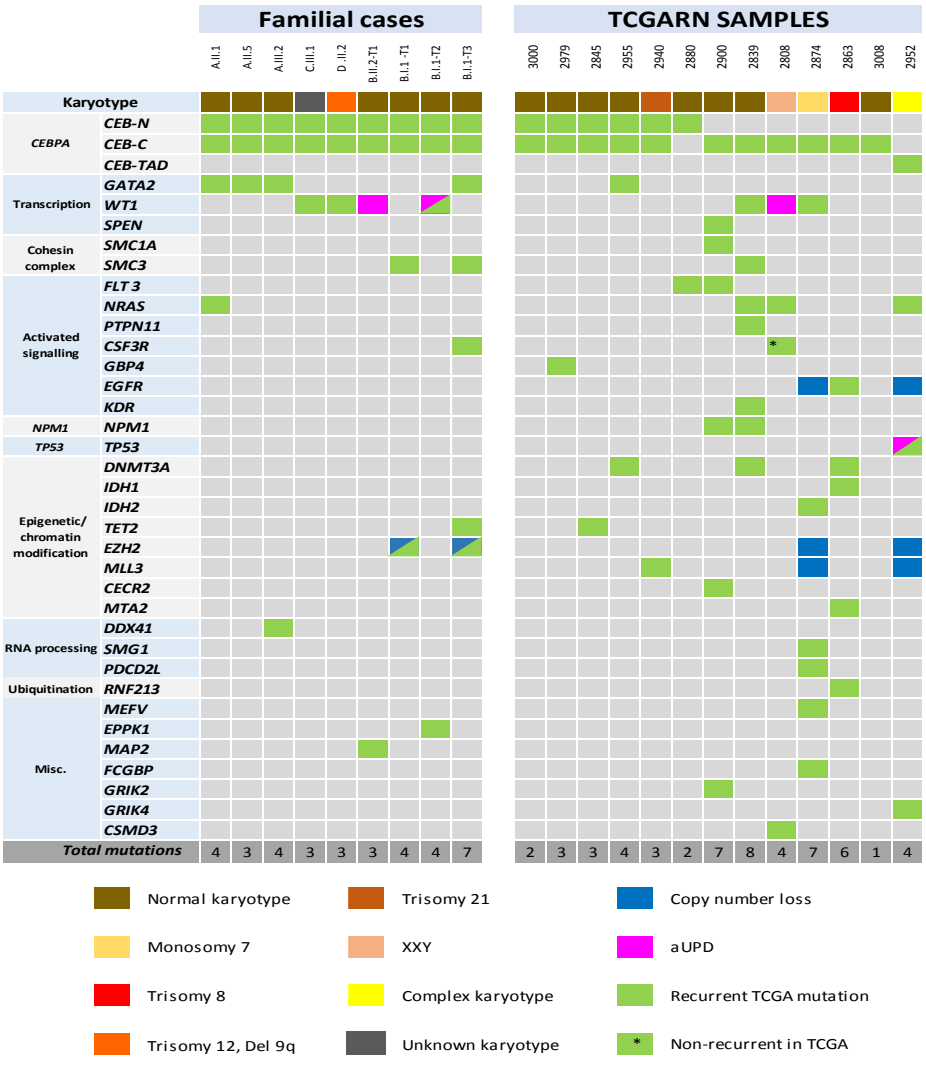

Familial cases
TCGARN SAMPLES
A.II.1
A.II.5
A.III.2
C.III.1
D . II.2
B.II.2-T1
B.I.1 -T1
B.I.1-T2
B.I.1-T3
3000
2979
2845
2955
2940
2880
2900
2839
2808
2874
2863
3008
2952
Karyotype
CEBPA
CEB-N
CEB-C
CEB-TAD
Transcription
GATA2
WT1
SPEN
Cohesin complex
SMC1A
SMC3
Activated signalling
FLT 3
NRAS
PTPN11
CSF3R
GBP4
EGFR
KDR
NPM1
NPM1
TP53
TP53
Epigenetic/ chromatin modification
DNMT3A
IDH1
IDH2
TET2
EZH2
MLL3
CECR2
MTA2
RNA processing
DDX41
SMG1
PDCD2L
Ubiquitination
RNF213
Misc.
MEFV
EPPK1
MAP2
FCGBP
GRIK2
GRIK4
CSMD3
Total mutations
4 3 4 3 3 3 4 4 7
2 3 3 4 3 2 7 8 4 7 6 1 4
Normal karyotype
Trisomy 21
Copy number loss
Monosomy 7
XXY
aUPD
Trisomy 8
Complex karyotype
Recurrent TCGA mutation
Trisomy 12, Del 9q
Unknown karyotype
Non-recurrent in TCGA

Figure 3.3 Molecular profiling of familial AML with germline *CEBPA* mutations

(A) Distribution of germline and acquired CEBPA mutations in familial AML. Transactivation domain (TAD) 1 amino acids (AA) 70-97; p30 start codon (ATG), AA 120; TAD2, AA 126-200; DNA-binding domain (DBD), AA 278-306; Leucine zipper domain (LZD), AA 307-358. All germline mutations were localised to the N-terminal, causing a frameshift preceding the internal *CEBPA*-p30 start codon. Somatic mutations (detailed in Table 3.2) clustered in the C-terminal of the gene, with a hotspot located at residue K313.

(B) Genetic profiling of familial and sporadic *CEBPA*-mutated AML. Molecular profiles are shown for familial and sporadic *CEBPA*-mutated AML samples (the latter analyzed within the TCGA consortium, 2013). In general, sporadic *CEBPA*sm tumours (predominantly with a lone C-terminal mutation) harboured more mutations than primary familial or sporadic *CEBPA*dm AML. *GATA2* and *WT1* mutations were most frequently observed in familial AML and were mutually exclusive. Identical *CSF3R* (p.T618I) mutations were detected in both familial and sporadic *CEBPA*-mutated AML, this recurrent mutation was previously reported in chronic neutrophilic leukemia (Maxson *et al*., 2013; Pardanani *et al*., 2013) Regions of chromosomal loss and aUPD were identified using WES data in familial AML, whilst high-resolution SNP array was used in the TCGA cohort.

3.4.3 WES reveals recurrent intra-familial somatic mutations

In pedigrees A and B, WES was performed on 2 or more affected members, enabling an intra-familial comparison of somatic leukemic variants. Three members of pedigree A acquired one or more *GATA2* mutations, all of which localised to the first zinc finger domain (A.II.1, p.L321F; A.II.5, p.L321F and p.R330Q; A.III.1, p.N317I. This was unlikely to occur by chance, considering 30% of *CEBPA*dm cases harbour *GATA2* mutations and the probability of all 3 individuals acquiring a similar lesion is therefore 0.3^3=0.027 (or 2.7%). In pedigree B, a mother (B.I.1) and son (B.II.2) both acquired uniparental disomy (aUPD) of chromosome 11p, with an underlying *WT1* mutation detected within the mother's leukemic DNA (*WT1* p.D223-S233dup). These findings suggest that family members may acquire somatic lesions in a non-random manner, possibly governed by features within their mutual genotype.

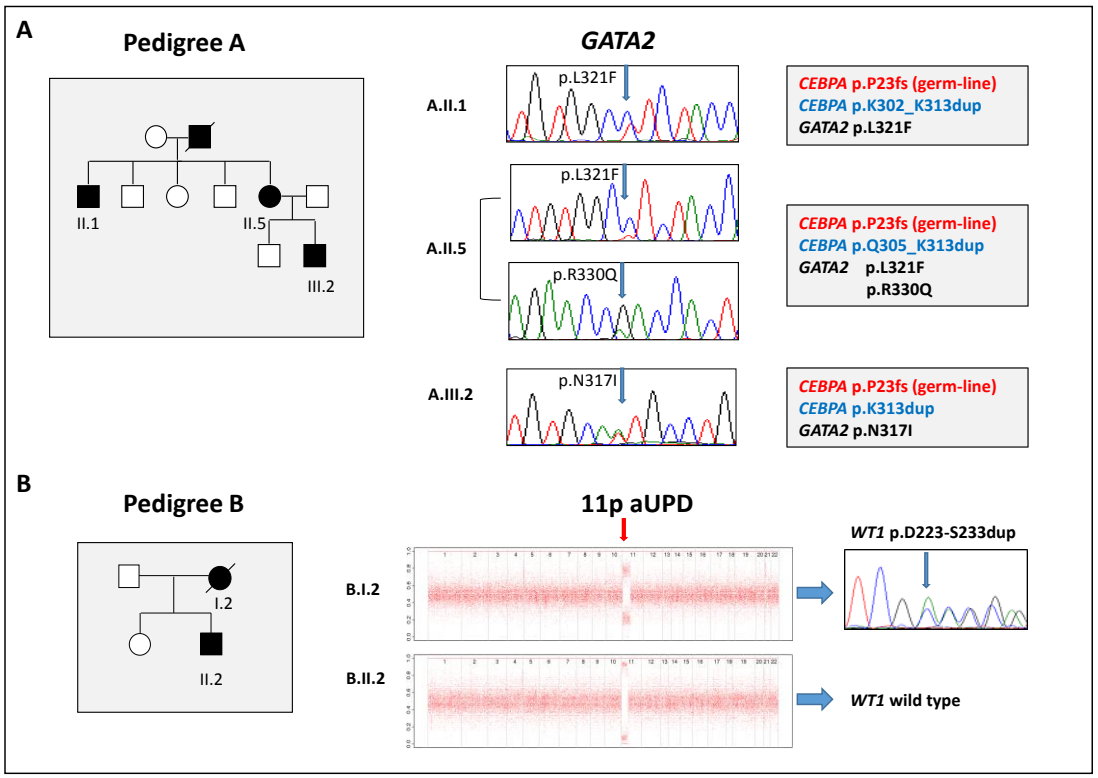


Figure 3.4 Recurrent genetic lesions acquired within family members with germline *CEBPA* mutations.

(A) Somatic *GATA2* mutations were detected in the ZF1 domain within three family members from pedigree A. Notably, individual A.II.5 acquired two mutations, p.L321F (as detected in her brother) and p.R330Q, while her child (A.III.2) acquired a unique mutation p.N317I. (B) Both mother and son in pedigree B developed aUPD of chromosome 11p, with an underlying *WT1* mutation detected in the maternal tumour sample.

3.4.4 Instability of somatic *CEBPA* mutations at disease recurrence

In addition to molecular profiling of familial AML, a further aim of this study was to examine both the clinical and molecular basis of disease progression. Sanger sequencing of *CEBPA* at diagnosis and recurrence revealed discordant somatic mutations in 83% of tumour pairs (n=5), suggesting the presentation of new leukaemic episodes (Table 3.4). Novel mutations were predominantly acquired within the C- terminal (B.I.1, E.I.1 and F.IV.2), although one patient (B.II.2) acquired a somatic TAD2 mutation, while another was left with only the germline *CEBPA* mutation at relapse (E.II.1).

To further investigate the pattern of disease progression in B.II.2, a combination of 10 WES variants from the diagnostic tumour (B.II.2-T1) were tested at relapse (B.II.2-T2). Notably, all 10 variants (7 clonal and 3 subclonal) were absent at recurrence, this again indicated the onset of a new disease episode with a distinct complement of mutations (Appendix 2, Table 3).

The combination of N-terminal (p.H24fsX84) and somatic TAD2 (p.L146fsX24) frameshift *CEBPA* mutations was an unusual finding in the relapse sample of B.II.2 and the allelic distribution of these lesions was subsequently assessed by cloning (Original TA™ kit, Invitrogen). A single PCR fragment encompassing both mutations was ligated to the pCR2.1 vector and transformed into competent *E.coli*. Twenty-four colonies were bi-directionally Sanger sequenced, revealing a predominantly bi-allelic distribution; an isolated somatic TAD2 mutation was detected in 46% of colonies (n=11), while the germline mutation was present in 33% (n=8), Figure 3.5. This distribution predicted expression of the p30 isoform from the germline mutated allele, while the somatic TAD2 mutation putatively led to NMD of the second transcript. A single colony (4%) harboured both mutations (suggesting a mono-allelic distribution resulting in haploinsufficiency) and a further 17% of colonies (4/24) had neither mutation. It is therefore possible that both haploinsufficiency and translation of the dominant negative p30 isoform were both pathogenic mechanisms causing CEBPA dysfunction within this disease episode.

Table 3.4 Discordance of somatic *CEBPA* mutations at diagnosis and relapse

Patient	*CEBPA* somatic mutation-diagnosis		*CEBPA* somatic mutation-relapse		Time to relapse, yrs
	Amino acid	cDNA mutation	Amino acid	cDNA mutation	
B.I.1	N281RfsX38	991_992insGA	K313del	1087_1089del	14
B.II.2	R306dup	1067_1068insGCG	L146PfsX24	586_587insC	1.3
E.I.1	R306ins18bp	1067_1068ins18bp	Q312dup	1084_1086dup	6.7
E.II.1	K313dup	1087_1089dup	Wild type	Wild type	6.6
F.IV.2	E309_Q311delinsLEAK	1075_1081delinsCTGGAGGCCA	N307_T318dup	1070_1105dup	19.9*
G.II.2	K313dup	1087_1089dup	K313dup	1087_1089dup	2.3

Key: * 3rd relapse

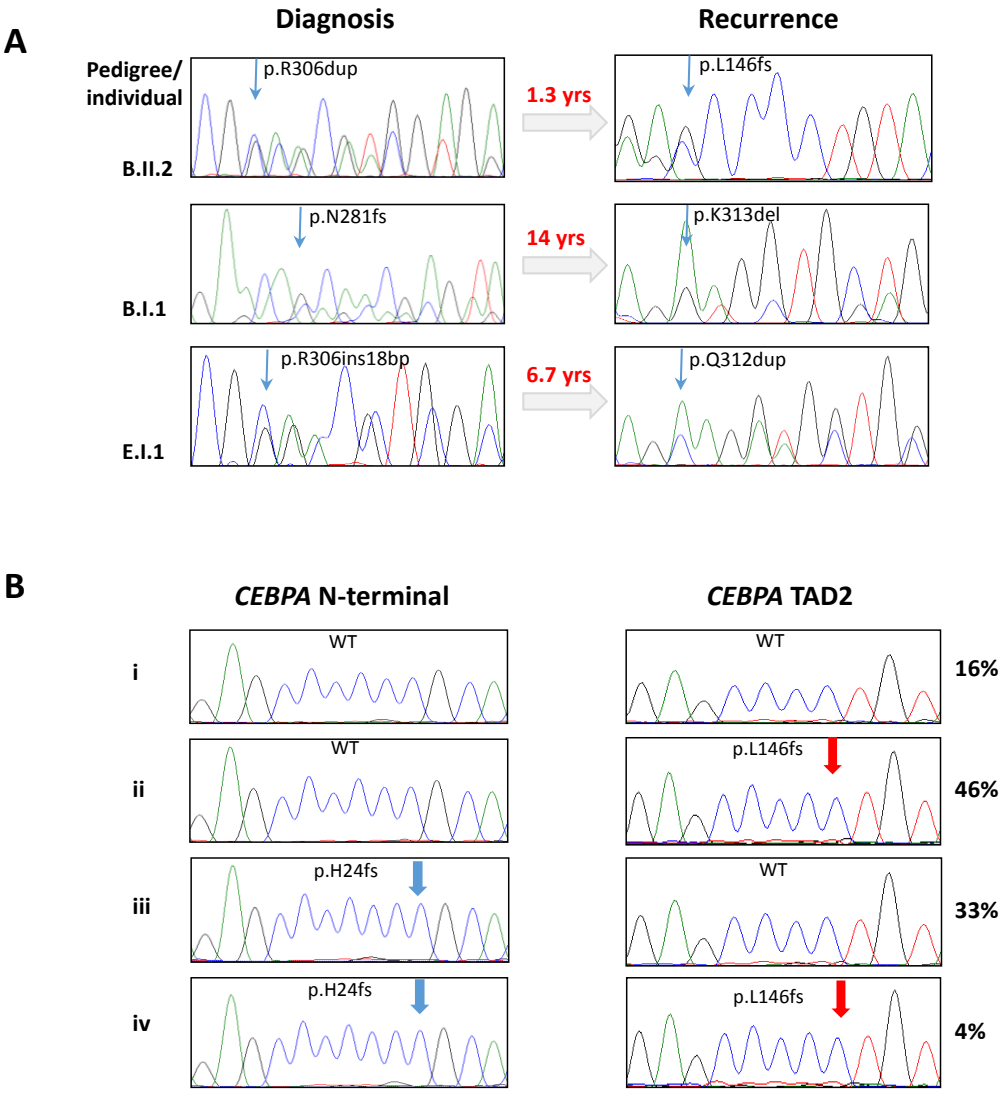


Figure 3.5 Sanger sequencing reveals variation of somatic *CEBPA* mutations at recurrence.

(A) Sanger sequencing of three paired diagnostic and relapse tumours showing instability of somatic *CEBPA* mutations at early and late (>3 yrs) disease recurrence. (B) Cloning and sequencing of transformed colonies from disease recurrence in B.II.2 revealed the germline (blue arrow) and somatic TAD2 mutations (red arrow) had a predominantly biallelic distribution as shown in B(ii) and B(iii)..

3.4.5 Deep sequencing confirms the tumour specificity of somatic *CEBPA* mutations

To prove that disease recurrence represented the onset of new, independent leukaemic episodes, we sought to verify the absence of relapse-specific mutations in pre-existing clones. Targeted deep sequencing of somatic *CEBPA* mutations was performed in paired diagnostic and relapse tumours from B.I.1 and E.I.1. This analysis confirmed the absence of relapse-associated mutations in diagnostic tumours, to sequencing depths exceeding 16,000x (B.I.1) and 143,000x (E.I.1). Furthermore, somatic *CEBPA* mutations from the diagnostic tumour were not detected at relapse with sequencing depths exceeding 11,000x (B.I.1) and 212,000x (E.I.1), suggesting the primary episode had been cured. These data highlighted the tumour specificity of somatic *CEBPA* mutations, supporting the hypothesis that recurrence was initiated by new, distinct leukemic episodes.

3.4.6 Molecular profiling reveals two distinct models of AML recurrence

We next performed integrated whole-exome and deep sequencing analysis of 3 consecutive tumour episodes in patient B.I.1, to comprehensively characterise the pattern of disease progression over a period of 17 years. As shown in figure 3.2 patient B.I.1 first presented at 23 years (T1), achieving a 14 year remission following conventional chemotherapy. She developed AML recurrence at 37 years (T2) and underwent autologous stem cell transplantation (with CD34+ cells harvested during CR1), followed by a third presentation at 40 years (T3).

WES data revealed a high degree of semblance between T1 and T3, with both tumours sharing clonal loss of chromosome 7q (~50Mb deletion, Figure 3.6 A-B) and 12 identical mutations, incuding 10 coding variants (Fig 3.6 C, Appendix 2 Table 3). The region of 7q loss in T1 and T3 encompassed *EZH2* and, significantly, both tumours demonstrated different *EZH2* mutations, suggesting convergent evolution within these disease episodes. T1 harboured the *EZH2* mutation p.E249K (WES VAF 15%), while in T3, the p.R502Q substitution (WES VAF 50%) was detected. The presence of *EZH2* mutations with 7q loss is reminiscent of similar findings reported in sporadic MDS/AML and further illustrates the co-operation between these lesions (Wang *et al.,* 2013).

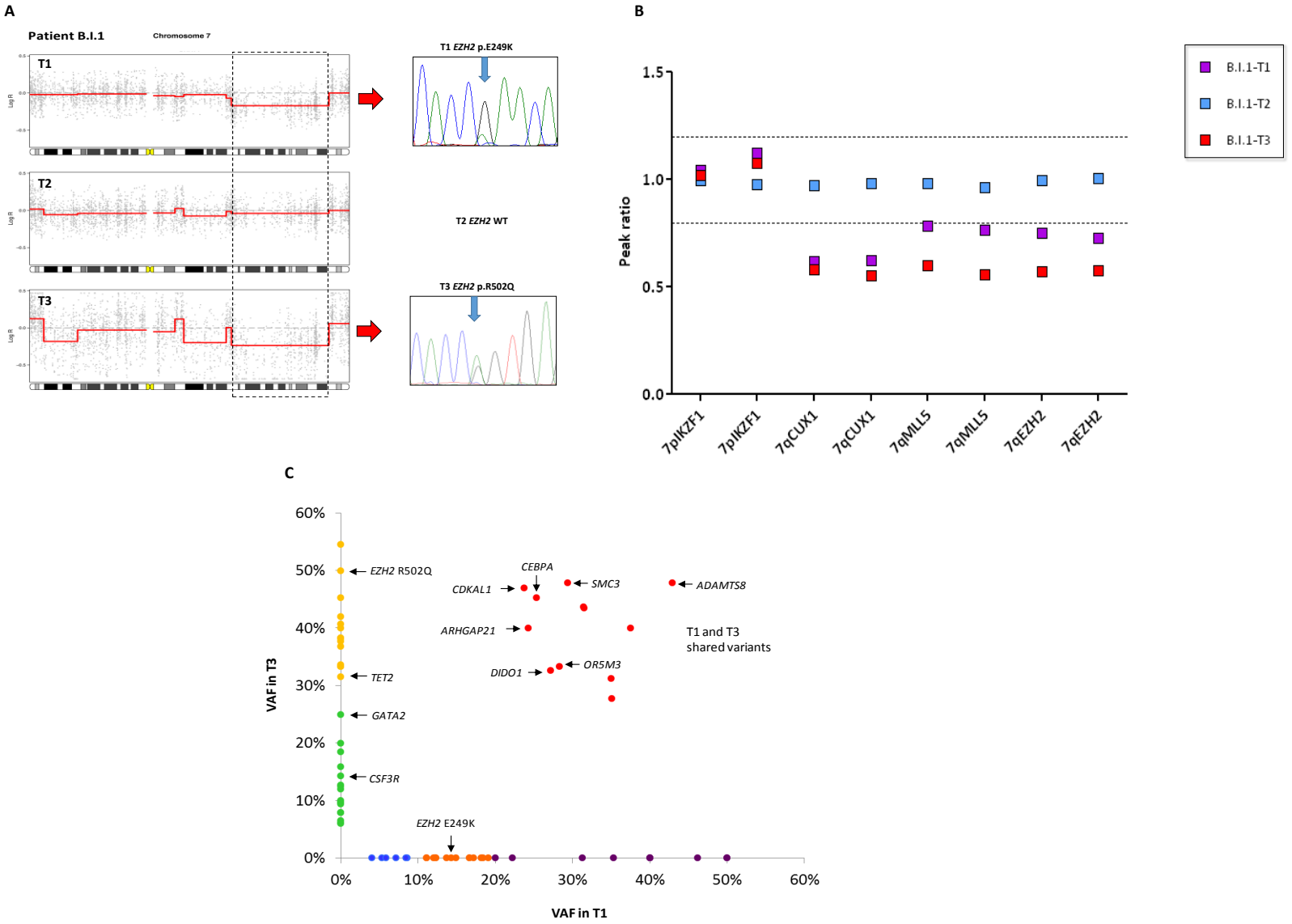

A
Patient B.I.1
Chromosome 7
T1
T2
T3
T1 EZH2 p.E249K
T2 EZH2 WT
T3 EZH2 p.R502Q
B
Peak ratio
1.5
1.0
0.5
0.0
7pIKZF1
7pIKZF1
7qCUX1
7qCUX1
7qMLL5
7qMLL5
7qEZH2
7qEZH2
B.I.1-T1
B.I.1-T2
B.I.1-T3
C
VAF in T3
VAF in T1
60%
50%
40%
30%
20%
10%
0%
EZH2 R502Q
TET2
GATA2
CSF3R
EZH2 E249K
CEBPA
CDKAL1
SMC3
ADAMTS8
ARHGAP21
DIDO1
OR5M3
T1 and T3
shared variants

Figure 3.6 Molecular and copy number profiling of sequential tumour episodes

(A) WES reads demonstrated chromosome 7q loss in B.I.1 for the first (T1) and third (T3) disease episodes, but not the intervening tumour (T2). The deleted segment was ~50Mb in length (segment coordinates: chr7:100,411,278–148,904,343) and encompassed *EZH2*, with different mutations of this gene detected in each tumour. (B) MLPA was used to verify this chromosomal deletion with multiple genetic probes distributed throughout chromosome 7q. Heterozygous loss was demonstrated by reduced peak ratios (0.55-0.78) in T1 and T3, and copy neutral ratios in T2 (0.97-1.00). The dotted lines show the peak ratio range for copy neutrality (copy number =2, peak ratios=0.8-1.2). (C) The scatter plot shows WES somatic variant allele frequency (VAF) measurements for T1 and T3. Variants shown in red were shared by T1 and T3, with further evolution of T3 represented by novel clonal (yellow) and subclonal (green) mutations on the vertical axis. Clonal (purple) and subclonal variants (orange and blue) unique to T1 are represented on the horizontal axis.

T2 was molecularly and structurally distinct from T1 and T3, with no shared somatic mutations or loss of heterozygosity (LOH). This tumour was characterised by 11p aUPD with 16 novel non-synonymous mutations (including *CEBPA* p.K313del and *WT1* p.D223-S233dup), consistent with the emergence of a novel leukemia.

To further characterise the ancestral relationship between T1, T2 and T3, deep sequencing was used to track a selection of 28 variants across each disease episode (Figure 3.7 A). Sequencing depths exceeded 70,000x for all samples, providing a variant detection sensitivity of 0.02% (15-20 supporting reads for 70,000-100,000x depth), with variant allele frequency (VAF) measurements supporting WES data (Pearson correlation coefficient R^2=0.74, Figure 3.7 B). All T1 and T3 somatic variants were absent from T2, confirming that this was an independent leukemic episode (Figure 3.8 A). By tracing variants across T1 and T3, deep sequencing detected the *TET2* p.R571X mutation in an estimated 3% of cells in T1, in contrast, the limited sensitivity of WES (mean sequencing depth 93x) had suggested the mutation was specific to T3, Figure 3.8 B. We concluded that this nascent *TET2* mutated subpopulation appeared to survive chemotherapy and markedly expanded 3 years post autologous transplantation, to re-emerge as the T3 founding clone. T3 was characterised by chemotherapy-resistance, indicating that this tumour was biologically distinct from T1 and T2.

Novel mutations which likely promoted the clonal expansion and genetic heterogeneity of T3, included the clonal *EZH2* mutation (p.R502Q) and sub-clonal *GATA2* (p.L321P) and *CSF3R* (p.T618I) mutations. Collectively, these findings demonstrated a unique model of disease progression in B.I.1, as shown in Figure 3.9. This was characterised by a novel leukaemic episode initiating the first recurrence (T2), followed 3 years later, by expansion of a latent subclone triggering T3. While novel leukaemic episodes have not previously been reported, evolution and expansion of pre-existing subclones is a common mode of relapse in sporadic AML (Ding *et al*., 2012).

A

Gene	Transcript ID	cDNA	Ref	Alt	Codon	aa change	VAF % (TR)	Mutation present in	Total reads T1	Total reads T2	Total reads T3
EZH2(i)	ENST00000320356	867	C	T	249	E>K	17.62%	T1	53745	45576	48225
PCYOX1L	ENST00000274569	300	G	A	80	G>R	11.63%	T1	62749	66169	34751
NUP35	ENST00000295119	1043	A	G	314	S>G	12.91%	T1	20129	31855	29124
LIPG	ENST00000261292	1458	C	T	402	T>T	11.62%	T1	105764	141267	77343
ZEB1	ENST00000361642	1323	G	A	420	A>A	12.50%	T1	306081	190791	253023
WFDC6	ENST00000372665	72	G	A	24	H>H	3.32%	T1	252327	175331	275014
PRICKLE2	ENST00000295902	1059	G	A	158	P>L	9.29%	T1	92249	71724	93423
ARHGEF9	ENST00000253401	1843	C	A	348	V>F	27.18%	T2	45367	72483	62346
DHX37	ENST00000308736	2290	C	T	731	V>M	27.89%	T2	10074	23704	15538
ABCF3	ENST00000292808	44	T	C	4	C>R	22.67%	T2	49144	69963	47059
ERC2	ENST00000288221	2066	G	A	604	R>*	27.25%	T2	99330	146487	107979
OR5AC2	ENST00000358642	595	C	T	199	L>L	27.02%	T2	172046	212518	194697
CEBPA	ENST00000498907	1090	CCTT	C	314	KV>VL	23.97%	T2	16173	11950	3258
EZH2(ii)	ENST00000320356	1627	C	T	502	R>Q	69.85%	T3	15664	19840	14295
LIMCH1	ENST00000313860	811	C	G	253	R>G	46.87%	T3	32301	27806	24039
KIF19	ENST00000389916	1504	C	G	456	R>G	38.38%	T3	70800	53893	31291
GATA2	ENST00000341105	1294	A	G	321	L>P	29.43%	T3	64679	44298	27114
CSF3R	ENST00000331941	1853	G	A	618	T>I	4.62%	T3	29830	50096	243251
MTOR	ENST00000361445	5453	C	T	1792	A>A	37.76%	T3	37658	58473	422610
ALDH4A1	ENST00000290597	690	C	T	220	A>A	43.43%	T3	67555	104533	316427
MAP3K12	ENST00000267079	Splice site	C	T	NA		36.67%	T3	19790	30036	250846
TET2	**ENST00000513237**	**2508**	**C**	**T**	**571**	**R>***	**0.78% (T1); 45.26% (T3)**	**T1 and T3**	**70076**	**55371**	**47950**
GRPEL2	ENST00000329271	452	G	A	114	V>V	31.64% (T1); 45.48% (T3)	T1 and T3	152983	217321	117270
SMC3	ENST00000361804	1268	G	A	381	R>Q	27.48% (T1); 40.11% (T3)	T1 and T3	1139	1134	1259
PKD1L1	ENST00000289672	5517	G	A	1822	G>G	30.97% (T1); 44.12% (T3)	T1 and T3	46726	51404	65665
DIDO1	ENST00000266070	4291	AAG	A	1322	AP>AL|AK	32.36% (T1); 46.65% (T3)	T1 and T3	64898	59009	19950
CDKAL1	ENST00000274695	1324	TC	T	386	F>F	30.93% (T1); 42.41% (T3)	T1 and T3	11233	11610	13779
CEBPA	ENST00000498907	992	T	TTC	281	N>R	27.45% (T1); 36.87% (T3)	T1 and T3	16680	11042	3563
								Average read number	***70,971***	***73,417***	***101,467***

B

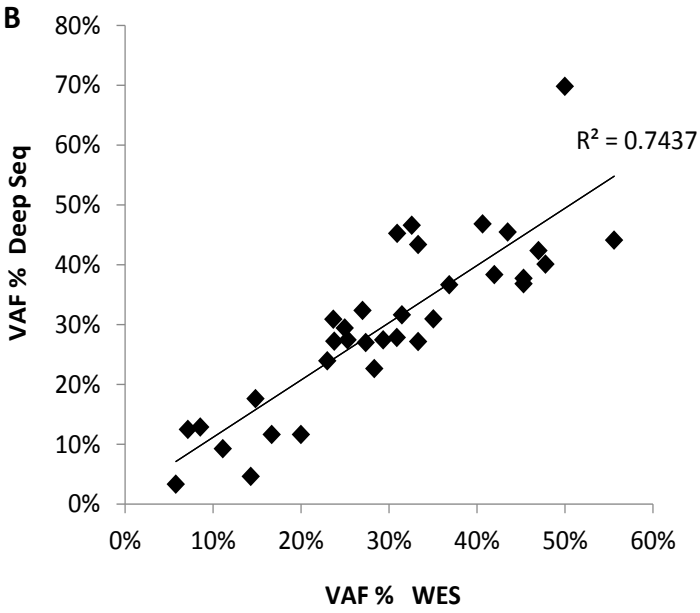


Figure 3.7 Deep sequencing of 28 variants to define pattern of clonal evolution in B.I.1.

(A) Deep sequencing was performed for 28 somatic variants across the 3 consecutive tumours, T1-T3. The variants are grouped according to the tumour samples in which they were initially identified using WES. (B) Comparison of WES and deep sequencing VAFs for all 28 variants, showed concordance of values between both methods, confirming the accuracy and validity of results.

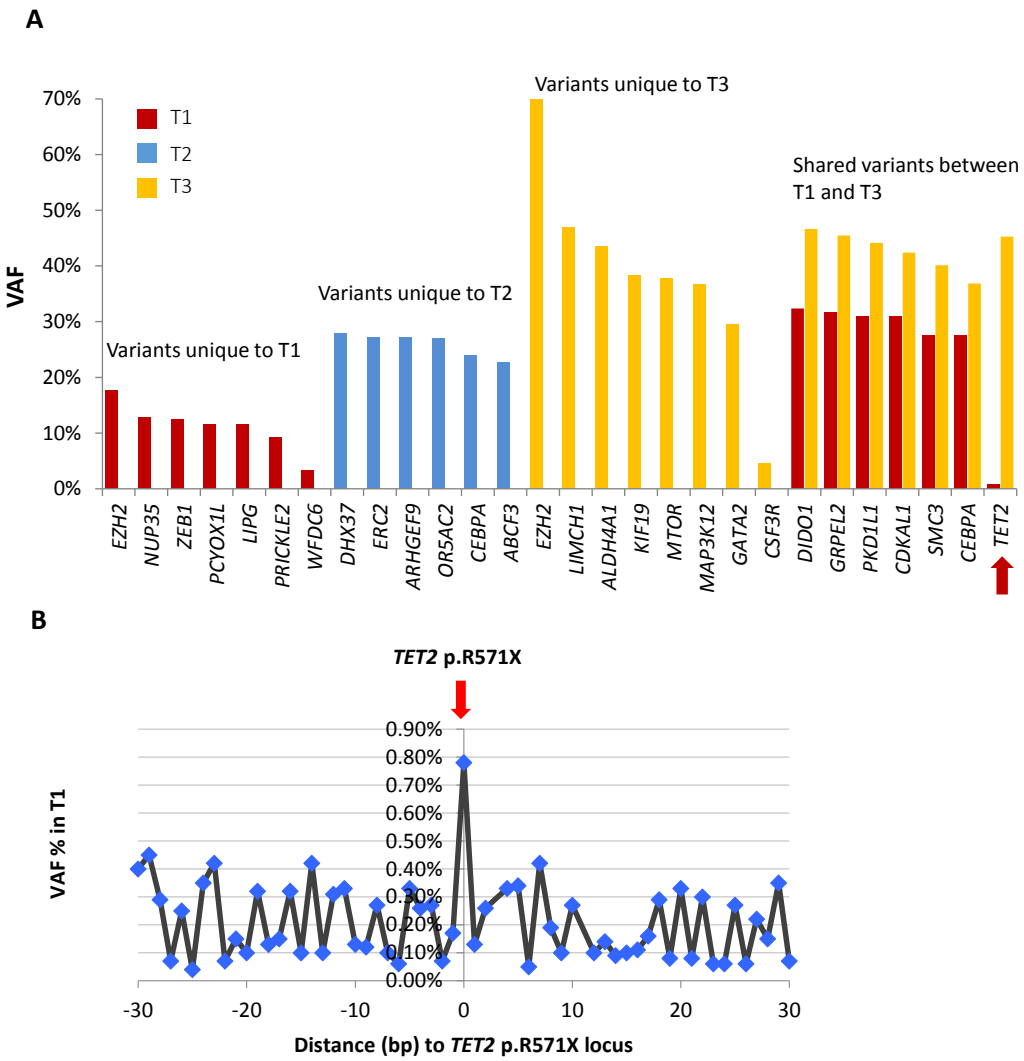


Figure 3.8 Deep sequencing of sequential disease episodes in B.I.1.

(A) Histogram showing VAFS derived by deep sequencing of 28 variants across 3 consecutive tumours (T1-T3). The second tumour (T2) was genetically distinct, with a unique complement of mutations. In contrast, T1 and T3 shared multiple mutations, including *TET2* p.R571X. This represented a minor subclonal variant in T1 but expanded to become a founding clonal variant in T3 (red arrow). (B) Analysis of nucleotide variation surrounding *TET2* p.R571X in T1. The mutation VAF was significantly higher than the surrounding 'background noise' frequency (0.20 ± 0.24%, $p=7.81^{-19}$) indicating that this was a genuine subclonal variant.

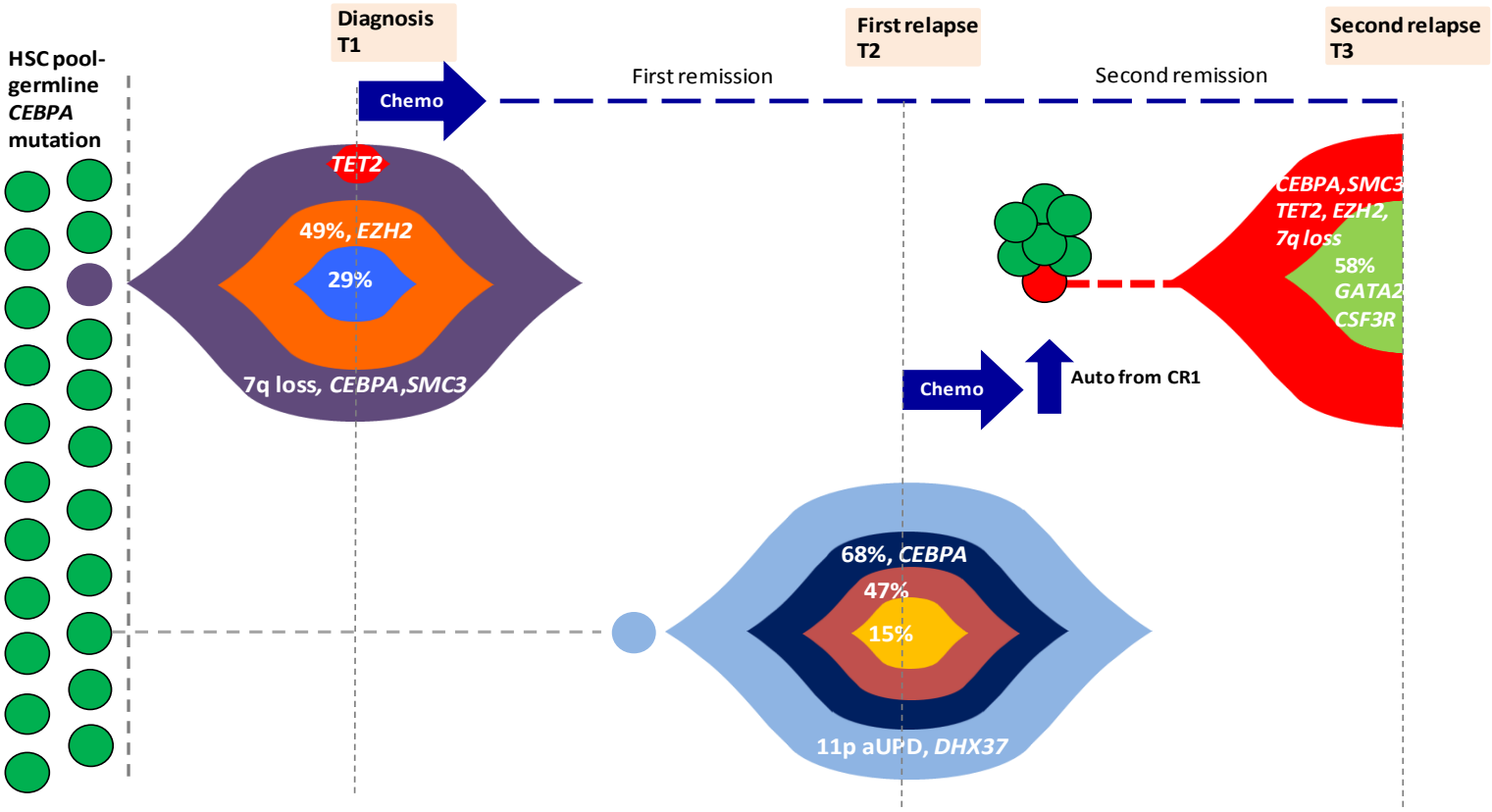


Figure 3.9 Model of disease progression in B.I.1

The T1 founding clone was characterised by 7q loss with 8 non-synonymous mutations including *CEBPA* (p.N281fs) and *SMC3* (p.R381Q). The patient received conventional chemotherapy and stem cells were harvested in CR1 (for future use). T3 presented 3 years after autologous transplantation of these cells, demonstrating expansion of a latent T1-derived subclone, sharing 10 somatic coding variants and the subclonal *TET2* mutation (shown in red). Novel clonal variants in T3 included *EZH2* (p.R502Q) with a new subpopulation harbouring *GATA2* and *CSF3R* mutations. The intervening tumour, T2, was distinct from T1 and T3, characterised by 11p aUPD and 16 novel non-synonymous variants. The percentage of cells in each tumour subpopulation is shown relative to the dominant tumour clone.

3.5 Limitations of Study

The rarity of familial leukaemia creates a difficult foundation upon which to establish research efforts. Collection and investigation of samples relies on the availability of historical samples and development of multiple collaborations with research groups. Whilst extensive efforts were made to achieve both in this study, sample numbers were limited. Consequently, the identification of recurrent molecular events in familial AML was restricted to common variants, such as *GATA2* and *WT1* mutations. A larger cohort would be required to establish less frequent molecular associations in this subgroup. The convergence of *EZH2* mutations in B.I.1-T1 and T3 suggests that this may be an important co-operating lesion. If so, it is likely that this association will also be relevant to sporadic *CEBPA*-mutated AML, as with *GATA2* and *WT1* mutations. Investigation of a larger cohort of sporadic and/or familial *CEBPA*-mutated AML is needed to confirm this hypothesis, as previous studies (Green *et al*., 2013; Grief *et al*., 2013; Fasan *et al*., 2013; Fasan *et al*., 2014) have employed targeted or Sanger sequencing approaches, which have not included assessment of *EZH2*.

The use of remission samples to represent germline molecular profiles may be challenged by the persistence of latent pre-leukaemic mutations, as recently described in AML (Corces-Zimmerman *et al*., 2014; Klco *et al*., 2014; Shlush *et al*., 2014). This is an evolving field and a limited series of genes has, to date, been associated with clonal latency either in a pre-leukaemic phase or during remission. To ensure all efforts were made to identify these mutations, candidate pre-leukaemic genes were independently analysed in WES data for all tumour and remission samples. While extraction of germline DNA from cultured fibroblasts has been recommended (Nickels *et al*., 2013), in practice this would considerably limit the testing of historical samples in our familial cohort and hence remission samples were implemented with the additional testing described above.

The clinical outcomes of sporadic *CEBPA*sm and *CEBPA*dm have been well defined. To compare clinical outcomes of familial AML with sporadic AML, it was not possible to select a precisely age matched cohort, although the median ages of sporadic and familial patients were comparable. As the familial cohort were treated in multiple centres from

1963-2012 treatment approaches differed, however, with the exception of one patient, all received standard regimens used in current practice. The percentage of patients receiving intensive consolidation with autologous or allogeneic HSCT in CR1 was comparable across sporadic and familial cohorts, although autologous HSCT appeared more frequent in familial patients (due to the limited cohort size the percentage variation numerically represented 2 patients). Schlenk *et al*. (2013) recently demonstrated no difference in OS for sporadic *CEBPA*dm patients receiving autologous or allogeneic HSCT in CR1 compared with chemotherapy. As the clinical demographics of familial and sporadic cohorts were otherwise reasonably matched, the interpretation of OS in familial AML is unlikely to have been biased by this minor variation in consolidation therapy.

3.6 Discussion

This multi-centre study documents our collective experience of managing this rare subgroup of patients, providing extended clinical follow up and molecular profiling of familial AML at diagnosis and relapse. Germline *CEBPA* mutations are highly penetrant (only 3 asymptomatic carriers were identified) and, typical of inherited cancers, individuals presented at an early age. In contrast to other leukaemia-predisposition syndromes, germline *CEBPA* mutations are associated with the occurrence of *de novo* AML, without a preceding dysplastic or cytopenic phase and the molecular and clinical profiles of familial tumours corresponded to sporadic *CEBPA*dm AML, with favourable long term outcomes.

3.6.1 A new model of disease progression in familial AML

For the first time, we describe the molecular events underlying disease progression and evolution in familial AML. In contrast to multiple reports illustrating the stability of *CEBPA* mutations in sporadic AML (Tiesmeyer *et al*., 2003; Shih *et al*., 2006; Hollink *et al*., 2011), the opposite was observed in familial AML, with the majority of cases demonstrating variation of somatic *CEBPA* mutations at disease recurrence. Deep sequencing confirmed that these mutations were not derived from pre-existing subclones, but had developed *de novo* suggesting the occurrence of novel, independent leukemic episodes, arising from highly penetrant germline mutations.

This unique model of disease recurrence was distinguished from conventional patterns of AML relapse (Ding *et al.,* 2012) by in-depth molecular profiling of three consecutive leukemic episodes within a single patient. We demonstrated that a novel leukemic clone initiated the first episode of disease recurrence (T2), whilst the second (T3) evolved from a subclone of the primary tumour (as observed in sporadic AML). It is noteworthy that although the second and third disease episodes were independent clonal expansions, *WT1* and *TET2* mutations, in T2 and T3 respectively, targeted a common pathway of DNA hydroxymethylation (Rampal *et al.,* 2014). Functional interdependence of these proteins was further demonstrated by Wang *et al* (2015) who reported that WT1 physically binds and recruits TET2 to its target genes for activation of expression.

Leukaemic variants from the diagnostic episode were not detected with deep sequencing at first recurrence, suggesting that leukaemia-initiating cells were re-introduced following autologous transplantation of CR1 CD34+ cells, during CR2. Since the existence of residual primary leukemic clones could not be fully excluded, the expansion of an independent leukemic episode may be explained by novel clones outcompeting pre-existing clones, due to the high penetrance of germline *CEBPA* mutations. Further analysis of leukaemia-predisposition syndromes will elucidate if this phenomenon is unique to germline *CEBPA* mutations, or whether it universally applies to all familial leukaemia subtypes. It seems likely that the variable penetrance of *RUNX1* and *GATA2* mutations (40-80%) may favour conventional models of relapse, with expansion of pre-existing clones.

3.6.2 Convergence of disease evolution in family members

The acquisition of shared genetic lesions amongst family members, suggests the presence of a selective pressure leading to the intra-familial convergence of tumours. Although studies are limited to date, this phenomenon does not appear to be restricted to germline *CEBPA* mutations. An analogous observation was reported within a *GATA2*-mutated pedigree, where two cousins developed MDS/AML, both acquiring identical somatic *ASXL1* mutations (p.G646WfsX12) and monosomy 7 (Bödör *et al.,* 2012). Enrichment of somatic *ASXL1* mutations was later reported in 30% of germline *GATA2*-mutated patients, coinciding with progression to a proliferative MDS phenotype (West *et al.,* 2014). More recently, Yoshimi *et al*. (2014) reported the inter- and intra-familial acquisition of somatic *CDC25C* mutations in families with germline *RUNX1* mutations. These observations suggest it is possible that germline mutations may influence the acquisition or selection of specific co-operating mutations and that the susceptibility to mutation acquisition may be governed by inherited factors, often shared within families.

3.6.3 Conclusion

From a clinical perspective, our data reinforce the need for long term surveillance of families with germline *CEBPA* mutations and, crucially, highlight the need for re-assessment of somatic mutations at relapse. Although independent disease episodes arising from new leukemic clones may retain chemotherapy-sensitivity, allogeneic HSCT remains an important consideration for the prevention of future recurrence. Comprehensive genetic counselling and screening is essential prior to the selection of related stem cell donors, with the possibility of identifying asymptomatic mutation carriers often weighing significant anxiety upon families.

The accurate identification of familial leukaemia requires clinical vigilance and appropriate investigation. Earlier reports suggest that 5-10% of all *CEBPA*-mutated AML cases harbour germline mutations (Pabst *et al.,* 2008; Taskesen *et al.,* 2011). While evaluation of family history is of paramount importance, in practice this information may not be readily available. Germline *CEBPA* mutations are unique amongst leukaemia-predisposition syndromes, presenting with *de novo* AML in the absence of prodromal cytopenias, which significantly increases the likelihood that these mutations may go undetected. We therefore advocate a thorough investigation of family history and germline DNA assessment of patients younger than 50 years, presenting with *CEBPA*dm AML.

By demonstrating a new model of disease progression, this study reveals important differences underlying disease evolution in sporadic and familial AML, providing a foundation for evidence-based genetic counselling and tailored management of this unique patient population. It is hoped that with improved clinical awareness and ongoing research efforts, comprehensive characterisation of familial *CEBPA*-mutated AML will lead to further valuable disease insights with translational significance.

Chapter 4

Convergence of Somatic JAK-STAT Mutations in a Novel *RUNX1*-Mutated Pedigree

4.1 Introduction

RUNX1 is a definitive regulator of haematopoiesis and is subject to recurrent mutations in both MDS and AML. Germline mutations are associated with an autosomal dominant propensity to developing familial platelet disorder (FPD) and myelodysplasia/acute myeloid leukaemia (MDS/AML). Since the initial report of germline *RUNX1* mutations (Song *et al*., 1999), over 30 pedigrees have now been described in the literature and, although rare, the true incidence of this syndrome is unknown, with published pedigrees likely to represent the 'tip of the iceberg'.

The majority of germline mutations cluster in the highly conserved RHD, involved with DNA binding. Significant variation of mutation penetrance and disease phenotypes has been observed both within and between families. Approximately 40% of carriers develop overt malignancy, whilst others present with mild to moderate thrombocytopenia (FPD) or retain normal peripheral blood counts. This suggests that the germline mutation alone is unable to initiate disease, but that it acts synergistically with subsequent molecular events to initiate clonal expansion and proliferation.

MDS and AML are the most common haematological malignancies affecting *RUNX1*-mutated families, although ALL (in particular T-ALL) and biphenotypic leukaemia may also occur, as described in Chapter 1. The landscape of somatic mutations in FPD/AML lacks detailed characterisation, although current reports describe frequent loss of the WT *RUNX1* allele caused by somatic mutation or deletion, alternatively duplication or aUPD of the germline mutant allele may also occur (Preudhomme *et al*., 2009, Antony-Debré *et al*; 2015). Mutation of *CBL*, an E3 ubiquitin ligase, has been reported in a case of FPD; notably, aUPD of chromosome 11q, created a second 'hit' causing homozygosity of this mutation and accelerating transformation to CMML (Shiba *et al*., 2011). Additional reports describe the acquisition of *ASXL1* mutation in a patient with T-ALL (Prebet *et al*., 2013) and Yoshimi *et al* (2014) recently observed the inter- and intra-familial recurrence of *CDC25C* mutations in over 50% of germline *RUNX1*-mutated cases with isolated thrombocytopenia or MDS/AML. Somatic *GATA2* mutations were also detected in 3/13 individuals, 2 of whom presented with AML, suggesting a potent oncogenic signal in combination with mutant *RUNX1*.

In the majority of families, tumour DNA is available from only one or two affected relatives, limiting the opportunity to examine intra-familial tumour evolution. As discussed in Chapter 1, somatic mutation acquisition and clonal evolution appear to be non-random events (as suggested by the limited mutation profile within tumours), although the factors governing mutagenesis remain unclear. In this study, comprehensive genetic profiling of tumours across multiple family members provided a rare insight into leukaemic evolution, helping to elucidate whether inherited genetic factors may govern somatic mutation acquisition. In broader terms, this hypothesis would suggest a hierarchical model of tumour evolution (reviewed by Nguyen *et al*., 2012), whereby clonal selection is pre-determined by inherited features.

4.2 Aims and Objectives

The objectives of this investigation were to perform in-depth genetic profiling of tumour samples from four siblings with an inherited *RUNX1* mutation. We hypothesised that intra-familial tumours may demonstrate convergent evolution, with the acquisition of specific somatic mutations governed by inherited factors. This analysis aimed to provide a foundation for future studies of intra-familial disease evolution, potentially highlighting important molecular associations and trends relevant to both familial and sporadic leukaemogenesis.

4.3 Materials and Methods

4.3.1 DNA extraction and Sanger sequencing

The pedigree investigated in this chapter is shown in Figure 4.1A with a summary of the clinical history in section 4.4.2. To summarise, three siblings presented with secondary AML, all aged 5 years. Sibling K.II.3 was later diagnosed with low-risk MDS at the age of 14 years. Both parents were asymptomatic with normal full blood count profiles.

Genomic DNA was extracted from PB or BM in all four siblings and both parents. In K.II.3 and K.II.4, salivary DNA was also extracted using the Oragene (OG-500) kit, according to the manufacturer's protocol. PCR amplification of *RUNX1* exons 3-8 was performed using genomic DNA from all family members, followed by bidirectional Sanger sequencing. Sequences were aligned to the *RUNX1* reference sequence ENST00000300305 (hg19) using BioEdit software, with all variants confirmed in both forward and reverse directions.

4.3.2 In-depth genetic profiling of tumour samples

4.3.2.1 WES of pedigree K

Whole-exome capture libraries were constructed from 2-3ug of tumour DNA from siblings K.II.1-K.II.4, with DNA from both parents K.I.1 (mother) and K.I.2 (father) employed as a germline reference for all 4 children. Enriched exome libraries were multiplexed and sequenced on the HiSeq 2000 (Illumina) to generate 100bp paired-end reads. Analysis of raw WES data was performed by Dr Jun Wang (Centre for Molecular Oncology, BCI) with reads first aligned to the hg19 reference genome using BWA and GATK employed for base quality recalibration, PCR duplicate removal and indel realignment. Using maternal and paternal exomes as a reference, somatic variants were determined for all four siblings using Mutect (for SNVs) and GATK (for indels).

To determine somatic variants in the relapse sample (T2) of K.II.4, leukaemic BM DNA (post allograft) was compared with salivary DNA (also sampled post allograft) as a germline reference. In view of the potential mixed chimerism in both BM and saliva samples (the latter due to a mixture of buccal epithelial cells and lymphocytes), all

apparent 'somatic' mutations were then genotyped across both parents to ensure that host germline variants were excluded.

It was difficult to reliably exclude donor polymorphisms from genuine tumour-associated somatic variants and we therefore focussed our analysis on tracing known somatic mutations that were shared between T1 and T2. To decipher the pattern of clonal evolution, we first adjusted all VAFs to the germline *RUNX1* mutation in each tumour, this aimed to correct for variations in tumour purity. As there was no evidence of LOH affecting *RUNX1* in either disease episode, all observed VAFs were corrected to the expected *RUNX1* VAF 0.5 or 50%. In the relapse sample (post allograft), it was assumed that all tumour cells harboured the germline *RUNX1* mutation, which therefore represented a cancer cell fraction (CCF) of 1 (or 100%). The following formulae summarise these calculations:

- Corrected VAF = 0.5(Observed VAF/*RUNX1* VAF)
- CCF of variant = % of cells with variant / % tumour purity
- CCF of heterozygous variant = VAF x 2 / % tumour purity
- CCF of homozygous variant = VAF / % tumour purity

4.3.2.2 Functional annotation of mutations

All novel (non-polymorphic) variants were subjected to gene ontology and mutation prediction algorithms to establish the functional relevance of the gene and the predicted impact of mutations on protein function. This analysis played an integral role in detecting variants within common signalling pathways as will be discussed later in this chapter.

4.3.2.3 Detection of loss of heterozygosity in WES data

Regions of LOH were identified by analysis of logR ratios (LRR) which were calculated from depth normalised reads for all variants in tumour and normal samples (to measure the intensity of allelic representation). Copy neutrality was signified by an LRR of 0, whilst negative and positive deviations represented chromosomal losses and gains, respectively. Alternative allele (or B-allele) frequency (BAF) files were also generated for all variants within LOH regions. Deviation of BAFs from the null value of 0.5 (signifying

germline heterozygous variants) suggested LOH within a chromosomal segment, with 0 representing the genotype AA (homozygous wild-type) and 1 representing homozygosity of the alternative allele (BB) in a 'pure' sample. Tumour profiles of the four siblings were then compared with parental DNA samples, to distinguish somatic changes. Regions of copy-neutral LOH (also known as acquired uniparental disomy, aUPD) occur by mitotic recombination, followed by selection for one of the products. This scenario is characterised by LOH with an LRR of 0 and divergence of BAFs towards 1 and 0.

4.3.2.4 Verification of chromosomal copy number with MLPA

MLPA was used to verify somatic copy number alterations (CNAs) with the probemix X060-X2 (MRC Holland), as described in Chapter 2. Relative copy number was calculated after intra-sample normalisation against control fragment peaks and inter-sample normalisation against control DNA samples. Although the blast percentage may be low in MDS, the majority of cases are molecularly indistinguishable from secondary AML, including those with blast counts <5%, suggesting that clonal haematopoiesis occurs even in the early stages of disease (Walter *et al*., 2012; Walter *et al*., 2013). MLPA peak ratio thresholds were therefore consistent across MDS and AML samples with a ratio of 1±0.2 representing a normal copy number of 2; ratios <0.8 and >1.2 indicated heterozygous deletions and amplifications, respectively.

4.4 Results

4.4.1 Identification of the germline *RUNX1* R201X mutation

Direct Sanger sequencing of *RUNX1* exons 3-8 revealed a nonsense mutation located adjacent to the RHD: C1046T, p.R201X (Figure 4.1B). This mutation was detected in tumour DNA from all four siblings (K.II.1-4), with further verification performed by testing salivary DNA from K.II.3 and K.II.4. The mutation was inherited from the children's mother (K.I.1), who at 44 years, remains an asymptomatic carrier with no evidence of peripheral cytopenias and normal tri-lineage haematopoiesis.

4.4.2 Clinical profile of siblings with MDS/AML

All four siblings presented with MDS or sAML between 2002 and 2012, their timelines of disease are summarised in Figure 4.1 C. Of note, the dizygotic twins (K.II.1 and K.II.2), presented in January 2002 within a period of two weeks. Both had hepatosplenomegaly with BM morphology confirming myelomonocytic AML with dysplastic features.

Sibling K.II.4 initially presented with thrombocytopenia at 1 year. She remained well until 5 years of age, when she developed fever, hepatosplenomegaly and monocytosis. BM morphology revealed myelomonocytic AML with dysplastic features, as in her twin siblings. Following intensive chemotherapy, her recovery was complicated by prolonged pancytopenia and she underwent allogeneic cord-blood HSCT in October 2012. After 15 months, she presented with AML relapse (T2) and achieved a second complete remission (CR2) with salvage chemotherapy. This was consolidated with a second allogeneic HSCT from a volunteer unrelated donor (VUD) and she continues in remission over a year post-transplant.

Sibling K.II.3 presented at 14 years, with mild anaemia and thrombocytopenia. The BM morphology was consistent with refractory cytopenia with multi-lineage dysplasia (RCMD) and she continues under 6-monthly clinical surveillance with no evidence of disease progression.

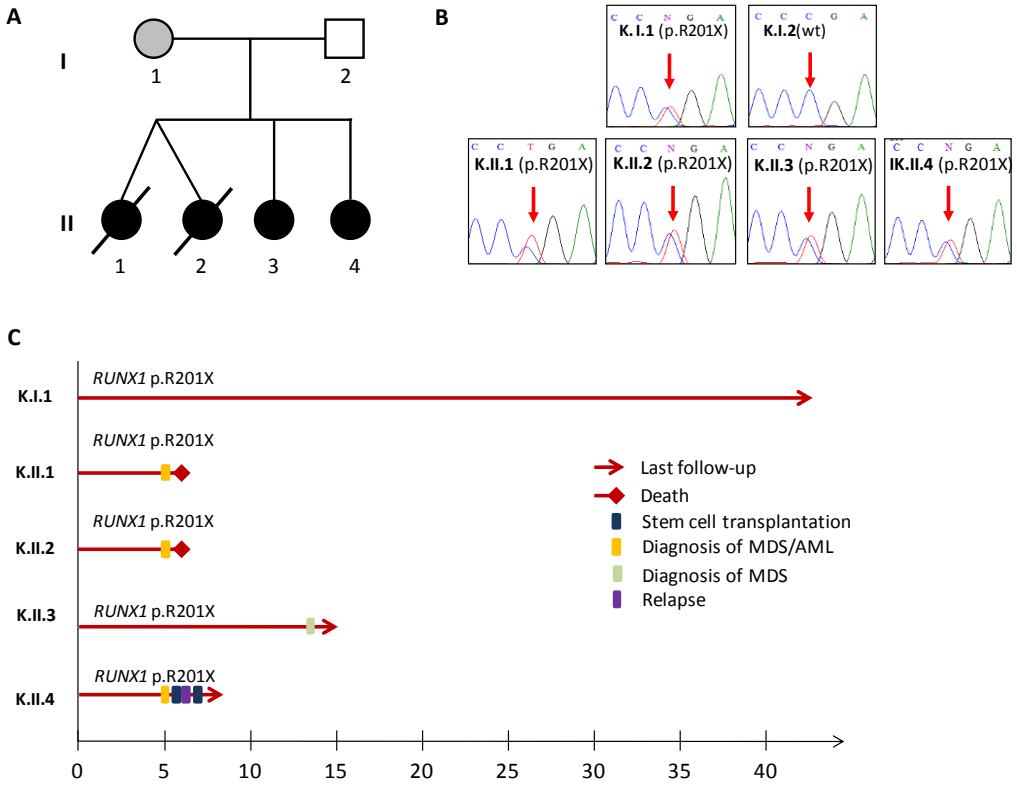


Figure 4.1 Characterisation of a novel *RUNX1*-mutated pedigree

(A) This pedigree had four affected siblings including dizygotic twins. (B) Sanger sequencing of RUNX1 identified the germline mutation p.R201X in all four siblings and their mother was an asymptomatic carrier (shaded grey). (C) Clinical timeline of pedigree K showing that siblings K.II.1, K.II.2 and K.II.4 all developed MDS transforming to AML at 5 years whilst the sibling K.II.3 remains under expectant follow up.

4.4.3 WES metrics for pedigree K

The mean sequencing coverage across all 6 family members was 89x and diagnostic tumour samples from the siblings demonstrated a mean coverage of 96x (Table 4.1). Following exclusion of inherited variants by comparison with maternal and paternal exomes, the median number of somatic variants detected in diagnostic samples from K.II.1-K.II.4 was 11 (range 10-11), with non-synonymous variants predominantly represented by base substitutions (n=39, 97% of variants, Appendix 3, Table 1). Since our WES bioinformatic pipeline was extensively verified in Chapter 3, selected somatic mutations were verified using Sanger sequencing, achieving concordance of 88%, (Appendix 3, Table 2).

The germline *RUNX1* mutation was confirmed in WES data from the diagnostic tumours of all 4 siblings and their mother, with VAFs ranging from 42-67%. The lower VAF of 11.6% in K.II.4-T2 indicated that <25% of BM cells were of host origin, indeed all of these were likely to be primitive cells heralding AML relapse. In the salivary DNA (sDNA) sample of K.II.4, the observed mutant *RUNX1* VAF of 35% most likely reflected contamination with wild-type donor lymphocytes, comprising approximately 30% of cells within the sample. These data are summarised in Table 4.2.

The mother was clinically asymptomatic, however to exclude somatic mutations associated with latent (pre-leukaemic) clonal expansions, WES reads for candidate genetic loci (in particular *DNMT3A*, *TET2*, *ASXL1*, *SMC1A, SMC3, EZH2* and *JAK2*) were compared with the hg19 reference genome. Novel (non-polymorphic) variants were not detected at an average sequencing depth of 80x, confirming the germline *RUNX1* mutation appeared to be the sole lesion of note in the maternal sample.

Table 4.1 WES coverage for pedigree K samples

Sample	Tissue source	Sequenced nucleotides (Gb)	Mean exome coverage	% bases ≥ 10X coverage
K.I.1	normal PB	4.07	80.3	95.9
K.I.2	normal PB	3.77	74.5	95.4
K.II.1	sAML BM	5.52	109.1	97.1
K.II.2	sAML BM	4.58	90.4	96.4
K.II.3	MDS BM	5.55	109.5	96.4
K.II.4-T1	Diagnostic sAML-BM	3.76	74.4	97.1
Mean		4.5	89.7	96.4

Table 4.2 WES VAFs of germline *RUNX1* mutation across family members

Patient	Age, yrs	Disease state	Sample	*RUNX1*	VAF (%)
K.I.1	44	Asymptomatic carrier	PB DNA	R201X	50
K.I.2	43	Healthy	PB DNA	WT	0
K.II.1	5	sAML	BM DNA	R201X	66.7
K.II.2	5	sAML	BM DNA	R201X	42.4
K.II.3	14	MDS	BM DNA	R201X	41
K.II.4-T1	5	sAML	BM DNA	R201X	52.3
K.II.4-T2	6	sAML-relapse post cord HSCT	BM DNA-relapse	R201X	11.6
K.II.4-sDNA	6	Salivary DNA post cord HSCT	Salivary DNA	R201X	34.2

4.4.4 Molecular variants detected by WES

The key somatic mutations identified within the siblings' tumour samples are detailed below and are summarised in Figure 4.2.

4.4.4.1 JAK-STAT signalling lesions

Analysis of WES data revealed molecular addiction to JAK2 signalling in siblings K.II.1, K.II.2 and K.II.4. The gain of function *JAK2* p.V617F mutation, frequently observed in myeloproliferative disorders, was detected in siblings K.II.2 and K.II.4. While in K.II.1, a unique missense mutation (p.R392Q, G1532A) was detected in *SH2B3*, also known as *LNK*, a lymphocyte specific adaptor protein. SH2B3 is a negative regulator of JAK2, binding both WT and mutant JAK2 isoforms and inhibiting their phosphorylation (Bersenev *et al*., 2008). The p.R392Q mutation was located within the highly conserved SH2 domain, through which JAK2 binds to SH2B3. This variant was predicted to have a disruptive or damaging effect on protein function (MutationTaster p=1.0000 4d.p.) and was therefore likely to promote JAK2 phosphorylation and signalling through loss of negative regulation. *JAK2* and *SH2B3* mutations were not detected in sibling K.II.3 or the children's mother (K.I.1).

4.4.4.2 *CDC27*: A component of the anaphase-promoting complex

The cell division cycle protein 27 homologue, *CDC27*, on 17q21.32, is one of several subunits comprising the anaphase promoting complex (APC) (Schreiber *et al*., 2011). CDC27 controls the metaphase-to-anaphase transition of the cell cycle and is regulated by spindle checkpoint proteins (OMIM 116946). The APC is an E3 ubiquitin ligase that targets cell cycle regulatory proteins for degradation by the proteasome, allowing progression through the cell cycle (Jörgensen *et al*., 2001). Mutations of *CDC27* are rarely observed in haematopoietic malignancies, although they occur frequently in solid tumours, particularly pancreatic cancer and melanoma, affecting over 20% of cases within the TCGA dataset (CBioportal, Gao *et al*., 2013).

Siblings, K.II.2 (sAML) and K.II.4 (MDS) both acquired somatic mutations of *CDC27*. In K.II.2, the novel variant *CDC27* p.I634T (T1905C, VAF 16.7%) was detected and predicted to have a damaging effect on protein function (probability 1.0000 4d.p., MutationTaster). A subclonal *CDC27* variant, p.I493V (A1481G) was detected in K.II.4

with VAF of 7.5%. Despite the latter being identified as a rare SNP (rs140737545) with a minor allele frequency (MAF) of 0.015 (1.5%), it was still predicted to have a deleterious effect on protein function (probability 0.9994), suggesting a pathogenic role for both variants and alluding to the acquisition of a common co-operating lesion in both tumours.

4.4.4.3 Retinoblastoma binding protein 8

Retinoblastoma binding protein 8 (*RBBP8*) is located on chromosome 18q11.2 and encodes a ubiquitously expressed nuclear protein. The gene is also known as *CTIP* (CTBP interacting protein) as it is complexed to the transcriptional co-repressor, *C-terminal binding protein* (CTBP). CTIP/RBBP8 is a multifunctional protein involved in both the detection of DNA damage and its repair by homologous recombination (HR) and microhomology-mediated end-joining (MMEJ), in addition to regulating transcription, (Yun *et al*., 2009). These multiple functions are controlled by interaction with several factors, including the MRN (MRE11-RAD50-NBS1) complex and the tumour suppressors, BRCA1 and retinoblastoma (RB1) (reviewed by You and Bailis, 2010).

A novel missense mutation of *RBBP8* was detected in K.II.2, p.E320G (A1307G) with an allelic frequency of 32.3%. This variant was located near Serine 327 (S327), which following phosphorylation, mediates the interaction of RBBP8 with BRCA1 and MRN (Yu *et al*., 2004; Chen *et al*., 2008). Using MutationTaster, the p.E320G variant was predicted to have a deleterious effect on protein function (probability 0.8415 4d.p), potentially disrupting the S327 locus and its subsequent role in DNA repair.

4.4.4.4 U2 small nuclear RNA auxiliary factor 2

U2 small nuclear RNA auxiliary factor 2 (*U2AF2*) is located on chromosome 19q13.42 and encodes a component of the spliceosome machinery, which is required for splicing of pre-mRNA. Each spliceosome is composed of small nuclear RNAs (snRNAs) and associated proteins, which combine to form small nuclear RNA-protein complexes (snRNPs). U2AF2 forms a heterodimer with U2AF1 and is responsible for the recognition of the 3'splice site and its nearby polypyrimidine tract, which is thought to be required for the subsequent recruitment of the U2snRNP (containing SF3A1 and SF3B1) forming the splicing A complex (reviewed by Wahl *et al*., 2009; Yoshida *et al*., 2011).

Spliceosome mutations comprise the most commonly mutated subgroup in MDS and are recognized as founding lesions in the majority of cases. In an extensive analysis of over 550 cases, mutations of *U2AF1* were found to be a recurrent event occurring at a frequency of ~6%, while *U2AF2* mutations were detected in <1% of cases, highlighting their relative rarity (Yoshida *et al*., 2011). The *U2AF2* mutation detected in K.II.4-T1 (p.Q147E [C479G] VAF 41.2%) was novel and predicted to be deleterious using MutationTaster (probability 1.000 to 4dp), suggesting disruption of spliceosome function via this component was important for leukaemogenesis.

4.4.5 Chromosomal LOH within tumour samples

All diagnostic BM samples were subjected to metaphase G-banding and fluorescence in situ hybridisation (FISH) analysis, for detection of chromosomal deletions, duplications and translocations in the initial diagnostic assessment at Semmelweis University, Budapest. Although AML in K.II. 4 was associated with a normal karyotype, tumour samples in her twin sisters (K.II.1 and K.II.2) revealed multiple copy gains and losses, which were verified using both WES and MLPA. WES provided a valuable method for further characterisation of LOH regions, assessing copy neutral LOH (or aUPD), which proved a highly significant feature within K.II.1 and K.II.2.

G-banding of BM metaphases was performed twice during the dysplastic phase of disease in K.II.1, with the first (S1) and second (S2) samples taken one month apart. Neither of these samples were available for WES. In S1, the karyotype 47,XX,+21[15]/47,XX,+21,del(2)(q33)[10] was detected, suggesting that del(2)(q33) had arisen in a subclone of cells with +21. After one month, the S2 karyotype had evolved further to 47,XX,+21[24]/47,XX,+21,del(2)(q33)[2], indicating expansion of the founding (47,XX,+21) clone.

A third sample (S3) taken at leukaemic transformation 4 months following S2, revealed complete absence of del2q33 (also confirmed by WES analysis). This suggested that isolated expansion of the +21 clone was associated with disease transformation. Verification of +21 in S3 was performed using MLPA (X060-MRC Holland) with multiple probes in *RUNX1* and *U2AF1*, both located on chromosome 21q (Figure 4.3 A). Notably,

the duplication specifically amplified the mutant *RUNX1* allele, increasing the germline mutation burden within leukaemic cells (VAF p.R201X=0.68).

LOH analysis in WES reads from K.II.2 revealed monosomy 7 and deletion of 9q22, the former was verified using MLPA with multiple chromosome 7 probes (Figure 4.3 B). While monosomy 7 is detected in approximately 10% of sAML (Grimwade *et al*., 1998), chromosome 9q deletions are relatively rare, occurring in only 2% of cases. Interestingly, 36-50% of all 9q deletions co-occur with t(8:21) (*RUNX1*:*RUNX1T1*), suggesting specific co-operation with the fusion gene (Grimwade *et al*., 1998; Schoch *et al*., 1996; Grimwade *et al*., 2001). The deleted region of chromosome 9q in K.II.2 (chr9:80,412,527-109,484,524) encompassed *SMC2*; this encodes a condensin complex component which is recurrently mutated in solid tumours, most commonly colorectal cancer (8% of cases in TCGA dataset, CBioportal). Although *SMC2* mutations have not been reported in myeloid malignancies to date, alternative components of the cohesin complex, encoded by *SMC1A* and *SMC3* comprise approximately 10% of mutations in AML (TCGARN, 2012).

4.4.6 Copy-neutral LOH increases mutation dosage

Regions of copy neutral LOH (also known as aUPD) were detected in tumour samples from K.II.1 and K.II.2; both regions encompassed mutations within the JAK-STAT pathway, increasing the mutant allelic burden. In K.II.1, aUPD of chromosome 12q caused homozygosity of the *SH2B3* mutation (VAF 67.2%), while In K.II.2 aUPD of chromosome 9p led to homozygosity of the *JAK2* p.V617F mutation (VAF 68.2%), as shown in Figure 4.3 C. Although 12q aUPD is an infrequent event in myeloid malignancy, 9p aUPD occurs in approximately 30% of sporadic V617F myeloproliferative disorders, most commonly polycythaemia rubra vera (PRV) and post-PRV myelofibrosis (reviewed by Viny and Levine, 2014).

Collectively these data revealed multiple genetic lesions within tumour samples from the 3 siblings K.II.1, K.II.2 and K.II.4. A combination of mutations, chromosomal CNAs and aUPD were detected, with the genetic profiles for each sibling summarised in Figure 4.2. LOH analysis from WES data and verification of CNAs using MLPA is shown in Figure 4.3.

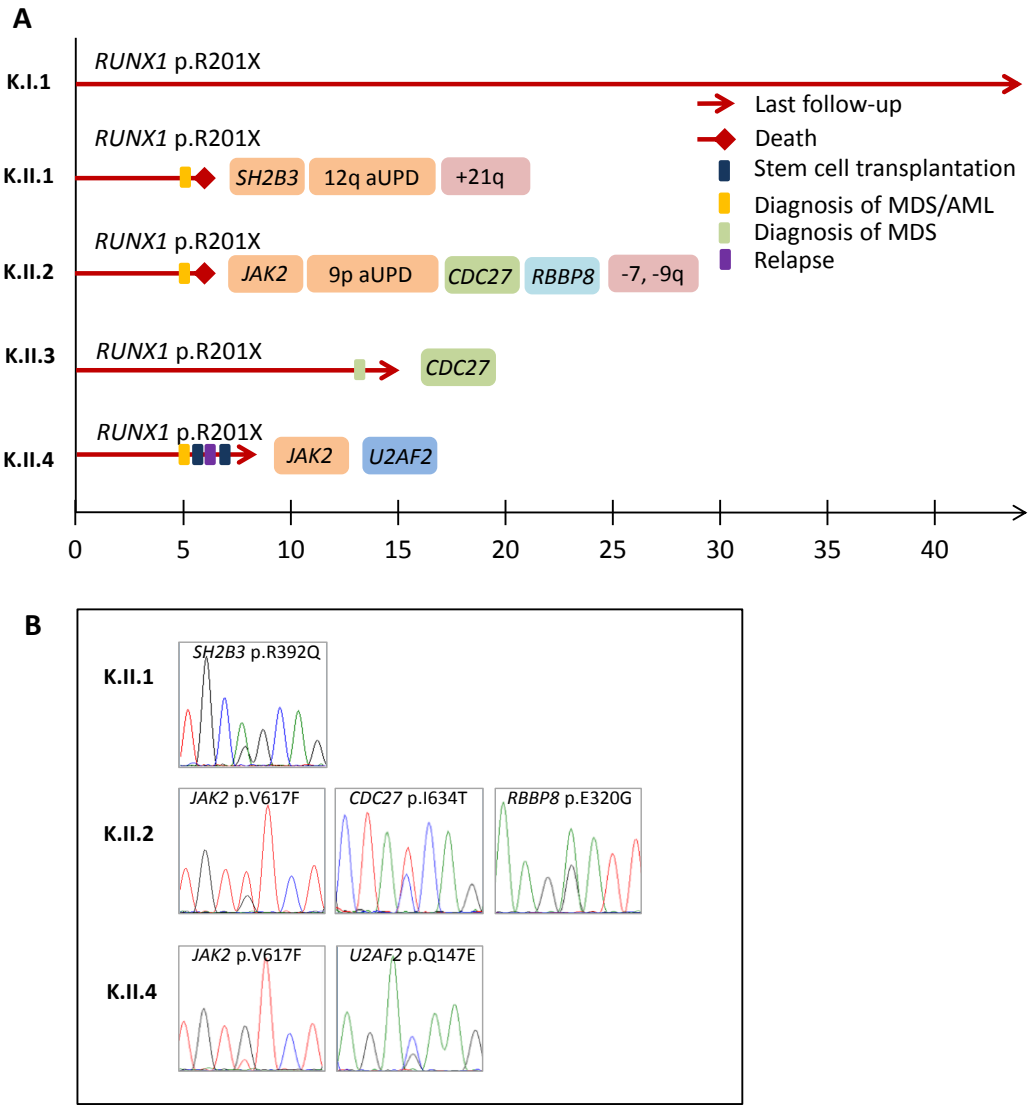
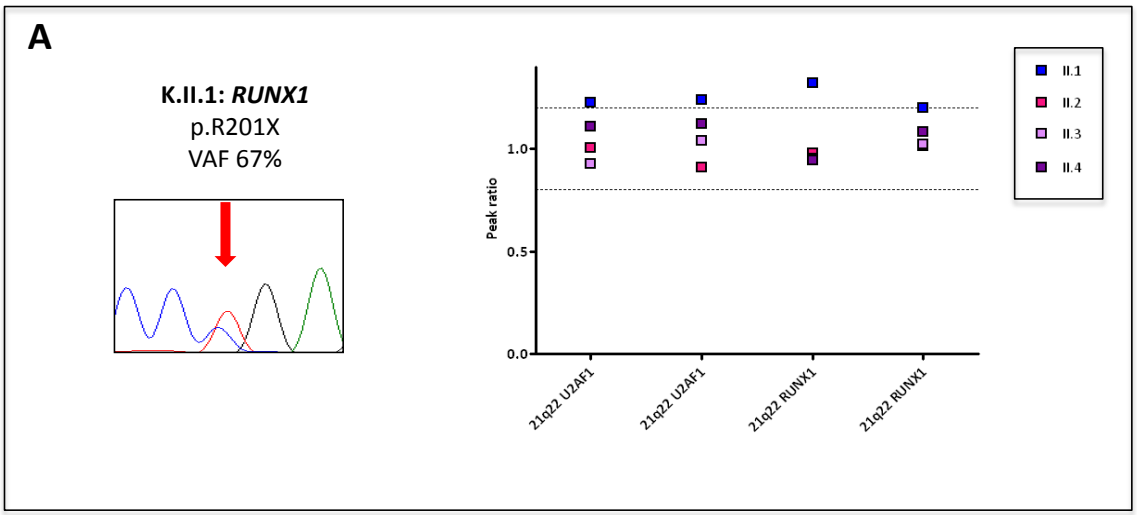
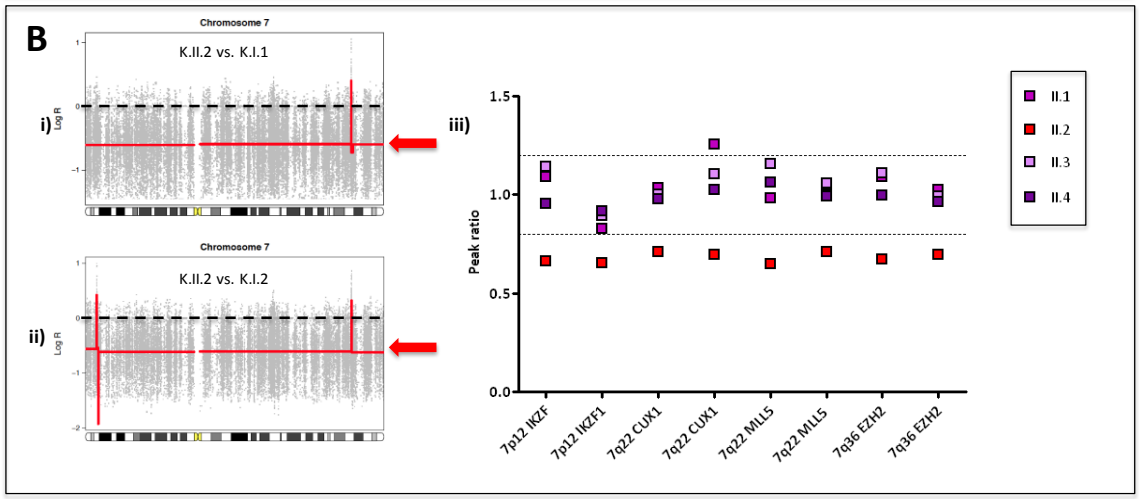


Figure 4.2 Somatic mutations and LOH in tumour samples from family K.

(A) Summary of the key genetic lesions detected in all four siblings with the timeline of clinical events. (B) Sanger sequencing of mutations detected in K.II.1, K.II.2 and K.II.4. The *CDC27* mutation in K.II.3 could not be verified by Sanger sequencing due to its low VAF (7%).


A
K.II.1: RUNX1
p.R201X
VAF 67%
Peak ratio
1.0
0.5
0.0
II.1
II.2
II.3
II.4
21q22 U2AF1
21q22 U2AF1
21q22 RUNX1
21q22 RUNX1


B
Chromosome 7
K.II.2 vs. K.I.1
i)
Log R
Chromosome 7
K.II.2 vs. K.I.2
ii)
Log R
iii)
Peak ratio
1.5
1.0
0.5
0.0
II.1
II.2
II.3
II.4
7p12 IKZF
7p12 IKZF1
7q22 CUX1
7q22 CUX1
7q22 MLL5
7q22 MLL5
7q36 EZH2
7q36 EZH2

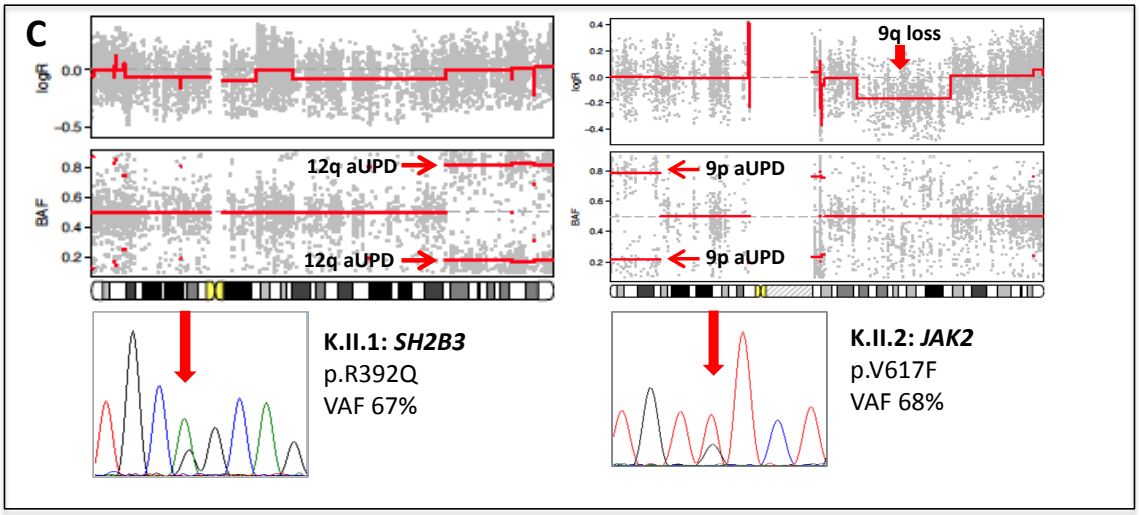

C
logR
BAF
12q aUPD
12q aUPD
9q loss
9p aUPD
9p aUPD
K.II.1: SH2B3
p.R392Q
VAF 67%
K.II.2: JAK2
p.V617F
VAF 68%

Figure 4.3 Somatic CNAs and aUPD detected in twin-sibling tumours.

(A) The germline *RUNX1* pR201X mutation dosage was increased due to amplification of 21q in tumour DNA of K.II.1. This was demonstrated by the elevated peak ratio (>1.2) for 21q MLPA probes in K.II.1, compared with those of her siblings. The dotted lines show the peak ratio thresholds representing normal copy number (peak ratio 0.8-1.2) (B) (i and ii) The LRR of germline variants across chromosome 7 revealed heterozygous deletion in tumour DNA of K.II.2 compared with each parent, this was subsequently verified by MLPA (iii), with peak ratios for 7p and 7q probes ranging between 0.65-0.70. (C) In K.II.1, BAF values for germline variants on chromosome 12q demonstrated aUPD, causing homozygosity of the *SH2B3* mutation p.R392Q. In K.II.2, similar analysis revealed 9p aUPD, increasing the *JAK2* p.V617F burden, 9q loss was also detected in this tumour, as seen by the reduced LRR of variants in this region.

4.4.7 Deciphering the sequence of genetic events in K.II.1 and K.II.2

The detection of multiple somatic chromosomal lesions within the tumour DNA of twins K.II.1 and K.II.2, provided an opportunity to elucidate the sequence in which these lesions were acquired. In brief, Kernel density plots (Figure 4.4 A and B) were used to determine the peak BAF of germline SNPs within each chromosomal region of interest, enabling the calculation of tumour purity and subsequent estimation of CCF or clonality (see section 4.3.1.2). Importantly, in all calculations, the specific allelic representation occurring in each aberration was considered, with 'A' as the reference and 'B' as the variant allele. For example, 'ABB', 'BB' and 'A' corresponded to amplification, aUPD and deletion of the B-allele, respectively. After determining the clonality of each lesion, it was possible to extrapolate a timeline of genetic evolution, with dominant clonal lesions considered early events.

In K.II.1, the tumour purity was estimated to be 73% with 12q aUPD harbouring the greatest CCF, this was followed by gain of 21q in 68.5% of cancer cells. In K.II.2, the tumour purity was estimated as 68.4%, with monosomy 7 comprising the highest CCF, followed by 9p aUPD and del9q. In both K.II.1 and K.II.2, homozygous VAFs of *SH2B3* and *JAK2* mutations, suggested their acquisition preceded mitotic recombination of 12q and 9p respectively, thereby permitting the sequence of events to be determined as shown in Figure 4.4 C. In K.II.4, the CCF and clonality of mutations was determined according to tumour VAFs alone, as LOH was not detected.

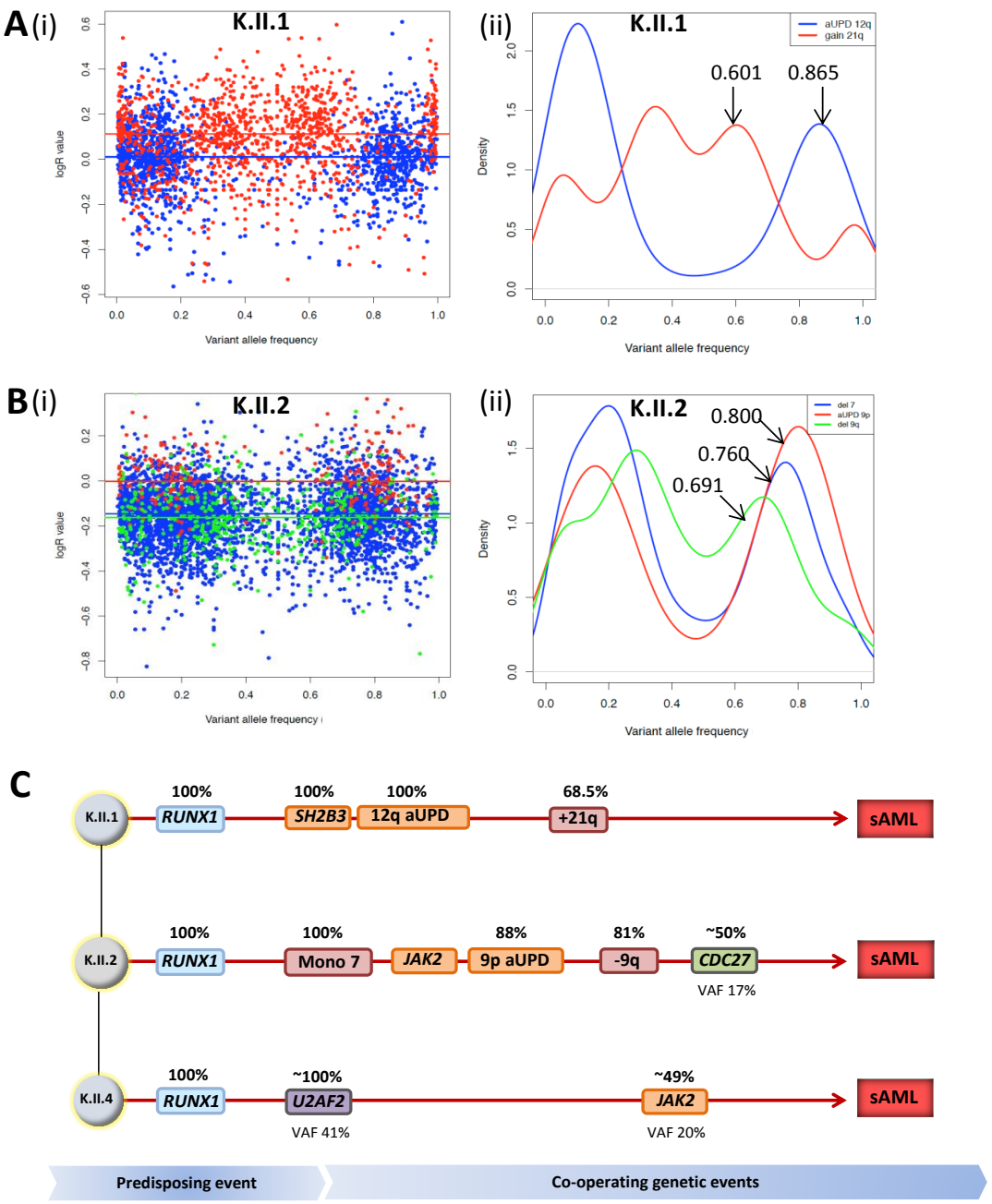


Figure 4.4 Sequential acquisition of CNAs and aUPD during leukaemic evolution.

(A) i). LRR for germline variants encompassed within regions of 12q aUPD (blue) and gain of 21q (red) in K.II.1. ii) BAF density plots of germline variants within the corresponding regions of LOH, the peak density BAF is shown for each cluster. (B) i) LRR for germline variants located on chromosome 7 (monosomy 7, blue) and regions of 9p aUPD (red) and deletion of 9q (green) in K.II.2. ii) BAF density plots of variants within the corresponding regions, with the peak density BAF shown for each cluster. (C) From the calculations in (A) and (B) the CCF for each LOH segment was calculated, allowing the chronological order of events to be determined, as shown in the timeline. In K.II.4, LOH was not detected at diagnosis and variant CCFs were estimated using the mutation VAF (VAF).

4.4.8 Clonal evolution at relapse in K.II.4

As mentioned previously, sibling K.II.4 presented with sAML at 5 years and underwent allogeneic HSCT from double cord-blood units post induction chemotherapy. AML relapse (T2) occurred after 13 months and WES of T2 BM DNA was performed, with salivary DNA (collected post allograft) serving as a germline control.

Comparison of T2 against the reference genome identified four clusters of variants, with mean VAFs corresponding to 10%, 40%, 80% and 100%. The germline *RUNX1* VAF in T2 was 11.6%, this suggested that <25% of cells were of host origin and ~75-80% of cells originated from the dominant cord-blood unit. The clusters represented heterozygous and homozygous variants within host and donor cells, confirming that relapse had occurred in the context of mixed chimerism. Somatic mutations were then detected by comparing T2 with salivary DNA and, in light of the low tumour purity, we focused on somatic variants with VAFs ranging between 5-25% (to include heterozygous and homozygous host variants). Residual host germline variants were excluded by comparison with both parental exomes. Unfortunately, over 150 apparently 'somatic' variants remained, the majority of which were annotated SNPs, most likely representing germline donor variants. This prevented any meaningful interpretation of novel, relapse-specific, mutations. We therefore selectively traced the clonal evolution of pre-existing mutations.

To determine the pattern of evolution in individual K.II.4, we compared purity-corrected VAFs for mutations at diagnosis (T1) and relapse (T2). In T1, the *U2AF2* p.Q147E mutation represented a dominant clonal lesion with a corrected VAF of 38% (suggesting that 76% of cells contained this heterozygous mutation). In contrast, *JAK2* V617F had a corrected T1 VAF of 19%, suggesting heterozygosity in 38% of cells. These calculations suggested that the *JAK2* mutation, most likely developed within a subclone of the *U2AF2*-mutated founding clone.

Both lesions recurred in T2, although the *JAK2* V617F mutation burden increased (corrected VAF 43%), while the mutant *U2AF2* level appeared to decline (corrected VAF 28.5%), as shown in Figure 4.5. This clonal divergence was an unexpected finding,

considering that both lesions were likely to originate from the same subpopulation of cells, potential explanations for this are discussed in section 4.6.5.

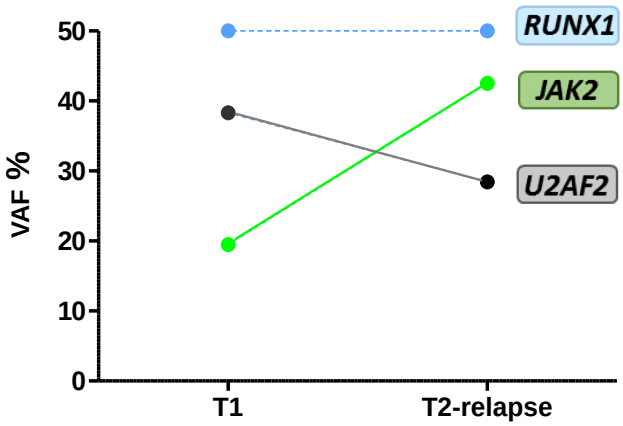


Figure 4.5 Corrected VAFs for *U2AF2* and *JAK2* mutations at diagnosis (T1) and relapse (T2) in K.II.4.

In T1, mutation of *U2AF2* appeared the dominant clonal lesion (corrected VAF 38%) with *JAK2* V617F representing a subclonal event (corrected VAF 19%). In T2, the opposite was observed, with an increase in the *JAK2* mutation burden (corrected VAF 43%), while the *U2AF2* VAF declined (corrected VAF 28%).

4.4.9 Clinical outcomes in family K

The acquisition of JAK-STAT signalling mutations in 3 siblings appeared to confer an aggressive disease phenotype with chemotherapy-refractory features. The non-identical twins, K.II.1 and K.II.2, both presented with sAML in January 2002 and were treated with intensive combination chemotherapy. K.II.1 achieved CR but died following disease relapse one year later. K.II.2 failed to achieve CR and died from chemotherapy-refractory disease 8 months after her initial presentation. Their sibling, K.II.4, also presented at 5 years with sAML, which clinically and morphologically resembled the disease phenotype in both twins. K.II.4 received induction chemotherapy comprising cytarabine, idarubicin and etoposide (AIE), with recovery complicated by delayed haematopoietic reconstitution, requiring unrelated allogeneic HSCT. Engraftment occurred 34 days post double cord blood HSCT, but was followed by relapse after 17 months. Following re-induction chemotherapy, a second VUD HSCT was performed in June 2014 and she continues in remission with full donor chimerism. Sibling K.II.3, the only sibling without a JAK-STAT mutation, is now 14 years of age and remains clinically asymptomatic, with mild thrombocytopenia and neutropenia. BM examination in K.II.3 revealed evidence of dysplasia (suggesting RCMD) and she is currently managed expectantly. The mother, K.I.1 remains an asymptomatic healthy carrier, with normal peripheral blood counts and BM morphology. Sanger sequencing of maternal PB DNA confirmed the absence of *SH2B3* (all exons) and *JAK2* (exon 14) mutations.

4.4.10 *JAK2* haplotype analysis

Increasing evidence suggests that germline polymorphisms (representing a specific haplotype) may be associated with the risk of somatic mutation acquisition. In 2009, several studies reported that the *JAK2* 46/1 haplotype was significantly associated with the acquisition of somatic *JAK2* V617F mutation and a susceptibility to MPNs (Jones *et al*., 2009; Kilpivaara *et al*., 2009; Olcaydu *et al*., 2009). In light of these reports, the haplotype of all 4 siblings and their parents was assessed by Sanger sequencing of the *JAK2* SNP rs12340895, which is in complete linkage disequilibrium (LD) with the 46/1 allele.

All 4 children were heterozygous for the 46/1 tagged SNP (rs12340895 G/C). Testing of their mother revealed homozygosity of the 46/1 (GG) allele, while their father was homozygous for the non-46/1 allele (CC), Figure 4.6 B. The V617F mutation in K.II.2, occurred on the 46/1 allele with both respective variants appearing homozygous due to 9p aUPD. Of note, homozygous and heterozygous 46/1 status are associated with an increased predisposition to *JAK2* V617F acquisition and subsequent development of MPNs (Kilpivaara *et al*., 2009). The absence of disease in the mother, in addition to acquisition of an *SH2B3* mutation in K.II.1, (not, as yet, associated with the 46/1 haplotype) suggests that additional factors governed JAK-STAT convergence within this family.

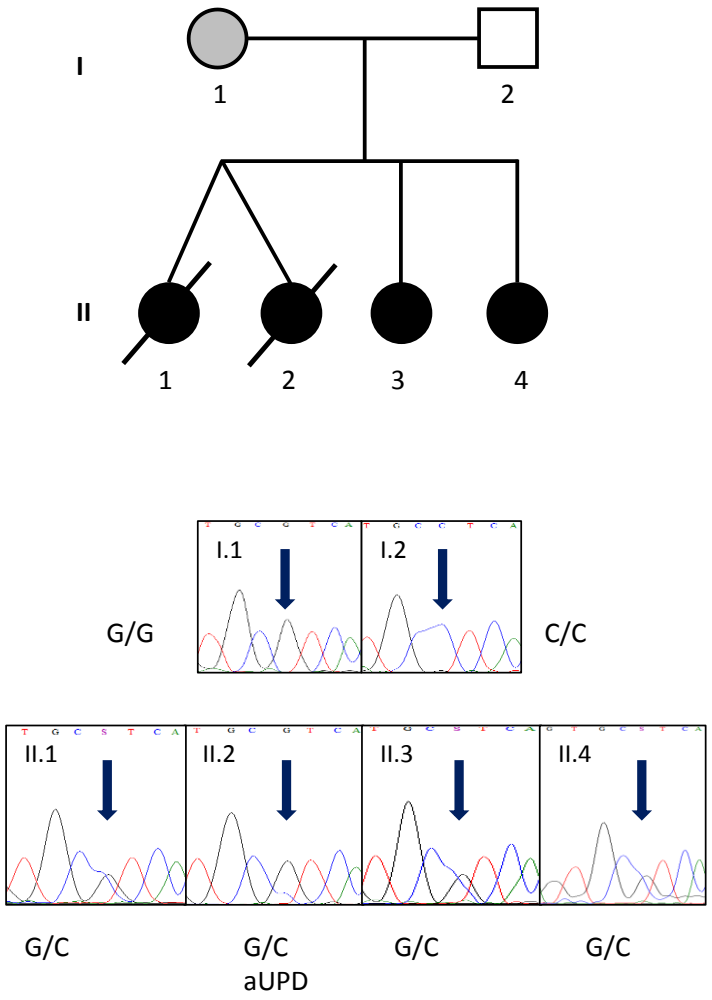


Figure 4.6 Sanger sequencing of JAK2 46/1 tagged SNP rs12340895 in pedigree K.

Seqencing of rs12340895, revealed the mother (K.I.1) was homozygous for the 46/1 haplotype (GG) and the father was homozygous for the non-46/1 allele (CC). Consequently, all four children were heterozygous for the 46/1 haplotype which is associated with an increased frequency of *JAK2* V617F acquisition.

4.4.11 Sporadic AML with *RUNX1* and *JAK2* mutations

JAK2 and *SH2B3* mutations are rare in sporadic AML, occurring in <1% of the TCGA cohort (TCGARN, 2013). The findings in family K suggested molecular co-operation between the RUNX1 and JAK-STAT signaling pathways. To test this hypothesis, we identified 20 cases of *RUNX1* mutated AML (16 sporadic and 4 cases of familial AML (unrelated to pedigree K). Sanger sequencing was performed for *JAK2* V617F and all coding exons of *SH2B3*. A single *JAK2* V617F mutant was identified, suggesting co-occurrence with *RUNX1* mutation in 5% of cases. This is similar to findings in the TCGARN cohort (2013) where single *SH2B3* and *JAK2* V617F mutations were identified within the entire cohort, both co-occurring with mutations of *RUNX1* at a frequency of 6% (1/16 sporadic *RUNX1*-mutated cases).

4.5 Limitations of study

This study provided a rare opportunity to study historical samples from this novel FPD-AML pedigree, with multiple affected members. While interesting observations were made, the lack of paired tumour/remission and sequential disease samples limited more comprehensive investigation. Maternal and paternal exomes were used as a germline reference for tumour samples from all four siblings, to help distinguish somatic and germline variants. Since the mother was an asymptomatic carrier with normal peripheral blood count indices and BM morphology, the use of maternal PB DNA as a germline reference assumed that somatic pre-leukaemic mutations were absent. In light of recent publications highlighting clonal haematopoiesis with normal haematological parameters (discussed in Chapter 1), such lesions cannot be fully excluded. Given that identical somatic mutations may be acquired amongst family members (as described above and in Chapter 3), there is a possibility that shared somatic lesions between the mother and children may pass undetected by our analytical pipeline.

Full evaluation of clonal evolution in K.II.4 proved difficult due variation of tumour purity between the disease episodes and 'contamination' of T2 with donor DNA following cord-blood HSCT. To help correct for variations in tumour purity, the VAFs of mutations detected in both T1 and T2 were adjusted to the germline *RUNX1* VAF in the corresponding tumour. In T1, there was no evidence of LOH at the *RUNX1* locus and this was assumed to be the same for T2, although it was not possible to confirm this due to the low tumour purity of the sample.

Finally, the finding of recurrent somatic JAK-STAT mutations in a *RUNX1* mutated pedigree suggests specific molecular co-operation between these lesions. We have attempted to investigate this further in a limited cohort (n=20) of *RUNX1* mutated AML, although extensive testing of a larger cohort is required to verify both the frequency of co-occurrence and associated survival outcomes within this subgroup of MDS/AML.

4.6 Discussion

In this study we have implemented WES to identify recurrent somatic mutations and LOH in multiple family members harbouring the germline *RUNX1* mutation p.R201X. This has revealed a pattern of non-random somatic mutation acquisition with concurrent LOH increasing the mutation dosage within tumours.

4.6.1 JAK-STAT signalling activation within Pedigree K

Molecular addiction to JAK-STAT signaling was demonstrated in 3 siblings with MDS/AML within this novel *RUNX1*-mutated pedigree. The siblings K.II.1, K.II.2 and K.II.4 acquired mutations in *JAK2* and *SH2B3*, both of which are infrequently mutated in sporadic MDS/AML (*JAK2* <5%, *SH2B3* <1% [TCGARN, 2013]). Notably, mutations in both genes were predicted to upregulate JAK-STAT signaling. This is well described for the *JAK2* V617F mutation which disrupts auto-inhibitory activity, leading to constitutive activation of the kinase domain. Activation of JAK2 promotes phosphorylation of the transcription factors STAT3 and STAT5, enabling formation of stable homodimers and heterodimers which subsequently translocate to the nucleus, whereupon they regulate expression of genes associated with cell proliferation, differentiation, and apoptosis (reviewed in Quintas-Cardama and Verstovsek, 2013). The acquisition of V617F in K.II.2 and K.II.4 therefore served as a proliferative signal to expedite disease transformation, with 9p aUPD further augmenting this effect in K.II.2.

SH2B3 is highly expressed in multipotent haematopoietic stem cells and progenitor cells. Activation of the JAK-STAT pathway induces SH2B3, which then inhibits several tyrosine kinases and cytokine signaling pathways (JAK, KIT, MPL, EPOR, PDGFRA/B, FLT3). Several studies have demonstrated the critical role of SH2B3 in regulating HSC self-renewal and quiescence. Mice lacking SH2B3 had an expanded HSC pool with enhanced self-renewal and quiescence, mediated by thrombopoietin (TPO) and JAK2 signalling (Buza-Vidas *et al*., 2006; Seita *et al*., 2007; Bersenev *et al*., 2008). Homozygous knockout mice also displayed abnormal B-lymphopoiesis, megakaryopoiesis and erythropoiesis (Tong *et al*., 2004; Tong *et al*., 2005), illustrating the extent of haematopoietic dysregulation in the absence of SH2B3.

Members of the SH2B family directly interact with JAK2 via the SH2 domain and the phosphorylated tyrosine residue 813 (Y813). This association regulates both JAK2 activity and the subsequent phosphorylation of the adaptor proteins (Kurzer *et al*., 2006). Baran-Marszak *et al* (2010) have further confirmed binding of SH2B3 to JAK2 (both WT and V617F) via an intact SH2 domain, leading to constitutive phosphorylation of SH2B3. Targeted disruption of the SH2 domain (with the mutation p.R364M) revealed interaction with JAK2 may still occur through a putative binding region in the N-terminal. Of note, inhibition of JAK2-mediated proliferation was lost in the R364M cells, suggesting the SH2 domain is essential for mediating the adaptor protein's regulatory function.

SH2B3 mutations occur in <5% of chronic MPNs (Oh *et al*., 2010), with no observed predilection for disease phenotype. The majority of *SH2B3* mutations are missense substitutions in the Pleckstrin-homology (PH) domain, with a hotspot spanning from Glu208 to Asp234. Notably *SH2B3* mutations were detected in 13% of MPN patients in blast phase, suggesting an association with high grade transformation (Pardanani *et al*., 2010). Although *SH2B3* (*LNK*) mutations occur in <1% of sporadic AML and MDS (TCGARN, 2013; Malcovati *et al*., 2014), they occur more frequently in early T-cell precursor ALL, with mutations reported in 6% of cases (Zhang *et al*., 2012). A recent report also suggested a tumour-suppressor role for SH2B3, with multiple members of a consanguineous family bearing homozygous PH domain mutations and presenting with autoimmune disease and ALL (Perez-Garcia *et al*., 2013).

In the current study, the somatic *SH2B3* mutation detected in K.II.1 (p.R392Q) targeted the SH2 domain and was predicted to cause a loss of function. The subsequent acquisition of 12q UPD, led to homozygosity and complete disruption of SH2B3 in tumour cells. This feasibly provided a proliferative signal by abolishing negative regulation of JAK2 signalling and co-operating with RUNX1 dysfunction to accelerate leukemogenesis.

4.6.2 JAK-STAT mutations may co-operate with RUNX1 dysfunction

The apparent clustering of somatic JAK-STAT mutations within the siblings' tumours suggests that these lesions are non-randomly acquired (and positively selected) in the context of RUNX1 disruption. Co-occurrence of lesions within these key pathways is also observed in t(8:21) AML, where 3-8% of cases harbour the *JAK2* V617F mutation (Krauth *et al*., 2014; Schnittger *et al*., 2007a and b; Illmer *et al*., 2007). Notably, acquisition of *RUNX1* mutations has also been reported following leukemic transformation of MPNs, particularly idiopathic myelofibrosis (IMF), where *JAK2* V617F positivity is maintained upon disease transformation (and is often subject to aUPD) (Beer *et al*., 2010; Engle *et al*., 2014). This suggests that co-operation between mutant *RUNX1* and *JAK2* occurs irrespective of the order in which these lesions are acquired, although the sequence of acquistion appears to influence the clinical phenotype observed. For example, mutation of *JAK2* followed by *RUNX1* corresponds to leukaemic transformation of MPNs, while founding *RUNX1* mutations are associated with primary dysplastic or leukaemic phenotypes. These findings are reminiscent of the recent report by Ortmann *et al* (2015) where the order of *JAK2* and *TET2* mutation acquisition influenced both the clinical phenotype (*JAK2* first cases were more likely to present with PRV) and the molecular and biological profiles of stem and progenitor cells.

JAK-STAT mutations appear to be enriched in AMKL associated with Down's syndrome (constitutional trisomy 21), where 35% of cases harbour lesions throughout the pathway: *JAK1* (4%), *JAK2* (8%), *JAK3* (12%), *SH2B3* (8%), and *MPL* (2%). All cases harboured acquired mutations in the transcription factor *GATA1*, although *RUNX1* mutations were not detected in this series (Yoshida *et al*., 2013). These findings suggest that JAK-STAT mutations are enriched as co-operating lesions in distinct subtypes of AML, particularly those associated with *RUNX1* mutations and translocation or trisomy of chromosome 21.

4.6.3 Non-random acquisition of mutations in family K

Over the last decade, molecular profiling of haematologic neoplasms has identified multiple recurrent genetic associations, however the factors governing the mutation acquisition remain unclear. Whilst mutagenesis has long been considered a random event that promotes clonal survival or proliferation, the findings in family K appear to suggest that the host genotype may govern the acquisition or selection of specific mutations.

Acquisition of the *JAK2* p.V617F mutation has previously been associated with the 46/1 haplotype, common in individuals of European ancestry, with mutations predominantly arising on the 46/1 allele (Jones *et al*., 2009; Kilpivaara *et al*., 2009; Olcaydu *et al*., 2009). All four siblings were found to be heterozygous for the 46/1 haplotype (GC). Surprisingly, the mother (K.I.1), was homozygous for the 46/1 allele (GG), but remained an asymptomatic carrier of the *RUNX1* mutation with WT *JAK2* (and *SH2B3*) status. This suggested that additional inherited factors were responsible for JAK-STAT mutation acquisition and disease development.

Indeed, the association between the 46/1 haplotype and *JAK2* mutation acquisition is not clearly defined. Two theories were proposed by Jones *et al* (2009): (i) the 46/1 allele may cause hypermutability (increasing the likelihood of *JAK2* mutation acquisition) and (ii) the 46/1 allele is in LD with an unknown factor, which increases the likelihood of developing disease. A combination of these concepts is also possible i.e. the 46/1 allele is in LD with an unknown factor which increases susceptibility to acquiring *JAK2* (and possibly *SH2B3*) mutations, hence increasing the risk of disease. The latter scenario, appears most likely in this family; the three siblings with MDS/AML inherited the 46/1 allele and an additional factor (of paternal origin) subsequently leading to JAK-STAT mutation acquisition and disease development. In contrast their mother, despite 46/1 homozygosity, appeared to lack the additional inherited variant and remained a healthy carrier.

4.6.4 Mutation dosage is increased by chromosomal LOH

In the twin siblings K.II.1 and K.II.2, acquired chromosomal LOH in tumour samples led to an increase of somatic and germline mutation dosage. Copy neutral-LOH (aUPD) affected chromosomes 12q and 9p increasing the dosage of somatic *SH2B3* and *JAK2* mutations in K.II.1 and K.II.2, respectively. The apparent homozygosity of these mutations suggested that they preceded the acquisition of UPD at the corresponding loci. Whilst 9p aUPD is well recognised in JAK2-V617F positive MPNs, 12q aUPD associated with *SH2B3* mutations has not previously been reported. Furthermore, gain of 21q in K.II.1, led to amplification of the germline *RUNX1* mutation, consistent with previous reports of leukaemic transformation in FPD (Preudhomme *et al*., 2009; Antony-Debré *et al*., 2015). Our calculations of CCF suggested that this occurred as a later event in disease progression, following acquisition of the *SH2B3* mutation and UPD of 12q.

4.6.5 Clonal evolution in sibling K.II.4

In K.II.4, disease relapse was associated with an increased *JAK2* V617F mutation burden, while the clonal *U2AF2* mutation reduced in frequency. The most likely explanation for these findings is that 9p aUPD, selectively increased the *JAK2* V617F mutation dosage at relapse. Acquired UPD is a common molecular mechanism observed at AML relapse (Raghavan *et al*., 2008) and frequently targets signalling lesions such as *FLT3*-ITD mutations (Griffiths *et al*., 2005; Meshinchi *et al.*, 2006; Gale *et al.*, 2008). The limited tumour purity (~20%) and donor contamination in T2 precluded LOH analysis and we were therefore unable to confirm this hypothesis.

An alternative possibility is that the *U2AF2* and *JAK2* mutations occurred in distinct clonal populations, both of which re-expanded at relapse, with the *JAK2*-mutated population gaining a selective advantage. This seems unlikely as studies have shown that AML relapse predominantly arises from evolution and expansion of a single clonal population (Ding *et al*., 2012), furthermore VAF calculations from the primary tumour suggested that *JAK2* V617F and *U2AF2* mutations co-occurred within a common subpopulation of cells.

4.6.7 Conclusion

The data in this study demonstrate convergence of somatic JAK-STAT mutations across three siblings with MDS/AML. In all three cases, JAK-STAT lesions corresponded with an aggressive disease phenotype. While the rarity of JAK-STAT mutations in sporadic MDS/AML has limited evaluation of their prognostic relevance, data from this family reveal poor clinical outcomes with evidence of overt chemotherapy-refractory disease or low-level persistence precipitating relapse.

Such a pronounced, non-random association has not previously been described within familial leukaemia and was possible through the availability of historical samples and the advent of WES technology. While the host factors giving rise to this predilection are not precisely defined, this study clearly illustrates how an individual's genotype may pre-determine the acquisition of molecular events. Furthermore, they suggest co-operation between RUNX1 dysfunction and JAK2 signalling in the development of MDS/AML, with the dosage of *JAK2*, *SH2B3* and *RUNX1* mutations amplified in tumours due to somatic LOH. Importantly, this association may also provide a valuable insight into the molecular co-operation occurring within sporadic haematological malignancies, elucidating the stepwise evolution of disease, disruption of key signalling pathways and, potentially, providing a foundation for the development of targeted therapies.

Chapter 5

Novel and Atypical Germline Mutations in Familial AML

5.1 Introduction

Familial MDS/AML is an under-reported entity with many patients unaware of their inherited predisposition. Clinical recognition of cases is often limited by the lack of relevant family history and variations in disease latency and/or phenotype. At present, this emerging field lacks consensus and uniformity in approaches to diagnostic investigation. The majority of investigators perform Sanger sequencing for all coding exons or mutation hotspots in the three established genes *RUNX1*, *CEBPA* and *GATA2*, with the option of testing telomerase mutations in cases with a preponderance to BM aplasia. Using this approach, it is possible to ascribe inherited mutations in < 50% of cases, suggesting novel genes or atypical lesions are responsible for disease predisposition in the remaining patients.

Identification of germline mutations is of critical importance for patients and their relatives; it provides a basis for screening and counseling, often with the potential to predict the impending clinical manifestations and disease prognosis. Furthermore, screening of family members allows the appropriate selection of related HSC donors for patients with MDS/AML requiring imminent HSCT. The majority of germline lesions associated with familial MDS/AML are heterozygous SNVs or short indels, however several studies suggest that screening algorithms should also incorporate techniques to assess CNAs in the candidate genes identified to date, particularly *RUNX1* and *GATA2*. Moreover, and outside the remit of this thesis, the possibility exists that mutations within the non-coding genome or epigenetic modifications may also be an important component in the aetiology of these diseases.

The initial report of *RUNX1* lesions in FPD/AML (Song *et al* 1999) described 6 families, with one harbouring a complete allelic deletion. Since this seminal report, germline *RUNX1* sequence variants have now been identified in over 30 families worldwide, in contrast, investigation of inherited *RUNX1* CNAs has been limited. To address this issue, Jongmans *et al* employed array CGH (aCGH) to study CNAs in families with a characteristic FPD-AML phenotype, where sequencing was wild-type. They reported one family with a germline duplication of *RUNX1* (with breakpoints in introns 1 and 6), while a second harboured an intragenic deletion with loss of exons 5 and 5b, both lesions segregated with disease.

This phenomenon is not limited to *RUNX1*, partial allelic deletions of *GATA2* involving the N-terminal exons 2 and 3 and the initial ATG transcription start site (TSS), have also been reported (Hsu *et al*., 2011). To date, these lesions are predominantly associated with the phenotype of MonoMac syndrome due to haploinsufficiency of *GATA2* and nonsense mediated decay (NMD) of the truncated transcript. As mentioned in Chapter 1, additional reports of MonoMAC phenotypes have identified mutations within an intronic *GATA2* promoter (Johnson *et al*., 2012), suggesting that non-coding lesions may also be capable of inducing disease. In contrast to *RUNX1* and *GATA2*, as yet, there have been no reports of germline *CEBPA* allelic deletions or non-coding lesions associated with familial AML.

With the advent of NGS technologies, we are now equipped with the tools to perform unbiased genomic investigation enabling the identification of novel inherited lesions. This type of analysis is aided by collections of multiple affected family members with corresponding clinical information. Unfortunately, in practice, this is often difficult, and research investigations focus primarily on the testing of single members across multiple families sharing a similar disease phenotype. Although this approach is not without difficulty, requiring a large number of families, it has led to the recent identification of novel leukaemia predisposing lesions in *ETV6* (Zhang *et al*., 2015; Noetzli *et al*., 2015) and the RNA splicing component, *DDX41* (Polprasert *et al*., 2015).

5.2 Aims and Objectives

The main objective of this chapter was to expand our existing knowledge of the germline variants predisposing to familial MDS/AML, aiming to further characterise the genetic landscape within this heterogeneous patient population. To address this purpose, the chapter is essentially divided into 2 parts; the first focuses on the detection of atypical lesions in known genes (in particular *RUNX1*, *CEBPA* and *GATA2*), while the second aims to characterise novel germline mutations using WES in families where no lesions have been detected within the known predisposition loci.

5.3 Materials and Methods

5.3.1 DNA Extraction and Sanger Sequencing

The Oragene OG-500 kit (DNA Genotek) was used for collection of saliva samples, with genomic DNA extraction performed according to the manufacturer's protocol. Extraction of genomic DNA from peripheral blood or bone marrow samples was performed using the DNeasy kit (Qiagen). PCR and Sanger sequencing of the entire coding region of *RUNX1*, *CEBPA*, *GATA2*, *TERC* and *ETV6* were performed for at least one affected member from each family.

5.3.2 Array CGH

Array CGH employs a whole genome oligonucleotide hybridisation technique to identify chromosomal losses and gains, as described in Chapter 2. In 10 pedigrees (L, M, P, R, S, T, V, W, X and Z) which were wild type for *RUNX1*, *CEBPA, GATA2, ETV6* and *TERC* using Sanger sequencing, a customised 2x400k oligo-array (Agilent) was used to identify CNAs in these and other genes associated with familial MDS/AML and IBMF syndromes (Appendix 1).

5.3.3 MLPA

An MLPA assay customised for familial MDS/AML loci (P437-A1 Probemix, MRC Hollland) was used for verification and detection of novel CNAs. This assay included multiple probes spanning all exons in *RUNX1*, *CEBPA*, *GATA2* and the telomerase genes, *TERC* and *TERT*. Additional probes identified the recurrent germline *GATA2* mutations T354M and R398W). Relative copy number was calculated following intra- and inter-sample normalisation of peak heights. As CNAs were germline in origin, affecting all cells, heterozygous deletions were represented by peak ratios of 0.4-0.6.

5.3.4 SNP array

SNP array profiling was performed to verify MLPA findings in individual N.II.1. This work was performed by Nicola Trim at the West Midlands Regional Genetics Laboratory, Birmingham. The Affymetrix Cytoscan HD array was used, this contains 750,000 unique SNP and 1.9 million oligonucleotide probes, enabling reliable detection of LOH. Data were analysed using the Affymetrix chromosome analysis suite (CHAS) v 2.1.

5.3.5 WES to detect novel germline variants

To detect novel germline mutatons in affected family memebers, WES of genomic DNA samples was performed using the Sure Select Human All Exon v5 library (Agilent). In pedigree P, WES of tumour and remission DNA from P.III.2 was performed, to identify both somatic and germline variants. In pedigrees Q-T, DNA from two affected family members was sequenced to detect shared germline variants potentially segregating with disease. A detailed description of this analysis is provided in Chapter 2.

5.3.5 Families investigated for novel and atypical variants

In total, 17 families were investigated in this chapter, as summarised in Table 2.1. In 15 families (L-Z) where *RUNX1*, *GATA2*, *CEBPA TERC* and *ETV6* mutations were not detected, CNAs were investigated using array CGH and MLPA. Both assays were enriched for probes in all genetic loci associated with familial MDS/AML, including the telomerase complex components (*TERC* and *TERT*). Importantly, since this work predated the identification of germline *ETV6* (Zhang *et al*., 2015; Noetzli *et al*., 2015) and *DDX41* mutations (Polprasert *et al*., 2015), copy number lesions within these genes were not assessed. WES was subsequently performed to identify novel candidate mutations in 5 pedigrees (P-T) where no candidate lesions were detected, these investigations are summarised in Figure 5.1.

Pedigree	Individual	Clinical details of samples tested	MLPA	CGH	SNP	WES
L	L.II.2	Thrombocytopenia (PB)				
	L.II.2	Thrombocytopenia (sDNA)				
M	M.II.5	a)AML (BM)				
	M.II.5	b)CR (PB)				
	M.III.2	a)AML (BM)				
	M.III.2	b)CR post SCT (PB)				
	M.III.3	Thrombocytopenia (PB)				
	M.III.4	Thrombocytopenia (PB)				
N	N.II.1	AML (BM)				
	N.II.2	Healthy (PB)				
O	O.II.1	Thrombocytopenia (PB)				
P	P.III.2	a) AML (BM)				
	P.III.2	b) CR (PB)				
	P.IV.1	Thrombocytopenia (PB)				
	P.IV.2	CMML -FFPE DNA from BM trephine				
Q	Q.I.1	AML relapse (BM)				
Q	Q.II.1	CR (PB)				
R	R.III.2	MDS (BM)				
R	R.III.3	MDS (BM)				
S	S.I.1	Macrocytic anaemia/Thrombocytopenia (PB)				
	S.II.1	Macrocytic anaemia/Thrombocytopenia (PB)				
T	T.II.1	MDS/AML (BM)				
	T.II.2	MDS/AML (BM)				
U	U.II.1	AML (BM)				
	U.II.2	AML (BM)				
V	V.II.1	MDS (BM)				
	V.II.2	MDS (BM)				
W	W.II.1	ALL (BM)				
X	X.II.1	AML (BM)				
Y	Y.II.1	AML (BM)				
Z	Z.II.2	MDS (BM)				

Figure 5.1 Integrated genetic investigation to identify novel germline lesions.

Analysis of sequencing variants and/or CNAs was performed using MLPA, aCGH, SNP array (SNP) and WES across affected family members in pedigrees L-Z. Key: PB, peripheral blood DNA; BM, bone marrow DNA; sDNA, salivary DNA.

.

5.4 Results

Part 1: The identification of atypical lesions within *RUNX1*, *CEBPA* and *GATA2*

5.4.1 Families with germline *RUNX1* deletions

The initial aims of this study were to define atypical lesions within known familial AML genes. As described above, aCGH and MLPA were employed to screen for CNAs within the pre-defined molecular landscape. This investigation identified germline deletions of *RUNX1* across three pedigrees (L, M, N,), and a further case (O.II.1) in whom there was no family history of disease, Figure 5.2.

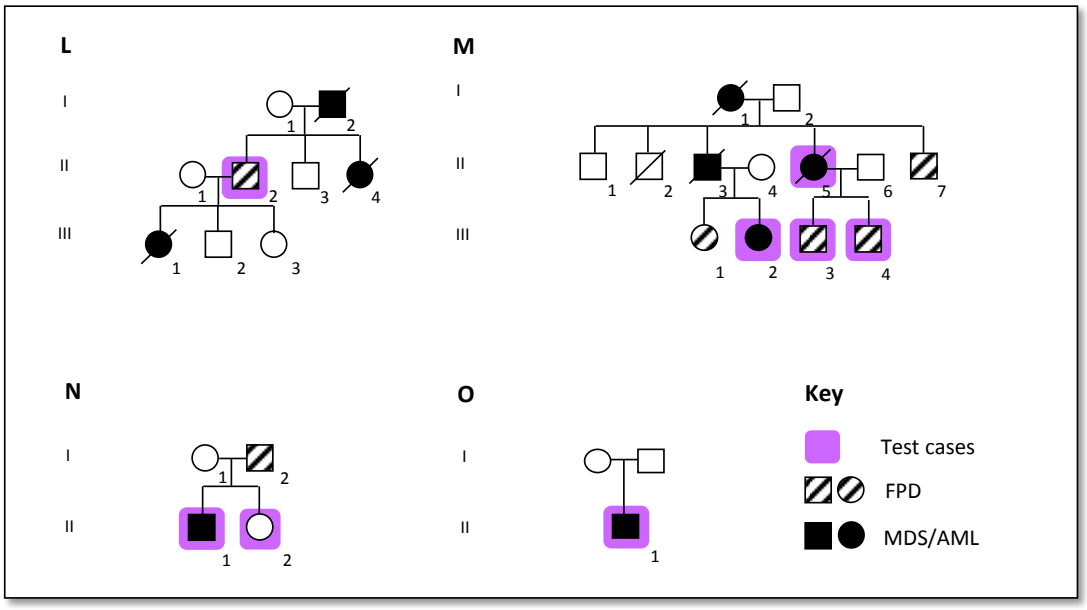


Figure 5.2 Pedigrees and disease phenotypes of cases demonstrating *RUNX1* deletions.

The individuals investigated in each family are highlighted, with disease phenotypes of thrombocytopenia (FPD) and MDS/AML differentiated by shading patterns.

Pedigree L: L.II.2 was investigated in this study; he presented in his early 50's with an asymptomatic mild thrombocytopenia and continues under regular clinical surveillance, with no evidence of transformation to MDS/AML after 5 years. His father (L.I.2) and sister (L.II.4) had both died from AML at 67 and 7 years, respectively. His eldest daughter (L.III.1) had died from AML at the age of 15 years, his remaining son (L.III.2) and daughter (L.III.3) are well with normal PB counts.

Pedigree M: Three generations of this family were affected with thrombocytopenia and/or MDS/AML. M.III.2, initially presented at 8 years with a moderate thrombocytopenia (platelet count 30-40), which subsequently evolved to NK-AML (FAB subtype M2) at 14 years. Molecular testing of leukaemic blasts revealed WT *FLT3* and *NPM1* status. CR was achieved post chemotherapy and was consolidated with allogeneic HSCT, but the patient later died from pneumonia associated with immunosuppressive therapy for GvHD.

The paternal aunt of M.III.2 (M.II.5) developed NK-AML (FAB subtype M4) at 38 years, with WT *FLT3* and *NPM1* status. She entered CR post chemotherapy and underwent allogeneic HSCT, but died three months later from microangiopathic haemolytic anaemia (MAHA). Her father (M.II.3) also presented with thrombocytopenia which subsequently evolved to AML, whilst her paternal grandmother (M.I.2) developed CMML, which later transformed to AML. Her sister (M.III.1), two male cousins (M.III.4 and M.III.5) and paternal uncle (M.II.7), remain under clinical follow up with asymptomatic thrombocytopenia.

Pedigree N: Individual N.II.1, the index case of this pedigree, presented with a bleeding tendency at the age of 6 years and was subsequently diagnosed with a platelet storage pool disorder. The child's father (N.I.1) had also been diagnosed with a storage pool disorder and remains under regular review. Notably, neither individual had thrombocytopenia.

At the age of 13 years, N.II.1 developed AML FAB subtype M2 with trisomy 19 detected upon cytogenetic analysis of BM blasts. Immunophenotyping revealed an aberrant T-cell population expressing CD2, CD5, CD7, weak CD4, intracellular CD3 and terminal

deoxynucleotidyl transferase (TdT) in 10% of cells. After failing to respond to ADE chemotherapy (daunorubicin, cytarabine and etoposide), FLAG-IDA (fludarabine, cytarabine, idarubicin and granulocyte-colony stimulating factor, G-CSF) was administered, resulting in morphological remission with 3% BM blasts detected after the first cycle. A second cycle of FLAG-IDA was administered and BM examination subsequently revealed expansion of the abnormal T-cell population, confirming negativity for CD3, CD8 and CD10 with 3% CD34 positivity. A new diagnosis of T-ALL was made and the patient continued to receive ALL-based therapy at the last follow up available. The child has one sibling who is asymptomatic (N.II.2) and was screened as a potential HSCT donor following this study.

Pedigree O: The index case in this pedigree (O.II.1) is 3 years and has no family history of disease. The child presented with global developmental delay, hypospadias and feeding problems and remains under clinical follow up with no evidence of cytopenias or dysplasia.

5.4.2 Detection of *RUNX1* deletions using array CGH

Array CGH analysis of PB and salivary DNA from L.II.2 revealed a 666kb deletion targeting the N-terminal of *RUNX1 and* encompassing exons 1 and 2. The co-ordinates from aCGH corresponded to chr21: 36389457- 37055677 and were later verified with PCR as chr21:36389492-3705605 (Figure 5.3 A and B). A compound duplication and deletion was identified in remission (germline) BM DNA from patient M.II.5, with the deleted region corresponding to chr21:36349450-36572837, also involving exons 1 and 2 (Figure 5.3 C). The duplication affected the non-coding region upstream of the 5'UTR, with the genomic loci chr21: 36594603- 36768032.

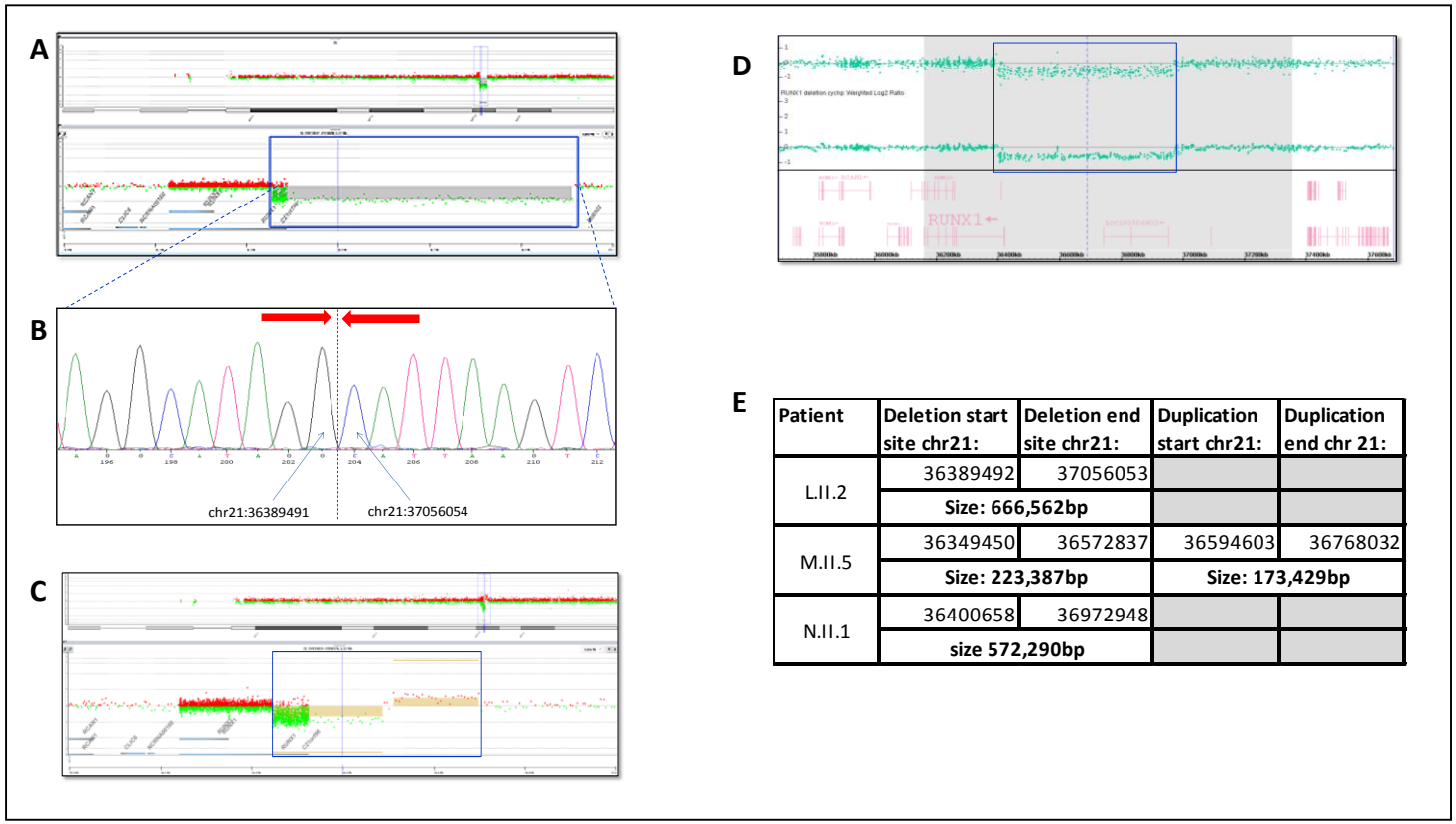
5.4.2.1 Verification of chromosomal breakpoints using nested PCR and Sanger sequencing

To verify the co-ordinates of the deleted genomic region in patient L.II.2, primers were designed to amplify a region spanning >1kb around the deletion breakpoints predicted by aCGH, (forwards sequence 5'CAACCCACACATAACAGCCC and reverse sequence 5' AACAGGCTCTTCAGTAGGCT). PCR was performed using HotStarTaq Plus DNA polymerase master mix (Qiagen) and touchdown thermal cycling parameters with an

annealing temperature gradient of 68-50^{o}C (Table 2.9). Internal primers were then used to sequence across the breakpoint region, confirming the precise co-ordinates of the deletion. The endpoint of the deletion (chr21:36389492) was within 50bp of the breakpoint predicted by aCGH, although the start site (chr21: 37056053) varied by > 300bp, this was likely to be due to the reduced density of probes within this non-coding region (numbers refer to reverse strand). Similar approaches for verification of the deletion in Pedigree M were unsuccessful, with repeated attempts yielding non-specific amplification products. This may have been caused by the complexity of the chromosomal lesion in this family, with the duplicated segment possibly re-annealing in the region of the deletion.

Further confirmatory analysis of the deletions in pedigrees L and M was performed using MLPA. Normalised *RUNX1* peak ratios revealed heterozygous deletions in PB and sDNA from L.II.2 and in multiple affected individuals from Pedigree M (M.II.5, M.III.2, M.III.3 and M.III.4). Novel *RUNX1* deletions were also detected in two additional indivduals, N.II.1 and O.II.1 (Figure 5.3 D-E and 5.4 A-B). Normal copy number status was found in the healthy sibling of N.II.1 (N.II.2). SNP array profiling of DNA from N.II.1 and N.II.2 verified these findings and resolved the genomic co-ordinates of the deletion, data courtesy of Nicola Trim at the West Midlands Regional Genetics Laboratory (WMRGL, Figure 5.3 D). As seen in Figure 5.3 E, the *RUNX1* deletions in Pedigrees L, M and N ranged from 311Kb-666Kb, although O.II.1 harboured a much larger deletion of chromosome 21q spanning 2.78Mb (information provided by referring clinician).

Comparison of breakpoints in families L, M and N revealed a common deleted segment spanning ~100kb (chr21: 36400658-36572837), this encompassed the distal (P1) promoter and first two exons. Although the precise genomic co-ordinates of the deletion in O.II.1 were not known, MLPA confirmed heterozygous loss of exons 1 and 2. Collectively, these data revealed a 'hotspot' for constitutional deletion, which predicted loss of the *RUNX1c* isoform (expressed from the P1 promoter and specifically containing exons 1 and 2). Since the proximal (P2) promoter remained intact, it was possible that mutant alleles may cause pathogenicity either by selective expression of *RUNX1b* (potentially causing a dominant negative effect) or haploinsufficiency.



E

Patient	Deletion start site chr21:	Deletion end site chr21:	Duplication start chr21:	Duplication end chr 21:
L.II.2	36389492	37056053		
	Size: 666,562bp			
M.II.5	36349450	36572837	36594603	36768032
	Size: 223,387bp		**Size: 173,429bp**	
N.II.1	36400658	36972948		
	size 572,290bp			

Figure 5.3 *RUNX1* deletions detected in pedigrees L,M and N.

Array CGH showing germline *RUNX1* deletion in Pedigree L (A) with Sanger sequencing identifying the precise breakpoint of the deletion (B). A compound germline deletion and duplication was detected in pedigree M using aCGH (B) and in pedigree N, germline *RUNX1* deletion was detected using SNP array (D). The predicted hg19 co-ordinates of these lesions are shown in (E).

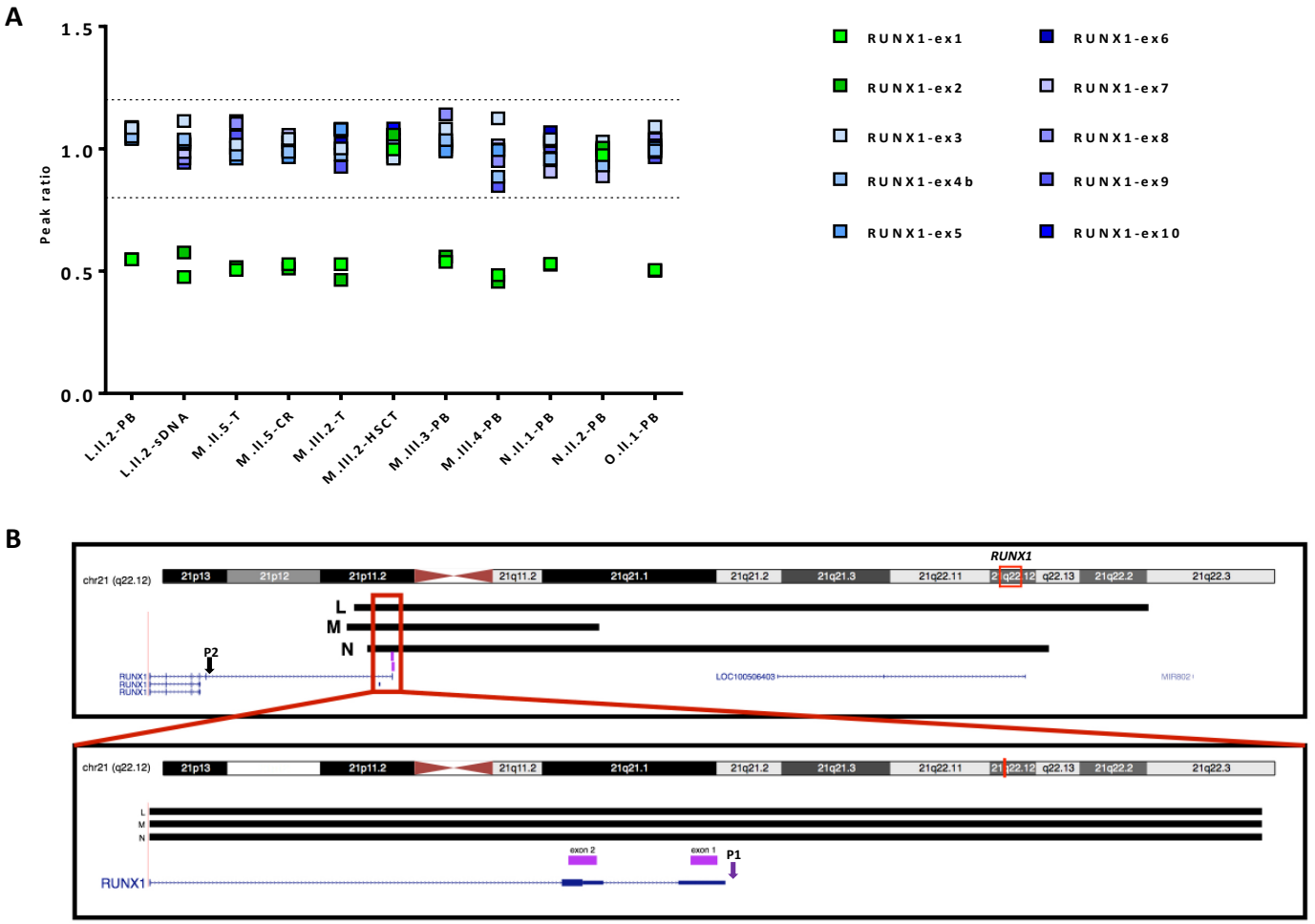


Figure 5.4 Verification of *RUNX1* deletions and alignment to reference genome

(A) MLPA analysis of individuals from pedigrees L-O showing heterozygous loss of *RUNX1* exons 1 and 2 (peak ratios=0.46-0.55), with normal peak ratios in remaining exons. (B) Alignment of deletions (black bar) to *RUNX1* hg19 co-ordinates, the common deleted region across pedigrees L, M and N is enlarged to show loss of the P1 promoter and exons 1-2. Key: PB, peripheral blood; sDNA, salivary DNA; Tm, tumour; CR, complete remission; post BMT, post allogeneic HSCT.

5.4.3 Clinical phenotypes associated with *RUNX1* deletions

To date, less than 5 FPD/AML families harbouring germline deletions have been described in the literature (Song *et al*., 1999; Jongmans *et al*., 2010; Buijs *et al*., 2010), with lesions varying from complete allelic loss to single exon deletions. Germline microdeletions of chromosome 21q22.11 (encompassing the *RUNX1* locus) have also been reported in <20 individuals. These larger germline deletions typically occur *de novo*, with no significant family history of haematological disease or thrombocytopenia. The phenotypic manifestations are heterogeneous, but frequently involve cranio-facial dysmorphia, developmental delay and urogenital abnormalities. Interestingly, MDS/AML is not a common feature within these patients, although mild platelet defects have been described.

The identification of *RUNX1* deletions across 4 families clearly emphasises the need for further characterisation of these lesions and inclusion of CNA assessment within current diagnostic algorithms for familial leukaemia. The intra- and inter-familial phenotypes varied significantly, including familial platelet disorder (with and without thrombocytopenia), MDS, CMML, AML, and T-ALL. Since the deletions universally targeted exons 1 and 2, it seems likely that the differential acquisition of secondary genetic lesions may influence disease phenotype, leading to the clinical heterogeneity observed. While this represents an important avenue for future work, first and foremost these results highlight the need for investigation of germline deletions within all families with thrombocytopenia and haematological malignancy.

5.4.4 Atypical germline mutations in *GATA2* and *CEBPA*

Our current knowledge of germline mutations is based on a limited number of pedigrees and the majority of lesions demonstrate a distinctive localisation, suggesting discrete hotspots within genes. With so few families described, these findings may easily lead to misinterpretation of data, inferring that only specific hotspots are associated with leukaemia predisposing mutations. It is therefore essential to ensure that all rare or atypical mutations are reported for a more comprehensive and balanced representation of germline lesions. Two atypical mutations in *GATA2* and *CEBPA* are described below, each associated with highly variable clinical manifestations and penetrance.

5.4.4.1 A novel compound *GATA2* mutation with variable clinical manifestations

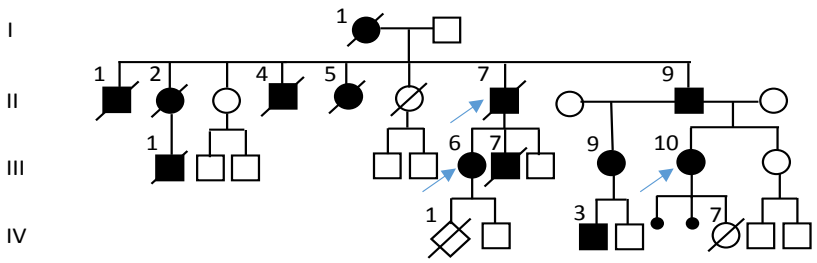
In 2011, Hahn *et al* described 5 pedigrees harbouring germline mutations within the transcription factor *GATA2*. Notably, all families had a uniform clinical phenotype of MDS/AML and mutations clustered within the second zinc finger domain (ZF2). This region is required for DNA binding and interaction with transcription factors such as PU.1 and appears to represent a hotspot for inherited lesions. Subsequent reports described variable disease phenotypes including MonoMAC (Hsu *et al*., 2011; Dickinson *et al*., 2011) and Emberger syndrome (Ostergaard *et al*., 2011), with germline mutations distributed throughout the gene, although the propensity for ZF2 mutations remained.

In 2012, we detected a novel, atypical germline *GATA2* mutation in 3 affected members (III.6, III.10 and II.7) of a family with multiple clinical manifestations, Figure 5.5 A and B (Mutsaers *et al*., 2013). The index patient (III.6) presented with features of MonoMAC syndrome (monocytopenia, NK-and B-cell cytopenias with recurrent bacterial sinopulmonary infections, mucocutaneous candidiasis and human papillomavirus (HPV) infection causing multiple plantar warts. She later developed multilineage MDS (RCMD), the treatment of which was limited by co-morbidity, including progressive deterioration in lung function. Her father, (II.7), presented with MDS and monosomy 7 but died from an EBV-related peripheral T-cell non-Hodgkin lymphoma of the nasopharynx, causing a rapidly progressive pancytopenia of unknown cause. Her cousin (III.10) presented with congenital lymphoedema, cervical dysplasia, followed by MDS with monosomy 7. Additional affected individuals in this family were the paternal grandmother (I.2) who died at 74 years with AML. The paternal aunt (II.2) suffered disseminated MAC infection in her mid-thirties and died at 42 years from Langerhans cell histiocytosis (LCH). A paternal uncle of the index case (II.1) was diagnosed with Emberger syndrome and both her cousin (III.1) and brother (III.7) died from AML at 21 and 14 years, respectively.

Sanger sequencing of PB and sDNA from III.6, III.10 and II.7 revealed a shared, unique compound variant in *GATA2*, comprising two novel mutations located within the infrequently mutated N-terminal (exon 2). The component germline mutations were a missense substitution p.H26P (A409C) and a single nucleotide deletion causing the frameshift p.G28AfsX52 (415delG), both lesions were predicted to have a damaging effect on protein structure using the MutationTaster algorithm (p=1.000 to 4dp for both).

Cloning of exon 2 amplicons (Original TA-kit, Invitrogen) followed by Sanger sequencing of transformed colonies revealed that both mutations were mono-allelic (or *in-cis*), as shown in Figure 5.5 C. This mono-allelic combination of mutations is predicted to cause haploinsufficiency of *GATA2*, due to NMD of the truncated transcript.

A



B

Individual	Clinical Phenotype
I.2	AML, died at 74 years.
II.1	Emberger syndrome with multiple verrucae and primary lymphoedema of lower limbs, died at 58 years.
II.2	Presented with leukaemia 35 years. History of recurrent bacterial infections and leucopenia since childhood. Disseminated *Mycobacterium avium complex* infection and died at 42 years from histiocytosis X.
II.4	History of pulmonary disease. Died from leukaemia at 13 years
II.5	History of pulmonary disease. Died from leukaemia at 7 years.
II.7	MDS with monosomy 7. EBV related T-cell NHL with progressive pancytopenia, died at 56 years.
III.1	Died from AML at 21 years.
III.6	Recurrent sinopulmonary infections, verrucae, bacterial pneumonia with progressive fibrosis. Nodular pulmonary infiltrates. MDS
III.7	MDS and monosomy 7 at 4 years. Multiple palmar warts. Interstitial pulmonary haemorrhage. AML at 9 years, sibling HSCT, died following relapse at 14 years.
III.9	Pulmonary alveolar proteinosis.
III.10	Congenital lymphoedema, cervical dysplasia, perianal fistulae, abscesses and warts, recurrent erysipelas and MDS with monosomy 7. Two miscarried pregnancies.
IV.1	Intra-uterine death at 24 weeks
IV.3	Emberger syndrome with multiple warts and immunodeficiency
IV.7	Born premature and died 6 days post partum

C

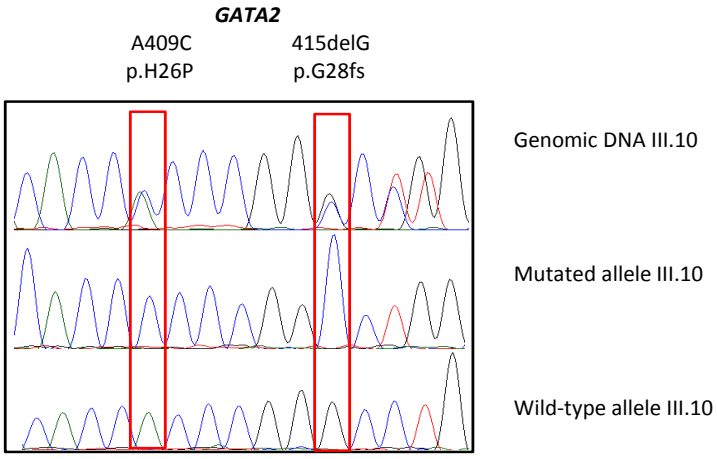


Figure 5.5 Compound germline *GATA2* mutation in a family with varying clinical phenotypes

(A) Pedigree of family showing affected individuals (B) Clinical presentation and phenotypes of disease across family members (C) Mono-allelic distribution of the compound germline mutation identified in individuals II.7, III.6 and III.10.

5.4.4.2 A novel C-terminal *CEBPA* mutation with variable penetrance

Chapter 3 of this thesis describes the characterisation of germline *CEBPA* mutations within 10 families with a clinical history of AML. All germline mutations within this cohort were frameshift lesions within the N-terminal, causing translation of the shorter p30 isoform. It is generally considered that the penetrance of germline *CEBPA* mutations approaches 100%, although our cohort identified 3 adult asymptomatic carriers suggesting that penetrance may be more variable than once considered.

Following completion of the study described in Chapter 3, we recently identified a family with a germline C-terminal *CEBPA* mutation (Figure 5.6 A). The index case of this pedigree was a child who developed NK-AML at 5 years. Diagnostic evaluation of tumour DNA identified two frameshift *CEBPA* mutations, p.I62AfsX95 (330-339del) and p.Q330RfsX74 and (1138dupGAAC), located within the N- and C-terminal domains respectively. Testing of remission DNA revealed persistence of the p.Q330RfsX74 mutation, suggesting it was of germline origin and the first C-terminal frameshift mutation associated with familial AML (Figure 5.6 B). Further testing of the child's mother (III.2, 38 years) and maternal grandfather (II.2, 68 years) confirmed that both were asymptomatic carriers. Individual I.4, the paternal uncle of II.2, developed 'acute leukaemia' at 53 years, suggesting that he too carried the mutation and that his brother, I.5 (the father of II.2) was an obligate carrier, who died from ischaemic heart disease at 75 years. Conceivably, 60% of mutation carriers in this family (both obligate and confirmed, n=3) had no history of haematological malignancy. This highlights the marked variation in penetrance associated with this particular mutation which, to date, has spared 2 generations from disease.

The clinical and molecular findings in this family emphasise the need for comprehensive testing of the entire *CEBPA* coding region at diagnosis and remission in apparent sporadic cases of *CEBPA*-mutated AML (particularly those with double mutations). While most cases of familial leukaemia present below the age of 50 years, the late onset of disease in I.4 (53 years), warrants consideration of a higher age threshold for germline mutation screening. Furthermore, these findings raise important questions regarding the penetrance of germline mutations and the molecular basis for asymptomatic carriage. This is a new area of research now being explored by my host group.

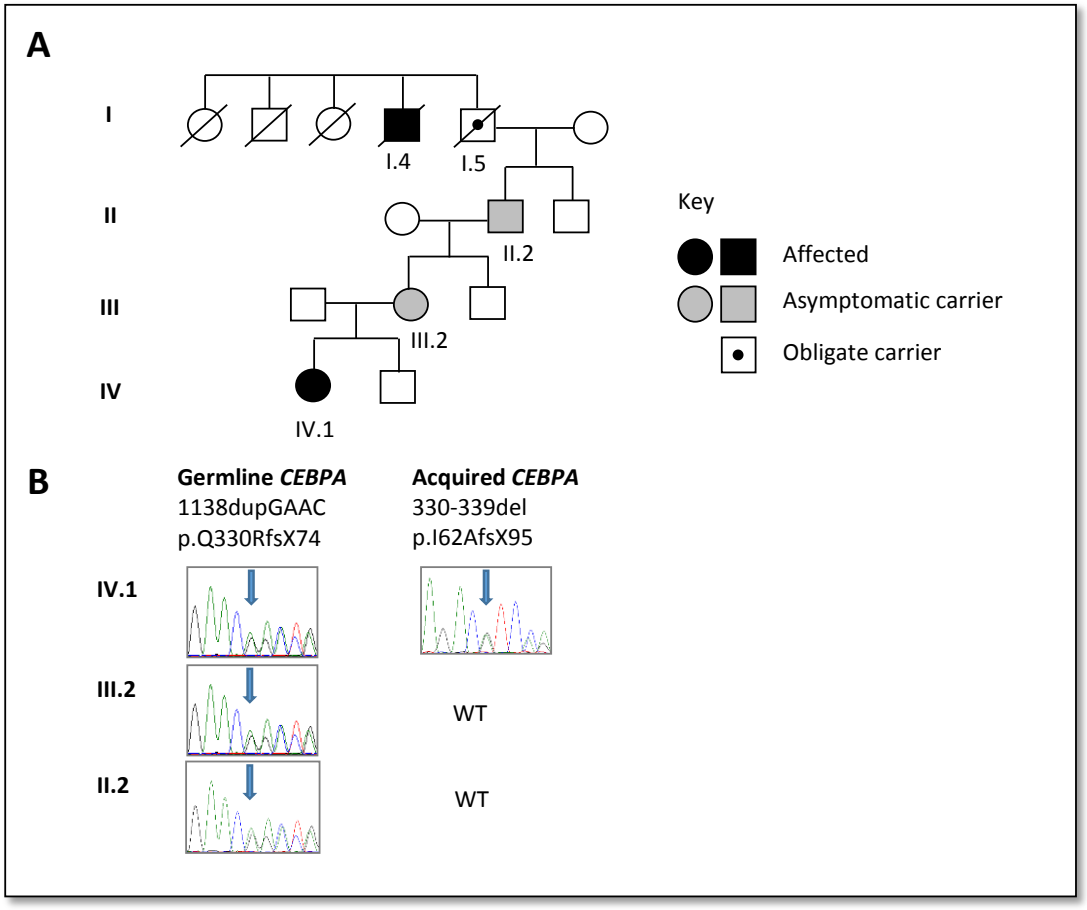


Figure 5.6 A germline C-terminal *CEBPA* mutation associated with variable penetrance.

(A) Pedigree of a novel AML pedigree with a germline C-terminal *CEBPA* mutation. Two asymptomatic carriers were confirmed and I.5 represented an obligate carrier. B) The germline p.Q330fs mutation was detected in indivduals IV.1, III.2 and II.2, with testing of tumour DNA in IV.1 revealing an acquired frameshift mutation within the N-terminal of the gene, preceding the internal start codon.

Part 2: The investigation of novel germline mutations using WES

5.4.5 Families investigated for novel germline mutations

The above examples illustrate the occurrence of atypical germline lesions in known familial MDS/AML genes. Throughout the course of this study, we also obtained material from eleven pedigrees where no recognised mutation was detected and the second part of this chapter describes preliminary studies that aimed to identify novel susceptibility loci using WES. Analysis was restricted to 5 families (P-T) as shown in Figure 5.7 and detailed below. Two or more samples were available from affected individuals within each family and the clinical history was highly suggestive of an inherited predisposition.

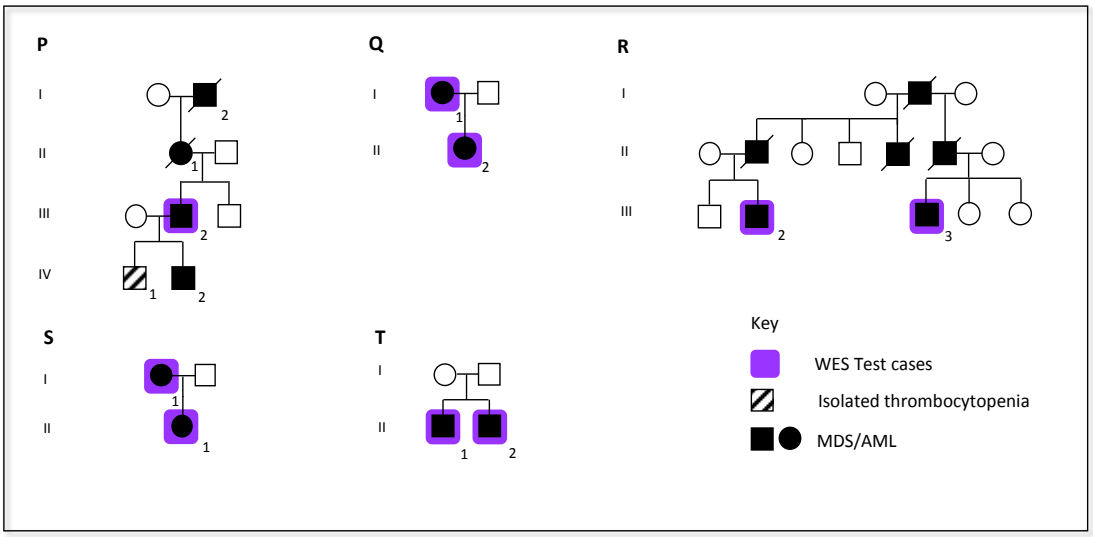


Figure 5.7 Families investigated for novel germline mutations using WES.

The figure shows the pedigrees and disease phenotypes identified across affected individuals. WES was performed using DNA samples from the highlighted individuals in each pedigree.

Pedigree P

A highly penetrant MDS/AML phenotype was observed in 5 individuals across four generations within this family. The pedigree is shown in Figure 5.7 and the index case was a male of 42 years (P.III.2) who presented with high-risk MDS (RAEB2, 16% BM blasts and virtually complete loss of megakaryopoiesis). He received standard AML induction chemotherapy, followed by an unrelated allogeneic VUD-HSCT and is currently in remission. Karyotypic analysis of BM was normal with wild-type *FLT3* and *NPM1* status. The patient's mother (P.II.1), was diagnosed with AML M2 at 37 years and achieved a 4 year remission following chemotherapy, she subsequently presented with MDS and died after 2 years. His maternal grandfather (P.I.2) died of acute leukemia in 1968 at the age of 50 years. The patient has two sons, the younger (P.IV.2) developed CMML (MDS) at 8 years and underwent VUD-HSCT. The older son (P.IV.1) presented with an isolated thrombocytopenia at 18 years and continues to be monitored for evidence of disease progression; there is no history to suggest platelet dysfunction or bleeding tendency in this individual.

Pedigree Q

The two affected cases in this family were mother and daughter, both presented with AML. The mother (Q.I.1) was diagnosed in 2010 at 53 years and received intensive chemotherapy. She failed to achieve a durable complete remission (CR1 8 months) and died from relapsed disease. Karyotypic analysis was suboptimal in both the diagnostic and relapse BM samples due to insufficient metaphases. In 2013, her daughter (Q.II.1) developed AML (FAB subtype M2) associated with a complex karyotype, including abnormalities of chromosomes 5 and 7. She was treated with intensive chemotherapy and underwent VUD-allogeneic HSCT in January 2014, further outcome data are unknown for this patient.

Pedigree R

This represented an extensive pedigree with multiple individuals affected with MDS and AML, often associated with monosomy 7. Two cousins (R.III.2 and R.III.3) presented with MDS and were investigated for shared genetic lesions in tumour DNA using WES.

Pedigree S

A mother and daughter presented with macrocytic anaemia and thrombocytopenia, associated with a normal karyotype. Both continue under clinical observation without requiring any treatment to date.

Pedigree T

Two brothers were affected in this family; the older sibling (T.II.1) presented with AML at 16 years and his younger sibling (T.II.2) was diagnosed with MDS at 11 months of age. Notably, both tumours had monosomy 7 suggesting this may be an important somatic co-operating event within this family.

5.4.5.1 WES metrics for Pedigrees P-T

Families P-T were investigated with, MLPA, aCGH (or both) to exclude CNAs in known leukaemia predisposition genes prior to WES (as summarised in Figure 5.1). WES was then performed using the SureSelect Human All Exon v5 library (Agilent). DNA from two affected individuals was sequenced in pedigrees Q-T to detect shared germline variants, whilst in pedigree P, tumour and remission DNA from the index case (P.III.2) were sequenced. The WES metrics for all samples are summarised in Table 5.1. An exhaustive validation was not completed for these data, as our informatics pipeline is now well established and credible (Okosun *et al*., 2013; Tawana *et al*., 2015). In total, 26 WES variants were verified with Sanger sequencing, achieving 94% concordance and confirming the data authenticity (Appendix 4, Table 2).

Table 5.1 WES metrics for samples from pedigrees P-T

Sample	Tissue source	Sequenced nucleotides (Gb)	Mean exome coverage	% bases ≥ 10x coverage
P.III.2	sAML-BM	3.79	74.93	96.3
P.III.2	CR1- PB	3.97	78.5	95.7
Q.I.1	AML relapse-BM	3.18	62.81	97.1
Q.II.2	CR1-PB	2.14	42.37	93.6
R.III.2	tumour-BM	4.67	92.25	98.8
R.III.3	tumour-BM	5.25	103.71	99.1
S.I.1	tumour-PB	4.89	96.56	98.8
S.II.1	tumour-PB	4.27	84.35	98.6
T.II.1	tumour-BM	2.46	48.57	91.9
T.II.2	tumour-BM	4.93	97.48	96.8

5.4.5.2 WES variant calling and data filtering

Germline variant calling was performed using GATK, all non-synonymous (protein altering) were selected and unique variants were distinguished by comparison with dbSNP137, TGP and HAPMAP projects. From the resulting novel, non-synonymous germline variants, functional annotation was performed to prioritise genes of pathogenic relevance. As discussed previously in Chapter 1, critical pathways disrupted in MDS and AML include epigenetic modifiers, RNA splicing machinery, transcription factors, DNA repair and signalling pathways, all of which provide a guide for the ranking and prioritisation of candidate genes.

Further filtering was performed using the variant prediction algorithm MutationTaster, followed by updated cross referencing with the ExAC browser (Broad Institute) to determine the allelic frequency of rare variants, as described in Chapter 2. To determine whether candidate genes were recurrently mutated in haematological malignancies, the COSMIC database, containing information from over 1900 myeloid and lymphoid tumours samples, was interrogated. These data were complemented by Pubmed and

OMIM (www.omim.org) searches, to identify additional disease associations or relevant information.

A further consideration regarding the pathogenicity of rare variants is their associated penetrance and allelic frequency within a 'healthy' population, for example, pathogenic variants with relatively low penetrance are likely to be associated with a higher frequency of asymptomatic carriage. The families investigated in this study appear to demonstrate an autosomal dominant pattern of disease inheritance, with no history of consanguinity. As with other leukaemia predisposition syndromes, variation in penetrance is likely to occur with novel mutations, but is difficult to assess accurately given the limited screening of family members and absence of comprehensive clinical histories.

The ExAC browser combines data from large sequencing consortia including the NHLBI and TGP. Since individuals with severe paediatric disease are excluded from this dataset, it may serve as a useful reference to determine MAFs within a control population. Importantly, it should be noted that rare pathogenic variants may still persist within this extensive database, due to a minority of asymptomatic carriers. As the number of genomes and exomes increases rapidly, universal exclusion of all rare variants may lead to false negative results. This is particularly relevant for autosomal recessive disorders (all of which are associated with asymptomatic heterozygous carriage) but may also apply to autosomal dominant alleles with variable penetrance. In general, rare variants with a MAF <1% (<0.01) should be considered for recessive disorders and a more strict cutoff of <0.1% (0.001) is useful for dominant disorders (Bamshad *et al*., 2011).

5.4.6 Molecular variants prioritised across Pedigrees P-T

WES of this preliminary cohort of individuals from 5 families revealed significant variation in the number of novel non-synonymous germline variants. As expected, the highest number of variants was detected with comparison of matched tumour and remission DNA for individual P.III.2 (n=86), followed by Pedigrees Q (mother and daughter, n=64); S (mother and daughter, n=63) and T (two siblings, n=23). The most distantly related family members harboured the least shared variants in Pedigree R (two cousins, n=10). These variants were subsequently ranked and a limited number (ranging from 2-6 per pedigree) were prioritised following functional annotation. Further evaluation involved MutationTaster prediction analysis (Schwarz *et al*.,2010) and an updated assessment of the allelic frequency in TGP and ExAC databases, as summarised in Table 5.2.

No recurrent genetic lesions were detected across the selected variants for these pedigrees, suggesting that some pedigrees may harbour private mutations or inherited CNAs in novel loci yet to be investigated, such as *ETV6* or *DDX41*. In pedigree P, all candidate variants were further assessed in both affected sons using Sanger sequencing. Tumour DNA was obtained from a paraffin embedded trephine sample from P.IV.2, who presented with CMML at 8 years, and PB DNA from P.IV.1, who presented with thrombocytopenia at 18 years. Of the 6 variants in *GATAD2B, MYST3*, *RASA3, PSMB3*, *UPF1* and *USP16*, only two segregated across all 3 individuals: *MYST3* and *GATAD2B* (Figure 5.8).

Table 5.2 Prioritised germline variants following WES of pedigrees P-T

Family	Variant*	Gene	Function	Amino acid	TGP	EXAC no. of alleles/homozygotes (allelic freq)	MutationTaster	COSMIC (Haematopoietic) **
P	1:153800510 T/G	*GATAD2B*	Transcriptional repressor -component of methyl CpG-binding protein 1 complex (MeCP1) - deacetylase	V105G	Unique	Unique	Disease causing (1.0000)	AML (4)
P	8:41789998 A/G	*MYST3*	Histone acetyltransferase , transcriptional activator of RUNX1 and RUNX2	M1914V	Unique	20/0 (0.0001648)	Disease causing (0.9645)	CLL(1) CML(1) HCL (1) FL (1)
P	13:114774873 C/T	*RASA3*	RAS GTPase-activating protein, inhibition of RAS signalling	A553V	Unique	3/0 (0.00002497)	Disease causing (1.0000)	HCL (1)
P	17:36909577 G/A	*PSMB3*	proteosome complex, DNA damage response, cell cycle arrest	V60I	Unique	Unique	Disease causing (1.0000)	N/A
P	19:18974413 A/G	*UPF1*	RNA helicase essential for nonsense mediated mRNA decay,	I923V	Unique	4/0 (0.00003449)	Disease causing (0.9886)	ALL (1) AML (1) MM (1) DLBL(1)
P	21:30426208 C/G	*USP16*	Deubiquitinates Histone 2A K119, considered to antagnoize self renewal/senescence in Downs's syndrome	T760S	Unique	Unique	Disease causing (1.0000)	MM (2) DLBL (1)
Q	9:742275 G/C	*KANK1*	AKA *ANKRD15*. Actin polymerisation and cell motility. Fusion with PDGFRB reported in MPN	C1256S	Unique	Unique	Disease causing (1.0000)	ALL (1) AML (1) Burkitt (1) lymphoid neoplasm (2)
Q	10:112641121 G/C	*PDCD4*	Tumour suppressor, binds to and inhibits eukaryotic translation initiation factor 4A1 (EIF4A1)	K58N	Unique	Unique	Disease causing (1.0000)	DLBL (1) MCL (1)
Q	11:125495891 G/A	*CHEK1*	Serine/threonine-protein kinase, checkpoint-mediated cell cycle arrest, activation of DNA repair	W79X	Polymorphism	15/0 (0.001411) rs556405813	Disease causing (1.0000) NMD	FL (1)
Q	16:89845230 C/G	*FANCA*	Component of Fanconi complex, DNA repair	A602G	Unique	Unique	novel polymorphism (1.0000)	CLL(2) ALL (1) PRV (1) AML (1) MM (1)
R	12:122968061 G/A	*ZCCHC8*	Component of nuclear exosome targeting complex (NEXT)- exosomal degradation of promoter upstream transcripts	G184E	Unique	Unique	Disease causing (1.0000)	CLL (1)
R	20:37187086 C/T	*RALGAPB*	Mitotic regulation, chromosomal alignment and cell cycle control	T1174M	Unique	8/0 (0.00006595)	Disease causing (1.0000)	CLL (1) NK-TCL (1) MM (2) DLBL (1) CNS-lymphoma (1) lymphoid neoplasm (1)
S	3:142218500 G/T	*ATR*	DNA damage, cell cycle regulation	Q1783H	Unique	1/0 (0.000008237)	Disease causing (0.9998)	HCL(2) LGL (1) DLBL (1) MZL (1) AML (2) MCL (1) FL (2)
S	6:30688326 C/T	*TUBB*	β-Tubulin component -β1 tubulin (*TUBB1*) mutations associated with inherited thrombocytopenia	Q15X	Unique	Unique	Disease causing (1.0000)	N/A
S	12:49420696 G/C	*KMT2D*	Histone methyltransferase, transcription regulation, recurrently mutated in lymphoid tumours. Germline mutations (predominantly truncating) associated with Kabuki syndrome	T5018R	Unique	Unique	Disease causing (0.9909)	300 mutations- majority DLBL and FL. MDS (34)
S	15:43749215 G/C	*TP53BP1*	DNA damage response. Cell cycle arrest and downstream phosphorylation of BRCA1 and CHEK2	P526A	Unique	1/0 (0.000008240)	polymorphism (0.9984)	ALL (2) AML (2) HCL (1)
S	17:37565756 C/A	*MED1*	DNA damage, p53 dep apoptosis	L906F	Unique	Unique	Disease causing (1.0000)	lymphoid neoplasm (2)
T	6:111685100 A/C	*REV3L*	DNA translesion repair- mutations associated with Mobius syndrome	K2279Q	Unique	Unique	Disease causing (0.9990)	ALL (1) CLL (1)
T	16:89967047 A/G	*TCF25*	Basic helix-loop-helix (bHLH) family of transcription factors, embryonic development, apoptosis	H409R	Unique	1/0 (0.000008369)	Disease causing (1.0000)	ALL (1)
T	16:89346343 C/T	*ANKRD11*	Interacts and inhibits transcriptional activation of p160 nuclear receptor coactivators (which co-operate with CBP for cell-cycle control, differentiation and apoptosis)	L2203F	Unique	2/0 (0.00001919)	polymorphism (1.0000)	ALL (4) AML (1) DLBL (1)
T	17:79519292 C/A	*C17ORF70*	Fanconi DNA damage response (Fanconi anemia-associated protein of 100 kDa)	A22E	Unique	Unique	Disease causing (0.7182)	ALL (1) IMF (1)

Key: * Variant details are shown chr: genome number reference/alteration. ** Haematopoietic and lymphoid samples tested in COSMIC (n=1982), CLL (chronic lymphocytic leukaemia); DLBL (diffuse large B-cell Lymphoma); FL (follicular lymphoma); HCL (hairy cell leukaemia); ALL (Acute lymphoblastic leukaemia); MCL (mantle cell lymphoma); LGL (large granular lymphocytic leukaemia); MZL (marginal zone lymphoma); AML (acute myeloid leukaemia); MDS (myelodysplastic syndrome); MM (multiple myeloma); IMF (idiopathic myelofibrosis); PRV (polycythaemia rubra vera); CML (chronic myeloid leukaemia)

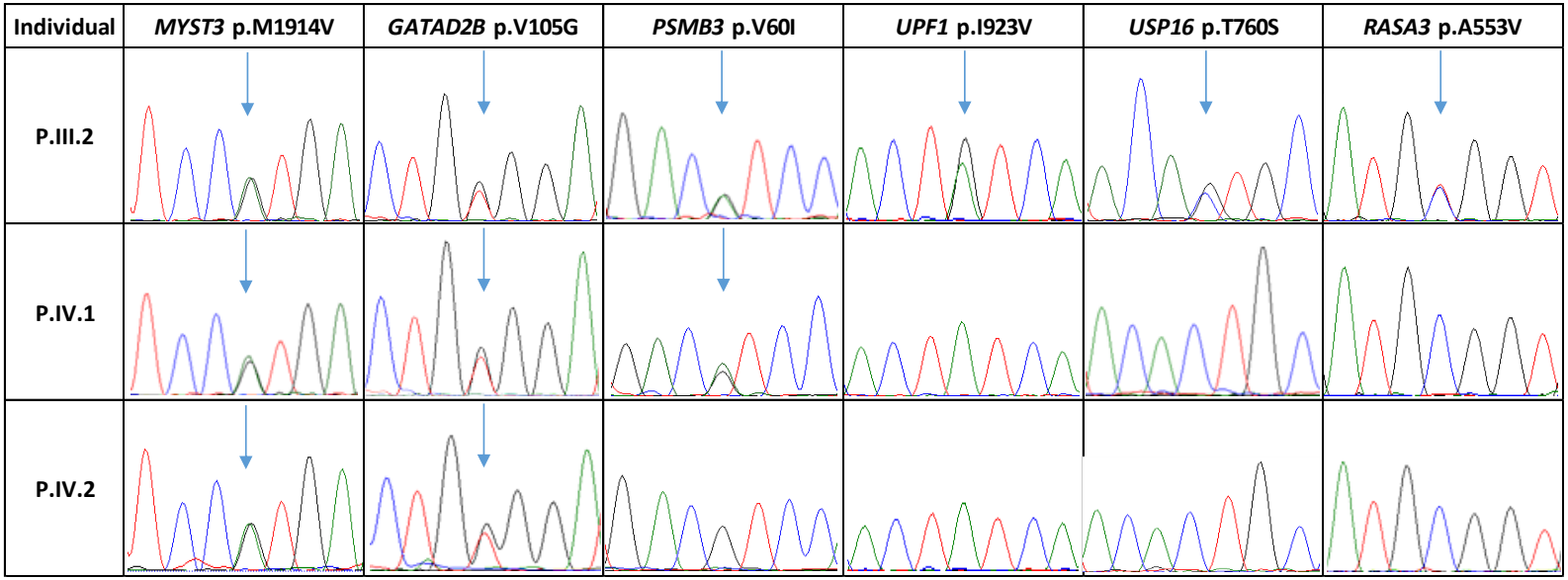


Figure 5.8 Segregation of germline variants across affected individuals in pedigree P

Sanger sequencing of prioritised germline variants across P.III.2 (AML, M2), P.IV.1 (thrombocytopenia) and P.IV.2 (CMML). Variants in the epigenetic modifiers *MYST3* and *GATAD2B* segregated with disease.

5.4.6.1 Candidate germline mutations in Pedigree P

MYST3

MYST3 (also known as *MOZ* or *KAT6A*) is located on chromosome 8p11.2 and belongs to the MYST family of proteins, all of which contain a highly conserved histone acetyltransferase (HAT) domain required for protein and lysine acetylation. It specifically interacts with RUNX1 (Kitabayashi *et al*., 2001a), PU.1 (Katsumoto *et al*., 2006), RUNX2 (Pelletier *et al*., 2002) and NF-kB (Chan *et al*., 2007) as a transcriptional coactivator, and therefore has a key role in haematopoietic development and HSC renewal, with Myst3-deficient mice demonstrating embryonic lethality (reviewed in Katsumoto *et al*., 2008).

MYST3 is recurrently amplified or translocated in sporadic AML, forming fusion genes with other HATs such as *CREBBP* (*CBP*), t(8;16)(p11;p13), (Borrow *et al*., 1996) and *P300*, t(8;22)(p11;q13), (Chaffanet *et al*., 2000; Kitabayashi *et al*., 2001b). Additional translocation partners include the nuclear receptor co-activators *TIF2*, (inv8)(p11;q13), (Carapeti *et al*., 1998; Liang *et al*., 1998) and *NCOA3*, t(8;20)(p11;q13), (Esteyries *et al*., 2008). Sporadic cases of AML harbouring these translocations have a monocytic or monoblastic AML phenotype reflecting the gene's synonym *monocytic leukaemia zinc finger protein* or *MOZ*. The majority of cases harbouring *MYST3* translocations are therapy–related, often displaying erythrophagocytosis with adverse clinical outcomes (Sun *et al*., 2001; Gervais *et al*., 2008).

Five different *MYST3*-*CREBBP* fusion transcripts have been detected to date, with *MYST3* breakpoints all occurring in the C-terminal of the gene (Gervais *et al*., 2008). The M1914V variant in Pedigree P was also situated in the C-terminal (exon 17), near the breakpoint of the type I translocation which juxtaposes exon 17 of *MYST3* with exon 2 of *CREBBP*. The M1914V variant targeted the methionine-rich region, adjacent to the putative *RUNX1* binding domain, and was predicted to be 'disease-causing'. Disruption of RUNX1 signalling may explain the significant reduction of megakaryopoiesis observed in the index patient, as well as the thrombocytopenia in his older son (akin to FPD in *RUNX1*-mutated pedigrees). Data from ExAC suggests that the M1914V variant has been observed in 20 alleles with a MAF of <0.0001, although it was not detected in the healthy control population of the TGP. These data are equivocal, indicating that M1914V may represent either a rare polymorphism or a low frequency pathogenic allele.

There have been no reports of *MYST3* mutations in AML to date, however AML is highly heterogeneous and mutations may be enriched within specific morphological subtypes. Since chromosomal translocations of *MYST3* are predominantly associated with monocytic leukaemia, we postulated that the C-terminal region of *MYST3* may represent a novel mutation target for monocytic subtypes of AML or CMML. PCR and bidirectional sequencing of the C-terminal exons 14-17 were performed for 11 cases of sporadic CMML and 34 cases of sporadic acute monocytic/monoblastic leukaemia (AML M5). In total, 10 PCR amplicons were investigated, with intronic and overlapping exonic primers designed according to the transcript ENST00000265713. No mutations were identified in this cohort, suggesting that the C-terminal is unlikely to be a molecular target in sporadic subtypes of monocytic MDS/AML.

GATAD2B

The second candidate mutation in this pedigree was the V105G substitution in GATA zinc finger domain containing 2B (*GATAD2B*). This gene encodes a subunit of the MeCP1-Mi-2/nucleosome remodelling and deacetylase complex (NuRD) responsible for silencing of methylated DNA by nucleosome remodelling and histone deacetylation (Feng *et al*., 2001; Brackertz *et al*., 2006). The NuRD complex is evolutionarily conserved and comprises a large macromolecule with multiple subunits, including histone deacetylases (HDAC1/2), ATP-dependent remodelling enzymes (CHD3/4), histone chaperones (RbAp46/48), CpG (methyl) binding domain proteins (MBD2/3), histone modifier proteins (MTA1/2/3) in addition to p66α and/or p66β (GATAD2A/GATAD2B) as shown in Figure 5.8. The latter are thought to have a repressive effect on gene expression by potentiating the action of the CpG binding protein, MBD2 (reviewed in Torchy *et al*., 2015).

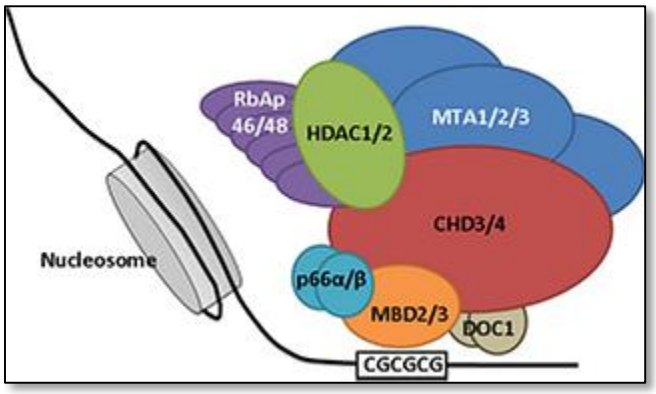


Figure 5.9 Schematic representation of the NuRD complex (Taken from Torchy *et al*., 2015).

The precise structure and subunit interactions are not yet defined but mass spectrometry data have revealed that the complex contains one CHD3/4, one MBD 2/3, three MTA 1/2/3, one HDAC 1/2, six RbAp 46/48, two p66α/β and two DOC1 subunits. Key: CHD; Chromodomain, helicase, DNA binding domain; MBD- methyl-binding domain; HDAC- histone deacetylase.

GATAD2B contains two highly conserved regions (ConsR1 spanning amino acids 165-190, and ConsR2 amino acids 340-480) which are essential for NuRD complex formation and mediating gene silencing (Brackertz *et al*., 2006). Loss of function, frameshift and nonsense mutations have been associated with a paediatric syndrome comprising intellectual disability, hypotonia, hypertelorism, strabismus and a characteristic 'tubular' nose with broad shaped tip (Willemsen *et al*., 2013). Notably somatic mutations within *GATAD2B* have been reported in 4 cases of AML (COSMIC), all of which harboured an identical nonsense mutation (p.R179X) within ConsR1. This mutation was also detected within a case of PRV with homozygous *JAK2* V617F mutation, suggesting disruption of histone deacetylation may be important for the pathogenesis of MPNs in addition to AML. The mutation in Pedigree P encoded a missense substitution p.V105G within the N-terminal of *GATAD2B* and has not previously been reported in the COSMIC, ExAC or TGP databases. It was predicted to be 'disease-causing' using MutationTaster analysis, although the precise functional consequences are not known.

Collectively these data suggest that *MYST3* and *GATAD2B* are potential MDS/AML susceptibility genes within pedigree P. Since both genes are epigenetic modifiers, these

findings, if verified in additional families, will introduce a novel genetic subgroup associated with MDS/AML predisposition.

5.4.6.2 Acquired genetic lesions in P.III.2 and P.IV.2

To further characterise the molecular profile of MDS/AML in Pedigree P, somatic mutations were identified by WES of tumour and remission DNA in P.III.2. In total, 21 non-synonymous somatic variants were detected in this patient, representing 19 SNVs and 2 indels. Sanger sequencing was used to verify selected variants (Appendix 4, Table 2) achieving concordance of 89% (8 out of 9 variants verified). The salient somatic mutations in this patient included a novel clonal splice site mutation in intron 4 of the chromatin adaptor protein plant homeodomainfinger protein 6 (*PHF6,* p.V126-3A>G, VAF 37%). This gene is recurrently mutated in both T-ALL (Van Vlierberghe *et al*., 2010) and adult AML (Van Vlierberghe *et al*., 2011) and, notably, PHF6 directly interacts with the NuRD chromatin remodelling complex to regulate transcription. The latter demonstrates convergence with the inherited *GATAD2B* mutation in this family, signifying germline and somatic disruption of NuRD function in the tumour sample of P.III.2. Additional somatic lesions included a nonsense mutation in the DNA hydroxymethylase *TET2* (p.R1486X, VAF 19%) and a missense mutation in the GTP-ase signalling gene, *KRAS* (p.Q61R, VAF 14%). WES and MLPA (X060-MRC Holland) failed to identify somatic CNAs in tumour DNA from P.III.2, confirming the key pathogenic lesions were those described above.

Unfortunately, we were unable to perform WES on CMML DNA from the younger son, P.IV.2. Instead, somatic CNAs were assessed using MLPA (X060 probemix- MRC Holland); the assay also contained a probe for detection of the recurrent *DNMT3A* p.R882H mutation, found in approximately 10% of CMML (Jankowska *et al*., 2011). A positive signal was detected from the *DNMT3A* probe and the pR882H mutation was subsequently verified with Sanger sequencing. MLPA also revealed heterozygous deletion of *ASXL1* in tumour DNA from P.IV.2, with reduced peak ratios in exons 1 and 8 (0.61 and 0.62, respectively; normal copy number peak ratio 0.8-1.2). This further emphasised the convergence of somatic and germline lesions upon epigenetic dysregulation. The somatic lesions in pedigree P are summarised in Figure 5.10.

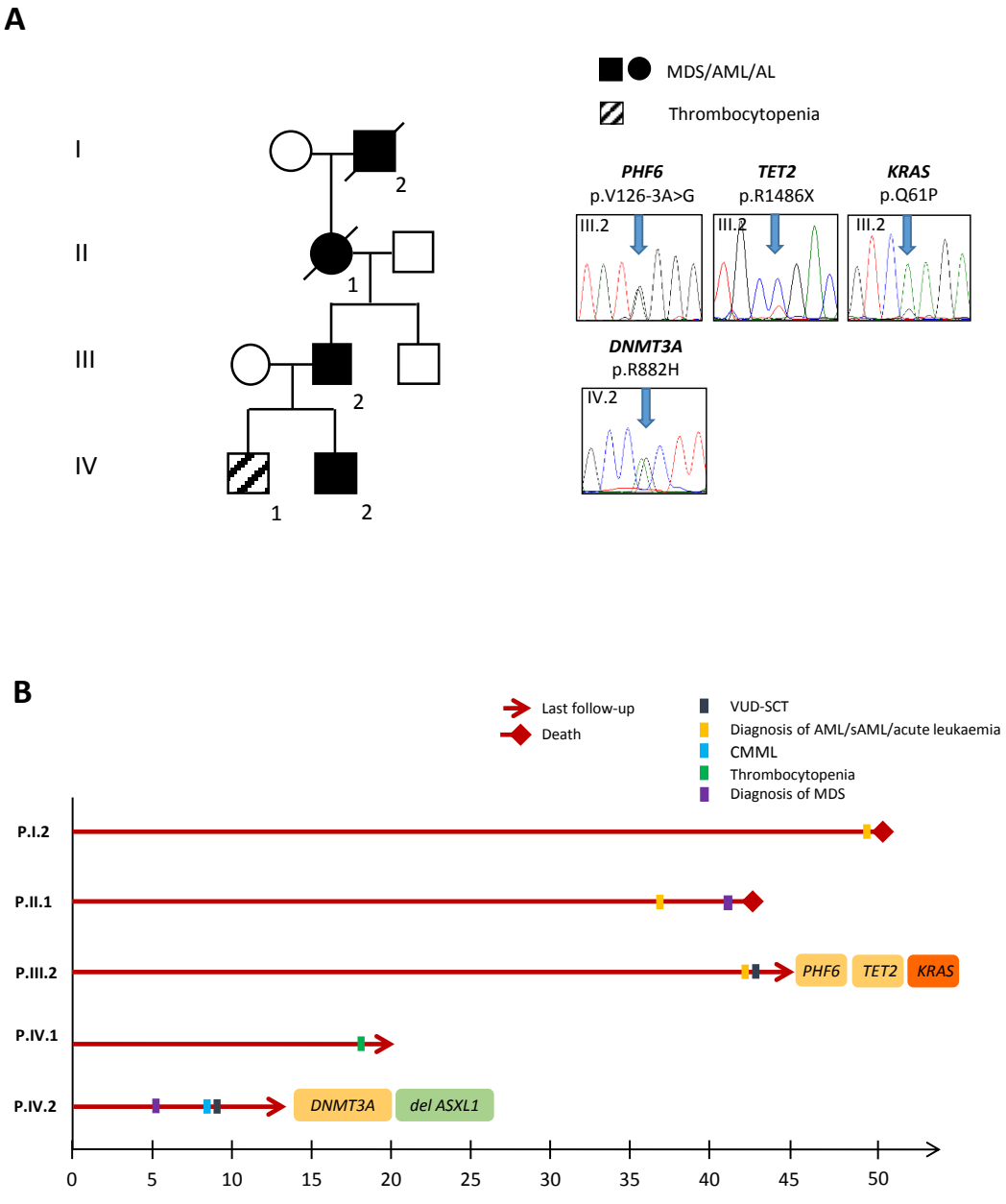


Figure 5.10 Somatic mutations and timeline of disease events in Pedigree P.

Five affected family members were identified across four generations. Somatic mutations detected in P.III.2 and P.IV.2 are shown, with several epigenetic mutations in *PHF6*, *TET2* and *DNMT3A*. (B) Summary of clinical timelines and genetic lesions acquired in P.III.2 and P.IV.2.

5.4.6.3 Candidate germline variants detected in Pedigrees Q-T

In Pedigrees Q-T, DNA from two affected family members was exome sequenced to identify the shared variants across both. The prioritised variants for each family are summarised in Table 5.2.

Pedigree Q

Following filtering and prioritisation of variants in Pedigree Q, several candidates emerged: *KANK1* (p.C1256S), *PDCD4* (p.K58N), *CHEK1* (p.W79X) and *FANCA* (p.A602G), although analysis with MutationTaster and ExAC revealed that variants in *FANCA* and *CHEK1* were likely to represent polymorphisms.

Variants in *KANK1 and PDCD4* remain potential candidate genes although little is known regarding these genes in the context of myeloid neoplasia. *KANK1, a*lso known as *ANKRD15*, is part of a family of proteins that regulate actin polymerisation and cell motility and was first described as a putative tumour suppressor in renal cell carcinoma (Sarkar *et al*., 2002; Roy *et al*., 2005). Reduced expression of *KANK1* has been described in myeloproliferative disorders (Kralovics *et al*., 2005) and more recently, translocation with *PDGFRA*, was detected in a myeloproliferative disorder with marked thrombocytosis (Medves *et al*., 2010). PDCD4 is a putative tumour suppressor, which inhibits translation through interaction with translation initiation factor eIF4A (a member of the DEAD-box protein family which functions as an RNA helicase) (Loh *et al*., 2009). Expression is down regulated in multiple solid tumours and in AML blasts (particularly in *NPM1*-mutated cases), due to upregulation of the microRNA miR-21, which directly suppresses *PDCD4* (Riccioni *et al*., 2015). As mutations within both genes are rare, limited conclusions can be drawn for both *KANK1* and *PDCD4* germline variants and further functional studies are required to elicit their pathogenicity.

Pedigree R

In pedigree R, 9 novel non-synonymous variants were shared between two first cousins with MDS. Following completion of our filtering pipeline, two variants ranked highly, in *ZCCHC8* and *RALGAPB*. *ZCCHC8* encodes a subunit of the recently characterised human nuclear exosome targeting complex (NEXT), which is required for the degradation of promoter upstream transcripts (Lubas *et al*., 2011). RNA processing pathways (including

RNA splicing/processing and nonsense mediated decay) are emerging as recurrent targets of mutation in sporadic and familial MDS/AML. Germline mutations of *DDX41*, the DEAD/H-box RNA helicase gene, were recently detected in familial MDS/AML patients with late onset disease, who subsequently developed somatic *DDX41* mutations in tumour samples, suggesting a tumour suppressor function (Polprasert *et al*., 2015). There have been no reports of *ZCCHC8* mutations in myeloid malignancies to date and the variant observed in this family is unique across both TGP and ExAC databases, suggesting either a private pathogenic mutation or rare polymorphism.

RALGAPB was recently shown to participate in mitotic regulation, demonstrating marked redistribution to the mitotic spindle where it co-localises with tubulin. Within the same study, targeted depletion of RALGAPB with silencing RNAs led to chromosomal misalignment and disruption of the metaphase to anaphase transition in HeLa cells (Personnic *et al*., 2014). Whilst there are no reports of *RALGAPB* mutations in myeloid malignancies to date, reports are emerging of recurrence within solid tumours, most commonly small cell lung cancer, in which 10% of cases are mutated (CBioportal, 2015). The mutation within Pedigree R was not observed within the TGP database but occurred 8 times within the ExAC browser, suggesting it may represent a rare polymorphism although the details of individuals harbouring mutations within this gene are not known.

Pedigree S

Variants in several candidate genes were observed in Pedigree S where both mother and daughter had macrocytic anaemia with a modest thrombocytopenia. The most prominent candidate was observed in the histone methyltransferase, *KMT2D*. Whilst *KMT2D* is frequently mutated in lymphoid neoplasms such as DLBL and FL (Pasqualucci *et al*.,2011; Okosun *et al*., 2013), a significant number of mutations have also been reported in MDS, suggesting a relevant role in the pathogenesis of myeloid malignancy. Germline mutations in *KMT2D* are associated with Kabuki syndrome and whilst the majority of these are nonsense or frameshift mutations causing haploinsufficiency, missense mutations have been reported with apparent asymptomatic parental carriers of children with KS (Micale *et al*., 2014). The *KMT2D* mutation in mother and daughter from Pedigree S was a unique C-terminal missense variant, not listed in COSMIC or ExAC and predicted to be damaging. If this mutation represents the pathogenic germline

allele, it would suggest that *KMT2D* mutations may display phenotypes other than Kabuki and conversely that Kabuki patients with certain mutations may be at increased risk of MDS/AML. Although, of note, there have been no reports of haematological malignancy in KS patients to date.

Other germline variants predicted to be damaging of note within this family involved the DNA damage response genes *ATR* and *MED1*, the former represents a more favourable candidate with multiple lesions described in haematological malignancies. The finding of a nonsense mutation in exon 1 of the β-tubulin component *TUBB*, may point to an alternative predisposing mechanism. Kunishima *et al* (2009) reported the case of an autosomal dominant thrombocytopenia in a Japanese mother and son, who both carried the *TUBB1* R318W mutation, which encodes β1 tubulin and is almost exclusively expressed in platelets and megakaryocytes. Whilst limited information is available regarding *TUBB*, the haploinsufficiency arising from this p.Q15X variant may be relevant to the disease phenotype observed within this family and warrants further investigation of cases with familial thrombocytopenia.

The germline *KMT2D* variant appears the most likely susceptibility candidate within this family, although several others are of interest and, at present, it is difficult to convincingly distinguish the pathogenic lesion in this family.

Pedigree T

The final pedigree investigated in this study, included two brothers with MDS/AML who presented at 11 months and 16 years. From 23 variants, *REV3L,* emerged as the favoured germline variant. *REV3L* encodes the catalytic subunit of DNA polymerase zeta and is involved with trans-lesion repair, essential for the replication of exogenously or endogenously damaged DNA (Lin *et al*., 1999, Van Sloun *et al*.,2002). *Rev3L* null mice demonstrate embryonic lethality, however heterozygous deficiency results in normal development with increased levels of p53 (Van Sloun *et al*., 2002). Heterozygous *de novo REV3L* mutations were recently described in three unrelated patients with Mobius syndrome, a rare congenital disorder characterised by non-progressive facial and abducens cranial nerve palsies (Tomas-Roca *et al*.,2015). Notably, all 3 mutations described in this report localised to N-terminal, suggesting that this region may harbour

a unique role in neuronal development. In contrast, the unique mutation in the siblings from Pedigree T localised to the C-terminal polymerase domain, possibly leading to increased DNA damage or chromosomal instability (as suggested by the occurrence of monosomy 7 in both tumours).

5.5 Limitations of this study

This study aimed firstly to identify novel or atypical lesions within known familial MDS/AML loci and secondly to detect novel genetic susceptibility candidates using WES. While deletions within *RUNX1* were successfully identified, following recent reports of the novel familial leukaemia genes, *ETV6* and *DDX41*, it should be noted that CNAs at these loci were not comprehensively investigated, although coding variants were not detected.

WES excluded germline SNVs and small Indels within known loci for Pedigrees P-T, however this method has limited sensitivity for detecting larger indel mutations, which are best characterised with Sanger sequencing. WES includes minimal or negligible coverage of regulatory promoter and intronic regions, which may also harbour pathogenic germline mutations, as observed in the intronic *GATA2* promoter (Hsu *et al*., 2013), such mutations are optimally detected with whole genome sequencing approaches. Furthermore, inherited or *de novo* epigenetic modifications of non-coding regions remain to be explored. This is an important line of investigation with the potential for genetic silencing due to promoter hypermethylation, as previously reported for *RUNX1* (Webber *et al*., 2013) *GATA2* (Celton *et al*., 2014; Shih *et al*., 2014) and *CEBPA* (Wouters *et al*., 2007; Lin *et al*., 2011; Fasan *et al*., 2013).

5.6 Discussion

5.6.1 Familial platelet disorder and myeloid malignancy with *RUNX1* deletions

The detection of *RUNX1* deletions in four pedigrees with thrombocytopenia and/or acute leukaemia highlights the recurrence of pathogenic copy number losses within this gene. This has important implications for screening of families and sheds further light into the mechanisms underlying disease initiation within these families. Our results suggest that copy number analysis of candidate genetic loci should be performed for all patients presenting with probable inherited thrombocytopenia and MDS/AML. This may involve a customised genome wide approach, such as array CGH, or targeted analysis using MLPA.

It is noteworthy that deletions across all four families led to loss of the *RUNX1* N-terminal exons 1 and 2, most likely causing preferential expression of the shorter *RUNX1b* isoform, transcribed from the proximal P2 promoter. In mice, Runx1b is critical for haematopoietic development and studies have shown that even a 12-hour pulse of this may rescue *Runx1* -/- blast colony development, permitting the formation of definitive haematopoietic progenitors (Hoogenkamp *et al*., 2009). The P2 promoter is active throughout primitive and definitive haematopoiesis during embryogenesis. These stages describe the emergence of haematopoietic precursors from the haemogenic endothelium and the development of committed haematopoietic progenitors, respectively (Bee *et al*., 2009; Challen *et al*., 2010). Activity of the distal P1 promoter is dominant in adult mice, although targeted knockout of the distal coding region revealed no evidence of mid-gestation or adult lethality, suggesting *RUNX1c* was dispensable for haematopoietic maintenance (Sroczynska *et al*., 2010).

Emerging data appear to suggest that the growth of human lymphoblastoid cell lines is suppressed by RUNX1c but not by RUNX1b, suggesting that selective expression of the latter with N-terminal deletions may provide a proliferative advantage (Brady *et al*., 2013). Collectively, our findings indicate that heterozygous loss of the P1 promoter within FPD/AML pedigrees may cause mild to moderate defects in haematopoiesis, which most frequently involve the megakaryocytic lineage, manifesting as familial

platelet disorder. It is likely that this ongoing disruption generates a selective pressure that predisposes to the acquisition of secondary mutations and clonal evolution leading to MDS/acute leukaemia. Whilst it has not been possible to investigate the spectrum of somatic mutations, this remains an avenue for further investigation. Given the heterogeneity of disease phenotypes across *RUNX1*-deleted pedigrees, one may predict a diverse spectrum of secondary mutations in affected individuals.

5.6.2 Novel disease insights with atypical lesions in *GATA2* and *CEBPA*

The atypical *GATA2* and *CEBPA* lesions described within this chapter provide unique insights into disease pathogenesis. The novel compound germline *GATA2* mutation was predicted to cause haploinsufficiency due to NMD and was associated with marked variation of clinical phenotypes across family members. This clinical heterogeneity was likely to be caused by the acquisition of secondary mutations, as haploinsufficient mutations lack a dominant negative effect which specifically disrupts protein structure and function.

The atypical germline *CEBPA* mutation provided important insights into this distinct subset of familial leukaemia, confirming that inherited lesions are not confined to the N-terminal of the gene. This novel C-terminal frameshift mutation was associated with three asymptomatic carriers and two affected individuals in whom disease latency differed by almost 50 years; accordingly, the estimated mutation penetrance was only 40%. The increased frequency of asymptomatic carriers may be a unique property of germline C-terminal *CEBPA* mutations, furthermore, this finding potentially explains the strong bias towards the detection of germline N-terminal mutations, which are typically associated with complete penetrance. It should be noted that in the majority of families with N-terminal mutations, estimations of penetrance are tempered by limited testing of extended family members, possibly leading to an overestimation of disease frequency.

Asymptomatic carriage across the mother, maternal grandfather and great-grandfather in this novel pedigree, suggested that disease onset followed a non-linear course and was likely to be governed by a complex network of factors regulating gene expression and the acquisition of somatic mutations. Elucidating these associations in future work

will provide critical information with which to inform families and counsel asymptomatic carriers.

5.6.3 The challenges of investigating unknown germline mutations

In addition to identifying atypical lesions in established familial leukaemia genes, the second part of this chapter focussed on the investigation of novel susceptibility loci identified through WES. The identification of novel genetic candidates requires demonstration of mutation recurrence across =/>2 families, or segregation with disease in multiple affected members of large pedigrees. Since sample availability is often limited for historical cases from large pedigrees, success in discovery genetics often relies on the investigation of multiple pedigrees with a homogeneous phenotype. Such investigations are hampered by the rarity of familial cases, clinical heterogeneity and costs of sequencing; often requiring collaboration between large international consortia to power these studies. Whilst several candidate loci were detected within this investigation, limited conclusions can be drawn to the lack of mutation recurrence within the cohort. As further cases are investigated by our group, variants will be cross referenced with the data from these pedigrees in an attempt to distinguish novel susceptibility loci.

Perhaps the most interesting findings were in pedigree P, where epigenetic dysregulation was observed in prioritised germline variants (*MYST3* and *GATAD2B*) in addition to somatic variants within father and son. Importantly, combined analysis of germline and somatic variants in the father's (P.III.2) tumour DNA revealed convergence upon NuRD disruption, with both germline (*GATAD2B*) and somatic (*PHF6*) mutations targeting this chromatin remodelling complex. A nonsense mutation of *TET2* (p.R1486X) was also detected in tumour DNA from P.III.2, while in the son (P.IV.2), analysis of CMML tumour DNA revealed the canonical *DNMT3A* p.R882H mutation and heterozygous deletion of *ASXL1*.

PHF6 is frequently mutated in paediatric (16% of cases) and adult (38% of cases) T-ALL, but only 3% of AML cases harbour mutations within this gene (TCGARN, 2013). Data from the TCGA series were interrogated to assess the molecular profiles of *GATAD2B* and *PHF6* mutated AML. Notably, a single *GATAD2B* mutation was identified (R179X) and this

co-occurred with a missense *PHF6* mutation (C242Y), further suggesting molecular co-operation may occur between these rare lesions (CBioportal).

Somatic *TET2* mutations occur in 10-20% of sporadic AML, with tumours characterised by reduced 5-hMC, increased DNA methylation and epigenetic silencing (Figueroa *et al*., 2010; Ko *et al*., 2010; Rampal *et al*., 2014). *DNMT3A* mutations are detected in approximately 20% of AML (Ley *et al*., 2010) and 10% of CMML cases (Jankowska *et al*., 2011). Furthermore, Xu *et al* (2014) recently demonstrated the recurrent R882H lesion (observed in P.IV.2) led to the development of CMML in a transgenic mouse model. ASXL1 is a putative epigenetic modifier involved in PRC2-mediated gene repression (Abdel-Wahab *et al*., 2012). Recurrent somatic loss-of- function *ASXL1* mutations are detected in 20% of MDS and 10% of AML cases, with enrichment in MDS/MPN crossover syndromes, including CMML, where mutations occur in 40-50% of cases (Gelsi-Boyer *et al*., 2009; Boultwood *et al*., 2010; Jankowska *et al*., 2011). These epigenetic lesions highlight the apparent convergence of somatic mutations in pedigree P and, in light of their recurrence in sporadic AML and CMML, suggest a direct effect upon the disease phenotypes observed within this family.

5.6.4 Conclusion

With ongoing characterisation of the human genome and exome across multiple diseases, it is likely that novel familial leukaemia susceptibility genes will soon emerge and it is possible that analysis of somatic variants may reveal important clues or novel insights into molecular co-operation. In contrast to sporadic leukaemia, where molecular lesions are repeatedly characterised in large clinical trials over several years, novel candidate genes in familial leukaemia often have an immediate translational impact enabling further genetic counselling and testing of patients and their families. It is therefore essential that reports of novel lesions are verified across multiple pedigrees, with appropriate measures to exclude polymorphic germline alleles. Whilst this once appeared an almost impossible task, given the rarity of families and the genetic technologies available, rapid technological advances and the development of large sequencing databases have made this increasingly practical and possible.

Chapter 6

Discussion

6 Discussion

With the advent of NGS technologies, the focus of genetic investigation in haematological malignancies has now shifted towards the comprehensive characterisation of tumours and disease evolution. With an integrated application of sequencing and copy number analysis, this thesis describes the molecular profiling of familial leukaemia, patterns of disease progression and the identification of novel germline lesions which will help pave the way to improved understanding of this rare patient population.

6.1 Models of disease progression in sporadic and familial AML

One of the key findings in this thesis was elucidating the pattern of disease evolution in familial *CEBPA*-mutated AML. By combining sequencing technologies we revealed that the instability of *CEBPA* mutations at relapse represented the onset of novel leukaemic episodes. Notably, we also observed an episode of relapse initiated by a minor pre-existing *TET2* mutated subclone, reminiscent of studies describing subclonal evolution in sporadic AML. Ding *et al.* (2012) elegantly established two models of clonal evolution initiating sporadic AML relapse, the first described expansion of the founding tumour clone (linear evolution) and the second depicted expansion of a pre-existing subclone (branching evolution). Both models were associated with the acquisition of novel genetic lesions promoting clonal selection and proliferation.

These models of evolution suggested that LICs could remain latent for months to years following complete remission (using standard disease assessment methods) and that latent pre-leukaemic cells were capable of normal haematopoiesis, but had the potential to re-expand and instigate relapse. Expanding upon these concepts, several groups subsequently characterised persistent LIC populations during remission. Notably, these 'pre-leukaemic' cells, frequently harboured mutations within the epigenetic modifiers *DNMT3A* and *TET2.* Pre-leukaemic clones are capable of normal haematopoiesis but display a multilineage repopulation advantage over normal HSCs suggesting ongoing clonal selection with a propensity to initiate relapse (Shlush *et al*., 2014; Corces-Zimmerman *et al*., 2013; Klco *et al*., 2014). Collectively, these studies have led to a

paradigm shift in our understanding of AML. As we now move towards an era of high-throughput cancer diagnostics, the recognition of pre-leukaemic cells and subclonal evolution highlights the need for comprehensive molecular profiling of tumours, to define both clonal and subclonal lesions. This may significantly enhance the sensitivity and accuracy of MRD monitoring and may also provide novel prognostic insights.

6.2 A new model of disease progression in familial AML

As we begin to elucidate the pattern of disease evolution in sporadic AML, similar questions can be posed of familial AML. Critically, it is essential to determine whether germline mutations confer unique properties that distinguish the biology of familial and sporadic forms of AML. It would be logical to hypothesise that all germline mutations are essentially pre-leukaemic variants which affect not only a small fraction, but the entire HSC population and that the successive acquisition of mutations within a subset of these stem/progenitor cells, finally initiates leukaemia.

Following remission of familial AML (without allogeneic HSCT), germline mutations persist in all HSCs and pose a threat of disease recurrence. Theoretically, this may arise from the initial LIC population/s (if persistent) or from independent clonal events. The latter appears most likely to occur with highly penetrant germline mutations, which may confer a greater propensity to somatic mutation acquisition or demonstrate more potent leukaemia-initiating signals following the acquisition of specific mutations.

To date, clinical and scientific investigation of familial leukaemia cases has not addressed the key issues of disease evolution and progression which clearly have important connotations for the counselling and management of patients and their families. The main objectives of Chapter 3 were therefore to perform an in-depth analysis of clinical outcomes and disease evolution in a subset of familial AML associated with germline *CEBPA* mutations.

Familial *CEBPA*-mutated AML was a prime candidate for this investigation, providing a unique opportunity for direct comparison with sporadic *CEBPA*-mutated AML. Both familial and sporadic *CEBPA*-mutated tumours are relatively homogeneous, causing *de*

novo AML, predominantly of FAB subtypes M1 and M2. Furthermore sporadic *CEBPA*-mutated AML is now a well-characterised entity, with clear definition of prognostic outcomes in single and double-mutated cases (Dufour *et al*., 2010; Green *et al*., 2010; Fasan *et al*., 2014). By collaborating with Prof. Rosemary Gale and Dr. Robert Hills, we were able to compare familial AML outcomes with those of a sporadic *CEBPA*-mutated cohort treated within the national MRC AML 10 and 12 trials. Importantly, several studies have also examined mutation profiles of *CEBPA*-mutated tumours at diagnosis and relapse (Tiesmeier *et al*., 2003; Shih *et al*., 2006; Hollink *et al*., 2011) and this provided a useful foundation for comparison with disease recurrence in familial AML.

Whole-exome profiling revealed that familial tumours harboured a similar mutation profile to that of sporadic *CEBPA*dm AML, with *GATA2* and *WT1* mutations most frequently observed (Grief *et al*., 2012; Green *et al*., 2013; Fasan *et al*., 2014). The discordance of somatic *CEBPA* mutations in over 80% of tumours at diagnosis and relapse was an unexpected finding which signified the occurrence of entirely new leukaemic episodes. Notably, recurrence was also observed post autologous-SCT, with branching evolution from a pre-existing *TET2*-mutated subclone. As AML had first presented 17 years earlier and was followed by an independent leukaemic episode, the most likely explanation for the final disease episode was persistence of nascent chemotherapy-resistant LICs within the autologous stem cell harvest. Indeed, it has been reported that *TET2* mutations are associated with a reduced frequency of CR and shorter DFS in AML (Metzeler *et al*., 2011), suggesting they may confer a degree of chemotherapy-resistance, as observed in this patient.

This novel pattern of disease evolution, combining independent leukaemic episodes, has not previously been described and appears unique to familial AML, signifying the first key difference to be reported in the biology of familial and sporadic forms of disease. Importantly, relapse was associated with durable responses to secondary therapies, and this information may influence subsequent decisions to proceed with allogeneic HSCT, particularly in the absence of suitably matched sibling or unrelated donors, thereby avoiding unnecessary and high-risk procedures. Since the identification of novel leukaemic episodes is easily distinguished by comparison of somatic *CEBPA* mutations at

diagnosis and relapse, our data provide strong rationale for performing this rapid, cost-effective analysis for a useful insight into disease evolution at relapse.

6.3 The molecular convergence of tumours within families

The existence of non-random mutations in critical regulators of cell growth and differentiation is a key theme in cancer pathogenesis and this is most aptly demonstrated in leukaemogenesis, where a relatively limited number of genes are prone to recurrent mutation or translocation in the majority of cases. In recent years the use of targeted gene panels or WES/WGS has provided an invaluable insight into the patterns of mutation acquisition, further demonstrating this apparent non-random selection of lesions. Whilst some mutations are mutually exclusive indicating similar functions (e.g. *WT1* and *TET*2, or *IDH1/2* and *TET2*), other lesions often co-occur, suggesting synergistic effects (e.g. *CEBPA* and *GATA2*, *FLT3*-ITD and *NPM1* or *c-KIT* and CBF gene fusions).

In sporadic MDS/AML, we are unable to predict the factors governing mutation acquisition or the subsequent risk of disease within asymptomatic mutation carriers. For example, it is estimated that 5% of healthy adults > 70 years harbour *DNMT3A* or *TET2* mutations (Busque *et al*., 2013; Xie *et al*., 2014), notably, after 5 years of follow up, only 1 out of 7 cases developed disease *(JAK2* V617F positive ET) (Busque *et al*., 2013). Host genetic factors are often considered to contribute to disease development (evidenced by the wealth of genome-wide association studies [GWAS]); although these variants are not considered to be directly causal, they may influence a trait, or co-localise with common lesions to increase the individual's risk. Therefore, we may consider that a unique combination of inherited and somatic variants ultimately leads to disease initiation.

In this thesis, the molecular profiling of tumours across multiple families suggests that host genetic factors may pre-determine the acquisition of somatic lesions (targeting specific genes (*GATA2* in pedigree A), signalling pathways (JAK-STAT in pedigree K) or chromosomal LOH (11p aUPD in pedigree B), which subsequently mediate disease development. Whilst these events may represent a chance occurrence, it appears increasingly likely that they provide further evidence of the non-random acquisition of

mutations across familial tumours. Perhaps more significantly, they suggest that some family members may develop similar disease profiles. Whilst these data are preliminary observations and certainly do not preclude independent investigation of individuals, they provide a new insights into familial leukaemogenesis creating a foundation for future work in familial and sporadic disease.

6.4 Challenges in the clinical recognition of leukaemia predisposition syndromes

Perhaps one of the biggest clinical challenges is the accurate recognition of familial disease and the identification of germline mutations. This task is made increasingly complex by the absence of a comprehensive family history, atypical disease manifestations or an, as yet, undiscovered genetic lesion. Throughout the course of this project, I have often heard from clinicians that the inherited predisposition only becomes apparent when multiple family members are diagnosed with leukaemia, leading to missed opportunities in the management of patients and their families, particularly with consideration to related HSC donors. At present, there is little practical information or guidance on how diagnostic evaluation should be performed with an absence of nationally or internationally defined criteria or guidelines, particularly within the basis of clinical trials. This is further exacerbated by the variable availability of screening tests within accredited diagnostic laboratories, leading many clinicians to approach research laboratories for assistance.

This can be a highly distressing and difficult experience for patients and their relatives, who often fear an inherited predisposition without the medical and scientific knowledge to confirm or treat this pro-actively. Families and clinicians are often left perplexed when, for example, the presence of an obvious FPD-AML phenotype, as in Pedigrees L and M, fails to identify sequence variations in *RUNX1*. By using techniques such as MLPA and aCGH, this thesis highlights the significance of investigating germline CNAs in known genetic loci, and it is hoped that this will lead to a more comprehensive screening algorithm for future clinical implementation.

As the molecular landscape of familial leukaemia rapidly expands, information is now filtering into clinical circles and modification of screening assays should aim to assimilate these findings at an equal pace. Within the last 3 years alone, several candidate genes have been reported including germline *ETV6* (Zhang *et al*., 2014) and *DDX41* (Polprasert *et al*., 2015) mutations. Whilst Sanger sequencing of multiple genes may be time consuming and require large amounts of DNA, the advent of NGS targeted gene panels provide an opportunity for massively parallel sequencing of multiple loci simultaneously, reducing time, costs and DNA requirements.

Clinical vigilance is equally important in the recognition of familial leukaemia. It is entirely plausible that, to date, we have identified only the 'tip of the iceberg' and routine testing of remission or germline DNA may help to identify many more cases. This represents a paradigm shift in current investigation, with appreciable logistical difficulties. Testing of remission PB or BM DNA appears the most practical option for patients presenting with *de novo* AML, while in sAML, the potential persistence of dysplastic clones warrants investigation of skin or salivary DNA to confirm germline mutation status. It is essential that all patients and guardians are appropriately counselled prior to proceeding with testing, to ensure that individuals are fully aware of the implications and are receptive to subsequent counselling and explanation in the event of positive results (Nickels *et al*., 2013). The clinical attentiveness of a colleague recently led to the detection of a germline C-terminal *CEBPA* mutation in a child of 5 years, who presented with AML in the absence of a positive family history, as described in Chapter 5. Suspicion of a germline mutation was prompted by the infrequent presentation of *CEBPA*dm AML, which occurs in <5% of paediatric AML cases (Ho *et al*., 2009; Liang *et al*., 2005). Only after further information from the child's grandparents, was it discovered that her maternal great-uncle had also developed leukaemia at 53 years. Remarkably, the child's mother, maternal grandfather and maternal great-grandfather had remained asymptomatic carriers aged 39-75 years. Detection of germline mutations is highly relevant for the management of AML, which often entails consideration of sibling allogeneic HSCT. Indeed, reports of donor derived leukaemia, serve as an important reminder of the potential consequences when germline mutations are not considered and tested (Xiao *et al*., 2011).

6.5 Suggestions for future work

As with all rare diseases, research in familial leukaemia significantly lags behind work in sporadic MDS/AML. Whilst investigation of individual families is useful for discovery purposes, consolidating the scientific impact and clinical significance of these results requires their replication across multiple families. To achieve this end, a collaborative research network including clinicians and scientists, has been founded at Barts Cancer Institute, and aims to identify, investigate, counsel and treat patients in a multi-disciplinary approach with a view to powering larger scale studies. Future goals are to develop and optimise a comprehensive screening program for this rare patient population, to further characterise disease evolution and molecular co-operation across multiple genetic subtypes, to uncover the factors governing mutation penetrance, and identify novel candidate genes. The integration of this work will build a foundation for the personalised management of familial leukaemia, with novel molecular insights that may ultimately lead to the implementation of targeted therapies for cancer prevention and treatment.

6.6 Conclusion

This thesis describes the profiling of somatic and germline mutations in multiple families with MDS/AML, revealing a new model of disease progression, molecular convergence of tumours within families, and novel germline lesions in known and unknown genetic loci. This work has made a significant contribution to our understanding of familial leukaemia, whilst importantly highlighting avenues for further investigation.

References

Abdel-Wahab, O. and Levine, R. L. (2013) 'Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia', *Blood*, 121(18), 3563-72.

Abdel-Wahab, O., Adli, M., LaFave, L. M., Gao, J., Hricik, T., Shih, A. H., Pandey, S., Patel, J. P., Chung, Y. R., Koche, R., Perna, F., Zhao, X., Taylor, J. E., Park, C. Y., Carroll, M., Melnick, A., Nimer, S. D., Jaffe, J. D., Aifantis, I., Bernstein, B. E. and Levine, R. L. (2012) 'ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression', *Cancer Cell*, 22(2), 180-93.

Andor, N., Harness, J. V., Müller, S., Mewes, H. W. and Petritsch, C. (2014) 'EXPANDS: expanding ploidy and allele frequency on nested subpopulations', *Bioinformatics*, 30(1), 50-60.

Antoniou, A. C., Casadei, S., Heikkinen, T., Barrowdale, D., Pylkäs, K., Roberts, J., Lee, A., Subramanian, D., De Leeneer, K., Fostira, F., Tomiak, E., Neuhausen, S. L., Teo, Z. L., Khan, S., Aittomäki, K., Moilanen, J. S., Turnbull, C., Seal, S., Mannermaa, A., Kallioniemi, A., Lindeman, G. J., Buys, S. S., Andrulis, I. L., Radice, P., Tondini, C., Manoukian, S., Toland, A. E., Miron, P., Weitzel, J. N., Domchek, S. M., Poppe, B., Claes, K. B., Yannoukakos, D., Concannon, P., Bernstein, J. L., James, P. A., Easton, D. F., Goldgar, D. E., Hopper, J. L., Rahman, N., Peterlongo, P., Nevanlinna, H., King, M. C., Couch, F. J., Southey, M. C., Winqvist, R., Foulkes, W. D. and Tischkowitz, M. (2014) 'Breast-cancer risk in families with mutations in PALB2', *N Engl J Med*, 371(6), 497-506.

Antony-Debré, I., Duployez, N., Bucci, M., Geffroy, S., Micol, J. B., Renneville, A., Boissel, N., Dhédin, N., Réa, D., Nelken, B., Berthon, C., Leblanc, T., Mozziconacci, M. J., Favier, R., Heller, P. G., Abdel-Wahab, O., Raslova, H., Latger-Cannard, V. and Preudhomme, C. (2015) 'Somatic mutations associated with leukemic progression of familial platelet disorder with predisposition to acute myeloid leukemia', *Leukemia*.

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A. and Shendure, J. (2011) 'Exome sequencing as a tool for Mendelian disease gene discovery', *Nat Rev Genet*, 12(11), 745-55.

Bao, L., Pu, M. and Messer, K. (2014) 'AbsCN-seq: a statistical method to estimate tumor purity, ploidy and absolute copy numbers from next-generation sequencing data', *Bioinformatics*.

Baran-Marszak, F., Magdoud, H., Desterke, C., Alvarado, A., Roger, C., Harel, S., Mazoyer, E., Cassinat, B., Chevret, S., Tonetti, C., Giraudier, S., Fenaux, P., Cymbalista, F., Varin-Blank, N., Le Bousse-Kerdilès, M. C., Kiladjian, J. J. and Velazquez, L. (2010) 'Expression level and differential JAK2-V617F-binding of the adaptor protein Lnk regulates JAK2-mediated signals in myeloproliferative neoplasms', *Blood*, 116(26), 5961-71.

Basecke, J., Whelan, J. T., Griesinger, F. and Bertrand, F. E. (2006) 'The MLL partial tandem duplication in acute myeloid leukaemia', *Br J Haematol*, 135(4), 438-49.

Bee, T., Liddiard, K., Swiers, G., Bickley, S. R., Vink, C. S., Jarratt, A., Hughes, J. R., Medvinsky, A. and de Bruijn, M. F. (2009) 'Alternative Runx1 promoter usage in mouse developmental hematopoiesis', *Blood Cells Mol Dis*, 43(1), 35-42.

Bee, T., Swiers, G., Muroi, S., Pozner, A., Nottingham, W., Santos, A. C., Li, P. S., Taniuchi, I. and de Bruijn, M. F. (2010) 'Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis', *Blood*, 115(15), 3042-50.

Beer, P. A., Delhommeau, F., LeCouédic, J. P., Dawson, M. A., Chen, E., Bareford, D., Kusec, R., McMullin, M. F., Harrison, C. N., Vannucchi, A. M., Vainchenker, W. and Green, A. R. (2010) 'Two routes to leukemic transformation after a JAK2 mutation-positive myeloproliferative neoplasm', *Blood*, 115(14), 2891-900.

Bejar, R. and Steensma, D. P. (2014) 'Recent developments in myelodysplastic syndromes', *Blood*, 124(18), 2793-803.

Bejar, R., Levine, R. and Ebert, B. L. (2011a) 'Unraveling the molecular pathophysiology of myelodysplastic syndromes', *J Clin Oncol*, 29(5), 504-15.

Bejar, R., Stevenson, K., Abdel-Wahab, O., Galili, N., Nilsson, B., Garcia-Manero, G., Kantarjian, H., Raza, A., Levine, R. L., Neuberg, D. and Ebert, B. L. (2011b) 'Clinical effect of point mutations in myelodysplastic syndromes', *N Engl J Med*, 364(26), 2496-506.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R. and Sultan, C. (1976) 'Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group', *Br J Haematol*, 33(4), 451-8.

Bereshchenko, O., Mancini, E., Moore, S., Bilbao, D., Månsson, R., Luc, S., Grover, A., Jacobsen, S. E., Bryder, D. and Nerlov, C. (2009) 'Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPalpha mutant AML', *Cancer Cell*, 16(5), 390-400.

Bernt, K. M., Zhu, N., Sinha, A. U., Vempati, S., Faber, J., Krivtsov, A. V., Feng, Z., Punt, N., Daigle, A., Bullinger, L., Pollock, R. M., Richon, V. M., Kung, A. L. and Armstrong, S. A. (2011) 'MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L', *Cancer Cell*, 20(1), 66-78.

Bersenev, A., Wu, C., Balcerek, J. and Tong, W. (2008) 'Lnk controls mouse hematopoietic stem cell self-renewal and quiescence through direct interactions with JAK2', *J Clin Invest*, 118(8), 2832-44.

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. and Raslova, H. (2014)

'Thrombocytopenia-associated mutations in the ANKRD26 regulatory region induce MAPK hyperactivation', *J Clin Invest*, 124(2), 580-91.

Bödör, C., Grossmann, V., Popov, N., Okosun, J., O'Riain, C., Tan, K., Marzec, J., Araf, S., Wang, J., Lee, A. M., Clear, A., Montoto, S., Matthews, J., Iqbal, S., Rajnai, H., Rosenwald, A., Ott, G., Campo, E., Rimsza, L. M., Smeland, E. B., Chan, W. C., Braziel, R. M., Staudt, L. M., Wright, G., Lister, T. A., Elemento, O., Hills, R., Gribben, J. G., Chelala, C., Matolcsy, A., Kohlmann, A., Haferlach, T., Gascoyne, R. D. and Fitzgibbon, J. (2013) 'EZH2 mutations are frequent and represent an early event in follicular lymphoma', *Blood*, 122(18), 3165-8.

Bödör, C., Renneville, A., Smith, M., Charazac, A., Iqbal, S., Etancelin, P., Cavenagh, J., Barnett, M. J., Kramarzová, K., Krishnan, B., Matolcsy, A., Preudhomme, C., Fitzgibbon, J. and Owen, C. (2012) 'Germ-line GATA2 p.THR354MET mutation in familial myelodysplastic syndrome with acquired monosomy 7 and ASXL1 mutation demonstrating rapid onset and poor survival', *Haematologica*, 97(6), 890-4.

Borkhardt, A., Cazzaniga, G., Viehmann, S., Valsecchi, M. G., Ludwig, W. D., Burci, L., Mangioni, S., Schrappe, M., Riehm, H., Lampert, F., Basso, G., Masera, G., Harbott, J. and Biondi, A. (1997) 'Incidence and clinical relevance of TEL/AML1 fusion genes in children with acute lymphoblastic leukemia enrolled in the German and Italian multicenter therapy trials. Associazione Italiana Ematologia Oncologia Pediatrica and the Berlin-Frankfurt-Münster Study Group', *Blood*, 90(2), 571-7.

Borrow, J., Stanton, V. P., Andresen, J. M., Becher, R., Behm, F. G., Chaganti, R. S., Civin, C. I., Disteche, C., Dubé, I., Frischauf, A. M., Horsman, D., Mitelman, F., Volinia, S., Watmore, A. E. and Housman, D. E. (1996) 'The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein', *Nat Genet*, 14(1), 33-41.

Boultwood, J., Lewis, S. and Wainscoat, J. S. (1994) 'The 5q-syndrome', *Blood*, 84(10), 3253-60.

Boultwood, J., Perry, J., Pellagatti, A., Fernandez-Mercado, M., Fernandez-Santamaria, C., Calasanz, M. J., Larrayoz, M. J., Garcia-Delgado, M., Giagounidis, A., Malcovati, L., Della Porta, M. G., Jädersten, M., Killick, S., Hellström-Lindberg, E., Cazzola, M. and Wainscoat, J. S. (2010) 'Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia', *Leukemia*, 24(5), 1062-5.

Bowie, M. B., McKnight, K. D., Kent, D. G., McCaffrey, L., Hoodless, P. A. and Eaves, C. J. (2006) 'Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect', *J Clin Invest*, 116(10), 2808-16.

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. and Helin, K. (2006) 'Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions', *Genes Dev*, 20(9), 1123-36.

Brackertz, M., Gong, Z., Leers, J. and Renkawitz, R. (2006) 'p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction', *Nucleic Acids Res*, 34(2), 397-406.

Brady, G., Elgueta Karstegl, C. and Farrell, P. J. (2013) 'Novel function of the unique N-terminal region of RUNX1c in B cell growth regulation', *Nucleic Acids Res*, 41(3), 1555-68.

Breems, D. A., Van Putten, W. L., Huijgens, P. C., Ossenkoppele, G. J., Verhoef, G. E., Verdonck, L. F., Vellenga, E., De Greef, G. E., Jacky, E., Van der Lelie, J., Boogaerts, M. A. and Löwenberg, B. (2005) 'Prognostic index for adult patients with acute myeloid leukemia in first relapse', *J Clin Oncol*, 23(9), 1969-78.

Büchner, T., Berdel, W. E., Haferlach, C., Haferlach, T., Schnittger, S., Müller-Tidow, C., Braess, J., Spiekermann, K., Kienast, J., Staib, P., Grüneisen, A., Kern, W., Reichle, A., Maschmeyer, G., Aul, C., Lengfelder, E., Sauerland, M. C., Heinecke, A., Wörmann, B. and Hiddemann, W. (2009b) 'Age-related risk profile and chemotherapy dose response in acute myeloid leukemia: a study by the German Acute Myeloid Leukemia Cooperative Group', *J Clin Oncol*, 27(1), 61-9.

Busque, L., Patel, J. P., Figueroa, M. E., Vasanthakumar, A., Provost, S., Hamilou, Z., Mollica, L., Li, J., Viale, A., Heguy, A., Hassimi, M., Socci, N., Bhatt, P. K., Gonen, M., Mason, C. E., Melnick, A., Godley, L. A., Brennan, C. W., Abdel-Wahab, O. and Levine, R. L. (2012) 'Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis', *Nat Genet*, 44(11), 1179-81.

Buza-Vidas, N., Antonchuk, J., Qian, H., Månsson, R., Luc, S., Zandi, S., Anderson, K., Takaki, S., Nygren, J. M., Jensen, C. T. and Jacobsen, S. E. (2006) 'Cytokines regulate postnatal hematopoietic stem cell expansion: opposing roles of thrombopoietin and LNK', *Genes Dev*, 20(15), 2018-23.

Carapeti, M., Aguiar, R. C., Goldman, J. M. and Cross, N. C. (1998) 'A novel fusion between MOZ and the nuclear receptor coactivator TIF2 in acute myeloid leukemia', *Blood*, 91(9), 3127-33.

Cazzola, M., Della Porta, M. G. and Malcovati, L. (2013) 'The genetic basis of myelodysplasia and its clinical relevance', *Blood*, 122(25), 4021-34.

Celton, M., Forest, A., Gosse, G., Lemieux, S., Hebert, J., Sauvageau, G. and Wilhelm, B. T. (2014) 'Epigenetic regulation of GATA2 and its impact on normal karyotype acute myeloid leukemia', *Leukemia*, 28(8), 1617-26.

Chaffanet, M., Gressin, L., Preudhomme, C., Soenen-Cornu, V., Birnbaum, D. and Pébusque, M. J. (2000) 'MOZ is fused to p300 in an acute monocytic leukemia with t(8;22)', *Genes Chromosomes Cancer*, 28(2), 138-44.

Chakrabarti, S. R. and Nucifora, G. (1999) 'The leukemia-associated gene TEL encodes a transcription repressor which associates with SMRT and mSin3A', *Biochem Biophys Res Commun*, 264(3), 871-7.

Challen, G. A. and Goodell, M. A. (2010) 'Runx1 isoforms show differential expression patterns during hematopoietic development but have similar functional effects in adult hematopoietic stem cells', *Exp Hematol*, 38(5), 403-16.

Chan, E. M., Chan, R. J., Comer, E. M., Goulet, R. J., Crean, C. D., Brown, Z. D., Fruehwald, A. M., Yang, Z., Boswell, H. S., Nakshatri, H. and Gabig, T. G. (2007) 'MOZ and MOZ-CBP cooperate with NF-kappaB to activate transcription from NF-kappaB-dependent promoters', *Exp Hematol*, 35(12), 1782-92.

Chen, L., Nievera, C. J., Lee, A. Y. and Wu, X. (2008) 'Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair', *J Biol Chem*, 283(12), 7713-20.

Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. and Speck, N. A. (2009) 'Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter', *Nature*, 457(7231), 887-91.

Cheshier, S. H., Morrison, S. J., Liao, X. and Weissman, I. L. (1999) 'In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells', *Proc Natl Acad Sci USA*, 96(6), 3120-5.

Cheson, B. D., Bennett, J. M., Kopecky, K. J., Büchner, T., Willman, C. L., Estey, E. H., Schiffer, C. A., Doehner, H., Tallman, M. S., Lister, T. A., Lo-Coco, F., Willemze, R., Biondi, A., Hiddemann, W., Larson, R. A., Löwenberg, B., Sanz, M. A., Head, D. R., Ohno, R., Bloomfield, C. D., LoCocco, F. and International Working Group for Diagnosis, S. o. R. C., T.eatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia (2003) 'Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia', *J Clin Oncol*, 21(24), 4642-9.

Chou, S. H., Ko, B. S., Chiou, J. S., Hsu, Y. C., Tsai, M. H., Chiu, Y. C., Yu, I. S., Lin, S. W., Hou, H. A., Kuo, Y. Y., Lin, H. M., Wu, M. F., Chou, W. C. and Tien, H. F. (2012) 'A knock-in Npm1 mutation in mice results in myeloproliferation and implies a perturbation in hematopoietic microenvironment', *PLoS One*, 7(11), e49769.

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E. S. and Getz, G. (2013) 'Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples', *Nat Biotechnol*, 31(3), 213-9.

Collin, M., Dickinson, R. and Bigley, V. (2015) 'Haematopoietic and immune defects associated with GATA2 mutation', *Br J Haematol*, 169(2), 173-87.

Cooper, S., Guo, H. and Friedman, A. D. (2015) 'The +37 kb Cebpa Enhancer Is Critical for Cebpa Myeloid Gene Expression and Contains Functional Sites that Bind SCL, GATA2, C/EBPα, PU.1, and Additional Ets Factors', *PLoS One*, 10(5), e0126385.

Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C. and Majeti, R. (2014) 'Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission', *Proc Natl Acad Sci USA*, 111(7), 2548-53.

Dayem Ullah, A. Z., Lemoine, N. R. and Chelala, C. (2012) 'SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update)', *Nucleic Acids Res*, 40(Web Server issue), W65-70.

Debeljak, M., Kitanovski, L., Pajič, T. and Jazbec, J. (2013) 'Concordant acute myeloblastic leukemia in monozygotic twins with germline and shared somatic mutations in the gene for CCAAT-enhancer-binding protein α with 13 years difference at onset', *Haematologica*, 98(7), e73-4.

Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J. P., Robert, F., Alberdi, A., Lécluse, Y., Plo, I., Dreyfus, F. J., Marzac, C., Casadevall, N., Lacombe, C., Romana, S. P., Dessen, P., Soulier, J., Viguié, F., Fontenay, M., Vainchenker, W. and Bernard, O. A. (2009) 'Mutation in TET2 in myeloid cancers', *N Engl J Med*, 360(22), 2289-301.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D. and Daly, M. J. (2011) 'A framework for variation discovery and genotyping using next-generation DNA sequencing data', *Nat Genet*, 43(5), 491-8.

Dickinson, R. E., Griffin, H., Bigley, V., Reynard, L. N., Hussain, R., Haniffa, M., Lakey, J. H., Rahman, T., Wang, X. N., McGovern, N., Pagan, S., Cookson, S., McDonald, D., Chua, I., Wallis, J., Cant, A., Wright, M., Keavney, B., Chinnery, P. F., Loughlin, J., Hambleton, S., Santibanez-Koref, M. and Collin, M. (2011) 'Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency.', *Blood*, 118(10), 2656-8.

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D., McMichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., Schmidt, H., Kalicki-Veizer, J., Magrini, V. J., Cook, L., McGrath, S. D., Vickery, T. L., Wendl, M. C., Heath, S., Watson, M. A., Link, D. C., Tomasson, M. H., Shannon, W. D., Payton, J. E., Kulkarni, S., Westervelt, P., Walter, M. J., Graubert, T. A., Mardis, E. R., Wilson, R. K. and DiPersio, J. F. (2012) 'Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing', *Nature*, 481(7382), 506-10.

Dowton, S. B., Beardsley, D., Jamison, D., Blattner, S. and Li, F. P. (1985) 'Studies of a familial platelet disorder', *Blood*, 65(3), 557-63.

Dufour, A., Schneider, F., Metzeler, K. H., Hoster, E., Schneider, S., Zellmeier, E., Benthaus, T., Sauerland, M. C., Berdel, W. E., Büchner, T., Wörmann, B., Braess, J., Hiddemann, W., Bohlander, S. K. and Spiekermann, K. (2010) 'Acute myeloid leukemia with biallelic CEBPA

gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome', *J Clin Oncol*, 28(4), 570-7.

Engle, E. K., Fisher, D. A., Miller, C. A., McLellan, M. D., Fulton, R. S., Moore, D. M., Wilson, R. K., Ley, T. J. and Oh, S. T. (2015) 'Clonal evolution revealed by whole genome sequencing in a case of primary myelofibrosis transformed to secondary acute myeloid leukemia', *Leukemia*, 29(4), 869-76.

Esteyries, S., Perot, C., Adelaide, J., Imbert, M., Lagarde, A., Pautas, C., Olschwang, S., Birnbaum, D., Chaffanet, M. and Mozziconacci, M. J. (2008) 'NCOA3, a new fusion partner for MOZ/MYST3 in M5 acute myeloid leukemia', *Leukemia*, 22(3), 663-5.

Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A., Bigerna, B., Pacini, R., Pucciarini, A., Liso, A., Vignetti, M., Fazi, P., Meani, N., Pettirossi, V., Saglio, G., Mandelli, F., Lo-Coco, F., Pelicci, P. G. and Martelli, M. F. (2005) 'Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype', *N Engl J Med*, 352(3), 254-66.

Fasan, A., Alpermann, T., Haferlach, C., Grossmann, V., Roller, A., Kohlmann, A., Eder, C., Kern, W., Haferlach, T. and Schnittger, S. (2013) 'Frequency and prognostic impact of CEBPA proximal, distal and core promoter methylation in normal karyotype AML: a study on 623 cases', *PLoS One*, 8(2), e54365.

Fasan, A., Eder, C., Haferlach, C., Grossmann, V., Kohlmann, A., Dicker, F., Kern, W., Haferlach, T. and Schnittger, S. (2013) 'GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis', *Leukemia*, 27(2), 482-5.

Fasan, A., Haferlach, C., Alpermann, T., Jeromin, S., Grossmann, V., Eder, C., Weissmann, S., Dicker, F., Kohlmann, A., Schindela, S., Kern, W., Haferlach, T. and Schnittger, S. (2013) 'The role of different genetic subtypes of CEBPA mutated AML', *Leukemia*.

Federici, L. and Falini, B. (2013) 'Nucleophosmin mutations in acute myeloid leukemia: A tale of protein unfolding and mislocalization', *Protein Sci*, 22(5), 545-56.

Fenaux, P., Mufti, G. J., Hellstrom-Lindberg, E., Santini, V., Finelli, C., Giagounidis, A., Schoch, R., Gattermann, N., Sanz, G., List, A., Gore, S. D., Seymour, J. F., Bennett, J. M., Byrd, J., Backstrom, J., Zimmerman, L., McKenzie, D., Beach, C., Silverman, L. R. and Group, I. V. H.-R. M. S. S. (2009) '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*, 10(3), 223-32.

Fenaux, P., Mufti, G. J., Hellström-Lindberg, E., Santini, V., Gattermann, N., Germing, U., Sanz, G., List, A. F., Gore, S., Seymour, J. F., Dombret, H., Backstrom, J., Zimmerman, L., McKenzie, D., Beach, C. L. and Silverman, L. R. (2010a) 'Azacitidine prolongs overall survival compared with conventional care regimens in elderly patients with low bone marrow blast count acute myeloid leukemia', *J Clin Oncol*, 28(4), 562-9.

Feng, Q. and Zhang, Y. (2001) 'The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes', *Genes Dev*, 15(7), 827-32.

Fernandez, H. F., Sun, Z., Yao, X., Litzow, M. R., Luger, S. M., Paietta, E. M., Racevskis, J., Dewald, G. W., Ketterling, R. P., Bennett, J. M., Rowe, J. M., Lazarus, H. M. and Tallman, M. S. (2009a) 'Anthracycline dose intensification in acute myeloid leukemia', *N Engl J Med*, 361(13), 1249-59.

Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H. F., Tallman, M. S., Sun, Z., Wolniak, K., Peeters, J. K., Liu, W., Choe, S. E., Fantin, V. R., Paietta, E., Löwenberg, B., Licht, J. D., Godley, L. A., Delwel, R., Valk, P. J., Thompson, C. B., Levine, R. L. and Melnick, A. (2010) 'Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation', *Cancer Cell*, 18(6), 553-67.

Gaidzik, V. I., Bullinger, L., Schlenk, R. F., Zimmermann, A. S., Röck, J., Paschka, P., Corbacioglu, A., Krauter, J., Schlegelberger, B., Ganser, A., Späth, D., Kündgen, A., Schmidt-Wolf, I. G., Götze, K., Nachbaur, D., Pfreundschuh, M., Horst, H. A., Döhner, H. and Döhner, K. (2011) 'RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group'*, J Clin Oncol*, 29(10), 1364-72.

Gaidzik, V. I., Schlenk, R. F., Moschny, S., Becker, A., Bullinger, L., Corbacioglu, A., Krauter, J., Schlegelberger, B., Ganser, A., Döhner, H., Döhner, K. and Group, G.-A. A. S. (2009) 'Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group', *Blood*, 113(19), 4505-11.

Gale, R. E., Green, C., Allen, C., Mead, A. J., Burnett, A. K., Hills, R. K., Linch, D. C. and Party, M. R. C. A. L. W. (2008) 'The impact of FLT3 internal tandem duplication mutant level, number, size, and interaction with NPM1 mutations in a large cohort of young adult patients with acute myeloid leukemia', *Blood*, 111(5), 2776-84.

Gale, R. E., Hills, R., Pizzey, A. R., Kottaridis, P. D., Swirsky, D., Gilkes, A. F., Nugent, E., Mills, K. I., Wheatley, K., Solomon, E., Burnett, A. K., Linch, D. C., Grimwade, D. and Party, N. A. L. W. (2005) 'Relationship between FLT3 mutation status, biologic characteristics, and response to targeted therapy in acute promyelocytic leukemia', *Blood*, 106(12), 3768-76.

Gelsi-Boyer, V., Trouplin, V., Adélaïde, J., Bonansea, J., Cervera, N., Carbuccia, N., Lagarde, A., Prebet, T., Nezri, M., Sainty, D., Olschwang, S., Xerri, L., Chaffanet, M., Mozziconacci, M. J., Vey, N. and Birnbaum, D. (2009) 'Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia', *Br J Haematol*, 145(6), 788-800.

Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landén, M., Höglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., Sklar, P.,

Grönberg, H., Hultman, C. M. and McCarroll, S. A. (2014) 'Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence', *N Engl J Med*, 371(26), 2477-87.

Gervais, C., Murati, A., Helias, C., Struski, S., Eischen, A., Lippert, E., Tigaud, I., Penther, D., Bastard, C., Mugneret, F., Poppe, B., Speleman, F., Talmant, P., VanDen Akker, J., Baranger, L., Barin, C., Luquet, I., Nadal, N., Nguyen-Khac, F., Maarek, O., Herens, C., Sainty, D., Flandrin, G., Birnbaum, D., Mozziconacci, M. J., Lessard, M. and Hématologique, G. F. d. C. (2008) 'Acute myeloid leukaemia with 8p11 (MYST3) rearrangement: an integrated cytologic, cytogenetic and molecular study by the groupe francophone de cytogénétique hématologique', *Leukemia*, 22(8), 1567-75.

Ghozi, M. C., Bernstein, Y., Negreanu, V., Levanon, D. and Groner, Y. (1996) 'Expression of the human acute myeloid leukemia gene AML1 is regulated by two promoter regions', *Proc Natl Acad Sci USA*, 93(5), 1935-40.

Gilliland, D. G. (2002) 'Molecular genetics of human leukemias: new insights into therapy', *Semin Hematol*, 39(4 Suppl 3), 6-11.

Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K. A., Rout, R., Chaudhury, S., Gilkes, A., Knapper, S., Beldjord, K., Begum, S., Rose, S., Geddes, N., Griffiths, M., Standen, G., Sternberg, A., Cavenagh, J., Hunter, H., Bowen, D., Killick, S., Robinson, L., Price, A., Macintyre, E., Virgo, P., Burnett, A., Craddock, C., Enver, T., Jacobsen, S. E., Porcher, C. and Vyas, P. (2011) 'Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia', *Cancer Cell*, 19(1), 138-52.

Golub, T. R., Barker, G. F., Bohlander, S. K., Hiebert, S. W., Ward, D. C., Bray-Ward, P., Morgan, E., Raimondi, S. C., Rowley, J. D. and Gilliland, D. G. (1995) 'Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia', *Proc Natl Acad Sci USA*, 92(11), 4917-21.

Goyama, S., Huang, G., Kurokawa, M. and Mulloy, J. C. (2015) 'Posttranslational modifications of RUNX1 as potential anticancer targets', *Oncogene*, 34(27), 3483-92.

Green, C. L., Koo, K. K., Hills, R. K., Burnett, A. K., Linch, D. C. and Gale, R. E. (2010) 'Prognostic significance of CEBPA mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double CEBPA mutations and the interaction with FLT3 and NPM1 mutations', *J Clin Oncol*, 28(16), 2739-47.

Green, C. L., Tawana, K., Hills, R. K., Bödör, C., Fitzgibbon, J., Inglott, S., Ancliff, P., Burnett, A. K., Linch, D. C. and Gale, R. E. (2013) 'GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations', *Br J Haematol*, 161(5), 701-5.

Greenberg, P. L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Solé, F., Bennett, J. M., Bowen, D., Fenaux, P., Dreyfus, F., Kantarjian, H., Kuendgen, A., Levis, A., Malcovati, L., Cazzola, M., Cermak, J., Fonatsch, C., Le Beau, M. M., Slovak, M. L., Krieger, O., Luebbert, M., Maciejewski, J., Magalhaes, S. M., Miyazaki, Y., Pfeilstöcker, M., Sekeres, M., Sperr, W. R., Stauder, R., Tauro, S., Valent, P., Vallespi, T., van de Loosdrecht, A. A., Germing, U. and Haase,

D. (2012) 'Revised international prognostic scoring system for myelodysplastic syndromes', *Blood*, 120(12), 2454-65.

Greif, P. A., Dufour, A., Konstandin, N. P., Ksienzyk, B., Zellmeier, E., Tizazu, B., Sturm, J., Benthaus, T., Herold, T., Yaghmaie, M., Dörge, P., Hopfner, K. P., Hauser, A., Graf, A., Krebs, S., Blum, H., Kakadia, P. M., Schneider, S., Hoster, E., Schneider, F., Stanulla, M., Braess, J., Sauerland, M. C., Berdel, W. E., Büchner, T., Woermann, B. J., Hiddemann, W., Spiekermann, K. and Bohlander, S. K. (2012) 'GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia', *Blood*, 120(2), 395-403.

Griffiths, M., Mason, J., Rindl, M., Akiki, S., McMullan, D., Stinton, V., Powell, H., Curtis, A., Bown, N. and Craddock, C. (2005) 'Acquired isodisomy for chromosome 13 is common in AML, and associated with FLT3-itd mutations', *Leukemia*, 19(12), 2355-8.

Grimwade, D. (2012) 'The changing paradigm of prognostic factors in acute myeloid leukaemia', *Best Pract Res Clin Haematol*, 25(4), 419-25.

Grimwade, D. and Freeman, S. D. (2014) 'Defining minimal residual disease in acute myeloid leukemia: which platforms are ready for "prime time"?', *Blood*, 124(23), 3345-55.

Grimwade, D., Hills, R. K., Moorman, A. V., Walker, H., Chatters, S., Goldstone, A. H., Wheatley, K., Harrison, C. J., Burnett, A. K. and Group, N. C. R. I. A. L. W. (2010a) 'Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials', *Blood*, 116(3), 354-65.

Grimwade, D., Vyas, P. and Freeman, S. (2010b) 'Assessment of minimal residual disease in acute myeloid leukemia', *Curr Opin Oncol*, 22(6), 656-63.

Grimwade, D., Walker, H., Harrison, G., Oliver, F., Chatters, S., Harrison, C. J., Wheatley, K., Burnett, A. K., Goldstone, A. H. and Party, M. R. C. A. L. W. (2001) 'The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial', *Blood*, 98(5), 1312-20.

Grimwade, D., Walker, H., Oliver, F., Wheatley, K., Harrison, C., Harrison, G., Rees, J., Hann, I., Stevens, R., Burnett, A. and Goldstone, A. (1998) 'The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties', *Blood*, 92(7), 2322-33.

Grossmann, V., Kohlmann, A., Eder, C., Haferlach, C., Kern, W., Cross, N. C., Haferlach, T. and Schnittger, S. (2011) 'Molecular profiling of chronic myelomonocytic leukemia reveals diverse mutations in >80% of patients with TET2 and EZH2 being of high prognostic relevance', *Leukemia*, 25(5), 877-9.

Grove, C. S. and Vassiliou, G. S. (2014) 'Acute myeloid leukaemia: a paradigm for the clonal evolution of cancer?', *Dis Model Mech*, 7(8), 941-51.

Growney, J. D., Shigematsu, H., Li, Z., Lee, B. H., Adelsperger, J., Rowan, R., Curley, D. P., Kutok, J. L., Akashi, K., Williams, I. R., Speck, N. A. and Gilliland, D. G. (2005) 'Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype', *Blood*, 106(2), 494-504.

Guglielmelli, P., Biamonte, F., Score, J., Hidalgo-Curtis, C., Cervantes, F., Maffioli, M., Fanelli, T., Ernst, T., Winkelman, N., Jones, A. V., Zoi, K., Reiter, A., Duncombe, A., Villani, L., Bosi, A., Barosi, G., Cross, N. C. and Vannucchi, A. M. (2011) 'EZH2 mutational status predicts poor survival in myelofibrosis', *Blood*, 118(19), 5227-34.

Guo, H., Ma, O., Speck, N. A. and Friedman, A. D. (2012) 'Runx1 deletion or dominant inhibition reduces Cebpa transcription via conserved promoter and distal enhancer sites to favor monopoiesis over granulopoiesis', *Blood*, 119(19), 4408-18.

Hahn, C. N., Chong, C. E., Carmichael, C. L., Wilkins, E. J., Brautigan, P. J., Li, X. C., Babic, M., Lin, M., Carmagnac, A., Lee, Y. K., Kok, C. H., Gagliardi, L., Friend, K. L., Ekert, P. G., Butcher, C. M., Brown, A. L., Lewis, I. D., To, L. B., Timms, A. E., Storek, J., Moore, S., Altree, M., Escher, R., Bardy, P. G., Suthers, G. K., D'Andrea, R. J., Horwitz, M. S. and Scott, H. S. (2011) 'Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia.', *Nat Genet*, 43(10), 1012-7.

Harrison, C. J., Hills, R. K., Moorman, A. V., Grimwade, D. J., Hann, I., Webb, D. K., Wheatley, K., de Graaf, S. S., van den Berg, E., Burnett, A. K. and Gibson, B. E. (2010) 'Cytogenetics of childhood acute myeloid leukemia: United Kingdom Medical Research Council Treatment trials AML 10 and 12', *J Clin Oncol*, 28(16), 2674-81.

Heiss, N. S., Knight, S. W., Vulliamy, T. J., Klauck, S. M., Wiemann, S., Mason, P. J., Poustka, A. and Dokal, I. (1998) 'X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions', *Nat Genet*, 19(1), 32-8.

Ho, C. Y., Otterud, B., Legare, R. D., Varvil, T., Saxena, R., DeHart, D. B., Kohler, S. E., Aster, J. C., Dowton, S. B., Li, F. P., Leppert, M. and Gilliland, D. G. (1996) 'Linkage of a familial platelet disorder with a propensity to develop myeloid malignancies to human chromosome 21q22.1-22.2', *Blood*, 87(12), 5218-24.

Ho, P. A., Alonzo, T. A., Gerbing, R. B., Pollard, J., Stirewalt, D. L., Hurwitz, C., Heerema, N. A., Hirsch, B., Raimondi, S. C., Lange, B., Franklin, J. L., Radich, J. P. and Meshinchi, S. (2009) 'Prevalence and prognostic implications of CEBPA mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group', *Blood*, 113(26), 6558-66.

Hollink, I. H., van den Heuvel-Eibrink, M. M., Arentsen-Peters, S. T., Zimmermann, M., Peeters, J. K., Valk, P. J., Balgobind, B. V., Sonneveld, E., Kaspers, G. J., de Bont, E. S., Trka, J., Baruchel, A., Creutzig, U., Pieters, R., Reinhardt, D. and Zwaan, C. M. (2011) 'Characterization

of CEBPA mutations and promoter hypermethylation in pediatric acute myeloid leukemia', *Haematologica*, 96(3), 384-92.

Hoogenkamp, M., Lichtinger, M., Krysinska, H., Lancrin, C., Clarke, D., Williamson, A., Mazzarella, L., Ingram, R., Jorgensen, H., Fisher, A., Tenen, D. G., Kouskoff, V., Lacaud, G. and Bonifer, C. (2009) 'Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program', *Blood*, 114(2), 299-309.

Horan, J. T., Logan, B. R., Agovi-Johnson, M. A., Lazarus, H. M., Bacigalupo, A. A., Ballen, K. K., Bredeson, C. N., Carabasi, M. H., Gupta, V., Hale, G. A., Khoury, H. J., Juckett, M. B., Litzow, M. R., Martino, R., McCarthy, P. L., Smith, F. O., Rizzo, J. D. and Pasquini, M. C. (2011) 'Reducing the risk for transplantation-related mortality after allogeneic hematopoietic cell transplantation: how much progress has been made?', *J Clin Oncol*, 29(7), 805-13.

Hsu, A. P., Johnson, K. D., Falcone, E. L., Sanalkumar, R., Sanchez, L., Hickstein, D. D., Cuellar-Rodriguez, J., Lemieux, J. E., Zerbe, C. S., Bresnick, E. H. and Holland, S. M. (2013) 'GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome', *Blood*, 121(19), 3830-7, S1-7.

Hsu, A. P., Sampaio, E. P., Khan, J., Calvo, K. R., Lemieux, J. E., Patel, S. Y., Frucht, D. M., Vinh, D. C., Auth, R. D., Freeman, A. F., Olivier, K. N., Uzel, G., Zerbe, C. S., Spalding, C., Pittaluga, S., Raffeld, M., Kuhns, D. B., Ding, L., Paulson, M. L., Marciano, B. E., Gea-Banacloche, J. C., Orange, J. S., Cuellar-Rodriguez, J., Hickstein, D. D. and Holland, S. M. (2011) 'Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome.', *Blood*, 118(10), 2653-5.

Huang, G., Zhao, X., Wang, L., Elf, S., Xu, H., Sashida, G., Zhang, Y., Liu, Y., Lee, J., Menendez, S., Yang, Y., Yan, X., Zhang, P., Tenen, D. G., Osato, M., Hsieh, J. J. and Nimer, S. D. (2011) 'The ability of MLL to bind RUNX1 and methylate H3K4 at PU.1 regulatory regions is impaired by MDS/AML-associated RUNX1/AML1 mutations', *Blood*, 118(25), 6544-52.

Hyde, R. K. and Liu, P. P. (2011) 'GATA2 mutations lead to MDS and AML', *Nat Genet*, 43(10), 926-7.

Ichikawa, M., Asai, T., Chiba, S., Kurokawa, M. and Ogawa, S. (2004a) 'Runx1/AML-1 ranks as a master regulator of adult hematopoiesis', *Cell Cycle*, 3(6), 722-4.

Ichikawa, M., Asai, T., Saito, T., Seo, S., Yamazaki, I., Yamagata, T., Mitani, K., Chiba, S., Ogawa, S., Kurokawa, M. and Hirai, H. (2004b) 'AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis', *Nat Med*, 10(3), 299-304.

Ichikawa, M., Goyama, S., Asai, T., Kawazu, M., Nakagawa, M., Takeshita, M., Chiba, S., Ogawa, S. and Kurokawa, M. (2008) 'AML1/Runx1 negatively regulates quiescent hematopoietic stem cells in adult hematopoiesis', *J Immunol*, 180(7), 4402-8.

Illmer, T., Schaich, M., Ehninger, G., Thiede, C. and group, D. A. s. (2007) 'Tyrosine kinase mutations of JAK2 are rare events in AML but influence prognosis of patients with CBF-leukemias', *Haematologica*, 92(1), 137-8.

Imperato, M. R., Cauchy, P., Obier, N. and Bonifer, C. (2015) 'The RUNX1-PU.1 axis in the control of hematopoiesis', *Int J Hematol*, 101(4), 319-29.

Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tomiyama, H., Saito, N., Fukata, M., Miyamoto, T., Lyons, B., Ohshima, K., Uchida, N., Taniguchi, S., Ohara, O., Akashi, K., Harada, M. and Shultz, L. D. (2007) 'Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region', *Nat Biotechnol*, 25(11), 1315-21.

Itzykson, R. and Fenaux, P. (2013) 'Epigenetics of myelodysplastic syndromes', *Leukemia*.

Itzykson, R., Kosmider, O. and Fenaux, P. (2013a) 'Somatic mutations and epigenetic abnormalities in myelodysplastic syndromes', *Best Pract Res Clin Haematol*, 26(4), 355-64.

Itzykson, R., Kosmider, O., Renneville, A., Gelsi-Boyer, V., Meggendorfer, M., Morabito, M., Berthon, C., Adès, L., Fenaux, P., Beyne-Rauzy, O., Vey, N., Braun, T., Haferlach, T., Dreyfus, F., Cross, N. C., Preudhomme, C., Bernard, O. A., Fontenay, M., Vainchenker, W., Schnittger, S., Birnbaum, D., Droin, N. and Solary, E. (2013b) 'Prognostic score including gene mutations in chronic myelomonocytic leukemia', *J Clin Oncol*, 31(19), 2428-36.

Itzykson, R., Kosmider, O., Renneville, A., Morabito, M., Preudhomme, C., Berthon, C., Adès, L., Fenaux, P., Platzbecker, U., Gagey, O., Rameau, P., Meurice, G., Oréar, C., Delhommeau, F., Bernard, O. A., Fontenay, M., Vainchenker, W., Droin, N. and Solary, E. (2013c) 'Clonal architecture of chronic myelomonocytic leukemias', *Blood*, 121(12), 2186-98.

Jaffe ES, et al eds (2001). *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon, France: IARC.

Jan, M., Snyder, T. M., Corces-Zimmerman, M. R., Vyas, P., Weissman, I. L., Quake, S. R. and Majeti, R. (2012) 'Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia', *Sci Transl Med*, 4(149), 149ra118.

Jankowska, A. M., Makishima, H., Tiu, R. V., Szpurka, H., Huang, Y., Traina, F., Visconte, V., Sugimoto, Y., Prince, C., O'Keefe, C., Hsi, E. D., List, A., Sekeres, M. A., Rao, A., McDevitt, M. A. and Maciejewski, J. P. (2011a) 'Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A', *Blood*, 118(14), 3932-41.

Jerez, A., Sugimoto, Y., Makishima, H., Verma, A., Jankowska, A. M., Przychodzen, B., Visconte, V., Tiu, R. V., O'Keefe, C. L., Mohamedali, A. M., Kulasekararaj, A. G., Pellagatti, A., McGraw, K., Muramatsu, H., Moliterno, A. R., Sekeres, M. A., McDevitt, M. A., Kojima, S., List, A., Boultwood, J., Mufti, G. J. and Maciejewski, J. P. (2012) 'Loss of heterozygosity in 7q myeloid disorders: clinical associations and genomic pathogenesis', *Blood*, 119(25), 6109-17.

Johnson, K. D., Hsu, A. P., Ryu, M. J., Wang, J., Gao, X., Boyer, M. E., Liu, Y., Lee, Y., Calvo, K. R., Keles, S., Zhang, J., Holland, S. M. and Bresnick, E. H. (2012) 'Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity', *J Clin Invest*, 122(10), 3692-704.

Jones, A. V., Chase, A., Silver, R. T., Oscier, D., Zoi, K., Wang, Y. L., Cario, H., Pahl, H. L., Collins, A., Reiter, A., Grand, F. and Cross, N. C. (2009) 'JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms', *Nat Genet*, 41(4), 446-9.

Jongmans, M. C., Kuiper, R. P., Carmichael, C. L., Wilkins, E. J., Dors, N., Carmagnac, A., Schouten-van Meeteren, A. Y., Li, X., Stankovic, M., Kamping, E., Bengtsson, H., Schoenmakers, E. F., van Kessel, A. G., Hoogerbrugge, P. M., Hahn, C. N., Brons, P. P., Scott, H. S. and Hoogerbrugge, N. (2010) 'Novel RUNX1 mutations in familial platelet disorder with enhanced risk for acute myeloid leukemia: clues for improved identification of the FPD/AML syndrome', *Leukemia*, 24(1), 242-6.

Jörgensen, P. M., Gräslund, S., Betz, R., Ståhl, S., Larsson, C. and Höög, C. (2001) 'Characterisation of the human APC1, the largest subunit of the anaphase-promoting complex', *Gene*, 262(1-2), 51-9.

Katsumoto, T., Aikawa, Y., Iwama, A., Ueda, S., Ichikawa, H., Ochiya, T. and Kitabayashi, I. (2006) 'MOZ is essential for maintenance of hematopoietic stem cells', *Genes Dev*, 20(10), 1321-30.

Katsumoto, T., Yoshida, N. and Kitabayashi, I. (2008) 'Roles of the histone acetyltransferase monocytic leukemia zinc finger protein in normal and malignant hematopoiesis', *Cancer Sci*, 99(8), 1523-7.

Killick, S. B., Carter, C., Culligan, D., Dalley, C., Das-Gupta, E., Drummond, M., Enright, H., Jones, G. L., Kell, J., Mills, J., Mufti, G., Parker, J., Raj, K., Sternberg, A., Vyas, P., Bowen, D. and Haematology, B. C. f. S. i. (2014) 'Guidelines for the diagnosis and management of adult myelodysplastic syndromes', *Br J Haematol*, 164(4), 503-25.

Kilpivaara, O., Mukherjee, S., Schram, A. M., Wadleigh, M., Mullally, A., Ebert, B. L., Bass, A., Marubayashi, S., Heguy, A., Garcia-Manero, G., Kantarjian, H., Offit, K., Stone, R. M., Gilliland, D. G., Klein, R. J. and Levine, R. L. (2009) 'A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms', *Nat Genet*, 41(4), 455-9.

Kirstetter, P., Schuster, M. B., Bereshchenko, O., Moore, S., Dvinge, H., Kurz, E., Theilgaard-Mönch, K., Månsson, R., Pedersen, T. A., Pabst, T., Schrock, E., Porse, B. T., Jacobsen, S. E., Bertone, P., Tenen, D. G. and Nerlov, C. (2008) 'Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells', *Cancer Cell*, 13(4), 299-310.

Kitabayashi, I., Aikawa, Y., Nguyen, L. A., Yokoyama, A. and Ohki, M. (2001a) 'Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein', *EMBO J*, 20(24), 7184-96.

Kitabayashi, I., Aikawa, Y., Yokoyama, A., Hosoda, F., Nagai, M., Kakazu, N., Abe, T. and Ohki, M. (2001b) 'Fusion of MOZ and p300 histone acetyltransferases in acute monocytic leukemia with a t(8;22)(p11;q13) chromosome translocation', *Leukemia*, 15(1), 89-94.

Kitabayashi, I., Yokoyama, A., Shimizu, K. and Ohki, M. (1998) 'Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation', *EMBO J*, 17(11), 2994-3004.

Klco, J. M., Spencer, D. H., Miller, C. A., Griffith, M., Lamprecht, T. L., O'Laughlin, M., Fronick, C., Magrini, V., Demeter, R. T., Fulton, R. S., Eades, W. C., Link, D. C., Graubert, T. A., Walter, M. J., Mardis, E. R., Dipersio, J. F., Wilson, R. K. and Ley, T. J. (2014) 'Functional heterogeneity of genetically defined subclones in acute myeloid leukemia', *Cancer Cell*, 25(3), 379-92.

Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., An, J., Lamperti, E. D., Koh, K. P., Ganetzky, R., Liu, X. S., Aravind, L., Agarwal, S., Maciejewski, J. P. and Rao, A. (2010) 'Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2', *Nature*, 468(7325), 839-43.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L. and Wilson, R. K. (2012) 'VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing', *Genome Res*, 22(3), 568-76.

Kon, A., Shih, L. Y., Minamino, M., Sanada, M., Shiraishi, Y., Nagata, Y., Yoshida, K., Okuno, Y., Bando, M., Nakato, R., Ishikawa, S., Sato-Otsubo, A., Nagae, G., Nishimoto, A., Haferlach, C., Nowak, D., Sato, Y., Alpermann, T., Nagasaki, M., Shimamura, T., Tanaka, H., Chiba, K., Yamamoto, R., Yamaguchi, T., Otsu, M., Obara, N., Sakata-Yanagimoto, M., Nakamaki, T., Ishiyama, K., Nolte, F., Hofmann, W. K., Miyawaki, S., Chiba, S., Mori, H., Nakauchi, H., Koeffler, H. P., Aburatani, H., Haferlach, T., Shirahige, K., Miyano, S. and Ogawa, S. (2013) 'Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms', *Nat Genet*, 45(10), 1232-7.

Konopleva, M. Y. and Jordan, C. T. (2011) 'Leukemia stem cells and microenvironment: biology and therapeutic targeting', *J Clin Oncol*, 29(5), 591-9.

Kottaridis, P. D., Gale, R. E., Frew, M. E., Harrison, G., Langabeer, S. E., Belton, A. A., Walker, H., Wheatley, K., Bowen, D. T., Burnett, A. K., Goldstone, A. H. and Linch, D. C. (2001) 'The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials', *Blood*, 98(6), 1752-9.

Kralovics, R., Teo, S. S., Buser, A. S., Brutsche, M., Tiedt, R., Tichelli, A., Passamonti, F., Pietra, D., Cazzola, M. and Skoda, R. C. (2005) 'Altered gene expression in myeloproliferative

disorders correlates with activation of signaling by the V617F mutation of Jak2', *Blood*, 106(10), 3374-6.

Krauth, M. T., Eder, C., Alpermann, T., Bacher, U., Nadarajah, N., Kern, W., Haferlach, C., Haferlach, T. and Schnittger, S. (2014) 'High number of additional genetic lesions in acute myeloid leukemia with t(8;21)/RUNX1-RUNX1T1: frequency and impact on clinical outcome', *Leukemia*, 28(7), 1449-58.

Kunishima, S., Kobayashi, R., Itoh, T. J., Hamaguchi, M. and Saito, H. (2009) 'Mutation of the beta1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly', *Blood*, 113(2), 458-61.

Kunishima, S., Nishimura, S., Suzuki, H., Imaizumi, M. and Saito, H. (2014) 'TUBB1 mutation disrupting microtubule assembly impairs proplatelet formation and results in congenital macrothrombocytopenia', *Eur J Haematol*, 92(4), 276-82.

Kurzer, J. H., Saharinen, P., Silvennoinen, O. and Carter-Su, C. (2006) 'Binding of SH2-B family members within a potential negative regulatory region maintains JAK2 in an active state', *Mol Cell Biol*, 26(17), 6381-94.

Kwiatkowski, B. A., Bastian, L. S., Bauer, T. R., Tsai, S., Zielinska-Kwiatkowska, A. G. and Hickstein, D. D. (1998) 'The ets family member Tel binds to the Fli-1 oncoprotein and inhibits its transcriptional activity', *J Biol Chem*, 273(28), 17525-30.

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M. and Maglott, D. R. (2014) 'ClinVar: public archive of relationships among sequence variation and human phenotype', *Nucleic Acids Res*, 42(Database issue), D980-5.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nat Methods*, 9(4), 357-9.

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A. and Dick, J. E. (1994) 'A cell initiating human acute myeloid leukaemia after transplantation into SCID mice', *Nature*, 367(6464), 645-8.

Lee, S. J., Klein, J., Haagenson, M., Baxter-Lowe, L. A., Confer, D. L., Eapen, M., Fernandez-Vina, M., Flomenberg, N., Horowitz, M., Hurley, C. K., Noreen, H., Oudshoorn, M., Petersdorf, E., Setterholm, M., Spellman, S., Weisdorf, D., Williams, T. M. and Anasetti, C. (2007) 'High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation', *Blood*, 110(13), 4576-83.

Levanon, D., Bernstein, Y., Negreanu, V., Ghozi, M. C., Bar-Am, I., Aloya, R., Goldenberg, D., Lotem, J. and Groner, Y. (1996) 'A large variety of alternatively spliced and differentially expressed mRNAs are encoded by the human acute myeloid leukemia gene AML1', *DNA Cell Biol*, 15(3), 175-85.

Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., Harris, C. C., Lichti, C. F., Townsend, R. R., Fulton, R. S., Dooling, D. J., Koboldt, D. C., Schmidt, H., Zhang, Q., Osborne, J. R., Lin, L., O'Laughlin, M., McMichael, J. F., Delehaunty, K. D., McGrath, S. D., Fulton, L. A., Magrini, V. J., Vickery, T. L., Hundal, J., Cook, L. L., Conyers, J. J., Swift, G. W., Reed, J. P., Alldredge, P. A., Wylie, T., Walker, J., Kalicki, J., Watson, M. A., Heath, S., Shannon, W. D., Varghese, N., Nagarajan, R., Westervelt, P., Tomasson, M. H., Link, D. C., Graubert, T. A., DiPersio, J. F., Mardis, E. R. and Wilson, R. K. (2010) 'DNMT3A mutations in acute myeloid leukemia', *N Engl J Med*, 363(25), 2424-33.

Ley, T. J., Mardis, E. R., Ding, L., Fulton, B., McLellan, M. D., Chen, K., Dooling, D., Dunford-Shore, B. H., McGrath, S., Hickenbotham, M., Cook, L., Abbott, R., Larson, D. E., Koboldt, D. C., Pohl, C., Smith, S., Hawkins, A., Abbott, S., Locke, D., Hillier, L. W., Miner, T., Fulton, L., Magrini, V., Wylie, T., Glasscock, J., Conyers, J., Sander, N., Shi, X., Osborne, J. R., Minx, P., Gordon, D., Chinwalla, A., Zhao, Y., Ries, R. E., Payton, J. E., Westervelt, P., Tomasson, M. H., Watson, M., Baty, J., Ivanovich, J., Heath, S., Shannon, W. D., Nagarajan, R., Walter, M. J., Link, D. C., Graubert, T. A., DiPersio, J. F. and Wilson, R. K. (2008) 'DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome', *Nature*, 456(7218), 66-72.

Liang, D. C., Shih, L. Y., Huang, C. F., Hung, I. J., Yang, C. P., Liu, H. C., Jaing, T. H., Wang, L. Y. and Chang, W. H. (2005) 'CEBPalpha mutations in childhood acute myeloid leukemia', *Leukemia*, 19(3), 410-4.

Liang, J., Prouty, L., Williams, B. J., Dayton, M. A. and Blanchard, K. L. (1998) 'Acute mixed lineage leukemia with an inv(8)(p11q13) resulting in fusion of the genes for MOZ and TIF2', *Blood*, 92(6), 2118-22.

Liew, E. and Owen, C. (2011) 'Familial myelodysplastic syndromes: a review of the literature', *Haematologica*, 96(10), 1536-42.

Lin, T. C., Hou, H. A., Chou, W. C., Ou, D. L., Yu, S. L., Tien, H. F. and Lin, L. I. (2011) 'CEBPA methylation as a prognostic biomarker in patients with de novo acute myeloid leukemia', *Leukemia*, 25(1), 32-40.

Lin, W., Wu, X. and Wang, Z. (1999) 'A full-length cDNA of hREV3 is predicted to encode DNA polymerase zeta for damage-induced mutagenesis in humans', *Mutat Res*, 433(2), 89-98.

Lindsley, R. C. and Ebert, B. L. (2013) 'The biology and clinical impact of genetic lesions in myeloid malignancies', *Blood*, 122(23), 3741-8.

List, A., Kurtin, S., Roe, D. J., Buresh, A., Mahadevan, D., Fuchs, D., Rimsza, L., Heaton, R., Knight, R. and Zeldis, J. B. (2005) 'Efficacy of lenalidomide in myelodysplastic syndromes', *N Engl J Med*, 352(6), 549-57.

Loh, P. G., Yang, H. S., Walsh, M. A., Wang, Q., Wang, X., Cheng, Z., Liu, D. and Song, H. (2009) 'Structural basis for translational inhibition by the tumour suppressor Pdcd4', *EMBO J*, 28(3), 274-85.

LORENZ, E., UPHOFF, D., REID, T. R. and SHELTON, E. (1951) 'Modification of irradiation injury in mice and guinea pigs by bone marrow injections', *J Natl Cancer Inst*, 12(1), 197-201.

Lu, C., Ward, P. S., Kapoor, G. S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C. R., Khanin, R., Figueroa, M. E., Melnick, A., Wellen, K. E., O'Rourke, D. M., Berger, S. L., Chan, T. A., Levine, R. L., Mellinghoff, I. K. and Thompson, C. B. (2012) 'IDH mutation impairs histone demethylation and results in a block to cell differentiation', *Nature*, 483(7390), 474-8.

Lubas, M., Christensen, M. S., Kristiansen, M. S., Domanski, M., Falkenby, L. G., Lykke-Andersen, S., Andersen, J. S., Dziembowski, A. and Jensen, T. H. (2011) 'Interaction profiling identifies the human nuclear exosome targeting complex', *Mol Cell*, 43(4), 624-37.

Makishima, H., Yoshida, K., Nguyen, N., Przychodzen, B., Sanada, M., Okuno, Y., Ng, K. P., Gudmundsson, K. O., Vishwakarma, B. A., Jerez, A., Gomez-Segui, I., Takahashi, M., Shiraishi, Y., Nagata, Y., Guinta, K., Mori, H., Sekeres, M. A., Chiba, K., Tanaka, H., Muramatsu, H., Sakaguchi, H., Paquette, R. L., McDevitt, M. A., Kojima, S., Sauntharajah, Y., Miyano, S., Shih, L. Y., Du, Y., Ogawa, S. and Maciejewski, J. P. (2013) 'Somatic SETBP1 mutations in myeloid malignancies', *Nat Genet*, 45(8), 942-6.

Malcovati, L., Papaemmanuil, E., Ambaglio, I., Elena, C., Gallì, A., Della Porta, M. G., Travaglino, E., Pietra, D., Pascutto, C., Ubezio, M., Bono, E., Da Vià, M. C., Brisci, A., Bruno, F., Cremonesi, L., Ferrari, M., Boveri, E., Invernizzi, R., Campbell, P. J. and Cazzola, M. (2014) 'Driver somatic mutations identify distinct disease entities within myeloid neoplasms with myelodysplasia', *Blood*, 124(9), 1513-21.

Maloney, K., McGavran, L., Murphy, J., Odom, L., Stork, L., Wei, Q. and Hunger, S. (1999) 'TEL-AML1 fusion identifies a subset of children with standard risk acute lymphoblastic leukemia who have an excellent prognosis when treated with therapy that includes a single delayed intensification', *Leukemia*, 13(11), 1708-12.

Marcucci, G., Haferlach, T. and Döhner, H. (2011) 'Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications', *J Clin Oncol*, 29(5), 475-86.

Marcucci, G., Maharry, K., Wu, Y. Z., Radmacher, M. D., Mrózek, K., Margeson, D., Holland, K. B., Whitman, S. P., Becker, H., Schwind, S., Metzeler, K. H., Powell, B. L., Carter, T. H., Kolitz, J. E., Wetzler, M., Carroll, A. J., Baer, M. R., Caligiuri, M. A., Larson, R. A. and Bloomfield, C. D. (2010) 'IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study', *J Clin Oncol*, 28(14), 2348-55.

Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., McLellan, M. D., Chen, K., Koboldt, D. C., Fulton, R. S., Delehaunty, K. D., McGrath, S. D., Fulton, L. A., Locke, D. P., Magrini, V. J., Abbott, R. M., Vickery, T. L., Reed, J. S., Robinson, J. S., Wylie, T., Smith, S. M., Carmichael, L., Eldred, J. M., Harris, C. C., Walker, J., Peck, J. B., Du, F., Dukes, A. F., Sanderson, G. E., Brummett, A. M., Clark, E., McMichael, J. F., Meyer, R. J., Schindler, J. K., Pohl, C. S., Wallis, J. W., Shi, X., Lin, L., Schmidt, H., Tang, Y., Haipek, C., Wiechert, M. E., Ivy, J. V., Kalicki, J., Elliott, G., Ries, R. E., Payton, J. E., Westervelt, P., Tomasson, M. H., Watson, M. A., Baty, J., Heath,

S., Shannon, W. D., Nagarajan, R., Link, D. C., Walter, M. J., Graubert, T. A., DiPersio, J. F., Wilson, R. K. and Ley, T. J. (2009) 'Recurring mutations found by sequencing an acute myeloid leukemia genome.', *N Engl J Med*, 361(11), 1058-66.

Martelotto, L. G., Ng, C. K., De Filippo, M. R., Zhang, Y., Piscuoglio, S., Lim, R. S., Shen, R., Norton, L., Reis-Filho, J. S. and Weigelt, B. (2014) 'Benchmarking mutation effect prediction algorithms using functionally validated cancer-related missense mutations', *Genome Biol*, 15(10), 484.

Maximow A. The Lymphocyte as a stem cell common to different blood elements in embryonic development and during the post-fetal life of mammals (1909). Originally in German: *Folia Haematologica* 8.1909, 125-134. English translation: *Cell Ther Transplant*. 2009,1:e.000032.01. doi:10.3205/ctt-2009-en-000032.01

Maxson, J. E., Gotlib, J., Pollyea, D. A., Fleischman, A. G., Agarwal, A., Eide, C. A., Bottomly, D., Wilmot, B., McWeeney, S. K., Tognon, C. E., Pond, J. B., Collins, R. H., Goueli, B., Oh, S. T., Deininger, M. W., Chang, B. H., Loriaux, M. M., Druker, B. J. and Tyner, J. W. (2013) 'Oncogenic CSF3R mutations in chronic neutrophilic leukemia and atypical CML', *N Engl J Med*, 368(19), 1781-90.

Medeiros, B. C., Minden, M. D., Schuh, A. C., Schimmer, A. D., Yee, K., Lipton, J. H., Messner, H. A., Gupta, V., Chun, K., Xu, W., Das, P., Kamel-Reid, S. and Brandwein, J. M. (2007) 'Characteristics and outcomes of acute myelogenous leukemia patients with very late relapse (>5 years)', *Leuk Lymphoma*, 48(1), 65-71.

Medves, S., Duhoux, F. P., Ferrant, A., Toffalini, F., Ameye, G., Libouton, J. M., Poirel, H. A. and Demoulin, J. B. (2010) 'KANK1, a candidate tumor suppressor gene, is fused to PDGFRB in an imatinib-responsive myeloid neoplasm with severe thrombocythemia', *Leukemia*, 24(5), 1052-5.

Mendler, J. H., Maharry, K., Radmacher, M. D., Mrózek, K., Becker, H., Metzeler, K. H., Schwind, S., Whitman, S. P., Khalife, J., Kohlschmidt, J., Nicolet, D., Powell, B. L., Carter, T. H., Wetzler, M., Moore, J. O., Kolitz, J. E., Baer, M. R., Carroll, A. J., Larson, R. A., Caligiuri, M. A., Marcucci, G. and Bloomfield, C. D. (2012) 'RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and MicroRNA expression signatures', *J Clin Oncol*, 30(25), 3109-18.

Meshinchi, S., Alonzo, T. A., Stirewalt, D. L., Zwaan, M., Zimmerman, M., Reinhardt, D., Kaspers, G. J., Heerema, N. A., Gerbing, R., Lange, B. J. and Radich, J. P. (2006) 'Clinical implications of FLT3 mutations in pediatric AML', *Blood*, 108(12), 3654-61.

Metzeler, K. H., Maharry, K., Radmacher, M. D., Mrózek, K., Margeson, D., Becker, H., Curfman, J., Holland, K. B., Schwind, S., Whitman, S. P., Wu, Y. Z., Blum, W., Powell, B. L., Carter, T. H., Wetzler, M., Moore, J. O., Kolitz, J. E., Baer, M. R., Carroll, A. J., Larson, R. A., Caligiuri, M. A., Marcucci, G. and Bloomfield, C. D. (2011) 'TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study', *J Clin Oncol*, 29(10), 1373-81.

Micale, L., Augello, B., Maffeo, C., Selicorni, A., Zucchetti, F., Fusco, C., De Nittis, P., Pellico, M. T., Mandriani, B., Fischetto, R., Boccone, L., Silengo, M., Biamino, E., Perria, C., Sotgiu, S., Serra, G., Lapi, E., Neri, M., Ferlini, A., Cavaliere, M. L., Chiurazzi, P., Monica, M. D., Scarano, G., Faravelli, F., Ferrari, P., Mazzanti, L., Pilotta, A., Patricelli, M. G., Bedeschi, M. F., Benedicenti, F., Prontera, P., Toschi, B., Salviati, L., Melis, D., Di Battista, E., Vancini, A., Garavelli, L., Zelante, L. and Merla, G. (2014) 'Molecular analysis, pathogenic mechanisms, and readthrough therapy on a large cohort of Kabuki syndrome patients', *Hum Mutat*, 35(7), 841-50.

Michaud, J., Wu, F., Osato, M., Cottles, G. M., Yanagida, M., Asou, N., Shigesada, K., Ito, Y., Benson, K. F., Raskind, W. H., Rossier, C., Antonarakis, S. E., Israels, S., McNicol, A., Weiss, H., Horwitz, M. and Scott, H. S. (2002) 'In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis', *Blood*, 99(4), 1364-72.

Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D. and Hess, J. L. (2002) 'MLL targets SET domain methyltransferase activity to Hox gene promoters', *Mol Cell*, 10(5), 1107-17.

Miyoshi, H., Ohira, M., Shimizu, K., Mitani, K., Hirai, H., Imai, T., Yokoyama, K., Soeda, E. and Ohki, M. (1995) 'Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia', *Nucleic Acids Res*, 23(14), 2762-9.

Moldovan, G. L. and D'Andrea, A. D. (2009) 'How the fanconi anemia pathway guards the genome', *Annu Rev Genet*, 43, 223-49.

Morin, R. D., Johnson, N. A., Severson, T. M., Mungall, A. J., An, J., Goya, R., Paul, J. E., Boyle, M., Woolcock, B. W., Kuchenbauer, F., Yap, D., Humphries, R. K., Griffith, O. L., Shah, S., Zhu, H., Kimbara, M., Shashkin, P., Charlot, J. F., Tcherpakov, M., Corbett, R., Tam, A., Varhol, R., Smailus, D., Moksa, M., Zhao, Y., Delaney, A., Qian, H., Birol, I., Schein, J., Moore, R., Holt, R., Horsman, D. E., Connors, J. M., Jones, S., Aparicio, S., Hirst, M., Gascoyne, R. D. and Marra, M. A. (2010) 'Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin', *Nat Genet*, 42(2), 181-5.

Murati, A., Brecqueville, M., Devillier, R., Mozziconacci, M. J., Gelsi-Boyer, V. and Birnbaum, D. (2012) 'Myeloid malignancies: mutations, models and management', *BMC Cancer*, 12, 304.

Mutsaers, P. G., van de Loosdrecht, A. A., Tawana, K., Bödör, C., Fitzgibbon, J. and Menko, F. H. (2013b) 'Highly variable clinical manifestations in a large family with a novel GATA2 mutation', *Leukemia*, 27(11), 2247-8.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M. and Canaani, E. (2002) 'ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation', *Mol Cell*, 10(5), 1119-28.

Nakao, M., Yokota, S., Iwai, T., Kaneko, H., Horiike, S., Kashima, K., Sonoda, Y., Fujimoto, T. and Misawa, S. (1996) 'Internal tandem duplication of the flt3 gene found in acute myeloid leukemia', *Leukemia*, 10(12), 1911-8.

Nanri, T., Uike, N., Kawakita, T., Iwanaga, E., Mitsuya, H. and Asou, N. (2010) '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*, 49(3), 237-41.

Nerlov, C. (2007) 'The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control', *Trends Cell Biol*, 17(7), 318-24.

Network, C. G. A. (2012) 'Comprehensive molecular characterization of human colon and rectal cancer', *Nature*, 487(7407), 330-7.

Network, T. C. G. A. R. (2013a) 'Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia', *N Engl J Med*.

Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., Lee, C., Turner, E. H., Smith, J. D., Rieder, M. J., Yoshiura, K., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D. A., Bamshad, M. J. and Shendure, J. (2010b) 'Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome', *Nat Genet*, 42(9), 790-3.

Nguyen, L. V., Vanner, R., Dirks, P. and Eaves, C. J. (2012) 'Cancer stem cells: an evolving concept', *Nat Rev Cancer*, 12(2), 133-43.

NICE technology appraisal guidance TA218, March 2011, updated February 2014. Azacitidine for the treatment of myelodysplastic syndromes, chronic myelomonocytic leukaemia and acute myeloid leukaemia.

Nickels, E. M., Soodalter, J., Churpek, J. E. and Godley, L. A. (2013) 'Recognizing familial myeloid leukemia in adults', *Ther Adv Hematol*, 4(4), 254-69.

Noetzli, L., Lo, R. W., Lee-Sherick, A. B., Callaghan, M., Noris, P., Savoia, A., Rajpurkar, M., Jones, K., Gowan, K., Balduini, C. L., 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, F. G., Rowley, J. W., Weyrich, A. S., Kahr, W. H., Porter, C. C. and Di Paola, J. (2015) 'Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia', *Nat Genet*, 47(5), 535-8.

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, L. D., Klersy, C., Civaschi, E., Balduini, C. L. and Savoia, A. (2011) 'Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families', *Blood*, 117(24), 6673-80.

Nowell, P. C. (1976) 'The clonal evolution of tumor cell populations', *Science*, 194(4260), 23-8.

Ofran, Y. and Rowe, J. M. (2012) 'Treatment for relapsed acute myeloid leukemia: what is new?', *Curr Opin Hematol*, 19(2), 89-94.

Oh, S. T., Simonds, E. F., Jones, C., Hale, M. B., Goltsev, Y., Gibbs, K. D., Merker, J. D., Zehnder, J. L., Nolan, G. P. and Gotlib, J. (2010) 'Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms', *Blood*, 116(6), 988-92.

Okosun, J., Bödör, C., Wang, J., Araf, S., Yang, C. Y., Pan, C., Boller, S., Cittaro, D., Bozek, M., Iqbal, S., Matthews, J., Wrench, D., Marzec, J., Tawana, K., Popov, N., O'Riain, C., O'Shea, D., Carlotti, E., Davies, A., Lawrie, C. H., Matolcsy, A., Calaminici, M., Norton, A., Byers, R. J., Mein, C., Stupka, E., Lister, T. A., Lenz, G., Montoto, S., Gribben, J. G., Fan, Y., Grosschedl, R., Chelala, C. and Fitzgibbon, J. (2014) 'Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma', *Nat Genet*, 46(2), 176-81.

Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G. and Downing, J. R. (1996a) 'AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis', *Cell*, 84(2), 321-30.

Olcaydu, D., Harutyunyan, A., Jäger, R., Berg, T., Gisslinger, B., Pabinger, I., Gisslinger, H. and Kralovics, R. (2009) 'A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms', *Nat Genet*, 41(4), 450-4.

Ortmann, C. A., Kent, D. G., Nangalia, J., Silber, Y., Wedge, D. C., Grinfeld, J., Baxter, E. J., Massie, C. E., Papaemmanuil, E., Menon, S., Godfrey, A. L., Dimitropoulou, D., Guglielmelli, P., Bellosillo, B., Besses, C., Döhner, K., Harrison, C. N., Vassiliou, G. S., Vannucchi, A., Campbell, P. J. and Green, A. R. (2015) 'Effect of mutation order on myeloproliferative neoplasms', *N Engl J Med*, 372(7), 601-12.

Ostergaard, P., Simpson, M. A., Connell, F. C., Steward, C. G., Brice, G., Woollard, W. J., Dafou, D., Kilo, T., Smithson, S., Lunt, P., Murday, V. A., Hodgson, S., Keenan, R., Pilz, D. T., Martinez-Corral, I., Makinen, T., Mortimer, P. S., Jeffery, S., Trembath, R. C. and Mansour, S. (2011) 'Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome).', *Nat Genet*, 43(10), 929-31.

Owen, C. J., Toze, C. L., Koochin, A., Forrest, D. L., Smith, C. A., Stevens, J. M., Jackson, S. C., Poon, M. C., Sinclair, G. D., Leber, B., Johnson, P. R., Macheta, A., Yin, J. A., Barnett, M. J., Lister, T. A. and Fitzgibbon, J. (2008) 'Five new pedigrees with inherited RUNX1 mutations causing familial platelet disorder with propensity to myeloid malignancy.', *Blood*, 112(12), 4639-45.

Pabst, T., Eyholzer, M., Haefliger, S., Schardt, J. and Mueller, B. U. (2008) 'Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia', *J Clin Oncol*, 26(31), 5088-93.

Pabst, T., Mueller, B. U., Zhang, P., Radomska, H. S., Narravula, S., Schnittger, S., Behre, G., Hiddemann, W. and Tenen, D. G. (2001) 'Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia', *Nat Genet*, 27(3), 263-70.

Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J. S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., Godfrey, A. L., Rapado, I., Cvejic, A., Rance, R., McGee, C., Ellis, P., Mudie, L. J., Stephens, P. J., McLaren, S., Massie, C. E., Tarpey, P. S., Varela, I., Nik-Zainal, S., Davies, H. R., Shlien, A., Jones, D., Raine, K., Hinton, J., Butler, A. P., Teague, J. W., Baxter, E. J., Score, J., Galli, A., Della Porta, M. G., Travaglino, E., Groves, M., Tauro, S., Munshi, N. C., Anderson, K. C., El-Naggar, A., Fischer, A., Mustonen, V., Warren, A. J., Cross, N. C., Green, A. R., Futreal, P. A., Stratton, M. R., Campbell, P. J. and Consortium, C. M. D. W. G. o. t. I. C. G. (2011) 'Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts', *N Engl J Med*, 365(15), 1384-95.

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C. J., Ellis, P., Wedge, D. C., Pellagatti, A., Shlien, A., Groves, M. J., Forbes, S. A., Raine, K., Hinton, J., Mudie, L. J., McLaren, S., Hardy, C., Latimer, C., Della Porta, M. G., O'Meara, S., Ambaglio, I., Galli, A., Butler, A. P., Walldin, G., Teague, J. W., Quek, L., Sternberg, A., Gambacorti-Passerini, C., Cross, N. C., Green, A. R., Boultwood, J., Vyas, P., Hellstrom-Lindberg, E., Bowen, D., Cazzola, M., Stratton, M. R., Campbell, P. J. and Consortium, C. M. D. W. G. o. t. I. C. G. (2013) 'Clinical and biological implications of driver mutations in myelodysplastic syndromes', *Blood*, 122(22), 3616-27; quiz 3699.

Pardanani, A., Lasho, T. L., Laborde, R. R., Elliott, M., Hanson, C. A., Knudson, R. A., Ketterling, R. P., Maxson, J. E., Tyner, J. W. and Tefferi, A. (2013) 'CSF3R T618I is a highly prevalent and specific mutation in chronic neutrophilic leukemia', *Leukemia*, 27(9), 1870-3.

Pardanani, A., Lasho, T., Finke, C., Oh, S. T., Gotlib, J. and Tefferi, A. (2010) 'LNK mutation studies in blast-phase myeloproliferative neoplasms, and in chronic-phase disease with TET2, IDH, JAK2 or MPL mutations', *Leukemia*, 24(10), 1713-8.

Pasqualucci, L., Trifonov, V., Fabbri, G., Ma, J., Rossi, D., Chiarenza, A., Wells, V. A., Grunn, A., Messina, M., Elliot, O., Chan, J., Bhagat, G., Chadburn, A., Gaidano, G., Mullighan, C. G., Rabadan, R. and Dalla-Favera, R. (2011) 'Analysis of the coding genome of diffuse large B-cell lymphoma', *Nat Genet*, 43(9), 830-7.

Pasquet, M., Bellanné-Chantelot, C., Tavitian, S., Prade, N., Beaupain, B., Larochelle, O., Petit, A., Rohrlich, P., Ferrand, C., Van Den Neste, E., Poirel, H. A., Lamy, T., Ouachée-Chardin, M., Mansat-De Mas, V., Corre, J., Récher, C., Plat, G., Bachelerie, F., Donadieu, J. and Delabesse, E. (2013) 'High frequency of GATA2 mutations in patients with mild chronic neutropenia

evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia', *Blood*, 121(5), 822-9.

Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C. and Weissman, I. L. (2005) 'Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates', *J Exp Med*, 202(11), 1599-611.

Patel, J. P., Gönen, M., Figueroa, M. E., Fernandez, H., Sun, Z., Racevskis, J., Van Vlierberghe, P., Dolgalev, I., Thomas, S., Aminova, O., Huberman, K., Cheng, J., Viale, A., Socci, N. D., Heguy, A., Cherry, A., Vance, G., Higgins, R. R., Ketterling, R. P., Gallagher, R. E., Litzow, M., van den Brink, M. R., Lazarus, H. M., Rowe, J. M., Luger, S., Ferrando, A., Paietta, E., Tallman, M. S., Melnick, A., Abdel-Wahab, O. and Levine, R. L. (2012) 'Prognostic relevance of integrated genetic profiling in acute myeloid leukemia', *N Engl J Med*, 366(12), 1079-89.

Pelletier, N., Champagne, N., Stifani, S. and Yang, X. J. (2002) 'MOZ and MORF histone acetyltransferases interact with the Runt-domain transcription factor Runx2', *Oncogene*, 21(17), 2729-40.

Perez-Garcia, A., Ambesi-Impiombato, A., Hadler, M., Rigo, I., LeDuc, C. A., Kelly, K., Jalas, C., Paietta, E., Racevskis, J., Rowe, J. M., Tallman, M. S., Paganin, M., Basso, G., Tong, W., Chung, W. K. and Ferrando, A. A. (2013) 'Genetic loss of SH2B3 in acute lymphoblastic leukemia', *Blood*, 122(14), 2425-32.

Personnic, N., Lakisic, G., Gouin, E., Rousseau, A., Gautreau, A., Cossart, P. and Bierne, H. (2014) 'A role for Ral GTPase-activating protein subunit β in mitotic regulation', *FEBS J*, 281(13), 2977-89.

Peterson, L. F. and Zhang, D. E. (2004) 'The 8;21 translocation in leukemogenesis', *Oncogene*, 23(24), 4255-62.

Piazza, R., Valletta, S., Winkelmann, N., Redaelli, S., Spinelli, R., Pirola, A., Antolini, L., Mologni, L., Donadoni, C., Papaemmanuil, E., Schnittger, S., Kim, D. W., Boultwood, J., Rossi, F., Gaipa, G., De Martini, G. P., di Celle, P. F., Jang, H. G., Fantin, V., Bignell, G. R., Magistroni, V., Haferlach, T., Pogliani, E. M., Campbell, P. J., Chase, A. J., Tapper, W. J., Cross, N. C. and Gambacorti-Passerini, C. (2013) 'Recurrent SETBP1 mutations in atypical chronic myeloid leukemia', *Nat Genet*, 45(1), 18-24.

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. and Balduini, C. L. (2011) '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*, 88(1), 115-20.

Pløen, G. G., Nederby, L., Guldberg, P., Hansen, M., Ebbesen, L. H., Jensen, U. B., Hokland, P. and Aggerholm, A. (2014) 'Persistence of DNMT3A mutations at long-term remission in adult patients with AML', *Br J Haematol*, 167(4), 478-86.

Polprasert, C., Schulze, I., Sekeres, M. A., Makishima, H., Przychodzen, B., Hosono, N., Singh, J., Padgett, R. A., Gu, X., Phillips, J. G., Clemente, M., Parker, Y., Lindner, D., Dienes, B., Jankowsky, E., Sauntharajah, Y., Du, Y., Oakley, K., Nguyen, N., Mukherjee, S., Pabst, C., Godley, L. A., Churpek, J. E., Pollyea, D. A., Krug, U., Berdel, W. E., Klein, H. U., Dugas, M., Shiraishi, Y., Chiba, K., Tanaka, H., Miyano, S., Yoshida, K., Ogawa, S., Müller-Tidow, C. and Maciejewski, J. P. (2015) 'Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms', *Cancer Cell*, 27(5), 658-70.

Prebet, T., Carbuccia, N., Raslova, H., Favier, R., Rey, J., Arnoulet, C., Vey, N., Vainchenker, W., Birnbaum, D. and Mozziconacci, M. J. (2013) 'Concomitant germ-line RUNX1 and acquired ASXL1 mutations in a T-cell acute lymphoblastic leukemia', *Eur J Haematol*, 91(3), 277-9.

Preudhomme, C., Renneville, A., Bourdon, V., Philippe, N., Roche-Lestienne, C., Boissel, N., Dhedin, N., André, J. M., Cornillet-Lefebvre, P., Baruchel, A., Mozziconacci, M. J. and Sobol, H. (2009) 'High frequency of RUNX1 biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder', *Blood*, 113(22), 5583-7.

Quintás-Cardama, A. and Verstovsek, S. (2013) 'Molecular pathways: Jak/STAT pathway: mutations, inhibitors, and resistance', *Clin Cancer Res*, 19(8), 1933-40.

Raghavan, M., Smith, L. L., Lillington, D. M., Chaplin, T., Kakkas, I., Molloy, G., Chelala, C., Cazier, J. B., Cavenagh, J. D., Fitzgibbon, J., Lister, T. A. and Young, B. D. (2008) 'Segmental uniparental disomy is a commonly acquired genetic event in relapsed acute myeloid leukemia', *Blood*, 112(3), 814-21.

Rampal, R., Alkalin, A., Madzo, J., Vasanthakumar, A., Pronier, E., Patel, J., Li, Y., Ahn, J., Abdel-Wahab, O., Shih, A., Lu, C., Ward, P. S., Tsai, J. J., Hricik, T., Tosello, V., Tallman, J. E., Zhao, X., Daniels, D., Dai, Q., Ciminio, L., Aifantis, I., He, C., Fuks, F., Tallman, M. S., Ferrando, A., Nimer, S., Paietta, E., Thompson, C. B., Licht, J. D., Mason, C. E., Godley, L. A., Melnick, A., Figueroa, M. E. and Levine, R. L. (2014) 'DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia', *Cell Rep*, 9(5), 1841-55.

Reed-Inderbitzin, E., Moreno-Miralles, I., Vanden-Eynden, S. K., Xie, J., Lutterbach, B., Durst-Goodwin, K. L., Luce, K. S., Irvin, B. J., Cleary, M. L., Brandt, S. J. and Hiebert, S. W. (2006) 'RUNX1 associates with histone deacetylases and SUV39H1 to repress transcription', *Oncogene*, 25(42), 5777-86.

Reid, S., Renwick, A., Seal, S., Baskcomb, L., Barfoot, R., Jayatilake, H., Pritchard-Jones, K., Stratton, M. R., Ridolfi-Lüthy, A., Rahman, N., (UK), B. C. S. C. and Collaboration, F. W. T. (2005) 'Biallelic BRCA2 mutations are associated with multiple malignancies in childhood including familial Wilms tumour', *J Med Genet*, 42(2), 147-51.

Reid, S., Schindler, D., Hanenberg, H., Barker, K., Hanks, S., Kalb, R., Neveling, K., Kelly, P., Seal, S., Freund, M., Wurm, M., Batish, S. D., Lach, F. P., Yetgin, S., Neitzel, H., Ariffin, H., Tischkowitz, M., Mathew, C. G., Auerbach, A. D. and Rahman, N. (2007) 'Biallelic mutations

in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer', *Nat Genet*, 39(2), 162-4.

Renneville, A., Mialou, V., Philippe, N., Kagialis-Girard, S., Biggio, V., Zabot, M. T., Thomas, X., Bertrand, Y. and Preudhomme, C. (2009) 'Another pedigree with familial acute myeloid leukemia and germline CEBPA mutation', *Leukemia*, 23(4), 804-6.

Ribeiro, A. F., Pratcorona, M., Erpelinck-Verschueren, C., Rockova, V., Sanders, M., Abbas, S., Figueroa, M. E., Zeilemaker, A., Melnick, A., Löwenberg, B., Valk, P. J. and Delwel, R. (2012) 'Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia', *Blood*, 119(24), 5824-31.

Riccioni, R., Lulli, V., Castelli, G., Biffoni, M., Tiberio, R., Pelosi, E., Lo-Coco, F. and Testa, U. (2015) 'miR-21 is overexpressed in NPM1-mutant acute myeloid leukemias', *Leuk Res*, 39(2), 221-8.

Roboz, G. J. (2011) 'Novel approaches to the treatment of acute myeloid leukemia', *Hematology Am Soc Hematol Educ Program*, 2011, 43-50.

Rodrigues, N. P., Boyd, A. S., Fugazza, C., May, G. E., Guo, Y., Tipping, A. J., Scadden, D. T., Vyas, P. and Enver, T. (2008) 'GATA-2 regulates granulocyte-macrophage progenitor cell function.', *Blood*, 112(13), 4862-73.

Rodríguez-Paredes, M. and Esteller, M. (2011) 'Cancer epigenetics reaches mainstream oncology', *Nat Med*, 17(3), 330-9.

Romana, S. P., Mauchauffé, M., Le Coniat, M., Chumakov, I., Le Paslier, D., Berger, R. and Bernard, O. A. (1995) 'The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion', *Blood*, 85(12), 3662-70.

Rowe, J. M. and Tallman, M. S. (2010) 'How I treat acute myeloid leukemia', *Blood*, 116(17), 3147-56.

Roy, B. C., Aoyagi, T., Sarkar, S., Nomura, K., Kanda, H., Iwaya, K., Tachibana, M. and Kiyama, R. (2005) 'Pathological characterization of Kank in renal cell carcinoma', *Exp Mol Pathol*, 78(1), 41-8.

Sarkar, S., Roy, B. C., Hatano, N., Aoyagi, T., Gohji, K. and Kiyama, R. (2002) 'A novel ankyrin repeat-containing gene (Kank) located at 9p24 is a growth suppressor of renal cell carcinoma', *J Biol Chem*, 277(39), 36585-91.

Sasaki, M. S. and Tonomura, A. (1973) 'A high susceptibility of Fanconi's anemia to chromosome breakage by DNA cross-linking agents', *Cancer Res*, 33(8), 1829-36.

Savage, S. A., Giri, N., Baerlocher, G. M., Orr, N., Lansdorp, P. M. and Alter, B. P. (2008) 'TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita', *Am J Hum Genet*, 82(2), 501-9.

Schlenk, R. F., Döhner, K., Krauter, J., Fröhling, S., Corbacioglu, A., Bullinger, L., Habdank, M., Späth, D., Morgan, M., Benner, A., Schlegelberger, B., Heil, G., Ganser, A., Döhner, H. and Group, G.-A. A. M. L. S. (2008) 'Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia', *N Engl J Med*, 358(18), 1909-18.

Schlenk, R. F., Taskesen, E., van Norden, Y., Krauter, J., Ganser, A., Bullinger, L., Gaidzik, V. I., Paschka, P., Corbacioglu, A., Göhring, G., Kündgen, A., Held, G., Götze, K., Vellenga, E., Kuball, J., Schanz, U., Passweg, J., Pabst, T., Maertens, J., Ossenkoppele, G. J., Delwel, R., Döhner, H., Cornelissen, J. J., Döhner, K. and Löwenberg, B. (2013) 'The value of allogeneic and autologous hematopoietic stem cell transplantation in prognostically favorable acute myeloid leukemia with double mutant CEBPA', *Blood*, 122(9), 1576-82.

Schnittger, S., Bacher, U., Kern, W., Haferlach, C. and Haferlach, T. (2007a) 'JAK2 seems to be a typical cooperating mutation in therapy-related t(8;21)/ AML1-ETO-positive AML', *Leukemia*, 21(1), 183-4.

Schnittger, S., Bacher, U., Kern, W., Haferlach, T. and Haferlach, C. (2007b) 'JAK2V617F as progression marker in CMPD and as cooperative mutation in AML with trisomy 8 and t(8;21): a comparative study on 1103 CMPD and 269 AML cases', *Leukemia*, 21(8), 1843-5.

Schnittger, S., Eder, C., Jeromin, S., Alpermann, T., Fasan, A., Grossmann, V., Kohlmann, A., Illig, T., Klopp, N., Wichmann, H. E., Kreuzer, K. A., Schmid, C., Staib, P., Peceny, R., Schmitz, N., Kern, W., Haferlach, C. and Haferlach, T. (2013) 'ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome', *Leukemia*, 27(1), 82-91.

Schnittger, S., Kern, W., Tschulik, C., Weiss, T., Dicker, F., Falini, B., Haferlach, C. and Haferlach, T. (2009) 'Minimal residual disease levels assessed by NPM1 mutation-specific RQ-PCR provide important prognostic information in AML', *Blood*, 114(11), 2220-31.

Schoch, C., Haase, D., Haferlach, T., Gudat, H., Büchner, T., Freund, M., Link, H., Lengfelder, E., Wandt, H., Sauerland, M. C., Löffler, H. and Fonatsch, C. (1996) 'Fifty-one patients with acute myeloid leukemia and translocation t(8;21)(q22;q22): an additional deletion in 9q is an adverse prognostic factor', *Leukemia*, 10(8), 1288-95.

Schouten, J. P., McElgunn, C. J., Waaijer, R., Zwijnenburg, D., Diepvens, F. and Pals, G. (2002) 'Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification', *Nucleic Acids Res*, 30(12), e57.

Schreiber, A., Stengel, F., Zhang, Z., Enchev, R. I., Kong, E. H., Morris, E. P., Robinson, C. V., da Fonseca, P. C. and Barford, D. (2011) 'Structural basis for the subunit assembly of the anaphase-promoting complex', *Nature*, 470(7333), 227-32.

Schwarz, J. M., Cooper, D. N., Schuelke, M. and Seelow, D. (2014) 'MutationTaster2: mutation prediction for the deep-sequencing age', *Nat Methods*, 11(4), 361-2.

Schwarz, J. M., Rödelsperger, C., Schuelke, M. and Seelow, D. (2010) 'MutationTaster evaluates disease-causing potential of sequence alterations', *Nat Methods*, 7(8), 575-6.

Scott, E. W., Simon, M. C., Anastasi, J. and Singh, H. (1994) 'Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages', *Science*, 265(5178), 1573-7.

Sellick, G. S., Spendlove, H. E., Catovsky, D., Pritchard-Jones, K. and Houlston, R. S. (2005) 'Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia', *Leukemia*, 19(7), 1276-8.

Shiba, N., Hasegawa, D., Park, M. J., Murata, C., Sato-Otsubo, A., Ogawa, C., Manabe, A., Arakawa, H., Ogawa, S. and Hayashi, Y. (2012) 'CBL mutation in chronic myelomonocytic leukemia secondary to familial platelet disorder with propensity to develop acute myeloid leukemia (FPD/AML)', *Blood*, 119(11), 2612-4.

Shih, A. H., Jiang, Y., Meydan, C., Shank, K., Pandey, S., Barreyro, L., Antony-Debre, I., Viale, A., Socci, N., Sun, Y., Robertson, A., Cavatore, M., de Stanchina, E., Hricik, T., Rapaport, F., Woods, B., Wei, C., Hatlen, M., Baljevic, M., Nimer, S. D., Tallman, M., Paietta, E., Cimmino, L., Aifantis, I., Steidl, U., Mason, C., Melnick, A. and Levine, R. L. (2015) 'Mutational cooperativity linked to combinatorial epigenetic gain of function in acute myeloid leukemia', *Cancer Cell*, 27(4), 502-15.

Shih, L. Y., Liang, D. C., Huang, C. F., Wu, J. H., Lin, T. L., Wang, P. N., Dunn, P., Kuo, M. C. and Tang, T. C. (2006) '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*, 20(4), 604-9.

Shlush, L. I., Zandi, S., Mitchell, A., Chen, W. C., Brandwein, J. M., Gupta, V., Kennedy, J. A., Schimmer, A. D., Schuh, A. C., Yee, K. W., McLeod, J. L., Doedens, M., Medeiros, J. J., Marke, R., Kim, H. J., Lee, K., McPherson, J. D., Hudson, T. J., Brown, A. M., Yousif, F., Trinh, Q. M., Stein, L. D., Minden, M. D., Wang, J. C., Dick, J. E. and Consortium, H. P.-L. G. P. (2014) 'Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia', *Nature*, 506(7488), 328-33.

Smith, M. L., Cavenagh, J. D., Lister, T. A. and Fitzgibbon, J. (2004) 'Mutation of CEBPA in familial acute myeloid leukemia.', *N Engl J Med*, 351(23), 2403-7.

Song, W. J., Sullivan, M. G., Legare, R. D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I. C., Haworth, C., Hock, R., Loh, M., Felix, C., Roy, D. C., Busque, L., Kurnit, D., Willman, C., Gewirtz, A. M., Speck, N. A., Bushweller, J. H., Li, F. P., Gardiner, K., Poncz, M., Maris, J. M. and Gilliland, D. G. (1999) 'Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia', *Nat Genet*, 23(2), 166-75.

Sroczynska, P., Lancrin, C., Kouskoff, V. and Lacaud, G. (2009) 'The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis', *Blood*, 114(26), 5279-89.

Stelljes, M., Corbacioglu, A., Schlenk, R. F., Döhner, K., Frühwald, M. C., Rossig, C., Ehlert, K., Silling, G., Müller-Tidow, C., Juergens, H., Döhner, H., Berdel, W. E., Kienast, J. and Koschmieder, S. (2011) 'Allogeneic stem cell transplant to eliminate germline mutations in the gene for CCAAT-enhancer-binding protein α from hematopoietic cells in a family with AML', *Leukemia*, 25(7), 1209-10.

Stenson, P. D., Mort, M., Ball, E. V., Shaw, K., Phillips, A. and Cooper, D. N. (2014) 'The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine', *Hum Genet*, 133(1), 1-9.

Stockley, J., Morgan, N. V., Bem, D., Lowe, G. C., Lordkipanidzé, M., Dawood, B., Simpson, M. A., Macfarlane, K., Horner, K., Leo, V. C., Talks, K., Motwani, J., Wilde, J. T., Collins, P. W., Makris, M., Watson, S. P., Daly, M. E. and Group, U. G. a. P. o. P. S. (2013) 'Enrichment of FLI1 and RUNX1 mutations in families with excessive bleeding and platelet dense granule secretion defects', *Blood*, 122(25), 4090-3.

Sun, T. and Wu, E. (2001) 'Acute monoblastic leukemia with t(8;16): a distinct clinicopathologic entity; report of a case and review of the literature', *Am J Hematol*, 66(3), 207-12.

Sun, W. and Downing, J. R. (2004) 'Haploinsufficiency of AML1 results in a decrease in the number of LTR-HSCs while simultaneously inducing an increase in more mature progenitors', *Blood*, 104(12), 3565-72.

Swerdlow SH et al, eds (2008). *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon, France: IARC.

Tahirov, T. H., Inoue-Bungo, T., Morii, H., Fujikawa, A., Sasaki, M., Kimura, K., Shiina, M., Sato, K., Kumasaka, T., Yamamoto, M., Ishii, S. and Ogata, K. (2001) 'Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFbeta', *Cell*, 104(5), 755-67.

Taskesen, E., Bullinger, L., Corbacioglu, A., Sanders, M. A., Erpelinck, C. A., Wouters, B. J., van der Poel-van de Luytgaarde, S. C., Damm, F., Krauter, J., Ganser, A., Schlenk, R. F., Löwenberg, B., Delwel, R., Döhner, H., Valk, P. J. and Döhner, K. (2011) 'Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity', *Blood*, 117(8), 2469-75.

Taussig, D. C., Vargaftig, J., Miraki-Moud, F., Griessinger, E., Sharrock, K., Luke, T., Lillington, D., Oakervee, H., Cavenagh, J., Agrawal, S. G., Lister, T. A., Gribben, J. G. and Bonnet, D. (2010)

'Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction', *Blood*, 115(10), 1976-84.

Tawana, K., Wang, J., Renneville, A., Bödör, C., Hills, R., Loveday, C., Savic, A., Van Delft, F. W., Treleaven, J., Georgiades, P., Uglow, E., Asou, N., Uike, N., Debeljak, M., Jazbec, J., Ancliff, P., Gale, R., Thomas, X., Mialou, V., Döhner, K., Bullinger, L., Mueller, B., Pabst, T., Stelljes, M., Schlegelberger, B., Wozniak, E., Iqbal, S., Okosun, J., Araf, S., Frank, A. K., Lauridsen, F. B., Porse, B., Nerlov, C., Owen, C., Dokal, I., Gribben, J., Smith, M., Preudhomme, C., Chelala, C., Cavenagh, J. and Fitzgibbon, J. (2015) 'Disease evolution and outcomes in familial AML with germline CEBPA mutations', *Blood*, 126(10), 1214-23.

Thiede, C., Koch, S., Creutzig, E., Steudel, C., Illmer, T., Schaich, M. and Ehninger, G. (2006) 'Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML)', *Blood*, 107(10), 4011-20.

Thiede, C., Steudel, C., Mohr, B., Schaich, M., Schäkel, U., Platzbecker, U., Wermke, M., Bornhäuser, M., Ritter, M., Neubauer, A., Ehninger, G. and Illmer, T. (2002) 'Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis', *Blood*, 99(12), 4326-35.

Thol, F., Bollin, R., Gehlhaar, M., Walter, C., Dugas, M., Suchanek, K. J., Kirchner, A., Huang, L., Chaturvedi, A., Wichmann, M., Wiehlmann, L., Shahswar, R., Damm, F., Göhring, G., Schlegelberger, B., Schlenk, R., Döhner, K., Döhner, H., Krauter, J., Ganser, A. and Heuser, M. (2014) 'Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications', *Blood*, 123(6), 914-20.

Thol, F., Damm, F., Lüdeking, A., Winschel, C., Wagner, K., Morgan, M., Yun, H., Göhring, G., Schlegelberger, B., Hoelzer, D., Lübbert, M., Kanz, L., Fiedler, W., Kirchner, H., Heil, G., Krauter, J., Ganser, A. and Heuser, M. (2011) 'Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia', *J Clin Oncol*, 29(21), 2889-96.

Thol, F., Suchanek, K. J., Koenecke, C., Stadler, M., Platzbecker, U., Thiede, C., Schroeder, T., Kobbe, G., Kade, S., Löffeld, P., Banihosseini, S., Bug, G., Ottmann, O., Hofmann, W. K., Krauter, J., Kröger, N., Ganser, A. and Heuser, M. (2013) 'SETBP1 mutation analysis in 944 patients with MDS and AML', *Leukemia*, 27(10), 2072-5.

Tiesmeier, J., Czwalinna, A., Müller-Tidow, C., Krauter, J., Serve, H., Heil, G., Ganser, A. and Verbeek, W. (2003) 'Evidence for allelic evolution of C/EBPalpha mutations in acute myeloid leukaemia', *Br J Haematol*, 123(3), 413-9.

Tipping, A. J., Pina, C., Castor, A., Hong, D., Rodrigues, N. P., Lazzari, L., May, G. E., Jacobsen, S. E. and Enver, T. (2009) 'High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle', *Blood*, 113(12), 2661-72.

Tomas-Roca, L., Tsaalbi-Shtylik, A., Jansen, J. G., Singh, M. K., Epstein, J. A., Altunoglu, U., Verzijl, H., Soria, L., van Beusekom, E., Roscioli, T., Iqbal, Z., Gilissen, C., Hoischen, A., de Brouwer, A. P., Erasmus, C., Schubert, D., Brunner, H., Pérez Aytés, A., Marin, F., Aroca, P.,

Kayserili, H., Carta, A., de Wind, N., Padberg, G. W. and van Bokhoven, H. (2015) 'De novo mutations in PLXND1 and REV3L cause Möbius syndrome', *Nat Commun*, 6, 7199.

Tong, W. and Lodish, H. F. (2004) 'Lnk inhibits Tpo-mpl signaling and Tpo-mediated megakaryocytopoiesis', *J Exp Med*, 200(5), 569-80.

Tong, W., Zhang, J. and Lodish, H. F. (2005) 'Lnk inhibits erythropoiesis and Epo-dependent JAK2 activation and downstream signaling pathways', *Blood*, 105(12), 4604-12.

Torchy, M. P., Hamiche, A. and Klaholz, B. P. (2015) 'Structure and function insights into the NuRD chromatin remodeling complex', *Cell Mol Life Sci*, 72(13), 2491-507.

Tsai, F. Y. and Orkin, S. H. (1997) 'Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation', *Blood*, 89(10), 3636-43.

Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W. and Orkin, S. H. (1994) 'An early haematopoietic defect in mice lacking the transcription factor GATA-2', *Nature*, 371(6494), 221-6.

Van den Berghe, H., Cassiman, J. J., David, G., Fryns, J. P., Michaux, J. L. and Sokal, G. (1974) 'Distinct haematological disorder with deletion of long arm of no. 5 chromosome', *Nature*, 251(5474), 437-8.

Van Sloun, P. P., Varlet, I., Sonneveld, E., Boei, J. J., Romeijn, R. J., Eeken, J. C. and De Wind, N. (2002) 'Involvement of mouse Rev3 in tolerance of endogenous and exogenous DNA damage', *Mol Cell Biol*, 22(7), 2159-69.

Van Vlierberghe, P., Palomero, T., Khiabanian, H., Van der Meulen, J., Castillo, M., Van Roy, N., De Moerloose, B., Philippé, J., González-García, S., Toribio, M. L., Taghon, T., Zuurbier, L., Cauwelier, B., Harrison, C. J., Schwab, C., Pisecker, M., Strehl, S., Langerak, A. W., Gecz, J., Sonneveld, E., Pieters, R., Paietta, E., Rowe, J. M., Wiernik, P. H., Benoit, Y., Soulier, J., Poppe, B., Yao, X., Cordon-Cardo, C., Meijerink, J., Rabadan, R., Speleman, F. and Ferrando, A. (2010) 'PHF6 mutations in T-cell acute lymphoblastic leukemia', *Nat Genet*, 42(4), 338-42.

Van Vlierberghe, P., Patel, J., Abdel-Wahab, O., Lobry, C., Hedvat, C. V., Balbin, M., Nicolas, C., Payer, A. R., Fernandez, H. F., Tallman, M. S., Paietta, E., Melnick, A., Vandenberghe, P., Speleman, F., Aifantis, I., Cools, J., Levine, R. and Ferrando, A. (2011) 'PHF6 mutations in adult acute myeloid leukemia', *Leukemia*, 25(1), 130-4.

Vardiman, J. W., Thiele, J., Arber, D. A., Brunning, R. D., Borowitz, M. J., Porwit, A., Harris, N. L., Le Beau, M. M., Hellström-Lindberg, E., Tefferi, A. and Bloomfield, C. D. (2009) 'The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes', *Blood*, 114(5), 937-51.

Vassiliou, G. S., Cooper, J. L., Rad, R., Li, J., Rice, S., Uren, A., Rad, L., Ellis, P., Andrews, R., Banerjee, R., Grove, C., Wang, W., Liu, P., Wright, P., Arends, M. and Bradley, A. (2011)

'Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice', *Nat Genet*, 43(5), 470-5.

Venkatraman, E. S. and Olshen, A. B. (2007) 'A faster circular binary segmentation algorithm for the analysis of array CGH data', *Bioinformatics*, 23(6), 657-63.

Verma, D., Kantarjian, H., Faderl, S., O'Brien, S., Pierce, S., Vu, K., Freireich, E., Keating, M., Cortes, J. and Ravandi, F. (2010) 'Late relapses in acute myeloid leukemia: analysis of characteristics and outcome', *Leuk Lymphoma*, 51(5), 778-82.

Vicente, C., Conchillo, A., García-Sánchez, M. A. and Odero, M. D. (2012) 'The role of the GATA2 transcription factor in normal and malignant hematopoiesis', *Crit Rev Oncol Hematol*, 82(1), 1-17.

Vinh, D. C., Patel, S. Y., Uzel, G., Anderson, V. L., Freeman, A. F., Olivier, K. N., Spalding, C., Hughes, S., Pittaluga, S., Raffeld, M., Sorbara, L. R., Elloumi, H. Z., Kuhns, D. B., Turner, M. L., Cowen, E. W., Fink, D., Long-Priel, D., Hsu, A. P., Ding, L., Paulson, M. L., Whitney, A. R., Sampaio, E. P., Frucht, D. M., DeLeo, F. R. and Holland, S. M. (2010) 'Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia', *Blood*, 115(8), 1519-29.

Viny, A. D. and Levine, R. L. (2014) 'Genetics of myeloproliferative neoplasms', *Cancer J*, 20(1), 61-5.

Visconte, V., Makishima, H., Maciejewski, J. P. and Tiu, R. V. (2012) 'Emerging roles of the spliceosomal machinery in myelodysplastic syndromes and other hematological disorders', *Leukemia*, 26(12), 2447-54.

Vulliamy, T. J., Kirwan, M. J., Beswick, R., Hossain, U., Baqai, C., Ratcliffe, A., Marsh, J., Walne, A. and Dokal, I. (2011) 'Differences in disease severity but similar telomere lengths in genetic subgroups of patients with telomerase and shelterin mutations', *PLoS One*, 6(9), e24383.

Vulliamy, T. J., Marrone, A., Knight, S. W., Walne, A., Mason, P. J. and Dokal, I. (2006) 'Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation', *Blood*, 107(7), 2680-5.

Vulliamy, T., Beswick, R., Kirwan, M., Marrone, A., Digweed, M., Walne, A. and Dokal, I. (2008) 'Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita', *Proc Natl Acad Sci USA*, 105(23), 8073-8.

Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P. J. and Dokal, I. (2001) 'The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita', *Nature*, 413(6854), 432-5.

Wahl, M. C., Will, C. L. and Lührmann, R. (2009) 'The spliceosome: design principles of a dynamic RNP machine', *Cell*, 136(4), 701-18.

Walne, A. J., Vulliamy, T., Beswick, R., Kirwan, M. and Dokal, I. (2010) 'Mutations in C16orf57 and normal-length telomeres unify a subset of patients with dyskeratosis congenita, poikiloderma with neutropenia and Rothmund-Thomson syndrome', *Hum Mol Genet*, 19(22), 4453-61.

Walne, A. J., Vulliamy, T., Kirwan, M., Plagnol, V. and Dokal, I. (2013) 'Constitutional mutations in RTEL1 cause severe dyskeratosis congenita', *Am J Hum Genet*, 92(3), 448-53.

Walne, A. J., Vulliamy, T., Marrone, A., Beswick, R., Kirwan, M., Masunari, Y., Al-Qurashi, F. H., Aljurf, M. and Dokal, I. (2007) 'Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10', *Hum Mol Genet*, 16(13), 1619-29.

Walter, M. J., Shen, D., Ding, L., Shao, J., Koboldt, D. C., Chen, K., Larson, D. E., McLellan, M. D., 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, J. L., Eades, W., Tomasson, M., Westervelt, P., DiPersio, J. F., Link, D. C., Mardis, E. R., Ley, T. J., Wilson, R. K. and Graubert, T. A. (2012) 'Clonal architecture of secondary acute myeloid leukemia', *N Engl J Med*, 366(12), 1090-8.

Walter, M. J., Shen, D., Shao, J., Ding, L., White, B. S., Kandoth, C., Miller, C. A., Niu, B., McLellan, M. D., Dees, N. D., Fulton, R., Elliot, K., Heath, S., Grillot, M., Westervelt, P., Link, D. C., DiPersio, J. F., Mardis, E., Ley, T. J., Wilson, R. K. and Graubert, T. A. (2013) 'Clonal diversity of recurrently mutated genes in myelodysplastic syndromes', *Leukemia*, 27(6), 1275-82.

Wang, F., Travins, J., DeLaBarre, B., Penard-Lacronique, V., Schalm, S., Hansen, E., Straley, K., Kernytsky, A., Liu, W., Gliser, C., Yang, H., Gross, S., Artin, E., Saada, V., Mylonas, E., Quivoron, C., Popovici-Muller, J., Saunders, J. O., Salituro, F. G., Yan, S., Murray, S., Wei, W., Gao, Y., Dang, L., Dorsch, M., Agresta, S., Schenkein, D. P., Biller, S. A., Su, S. M., de Botton, S. and Yen, K. E. (2013) 'Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation', *Science*, 340(6132), 622-6.

Wang, Q., Stacy, T., Miller, J. D., Lewis, A. F., Gu, T. L., Huang, X., Bushweller, J. H., Bories, J. C., Alt, F. W., Ryan, G., Liu, P. P., Wynshaw-Boris, A., Binder, M., Marín-Padilla, M., Sharpe, A. H. and Speck, N. A. (1996) 'The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo', *Cell*, 87(4), 697-708.

Wang, X., Dai, H., Wang, Q., Xu, Y., Wang, Y., Sun, A., Ruan, J., Chen, S. and Wu, D. (2013) 'EZH2 mutations are related to low blast percentage in bone marrow and -7/del(7q) in de novo acute myeloid leukemia', *PLoS One*, 8(4), e61341.

Wang, Y., Xiao, M., Chen, X., Chen, L., Xu, Y., Lv, L., Wang, P., Yang, H., Ma, S., Lin, H., Jiao, B., Ren, R., Ye, D., Guan, K. L. and Xiong, Y. (2015) 'WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation', *Mol Cell*, 57(4), 662-73.

Ward, P. S., Patel, J., Wise, D. R., Abdel-Wahab, O., Bennett, B. D., Coller, H. A., Cross, J. R., Fantin, V. R., Hedvat, C. V., Perl, A. E., Rabinowitz, J. D., Carroll, M., Su, S. M., Sharp, K. A., Levine, R. L. and Thompson, C. B. (2010) 'The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate', *Cancer Cell*, 17(3), 225-34.

Webber, B. R., Iacovino, M., Choi, S. H., Tolar, J., Kyba, M. and Blazar, B. R. (2013) 'DNA methylation of Runx1 regulatory regions correlates with transition from primitive to definitive hematopoietic potential in vitro and in vivo', *Blood*, 122(17), 2978-86.

Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., Wartman, L. D., Lamprecht, T. L., Liu, F., Xia, J., Kandoth, C., Fulton, R. S., McLellan, M. D., Dooling, D. J., Wallis, J. W., Chen, K., Harris, C. C., Schmidt, H. K., Kalicki-Veizer, J. M., Lu, C., Zhang, Q., Lin, L., O'Laughlin, M. D., McMichael, J. F., Delehaunty, K. D., Fulton, L. A., Magrini, V. J., McGrath, S. D., Demeter, R. T., Vickery, T. L., Hundal, J., Cook, L. L., Swift, G. W., Reed, J. P., Alldredge, P. A., Wylie, T. N., Walker, J. R., Watson, M. A., Heath, S. E., Shannon, W. D., Varghese, N., Nagarajan, R., Payton, J. E., Baty, J. D., Kulkarni, S., Klco, J. M., Tomasson, M. H., Westervelt, P., Walter, M. J., Graubert, T. A., DiPersio, J. F., Ding, L., Mardis, E. R. and Wilson, R. K. (2012) 'The origin and evolution of mutations in acute myeloid leukemia', *Cell*, 150(2), 264-78.

West, R. R., Hsu, A. P., Holland, S. M., Cuellar-Rodriguez, J. and Hickstein, D. D. (2014) 'Acquired ASXL1 mutations are common in patients with inherited GATA2 mutations and correlate with myeloid transformation', *Haematologica*, 99(2), 276-81.

Wheatley, K., Brookes, C. L., Howman, A. J., Goldstone, A. H., Milligan, D. W., Prentice, A. G., Moorman, A. V., Burnett, A. K. and Subgroup, U. K. N. C. R. I. H. O. C. S. G. a. A. M. L. (2009) 'Prognostic factor analysis of the survival of elderly patients with AML in the MRC AML11 and LRF AML14 trials', *Br J Haematol*, 145(5), 598-605.

Wheatley, K., Burnett, A. K., Goldstone, A. H., Gray, R. G., Hann, I. M., Harrison, C. J., Rees, J. K., Stevens, R. F. and Walker, H. (1999) 'A simple, robust, validated and highly predictive index for the determination of risk-directed therapy in acute myeloid leukaemia derived from the MRC AML 10 trial. United Kingdom Medical Research Council's Adult and Childhood Leukaemia Working Parties', *Br J Haematol*, 107(1), 69-79.

Whitman, S. P., Archer, K. J., Feng, L., Baldus, C., Becknell, B., Carlson, B. D., Carroll, A. J., Mrózek, K., Vardiman, J. W., George, S. L., Kolitz, J. E., Larson, R. A., Bloomfield, C. D. and Caligiuri, M. A. (2001) 'Absence of the wild-type allele predicts poor prognosis in adult de novo acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of FLT3: a cancer and leukemia group B study', *Cancer Res*, 61(19), 7233-9.

Willemsen, M. H., Nijhof, B., Fenckova, M., Nillesen, W. M., Bongers, E. M., Castells-Nobau, A., Asztalos, L., Viragh, E., van Bon, B. W., Tezel, E., Veltman, J. A., Brunner, H. G., de Vries, B. B., de Ligt, J., Yntema, H. G., van Bokhoven, H., Isidor, B., Le Caignec, C., Lorino, E., Asztalos, Z., Koolen, D. A., Vissers, L. E., Schenck, A. and Kleefstra, T. (2013) 'GATAD2B loss-of-function

mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in Drosophila', *J Med Genet*, 50(8), 507-14.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C. F., Eshkind, L., Bockamp, E., Lió, P., Macdonald, H. R. and Trumpp, A. (2008) 'Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair', *Cell*, 135(6), 1118-29.

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P. M., Kinston, S., Ouwehand, W. H., Dzierzak, E., Pimanda, J. E., de Bruijn, M. F. and Göttgens, B. (2010) 'Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators', *Cell Stem Cell*, 7(4), 532-44.

Wouters, B. J., Jordà, M. A., Keeshan, K., Louwers, I., Erpelinck-Verschueren, C. A., Tielemans, D., Langerak, A. W., He, Y., Yashiro-Ohtani, Y., Zhang, P., Hetherington, C. J., Verhaak, R. G., Valk, P. J., Löwenberg, B., Tenen, D. G., Pear, W. S. and Delwel, R. (2007) 'Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1', *Blood*, 110(10), 3706-14.

Wu, N. and Yu, H. (2012) 'The Smc complexes in DNA damage response', *Cell Biosci*, 2, 5.

Xiao, H., Shi, J., Luo, Y., Tan, Y., He, J., Xie, W., Zhang, L., Wang, Y., Liu, L., Wu, K., Yu, X., Cai, Z., Lin, M., Ye, X. and Huang, H. (2011) 'First report of multiple CEBPA mutations contributing to donor origin of leukemia relapse after allogeneic hematopoietic stem cell transplantation', *Blood*, 117(19), 5257-60.

Xie, M., Lu, C., Wang, J., McLellan, M. D., Johnson, K. J., Wendl, M. C., McMichael, J. F., Schmidt, H. K., Yellapantula, V., Miller, C. A., Ozenberger, B. A., Welch, J. S., Link, D. C., Walter, M. J., Mardis, E. R., Dipersio, J. F., Chen, F., Wilson, R. K., Ley, T. J. and Ding, L. (2014) 'Age-related mutations associated with clonal hematopoietic expansion and malignancies', *Nat Med*, 20(12), 1472-8.

Xu, J., Wang, Y. Y., Dai, Y. J., Zhang, W., Zhang, W. N., Xiong, S. M., Gu, Z. H., Wang, K. K., Zeng, R., Chen, Z. and Chen, S. J. (2014) 'DNMT3A Arg882 mutation drives chronic myelomonocytic leukemia through disturbing gene expression/DNA methylation in hematopoietic cells', *Proc Natl Acad Sci USA*, 111(7), 2620-5.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., Ito, S., Yang, C., Xiao, M. T., Liu, L. X., Jiang, W. Q., Liu, J., Zhang, J. Y., Wang, B., Frye, S., Zhang, Y., Xu, Y. H., Lei, Q. Y., Guan, K. L., Zhao, S. M. and Xiong, Y. (2011) 'Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases', *Cancer Cell*, 19(1), 17-30.

Yakoub-Agha, I., Mesnil, F., Kuentz, M., Boiron, J. M., Ifrah, N., Milpied, N., Chehata, S., Esperou, H., Vernant, J. P., Michallet, M., Buzyn, A., Gratecos, N., Cahn, J. Y., Bourhis, J. H., Chir, Z., Raffoux, C., Socié, G., Golmard, J. L., Jouet, J. P. and Therapy, F. S. o. B. M. T. a. C. (2006) 'Allogeneic marrow stem-cell transplantation from human leukocyte antigen-identical siblings versus human leukocyte antigen-allelic-matched unrelated donors (10/10) in

patients with standard-risk hematologic malignancy: a prospective study from the French Society of Bone Marrow Transplantation and Cell Therapy', *J Clin Oncol*, 24(36), 5695-702.

Yamagata, T., Maki, K. and Mitani, K. (2005) 'Runx1/AML1 in normal and abnormal hematopoiesis', *Int J Hematol*, 82(1), 1-8.

Yamaguchi, H., Calado, R. T., Ly, H., Kajigaya, S., Baerlocher, G. M., Chanock, S. J., Lansdorp, P. M. and Young, N. S. (2005) 'Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia', *N Engl J Med*, 352(14), 1413-24.

Yan, J., Liu, Y., Lukasik, S. M., Speck, N. A. and Bushweller, J. H. (2004) 'CBFbeta allosterically regulates the Runx1 Runt domain via a dynamic conformational equilibrium', *Nat Struct Mol Biol*, 11(9), 901-6.

Yan, X. J., Xu, J., Gu, Z. H., Pan, C. M., Lu, G., Shen, Y., Shi, J. Y., Zhu, Y. M., Tang, L., Zhang, X. W., Liang, W. X., Mi, J. Q., Song, H. D., Li, K. Q., Chen, Z. and Chen, S. J. (2011) 'Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia', *Nat Genet*, 43(4), 309-15.

Yates, J. W., Wallace, H. J., Ellison, R. R. and Holland, J. F. (1973) 'Cytosine arabinoside (NSC-63878) and daunorubicin (NSC-83142) therapy in acute nonlymphocytic leukemia', *Cancer Chemother Rep*, 57(4), 485-8.

Ye, K., Schulz, M. H., Long, Q., Apweiler, R. and Ning, Z. (2009) 'Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads', *Bioinformatics*, 25(21), 2865-71.

Ye, M., Zhang, H., Amabile, G., Yang, H., Staber, P. B., Zhang, P., Levantini, E., Alberich-Jordà, M., Zhang, J., Kawasaki, A. and Tenen, D. G. (2013) 'C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence', *Nat Cell Biol*, 15(4), 385-94.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., Chalkidis, G., Suzuki, Y., Shiosaka, M., Kawahata, R., Yamaguchi, T., Otsu, M., Obara, N., Sakata-Yanagimoto, M., Ishiyama, K., Mori, H., Nolte, F., Hofmann, W. K., Miyawaki, S., Sugano, S., Haferlach, C., Koeffler, H. P., Shih, L. Y., Haferlach, T., Chiba, S., Nakauchi, H., Miyano, S. and Ogawa, S. (2011) 'Frequent pathway mutations of splicing machinery in myelodysplasia', *Nature*, 478(7367), 64-9.

Yoshida, K., Toki, T., Okuno, Y., Kanezaki, R., Shiraishi, Y., Sato-Otsubo, A., Sanada, M., Park, M. J., Terui, K., Suzuki, H., Kon, A., Nagata, Y., Sato, Y., Wang, R., Shiba, N., Chiba, K., Tanaka, H., Hama, A., Muramatsu, H., Hasegawa, D., Nakamura, K., Kanegane, H., Tsukamoto, K., Adachi, S., Kawakami, K., Kato, K., Nishimura, R., Izraeli, S., Hayashi, Y., Miyano, S., Kojima, S., Ito, E. and Ogawa, S. (2013) 'The landscape of somatic mutations in Down syndrome-related myeloid disorders', *Nat Genet*, 45(11), 1293-9.

Yoshimi, A., Toya, T., Kawazu, M., Ueno, T., Tsukamoto, A., Iizuka, H., Nakagawa, M., Nannya, Y., Arai, S., Harada, H., Usuki, K., Hayashi, Y., Ito, E., Kirito, K., Nakajima, H., Ichikawa, M.,

Mano, H. and Kurokawa, M. (2014) 'Recurrent CDC25C mutations drive malignant transformation in FPD/AML', *Nat Commun*, 5, 4770.

You, Z. and Bailis, J. M. (2010) 'DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints', *Trends Cell Biol*, 20(7), 402-9.

Yu, X. and Chen, J. (2004) 'DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains', *Mol Cell Biol*, 24(21), 9478-86.

Yun, M. H. and Hiom, K. (2009) 'CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle', *Nature*, 459(7245), 460-3.

Zelent, A., Greaves, M. and Enver, T. (2004) 'Role of the TEL-AML1 fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia', *Oncogene*, 23(24), 4275-83.

Zhang, D. E., Zhang, P., Wang, N. D., Hetherington, C. J., Darlington, G. J. and Tenen, D. G. (1997) 'Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice', *Proc Natl Acad Sci USA*, 94(2), 569-74.

Zhang, J., Ding, L., Holmfeldt, L., Wu, G., Heatley, S. L., Payne-Turner, D., Easton, J., Chen, X., Wang, J., Rusch, M., Lu, C., Chen, S. C., Wei, L., Collins-Underwood, J. R., Ma, J., Roberts, K. G., Pounds, S. B., Ulyanov, A., Becksfort, J., Gupta, P., Huether, R., Kriwacki, R. W., Parker, M., McGoldrick, D. J., Zhao, D., Alford, D., Espy, S., Bobba, K. C., Song, G., Pei, D., Cheng, C., Roberts, S., Barbato, M. I., Campana, D., Coustan-Smith, E., Shurtleff, S. A., Raimondi, S. C., Kleppe, M., Cools, J., Shimano, K. A., Hermiston, M. L., Doulatov, S., Eppert, K., Laurenti, E., Notta, F., Dick, J. E., Basso, G., Hunger, S. P., Loh, M. L., Devidas, M., Wood, B., Winter, S., Dunsmore, K. P., Fulton, R. S., Fulton, L. L., Hong, X., Harris, C. C., Dooling, D. J., Ochoa, K., Johnson, K. J., Obenauer, J. C., Evans, W. E., Pui, C. H., Naeve, C. W., Ley, T. J., Mardis, E. R., Wilson, R. K., Downing, J. R. and Mullighan, C. G. (2012) 'The genetic basis of early T-cell precursor acute lymphoblastic leukaemia', *Nature*, 481(7380), 157-63.

Zhang, M. Y., Churpek, J. E., Keel, S. B., Walsh, T., Lee, M. K., Loeb, K. R., Gulsuner, S., Pritchard, C. C., Sanchez-Bonilla, M., Delrow, J. J., Basom, R. S., Forouhar, M., Gyurkocza, B., Schwartz, B. S., Neistadt, B., Marquez, R., Mariani, C. J., Coats, S. A., Hofmann, I., Lindsley, R. C., Williams, D. A., Abkowitz, J. L., Horwitz, M. S., King, M. C., Godley, L. A. and Shimamura, A. (2015) 'Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy', *Nat Genet*, 47(2), 180-5.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M. L., Dayaram, T., Owens, B. M., Shigematsu, H., Levantini, E., Huettner, C. S., Lekstrom-Himes, J. A., Akashi, K. and Tenen, D. G. (2004) 'Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha', *Immunity*, 21(6), 853-63.

Zhao, X., Chen, A., Yan, X., Zhang, Y., He, F., Hayashi, Y., Dong, Y., Rao, Y., Li, B., Conway, R. M., Maiques-Diaz, A., Elf, S. E., Huang, N., Zuber, J., Xiao, Z., Tse, W., Tenen, D. G., Wang, Q., Chen, W., Mulloy, J. C., Nimer, S. D. and Huang, G. (2014) 'Downregulation of RUNX1/CBFβ by MLL fusion proteins enhances hematopoietic stem cell self-renewal', *Blood*, 123(11), 1729-38.

Zhao, X., Jankovic, V., Gural, A., Huang, G., Pardanani, A., Menendez, S., Zhang, J., Dunne, R., Xiao, A., Erdjument-Bromage, H., Allis, C. D., Tempst, P. and Nimer, S. D. (2008) 'Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity', *Genes Dev*, 22(5), 640-53.

Zheng, R., Bailey, E., Nguyen, B., Yang, X., Piloto, O., Levis, M. and Small, D. (2011) 'Further activation of FLT3 mutants by FLT3 ligand', *Oncogene*, 30(38), 4004-14.

Zhi, J. and Hatchwell, E. (2008) 'Human MLPA Probe Design (H-MAPD): a probe design tool for both electrophoresis-based and bead-coupled human multiplex ligation-dependent probe amplification assays', *BMC Genomics*, 9, 407.

Zhong, F., Savage, S. A., Shkreli, M., Giri, N., Jessop, L., Myers, T., Chen, R., Alter, B. P. and Artandi, S. E. (2011) 'Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita', *Genes Dev*, 25(1), 11-6.

Appendix 1

Supplementary Data for Chapter 2: Materials and Methods

1. Genes with a high density of probes in customised array CGH protocol

(Agilent 2x 400k Design ID 045154)

- *RUNX1*
- *GATA2*
- *CEBPA*
- *SRP9*
- *SRP14*
- *SRP19*
- *SRP54*
- *SRP68*
- *SRP72*
- *MPL*
- *DKC1*
- *TERT*
- *TERC*
- *NOP10*
- *NHP2*
- *TINF2*
- *CTC1*
- *C16ORF57*
- *TCAB1*

2. Consortia contributing to ExAC Browser database (http://exac.broadinstitute.org/)

- 1000 Genomes
- Bulgarian Trios
- Finland-United States Investigation of NIDDM Genetics (FUSION)
- GoT2D
- Inflammatory Bowel Disease
- METabolic Syndrome In Men (METSIM)
- Jackson Heart Study
- Myocardial Infarction Genetics Consortium:
    - Italian Atherosclerosis, Thrombosis, and Vascular Biology Working Group
    - Ottawa Genomics Heart Study
    - Pakistan Risk of Myocardial Infarction Study (PROMIS)
    - Precocious Coronary Artery Disease Study (PROCARDIS)
    - Registre Gironi del COR (REGICOR)
- NHLBI-GO Exome Sequencing Project (ESP)
- National Institute of Mental Health (NIMH) Controls
- SIGMA-T2D
- Sequencing in Suomi (SISu)
- Swedish Schizophrenia & Bipolar Studies
- T2D-GENES
- Schizophrenia Trios from Taiwan
- The Cancer Genome Atlas (TCGA)
- Tourette Syndrome Association International Consortium for Genomics (TSAICG)

Appendix 2

Supplementary Data for Chapter 3: Clinical and molecular profiling of families with germline *CEBPA* mutations

Table 1 Total number of somatic variants detected in *CEBPA*-mutated AML samples using WES*

Sample	Total SNV	Total Indel	Coding mutations			UTR	
			Non-synonymous SNV	Synonymous SNV	Indel	SNV	Indel
A.II.1	21	9	13	5	7	3	2
A.II.5	17	12	8	5	2	4	10
A.III.2	19	6	9	3	3	7	3
B.I.1-T1	30	15	16	10	8	4	7
B.I.1-T2	19	11	13	3	4**	3	7
B.I.1-T3	37	9	24	8	4	5	5
B.II.2	22	6	12	8	3	2	3
C.III.1	30	6	15	8	2†	7	4
D.II.2	23	8	13	7	4	3	4
Total	**218**	**82**	**123**	**57**	**37**	**38**	**45**

*Results include all variants generated by bioinformatics pipeline

** Excluding *WT1* variant; this was detected by localised analysis of WES alignments, followed by Sanger sequencing

† Excluding *CEBPA* C-terminal variant; this was detected by Sanger Sequencing alone (39bp duplication)

Table 2 Somatic coding variants for pedigrees A, C and D *

Sample	Ensembl Gene ID	Ensembl Transcript ID	Gene Symbol	Chromosome	Genomic position	Nucleotide change (ref/variant)	cDNA position	aa position	aa change	Mutation type	EXPANDS analysis clonal/subclonal	VAF
A.II.1	ENSG00000185477	ENST00000333209	GPRIN3	chr4	90169380	C/T	2401	628	V>I	missense- non-syn	clonal	0.493
A.II.1	ENSG00000179348	ENST00000341105	GATA2	chr3	128202759	G/A	1293	321	L>F	missense- non-syn	clonal	0.444
A.II.1	ENSG00000100578	ENST00000261244	KIAA0586	chr14	58924643	G/A	1804	510	R>H	missense- non-syn	clonal	0.390
A.II.1	ENSG00000154258	ENST00000340001	ABCA9	chr17	67016593	C/A	2679	846	V>F	missense- non-syn	clonal	0.388
A.II.1	ENSG00000113924	ENST00000283871	HGD	chr3	120360494	G/A	1281	274	P>L	missense- non-syn	clonal	0.349
A.II.1	ENSG00000123989	ENST00000243776	CHPF	chr2	220406906	C/A	569	107	R>L	missense- non-syn	subclonal	0.097
A.II.1	ENSG00000163386	ENST00000342960	NBPF10	chr1	145323666	A/G	3538	1168	D>G	missense- non-syn	subclonal	0.085
A.II.1	ENSG00000163386	ENST00000342960	NBPF10	chr1	145323667	C/G	3539	1168	D>E	missense- non-syn	subclonal	0.085
A.II.1	ENSG00000213281	ENST00000369535	NRAS	chr1	115258747	C/G	289	12	G>A	missense- non-syn	subclonal	0.077
A.II.1	ENSG00000146233	ENST00000275016	CYP39A1	chr6	46620199	C/A	325	41	V>F	missense- non-syn	subclonal	0.077
A.II.1	ENSG00000205277	ENST00000379442	MUC12	chr7	100643376	G/C	9961	3321	G>R	missense- non-syn	subclonal	0.069
A.II.1	ENSG00000090863	ENST00000205061	GLG1	chr16	74486029	C/A	3587	1179	R>S	missense- non-syn	subclonal	0.064
A.II.1	ENSG00000144909	ENST00000296220	OSBPL11	chr3	125249417	C/A	2476	729	G>V	missense- non-syn	subclonal	0.060
A.II.1	ENSG00000090061	ENST00000389879	CCNK	chr14	99976750	C/T	1497	458	Y>Y	missense-syn	clonal	0.467
A.II.1	ENSG00000166682	ENST00000545579	TMPRSS5	chr11	113570709	A/G	307	18	P>P	missense-syn	clonal	0.381
A.II.1	ENSG00000204544	ENST00000376296	MUC21	chr6	30955008	A/G	1297	352	T>T	missense-syn	clonal	0.192
A.II.1	ENSG00000144381	ENST00000345042	HSPD1	chr2	198363501	C/T	189	24	R>R	missense-syn	subclonal	0.070
A.II.1	ENSG00000196497	ENST00000354464	IPO4	chr14	24652520	C/A	2294	721	L>L	missense-syn	subclonal	0.067
A.II.1	ENSG00000112983	ENST00000254900	BRD8	chr5	137503737	C/CA	1045	225	G>W	insertion(fs)	clonal	0.381
A.II.1	ENSG00000173442	ENST00000533364	EHBP1L1	chr11	65359237	A/AC	378	22	H>H	insertion(fs)	clonal	0.435
A.II.1	ENSG00000100285	ENST00000328842	NEFH	chr22	29885580	AAGGCCAAGTCCCCAGAGAAGGAAG/A	2022	651	KAKSPEKEE>KAKSPEKAK	deletion	clonal	0.558
A.II.1	ENSG00000155087	ENST00000285402	ODF1	chr8	103573010	CTGCAACCCCTGCAGCCCCTGCAACCCG/C	807	217	PCNPCSPCNP>PCSPYDPCNP	deletion	clonal	0.392
A.II.1	ENSG00000245848	ENST00000498907	CEBPA	chr19	33792382	C/CTTCTGCTGCGTCTCCACGTTGCGCTGCTTGGCCTTT	1089	313	KVLELTSDNDRLR>KKAKQRNVETQQK	insertion	subclonal†	0.101†
A.II.5	ENSG00000167524	ENST00000301037	SGK494	chr17	26938181	G/A	1023	331	L>F	missense- non-syn	clonal	0.515
A.II.5	ENSG00000004777	ENST00000007510	ARHGAP33	chr19	36271122	C/G	655	171	H>D	missense- non-syn	clonal	0.377
A.II.5	ENSG00000179348	ENST00000341105	GATA2	chr3	128202731	C/T	1321	330	R>Q	missense- non-syn	clonal	0.317
A.II.5	ENSG00000135636	ENST00000410020	DYSF	chr2	71896337	G/C	5783	1881	G>A	missense- non-syn	clonal	0.272
A.II.5	ENSG00000215041	ENST00000399464	NEURL4	chr17	7224982	G/A	3003	999	P>L	missense- non-syn	clonal	0.241
A.II.5	ENSG00000116117	ENST00000406610	PARD3B	chr2	206305364	C/A	3219	1004	N>K	missense- non-syn	clonal	0.162
A.II.5	ENSG00000179348	ENST00000341105	GATA2	chr3	128202759	G/A	1293	321	L>F	missense- non-syn	subclonal	0.159
A.II.5	ENSG00000237763	ENST00000332554	AMY1A	chr1	104256478	T/A	711	216	L>H	missense- non-syn	subclonal	0.122
A.II.5	ENSG00000157600	ENST00000372068	TMEM164	chrX	109416613	C/T	1167	276	Y>Y	missense-syn	clonal	0.413
A.II.5	ENSG00000170465	ENST00000252250	KRT6C	chr12	52867105	G/A	465	139	T>T	missense-syn	clonal	0.278
A.II.5	ENSG00000004897	ENST00000066544	CDC27	chr17	45234725	T/C	595	167	T>T	missense-syn	subclonal	0.13
A.II.5	ENSG00000237541	ENST00000374940	HLA-DQA2	chr6	32713827	C/T	693	197	D>D	missense-syn	subclonal	0.087
A.II.5	ENSG00000143702	ENST00000366542	CEP170	chr1	243333027	A/G	1798	582	R>R	missense-syn	subclonal	0.058
A.II.5	ENSG00000245848	ENST00000498907	CEBPA	chr19	33792381	C/CTTCTGCTGCGTCTCCACGTTGCGCTG	1090	314	VLELTSDND>QRNVETQQK	insertion	clonal	0.227
A.III.2	ENSG00000101825	ENST00000217939	MXRA5	chrX	3242301	G/T	1580	475	S>R	missense- non-syn	clonal	0.769
A.III.2	ENSG00000179348	ENST00000341105	GATA2	chr3	128202770	T/A	1282	317	N>I	missense- non-syn	clonal	0.520
A.III.2	ENSG00000182473	ENST00000406660	EXOC7	chr17	74097326	C/T	485	148	R>K	missense- non-syn	clonal	0.431
A.III.2	ENSG00000154429	ENST00000366687	CCSAP	chr1	229478136	C/A	129	26	C>F	missense- non-syn	subclonal 2	0.143
A.III.2	ENSG00000076928	ENST00000599846	ARHGEF1	chr19	42410855	G/T	2949	942	E>*	nonsense	subclonal 2	0.136
A.III.2	ENSG00000240386	ENST00000334371	LCE1F	chr1	152748877	C/A	30	10	C>*	nonsense	subclonal 2	0.103
A.III.2	ENSG00000183258	ENST00000330503	DDX41	chr5	176940016	G/T	1352	451	P>Q	missense- non-syn	subclonal 2	0.073
A.III.2	ENSG00000152223	ENST00000282041	EPG5	chr18	43450561	C/A	6231	2066	D>Y	missense- non-syn	subclonal 2	0.067
A.III.2	ENSG00000160716	ENST00000368476	CHRNB2	chr1	154544183	C/T	1148	295	P>L	missense- non-syn	subclonal 2	0.045
A.III.2	ENSG00000108242	ENST00000285979	CYP2C18	chr10	96480153	GA/G	1019	274	E>E	deletion (fs)	clonal	0.516
A.III.2	ENSG00000245848	ENST00000498907	CEBPA	chr19	33792381	C/CCTT	1090	314	VL>KV	insertion	clonal	0.374
A.III.2	ENSG00000196156	ENST00000391356	KRTAP4-3	chr17	39324229	T/TGCAGCAGGTGGTCAG	196	66	RTTCCR>LTTCCR	insertion	clonal	0.360
C.III.1	ENSG00000228344	ENST00000458396	AP001496.1	chr18	5233143	C/T	269	90	A>V	missense- non-syn	N/A**	0.474
C.III.1	ENSG00000184937	ENST00000332351	WT1	chr11	32456326	G/T	851	189	S>*	nonsense	N/A**	0.472
C.III.1	ENSG00000130714	ENST00000372228	POMT1	chr9	134387477	C/T	1281	368	P>S	missense- non-syn	N/A**	0.422
C.III.1	ENSG00000166831	ENST00000300069	RBPMS2	chr15	65041629	C/T	NA	NA	splice site	splice site	N/A**	0.348
C.III.1	ENSG00000234560	ENST00000431524	OR10G8	chr11	123900388	C/G	92	20	A>G	missense- non-syn	N/A**	0.179
C.III.1	ENSG00000137868	ENST00000432245	STRA6	chr15	74487969	A/T	700	152	L>M	missense- non-syn	N/A**	0.13
C.III.1	ENSG00000145113	ENST00000462323	MUC4	chr3	195512606	G/C	5917	1949	H>D	missense- non-syn	N/A**	0.115
C.III.1	ENSG00000151067	ENST00000399634	CACNA1C	chr12	2797817	G/T	6202	2068	E>*	nonsense	N/A**	0.115
C.III.1	ENSG00000139405	ENST00000552495	C12orf52	chr12	113624792	G/T	785	105	E>*	nonsense	N/A**	0.11
C.III.1	ENSG00000161640	ENST00000447370	SIGLEC11	chr19	50463839	G/C	521	144	L>V	missense- non-syn	N/A**	0.068
C.III.1	ENSG00000204442	ENST00000375915	FAM155A	chr13	108518717	C/G	367	76	Q>H	missense- non-syn	N/A**	0.066
C.III.1	ENSG00000141519	ENST00000374877	CCDC40	chr17	78064092	G/A	3014	996	R>H	missense- non-syn	N/A**	0.062
C.III.1	ENSG00000213401	ENST00000357916	MAGEA12	chrX	151900122	C/T	835	227	V>M	missense- non-syn	N/A**	0.06
C.III.1	ENSG00000213401	ENST00000357916	MAGEA12	chrX	151900169	T/A	788	211	K>I	missense- non-syn	N/A**	0.056
C.III.1	ENSG00000159363	ENST00000341676	ATP13A2	chr1	17312585	C/A	3549	1124	P>P	missense-syn	N/A**	0.099
C.III.1	ENSG00000172519	ENST00000308940	OR10H5	chr19	15905326	G/A	566	156	G>G	missense-syn	N/A**	0.093
C.III.1	ENSG00000204390	ENST00000375653	HSPA1L	chr6	31779678	C/T	413	24	Q>Q	missense-syn	N/A**	0.085
C.III.1	ENSG00000172519	ENST00000308940	OR10H5	chr19	15905398	C/T	638	180	H>H	missense-syn	N/A**	0.077
C.III.1	ENSG00000119227	ENST00000412723	PIGZ	chr3	196678780	G/T	270	41	L>L	missense-syn	N/A**	0.06
C.III.1	ENSG00000256453	ENST00000542735	DND1	chr5	140052469	T/C	209	55	P>P	missense-syn	N/A**	0.044
C.III.1	ENSG00000122515	ENST00000414051	ZMIZ2	chr7	4480584	G/T	2337	779	P>P	missense-syn	N/A**	0.037
C.III.1	ENSG00000227184	ENST00000525985	EPPK1	chr8	144940267	G/A	7227	2385	G>G	missense-syn	N/A**	0.031
C.III.1	ENSG00000165046	ENST00000379957	LETM2	chr8	38251663	GT/G	676	183	VF>VS	deletion (fs)	N/A**	0.529
D.II.2	ENSG00000184937	ENST00000332351	WT1	chr11	32421575	G/T	1302	339	Y>*	nonsense	clonal	0.478
D.II.2	ENSG00000090920	ENST00000221347	FCGBP	chr19	40408120	G/A	4609	1534	P>L	missense- non-syn	clonal	0.467
D.II.2	ENSG00000173838	ENST00000311269	MARCH10	chr17	60814250	C/T	1264	327	A>T	missense- non-syn	clonal	0.442
D.II.2	ENSG00000255569	ENST00000542354	TRAV1-1	chr14	22090620	C/T	292	76	R>C	missense- non-syn	clonal	0.390
D.II.2	ENSG00000147687	ENST00000519232	TATDN1	chr8	125535198	C/A	156	53	R>M	missense- non-syn	subclonal	0.086
D.II.2	ENSG00000110274	ENST00000278935	CEP164	chr11	117253605	G/T	1818	557	E>D	missense- non-syn	subclonal	0.083
D.II.2	ENSG00000198788	ENST00000441003	MUC2	chr11	1093324	G/A	5170	1715	G>S	missense- non-syn	subclonal	0.083
D.II.2	ENSG00000132406	ENST00000382753	TMEM128	chr4	4239589	C/T	482	158	V>I	missense- non-syn	subclonal	0.067
D.II.2	ENSG00000175707	ENST00000320567	C1orf172	chr1	27278630	C/A	331	81	W>L	missense- non-syn	subclonal	0.063
D.II.2	ENSG00000139515	ENST00000381033	PDX1	chr13	28494634	C/A	478	120	P>Q	missense- non-syn	subclonal	0.056
D.II.2	ENSG00000168930	ENST00000329758	TRIM49	chr11	89531780	G/C	1206	293	H>D	missense- non-syn	subclonal	0.053
D.II.2	ENSG00000151067	ENST00000327702	CACNA1C	chr12	2602528	G/T	1089	363	E>D	missense- non-syn	subclonal	0.046
D.II.2	ENSG00000168930	ENST00000329758	TRIM49	chr11	89531764	T/C	1222	298	N>S	missense- non-syn	subclonal	0.034
D.II.2	ENSG00000170214	ENST00000306675	ADRA1B	chr5	159399022	C/A	1209	362	I>I	missense-syn	clonal	0.535
D.II.2	ENSG00000135899	ENST00000540870	SP110	chr2	231067434	T/C	1185	309	S>S	missense-syn	clonal	0.442
D.II.2	ENSG00000135624	ENST00000258091	CCT7	chr2	73474943	G/A	813	224	G>G	missense-syn	clonal	0.328
D.II.2	ENSG00000222036	ENST00000409832	POTEG	chr14	19553494	G/A	130	26	K>K	missense-syn	subclonal	0.154
D.II.2	ENSG00000107159	ENST00000378357	CA9	chr9	35674208	T/C	356	84	P>P	missense-syn	subclonal	0.104
D.II.2	ENSG00000114859	ENST00000265593	CLCN2	chr3	184072424	C/A	1591	473	V>V	missense-syn	subclonal	0.057
D.II.2	ENSG00000186803	ENST00000357374	IFNA10	chr9	21206569	C/A	574	176	S>S	missense-syn	subclonal	0.079
D.II.2	ENSG00000092853	ENST00000544356	CLSPN	chr1	36214420	TC/T	2052	658	SE>SK|S	deletion (fs)	clonal	0.375
D.II.2	ENSG00000245848	ENST00000498907	CEBPA	chr19	33792381	C/CCTT	1090	314	VL>KV	insertion	clonal	0.303
D.II.2	ENSG00000160202	ENST00000398133	CRYAA	chr21	44589920	G/GC	210	29	R>R	insertion (fs)	clonal	0.333

* Exclusion of variants confirmed as sequencing artifacts by manual review of WES data and Sanger sequencing

** EXPANDS analysis could not be applied as LOH calculations were not possible for the tumour: remission comparison

† *CEBPA* mutation VAF underestimated due to size of insertion

Table 3 Somatic coding variants for B.I.1 T1-T3 and B.II.2

Sample	Ensembl Gene ID	Ensembl Transcript ID	Gene Symbol	Chromosome	Genomic position	Nucleotide change (ref/variant)	cDNA position	aa position	aa change	Mutation type	EXPANDS analysis clonal/subclonal	VAF
B.I.1-T1	ENSG00000134917	ENST00000414575	ADAMTS8	chr11	130286202	C/T	1207	403	D>N	missense- non-syn	clonal	0.429
B.I.1-T1	ENSG00000108055	ENST00000361804	SMC3	chr10	112343991	G/A	1268	381	R>Q	missense- non-syn	clonal	0.294
B.I.1-T1	ENSG00000174937	ENST00000312240	OR5M3	chr11	56237948	T/C	67	9	E>G	missense- non-syn	clonal	0.283
B.I.1-T1	ENSG00000107863	ENST00000396432	ARHGAP21	chr10	24873709	G/A	5996	1837	R>W	missense- non-syn	clonal	0.243
B.I.1-T1	ENSG00000145882	ENST00000274569	PCYOX1L	chr5	148742349	G/A	300	80	G>R	missense- non-syn	subclonal 1	0.2
B.I.1-T1	ENSG00000162390	ENST00000343744	ACOT11	chr1	55068352	G/A	1122	347	R>Q	missense- non-syn	subclonal 1	0.184
B.I.1-T1	ENSG00000170421	ENST00000546542	KRT8	chr12	53298675	A/C	550	109	S>A	missense- non-syn	subclonal 1	0.182
B.I.1-T1	ENSG00000134121	ENST00000453040	CHL1	chr3	239348	G/A	22	7	G>R	missense- non-syn	subclonal 1	0.171
B.I.1-T1	ENSG00000106462	ENST00000320356	EZH2	chr7	148523708	C/T	867	249	E>K	missense- non-syn	subclonal 1	0.149
B.I.1-T1	ENSG00000156564	ENST00000338305	LRFN2	chr6	40359815	C/T	2780	746	R>Q	missense- non-syn	subclonal 1	0.123
B.I.1-T1	ENSG00000163637	ENST00000295902	PRICKLE2	chr3	64142965	G/A	1059	158	P>L	missense- non-syn	subclonal 1	0.111
B.I.1-T1	ENSG00000163002	ENST00000295119	NUP35	chr2	184025818	A/G	1043	314	S>G	missense- non-syn	subclonal 2	0.086
B.I.1-T1	ENSG00000127080	ENST00000287996	IPPK	chr9	95411836	G/C	590	105	L>V	missense- non-syn	subclonal 2	0.084
B.I.1-T1	ENSG00000204514	ENST00000435989	ZNF814	chr19	58385546	G/T	1484	404	D>E	missense- non-syn	subclonal 2	0.053
B.I.1-T1	ENSG00000185133	ENST00000331075	INPP5J	chr22	31521029	C/T	353	102	R>C	missense- non-syn	subclonal 2	0.04
B.I.1-T1	ENSG00000158683	ENST00000289672	PKD1L1	chr7	47880145	G/A	5517	1822	G>S	missense-syn	clonal	0.351
B.I.1-T1	ENSG00000164284	ENST00000329271	GRPEL2	chr5	148730509	G/A	452	114	V>V	missense-syn	clonal	0.315
B.I.1-T1	ENSG00000203780	ENST00000368692	FANK1	chr10	127668764	C/T	48	16	V>V	missense-syn	clonal	0.314
B.I.1-T1	ENSG00000086102	ENST00000379521	NFX1	chr9	33351740	A/C	2668	869	A>A	missense-syn	clonal	0.222
B.I.1-T1	ENSG00000101670	ENST00000261292	LIPG	chr18	47109974	C/T	1458	402	T>T	missense-syn	subclonal 1	0.167
B.I.1-T1	ENSG00000067798	ENST00000228327	NAV3	chr12	78362426	C/T	788	205	H>H	missense-syn	subclonal 1	0.137
B.I.1-T1	ENSG00000228120	ENST00000433840	AP001631.10	chr21	44579820	G/A	408	74	P>P	missense-syn	subclonal 1	0.122
B.I.1-T1	ENSG00000204525	ENST00000539307	HLA-C	chr6	31238931	G/A	649	217	L>L	missense-syn	subclonal 1	0.12
B.I.1-T1	ENSG00000148516	ENST00000361642	ZEB1	chr10	31809520	G/A	1323	420	A>A	missense-syn	subclonal 2	0.071
B.I.1-T1	ENSG00000243543	ENST00000372665	WFDC6	chr20	44167975	G/A	72	24	H>H	missense-syn	subclonal 2	0.058
B.I.1-T1	ENSG00000245848	ENST00000498907	CEBPA	chr19	33792479	T/TTC	992	281	N>R	insertion (fs)	clonal	0.253
B.I.1-T1	ENSG00000101191	ENST00000266070	DIDO1	chr20	61513343	AAG/A	4291	1322	AP>AL|AK	deletion (fs)	clonal	0.271
B.I.1-T1	ENSG00000145996	ENST00000274695	CDKAL1	chr6	21065380	TC/T	1324	386	F>F	deletion (fs)	clonal	0.237
B.I.1-T1	ENSG00000111880	ENST00000265607	RNGTT	chr6	89638988	TA/T	splice site	splice site	splice site	deletion (fs)	clonal	0.313
B.I.1-T2	ENSG00000131089	ENST00000253401	ARHGEF9	chrX	62885780	C/A	1843	348	V>F	missense- non-syn	clonal	0.333
B.I.1-T2	ENSG00000150990	ENST00000308736	DHX37	chr12	125441648	C/T	2290	731	V>M	missense- non-syn	clonal	0.309
B.I.1-T2	ENSG00000198077	ENST00000301146	CYP2A7	chr19	41387647	A/G	732	64	C>R	missense- non-syn	clonal	0.286
B.I.1-T2	ENSG00000161204	ENST00000292808	ABCF3	chr3	183904005	T/C	44	4	C>R	missense- non-syn	clonal	0.283
B.I.1-T2	ENSG00000147724	ENST00000160713	FAM135B	chr8	139263192	C/T	533	145	R>Q	missense- non-syn	clonal	0.281
B.I.1-T2	ENSG00000206102	ENST00000382822	KRTAP19-8	chr21	32410615	G/A	181	50	R>*	missense- non-syn	subclonal 1	0.269
B.I.1-T2	ENSG00000198077	ENST00000301146	CYP2A7	chr19	41387656	A/T	723	61	F>I	missense- non-syn	subclonal 1	0.267
B.I.1-T2	ENSG00000187672	ENST00000288221	ERC2	chr3	56044587	G/A	2066	604	R>*	missense- non-syn	subclonal 1	0.238
B.I.1-T2	ENSG00000163520	ENST00000295761	FBLN2	chr3	13679659	T/G	316	106	V>G	missense- non-syn	subclonal 1	0.216
B.I.1-T2	ENSG00000087495	ENST00000371015	PHACTR3	chr20	58330382	C/A	971	168	D>E	missense- non-syn	subclonal 2	0.135
B.I.1-T2	ENSG00000178809	ENST00000323819	TRIM73	chr7	75028240	T/C	223	8	L>P	missense- non-syn	subclonal 3	0.067
B.I.1-T2	ENSG00000227184	ENST00000525985	EPPK1	chr8	144940290	C/G	7204	2378	D>H	missense- non-syn	subclonal 3	0.067
B.I.1-T2	ENSG00000225190	ENST00000430334	PLEKHM1	chr17	43531297	G/A	2091	641	P>S	missense- non-syn	subclonal 3	0.058
B.I.1-T2	ENSG00000021645	ENST00000330071	NRXN3	chr14	78709649	C/T	213	71	G>G	missense-syn	subclonal 1	0.207
B.I.1-T2	ENSG00000189058	ENST00000421243	APOD	chr3	195295912	C/T	649	171	V>V	missense-syn	subclonal 1	0.242
B.I.1-T2	ENSG00000196578	ENST00000358642	OR5AC2	chr3	97806611	C/T	595	199	L>L	missense-syn	subclonal 1	0.274
B.I.1-T2	ENSG00000180592	ENST00000449193	SKIDA1	chr10	21805461	TGC/T	3544	431	GS>ER	deletion (fs)	clonal	0.333
B.I.1-T2	ENSG00000245848	ENST00000498907	CEBPA	chr19	33792381	CCTT/C	1090	314	KV>VL	deletion	subclonal 1	0.229
B.I.1-T3	ENSG00000106462	ENST00000320356	EZH2	chr7	148513776	C/T	1627	502	R>Q	missense- non-syn	clonal	0.5
B.I.1-T3	ENSG00000108055	ENST00000361804	SMC3	chr10	112343991	G/A	1268	381	R>Q	missense- non-syn	clonal	0.478
B.I.1-T3	ENSG00000134917	ENST00000414575	ADAMTS8	chr11	130286202	C/T	1207	403	D>N	missense- non-syn	clonal	0.478
B.I.1-T3	ENSG00000196169	ENST00000389916	KIF19	chr17	72346692	C/G	1504	456	R>G	missense- non-syn	clonal	0.42
B.I.1-T3	ENSG00000064042	ENST00000313860	LIMCH1	chr4	41621279	C/G	811	253	R>G	missense- non-syn	clonal	0.407
B.I.1-T3	ENSG00000144278	ENST00000392825	GALNT13	chr2	155099306	C/T	1141	192	R>C	missense- non-syn	clonal	0.4
B.I.1-T3	ENSG00000107863	ENST00000396432	ARHGAP21	chr10	24873709	G/A	5996	1837	R>W	missense- non-syn	clonal	0.4
B.I.1-T3	ENSG00000136040	ENST00000258526	PLXNC1	chr12	94631501	T/C	2291	681	I>T	missense- non-syn	clonal	0.383
B.I.1-T3	ENSG00000136040	ENST00000258526	PLXNC1	chr12	94631525	G/A	2315	689	R>Q	missense- non-syn	clonal	0.382
B.I.1-T3	ENSG00000152443	ENST00000317178	ZNF776	chr19	58265496	G/A	1261	333	R>H	missense- non-syn	clonal	0.377
B.I.1-T3	ENSG00000139625	ENST00000267079	MAP3K12	chr12	53880222	C/T	NA	NA	splice site	splice site	clonal	0.368
B.I.1-T3	ENSG00000130396	ENST00000351017	MLLT4	chr6	168352491	G/A	4629	1486	R>Q	missense- non-syn	clonal	0.336
B.I.1-T3	ENSG00000174937	ENST00000312240	OR5M3	chr11	56237948	T/C	67	9	E>G	missense- non-syn	clonal	0.333
B.I.1-T3	ENSG00000168769	ENST00000513237	TET2	chr4	106156747	C/T	2508	571	R>*	missense- non-syn	clonal	0.309
B.I.1-T3	ENSG00000179348	ENST00000341105	GATA2	chr3	128202758	A/G	1294	321	L>P	missense- non-syn	subclonal	0.25
B.I.1-T3	ENSG00000169814	ENST00000437172	BTD	chr3	15686958	C/T	1833	534	T>M	missense- non-syn	subclonal	0.185
B.I.1-T3	ENSG00000185963	ENST00000356884	BICD2	chr9	95526977	G/T	118	17	A>E	missense- non-syn	subclonal	0.159
B.I.1-T3	ENSG00000119535	ENST00000331941	CSF3R	chr1	36933434	G/A	1853	618	T>I	missense- non-syn	subclonal	0.143
B.I.1-T3	ENSG00000169704	ENST00000307395	GP9	chr3	128780901	G/T	541	107	A>S	missense- non-syn	subclonal	0.127
B.I.1-T3	ENSG00000072858	ENST00000264852	SIDT1	chr3	113251962	C/A	820	32	P>T	missense- non-syn	subclonal	0.125
B.I.1-T3	ENSG00000123643	ENST00000243389	SLC36A1	chr5	150844159	C/A	542	107	R>S	missense- non-syn	subclonal	0.1
B.I.1-T3	ENSG00000136231	ENST00000258729	IGF2BP3	chr7	23353160	A/G	1865	503	I>T	missense- non-syn	subclonal	0.098
B.I.1-T3	ENSG00000070413	ENST00000263196	DGCR2	chr22	19055730	C/A	459	71	G>W	missense- non-syn	subclonal	0.06
B.I.1-T3	ENSG00000158683	ENST00000289672	PKD1L1	chr7	47880145	G/A	5517	1822	G>G	missense-syn	subclonal	0.556
B.I.1-T3	ENSG00000198793	ENST00000361445	MTOR	chr1	11190823	C/T	5453	1792	A>A	missense-syn	subclonal	0.453
B.I.1-T3	ENSG00000203780	ENST00000368692	FANK1	chr10	127668764	C/T	48	16	V>V	missense-syn	subclonal	0.437
B.I.1-T3	ENSG00000164284	ENST00000329271	GRPEL2	chr5	148730509	G/A	452	114	V>V	missense-syn	subclonal	0.435
B.I.1-T3	ENSG00000159423	ENST00000290597	ALDH4A1	chr1	19209636	C/T	690	220	A>A	missense-syn	subclonal	0.333
B.I.1-T3	ENSG00000184408	ENST00000331113	KCND2	chr7	119915718	C/T	1997	344	Y>Y	missense-syn	subclonal	0.096
B.I.1-T3	ENSG00000236637	ENST00000421715	IFNA4	chr9	21187003	C/A	596	176	S>S	missense-syn	subclonal	0.079
B.I.1-T3	ENSG00000245848	ENST00000498907	CEBPA	chr19	33792479	T/TTC	992	281	NS>RT	insertion (fs)	subclonal	0.453
B.I.1-T3	ENSG00000101191	ENST00000266070	DIDO1	chr20	61513343	AAG/A	4291	1322	AP>AL|AK	deletion (fs)	subclonal	0.326
B.I.1-T3	ENSG00000145996	ENST00000274695	CDKAL1	chr6	21065380	TC/T	1324	386	F>F	deletion (fs)	subclonal	0.470
B.I.1-T1/T3	ENSG00000108055	ENST00000361804	SMC3	chr10	112343991	G/A	1268	381	R>Q	missense- non-syn	clonal	0.294/0.478
B.I.1-T1/T3	ENSG00000174937	ENST00000312240	OR5M3	chr11	56237948	T/C	67	9	E>G	missense- non-syn	clonal	0.283/0.333
B.I.1-T1/T3	ENSG00000134917	ENST00000414575	ADAMTS8	chr11	130286202	C/T	1207	403	D>N	missense- non-syn	clonal	0.429/0.478
B.I.1-T1/T3	ENSG00000107863	ENST00000396432	ARHGAP21	chr10	24873709	G/A	5996	1837	R>W	missense- non-syn	clonal	0.243/0.4
B.I.1-T1/T3	ENSG00000203780	ENST00000368692	FANK1	chr10	127668764	C/T	48	16	V>V	missense-syn	clonal	0.314/0.437
B.I.1-T1/T3	ENSG00000164284	ENST00000329271	GRPEL2	chr5	148730509	G/A	452	114	V>V	missense-syn	clonal	0.315/0.435
B.I.1-T1/T3	ENSG00000158683	ENST00000289672	PKD1L1	chr7	47880145	G/A	5517	1822	G>G	missense-syn	clonal	0.351/0.556
B.I.1-T1/T3	ENSG00000245848	ENST00000498907	CEBPA	chr19	33792479	T/TTC	992	281	NS>RT	insertion (fs)	clonal	0.253/0.453
B.I.1-T1/T3	ENSG00000101191	ENST00000266070	DIDO1	chr20	61513343	AAG/A	4291	1322	AP>AL|AK	deletion (fs)	clonal	0.271/0.326
B.I.1-T1/T3	ENSG00000145996	ENST00000274695	CDKAL1	chr6	21065380	TC/T	1324	386	F>F	deletion (fs)	clonal	0.237/0.47
B.II.2	ENSG00000114019	ENST00000514516	AMOTL2*	chr3	134078205	G/A	2377	733	A>V	missense-non-syn	clonal	0.49
B.II.2	ENSG00000070729	ENST00000251102	CNGB1*	chr16	57918271	C/T	3610	1185	A>T	missense-non-syn	clonal	0.455
B.II.2	ENSG00000078018	ENST00000199940	MAP2*	chr2	210588406	G/A	1704	422	G>R	missense-non-syn	clonal	0.408
B.II.2	ENSG00000070366	ENST00000263073	SMG6*	chr17	2186155	T/C	NA	NA		splice site	clonal	0.38
B.II.2	ENSG00000157103	ENST00000287766	SLC6A1*	chr3	11064099	T/C	1080	220	I>T	missense-non-syn	clonal	0.354
B.II.2	ENSG00000145113	ENST00000462323	MUC4	chr3	195505801	G/T	12722	4217	A>E	missense-non-syn	subclonal 2	0.143
B.II.2	ENSG00000257103	ENST00000433627	AC010614.2	chr19	34710362	G/C	1044	283	G>A	missense-non-syn	subclonal 2	0.136
B.II.2	ENSG00000145113	ENST00000462323	MUC4	chr3	195505762	A/G	12761	4230	L>P	missense-non-syn	subclonal 2	0.128
B.II.2	ENSG00000221957	ENST00000339924	KIR2DS4*	chr19	55346565	G/T	155	22	Q>H	missense-non-syn	subclonal 2	0.087
B.II.2	ENSG00000005801	ENST00000533036	ZNF195	chr11	3383228	G/T	373	87	C>*	nonsense	subclonal 2	0.083
B.II.2	ENSG00000102034	ENST00000308167	ELF4*	chrX	129208104	G/T	640	87	A>D	missense-non-syn	subclonal 2	0.079
B.II.2	ENSG00000034152	ENST00000529517	MAP2K3	chr17	21199051	T/C	426	72	I>T	missense-non-syn	subclonal 2	0.069
B.II.2	ENSG00000189190	ENST00000338230	ZNF600	chr19	53268975	A/G	2320	678	D>D	missense-syn	subclonal 2	0.434
B.II.2	ENSG00000139364	ENST00000299308	TMEM132B*	chr12	126138782	G/A	2771	921	A>A	missense-syn	subclonal 1	0.214
B.II.2	ENSG00000257103	ENST00000433627	AC010614.2	chr19	34710315	T/G	997	267	A>A	missense-syn	subclonal 2	0.13
B.II.2	ENSG00000174306	ENST00000309060	ZHX3	chr20	39832369	G/A	1188	396	V>V	missense-syn	subclonal 2	0.099
B.II.2	ENSG00000129255	ENST00000250124	MPDU1*	chr17	7489087	G/T	348	47	G>G	missense-syn	subclonal 2	0.089
B.II.2	ENSG00000145113	ENST00000462323	MUC4	chr3	195505836	G/A	12687	4205	H>H	missense-syn	subclonal 2	0.088
B.II.2	ENSG00000186715	ENST00000334998	MST1P9	chr1	17087593	C/T	72	24	Q>Q	missense-syn	subclonal 2	0.085
B.II.2	ENSG00000013573	ENST00000350437	DDX11	chr12	31244703	C/T	1338	380	A>A	missense-syn	subclonal 2	0.077
B.II.2	ENSG00000245848	ENST00000498907	CEBPA*	chr19	33792403	G/GCGC	1068	306	RR>RR	insertion (fs)	clonal	0.359
B.II.2	ENSG00000214681	ENST00000398778	IQCF5	chr3	51911117	TC/T	25	9	SR>SR|SG	deletion (fs)	clonal	0.813

Exclusion of variants confirmed as sequencing artifacts by manual review of WES data and Sanger sequencing
Using EXPANDS, tumor purity was estimated as 53% (T1), 65% (T2) and 65% (T3)
Key: * Variants Sanger sequenced at diagnosis and recurrence in B.II.2 to determine pattern of clonal evolution. With the exception of variants in subclone 2 (VAF<10%), all were verified in the diagnostic sample but were absent at relapse.

Table 4 Somatic WES variants verified with Sanger and/or deep sequencing

Sample	Gene symbol	Chromosome	Genomic position of mutation	Mutation type	Validation method
A.II.1	GPRIN3	chr4	90169380	SNV	Sanger
A.II.1	GATA2	chr3	128202759	SNV	Sanger
A.II.1	KIAA0586	chr14	58924643	SNV	Sanger
A.II.1	ABCA9	chr17	67016593	SNV	Sanger
A.II.1	HGD	chr3	120360494	SNV	Sanger
A.II.1	CCNK	chr14	99976750	SNV	Sanger
A.II.1	TKTL1	chrX	153533594	SNV	Sanger
A.II.1	ODF1	chr8	103573010	Indel	Sanger
A.II.1	CEBPA	chr19	33792382	Indel	Sanger
A.II.5	SGK494	chr17	26938181	SNV	Sanger
A.II.5	ARHGAP33	chr19	36271122	SNV	Sanger
A.II.5	GATA2	chr3	128202731	SNV	Sanger
A.II.5	DYSF	chr2	71896337	SNV	Sanger
A.II.5	NEURL4	chr17	7224982	SNV	Sanger
A.II.5	GATA2	chr3	128202759	SNV	Sanger
A.II.5	C10orf46	chr10	120445456	Indel	Sanger
A.II.5	PLCXD3	chr5	41312787	Indel	Sanger
A.II.5	CEBPA	chr19	33792381	Indel	Sanger
A.III.2	MXRA5	chrX	3242301	SNV	Sanger
A.III.2	GATA2	chr3	128202770	SNV	Sanger
A.III.2	EXOC7	chr17	74097326	SNV	Sanger
A.III.2	CYP2C18	chr10	96480153	SNV	Sanger
A.III.2	CEBPA	chr19	33792381	Indel	Sanger
A.III.2	SMCHD1	chr18	2784945	Indel	Sanger
B.I.1-T1	ADAMTS8	chr11	130286202	SNV	Sanger
B.I.1-T1	SMC3	chr10	112343991	SNV	Sanger/Deep Seq
B.I.1-T1	OR5M3	chr11	56237948	SNV	Sanger
B.I.1-T1	ARHGAP21	chr10	24873709	SNV	Sanger
B.I.1-T1	PCYOX1L	chr5	148742349	SNV	Sanger/Deep Seq
B.I.1-T1	ACOT11	chr1	55068352	SNV	Sanger
B.I.1-T1	KRT8	chr12	53298675	SNV	Sanger
B.I.1-T1	EZH2	chr7	148523708	SNV	Sanger/Deep Seq
B.I.1-T1	PRICKLE2	chr3	64142965	SNV	Deep Seq
B.I.1-T1	NUP35	chr2	184025818	SNV	Deep Seq
B.I.1-T1	PKD1L1	chr7	47880145	SNV	Sanger/Deep Seq
B.I.1-T1	GRPEL2	chr5	148730509	SNV	Sanger/Deep Seq
B.I.1-T1	FANK1	chr10	127668764	SNV	Sanger
B.I.1-T1	LIPG	chr18	47109974	SNV	Deep Seq
B.I.1-T1	ZEB1	chr10	31809520	SNV	Deep Seq
B.I.1-T1	WFDC6	chr20	44167975	SNV	Deep Seq
B.I.1-T1	SERINC2	chr1	31886750	Indel	Sanger
B.I.1-T1	UBE3B	chr12	109928630	Indel	Sanger
B.I.1-T1	RND3	chr2	151343986	Indel	Sanger
B.I.1-T1	CEBPA	chr19	33792479	Indel	Sanger/Deep Seq
B.I.1-T1	DIDO1	chr20	61513343	Indel	Sanger/Deep Seq
B.I.1-T1	CDKAL1	chr6	21065380	Indel	Sanger/Deep Seq
B.I.1-T2	ARHGEF9	chrX	62885780	SNV	Sanger/Deep Seq
B.I.1-T2	DHX37	chr12	125441648	SNV	Sanger/Deep Seq
B.I.1-T2	CYP2A7	chr19	41387647	SNV	Sanger
B.I.1-T2	ABCF3	chr3	183904005	SNV	Sanger/Deep Seq
B.I.1-T2	FAM135B	chr8	139263192	SNV	Sanger
B.I.1-T2	CYP2A7	chr19	41387656	SNV	Sanger
B.I.1-T2	ERC2	chr3	56044587	SNV	Sanger/Deep Seq
B.I.1-T2	NRXN3	chr14	78709649	SNV	Sanger
B.I.1-T2	APOD	chr3	195295912	SNV	Sanger
B.I.1-T2	OR5AC2	chr3	97806611	SNV	Sanger/Deep Seq
B.I.1-T2	KIAA1377	chr11	101868435	SNV	Sanger
B.I.1-T2	SKIDA1	chr10	21805461	Indel	Sanger
B.I.1-T2	CEBPA	chr19	33792381	Indel	Sanger/Deep Seq
B.I.1-T2	WT1	chr11	32450112	Indel	Sanger
B.I.1-T3	EZH2	chr7	148513776	SNV	Sanger/Deep Seq
B.I.1-T3	SMC3	chr10	112343991	SNV	Sanger/Deep Seq
B.I.1-T3	ADAMTS8	chr11	130286202	SNV	Sanger
B.I.1-T3	KIF19	chr17	72346692	SNV	Sanger/Deep Seq
B.I.1-T3	LIMCH1	chr4	41621279	SNV	Sanger/Deep Seq
B.I.1-T3	ARHGAP21	chr10	24873709	SNV	Sanger
B.I.1-T3	MAP3K12	chr12	53880222	SNV	Sanger/Deep Seq
B.I.1-T3	MLLT4	chr6	168352491	SNV	Sanger
B.I.1-T3	OR5M3	chr11	56237948	SNV	Sanger
B.I.1-T3	TET2	chr4	106156747	SNV	Sanger/Deep Seq
B.I.1-T3	GATA2	chr3	128202758	SNV	Sanger/Deep Seq
B.I.1-T3	BTD	chr3	15686958	SNV	Sanger
B.I.1-T3	CSF3R	chr1	36933434	SNV	Deep Seq
B.I.1-T3	PKD1L1	chr7	47880145	SNV	Sanger/Deep Seq
B.I.1-T3	MTOR	chr1	11190823	SNV	Sanger/Deep Seq
B.I.1-T3	FANK1	chr10	127668764	SNV	Sanger
B.I.1-T3	GRPEL2	chr5	148730509	SNV	Sanger/Deep Seq
B.I.1-T3	ALDH4A1	chr1	19209636	SNV	Sanger/Deep Seq
B.I.1-T3	CEBPA	chr19	33792479	Indel	Sanger/Deep Seq
B.I.1-T3	DIDO1	chr20	61513343	Indel	Sanger/Deep Seq
B.I.1-T3	CDKAL1	chr6	21065380	Indel	Sanger/Deep Seq
B.I.1-T3	SERINC2	chr1	31886750	Indel	Sanger
B.I.1-T3	C5orf35	chr5	56221321	Indel	Sanger
B.II.2	AMOTL2	chr3	134078205	SNV	Sanger
B.II.2	CNGB1	chr16	57918271	SNV	Sanger
B.II.2	MAP2	chr2	210588406	SNV	Sanger
B.II.2	SMG6	chr17	2186155	SNV	Sanger
B.II.2	AC010614.2	chr19	34710362	SNV	Sanger
B.II.2	SLC6A1	chr3	11064099	SNV	Sanger
B.II.2	ZNF600	chr19	53268975	SNV	Sanger
B.II.2	TMEM132B	chr12	126138782	SNV	Sanger
B.II.2	COL8A2	chr1	36563080	Indel	Sanger
B.II.2	CEBPA	chr19	33792403	Indel	Sanger
C.III.1	WT1	chr11	32456326	SNV	Sanger
C.III.1	POMT1	chr9	134387477	SNV	Sanger
C.III.1	OR10G8	chr11	123900388	SNV	Sanger
C.III.1	LETM2	chr8	38251663	SNV	Sanger
D.II.2	AP001496.1	chr18	5233143	SNV	Sanger
D.II.2	WT1	chr11	32421575	SNV	Sanger
D.II.2	FCGBP	chr19	40408120	SNV	Sanger
D.II.2	MARCH10	chr17	60814250	SNV	Sanger
D.II.2	TRAV1-1	chr14	22090620	SNV	Sanger
D.II.2	ADRA1B	chr5	159399022	SNV	Sanger
D.II.2	SP110	chr2	231067434	SNV	Sanger
D.II.2	CCT7	chr2	73474943	SNV	Sanger
D.II.2	CEBPA	chr19	33792381	Indel	Sanger
D.II.2	CRYAA	chr21	44589920	Indel	Sanger

107 variants were verified with Sanger +/- deep sequencing, achieving concordance of 83% for indels and 93% for S

Appendix 3

Supplementary Data for Chapter 4: Convergence of somatic JAK-STAT mutations in a novel *RUNX1*-mutated pedigree

Table 1 Somatic mutations identified in sibling tumour samples using WES*

Sample	Ensembl Gene ID	Ensembl Transcript ID	Gene Symbol	Chromosome	Genomic position	Nucleotide change (ref/variant)	cDNA position	aa position	aa change	Mutation type	VAF
K.II.1	ENSG00000111252	ENST00000341259	SH2B3	chr12	111885287	G/A	1532	392	R>Q	missense- non-syn	0.672
K.II.1	ENSG00000182704	ENST00000333090	TSKU	chr11	76507435	C/G	949	259	L>V	missense- non-syn	0.523
K.II.1	ENSG00000198911	ENST00000361204	SREBF2	chr22	42300883	C/A	3276	1037	A>D	missense- non-syn	0.422
K.II.1	ENSG00000179044	ENST00000314586	EXOC3L1	chr16	67218676	A/G	2184	648	L>P	missense- non-syn	0.389
K.II.1	ENSG00000168959	ENST00000305432	GRM5	chr11	88337977	G/A	1453	435	R>W	missense- non-syn	0.364
K.II.1	ENSG00000006283	ENST00000359106	CACNA1G	chr17	48697116	G/A	5854	1952	E>K	missense- non-syn	0.354
K.II.1	ENSG00000241935	ENST00000370646	HOGA1	chr10	99371366	G/A	1295	312	A>T	missense- non-syn	0.352
K.II.1	ENSG00000177689	ENST00000356790	MAGEB10	chrX	27839737	G/T	559	105	S>I	missense- non-syn	0.331
K.II.1	ENSG00000134222	ENST00000369903	PSRC1	chr1	109824633	G/C	267	43	R>G	missense- non-syn	0.200
K.II.1	ENSG00000111642	ENST00000309577	CHD4	chr12	6697051	C/T	3694	1177	R>H	missense- non-syn	0.056
K.II.2	ENSG00000096968	ENST00000381652	JAK2	chr9	5073770	G/T	2343	617	V>F	missense- non-syn	0.682
K.II.2	ENSG00000015475	ENST00000342111	BID	chr22	18222879	AC/A	377	96	S>V|S	frameshift	0.375
K.II.2	ENSG00000007171	ENST00000313735	NOS2	chr17	26093566	G/T	2450	739	S>Y	missense- non-syn	0.336
K.II.2	ENSG00000143889	ENST00000410076	HNRPLL	chr2	38809018	C/A	886	275	C>F	missense- non-syn	0.333
K.II.2	ENSG00000101773	ENST00000327155	RBBP8	chr18	20572749	A/G	1307	320	E>G	missense- non-syn	0.323
K.II.2	ENSG00000151012	ENST00000280612	SLC7A11	chr4	139104402	G/A	1253	325	L>F	missense- non-syn	0.288
K.II.2	ENSG00000005884	ENST00000320031	ITGA3	chr17	48165611	G/A	3398	1023	R>H	missense- non-syn	0.244
K.II.2	ENSG00000004897	ENST00000531206	CDC27	chr17	45214548	A/G	1905	634	I>T	missense- non-syn	0.167
K.II.2	ENSG00000197689	ENST00000329040	TBC1D29	chr17	28890301	G/A	460	104	S>N	missense- non-syn	0.093
K.II.2	ENSG00000103365	ENST00000309859	GGA2	chr16	23481460	C/A	1560	493	D>Y	missense- non-syn	0.081
K.II.3	ENSG00000132196	ENST00000254521	HSD17B7	chr1	162762514	C/T	156	34	A>V	missense- non-syn	0.495
K.II.3	ENSG00000253873	ENST00000398587	PCDHGA11	chr5	140801000	C/T	239	69	S>F	missense- non-syn	0.477
K.II.3	ENSG00000222036	ENST00000409832	POTEG	chr14	19574214	G/C	1323	424	S>T	missense- non-syn	0.200
K.II.3	ENSG00000244482	ENST00000245621	LILRA6	chr19	54745550	G/T	697	187	T>N	missense- non-syn	0.103
K.II.3	ENSG00000004897	ENST00000531206	CDC27	chr17	45219311	T/C	1481	493	I>V	missense- non-syn	0.075
K.II.3	ENSG00000072736	ENST00000329524	NFATC3	chr16	68156512	C/A	750	242	C>*	nonsense	0.049
K.II.3	ENSG00000079432	ENST00000160740	CIC	chr19	42796978	C/A	3476	1146	P>T	missense- non-syn	0.048
K.II.3	ENSG00000124103	ENST00000371328	C20orf106	chr20	55100084	C/A	543	74	Q>K	missense- non-syn	0.045
K.II.3	ENSG00000138823	ENST00000511045	MTTP	chr4	100532602	G/A	2075	688	G>S	missense- non-syn	0.042
K.II.4	ENSG00000169783	ENST00000355300	LINGO1	chr15	77906780	T/C	1521	490	Q>R	missense- non-syn	0.535
K.II.4	ENSG00000063244	ENST00000308924	U2AF2	chr19	56172508	C/G	479	147	Q>E	missense- non-syn	0.412
K.II.4	ENSG00000236669	ENST00000444154	AC006372.1	chr7	157318686	G/A	210	47	R>H	missense- non-syn	0.358
K.II.4	ENSG00000105655	ENST00000338128	ISYNA1	chr19	18547585	C/T	736	173	R>Q	missense- non-syn	0.349
K.II.4	ENSG00000130758	ENST00000253055	MAP3K10	chr19	40710439	G/A	1199	304	R>H	missense- non-syn	0.286
K.II.4	ENSG00000173040	ENST00000344408	EVC2	chr4	5682988	G/A	923	290	T>M	missense- non-syn	0.241
K.II.4	ENSG00000101343	ENST00000377340	CRNKL1	chr20	20024164	C/T	1459	476	R>Q	missense- non-syn	0.210
K.II.4	ENSG00000096968	ENST00000381652	JAK2	chr9	5073770	G/T	2343	617	V>F	missense- non-syn	0.200
K.II.4	ENSG00000028203	ENST00000397796	VEZT	chr12	95668655	C/T	986	329	P>L	missense- non-syn	0.193
K.II.4	ENSG00000011454	ENST00000373647	RABGAP1	chr9	125772742	G/C	1618	495	R>P	missense- non-syn	0.136
K.II.4	ENSG00000124103	ENST00000371328	C20orf106	chr20	55100084	C/A	543	74	Q>K	missense- non-syn	0.102

*Exclusion of variants deemed to be sequencing artifacts by manual review of WES and Sanger Sequencing

Table 2 Somatic WES variants verified in pedigree K with Sanger sequencing

Sample	Gene symbol	Chromosome	Genomic position of mutation	Mutation type	Validation method
K.II.1	*SH2B3*	chr12	111885287	SNV	Sanger
K.II.2	*JAK2*	chr9	5073770	SNV	Sanger
K.II.2	*RBBP8*	chr18	20572749	SNV	Sanger
K.II.2	*CDC27*	chr17	45214548	SNV	Sanger
K.II.3	*PCDHGA11*	chr5	140801000	SNV	Sanger
K.II.4	*U2AF2*	chr19	56172508	SNV	Sanger
K.II.4	*JAK2*	chr9	5073770	SNV	Sanger

Appendix 4

Supplementary data for Chapter 5: Novel and atypical germline mutations in familial AML

Table 1 Germline variants verified by Sanger sequencing in pedigrees P-T

Sample	Gene symbol	Chromosome	Genomic position of mutation	Mutation (ref/var)	Amino acid	VAF	Validation method
P.III.2 (CR/PB)	*GATAD2B*	chr1	153800510	A/C	V105G	0.42/0.40	Sanger
P.III.2 (CR/PB)	*MYST3*	chr8	41789998	T/C	M1914V	0.48/0.50	Sanger
P.III.2 (CR/PB)	*RASA3*	chr13	114774873	G/A	A553V	0.52/0.44	Sanger
P.III.2 (CR/PB)	*PSMB3*	chr17	36909577	G/A	V60I	0.55/0.62	Sanger
P.III.2 (CR/PB)	*UPF1*	chr19	18974413	A/G	I923V	0.47/0.63	Sanger
P.III.2 (CR/PB)	*USP16*	chr21	30426208	C/G	T760S	0.50/0.51	Sanger
Q.I.1/Q.II.1	*KANK1*	chr9	742275	G/C	C1256S	0.46/0.66	Sanger
Q.I.1/Q.II.1	*CHEK1*	chr11	125495891	G/A	W79X	0.71/0.44	Sanger
Q.I.1/Q.II.1	*FANCA*	chr16	89845230	G/C	A602G	0.42/0.59	Sanger
R.III.2/R.III.3	ARFGEF1	chr8	68213592	G/A	A60V	0.44/0.50	Sanger
R.III.2/R.III.3	ZCCHC8	chr12	122968061	C/T	G184E	0.44/0.47	Sanger
R.III.2/R.III.3	RALGAPB	chr20	37187086	C/T	T1174M	0.40/0.41	Sanger
R.III.2/R.III.3	ARVCF	chr22	19967551	C/T	V371M	0.49/0.42	Sanger
S.I.1/S.II.1	*BIRC6*	chr2	32702454	C/T	R2291C	0.52/0.38	Sanger
S.I.1/S.II.1	*ATR*	chr3	142218500	C/A	Q1783H	0.48/0.51	Sanger
S.I.1/S.II.1	*TUBB*	chr6	30688326	C/T	Q15X	0.50/0.30	Sanger
S.I.1/S.II.1	*NOD1*	chr7	30492324	C/T	R237C	0.43/0.39	Sanger
S.I.1/S.II.1	*BMP1*	chr8	22034583	C/T	F221L	0.40/0.51	Sanger
S.I.1/S.II.1	*KMT2D*	chr12	49420696	G/C	T5018R	0.51/0.51	Sanger
S.I.1/S.II.1	*TP53BP1*	chr15	43749215	G/C	P526A	0.55/0.43	Sanger
S.I.1/S.II.1	*MED1*	chr17	37565756	C/A	L906F	0.55/0.52	Sanger
S.I.1/S.II.1	*ACSS2*	chr20	33509227	C/T	R458C	0.52/0.42	Sanger
T.II.1/T.II.2	*REV3L*	chr6	111685100	T/G	K2279Q	0.52/0.54	Sanger
T.II.1/T.II.2	*TCF25*	chr16	89967047	A/G	H409R	0.45/0.49	Sanger
T.II.1/T.II.2	*ANKRD11*	chr16	89346343	C/T	L2203F	0.47/0.47	Sanger
T.II.1/ T.II.2	*C17ORF70*	chr17	79519292	C/A	A22E	0.57/0.41	Sanger

Table 2 Somatic coding variants detected in P.III.2*

Sample	Ensembl Gene ID	Ensembl Transcript ID	Gene Symbol	Chromosome	Genomic position	Nucleotide change (ref/variant)	cDNA position	aa position	aa change	Mutation type	VAF
P.III.2	ENSG00000110514	ENST00000311027	MADD	chr11	47303126	G/T	1456	431	A>S	missense- non-syn	0.5
P.III.2	ENSG00000156531	ENST00000332070	PHF6	chrX	133527937	A/G	SS	SS	N/A	Splice site	0.37
P.III.2	ENSG00000167210	ENST00000536736	LOXHD1	chr18	44146330	C/T	2327	776	R>H	missense- non-syn	0.31
P.III.2	ENSG00000147481	ENST00000276467	SNTG1	chr8	51362258	G/A	276	84	V>I	missense- non-syn	0.3
P.III.2	ENSG00000103479	ENST00000262133	RBL2	chr16	53501064	C/A	2066	653	T>N	missense- non-syn	0.27
P.III.2	ENSG00000111199	ENST00000261740	TRPV4	chr12	110230180	C/T	1962	627	G>S	missense- non-syn	0.25
P.III.2	ENSG00000157087	ENST00000342354	ATP2B2	chr3	10401723	G/A	2183	582	Q>*	nonsense	0.22
P.III.2	ENSG00000188505	ENST00000339852	NCCRP1	chr19	39691084	C/T	666	216	P>L	missense- non-syn	0.21
P.III.2	ENSG00000140157	ENST00000337451	NIPA2	chr15	23014486	G/A	852	80	A>V	missense- non-syn	0.2
P.III.2	ENSG00000171564	ENST00000302068	FGB	chr4	155488850	G/A	659	199	R>H	missense- non-syn	0.2
P.III.2	ENSG00000168769	ENST00000513237	TET2	chr4	106193931	C/T	5253	1486	R>*	nonsense	0.19
P.III.2	ENSG00000187905	ENST00000342608	AC002472.13	chr22	21402222	G/A	359	111	R>Q	missense- non-syn	0.16
P.III.2	ENSG00000133703	ENST00000256078	KRAS	chr12	25380276	T/C	246	61	Q>R	missense- non-syn	0.14
P.III.2	ENSG00000184185	ENST00000331718	KCNJ12	chr17	21318709	G/A	760	19	G>R	missense- non-syn	0.13
P.III.2	ENSG00000137210	ENST00000379542	TMEM14B	chr6	10756728	C/T	489	108	R>C	missense- non-syn	0.09
P.III.2	ENSG00000236032	ENST00000437310	OR5H14	chr3	97868656	C/T	487	143	R>W	missense- non-syn	0.08
P.III.2	ENSG00000130055	ENST00000374382	GDPD2	chrX	69646861	G/T	953	234	M>I	missense- non-syn	0.07
P.III.2	ENSG00000100197	ENST00000360608	CYP2D6	chr22	42523505	C/T	1232	373	G>S	missense- non-syn	0.07
P.III.2	ENSG00000144460	ENST00000272907	KIAA1486	chr2	226516156	G/A	2250	613	V>I	missense- non-syn	0.05
P.III.2	ENSG00000177426	ENST00000330513	TGIF1	chr18	3452222	CT/C	548	82	P>P	Indel	0.54
P.III.2	ENSG00000179979	ENST00000324803	CRIPAK	chr4	1388622	T/TCA	3283	108	LT>LT	Indel	0.4

*****shaded rows represent variants verified with Sanger sequencing