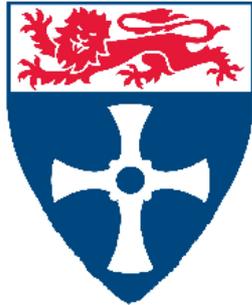


NEWCASTLE UNIVERSITY



Control of Lineage Commitment in Acute Leukaemia

Ricky Fong Tirtakusuma

Northern Institute for Cancer Research
Faculty of Medical Sciences
Doctor of Philosophy

September 2018

Abstract

Acute leukaemia with the t(4;11) translocation is strongly associated with pro B-acute lymphoblastic phenotype. Here is described a lineage switch from acute lymphoblastic leukaemia (ALL) to acute myeloid leukaemia (AML) which carries identical t(4;11) breakpoints that provides insight into regulation of lineage commitment and the haematopoietic origin of leukaemia.

Stable DNA microsatellite sequences argue against a therapy-related AML. Genome sequencing and RNAseq identified 12 novel and deleterious mutations unique to the AML. Immunoglobulin rearrangement analysis suggested the common cell of origin lied within a population prior to B cell differentiation. Sorting of haematopoietic stem/progenitor cell populations followed by multiplex PCR and next generation sequencing for the fusion and secondary mutations demonstrated the occurrence of the leukaemogenic *MLL-AF4* fusion gene in cell populations as early as the multipotent progenitor, MPP, population in both ALL and AML. In this most primitive population, the AML carries mutations in chromatin modulating genes CHD4 and PHF3, suggesting their importance in lineage commitment.

Knockdown CHD4 and PHF3 individually and in combination in the pro-B ALL t(4;11) SEM cell line resulted in ~3 fold higher expression of the myeloid cell surface marker CD33. Further analysis was performed using a recently described model of *MLL-AF4* leukaemogenesis consisting of CD34+ cord blood cells transduced with a chimeric *MLL-Af4* fusion gene. Knockdown of *CHD4* and *PHF3* resulted in loss of lymphoid differentiation potential *in vitro*.

Analysis of different PHF3 splice variants revealed that only mutation-carrying PHF3 variants increased CD33 on SEM cells and that a balance between PHF3 variants was required for the lineage fidelity.

This study suggests that the ALL and AML share a common primitive cell of origin and that mutations in CHD4 and PHF3 shift the lymphoid phenotype towards a myeloid lineage leukaemia.

Acknowledgment

I would like to express my biggest thanks to my supervisor Olaf Heidenreich for his continuous teaching, ideas, patience, massive support and care, which not just limited to the project. To make it shorter, his guidance is much more prominent than an excellent supervisor.

Thanks to Melanie, who was not only helping me a lot in the lab, but was also correcting my English in this thesis.

Also thanks to Natalia for providing many valuable ideas in this project, in particular, the multiplex PCR candidate genes on the haematopoietic hierarchy and PHF3 isoforms. I am sure these are highly essential points in this study, and at least these parts are not from Olaf's geniusness, but hers.

I want to thank Alex, for firstly, providing the primograft samples (together with Helen), and also always explaining me the subjects that I didn't even know where to start. His massive help was in particular at the beginning of the project, when I had the meetings with Olaf, didn't understand what he meant, Alex would always be able to re-explain it to me.

Massive thanks to Helen Blair for helping me transplant and harvest the primograft samples, also for her constant positive behaviour and support.

Thanks to Natalie, Katie, Sarah Fordham, Dan Coleman, Hesta, and Helen Marr that introduced and taught me to the lab work at the beginning of the project, and for making the lab extraordinarily nice and exciting. Hesta helped me a lot with the cell sorting. Sarah Fordham and James Allan analysed the microsatellite instability data and explained it to me in great detail. Helen taught me the immunophenotyping of the patient samples in a very comprehensive manner.

Thanks to Yuzhe, Hasan, Milene, Asmida, Eva, Peixun, Azira, Kasia, and Anja for not only being very nice, but also helping me with lab works, teaching some assays, and providing valuable discussion and ideas. Anja provided a highly valuable coffee automate. My huge thanks also to Lynne Minto, Marian, Liz, and Anne for being extremely helpful and keeping the lab running efficiently. Lynne, Marian, and Liz provided most of the patient samples. Also, Simon for reviewing this thesis.

Thanks to Martyna, Claire, Richard, Judith, Bailey, and Gary for always answering my questions about the basic protein works.

Thanks to Sarra Ryan and Paul Sinclair for teaching me the sequencing sample preparations, also still very helpful even though I ask plenty of questions!

To Rin, Louise, and Matt for helping analysing bioinformatics data.

Thanks to Paul Milne for helping the haematopoietic cell sorting and explaining me many things about the methods in flow cytometry.

Thanks to Claus Meyer and Rolf Marschalek for helping to identify the sequences of the MLLr samples. This is a huge help in determining the lineage switch case also finding the pre-leukaemic population.

Thanks to Ehud Shapiro's team, especially to Rivka Adar, who was very kindly helped me with the single cell assay and the troubleshooting. Also to Mulloy lab for the cord blood cells MLL/Af4 cells, especially to Shan Lin who explained to me very clearly how to work with the model.

Thanks to Gemma Llargues and Frederik as the source of ATP in the lab by always sharing the positivity and happy faces. I am sure they will still be positive even during torture.

Thanks to everyone in the NICR, including the past members, since I am sure I have asked many of the colleagues here for many things, including questions, reagents, chemicals, *etc.*

Lastly, thanks to the patients and the family. This study was coming from them and I hope I have not disappointed any of them and am grateful to them for giving their contributions and hopes.

Table of Contents

Abstract	i
Acknowledgment	iii
Abbreviation	xii
Chapter 1 Introduction	1
1.1. Haematopoiesis	1
1.2. Acute leukaemia	4
1.2.1. <i>Infant ALL</i>	5
1.2.2. <i>Paediatric acute myeloid leukaemia (AML)</i>	8
1.2.3. <i>MLL-rearranged leukaemia</i>	11
1.3. Leukaemia lineage switch.....	14
1.3.1. <i>Introduction to leukaemia lineage switch</i>	14
1.3.2. <i>Study cases</i>	15
1.4. Patient L826 and preliminary data	21
1.4.1. <i>Immunophenotypes</i>	22
1.4.2. <i>Chromosome study</i>	24
1.4.3. <i>Fusion gene breakpoint sequences</i>	25
1.4.4. <i>Microsatellite instability analysis</i>	25
1.4.5. <i>Whole genome, whole exome, and RNA sequencing</i>	28
1.5. Candidate driver genes.....	30
1.5.1. <i>ACAP1</i>	31
1.5.2. <i>CHD4</i>	32
1.5.3. <i>PHF3</i>	35
1.5.4. <i>PPP1R7</i>	37
Chapter 2 Materials and Methods	39
2.1. Materials	39
2.1.1. <i>Laboratory equipment</i>	39
2.1.2. <i>Chemicals and reagents</i>	40
2.1.3. <i>Buffers and media</i>	42
2.1.4. <i>Bacterial strains</i>	47
2.1.5. <i>Antibodies</i>	48
2.1.6. <i>Oligonucleotides</i>	49
2.1.7. <i>Software</i>	56
2.2. Cell culture methods	57
2.2.1. <i>Culture of cell lines</i>	57
2.2.2. <i>Culture of CD34+ cord blood MLL/Af4 cells</i>	58
2.2.3. <i>Cell counting and concentration determination</i>	59
2.2.4. <i>Thawing cells</i>	59
2.2.5. <i>Freezing cells</i>	59
2.2.6. <i>Lentivirus production</i>	60
2.2.7. <i>Lentiviral harvesting and concentrating</i>	60
2.2.8. <i>Lentiviral transduction</i>	60
2.2.9. <i>Single cell cloning by limiting dilution</i>	61
2.3. Flow cytometry and cell sorting.....	62
2.3.1. <i>Flow cytometry analysis</i>	62
2.3.2. <i>Cell sorting</i>	63
2.4. Molecular biology methods	66
2.4.1. <i>RNA isolation</i>	66
2.4.2. <i>cDNA synthesis</i>	67

2.4.3.	<i>DNA isolation</i>	67
2.4.4.	<i>Simultaneous DNA and RNA extraction</i>	68
2.4.5.	<i>PCR</i>	69
2.4.6.	<i>Multiplex PCR</i>	71
2.4.7.	<i>Barcoding PCR</i>	72
2.4.8.	<i>Nested PCR</i>	73
2.4.9.	<i>Quantitative PCR (qPCR)</i>	74
2.4.10.	<i>Agarose gel electrophoresis</i>	75
2.4.11.	<i>DNA polyacrylamide gel electrophoresis (DNA-PAGE)</i>	76
2.4.12.	<i>Staining polyacrylamide gel</i>	76
2.4.13.	<i>Whole genome amplification (WGA) single cells</i>	76
2.4.14.	<i>Production of competent bacteria</i>	77
2.4.15.	<i>Site-directed mutagenesis</i>	78
2.4.16.	<i>Hybridisation and phosphorylation pair of oligonucleotides</i>	79
2.4.17.	<i>Ligation</i>	80
2.4.18.	<i>Restriction endonuclease digestion</i>	81
2.4.19.	<i>Fill-in/removal DNA overhangs</i>	81
2.4.20.	<i>DNA dephosphorylation</i>	81
2.4.21.	<i>Gateway recombination</i>	81
2.4.22.	<i>Bacterial transformation of plasmid ligation and mutagenesis</i>	82
2.4.23.	<i>Bacterial transformation supercoiled DNA</i>	82
2.4.24.	<i>Miniprep plasmid isolation</i>	82
2.4.25.	<i>Maxiprep plasmid isolation</i>	83
2.5.	<i>Proteomic methods</i>	85
2.5.1.	<i>Protein isolation and quantification – western blotting</i>	85
2.5.2.	<i>Protein electrophoresis</i>	85
2.5.3.	<i>Immunoblotting</i>	85
2.5.4.	<i>Immunodetection</i>	86
2.5.5.	<i>Co-immunoprecipitation (Co-IP)</i>	86
Chapter 3 Pre-leukaemic populations of L826 and other MLLr leukaemia.....		88
3.1.	<i>Clonal V(D)J rearrangement in ALL and AML L826</i>	89
3.2.	<i>Pre-leukaemic populations L826 presentation and relapse</i>	94
3.3.	<i>Pre-leukaemic populations on the primograft L826</i>	103
3.4.	<i>Pre-leukaemic populations other primografts, t(4;11) and t(9;11)</i>	108
3.5.	<i>Pre-leukaemic population in other MLLr iALL</i>	111
3.5.1.	<i>Clinical characteristics MLLr iALL</i>	111
3.5.2.	<i>Haematopoietic populations in the MLLr iALL</i>	112
3.5.3.	<i>Pre-leukaemic populations on the MLLr iALL</i>	116
3.6.	<i>Discussion</i>	119
Chapter 4 Order of mutations candidate driver genes L826.....		122
4.1.	<i>Mutation acquisition analysis at the clonal level</i>	124
4.2.	<i>Mutational acquisition analysis at the single cell level</i>	127
4.3.	<i>Mutational acquisition analysis based on the haematopoietic hierarchy</i>	130
4.4.	<i>Secondary mutations on the L826 presentation, relapse, and remission</i>	135
4.5.	<i>Discussion</i>	137
Chapter 5 Functional evaluation of candidate driver genes		139
5.1.	<i>Over-expression of mutant candidate driver genes</i>	141
5.1.1.	<i>PHF3 over-expression</i>	142
5.1.2.	<i>CHD4 over-expression</i>	144
5.1.3.	<i>PPP1R7 over-expression</i>	148
5.2.	<i>Knockdown candidate driver genes on SEM cells</i>	151

5.2.1.	<i>Knockdown levels</i>	151
5.2.2.	<i>Immunophenotyping of the knockdown cells</i>	153
5.2.3.	<i>Combination PHF3 and CHD4 knockdown on SEM cells</i>	156
5.3.	Knockdown PHF3 and CHD4 on CD34+ cord blood MLL/Af4 cells.....	158
5.4.	Discussion.....	163
Chapter 6 PHF3 haematopoietic lineage control		165
6.1.	PHF3 variants	165
6.2.	The effect of different PHF3 variants on myeloid marker expression	168
6.3.	Downregulating CD33 on the SEM CD33+ cells.....	168
6.4.	<i>PHF3 knockdown on the AML cell line</i>	170
6.5.	Discussion.....	170
Chapter 7 Concluding discussion and future works		173
Appendix		178
Appendix A PCR MLL/AF4 L826 on different haematopoietic populations		179
A.1 <i>L826 Presentation</i>		179
A.2 <i>L826 Relapse and L826 ALL primograft</i>		180
Appendix B Overlaying DNase hypersensitivity, RNA-seq L826 and CHD4 ChIP-seq		182
B.1 <i>Hox cluster</i>		182
B.2 <i>EBF1</i>		183
B.3 <i>LEF1</i>		184
B.4 <i>PAX5</i>		185
Appendix C Vector maps		187
C.1 <i>LeGO PHF3 N-tag iCer2</i>		187
C.2 <i>pSIEW CHD4 C-tag eGFP</i>		188
C.3 <i>LeGO PPP1R7 N-tag iV2</i>		189
C.4 <i>LeGO ACAP1 N-tag iV2</i>		190
References		191

List of Figures

Figure 1-1 Human haematopoiesis scheme and the markers.....	3
Figure 1-2 Scheme protocol Interfant-06.....	8
Figure 1-3 MLL scheme.....	13
Figure 1-4 Breakpoints MLL/AF4 and AF4/MLL on L826 presentation and relapse.....	25
Figure 1-5 MSI analysis on L826 presentation, relapse, and remission	28
Figure 1-6 Identification of point mutations of non-synonymous mutation genes patient L826 by exome and RNA sequencing.....	30
Figure 1-7 ACAP1 scheme.....	31
Figure 1-8 CHD4 scheme.....	33
Figure 1-9 PHF3 scheme	35
Figure 1-10 PHF3 gene network based on transcriptional changes in the ALL and AML cases.....	37
Figure 1-11 PPP1R7 scheme.....	38
Figure 2-1 Plate design for single cell sorting	65
Figure 3-1 The schematic diagrams of IGH gene complex and the primer locations	90
Figure 3-2 Immunoglobulin rearrangement on presentation and relapse L826	93
Figure 3-3 Haematopoiesis scheme and the markers of the subsets	95
Figure 3-4 Distribution of haematopoietic populations and evaluation of MLL/AF4 on L826 presentation and relapse	99
Figure 3-5 Progenitor leukaemic populations in L826 presentation and relapse	102
Figure 3-6 L826 ALL primograft progenitor population distribution and the fusion gene	105
Figure 3-7 The distribution of haematopoietic subpopulations L826 primograft on the primary, secondary, and quaternary transplants	107
Figure 3-8 MLL/AF4 evaluation on HSC primograft L826 quaternary transplant.....	107
Figure 3-9 The distribution of haematopoietic subpopulations on the secondary transplant t(4;11) patient P929 and primary transplant t(9;11) patient A789 primograft	110
Figure 3-10 Haematopoietic progenitor populations on L876, LK124, LK228, LK230, L880, and LK271	116
Figure 3-11 The pre-leukaemic populations on t(4;11) and t(9;11) iALL.....	117
Figure 4-1 The order of mutation acquisition based on the individual clones derived from L826 relapse	125
Figure 4-2 Multiplex PCR candidate driver genes from amplified DNA single cell L826 relapse.....	128
Figure 4-3 The schematic order of mutations L826 relapse based on the haematopoietic hierarchy	133
Figure 4-4 Colony formation assay on CHD4-depleted SEM cells	134
Figure 5-1 Workflow over-expression of candidate genes analysis	141
Figure 5-2 PHF3 over-expression effects on lymphoid- or myeloid-associated genes on SEM cells.	143
Figure 5-3 CHD4 over-expression evaluation by FACS on eGFP expression.....	145
Figure 5-4 CHD4 over-expression effect on SEM cells transcripts	146
Figure 5-5 Co-IP EZH2 in CHD4 over-expression on 293T cells.....	147
Figure 5-6 PPP1R7 over-expression evaluation by FACS on VFP expression	149
Figure 5-7 PPP1R7 over-expression effect on SEM cells transcripts	150

<i>Figure 5-8 Candidate gene expression level after shRNA knockdown.....</i>	<i>152</i>
<i>Figure 5-9 Flow analysis CD33 on knockdown candidate genes on SEM cells</i>	<i>154</i>
<i>Figure 5-10 qRT-PCR following PHF3 and CHD4 knockdown.....</i>	<i>155</i>
<i>Figure 5-11 PHF3 and CHD4 combination knockdown effect on CD33 expression.....</i>	<i>156</i>
<i>Figure 5-12 Flow analysis CD34+ cord blood MLL/Af4 cells in myeloid culture and lymphoid culture</i>	<i>159</i>
<i>Figure 5-13 Flow analysis of CD34+ cord blood MLL/Af4 cells in myeloid culture control and lymphoid culture with shNTC, shPHF3, shCHD4, and combination shNTC, shPHF3, and shCHD4.....</i>	<i>160</i>
<i>Figure 5-14 Flow analysis CD34+ cord blood MLL/Af4 cells with shNTC and shPHF3, 32 days post-transduction.....</i>	<i>162</i>
<i>Figure 6-1 PHF3 variants</i>	<i>166</i>
<i>Figure 6-2 shRNA constructs target different PHF3 transcript variants</i>	<i>167</i>
<i>Figure 6-3 qRT-PCR knockdown different PHF3 variant on CD33 on SEM cells, day 7.....</i>	<i>168</i>
<i>Figure 6-4 CD33 expression following differential PHF3 variant knockdown in SEM cells</i>	<i>169</i>
<i>Figure 6-5 CD33 expression following knockdown of different PHF3 variants in Kasumi-1 cells.....</i>	<i>170</i>
<i>Figure 7-1 Model of CHD4 and PHF3 cooperation</i>	<i>176</i>

List of Tables

Table 1-1 Treatment results in infant ALL by several different protocols.....	6
Table 1-2 WHO AML classification.....	10
Table 1-3 Distribution of the MLL translocation partner gene and their type of leukaemia incidences.	12
Table 1-4 Summary study cases from different literatures.....	20
Table 1-5 Immunophenotype of patient L826 presenting with ALL and relapse AML	23
Table 1-6 Karyotype and FISH analysis on L826 presentation and relapse.....	24
Table 1-7 Non-synonymous mutated genes on L826 ALL presentation and AML relapse.....	29
Table 2-1 Instruments.....	39
Table 2-2 Experimental kits.....	40
Table 2-3 Cell culture media and supplements.....	41
Table 2-4 Growing media and the supplements.....	42
Table 2-5 Transfection reagents.....	43
Table 2-6 Flow cytometry and cell sorting media.....	43
Table 2-7 DNA electrophoresis buffers.....	43
Table 2-8 Bacterial culture and transformation buffer.....	44
Table 2-9 Western blot and Co-IP buffers.....	46
Table 2-10 Bacterial strains.....	47
Table 2-11 Fluorochrome labelled antibodies for flow cytometry.....	48
Table 2-12 Primary antibodies for western blotting.....	48
Table 2-13 Secondary antibodies for western blotting.....	48
Table 2-14 qRT-PCR primers.....	51
Table 2-15 Primers MLL/AF4 and MLL/AF9 for detecting common breakpoint in transcript.....	51
Table 2-16 Primers mutagenesis candidate genes.....	52
Table 2-17 Primers MiSeq MLL/AF4 and candidate genes for cDNA template.....	52
Table 2-18 Primers MiSeq MLL/AF4 and candidate genes for DNA template.....	53
Table 2-19 shRNA sequences on pLKO5d.SFFV.miR30n backbone.....	55
Table 2-20 Software.....	56
Table 2-21 Fluorescent channels for different fluorochromes on FACSCalibur.....	63
Table 2-22 Antibodies panel for haematopoietic sub-populations sorting on Astrios.....	64
Table 2-23 H ₂ O volume for cDNA dilution.....	67
Table 2-24 PCR components using HotStarTaq DNA Polymerase (QIAGEN).....	69
Table 2-25 Thermal cycle of PCR reaction using HotStarTaq DNA Polymerase (QIAGEN).....	69
Table 2-26 PCR components using Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB) ..	70
Table 2-27 Thermal cycle of PCR reaction using Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB)	70
Table 2-28 Primer components of the common primer MLL/AF4 and MLL/AF9 transcript breakpoints	71
Table 2-29 Multiplex PCR components.....	72
Table 2-30 Thermal cycle of the multiplex PCR.....	72
Table 2-31 PCR components for barcoding PCR using the Access Array Barcode Library for Illumina Sequencers-384, Single Direction (Fluidigm).....	72

<i>Table 2-32 Thermal cycle of barcoding PCR using the Access Array Barcode Library for Illumina Sequencers-384, Single Direction (Fluidigm).....</i>	<i>73</i>
<i>Table 2-33 PCR components second reaction of nested PCR.....</i>	<i>73</i>
<i>Table 2-34 Thermal cycle second PCR reaction of nested PCR.....</i>	<i>74</i>
<i>Table 2-35 qPCR components.....</i>	<i>74</i>
<i>Table 2-36 Thermal cycle of qPCR.....</i>	<i>75</i>
<i>Table 2-37 DNA-PAGE compositions.....</i>	<i>76</i>
<i>Table 2-38 WGA polymerase master mix.....</i>	<i>77</i>
<i>Table 2-39 Site-directed mutagenesis using KOD Hot Start DNA Polymerase (Merck Millipore).....</i>	<i>78</i>
<i>Table 2-40 Thermal cycle of site-directed mutagenesis using KOD Hot Start DNA Polymerase (Merck Millipore).....</i>	<i>79</i>
<i>Table 2-41 Hybridisation and phosphorylation reaction of oligonucleotides.....</i>	<i>79</i>
<i>Table 2-42 Ligation reaction of DNA or gene of interest into the vector.....</i>	<i>80</i>
<i>Table 2-43 Ligation reaction of oligo duplex into the vector.....</i>	<i>80</i>
<i>Table 2-44 Gateway LR clonase reaction.....</i>	<i>82</i>
<i>Table 3-1 Characteristics of patient infant ALL MLLr samples.....</i>	<i>112</i>
<i>Table 4-1 Conservation of mutated residues of candidate genes across different species.....</i>	<i>123</i>
<i>Table 4-2 Examples of read count and mutation percentage sequencing L826 single cells.....</i>	<i>129</i>
<i>Table 4-3 Sequencing reads candidate genes of L826 haematopoietic subsets.....</i>	<i>132</i>
<i>Table 4-4 Secondary mutations of L826 in different haematopoietic subpopulations.....</i>	<i>136</i>
<i>Table 5-1 RNAseq reads L826 on the assessed genes.....</i>	<i>139</i>

Abbreviation

2'-dCF	2'-deoxycoformycin
ACAP1	ArfGAP with coiled-coil, ankyrin repeat and PH domains 1
ADO	allelic drop-out
ALL	acute lymphoblastic leukaemia
AML	acute myeloid leukaemia
ANK	ankyrin
Arf	ADP-ribosylation factor
BAR	Bin/Amphiphysin/Rvs
BCR	B cell receptor
BM	bone marrow
bp	base pair
CAL	congenital acute leukaemia
CCDS	collaborative consensus coding sequence
CD	cluster of differentiation
CFP	cerulean fluorescent protein
CHD4	chromodomain helicase DNA-binding protein 4
ChIP-seq	chromatin immunoprecipitation sequencing
CMP	common myeloid progenitor
COG	children's oncology group
Co-IP	co-immunoprecipitation
CMP	common myeloid progenitor
Da	Dalton
DC	dendritic cells
DDR	DNA-damage response
DHS	DNaseI hypersensitive sites
DSB	double-strand break
EFS	event-free survival
eGFP	enhanced green fluorescent protein
ESC	embryonic stem cell
ETP	earliest thymic progenitor

FAB	French-American-British
FISH	fluorescent in situ hybridisation
GAP	GTPase-activating protein
GBM	glioblastoma multiforme
GMP	granulocyte-macrophage progenitor
HDAC	histone deacetylase
HSC	haematopoietic stem cell
iALL	infant ALL
Ig	immunoglobulin
IgH	immunoglobulin heavy chain
IRES	internal ribosome entry site
JILSG	Japanese infant leukemia study group
LMPP	lymphoid-primed multipotent progenitor
LSD1	lysine-specific histone demethylase 1
MDS	myelodysplastic syndrome
MEP	megakaryocyte-erythrocyte progenitor
MLL	mixed-lineage leukaemia
MLLg	MLL germline
MLLr	MLL rearrangement
MLP	multilymphoid progenitor
MPAL	mixed-phenotype acute leukaemia
MPP	multipotent progenitor
MS	microsatellite
MSI	microsatellite instability
NK	natural killer
nt	nucleotide
NuRD	nucleosome remodelling and histone deacetylase
PAGE	polyacrylamide gel electrophoresis
PB	peripheral blood
PH	pleckstrin homology
PHD	plant homeodomain

PHF3	PHD finger protein 3
Pol	polymerase
PP1	protein phosphatase 1
PPP1R7	protein phosphatase 1 regulatory subunit 7
PPR	poor prednisone response
RNAP II	RNA polymerase II
SFFV	spleen focus-forming virus
SPOC	spen paralog and ortholog C-terminal
SSB	single strand break
TAE	tris-acetate-EDTA
t-AML	therapy-related AML
TBE	tris-borate-EDTA
TCR	T cell receptor
tdTFP	tandem tomato fluorescent protein
TFIIS	transcription factor S-II
TFP	tomato fluorescent protein
T-LBL	T-cell lymphoblastic lymphoma
TLD	TFIIS-like domain
t-MDS	therapy-related myelodysplastic syndrome
t-MN	therapy-related myeloid neoplasms
TRC	the RNAi consortium
VFP	venus fluorescent protein
WBC	white blood cell
WES	whole exome sequencing
WGA	whole genome amplification
WHO	world health organization
WGS	whole genome sequencing
wt	wild-type

Chapter 1 Introduction

1.1. Haematopoiesis

Understanding physiological haematopoiesis is an important prerequisite to studying the control of lineage commitment in leukaemia. It may provide insight into which stage or differentiation step the aberrant leukemic development occurs.

Haematopoiesis is the formation process of all blood cells from haematopoietic stem cells (HSCs) by following controlled differentiation steps. It is highly regulated to maintain the steady-state by producing approximately 10^{12} new blood cells each day in adult humans[1, 2]. This high rate of regeneration is required to provide the continuous circulation and immune function under normal conditions and will be elevated amplification in haematological stress circumstances[3].

The HSC is located at the base of the haematopoiesis hierarchy (Figure 1-1). These cells have the potential to self-renew or reproduce the stem cells and to differentiate into any of the blood cells. When the cells differentiate, the process is accompanied by loss of self-renewal potential, but increasing lineage restriction. The HSC and progenitor cells are characterised by the expression of CD34, and when they differentiate into more mature cells, the CD34 expression is gradually lost[2].

HSCs differentiate into multipotent progenitor (MPP) cells. These cells lose the self-renewal ability, but are capable of differentiating into any haematopoietic cells.

MPPs differentiate into oligopotent progenitor cells, which have undergone a certain degree of lineage commitment [2]. This process is characterised by the expression of CD45RA, an isoform of CD45. CD45 is a surface marker that is expressed on all of the mature haematopoietic cells apart from erythrocytes and plasma cells, but expressed only at low levels on progenitor cells. However, the CD45RA isoform is exclusively presented on the oligopotent progenitor cells[2, 4].

The oligopotent progenitors are divided into lymphoid and myeloid progenitors. The lymphoid lineage originates from multipotent lymphoid progenitors (MLPs). These cells are capable of differentiating into all lymphoid cells, including the earliest thymic progenitors (ETPs) that give rise to T cells, and B_{NK} progenitors that give rise to B cells and natural killer (NK) cells[2]. In addition to this, Doulatov *et al* (2010) described that these cells could also generate some myeloid cells, including

monocytes, macrophages, and dendritic cells (DC), but not erythroid and megakaryocytes[5]. Furthermore, while MLP is defined by the presence of CD34+CD38-CD45RA+CD10+, Goardon *et al* (2011) described a population of CD34+CD38-CD45RA+CD10-, recognised as lymphoid-primed multipotent progenitors (LMPPs) as the population that could differentiate into lymphoid, macrophages, DC, and granulocytes in human[6]. On the other side, the myeloid lineage is represented by common myeloid progenitors (CMPs), granulocyte-macrophage progenitors (GMPs), and megakaryocyte-erythrocyte progenitors (MEPs), with the CMP being the early developmental intermediate of both GMP and MEP[2, 4]. Further, the GMPs give rise to granulocytes, monocytes, and DCs, and the MEPs generate erythrocytes and megakaryocytes/platelets. The granulocytes consist of neutrophils, eosinophils, and basophils.

These different subpopulations of haematopoiesis and their specific markers are summarised in Figure 1-1.

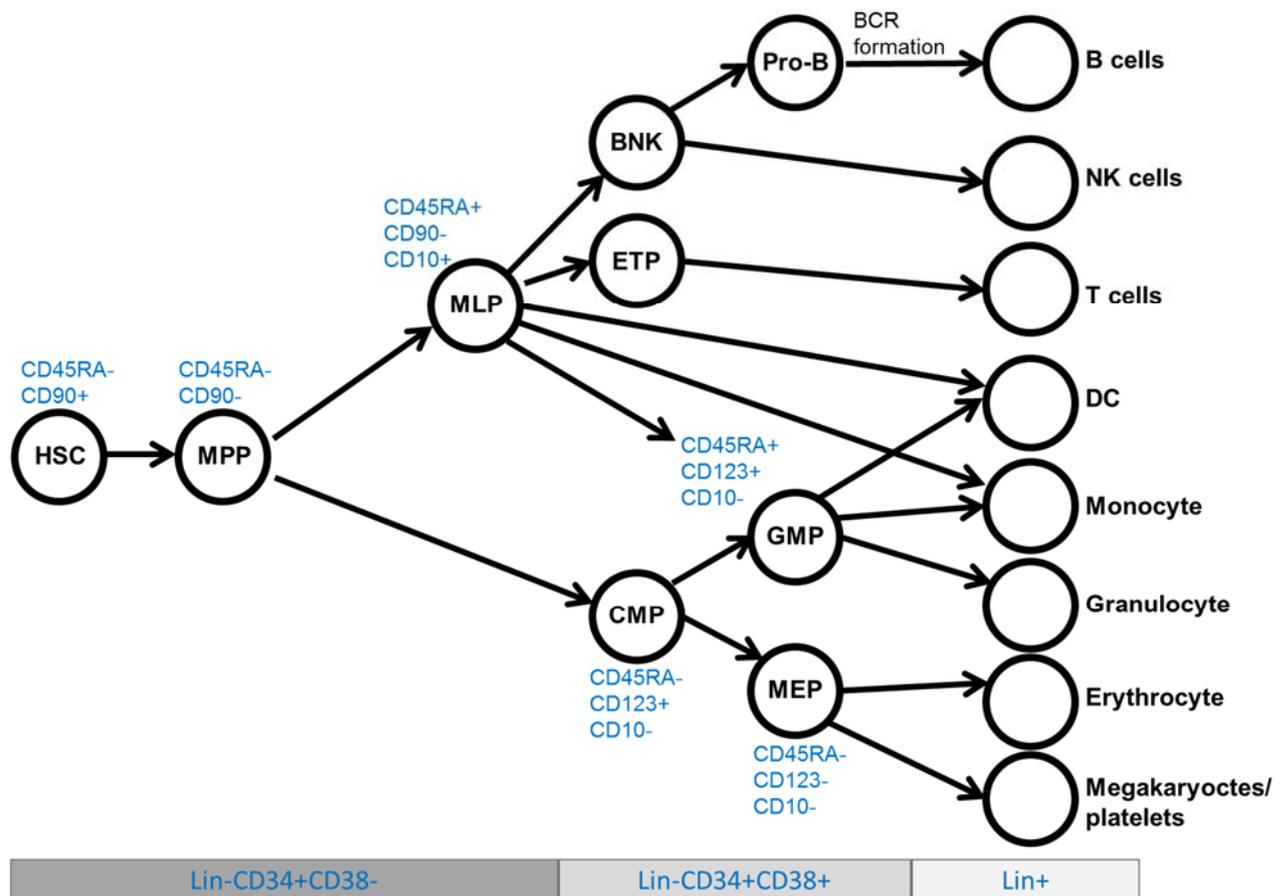


Figure 1-1 Human haematopoiesis scheme and the markers. HSC that is located at the apex of the hierarchy is described as CD34+CD38-CD90+CD45RA-. CD90 expression is lost in MPP. CD38 is gradually gained on the further differentiation states. MLP and GMP are characterised by the expression of CD45RA. MLP is determined to differentiate into lymphoid cells but remains to have the capacity to differentiate into some myeloid cells. Lin is a cocktail of cell surface markers of terminally differentiated cells such as B cell, T cell, DC, monocytes, granulocytes, megakaryocytes, and erythrocytes. B cell receptor (BCR) is formed during the B cell maturation. The figure is adapted from [1] and [2]. HSC = haematopoietic stem cell, MPP = multipotent progenitor, MLP = multilymphoid progenitor, BNK = B/NK cell progenitor, ETP = earliest thymic progenitor, CMP = common myeloid progenitor, GMP = granulocyte-macrophage progenitor, MEP = megakaryocyte-erythrocyte progenitor, NK = natural killer.

Lineage commitment is influenced by the presence of transcription factors. A change in the expression level of the transcription factor may lead to different lineage fate[7]. For instance, GATA1 is an important factor for MEP and GMP differentiation. High expression of GATA1 and FOG1 specify megakaryocyte-erythroid-restricted progenitors. However, moderate expression of GATA1 and C/EBP α or C/EBP β direct

granulocyte/eosinophil development[7, 8]. Furthermore, in addition to preventing the erythroid fate, C/EBP α can also inhibit the lymphoid development[7, 9]. It has been shown that over-expression of C/EBP α initiated myeloid differentiation in pro-B cells[10]. On the other side, EBF1 and PAX5 are known as the main regulators of lymphoid development. Enforced expression of Ebf1 in mouse promotes lineage commitment to B cells and inhibits myeloid development[11].

The understanding of haematopoiesis is expected to help to understand the development of leukaemia, what the gene regulatory networks are and how it affects the lineage fate choice and lineage commitment.

1.2. Acute leukaemia

Leukaemia is a type of blood cancer where the immature blood cells undergo aberrant proliferation, accumulate massively, and infiltrate into the bloodstream. Based on its cellular maturity and rate of clinical change, it can be divided into acute and chronic leukaemia. Acute leukaemias are characterised by the rapid proliferation of immature cells, while chronic leukaemias are characterised by a more gradual accumulation of terminally differentiated progeny[12]. Both groups are further classified based on the cell types into lymphoid and myeloid leukaemia.

The acute lymphoblastic leukaemia (ALL) occurs when the transforming abnormality happened within the lymphoid precursor cells and generates the excessive accumulation of the lymphoblasts. ALL is the dominant type of childhood cancer that accounts for up to 85% of childhood leukaemia cases with the peak age at 0 – 4 years old[13, 14]. Precursor B-cell lineage malignancy forms the majority (~85%) of cases, while the remaining includes mature B-ALL, T-ALL, and mixed phenotype acute leukaemia (MPAL) that co-expresses B- and T-cell markers or lymphoid and myeloid cell markers[15]. CD19 and cytoplasmic CD79a are expressed in all types of B-ALL[16]. The negative expression of CD10 characterise pro-B ALL sub-types and is mostly observed in infants with *MLL* rearrangement (11q23)[13].

Acute myeloid leukaemia (AML) develops within the myeloid precursor cells and can happen within the erythroid, granulocytic, monocytic, or megakaryocytic populations[17]. In contrast to ALL, AML is mostly found in adult patients where more than 40% of patients are >65 years and rarely diagnosed before the age of 40 years old[18]. Myeloperoxidase (MPO) is typically expressed in AML, as well as CD13, CD33, CD65, and CD117[19].

1.2.1. Infant ALL

1.2.1.1. Clinical features

Infant ALL (<1 year of age) is rare, accounting for <5% of childhood ALL. It is classified into two categories: infant ALL with MLL germline (MLLg) and MLL rearrangement (MLLr). The majority of infant ALL cases carry a rearrangement of MLL (80%) and are characterised at diagnosis by higher white blood cells (WBCs) count, and frequent involvement of extramedullary sites, in particular, central nervous system (CNS) and skin[20-26]. It also has a poor outcome with frequency of event-free survival (EFS) 37-49% regardless of the MLL translocation partner, whereas that of EFS MLLg is 60-74%[20-26]. The status of MLL is also used by each of the major cooperative groups, including Interfant, Children's Oncology Group (COG), and Japanese Infant Leukemia Study Group (JILSG) to risk-stratify infant ALL treatment.

Infant ALL MLLg has a higher incidence (~77%) in older infants (>6 month years old)[27]. Cytogenetics are shared with older children, albeit with different distribution, including lower frequency of the favourable abnormalities ETV6-RUNX1 and hyperdiploidy, and higher frequency of unfavourable abnormalities, including BCR-ABL1[27, 28]. Also, MLLg shows less risk of relapse incidences (~20%), in contrast to MLLr (~50%)[27, 28].

Among the different MLL translocation partners in infant ALL MLLr group, t(4;11)(q21;q23) accounts for about 49% of the total cases, followed by t(11;19)(q23;p13), t(9;11)(p22;q23), and t(10;11)(p12;q23) in about 22%, 17%, and 5%, respectively[25, 29]. The leukaemia cells are predominantly CD10- and may co-express one or more myeloid-associated antigens, in contrast to MLLg and childhood ALL that more commonly have a CD10+ immunophenotype[29, 30]. This indicates infant ALL MLLr may come from the immature B-cell precursor cells. Also, studies from Guthrie cards (neonatal blood spots) demonstrate the presence of *MLL* fusion sequences on the specimens, which suggest the *in utero* development or foetal haematopoiesis origin of MLLr ALL[31, 32]. However, the cell of origin of the leukaemia is still a question.

1.2.1.2. Infant specific ALL treatment protocols

Infant ALL was initially treated according to childhood ALL protocols, but was soon recognised to have an unfavourable prognosis, leading this group into high-risk stratification within these protocols[33]. Even though the outcome was sub-optimal, several key features of poor outcome were identified, including the presence of MLL

rearrangement[34], higher WBC[35], CD10-[35], age <6 months at diagnosis[36], and poor initial response to prednisone[36].

Given the rarity of infant ALL, a large international collaborative study group is required to provide a useful insight in determining the treatment protocol for infant leukaemia [37]. The first collaborative studies were named as Interfant-99 that included 482 patients by 17 study groups from 22 countries and resulted in an improved EFS when compared with other larger multicentre studies[37], summarised in Table 1-1.

	Date (year)	CR rate	EFS or survival timepoint	EFS rate (SE)	Survival rate (SE)	Patients enrolled
DFCI (1985–95)	1997	96%	4 year	54% (11)	–	23
Interfant-99	2007	94%	4 year	47% (2.6)	55% (2.7)	482
AIEOP-91/95	2006	96%	5 year	45% (95% CI 31–58)	–	52
BFM	1999	95%	6 year	43% (5)	48% (6)	105
EORTC-CLCG	1994	86%	4 year	43% (95% CI 24–62)	–	25
CCG-1953	2006	97%	5 year	42% (9)	45% (6)	115
CCG-1883	1999	97%	4 year	39% (4)	51% (4)	135
CCG-107	1999	94%	4 year	33% (5)	45% (5)	99
UKALL-92	2002	94%	5 year	33% (95% CI 23–44)	46% (95% CI 35–57)	86
POG 8493	1997	93%	4 year	28% (5)	–	82
POG alternating drugs	1998	94%	4 year	17% (8)	–	33

Table 1-1 Treatment results in infant ALL by several different protocols. DFCI = Dana-Farber Cancer Institute (USA, Canada), AIEOP = Associazione Italiana Ematologia Oncologia Pediatrica (Italy), BFM = Berlin-Frankfurt-Münster (Austria, Germany, Switzerland), EORTC-CLG = European Organisation for Research and Treatment of Cancer – Children’s Leukaemia Cooperative Group (France, Belgium, and Portugal), CCG = Children’s Cancer Group (US), UKALL = Medical Research Council United Kingdom, POG = Paediatric Oncology Group (USA). The table is taken from Pieters *et al*[37].

Interfant-99 data collected several factors that associated with the prognosis of the patients, including the role of sex, the age of diagnosis, WBC at diagnosis, CD10 expression, the status of MLL germline or rearranged, and prednisone response[37, 38]. High WBC, negative expression of CD10, MLL rearranged, and poor prednisone response (PPR) indicates a poor prognosis. These factors were further analysed with statistical Cox regression analysis that produced a new stratification risk group which has been used in the more recent protocol, Interfant-06. The classification includes[38]:

<i>Classification</i>	<i>Conditions</i>
<i>Low-risk (LR)</i>	MLL germline
<i>High-risk (HR)</i>	MLL-r, and diagnosed < 6 months (i.3. <183 days), and WBC $\geq 300 \times 10^9/L$ and/or poor prednisone response
<i>Medium-risk (MR)</i>	all other cases

Also, Interfant-06 includes a transition from “hybrid chemotherapy” of ALL- and AML-oriented drugs in Interfant-99 to an early intensification of AML-oriented drugs (cytarabine, daunorubicin or mitoxantrone, and etoposide)[39]. It was initiated since cytarabine showed a high sensitivity effect to the infant ALL cells, while the combination of high-dose methotrexate (ALL-oriented chemotherapy) and high-dose cytarabine (AML-oriented chemotherapy) in the delayed intensification of Interfant-99 did not benefit patients[25].

The Interfant-06 protocol consisted of two arms, the standard (control) arm that consists of induction (prednisone, dexamethasone, vincristine, cytarabine, daunorubicin, L-asparaginase), protocol IB (6-mercaptopurine, cytarabine, cyclophosphamide), MARMA (6-mercaptopurine, methotrexate, leucovorin rescue, prednisone, cytarabine, PEG-asparaginase), OCTADAD (dexamethasone, 6-thioguanine, vincristine, daunorubicin, PEG-asparaginase, cytarabine, cyclophosphamide), and maintenance (methotrexate, cytarabine, prednisolone, 6-mercaptopurine) and the experimental arm that consists of induction, ADE (cytarabine, daunorubicin, etoposide), MAE (cytarabine, mitoxantrone, etoposide), MARMA, OCTADA, and maintenance. Daunorubicin is included in OCTADAD, but not in OCTADA.

Patients with low risk were eligible for the standard arm, while medium- and high-risk patients were randomised for standard or experimental arm. Furthermore, the HR patients were eligible for stem cell transplantation (SCT), while the MR group were assessed for their MRD level[38, 40]. It is defined that the level of $\geq 10^{-4}$ (more than one leukaemic blast among 10^4 normal cells) was eligible for the SCT for the MR[38, 40]. A summary is depicted in Figure 1-2.

INTERFANT-06

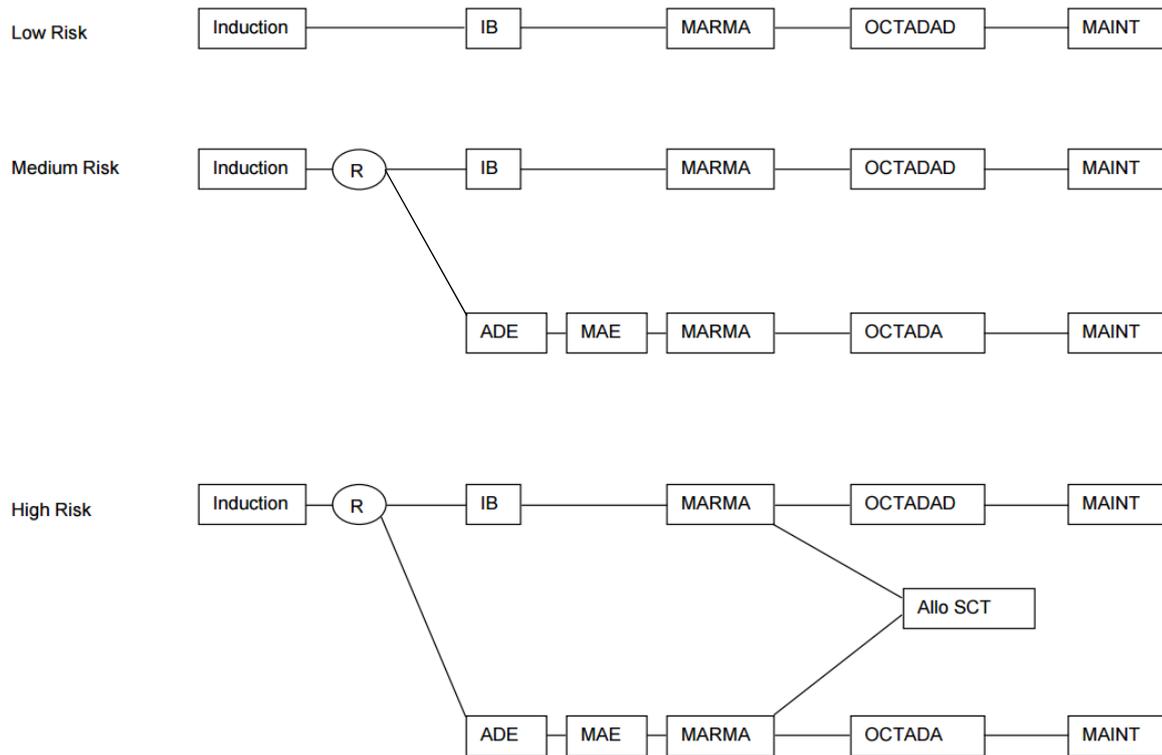


Figure 1-2 Scheme protocol Interfant-06. All LR is treated with standard arm (induction, IB, MARMA, OCTADAD, and maintenance). The MR and HR are randomised for standard or experimental arm (induction, ADE, MAE, MARMA, OCTADA, and maintenance). The HR is also eligible for SCT after MARMA or after receiving part of OCTADA(D). The figure is reproduced from [38].

At the time of this report writing, the protocol Interfant-06 trial has been closed, and the results are being analysed. Current interim guidance in the UK is to use the standard arm therapy.

1.2.2. Paediatric acute myeloid leukaemia (AML)

1.2.2.1. Epidemiology and clinical features

AML accounts for 15-20% of acute leukaemia in children[41]. In contrast to ALL, AML has higher incidence rate by age. Infant AML is a rare disease, accounting for only 0.8% among all AML age groups, rising to 1.5%, 4.3%, 23.1, and 28.7% in age groups 1-4, 5-19, 40-59, and 60-74 years, respectively[42]. It was observed that myelodysplastic syndrome (MDS) may precede AML in adult and elderly patients, but this is rare in paediatric AML[41]. Nevertheless, studies on genetic mutations of AML showed a similar mutation spectrum with different frequencies between adult and childhood[43, 44]. It also leads to different risk stratification groups for paediatric AML.

An extensive cytogenetic study of 729 paediatric AML patients (age 0 to 16 years) on the United Kingdom Medical Research Council trial allowed the evaluation of the chromosomal changes and the associated-risk[45]. This study showed that 75% of the cases had abnormal karyotypes and MLLr is the most frequent abnormality in paediatric AML, accounting for 16% of the cases, where it predominated in the infant cases. While the 11q23 rearrangement, apart from t(9;11), were categorised as high-risk in adult AML, this study observed the abnormalities as an intermediate outcome, including the t(9;11). Translocation t(8;21) and inv(16) had a favourable prognosis, whilst monosomy 7, abnormalities of 5q, and t(6;9) had the adverse outcome[45]. Recent large clinical trials for paediatric AML showed a complete response and overall survival of >90% and >60%, respectively[44, 46-52].

Initial classification by French-American-British (FAB) distinguish the AML into eight subtypes (M0 – M7) based on the morphology of the cells. Later, the better understanding of cytogenetic, immunophenotypic, and the disease biology of the AML develop a new classification by WHO[53]. The latest update in 2016 classifies the AML into AML with recurrent genetic abnormalities (e.g. t(8;21), inv(16), *PML-RARA*), AML with myelodysplasia-related changes, therapy-related myeloid neoplasms, and AML that do not fit to these criteria, recognised as AML not otherwise specified (AML NOS), summarised in Table 1-2[54].

Group	Subgroup
AML with recurrent genetic abnormalities	AML with t(8;21)(q22;q22.1)
	AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);CBFB-MYH11
	APL with PML-RARA (<i>FAB M3</i>)
	AML with t(9;11)(p21.3;q23.3);MLLT3-KMT2A
	AML with t(6;9)(p23;q34.1);DEK-NUP214
	AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2, MECOM
	AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);RBM15-MKL1
	Provisional entity: AML with BCR-ABL1
	AML with mutated NPM1
	AML with biallelic mutations of CEBPA
Provisional entity: AML with mutated RUNX1	
AML with myelodysplasia-related changes	
Therapy-related myeloid neoplasms	
AML, not otherwise specified (NOS)	AML with minimal differentiation (<i>FAB M0</i>)
	AML without maturation (<i>FAB M1</i>)
	AML with maturation (<i>FAB M2</i>)
	Acute myelomonocytic leukemia (<i>FAB M4</i>)
	Acute monoblastic/monocytic leukemia (<i>FAB M5</i>)
	Pure erythroid leukemia (<i>FAB M6</i>)
	Acute megakaryoblastic leukemia (<i>FAB M7</i>)
	Acute basophilic leukemia
	Acute panmyelosis with myelofibrosis

Table 1-2 WHO AML classification. The AML is classified based on the recurrent genetic abnormalities, the incidence of myelodysplasia, therapy-related myeloid neoplasms, and not otherwise specified (NOS). Several of the subgroups are overlapped with the FAB classification, indicated by FAB in italic annotation[54].

1.2.2.2. *Therapy-related AML (t-AML)*

Myeloid neoplasm in patients that had a history of cytotoxic agent exposure is classified as therapy-related myeloid neoplasms (t-MN) by the WHO. This condition consists of two groups, therapy-related myelodysplastic syndrome (t-MDS) and therapy-related AML (t-AML). The categorisation is solely based on the blast count where $\geq 20\%$ count is the t-AML, except for the neoplasms with t(8;21) or inv(16) that are diagnosed as t-AML regardless of the blast count[55]. These neoplasms are thought to be a consequence of mutational events induced by the cytotoxic agents[56]. They were found in various primary malignancies, but more frequently in Hodgkin lymphoma, non-Hodgkin lymphoma, myeloma, ALL, breast cancer, sarcoma, ovarian and testicular cancer[57-60]. A large-scale population study on 3,055 AML patients showed the t-AML accounted for 6.6% of the AML cases[58].

WHO further classified t-AML based on its major causative therapeutic exposures: alkylating agent-related and topoisomerase II inhibitor-related t-AML[61]. Alkylating agents act by transferring alkyl group to specific bases and cause inaccurate base pairing, which then creates single and double-strand breaks during the repair process[62, 63]. It has a latency of 4 to 7 years after the exposure[62]. Also, there is a high incidence of deletion of part or all chromosome 5 and/or 7 (*i.e.* -5/del(5q) and/or -7/del(7q)) in this group[57]. In contrast to the alkylating agent-related t-AML, the topoisomerase II inhibitor agent showed shorter latency within 2 to 3 years after the initial exposure[55]. Most of the cases carried the translocation involving 11q23, and less commonly 21q22 (*AML1*), 16q22 (*CBFB*), and 11p15.5 (*NUP98*)[64-66].

1.2.3. *MLL-rearranged leukaemia*

1.2.3.1. *Lineage specificity of rearranged 11q23*

As briefly mentioned in Section 1.2, the fusion partner of MLL is correlated with a certain lineage of leukaemia, which raises questions on whether mechanisms regulating lineage specificity are determined by 11q23 rearrangements.

A large-scale study of the *MLL* recombinome in 2017 by Meyer *et al.*[67] described in detail the distribution of the *MLL* partner and the associated leukaemia. It included 2,345 acute leukaemia patients consisting of infant, childhood, and adult patients and identified 135 different MLL rearrangements[67]. The study showed there were 35 recurrent translocation partner genes, and 9 of them account for more than 90% of rearrangements within the studied patients[67]. Lineage specificity of acute

leukaemia can be observed from these data. They are re-written and summarised in Table 1-3.

	ALL			AML		
	Infant	Paediatric	Adult	Infant	Paediatric	Adult
AF4	338/352 (96.0%)	139/152 (91.4%)	332/335 (99.1%)	4/352 (1.14%)	3/152 (1.97%)	3/335 (0.90%)
ENL	154/160 (96.3%)	56/78 (71.8%)	50/64 (78.1%)	2/160 (1.25%)	21/78 (26.9%)	14/64 (21.9%)
PTD	N/A	0/6 (0.00%)	1/101 (0.99%)	N/A	6/6 (100%)	98/101 (97.0%)
ELL	0/25 (0.00%)	0/24 (0.00%)	1/48 (2.08%)	24/25 (96.0%)	24/24 (100%)	45/48 (93.8%)
EPS15	16/18 (88.9%)	6/11 (54.5%)	4/9 (44.4%)	1/18 (5.56%)	5/11 (45.5%)	5/9 (55.6%)
AF1Q	1/14 (7.14%)	0/7 (0.00%)	0/2 (0.00%)	13/14 (92.9%)	7/7 (100%)	2/2 (100%)
SEPT6	0/5 (0.00%)	0/10 (0.00%)	0/2 (0.00%)	5/5 (100%)	10/10 (100%)	2/2 (100%)
MLLT6	N/A	1/3 (33.3%)	0/11 (0.00%)	N/A	2/3 (66.7%)	11/11 (100%)
SEPT9	0/2 (0.00%)	0/5 (0.00%)	0/6 (0.00%)	2/2 (100%)	5/5 (100%)	6/6 (100%)

the numbers = case/total case (percentage of the case)

Table 1-3 Distribution of the MLL translocation partner gene and their type of leukaemia

incidences. The overview includes only the recurrent translocation partner genes and the genes that showed distinct specificity of ALL or AML. AF4, ENL in infant, and EPS15 in infant yield ALL as the majority of the cases (blue highlight). On the other hand, PTD, ELL, AF1Q, SEPT6, MLLT6, and SEPT9 yield AML (green highlight). The overview includes ALL and AML but excludes 'others' category that consists of mixed lineage leukaemia, MDS, lymphoma, and no information. Infant = 0 – 1 year, paediatric = 1 – 18 years, adult = >18 years. PTD = partial tandem duplication. The data are adapted from Meyer *et al.*[67].

These data showed that there are at least 9 MLL fusion partners that are specifically associated with certain lineages of MLLr leukaemia, including AF4, ENL and EPS15 in the infant, which are associated with ALL, and PTD, ELL, AF1Q, SEPT6, MLLT6, and SEPT9 in AML.

How *MLL*-altered genes specify the lineage of leukaemia still needs to be determined. A possible mechanism is due to MLL roles that target distinct *HOX* genes, while *HOX* gene family is correlated with embryogenesis and haematopoiesis[68]. MLL consists of three AT-hooks that bind to target DNA, methyl-DNA binding domain (MBD), three PHDs (PHD1-3), one bromodomain followed by another PHD (PHD4), one transactivation domain (TAD), and one SET domain (Figure 1-3)[68].

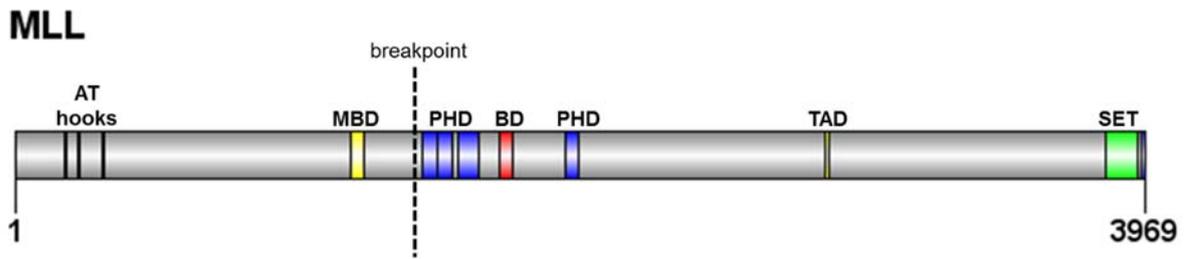


Figure 1-3 MLL scheme. MLL is a large protein (~4000 amino acids) that consists of three AT hooks, followed by MBD, PHD, BD, TAD, and SET domain. The major breakpoint cluster region is indicated on the dotted line. The chromosomal rearrangement will separate the MBD from the PHD domain.

It was observed that the third PHD of MLL interacts with CYP33[68, 69]. Over-expression of CYP33 in human chronic myelogenous leukaemia cells, K562, negatively affects the expression of *HOXC8* and *HOXC9*[69]. By comparing ChIP and gene expression datasets of MLL/AF4-harboured cells, the upregulation of *HOXA7* (+13.1 fold), *HOXA9* (+11.1 fold), and *HOXA10* (+10.1 fold) was identified [70, 71]. Moreover, the SET domain of MLL is known to act as a histone methylase on H3K4, and directly binds to *HOXA9* and *Hoxc8* promoters in human and mouse, respectively[68]. Methylation of H3K4 correlates with an active mark of transcription. Consequently, MLL may relate to the activation of *HOXA9* and *Hoxc8*[68]. Also, the SET domain has been demonstrated to interact with the chromatin remodelling NuRD complex, SWI/SNF, hSNF2H, and Sin3A [72]. In line with these data, expression of MLL/AF4 in human embryonic stem cells showed upregulation of global Hox gene expression[73]. Taken together, *MLL* rearrangement and its partners may deregulate the normal functions of MLL haematopoiesis. The presence of secondary mutations might create an additional dysregulation on the rearranged 11q23 leukaemia cells and lead to another lineage specificity.

1.2.3.2. Models of *MLLr* leukaemia

Several *MLLr* models have been developed to understand the oncogenic mechanisms of the rearrangement. The *MLL/AF9* fusion gene has been demonstrated to induce ALL or AML *in vitro* and *in vivo* in mice[74, 75]. It has been shown that microenvironmental factors impact on lineage commitments of *MLL/AF9* acute leukaemia, *i.e.* lymphoid growth factors induce ALL development, and myeloid factors induce AML development[75]. This principle suggests that introducing important regulators to *MLL/AF9* leukaemia stem cells (LSCs) can direct the leukaemia lineage commitment of the model.

Further to the MLL/AF9 model, Stavropoulou *et al.*[76] successfully expressed and characterised the fusion gene on two different mouse haematopoietic progenitor compartments, long-term HSC and GMP populations. They showed that introducing the fusion gene into these populations was sufficient to develop AML, but with different aggressiveness and outcome. MLL/AF9 long-term HSC developed more dispersed clonogenic growth of AML cells *in vitro*, indicating higher migratory capacity[76]. The fusion gene expression in this compartment was also more potent in inducing AML, by having shorter latency compared with the GMP population *in vivo*[76]. Moreover, they showed that bone marrow transplantation of only 125 MLL/AF9 long-term HSC cells were sufficient to induce AML, but required at least 2,500 transplanted cells in MLL/AF9 GMP cells transplantation[76].

Despite the success of MLL/AF9 model, different attempts to model MLL/AF4 to generate pro-B ALL have not been successful[77-81]. Recently, Lin *et al.*[82] developed a hybrid MLL/Af4 model that was derived from human N-terminal *MLL* and mouse C-terminal *Af4*. The fusion gene was transduced into human CD34+ cells using a retrovirus system, followed by *in vitro* culture or transplantation into immunodeficient mice (NSG mice). They showed that the cells were capable to expand into CD19+CD33- in the lymphoid culture, and the mouse transplanted cells reproducibly induced pro-B ALL phenotype[82]. Moreover, they also performed ChIP-seq and RNAseq and confirmed that the cells had highly similar DNA binding sites and gene expression signature with the human MLL/AF4 ALL.

We received the MLL/Af4 cells from the group (Mulloy lab) to explore further our study.

1.3. Leukaemia lineage switch

1.3.1. Introduction to leukaemia lineage switch

Leukaemia lineage switch is a condition where the patient is diagnosed with a certain lineage of leukaemia (lymphoid or myeloid) but then relapses with leukaemia of the opposite lineage[83, 84]. Although the underlying factors that cause lineage switch remain undetermined, several mechanisms have been suggested. Phenotypic conversion may occur at the multipotent progenitor cells that can differentiate into both lineages[83-89]. Intrinsic (genetic alterations) or extrinsic (chemo- and radiotherapy) factors could trigger a change in the differentiation program of those progenitor cells resulting in a phenotypic switch at relapse[83-89]. Alternatively, lineage switch may occur if there are 2 different *de novo* leukaemia clones. Therapy

at the primary diagnosis kills the dominant clone, followed by outgrowth of the secondary clone at relapse[83-89].

1.3.2. Study cases

A study looking at 239 childhood leukaemia patients with complete remission by Stass *et al.*[84] found that there were 89 relapsed in the bone marrow (BM), and 6 lineage switch from those relapse cases (6.7%). Among these 6 lineage switch cases, 5 cases were a conversion from ALL to AML, and 1 case was the reverse. In addition to its rare evidence, the reports[84, 86-97] also showed that the disease had a very poor prognosis with no standard treatment (most cases were resistant to chemotherapy at relapse). Conversion from ALL to AML forms the majority of cases and predominantly occurred in children. On the other hand conversion from AML to ALL predominately occurred in adults. Furthermore, lineage switch from T cell malignancy, such as T-cell lymphoblastic lymphoma (T-LBL) relapsed to AML was also reported[98, 99]. The conversions mostly arose less than 4 years after primary diagnosis. The reported study cases were collected and summarised in Table 1-4.

Case	Age initial diagnosis	Sex	Time to conversion from diagnosis (month)	Number haematological relapse at conversion	Diagnosis		Karyotype or cytogenetic abnormality		Notes	Ref.
					Pres.	Rel.	Pres.	Rel.		
1	Neonate	F	3	First	L1	M5	46,XX,t(1;6)(p36.2;q25.3), t(4;11)(q21;q23)	46,XX,t(1;6)(p36.2;q25.3), t(4;11)(q21;q23)	Lineage switch from a congenital acute leukaemia (CAL) is a rare case. CAL itself accounts for less than 1% of childhood leukaemia. The immunophenotypes on the presentation showed the lymphoid markers, also expressed dim CD33, but negative CD34.	[94]
2	Neonate	F	2	First	M5	Pro-B ALL	46,XX,ins(11;4)(q23;q21q25)	46,XX,ins(11;4)(q23;q21q25)	Up to 4% B cell precursor immunophenotypes coexisted in the blast presentation examination.	[100]
3	Neonate	M	0 (day 15)	First	Pro-B ALL	M5	MLL/ENL (RT-PCR)	MLL/ENL (RT-PCR)	The patient was treated based on Interfant-06, null response, and switched on day 15. The treatment was converted to the AML protocol and achieved CR.	[100]
4	Neonate	M	6	ND	B ALL	Mono-cytoid	MLL/ENL (RT-PCR)	MLL/ENL (RT-PCR)	The blast in the presentation showed lymphoid immunophenotypes that coexpressed CD34.	[101]
5	Neonate	F	0 (day 14)	First	M5	L2	t(4;11)(q21;q23)	t(4;11)(q21;q23)	At primary diagnosis, immunophenotypes of AML were identified, but negative for myeloperoxidase. The investigators concluded the case as the AML M5 that rarely carried t(4;11). Due to no chemotherapy treatment medicated, they suggested an evidence of leukaemic transformation in utero during foetal haematopoiesis.	[95]
6	21 d	M	± 12	First	Pre-B ALL	MPAL ¹	t(4;11)(q21;q23)	t(4;11)(q21;q23)	The MPAL relapse included monocytoid lineage (predominant) and lymphoid lineage (subdominant). The investigators showed previously undescribed MLL/AF4 sequence that had a breakpoint between MLL exon 12 and AF4 exon 4, found in both the presentation and relapse.	[93]
7	3 mo	F	2	First	L1	M4	t(4;11)(q21;q23)	t(4;11)(q21;q23)	The ALL presentation coexpressed CD33.	[102]

Case	Age initial diagnosis	Sex	Time to conversion from diagnosis (month)	Number haematological relapse at conversion	Diagnosis		Karyotype or cytogenetic abnormality		Notes	Ref.
					Pres.	Rel.	Pres.	Rel.		
8	<6 mo	M	6	First	Pro-B ALL	M5	47,XY,t(4;11)(q21;q23),+8	Complex karyotype including t(4;11)(q21;q23)	The patient was treated based on 1-ALL'96 BFM-HPG on the initial diagnosis, achieved CR	[100]
9	<6 mo	F	7	Second	Pro-B ALL	M5	46,XX,add(19)(p13 or q13)	46,XX,add(19)(p13 or q13)/48,idem,+21,+mar	The patient was treated based on Interfant-99, achieved CR, but relapsed with the same disease after six months latency. Following the relapse treatment, the patient underwent the switch to M5 22 days later.	[100]
10	<6 mo	M	0 (day 8)	First	Pre-B ALL	M5	t(11;19)(q23;p13)	t(11;19)(q23;p13)	The patient was treated based on Interfant-99 and achieved CR. However, undergoing another relapse three months later.	[100]
11	<6 mo	M	0 (day 8)	First	Pro-B ALL	M5	47,XY,+X,t(9;11)(p21;q23)	Unsuccessful karyotype	The patient was treated based on Interfant-99; death during induction. The karyotyping was not complete on the relapse. However, the presentation and relapse showed the same TCR rearrangement sequences.	[100]
12	<6 mo	ND	0 (day 15)	First	Pre-B ALL	M4	MLL/ENL (RT-PCR)	MLL/ENL (RT-PCR)	The patient was treated based on Interfant-99 on the initial diagnosis. After undergoing the switch to AML, the treatment was changed to AML protocol and achieved CR. However, the initial Pre-B ALL reappeared after three months latency.	[100]
13	9 mo	M	± 8	First	Pre-B ALL	M5b	48~50,xy,+2,+8,t(11;16)(q23;p13.3),+r,+mar	47~50,XY,+8,t(11;16)(q23;p13.3),+1~3mar	The t(11;16)(q23;p13.3) gave the product MLL/CBP that was previously described in t-AML.	[89]
14	18 mo	F	0 (day 13)	First	Pro-B ALL	M5	t(4;11)(q21;q23)	t(4;11)(q21;q23)	The patient was treated based on the high-risk group AIEOP-BFM-ALL-2000 on the initial diagnosis but relapsed with AML thirteen days later. The protocol was changed to the conventional AML-type treatment and achieved CR. However, the patient relapsed back with the initial pro-B ALL phenotype eight months afterwards.	[103]
15	3 y	F	± 14	First	L1	M1	46,XX,t(11;14)(p13;q13)/46,XX,del(6)(q?24)	46,XX,del(6)(q?24)	The 6q deletion clone in primary diagnosis was less than 15% of the cells.	[84]

Case	Age initial diagnosis	Sex	Time to conversion from diagnosis (month)	Number haematological relapse at conversion	Diagnosis		Karyotype or cytogenetic abnormality		Notes	Ref.
					Pres.	Rel.	Pres.	Rel.		
16	3 y 10 mo	M	138	Second	L1	M1	ND	t(4;11)(q21;q23)	t(4;11) was found in AML relapse. However, the cytogenetic identification had not been performed at primary state. This case demonstrated a lineage switch after 11 years interval.	[104]
17	4 y	M	19	First	T-ALL	M0	46,XY,del(12)(p12)	45,X,-Y,t(11;17)(q13;p12)/46,XY	The study included array comparative genomic hybridization (Array CGH) and whole exome sequencing (WES) on the presentation and relapse. The Array CGH showed heterozygous deletion accompanied by a small homozygous deletion, both encompassing 9p21.3, two heterozygous deletions at 7q34 and 11p13 on both the presentation and relapse. The WES listed mutated genes in presentation, relapse, also shared by both.	[105]
18	4 y	M	264	First	Pre-B ALL	AML	t(9;22)(q34;q11)	t(9;22)(q34;q11)	The case showed pre-B ALL switched to AML immunophenotypes after 22 years. The identical BCR-ABL1 breakpoint was identified on the presentation and relapse. Also, the relapse shared two identical immunoglobulin rearrangements with the presentation.	[106]
19	4 y	M	9	First	M5	pro-B ALL	Normal	Normal	The authors argued the case as a lineage switch because of the short latency (two months after treatment) compared with at least twelve months as in the secondary ALL.	[91]
20	6 y	M	13	Second	L1	Unclassifiable AML	47,XY,+C	47,XY,+C,13p+,del(17)(p11)/47,XY,+C,i(7q),del(17)(p11) ²	The lineage switch occurred one month after receiving 2'-dCF. The relapse blasts did not express any lymphoid markers but showed a monomorphic proliferation of immature myeloid cells with granules and moderate cytoplasm.	[84]
21	15 y	F	0 (day 8)	First	Common ALL	M5	46,XX,del(6)(q13q21 or q21q23)	46,XX,del(6)(q13q21 or q21q23)	The patient was treated based on ALLIC 02 but changed to AML protocol after the	[100]

Case	Age initial diagnosis	Sex	Time to conversion from diagnosis (month)	Number haematological relapse at conversion	Diagnosis		Karyotype or cytogenetic abnormality		Notes	Ref.
					Pres.	Rel.	Pres.	Rel.		
									lineage switch. She achieved CR, but death due to sepsis in CR.	
22	15 y	M	17	First	L2	M1	Normal	46,XY,del(7)(p14),9p+	The lineage switch occurred one week after receiving 2'-dCF. The investigators suspected the effect of this drug in cell differentiation program.	[84]
23	20 y	M	21	First	T-ALL	AML	52,XY,+?X,+8,+10,+11,+13,+19/46,XY	52,XY,+?X,+8,+10,+11,+13,+19/46,XY	The diagnosis showed the immature T lymphoid immunophenotypes (CD7+, CD4-, CD8-, CD1a-). Retrospect immunophenotype analysis on the non-WBC population (CD45dim) showed 1.2% CD34high population with possible coexpression of CD117, CD33, and CD19, which were expressed in the relapse.	[107]
24	21 y	F	ND	First	Pro-B ALL	MPAL	46,XX,t(4;11)(q21;q23)	46,XX,t(4;11)(q21;q23)/50,XX,t(4;11)(q21;q23),+8,+12,+13,+m/50,XX,t(4;11)(q21;q23),+8,+12,+13,+der(4)t(4;11)(q21;q23)	The MPAL relapse blasts consisted of about 10% lymphoid morphology cells and myeloid/monocytic of the remainings.	[108]
25	25 y	F	31	Second	Pro-B ALL	M5b	TAF15-ZNF384 (RT-PCR)	TAF15-ZNF384 (RT-PCR)	Conventional cytogenetics did not find t(12;17)(p13;q11) in the presentation but re-evaluation with PCR identified the translocation. CD33 and CD34 were expressed in the presentation.	[92]
26	27 y	M	67	Seventh	T-LBL ³	M5a	t(6;11)(q27;q23)	t(6;11)(q27;q23)	T-cell lymphoma that switched to AML is a highly rare case. The translocation in initial diagnosis and relapse had identical breakpoint sequences. Furthermore, the T-LBL coexpressed CD33. However, the AML showed no T-LBL markers.	[98]
27	31 y	M	14	First	T-LBL	M4	t(10;11)(q22;q23)	t(10;11)(q22;q23)	The initial diagnosis showed the clonal TCR gamma chain and a minor IgG kappa gene rearrangements. However, both rearrangements were not found on the relapse blasts.	[99]

Case	Age initial diagnosis	Sex	Time to conversion from diagnosis (month)	Number haematological relapse at conversion	Diagnosis		Karyotype or cytogenetic abnormality		Notes	Ref.
					Pres.	Rel.	Pres.	Rel.		
28	32 y	M	84	Second	T ALL	Myeloid sarcoma	ND	ND	The study declared the case as a lineage switch due to the identical TCR rearrangements. The patient was treated according to GMALL 07/2003 protocol during the T ALL diagnosis.	[109]
29	38 y	M	59	First	M4Eo	B-ALL	inv(16)(p13q22)	inv(16)(p13q22)	Eosinophilia was found in the AML diagnosis; the presentation also showed CD19 and TdT expression	[90]
30	40 y	F	42	First	Pro-B ALL	AML	46,XX,t(2;16)(p11;p11)	47,XX,t(2;16)(p11;p11),+12	Whole exome sequencing was performed on the presentation, relapse, and germline (saliva). The investigators found seven variants, including six variants (ETV6, JARID2, KLF4, PIK3C2A, PTPRG) presented at relapse, and one variant (SMC3) presented at both presentation and relapse.	[110]
31	46 y	M	10	First	M4	L2	Normal	Normal	A lineage switch from AML to ALL in adult was very rare. The authors argued the case was not therapy-related leukaemia because of the short latency (six months) between the treatment and the switch.	[96]
32	62 y	M	4	First	Erythro-leukaemia	Pre-B ALL	41,XY,-5,-7,add(11)(q23),der(14;21)(q10;q10),-16,add(17)(p13),-18,-20,-21,-22,+2mar	41,XY,add(1)(q21),-3,add(3)(q22),-4,-5,-7,der(9)t(9;11)(q34;q13),der(11)del(11)(p11.2)add(11)(q23),der(14;21)(q10;q10),-16,add(17)(p13),-18,-20,-22,+4mar	The first study case of lineage switch from erythroleukemia to ALL. Complex karyotypes on presentation and relapse indicated that the ALL was from the same clone with the AML.	[97]

Table 1-4 Summary study cases from different literatures. The switch from ALL to AML was the majority of the cases. Two T-LBL switches that carried the same translocation between presentation and relapse were also reported. Diagnosis classifications L- and M- refer to standard French–American–British (FAB) classification; ¹MPAL: mixed phenotype acute leukaemia; ²karyotype from the same clone; ³T-LBL: T-cell lymphoblastic lymphoma; ND: not defined; y: year; mo: month.

These collected reports showed rearrangement on 11q23 occurred on a high number of the cases.

MLL rearrangement (t(4;11) or t(9;11)) was found in all of the very rare congenital acute leukaemia (CAL) lineage switch patients and was also found in several childhood lineage switch cases. Interestingly, t(11;16) is uncommon and associated with therapy-related AML (t-AML) and was observed in both presentation and relapse of the lineage switch patient[89].

The two unique cases[98, 99] of T-LBL switched to AML also carried the 11q23 rearrangement, t(6;11)(q27;q23) and t(10;11)(q22;q23), respectively. Both studies showed the identical breakpoint in the presentation and relapse. Moreover, the study from Higuchi *et al.*, (2016) also showed the coexpression of the myeloid marker, CD33, in the initial T-LBL diagnosis, which suggested the involvement of myeloid precursor on the early malignancy[98]. However, it remained unclear if the AML relapse was due to clonal selection or phenotypic changes of the precursor cells.

One unique lineage switch was reported that included AML-M7 (acute megakaryoblastic leukaemia), relapse to ALL-L3, and back to AML-M7 in a 39 years old female patient[88]. Conversion from primary AML to ALL occurred within 1 year, and the second conversion to AML also happened within another year. Short interval neoplasms suggested a double lineage switch instead of secondary malignancies[88]. Complex karyotypes were found in the relapse states. This report showed that an oscillation between the 2 lineages could also occur in leukaemia.

Lastly, two reports suggested the effect of chemotherapy, in particular, adenosine deaminase inhibitor, 2'-deoxycoformycin (2'-dCF), as an extrinsic factor that caused a lineage switch[84, 111]. Even though the exact mechanism is still unknown, but gene activation through DNA methylation as a secondary effect of the drug has been proposed[84].

1.4. Patient L826 and preliminary data

Patient L826 (Newcastle-upon-Tyne Hospitals NHS Foundation Trust) presented with ALL at 9 months old with t(4;11). He was treated with Interfant-06 protocol and underwent complete remission, but relapsed with AML at 4 years old. In collaboration with Claus Meyer, Diagnostic Center of Acute Leukemia, Frankfurt, we determined that both initial presentation and relapse leukaemias shared an identical t(4;11) breakpoint. The presence of this identical breakpoint in both states has an

important meaning, determining that the current case is a lineage switch, instead of secondary neoplasm. Also, this data was supported by microsatellite instability (MSI) analysis which indicated no impairment of DNA mismatch repair in the relapse AML. While the presence of microsatellite instability (MSI) correlates with t-AML[112-114], these data imply that the relapse AML in patient L826 was not a therapy-related case.

1.4.1. Immunophenotypes

Patient L826 was diagnosed in presentation as ALL and relapse as AML. His white blood count in the ALL showed a total of 253×10^9 cells/l with a very high number (90%) of intermediate/small sized blast cells. Total white cell count in the AML was 17.99×10^9 cells/l of which 9.71×10^9 cells/l were blasts of an intermediate/large size. His bone marrow aspirate was clinically immunophenotyped by flow cytometry, as summarised in Table 1-5.

Marker	ALL	AML
CD19	90	9
CD5	2	<1
CD10	<1	<1
CD34	10	87
CD7	9	87
CD33	12	96
CD2	3	2
CD13	<1	2
Surface Ig lambda	<1	<1
Surface Ig kappa	<1	<1
CD14	<1	<1
CD3	3	nd
CD64	3	89
CD45	100	70
Glycophorin A	nd	2
CD15	30	<1
CD117	<1	89
HLA-DR	86	96
CD41a	<1	6
cytoCD3	<1	2
Anti-myeloperoxidase	2	<1
cyCD79a	1	<1
Tdt	3	<1

Table 1-5 Immunophenotype of patient L826 presenting with ALL and relapse AML. The shading on CD19 indicated the ALL marker. The immunophenotype also showed negative expression of surface Ig and CD10 which indicated pro-B ALL. The shading on CD33 and CD117 showed the AML markers. CD64 suggested the monocytoid morphology of the cells. Furthermore, co-expression of CD34 and CD7 indicated that the cells were immature. The anonymised data were obtained from Newcastle upon Tyne Hospitals.

High CD19 expression indicated the B ALL. The negativity of surface immunoglobulin lambda and kappa chains showed the immature phenotype, *i.e.* prior to complete B cell receptor formation. Furthermore, negative expression of CD10 concluded the characteristics of pro-B ALL. In addition to this, negative expression of myeloid markers (myeloperoxidase, CD13, CD33, CD117) and T ALL markers (CD3, CD2, CD5) suggested the disease to be exclusively pro-B ALL, rather than mixed phenotypic leukaemia.

In contrast, AML markers including CD33, CD64, and CD117 were highly expressed in the relapse. Also, the presence of CD64 indicated a tendency of monocytoid

differentiation. Similarly, the T ALL markers (CD3, CD2, CD5, CD10) and B ALL markers (CD79a, CD19, CD10, TdT) were not expressed. Intriguingly, the relapse also expressed CD7. Even though CD7 is a marker of T ALL, but its expression alone is insufficient for the categorisation. Instead, simultaneous expression of CD7 and CD34 in the AML, as in the relapse case, indicated immature blasts[115, 116].

1.4.2. Chromosome study

Chromosome study data were obtained from the Northern Genetics Service. Both presentation and relapse karyotyping showed t(4;11)(q21;q23) and confirmed the rearrangement of MLL by fluorescent in situ hybridisation (FISH). Details of the results were summarised in Table 1-6.

ALL	Karyotype	46, XY, t(4;11)(q21;q23)[3]/46, XY [2]
	FISH	MLL rearranged [89/100]
AML	Karyotype	46, XY, t(4;11)(q21;q23) [1] / 46, idem, del(5)(q2?q3?), add(10)(q) [7] / 47, idem, +der(4) t(4;11) [1] / 46, XY [2]
	FISH	MLL rearranged [92/100], extra 3'MLL signal [8/100]

Table 1-6 Karyotype and FISH analysis on L826 presentation and relapse. Karyotype L826 ALL presentation showed the translocation t(4;11) and the normal karyotype was also seen. The translocation was confirmed by MLL FISH that proved the MLL rearrangement, observed in 89/100 analysed cells; the remaining 11/100 cells are wild-type or non-leukaemic cells that are commonly also present on the analysed samples. The karyotype AML was more complex than the ALL. The translocation t(4;11) was observed, but seven of the examined cells also showed co-mutation deletion chromosome 5 and structural abnormality chromosome 10q. It was also found a cell with an extra copy of the der(4) t(4;11) indicating a subclonal event. FISH on the AML sample confirmed the MLL rearrangement as well as the additional copy of the MLL signal. This cytogenetics analysis was performed by the Northern Genetics Service, Newcastle.

Cytogenetic analysis on the ALL identified three cells with t(4;11)(q21;q23). This result was confirmed by *MLL* (11q23) breakapart FISH that showed MLL rearrangement in 89/100 cells.

The translocation was also found in the AML although several cells had more complex karyotypes. Seven cells showed deletion 5q and a structural abnormality of 10q along with the t(4;11). One cell showed t(4;11) with an extra copy of the der(4)t(4;11) chromosome. This result was confirmed by *MLL* FISH that showed the rearrangement in 92/100 cells also an additional copy 3'-*MLL* signal in 8/100 cells examined, consistent with the single karyotype der(4)t(4;11).

These data indicated there might be a shared clonal origin of the presentation and the relapse.

1.4.3. Fusion gene breakpoint sequences

The presence of t(4;11) in both presentation and relapse indicated that both states might have the same clonal origin. To confirm this, LDI-PCR followed by Sanger sequencing were conducted on the fusion products, *MLL/AF4* and its reciprocal *AF4/MLL*, performed by Dr Claus Meyer. The sequencing showed the same breakpoint as depicted in Figure 1-4.

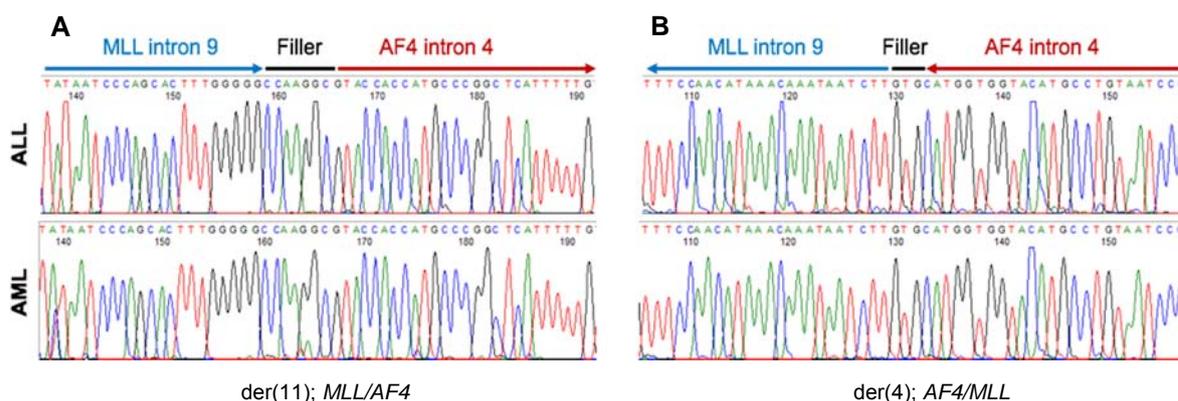


Figure 1-4 Breakpoints *MLL/AF4* and *AF4/MLL* on L826 presentation and relapse. Sanger sequencing genomic breakpoint der(11) (A) and der(4) (B) on L826 presentation (top) and relapse (bottom). Identical breakpoint sequences were seen on der(11) product, *MLL/AF4*, between ALL presentation and AML relapse, as well as on the der(4) product, *AF4/MLL*. The sequencing also showed the recombination *MLL/AF4* between *MLL* intron 9 and *AF4* intron 4, yielded the transcript product *MLL* exon 9 and *AF4* exon 5. Data obtained from Dr Claus Meyer, Diagnostic Center of Acute Leukemia, Frankfurt, Germany.

The identical breakpoints between the presentation and relapse confirmed that the relapse arose from the same origin as the presentation, instead of being a secondary leukaemia. It was further verified by the same random joining (filler) sequences appearing in both states, since it is extremely unlikely for two different leukaemias to share the same filler sequences. This result also showed the fusion gene product contained the truncated *MLL* gene (exon 1-9) and downstream of exon 5 of *AF4* (exon 5-21).

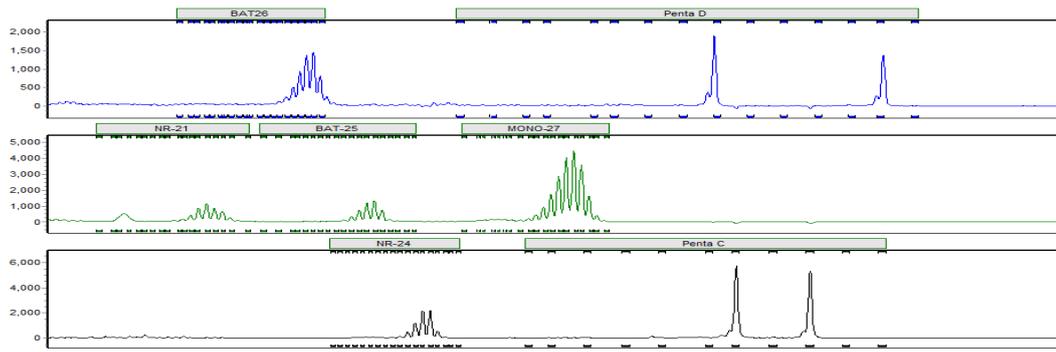
1.4.4. Microsatellite instability analysis

Microsatellites are short, repetitive regions of the DNA found throughout the genome. Due to their repetitive sequences, they are prone to slippage events during DNA replication which results in extra repeats being inserted, defined as microsatellite

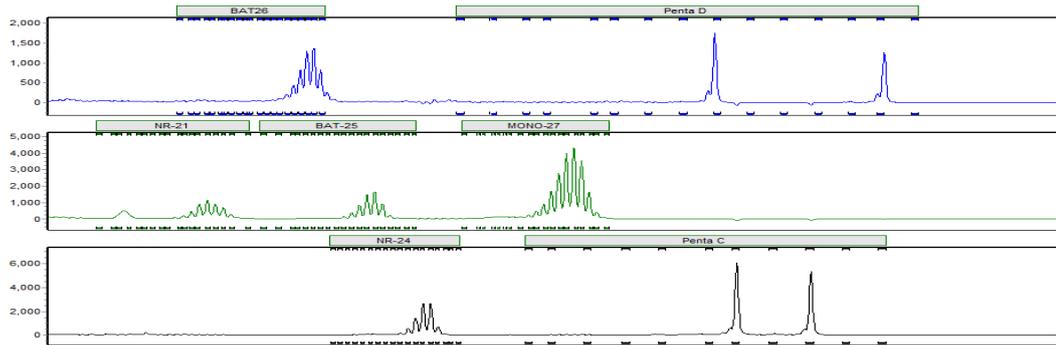
instability (MSI). These errors can be detected and removed by DNA mismatch repair mechanism. However, DNA-damaging chemotherapies (particularly alkylating agents, e.g. cyclophosphamide in Interfant-06), which are given during treatment for the primary malignancy, can cause defects in mismatch repair. Hence, MSI is accepted as a condition of secondary malignancy due to previous chemotherapy.

MSI is frequently found in t-AML and t-MDS, but it is uncommon in primary leukaemia[117]. To further validate that the relapse L826 was not secondary leukaemia due to t-AML, MSI analysis that included five mononucleotide microsatellite markers (BAT25, BAT26, NR-21, NR-24, and MONO-27) as described by Bacher *et al.*[118] was performed on the presentation and relapse, also the remission samples. DNA was extracted from the L826 samples and then assessed by fluorescent multiplex PCR for these microsatellite markers. The products were run on capillary electrophoresis and then analysed using Microsatellite Instability (MSI) Analysis System v1.2 (Promega). According to the current guidelines as defined by Umar *et al.*[119], L826 presentation, relapse, and remission had stable sequences on all of the markers, presented in Figure 1-5.

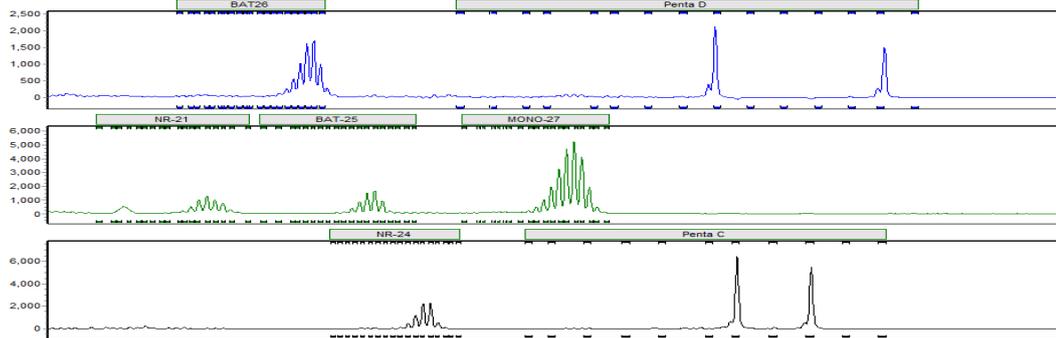
A. ALL Presentation



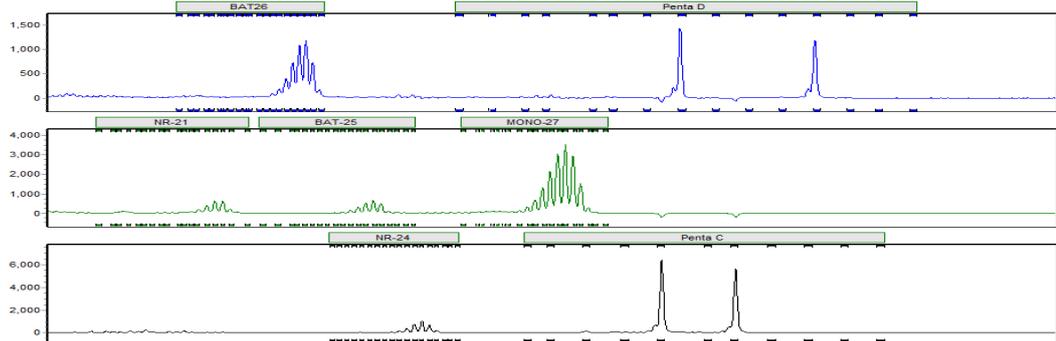
B. AML Relapse



C. Remission



D. Microsatellite stable control



E. Microsatellite unstable control

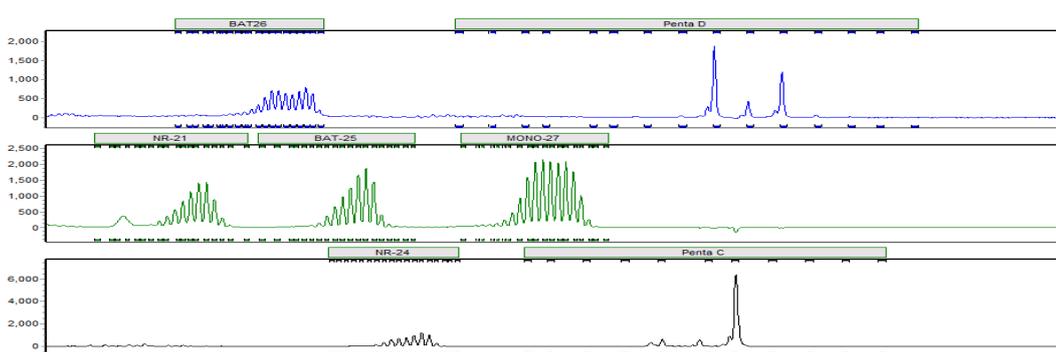


Figure 1-5 MSI analysis on L826 presentation, relapse, and remission. MSI analysis at five mononucleotide microsatellite markers (BAT25, BAT26, NR-21, NR-24, and MONO-27) using fluorescent multiplex PCR was assessed on L826 presentation (A), relapse (B), remission (C), and the analysis controls (D and E). MSI-positive is recognised for having longer or shorter length of microsatellite due to insertion (most of the cases) or deletion, respectively, as seen in the unstable control (E). Two pentanucleotide repeat markers, Penta C and Penta D, were included in the analysis to confirm the samples (presentation, relapse, and remission) were from the same individual, by having the same size of those markers. The analysis was performed using Microsatellite Instability (MSI) Analysis System v1.2 (Promega). Data obtained from Northern Genetics Service.

None of the tested samples showed microsatellite instability. This finding indicated a normal DNA mismatch repair in the relapse L826, suggesting it was not a t-AML.

1.4.5. Whole genome, whole exome, and RNA sequencing

Having validated the occurrence of lineage switch in patient L826, this case allowed us to study the control of lineage commitment. Whole genome sequencing (WGS) and whole exome sequencing (WES) were performed on the presentation, relapse, and remission L826 to identify candidate driver genes. Novel somatic variants including 16 in ALL and 98 in AML samples were found by the sequencing. Among these variants, 1 and 12 of them, respectively, were non-synonymous mutations and predicted to have a deleterious effect on the protein function by Condel scoring (a scoring to evaluate the deleterious probability impact of a variant based on weighted average of the normalized scores method)[120]. Furthermore, RNA sequencing confirmed that 11 of the 12 non-synonymous mutations unique to the AML were amongst the top 25% of expressed genes in both the AML also ALL. These variants are listed in Table 1-7.

ALL			AML		
Gene	Chromosome location	Mutation	Gene	Chromosome location	Mutation
CES1	16q12.2	G19V	ACAP1	17p13.1	R662P
			BACE2	21q22.3	R442P
			CACNB4	2q22-q23	G105R
			CEP164	11q23.3	R953Q
			CHD4	12p13	R1068H
			CHTF8	16q22.1	E27Q
			DHX36	3q25.2	S557G
			NCOA2	8q13.3	K640Q
			PHF3	6q12	K1119I
			PPP1R7	2q37.3	R199L
			SLC4A8	12q13.13	I772F
			ZNF136	19p13.2	L117V

Table 1-7 Non-synonymous mutated genes on L826 ALL presentation and AML relapse. The mutations were based on Exome sequencing comparing presentation, relapse, and remission. It was found there were 1 and 12 mutations in the ALL and AML, respectively. Among the 12 genes, 11 of them (highlighted) were confirmed to be expressed and located in the top 25% expressed genes in both ALL and AML L826. The sequences were aligned by Dr Sirintra Nakjang. The chromosome location and mutation were reviewed and summarised from the aligned sequences using Integrative Genomics Viewer Software[121].

In addition to these 12 variants, the exome and RNA sequencing also showed that these variants had 40 – 60% reads for the mutation. Considering the analysed samples contained 80% blast cells, it indicated heterozygosity of each mutation (Figure 1-6).

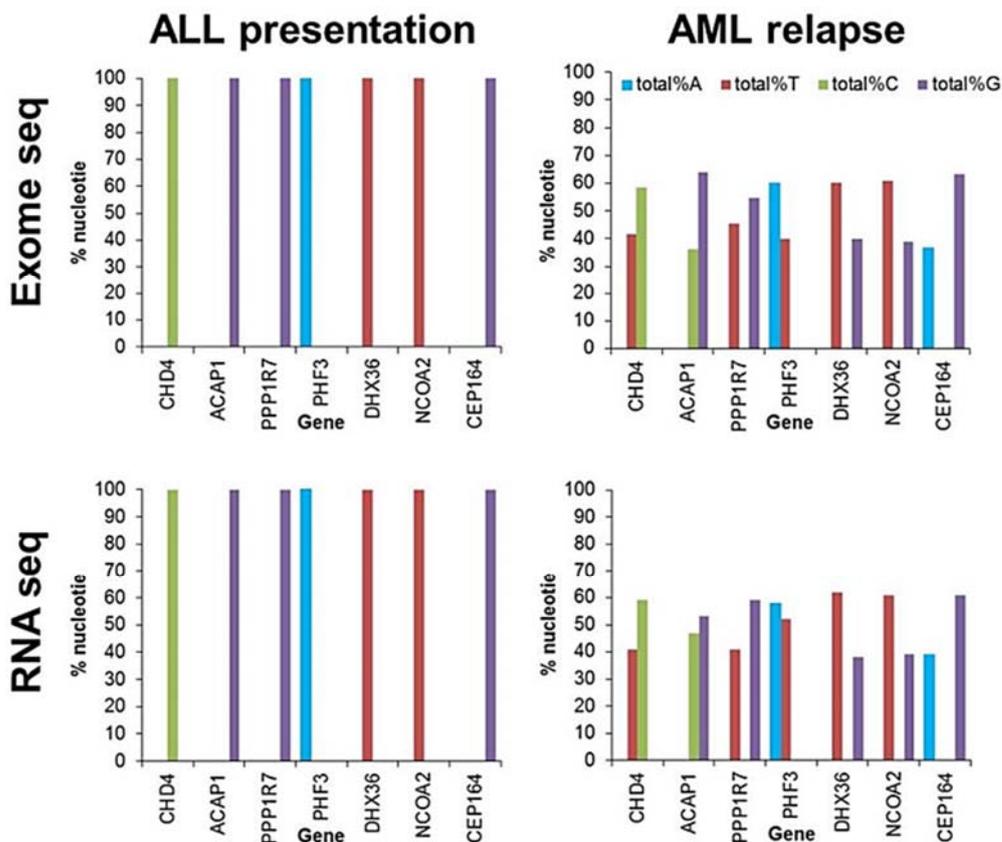


Figure 1-6 Identification of point mutations of non-synonymous mutation genes patient L826 by exome (top panel) and RNA sequencing (bottom panel). The graphs show the percentage of single-nucleotide where the point mutation is located. The left graphs are the reads at the ALL presentation that contain the wild-type genes; while the right graphs are the AML relapse with the mutated genes. The patterns of mutated genes at the AML demonstrate a ratio of 60/40 (reference/mutation), which indicates the mutations are heterozygous. Seven genes are shown as representatives, the other genes show the same pattern.

1.5. Candidate driver genes

Twelve mutated genes in L826 AML relapse are the candidate driver genes that are hypothesised to divert the leukaemia lineage commitment. Eleven of these candidate genes are expressed and located in the top 25% expressed genes in the AML L826 (Table 1-7). These mutated genes include genes that have important roles in cellular and molecular regulatory mechanisms, such as CHD4 (nucleosome remodelling complex, DNA repair[122-127], lineage fidelity[128-132]) ACAP1 (intracellular trafficking[133]), PPP1R7 (control of mitosis[134, 135]), PHF3 (potential role in transcription elongation[136], DNA repair[137]), DHX36 (RNA helicase – transcription regulation[138]), NCOA2 (transcriptional coactivator[139]), and CEP164 (microtubule organisation and chromosome segregation[140], DNA damage response[141, 142]).

These candidate genes were screened for the order of mutation acquisition, as described in Chapter 4. Based on that result, ACAP1, CHD4, PHF3, and PPP1R7 were studied more intensively in this thesis.

1.5.1. ACAP1

ArfGAP with coiled-coil, ankyrin repeat and PH domains 1 (ACAP1) is a member of Arf-GTPase-activating protein (GAP) family protein, a family that regulates ADP-ribosylation factors (Arfs)[133]. Arfs are a family of GTP-binding proteins that function as molecular switches in a variety of cellular events[133]. One member of Arf family is Arf6 that is known to have roles in GLUT4 trafficking[143], phagocytosis in macrophages[144], also modelling the plasma membrane and underlying cytoskeleton[145]. A study by Jackson *et al*[133] demonstrated Arf6 as the substrate of ACAP1, in which ACAP1 acted as a negative regulator of its activity.

ACAP1 consists of a Bin/Amphiphysin/Rvs (BAR) domain, a Pleckstrin homology (PH) domain, the Arf-GAP domain, and ankyrin (ANK)-repeats (Figure 1-7). Mutation L826 relapse R662P was found on an ANK repeat. A crystal structure of PAP, another member of Arf-GAP family, showed that the ANK repeats physically associate with the Arf GAP domain[133, 146]. Since ANK repeats and Arf-GAP domain are the conserved domains of AZAP subgroup of Arf-GAP family (where the ACAP1 and PAP are categorised in[147]), it may also imply that the ANK repeat and Arf-GAP interaction is a common mechanism in this subgroup[133].

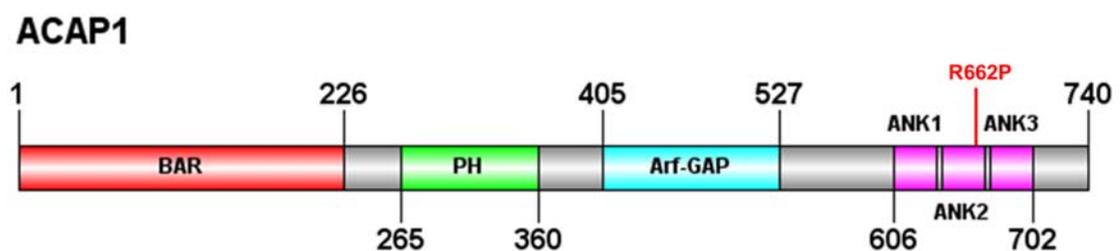


Figure 1-7 ACAP1 scheme. ACAP1 consists of BAR, PH, Arf-GAP domains, and ANK repeats. L826 relapse mutation occurred at R662P of ACAP1. The scheme is designed using IBS illustrator[148].

In addition to the known protein domains, an evaluation of ACAP1 protein expression showed that ACAP1 was widely expressed in different cells, including HeLa, CHO, Jurkat, lymphoblastoid C1R and JY cells, Daudi, B cells, HEK 293, NIH 3T3, but not in monocytes (myeloid cells)[133]. This expression profile may raise questions on the

effect of cellular organisation by ACAP1 mutation on the L826 relapse (*i.e.* myeloid leukaemia).

1.5.2. CHD4

Chromodomain helicase DNA-binding protein 4 (CHD4) was initially found in dermatomyositis-specific autoantigen[149, 150] and confirmed in 10-30% of the disease cases[151-153]. Interestingly, dermatomyositis is related with malignancies, particularly in ovary, lung, pancreas, stomach, colorectal, non-Hodgkin lymphomas and also ALL and AML[154-160].

CHD4 is recognised as a major subunit of nucleosome remodelling and histone deacetylase (NuRD) complex that also recruits HDAC1 and HDAC2, MBD2 and MBD3, MTA1, MTA2, and MTA3, RBBP4 and RBBP7, and it is also observed that it is involved in recruiting LSD1[155, 161]. CHD4 and other components of the NuRD complex work generally as transcriptional repressors through the chromatin remodelling ATPase activity of CHD4, histone deacetylation (HDAC), histone demethylation (LSD1), and DNA/promoter methylation[129, 155, 161, 162].

Looking in detail at its structure, CHD4 consists of two plant homeodomains (PHDs), two chromodomains, and an ATPase/helicase domain (helicase ATP-binding and helicase C-terminal; Figure 1-8) that apparently act in concert with each other or in combination with other NuRD components[129, 163, 164]. The PHDs interact with HDAC of the NuRD complex and also the histone tails, preferentially unmodified H3K4 and H3K9me[129, 163, 165]. The chromodomains have DNA binding activity and are required for nucleosome binding and nucleosome mobilisation through ATPase activity[165]. The ATPase/helicase domain is the part that promotes nucleosome mobilisation to maintain compact chromatin[129, 166]. Both PHDs and chromodomains work interdependently from the ATPase activity to regulate chromosome remodelling, shown by molecular shape reconstruction from small-angle X-ray scattering that revealed the extensive domain-domain interactions[164]. We found that the mutation of L826 was located in the ATPase/helicase domain (R1068H). Moreover, exome sequencing analysis on 52 serous, 23 clear-cell, 67 endometrioid, and 18 mixed-histology endometrial tumours, identified 17%, 4%, 7%, and 11% mutations in CHD4, respectively[167]. Intriguingly, 50% of the mutations were also located in the ATPase/helicase domain[167], which may indicate the importance of this domain. Furthermore, the mutation R1068H was also found in

congenital heart disease[168] and developmental disorder[169] cases, which indicate the probable change of function due to the mutation.

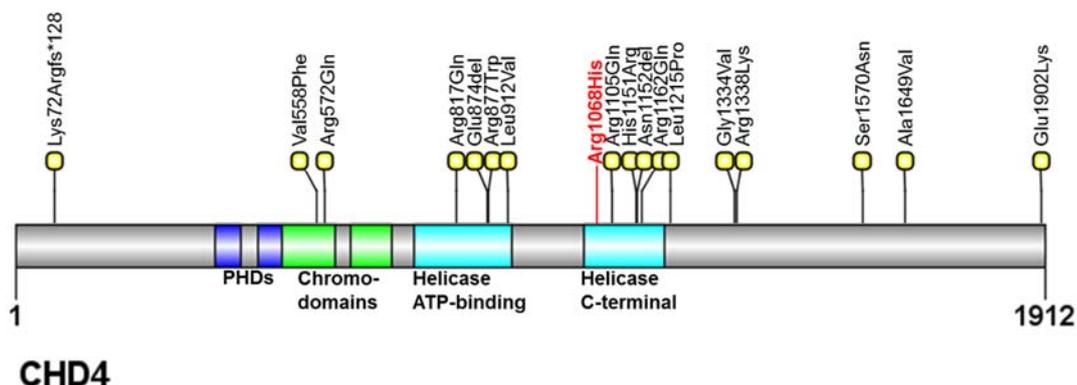


Figure 1-8 CHD4 scheme. CHD4 consists of two PHDs, two chromodomains, and ATPase/helicase domain that divided into helicase ATP-binding and helicase C-terminal. The mutation on L826 is located at Arg1068His. Other mutations found by other group[167] are denoted in the yellow circles.

The roles of CHD4 have been identified as (1) transcription regulation, such as in haematopoiesis (CD79a, Ikzf1, ROR γ)[128-132], cell maturation (Wnt and Tcf7, GATA3, γ -globin)[170, 171], and embryonic stem cells (ESCs) self-renewal (Tbx3, Klf4, Klf5)[172], also (2) genome maintenance (OGG1, p53, BRCA1, RPA2, RNF8, CDC25A)[122-127].

In lymphopoiesis, CHD4 was shown to interact with ROR γ and repress its activity[128]. ROR γ is a nuclear receptor that binds gene promoters as monomers to activate the target genes[128]. It is recognised to have important roles in thymocyte development and T-cell lymphoma formation[128]. It is also revealed that CHD4 expression is required for transition from double negative (CD4-CD8-) to double positive (CD4+ CD8+) stage and normal expression of CD4 during T cell development[131].

Studies in B cell lineage differentiation demonstrate that CHD4 is associated with CD79a, a gene that encodes Ig α (the trans-membrane component of the B cell receptor)[129, 130]. CHD4 acts to enhance hypermethylation of the CD79a promoter, which represses its transcription[129]. Knockdown of CHD4 substantially enhanced the chromatin accessibility of the promoter and increased the transcription of CD79a[129, 130]. Furthermore, a study on a patient with CD79a mutation shows a block from pro-B to pre-B transition[173]. As a result, NuRD is referred to as the gatekeeper for B-cell development.

In vivo study of Chd4 deletion in mouse BM revealed early loss of lymphoid and myeloid cells and compensated with erythroid cells accumulation[132]. The study confirmed that Chd4-depleted BM progressed only to erythroid commitment which arrests at the proerythroblast stage while restricting myeloid cells and their downstream differentiation pathways[132]. Furthermore, they also found that Chd4 expression was required to prime myeloid differentiation program on HSCs.

CHD4 is also known to influence lymphopoiesis by its interaction with Ikzf1[174]. Ikzf1 is required for the differentiation of multipotent cells into lymphoid cells, and also restricts the myeloid programming[175]. The study by Reynaud *et al*[176] showed that B cell development was completely blocked in the absence of Ikzf1. Taken together, we hypothesise that the mutation on CHD4 L826 might disrupt its interaction with IKZF1, thus affect the lineage fate.

In addition to the transcription regulation, the function of CHD4 is also widely studied in DNA-damage response (DDR) and cell cycle[122-127]. Polo *et al* (2010)[125] and Larsen *et al* (2010)[122] demonstrated rapid recruitment of CHD4 to the damage site upon DNA double-strand break (DSB) – inducing treatment. A further study identified that the recruitment promoted efficient ubiquitin conjugation by RNF8, which then mediated chromatin unfolding and recruitment of the other repair components, RNF168 and BRCA1[123].

As well as in DSB, the role of CHD4 was also identified during oxidative damage and single strand breaks (SSBs)[127]. By treating cells with H₂O₂ to induce oxidative damage, OGG1 (a key DNA repair enzyme involved in oxidative damage) strongly binds with CHD4 at the damage sites[127]. After that, CHD4 recruits repressive chromatin proteins, including DNMT1, DNMT3A, DNMT3B, EZH2, and EHMT2 to those sites[127]. Moreover, this study also demonstrated that the recruitment of EZH2, DNMTs, and EHMT2 by CHD4 was dependent its ATPase domain.

Lastly, a recent study by Sperlazza *et al*[177] investigated the importance of CHD4 on AML cells. They demonstrated the necessity of CHD4 for efficient DSB repair, where depletion of CHD4 yields a more susceptible effect of DNA-damaging agents, daunorubicin and cytarabine, both *in vitro* and *in vivo*. Furthermore, this study showed the knockdown of CHD4 reduced the capacity of AML cells to form colonies in soft agar. Conclusively, CHD4 is one of the main candidate genes in this study due to its known roles.

1.5.3. PHF3

PHD finger protein 3 (PHF3) was first identified as a tumour antigen (*i.e.* immunogenic) in 24 of 39 (61.53%) glioblastoma multiforme (GBM) patients, but in none of 14 healthy patients[136]. Further study by the same group in a higher number of patients found that there were 35 antibody responses against PHF3 from 62 glioblastoma patients (56.5%)[178]. In both studies, the presence of PHF3 autoantibodies was associated with a significantly better survival rate. The latter showed a median survival time of 14.5 months compared with 7.2 months for the patients with and without the autoantibody, respectively[178]. Moreover, they identified the wide expression of PHF3 in different tissues, including brain, heart, placenta, lung, liver, skeletal muscle, kidney, pancreas, and stomach, but the expression was significantly reduced in 15/21 glioblastoma, 2/2 glioblastoma cell lines, 2/2 anaplastic astrocytoma, and 2/2 astrocytoma WHO grade II[179].

In addition to the glioma cases, Sroczyńska *et al* (2014) showed the significance of PHF3 in human AML harbouring MLL/AF9 in the mouse model[180]. They performed epigenetic shRNA library targeting 319 known and candidate epigenetic regulators to screen for novel potential drug targets. They showed Phf3 was located at a high rank (4th rank) in the library indicating its role in cell survival and proliferation[180].

PHF3 is predicted to have a PHD finger domain, TFIIS homology domain, Spen paralog and ortholog C-terminal (SPOC) domain, and a proline-rich region (Figure 1-9). This protein consists of 2039 amino acids and has a total molecular weight of 229.45 kDa[136]. It is suggested to have a role as a transcription elongation factor due to homology in domains TFIIS[136]. In our study, exome sequencing data shows a point mutation from lysine to isoleucine on the 1119th amino acid of PHF3.

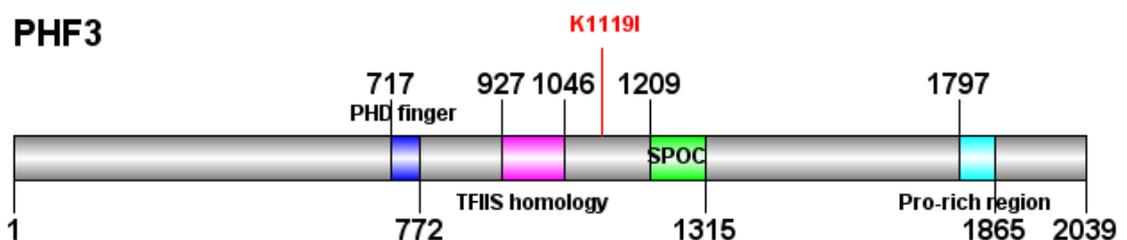


Figure 1-9 PHF3 scheme. PHF3 consists of PHD finger, TFIIS homology, SPOC, and proline-rich region. The mutation on L826 is located at K1119I.

A starting point of understanding PHF3 function is from its homologue in *Saccharomyces cerevisiae*, Bye1 that has the same domain organisation[181]. This

study demonstrated that Bye1 bound directly to the core of RNA polymerase II (RNAP II) through its TFIIIS-like domain (TLD)[181]. The investigators also analysed that the surface forming an interface between RNAP II and Bye1 TLD are well conserved in human RNAP II and PHF3, respectively, which suggests the same domain function of Bye1 and PHF3[181].

This hypothesis is supported by a study by Boeing *et al*[137] that examined the factors that responded to UV-induced DNA damage. DNA lesions lead to RNAP II stalling, resulting in a block of transcription elongation, which then is repaired by transcription-coupled nucleotide excision repair (NER)[137]. CSB is a protein that is recruited to the damage-stalled RNAP II and plays a key role in the NER[137]. This study included two interactome analyses: CSB and RNAP II interactomes upon UV irradiation, and showed PHF3 was recruited in both events. This investigation not only demonstrates the interactions of PHF3 but also indicates the role of PHF3 in DNA damage response[137].

Furthermore, PHF3 function is also studied from its mouse paralogue protein, Dido3, which has the same domain structure. Gatchalian *et al*[182] demonstrated that while the PHD domain of Dido3 could bind to H3K4me3, this binding was not found on PHF3[182].

In assistance with Dr Dan Williamson, we analysed the centrality of our mutated genes to variation in expression profiles of ALL and AML cases using an ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks) analysis[183, 184]. This algorithm allows the gene expression comparisons between one group (*i.e.* samples in ALL group) and the other group (*i.e.* samples in AML group). Genes that show consistent high expression pattern in certain group but not in the other group would be given higher score. By using Boolean logic, these consistent genes are collected and reconstructed into a network.

The expression data of ALL and AML cases were obtained from public genomics data repository GEO (Gene Expression Omnibus; NCBI), with accession number GSE11877 (207 samples) and GSE17855 (237 samples), respectively. The calculation was displayed in Cytoscape program[185]. When our 12 candidate mutations were overlaid on the analysis, PHF3 was found to have the highest number of degrees within the networks generated (Figure 1-10), implying a stronger

impact of this mutation on the switch from ALL to AML than our other 11 candidate mutations.

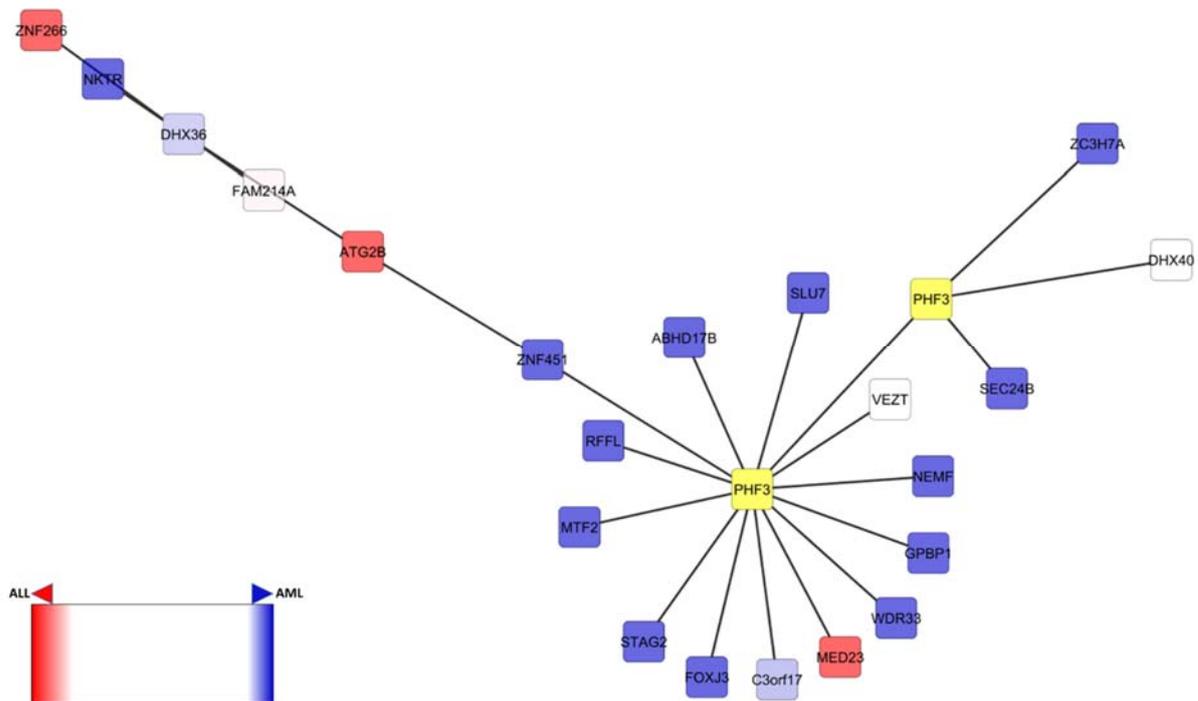


Figure 1-10 PHF3 gene network based on transcriptional changes in the ALL and AML cases.

The branch shows the relation among the genes referring to the transcriptional changes, and not to the protein functional interactions. The yellow node indicates the central node, in this case, is PHF3. Higher red intensity node indicates the gene with higher transcript expression in ALL cases. Higher blue intensity node indicates the gene with higher transcript expression in AML cases. This network indicates that the change in PHF3 expression is accompanied by those number of highly differentially expressed ALL-AML genes, which was much more than the other 11 other candidate genes.

Lastly, a translocation $t(6;15)(q12;q15)$, in which PHF3 is involved in one of these chromosomes, was identified in t-MDS[186]. This report hypothesised that one of the genes that were affected was PHF3. Taken together, these studies lead us to focus on PHF3 as the potential candidate driver gene.

1.5.4. PPP1R7

Protein phosphatase 1 regulatory subunit 7 (PPP1R7; 2q37.3, also called SDS22) encodes a protein that is part of the regulatory subunit of serine/threonine protein phosphatase 1 (PP1), consists of 360 residues and has the molecular weight of 41.6 kDa[187]. This protein is highly conserved and is also found in *Saccharomyces cerevisiae*, suggesting an important role [188]. It is composed almost entirely of 11 leucine-rich repeats (LRR) motifs (Figure 1-11). This motif consists of 22 residues with consensus sequences ExLxxLxxLxxLxLxxNxIxIxI (x is any residue) which

provides the essential binding region with the PP1[187]. L826 sequencing showed a mutation on a conserved residue, R199L that is located at the sixth LRR. This evidence might indicate an important effect of the mutation on PPP1R7 function.

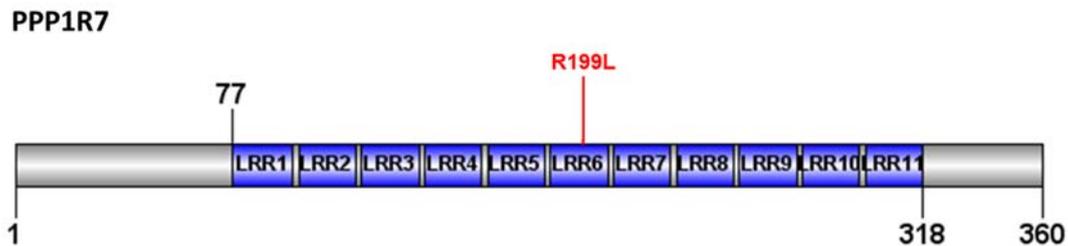


Figure 1-11 PPP1R7 scheme. PPP1R7 mainly consists of LRR repeats that allow the binding to PP1. The mutation R199L in L826 is found in the sixth LRR domain.

The interaction between PPP1R7 and PP1 is known to modulate the activity of Aurora-B kinase. While the Aurora-B is required for microtubule-kinetochore attachment, PPP1R7-PP1 reverses the phosphorylation of the kinase during anaphase, which then promotes the completion of mitosis[134, 135]. By depleting PPP1R7, it was shown a reduction of PP1 localisation at the kinetochore and increased the phosphorylated Aurora-B kinase[134].

In addition to counteracting Aurora-B phosphorylation, PPP1R7 is also known to interact with hADA3, a subunit of HAT complexes, which, interestingly, is involved in p53 acetylation[189, 190]. These observations suggest a role of PPP1R7 in cell cycle control.

Furthermore, a study on *Drosophila* showed *sds22* had a role as a tumour suppressor gene[191]. This study demonstrated the loss of *sds22* on activated Ras (Ras^{V12}) promoted tumour growth and invasive behaviour. Conversely, over-expression of *sds22* on Ras^{V12}*scrib*^{-/-} mutant cells suppressed the tumour growth, suggested by inhibiting myosin II and Jun N-terminal kinase (JNK) signalling[191].

Taken together, PPP1R7 may have an important role in cell growth and tumorigenesis.

Chapter 2 Materials and Methods

2.1. Materials

2.1.1. Laboratory equipment

<i>Instrument</i>	<i>Manufacturer</i>
<i>Centrifuges</i> Avanti J-26 XP centrifuge Centrifuge 5424 Centrifuge 5424R Heraus Multifuge 3SR+ Mistral 1000 Optima L-100 XP ultracentrifuge	Beckman Coulter Eppendorf Eppendorf Thermo Fisher Scientific MSE Beckman Coulter
<i>DNA gel imaging</i> GelDoc G:BOX	Bio-Rad Syngene
<i>Flow cytometer</i> Astrios EQ Cell Sorter FACSAria III FACSAria Fusion FACSCalibur FACSCanto II	Beckman Coulter Beckton Dickinson Beckton Dickinson Beckton Dickinson Beckton Dickinson
<i>Spectrophotometer</i> FLUOstar Omega Nanodrop 1000 spectrophotometer SmartSpec Plus Qubit 3.0 Fluorometer	BMG labtech Thermo Fisher Scientific Bio-Rad Thermo Fisher Scientific
<i>Thermal cycler</i> ABI 7900HT GeneAmp PCR System 2700 ViiA 7 Real-Time PCR System	Applied Biosystems Applied Biosystems Applied Biosystems
<i>Immunoblotting</i> Mini-PROTEAN Tetra Cell 2 X-Ray Film Processor JP-33	Bio-Rad JPI

Table 2-1 Instruments

2.1.2. Chemicals and reagents

2.1.2.1. General chemicals and reagents

All general chemicals and reagents were purchased from Sigma Aldrich, unless otherwise stated.

2.1.2.2. Experimental kits

Name	Manufacturer
AllPrep DNA/RNA Micro Kit	QIAGEN
AllPrep DNA/RNA Mini Kit	QIAGEN
BCA Protein Assay Kit	Santa Cruz Biotech
DNeasy Blood and Tissue Kit	QIAGEN
Endofree Plasmid Maxi Kit	QIAGEN
Gateway LR Clonase II Enzyme Kit	Invitrogen
HotStarTaq® DNA Polymerase Kit	QIAGEN
KOD Hot Start DNA Polymerase	Merck Millipore
Phusion® High-Fidelity PCR Master Mix with HF Buffer	NEB
Platinum® SYBR® Green SuperMix UDG	Applied Biosystems
QIAprep Miniprep Kit	QIAGEN
QIAquick Gel Extraction Kit	QIAGEN
QIAquick PCR Purification Kit	QIAGEN
REPLI-g Single Cell Kit	QIAGEN
RevertAid™ H Minus cDNA Synthesis Kit	Thermo Fisher Scientific
RNeasy Mini Kit	QIAGEN
RNeasy Plus Micro Kit	QIAGEN

Table 2-2 Experimental kits

2.1.2.3. *Cell culture media and supplements*

Name	Manufacturer
α -MEM, no nucleoside (22561021)	Thermo Fisher Scientific
DMEM (D5671)	Sigma Aldrich
GIBCO™ Foetal bovine serum	Thermo Fisher Scientific
IMDM (I3390)	Sigma Aldrich
L-Glutamine (G7513)	Sigma Aldrich
Polybrene (H9268)	Sigma Aldrich
Recombinant human FLT3LG (308-FK)	R&D Systems
Recombinant human IL-3 (203-IL)	R&D Systems
Recombinant human IL-6 (130-095-352)	MACS Miltenyi Biotec
Recombinant human IL-7 (207-IL)	R&D Systems
Recombinant human SCF (14-8449-62)	Affymetrix eBioscience
Recombinant human TPO (288-TP)	R&D Systems
RPMI 1640 (R8758)	Sigma Aldrich
Sodium pyruvate solution (S8636)	Sigma Aldrich
Trypsin-EDTA solution (T4174)	Sigma Aldrich

Table 2-3 Cell culture media and supplements

2.1.3. Buffers and media

2.1.3.1. Cell culture

Cells	Medium	Supplements
SEM, Kasumi-1	RPMI 1640	10% (v/v) FBS 2 mM L-glutamine
SKNO-1	RPMI 1640	20% (v/v) FBS 2 mM L-glutamine 10 ng/ml GM-CSF
293T	DMEM	10% (v/v) FBS 4 mM L-glutamine 1 mM sodium pyruvate
MS-5	α -MEM	10% (v/v) FBS 2 mM L-glutamine
Cord blood CD34 ⁺ MLL/Af4 – myeloid condition	IMDM	10% (v/v) FBS 2 mM L-glutamine 10 ng/ml SCF 10 ng/ml IL-3 10 ng/ml IL-6 10 ng/ml FLT-3L 10 ng/ml TPO
Cord blood CD34 ⁺ MLL/Af4 – lymphoid condition	α -MEM	10% (v/v) FBS 2 mM L-glutamine 10 ng/ml SCF 10 ng/ml FLT-3L 10 ng/ml IL-7 Co-culture with MS-5 cells

Table 2-4 Growing media and the supplements

2.1.3.2. *Mammalian cells transfection*

Name	Ingredients
2X HeBS	280 mM NaCl 50 mM HEPES 1.5 mM Na ₂ HPO ₄ pH 7.00
HEPES buffer solution	0.25% (v/v) 1 M HEPES pH 7.3
Polybrene solution, 8 mg/ml	8 mg/ml hexadimethrine bromide 0.9% (w/v) NaCl

Table 2-5 Transfection reagents

2.1.3.3. *Flow cytometry*

Name	Ingredients
Sort buffer	0.5% (v/v) filtered FBS 2 mM EDTA in Dulbecco's PBS
RF-10 media	10% (v/v) FBS 100 IU/ml penicillin-streptomycin 4 mM L-glutamine in RPMI 1640

Table 2-6 Flow cytometry and cell sorting media

2.1.3.4. *DNA electrophoresis*

Name	Ingredients
5X DNA loading dye	40% (v/v) glycerol 0.01% (w/v) bromophenol blue 10 mM EDTA pH 8.0
50X TAE	2 M tris 1 M acetic acid 50 mM EDTA pH 8.0
10X TBE	890 mM tris 890 mM boric acid 20 mM EDTA pH 8.0

Table 2-7 DNA electrophoresis buffers

2.1.3.5. *Bacterial culture*

Name	Ingredients
LB medium	1% (w/v) tryptone 0.5% (w/v) yeast extract 1% (w/v) NaCl pH 7.4
LB agar	LB media 1.5% (w/v) agar
TB medium	1.2% (w/v) tryptone 2.4% (w/v) yeast extract 0.4% (v/v) glycerol 17 mM KH ₂ PO ₄ 72 mM K ₂ HPO ₄ (potassium phosphate solution was autoclaved separately and added when the components were less than 60°C)
SOB media	2% (w/v) tryptone 0.5% (w/v) yeast extract 0.05% (w/v) NaCl 2.5 mM KCl pH 7.0 20 mM MgSO ₄ (magnesium sulphate solution was autoclaved separately and added before used)
SOC media	SOB media 20 mM glucose (glucose solution was sterilised through 0.22 µm filter and added when the components were less than 60°C)
Inoue transformation buffer	55 mM MnCl ₂ 15 mM CaCl ₂ 250 mM KCl 10 mM PIPES (0.5 M, pH 6.7)

Table 2-8 Bacterial culture and transformation buffer

2.1.3.6. *Western blotting and Co-IP*

Name	Ingredients
RIPA buffer	50 mM tris pH 7.4 150 mM NaCl 1% (v/v) Triton X-100 0.5% sodium deoxycholate 0.1% (w/v) SDS 1 tablet protease inhibitor cocktail (Roche) in a 50 ml solution
5X sample buffer	300 mM tris pH 6.8 10% (w/v) SDS 50% (v/v) glycerol 0.01% (w/v) bromophenol blue 5% β -mercaptoethanol
Co-IP lysis buffer	20 mM tris pH 7.5 150 mM NaCl 1 mM EDTA 1 mM EGTA 1% (v/v) Triton X-100 1 table protease inhibitor cocktail (Roche) in 50 ml solution
4X separating gel buffer	1.5 M tris 0.4% (w/v) SDS pH 8.8
4X stacking gel buffer	500 mM tris 0.4% (w/v) SDS pH 6.8
Electrophoresis buffer	25 mM tris 192 mM glycine 0.1% (w/v) SDS
Blotting buffer	25 mM tris 192 mM glycine 10% (v/v) methanol
TBS	20 mM tris 137 mM NaCl pH 7.6

Name	Ingredients
TBS-T	10 mM tris 150 mM NaCl 0.1% (v/v) Tween-20
Coomasie blue stain	0.1% (w/v) Coomassie G-250 10% (v/v) acetic acid 40% (v/v) methanol
Destaining Coomasie buffer	40% (v/v) methanol 10% (v/v) acetic acid

Table 2-9 Western blot and Co-IP buffers

2.1.4. Bacterial strains

Strain	Genotype	Vendor
Stbl3	F- <i>mcrB mrrhsdS20</i> (r _B ⁻ , m _B ⁻) <i>recA13 supE44 ara-14 galK2 lacY1 proA2 rpsL20</i> (Str ^R) <i>xyl-5 λ-leumtl-1</i>	Invitrogen
XL-1 Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac</i> [F' <i>proAB lac^fZΔM15 Tn 10</i> (Tet ^r)]	Agilent

Table 2-10 Bacterial strains

2.1.5. Antibodies

Antibodies for FACS staining, primary and secondary western blotting are listed:

Epitope	Clone	Manufacturer
hCD19-APC	SJ25C1	BD Biosciences
hCD19-APC/Cy7	SJ25C1	BioLegend
hCD19-BV421	HIB19	BioLegend
hCD19-PerCP-Cy5.5	SJ25C1	BD Biosciences
hCD33-APC/Cy7	P67.6	BioLegend
hCD33-BV421	WM53	BioLegend
hCD33-PE	WM53	BD Biosciences
hCD45-FITC	2D1	BD Biosciences
mCD45-PE-Cy7	30-F11	BD Biosciences
hCD117-BV421	104D2	BioLegend

Table 2-11 Fluorochrome labelled antibodies for flow cytometry

Epitope	Species	Dilution	Manufacturer	Catalogue
BMI1	Rabbit	1:5000	BETHYL	A301-694A
CHD4	Rabbit	1:10000	BETHYL	A301-081A
EZH2	Mouse	1:500	Merck Millipore	MABE 362
GAPDH	Mouse	1:320000	HyTest	5G4
IKZF1	Mouse	1:1000	Merck Millipore	MABE912
Strep-tag	Rabbit	1:1000	ABCAM	ab76949

Table 2-12 Primary antibodies for western blotting

Epitope	Species	Dilution	Manufacturer	Catalogue
Anti-mouse HRP	Goat	1:10000	DAKO	PO447
Anti-rabbit HRP	Goat	1:10000	DAKO	PO448

Table 2-13 Secondary antibodies for western blotting

2.1.6. Oligonucleotides

All of the oligonucleotides were purchased from Sigma Aldrich, desalted purification.

2.1.6.1. qRT-PCR primers

Name	Sequences (5'→3')
ACAP1	Fw : CTTCGTTGTCGGCATTGTG
	Rev: GGCTCACGGTGAATTTTCC
AF4/MLL	Fw : CAGAAGCCCACGGCTTATGT
	Rev: GGCAATACACAACTCAACAG
CD19	Fw : TGACCCACCAGGAGATTCTT
	Rev: CACGTTCCCGTACTGGTTCTG
CD33	Fw : CTCGTGCCCTGCACTTTCTT
	Rev: CCCGGAACCAGTAACCATGA
CD79A	Fw : GTGGCCCCCTGAGTTCTTG
	Rev: TCTTGTTACATTCTGGATGATCA
CEP164	Fw : GCCTGGACTTCGGTT
	Rev: TGTCTTCTATCCAGTGGTTGCT
CHD4	Fw : TGCTGACACAGTTATTATCTATGACTCTGA
	Rev: ACGCACGGGTCACAAACC
CSF1R	Fw : CCTCGCTTCCAAGAATTGCA
	Rev: CCATTGGTCAACAGCACGTT
CSF3R	Fw : CCCAGGCGATCTGCATACTT
	Rev: AACAAGCACAAAAGGCCATTG
DHX36	Fw : ATGCCTACAGTTAACCAGACACA
	Rev: ATACAGATGATAGCAATGACCAGG
EBF1	Fw : GAATTCACTACCGGCTTCAGCTT
	Rev: GTCAATGAGGCGCACGTAGA
EGR1	Fw : AAGTTTGCCAGGAGCGATGA
	Rev: CCGCAAGTGGATCTTGGTATG
GAPDH	Fw : GAAGGTGAAGGTCGGAGTC
	Rev: GAAGATGGTGATGGGATTTTC
HBA1	Fw : AAGTTTGCCAGGAGCGATGA
	Rev: CCGCAAGTGGATCTTGGTATG

Name	Sequences (5'→3')
HIPK1	Fw : CATTGTGCCTCCACTGAA
	Rev: CTGCTGGTTCTGGCTAAG
HOXA5	Fw : AGTCATGACAACATAGGCGGC
	Rev: CGGGTCAGGTAACGGTTGAA
HOXA6	Fw : CGGTTTACCCTTGGATGCA
	Rev: GCCCATGGCTCCCATACAC
HOXA7	Fw : GAGGCCAATTTCCGCATCTA
	Rev: GCGGTTGAAGTGGAACTCCTT
HOXA9	Fw : CCACCATCCCCGCACA
	Rev: TTTCCAAGGCAAACCTGTT
HOXC5	Fw : AGCCAATTCATTCTATAAGCAGAG
	Rev: CAATCCGCCGTAGCAGTA
IL6R	Fw : ACATTCACAACATGGATGGTCAA
	Rev: CGTGGATGACACAGTGATGCT
IL6ST	Fw : GGAAGCCCTGAATCCATAAAGG
	Rev: CTTTGGAAGGTGGAGCTTGTTT
ITGAL	Fw : TCTGCAGGGTCCCATGCT
	Rev: CATCAAACAGAAATACCAGGTCTACGT
KIT	Fw : GGACCAGGAGGGCAAGTCA
	Rev: GATAGCTTGCTTTGGACACAGACA
LEF1	Fw : CGGGTGGTGTGGACAGATC
	Rev: CCTGAATCCACCCGTGATG
MDK	Fw : AGTTTGGAGCCGACTGCAA
	Rev: CATCACACGCACCCAGTT
MLL/AF4	Fw : ACAGAAAAAAGTGGCTCCCCG
	Rev: TATTGCTGTCAAAGGAGGCGG
MT2A	Fw : CTCCAAGTCCCAGCGAACC
	Rev: CTTTACATCTGGGAGCGGGG
NCOA2	Fw : TGCGAATTTACAGAGCACTTTT
	Rev: GGAAAGGTCCAGCACCCAGTT
p16	Fw : TTCCTGGACACGCTGGTGGTG
	Rev: GGCATCTATGCGGGCATGGTTA

Name	Sequences (5'→3')
PAX5	Fw : GATGGTGCCTGGGAGTGAGT
	Rev: GGAGTCGTTGTACGAGGAATACTGA
PHF3	Fw : ATGGACCTGGGCTTGAAGT
	Rev: TGGTGGTGCACCTTTCAGGAG
PPP1R7	Fw : CAGGAGATGATGGAGGTTGACA
	Rev: CGATGCCACTGCTGTGTTTC
PRSS12	Fw : GTCAGCTTGGCTACAAGGGT
	Rev: AGTCAGCCAAGGACCTCTCA
ZFH3	Fw : CAACGCAGATAACGACAGT
	Rev: TGACCAGACCAGATGACAA
ZHX3	Fw : CCTCGTGGTCTCTTGTCTCTT
	Rev: TTGGCTGGCTCTCCTCTC

Table 2-14 qRT-PCR primers

2.1.6.2. Primers *MLLr* transcript breakpoint identification

Name	Sequences (5'→3')
MLL exon 8	Fw : GCCTCCACCACCAGAATCAG
MLL exon 9	Fw : CCGCCAAGTATCCCTGTAA
AF4 spanning exon 8 and 9	Rev: CGAGCATGGATGACGTTCT
AF9 spanning exon 10 and 11_A	Rev: TCACGATCTGCTGCAGAATGT
AF9 spanning exon 10 and 11_B	Rev: GGTTCACGATCTGCTGCAGAA

Table 2-15 Primers *MLL/AF4* and *MLL/AF9* for detecting common breakpoint in transcript

2.1.6.3. *Mutagenesis primers*

Name	Sequences (5' → 3')
ACAP1	Fw : CCTGTTCTGAAACcGGGAGCTGATCTGG
	Rev: CCAGATCAGCTCCCgGTTTCAGGAACAGG
CHD4	Fw : GGAGGGTGGGCATCaTGTA CTCTTTTCCC
	Rev: GGGAAAAGATGAGTACAtGATGCCACCCTCC
PHF3	Fw : CTTTTTGATCTCAACTGCAtAATCTGCATAGGTGGAATGG
	Rev: CCATTCGACCTATGCAGATTaTGCA GTTGAGATCAAAAAG
PPP1R7	Fw : GGGATCTAACCGCATCCtGGCAATCGAAAATATCG
	Rev: CGATATTTTCGATTGCCaGGATGCGGTTAGATCCC

Table 2-16 Primers mutagenesis candidate genes. The mutated nucleotide is indicated by the lower case.

2.1.6.4. *Fluidigm MiSeq candidate gene primers – cDNA template*

Name	Sequences (5' → 3')
ACAP1	Fw : AACTGACGACATGGTTCTACTGCACCACGCAACCATT
	Rev: TACGGTAGCAGAGACTTGGTCTGGAGAAGTCGCGGAAGATGTC
CHD4	Fw : AACTGACGACATGGTTCTACATATTGCTGCTGCAGAAAATGCT
	Rev: TACGGTAGCAGAGACTTGGTCTCATCTTGGTCATCTGGGAAAAGA
CEP164	Fw : AACTGACGACATGGTTCTACATTGAAACCAGAGCTAAAGATGTCAA
	Rev: TACGGTAGCAGAGACTTGGTCTTGCCTCTGCACATCAAGCA
CHTF8	Fw : AACTGACGACATGGTTCTACATGGGTGCTGATGGAGCTACA
	Rev: TACGGTAGCAGAGACTTGGTCTTGATCCCCAGGAGTGTGTTG
DHX36	Fw : AACTGACGACATGGTTCTACATGTTTCGGAAAATAGTAATTGCTACCA
	Rev: TACGGTAGCAGAGACTTGGTCTCCCACTCAGCGGACATTGTA
MLL/AF4	Fw : AACTGACGACATGGTTCTACAACAGAAAAAAGTGGCTCCCCG
	Rev: TACGGTAGCAGAGACTTGGTCTTATTGCTGTCAAAGGAGGCGG
NCOA2	Fw : AACTGACGACATGGTTCTACAGCCGTGAGCAGTGAGAGA
	Rev: TACGGTAGCAGAGACTTGGTCTGCAAGCTACCTGTGGAGTCTTTGT
PHF3	Fw : AACTGACGACATGGTTCTACACAGGAAGCAGCCATGGAGAT
	Rev: TACGGTAGCAGAGACTTGGTCTCTACAGGTGGTGCCATTCGA
PPP1R7	Fw : AACTGACGACATGGTTCTACACAGTAAAATTGAGAACTTAAGCAACTTACA
	Rev: TACGGTAGCAGAGACTTGGTCTCCAAAAACAACTCTCCAGGTT

Table 2-17 Primers MiSeq MLL/AF4 and candidate genes for cDNA template. The primers include CS1 (5'-AACTGACGACATGGTTCTACA-3') and CS2 (5'-TACGGTAGCAGAGACTTGGTCT-3') on the forward and reverse, respectively, that complement to the Fluidigm MiSeq barcode system.

2.1.6.5. *Fluidigm MiSeq candidate gene primers – DNA template*

Name	Sequences (5'→3')
ACAP1	Fw : ACACTGACGACATGGTTCTACACTGCACCACGCAACCATTTC
	Rev: TACGGTAGCAGAGACTTGGTCTGGCATTCTTACAGGGTGACGAT
BACE2	Fw : ACACTGACGACATGGTTCTACATTTGAGCGAGCCCATTTTGT
	Rev: TACGGTAGCAGAGACTTGGTCTGCAGCCCAATCCACTCTGTT
CACNB4	Fw : ACACTGACGACATGGTTCTACAGAGTCATCTCTGCAGGTTTCTTGA
	Rev: TACGGTAGCAGAGACTTGGTCTCAGAGCTGATTCCTGGCTTCA
CEP164	Fw : ACACTGACGACATGGTTCTACAGTCAGCCAGAAAATCCTGTCTCTT
	Rev: TACGGTAGCAGAGACTTGGTCTTGCCCTGGACAGATCAGATG
CHD4	Fw : ACACTGACGACATGGTTCTACAGCAGTGCCCTAATCAGAGCAT
	Rev: TACGGTAGCAGAGACTTGGTCTGTCCACATGATACCTGGGAAAAG
CHTF8	Fw : ACACTGACGACATGGTTCTACAATGCCCTAGCACTGCCCAATT
	Rev: TACGGTAGCAGAGACTTGGTCTCCCAGGAGGTTTCCAGCTAAT
DHX36	Fw : ACACTGACGACATGGTTCTACATCGGAAAATAGTAATTGCTACCAACA
	Rev: TACGGTAGCAGAGACTTGGTCTGCCTGGGCAACATAGTGAGACT
IL23A	Fw : ACACTGACGACATGGTTCTACATGGGAGACTCAGCAGATTCCA
	Rev: TACGGTAGCAGAGACTTGGTCTTGGAGATCTGAGTGCCATCCT
MLL/AF4	Fw : ACACTGACGACATGGTTCTACAAAATTATTTTTTGACCCCAACATCCT
	Rev: TACGGTAGCAGAGACTTGGTCTTTTGGGTGTGCATGCCTGTA
NCOA2	Fw : ACACTGACGACATGGTTCTACAGGCCGTGAGCAGTGAGAGA
	Rev: TACGGTAGCAGAGACTTGGTCTGGACTCCTGGCTCAGGTCTTT
PHF3	Fw : ACACTGACGACATGGTTCTACAAGCCGCCAATAAGTCATTGG
	Rev: TACGGTAGCAGAGACTTGGTCTGTCCACACGAAAGCTTAAAGACTGTGA
PPP1R7	Fw : ACACTGACGACATGGTTCTACAACGAAAAAATCTTCTTGGTCAACA
	Rev: TACGGTAGCAGAGACTTGGTCTGCTAGCTGCCTCTCCCTGAA
SLC4A8	Fw : ACACTGACGACATGGTTCTACACCCTTAGCCAACAAGGGATGA
	Rev: TACGGTAGCAGAGACTTGGTCTTCTGTTAATAATGACGGCTGTGA
ZNF136	Fw : ACACTGACGACATGGTTCTACAAAAATCCCTGGAGTGAACTCTGT
	Rev: TACGGTAGCAGAGACTTGGTCTTGGTGAGAACTGAAGGGTTTCC

Table 2-18 Primers MiSeq *MLL/AF4* and candidate genes for DNA template. The primers include CS1 (5'-ACACTGACGACATGGTTCTACA-3') and CS2 (5'-TACGGTAGCAGAGACTTGGTCT-3') on the forward and reverse, respectively, that complement to the Fluidigm MiSeq barcode system.

2.1.6.6. *shRNA sequences*

Name	Sequences (5' -> 3')	Target region
shACAP1_1	Fw: agcgcGCTGATGTCAACTGGGTCAATtagtgaagccacagatgtaATTGACCCAGTTGACATCAGct	CDS
	Rev: ggcaaGCTGATGTCAACTGGGTCAATtacatctgtggcttcactaATTGACCCAGTTGACATCAGCg	
shACAP1_2	Fw: agcgcTCACGCTAAATACGTGGAGAAtagtgaagccacagatgtaTTCTCCACGTATTTAGCGTGat	CDS
	Rev: ggcaaTCACGCTAAATACGTGGAGAAtacatctgtggcttcactaTTCTCCACGTATTTAGCGTGAg	
shACAP1_3	Fw: agcgcGCAGGAGATGAGACGTATCTTtagtgaagccacagatgtaAAGATACGTCTCATCTCCTGct	3'UTR, CDS
	Rev: ggcaaGCAGGAGATGAGACGTATCTTtacatctgtggcttcactaAAGATACGTCTCATCTCCTGCg	
shCEP164_1	Fw: agcgcGCAGTGAAAGTTCTGAATCTTtagtgaagccacagatgtaAAGATTCAGAACTTTCACTGct	CDS
	Rev: ggcaaGCAGTGAAAGTTCTGAATCTTtacatctgtggcttcactaAAGATTCAGAACTTTCACTGCg	
shCEP164_2	Fw: agcgcGCATTGTTTCCATCTGTCTTTtagtgaagccacagatgtaAAAGACAGATGGAAACAATGct	3'UTR, CDS
	Rev: ggcaaGCATTGTTTCCATCTGTCTTTtacatctgtggcttcactaAAAGACAGATGGAAACAATGCg	
shCEP164_3	Fw: agcgcTTACCTCTCTTCTCGTCAACAtagtgaagccacagatgtaTGTTGACGAGAAGAGAGTAat	CDS
	Rev: ggcaaTTACCTCTCTTCTCGTCAACAtacatctgtggcttcactaTGTTGACGAGAAGAGAGTAAg	
shCHD4_1	Fw: agcgcCGGGAGTTCACTACCAATAAtagtgaagccacagatgtaTTATTGGTACTGAACTCCCGct	CDS
	Rev: ggcaaCGGGAGTTCACTACCAATAAtacatctgtggcttcactaTTATTGGTACTGAACTCCCGCg	
shCHD4_2	Fw: agcgcTTCCTGCCAGGCTTGAAGAAAtagtgaagccacagatgtaTTTCTTCAAGCTGGCAGGAat	3'UTR, CDS
	Rev: ggcaaTTCCTGCCAGGCTTGAAGAAAtacatctgtggcttcactaTTTCTTCAAGCTGGCAGGAag	
shCHD4_3	Fw: agcgcCGAAGGTTTAAAGCTCTTAGAAtagtgaagccacagatgtaTTCTAAGAGCTTAAACCTTCgt	CDS
	Rev: ggcaaCGAAGGTTTAAAGCTCTTAGAAtacatctgtggcttcactaTTCTAAGAGCTTAAACCTTCGg	
shDHX36_1	Fw: agcgcTCCGCTGAGTGGGTTAGTAAAtagtgaagccacagatgtaTTTACTAACCCTCAGCGGat	CDS
	Rev: ggcaaTCCGCTGAGTGGGTTAGTAAAtacatctgtggcttcactaTTTACTAACCCTCAGCGGAg	
shDHX36_2	Fw: agcgcGCAGGAGAAGAACAATTTGAtagtgaagccacagatgtaTACAATTTGTTCTTCTCGTct	CDS
	Rev: ggcaaGCAGGAGAAGAACAATTTGAtacatctgtggcttcactaTACAATTTGTTCTTCTCGTCg	
shDHX36_3	Fw: agcgcCCCCTCTTTGGGAGTATATtagtgaagccacagatgtaAATATACTCCCAAAGAGTGGgt	3'UTR
	Rev: ggcaaCCCCTCTTTGGGAGTATATtagtgaagccacagatgtaAATATACTCCCAAAGAGTGGGg	
shNCOA2_1	Fw: agcgcGCACCTTGTGCTGCACAAAtagtgaagccacagatgtaTTTGTGCAGCAACAAGAGTGct	CDS
	Rev: ggcaaGCACCTTGTGCTGCACAAAtacatctgtggcttcactaTTTGTGCAGCAACAAGAGTGCg	
shNCOA2_2	Fw: agcgcATCCGTTCTCAGACTACTAAtagtgaagccacagatgtaATTAGTAGTCTGAGAACGGATt	CDS
	Rev: ggcaaATCCGTTCTCAGACTACTAAtacatctgtggcttcactaATTAGTAGTCTGAGAACGGATg	
shNCOA2_3	Fw: agcgcATTACCTTAGTGCACTTAGtagtgaagccacagatgtaCTAAGTTGCACTAAGGTGAAtt	3'UTR
	Rev: ggcaaATTACCTTAGTGCACTTAGtagtgaagccacagatgtaCTAAGTTGCACTAAGGTGAATg	
shPHF3_1	Fw: agcgcCCAGTCAAGTAGCGTTTCTTAtagtgaagccacagatgtaTAAGAAACGCTACTTGACTGGt	CDS
	Rev: ggcaaCCAGTCAAGTAGCGTTTCTTAtacatctgtggcttcactaTAAGAAACGCTACTTGACTGGg	
shPHF3_2	Fw: agcgcGCAACTGGATAGGCCATTTAAtagtgaagccacagatgtaTAAATGGCCTATCCAGTTGct	CDS
	Rev: ggcaaGCAACTGGATAGGCCATTTAAtacatctgtggcttcactaTAAATGGCCTATCCAGTTGCg	
shPHF3_3	Fw: agcgcCCTCGTTTAAATGGCACAGAAtagtgaagccacagatgtaTTCTTGTGCCATTAACGAGGgt	5'UTR, CDS
	Rev: ggcaaCCTCGTTTAAATGGCACAGAAtacatctgtggcttcactaTTCTTGTGCCATTAACGAGGg	
shPHF3_4	Fw: agcgcGCCAATAAGTCATTGGAGAAtagtgaagccacagatgtaTTCTCCAATGACTTATTGGCGt	CDS
	Rev: ggcaaGCCAATAAGTCATTGGAGAAtacatctgtggcttcactaTTCTCCAATGACTTATTGGCGg	
shPHF3_5	Fw: agcgcATCTATTGTTGGGCTTAATAtagtgaagccacagatgtaTAATTAAGCCCAACAATAGATt	CDS
	Rev: ggcaaATCTATTGTTGGGCTTAATAtacatctgtggcttcactaTAATTAAGCCCAACAATAGATg	
shPPP1R7_1	Fw: agcgcAGTTCTGGATGAACGACAATtagtgaagccacagatgtaGATTGTCGTTTCCAGAACTt	CDS
	Rev: ggcaaAGTTCTGGATGAACGACAATtagtgaagccacagatgtaGATTGTCGTTTCCAGAACTg	
shPPP1R7_2	Fw: agcgcGCAACTTACATCAACTACAGAtagtgaagccacagatgtaTCTGTAGTTGATGTAAGTTGct	CDS
	Rev: ggcaaGCAACTTACATCAACTACAGAtacatctgtggcttcactaTCTGTAGTTGATGTAAGTTGCg	

Name	Sequences (5' -> 3')	Target region
shPPP1R7_3	Fw: agcgcCAGTCACAAACCCAATGGCAAtagtgaagccacagatgtaTTGCCATTGGGTTTGTGACTGt	3'UTR
	Rev: ggcaaCAGTCACAAACCCAATGGCAAtacatctgtggccttcactaTTGCCATTGGGTTTGTGACTGg	
shNTC	Fw: agcgATCTCGCTTGGGCGAGAGTAAGtagtgaagccacagatgtaCTTACTCTCGCCCAAGCGAGAt	<i>Lupinus angustifolius</i>
	Rev: ggcaCTCTCGCTTGGGCGAGAGTAAGtacatctgtggccttcactaCTTACTCTCGCCCAAGCGAGAg	

Table 2-19 shRNA sequences on pLKO5d.SFFV.miR30n backbone. The upper case nucleotides on the sequences column indicate the target sequences. The lower case nucleotides on each end indicate the provided BsmBI restriction site. The lower case nucleotides in the middle sequences indicate the shRNA loop.

2.1.7. Software

Name	Manufacturer
CellQuest Pro	BD Biosciences
FACSDiva	BD Biosciences
FlowJo v10.0.8	FlowJo
MARS Data Analysis Software	BMG Labtech
Primer Express 2.0	Applied Biosystems
QuantStudio Real-Time PCR System	Applied Biosystems
SDS 2.3	Applied Biosystems

Table 2-20 Software

2.2. Cell culture methods

2.2.1. Culture of cell lines

2.2.1.1. Culture of suspension cells

The suspension cell lines were cultured in tissue culture flasks at 37°C, 5% CO₂ in humidified conditions. The cells were maintained at a concentration of 5 x 10⁵ to 2 x 10⁶ cells/ml by replacing media every 2 – 3 days. All of the cell lines were validated by the presence of specific fusion gene sequences.

2.2.1.2. Culture of adherent cells

The adherent cell lines were cultured in tissue culture dishes or flasks at 37°C, 5% CO₂ in humidified conditions. The cells were passaged when they reached 80 – 90% confluence by washing with PBS and then adding 1 volume (approximately 0.026 ml/cm²) of 0.5X trypsin-EDTA, incubated 1 – 10 min, depending on the cell line, at 37°C. The enzyme reaction was stopped by the addition of 1 volume of growth media; 1/10 of these cells were transferred into a new flask with pre-warmed culture medium.

2.2.2. Culture of CD34+ cord blood MLL/Af4 cells

2.2.2.1. Culture of CD34+ cord blood MLL/Af4 cells in myeloid condition

CD34+ cord blood MLL/Af4 cells were obtained from Mulloy's lab in frozen vials. The cells are immortal in myeloid condition media (composition was written in Table 2-4) as described previously[82].

After thawing, *i.e.* the recovery phase, the cells were grown at a concentration of $1 - 2 \times 10^6$ cells/ml. One week after the recovery, the cells were grown at a concentration of $0.25 - 0.5 \times 10^6$ cells/ml, adjusted once a week. Fresh media was added in the middle of the week. The cells had approximately three doublings per week.

2.2.2.2. Culture of CD34+ cord blood MLL/Af4 in lymphoid condition

Lymphoid culture components were described in Table 2-4. Co-culture with MS-5 cells was included. Lymphoid differentiation must be performed on the cord blood MLL/Af4 before they reach week 10. After this timepoint, no lymphoid differentiation could be achieved, as informed by Mulloy's lab.

MS-5 cells were grown at a concentration of $4 - 5 \times 10^3$ cells/cm² in α -MEM 10% FBS one day before the co-culture (it was preferable to use 12 well plate or larger surface area plate/flask due to the expansion of the cord blood cells after the long culture). On the day of co-culture, the MS-5 cells should reach 60-80% confluence. The cord blood cells were seeded at a concentration of $1 - 2 \times 10^5$ cells/ml onto the MS-5 layer in complete lymphoid growing media. During the co-culture, the media only required a 50% replacement, once a week.

Lymphoid differentiation was observed from week 2 – 3, and became the majority of the population on week 5 – 6. During this period, the MS-5 layer was effective at supporting growth for at least 19 days before it detached. When it started to detach, the cord blood cells were transferred onto a new MS-5 layer.

2.2.3. Cell counting and concentration determination

Cells numbers were counted using an Improved Neubauer haemocytometer (Hawksley). An aliquot of suspension cells was mixed with the viability stain, trypan blue (0.2% w/v), in a ratio 1:1, and 10 μ l of the mixture was loaded into the haemocytometer. The slide was transferred to an inverted microscope with 10x objective magnification for counting. For routine subculture, two quadrants of the haemocytometer were counted, while for more precise experiments, four quadrants were counted. The cell concentration (cell/ml) is determined by dividing the average number of cells (total number of cells/number of counted quadrants) by 0.1 mm³, or 10⁻⁴ ml (the volume of each quadrant), compensating for the dilution factor of the viability stain.

2.2.4. Thawing cells

Frozen cells were thawed rapidly in a 37°C water bath, the cryovial was cleaned with 70% ethanol, and diluted in 10 ml growth media. The suspensions were centrifuged at 300 g for 5 min to remove the supernatant with the cryoprotectant (DMSO). The cells were resuspended in fresh growth media at a concentration of 10⁶ cells/ml for cell lines, or 1 – 2 x 10⁶ cells/ml for patient materials and primograft cells.

The cell lines were cultured and routinely assessed to ensure a normal rate of growth prior to any experimental procedures.

2.2.5. Freezing cells

Cells were grown to reach the late log phase for cryopreservation. They were counted and centrifuged at 300 g for 5 min. The cells were resuspended in freezing medium at a concentration of 5 x 10⁶ to 10⁷ cells/ml and transferred to pre-labelled cryovials. They were placed in an insulated freezing box to provide a slow cooling and transferred to -80°C.

2.2.6. Lentivirus production

Lentivirus was produced in 293T cells by co-transfection of the lentiviral envelope (pMD2.G), packaging (pCMVdR8.91), and expression plasmids using calcium phosphate precipitation method. 293T cells with less than 16 passages were used for optimal lentiviral production. They were grown in 150 mm tissue culture dishes at a concentration of 3×10^6 cells in 30 ml medium the day prior to co-transfection. On the following day when the cells had reached 30 – 50% confluence, 60 µg of the lentiviral transfer vector containing expression cassette, 45 µg of pCMVdR8.91, and 15 µg of pMD2.G were mixed. The volume of the mixture was adjusted with HEPES buffer solution to a total volume of 750 µl. A volume of 750 µl of 0.5 M calcium chloride solution was added and mixed well. This solution was added to a prepared 1.5 ml of 2X HeBS pH 7.00 solution in a 5 ml vial. They were mixed by vortexing and left for 30 minutes at room temperature to allow the formation of the calcium phosphate precipitate. This suspension was then added dropwise and evenly on 293T cells. The cells were cultured for 6 – 14 hours at standard culture conditions. Afterwards, the medium was removed, the cells were washed with PBS, and 30 ml fresh growth medium was added. They were cultured for three days before harvesting the lentiviral particles (section 2.2.7).

2.2.7. Lentiviral harvesting and concentrating

Approximately 90 hours following co-transfection, the lentivirus particles were collected. The 30 ml supernatant of the transfected cells containing the viruses were collected in a conical tube. It was centrifuged at 600 g for 15 min at 4°C, and the supernatant was further filtered using PVDF membrane filter with 0.45 µm pore size to remove any residual debris.

The lentivirus particles were concentrated using a Beckman Optima L-100 XP Ultracentrifuge at 120,000 g, 4°C for 2 h in a swinging bucket rotor within a constant vacuum. Afterwards, the supernatant was removed by carefully decanting, and the viruses were resuspended in 1 ml of the experimented target cells growth media, aliquoted into several (up to fifteen) tubes, and stored at -80°C.

2.2.8. Lentiviral transduction

Lentiviral transduction (including transduction on cord blood cells) was performed using the spinoculation method. The cells were prepared and adjusted to a concentration of 10^6 cells/ml. Polybrene (stock concentration is 8 mg/ml) was supplemented to a final concentration of 8 µg/ml to assist in neutralising charge

repulsion between virions and sialic acid of the cells. The cells were plated into a 48 well plate with a volume of 500 μl /well (*i.e.* 5×10^5 cells/well). Different amounts of virus were added to the cells and spinoculated at 900 g, 32°C for 50 min, followed with overnight incubation at 37°C. The cells were transferred into 24 well plate on the next day and cultured for at least three days before conducting subsequent analysis.

2.2.9. Single cell cloning by limiting dilution

The suspension cells were seeded at a concentration of 5×10^5 cells/ml the day prior to cloning to provide good conditions for the cells and reliable counting on the day of the cloning.

On the next day, the cells were counted and serially diluted (10 fold for each dilution) to obtain a final concentration 100 – 500 cells/ml. From this lower density, the cells were poured into growing media to get 10 ml volumes with a final concentration of 3 cells/ml. These cells were then distributed into 96 well U-bottom cell culture plates with 100 μl volume per well (*i.e.* approximately 1 cell per 3 wells). They were cultured in the standard conditions and transferred into larger culture plates when the numbers of cells were sufficient (approximately 3 – 4 weeks).

2.3. Flow cytometry and cell sorting

2.3.1. Flow cytometry analysis

Flow cytometry analyses involved analysis of the endogenous fluorescent protein and immunophenotypes of the cells. They were performed using FACSCalibur (Becton Dickinson), or FACSCanto II (Becton Dickinson) when it required a blue laser.

2.3.1.1. Endogenous fluorescent protein analysis

The transduced cells that have fluorescent protein marker could be measured using flow cytometry. A small number of the cells (as few as 10^5 cells) were put into a FACS tube, diluted to 4 ml using PBS, and centrifuged (450 g for 4 min) to prevent colour interference from the growing media. The cells were resuspended in PBS at a maximum concentration of 10^6 cells/ml and measured using appropriate fluorescent channels (Table 2-21).

2.3.1.2. Surface marker staining and analysis

A small number of cells ($1 - 5 \times 10^5$ cells) were collected into FACS tubes, made up to 4 ml using PBS, and centrifuged at 450 g for 4 min. They were resuspended in 50 μ l of 0.2% PBSA (0.2% BSA in PBS) containing the labelled antibodies (antibody volume was based on the manufacturer's recommendation) and incubated for 20 min in the dark at room temperature. Subsequently, the samples were washed by adding 3.5 ml of 0.2% PBSA and centrifuged, to remove the untagged antibodies. The samples were resuspended in 500 μ l of 0.2% PBSA and measured using the appropriate fluorescent channels (Table 2-21).

Fluorochrome	Fluorescent channel
APC	FL4
eGFP	FL1
FITC	FL1
PE	FL2
PE-Cy7	FL3
PerCp-Cy5.5	FL3
RFP657	FL4
tdTomato	FL2
Venus	FL1

Table 2-21 Fluorescent channels for different fluorochromes on FACSCalibur. APC-Cy7 and BV421 were analysed using FACSCanto II due to no available filters and laser in FACSCalibur. Also, the fluorescent channels have been named according to the fluorochromes in FACSCanto II, apart from eGFP and Venus, tdTomato, RFP657, and BV421 that were analysed using FITC, PE, APC, and Pacific Blue channels, respectively.

2.3.2. Cell sorting

Cell sorting techniques were used to isolate fluorescent protein expressing cells, labelled cells populations, or single cells. Three different instruments were used: FACS Aria III (Becton Dickinson), Astrios EQ Cell Sorter (Beckman Coulter) and FACS Aria Fusion Cell Sorter (Becton Dickinson). The FACS Aria III was located in NICR and experiments were carried out by Hesta McNeill, while others are located with and performed by the Flow Cytometry Core Facility, Newcastle University. The sample labelling for haematopoietic sub-population sorting was carried out with assistance from Dr Paul Milne, who had previously optimised the antibody panel used.

2.3.2.1. *Fluorescent protein expressing cells sorting*

Before sorting, the population of fluorescent protein-expressing cells was evaluated (section 2.3.1), to have an estimate of the number of cells that would be collected. After that, the desired number of cells was centrifuged (300 g, 5 min) and resuspended in sorting buffer at a concentration of $1 - 2 \times 10^7$ cells/ml. The cells were then passed through a 40 μ m cell strainer to remove any clumped cells and transferred into a FACS tube to be sorted.

The cells were sorted and collected into a 1.5 ml microfuge tube, FACS tube, or 15 conical tube containing 500 μ l, 1 ml, or 2 ml culture media, respectively, depending

on the number of cells to be collected. Afterwards, the cells were cultured under standard conditions.

2.3.2.2. *Haematopoietic sub-population sorting*

Haematopoietic sub-population analyses were performed on primary and primograft patient samples. The sorting preparation was assisted by Dr Paul Milne. Up to 10^7 cells were prepared, centrifuged, and resuspended in 100 μ l sort buffer. The antibodies were added and incubated in the dark at 4°C for 30 min. The antibodies to be used are listed in Table 2-22. After staining, the cells were washed, resuspended in sort buffer, and passed through a 40 μ m cell strainer to remove any clumped cells and transfer into a FACS tube, and sorted.

The sorted cells were collected into 1.5 ml microfuge tubes containing 500 μ l RF-10 media. After sorting, they were added with additional 500 μ l media to dislodge any cells that may attach to the upper part of the microfuge containing no media, and continued with further applications.

Antibody	Fluorochrome	Astrios filter
CD56	FITC	488/530
CD34	PerCP-Cy5.5	488/695
CD117	PE	561/586
CD19	PE-CF594	561/620
CD38	PECy7	561/795
CD33	APC	642/671
CD90	A700	642/722
CD3	APC-Cy7	642/795
CD123	BV421	405/448
CD45RA	BV510	405/530
CD10	BV650	405/671
HLA-DR	BV786	405/795

Table 2-22 Antibodies panel for haematopoietic sub-populations sorting on Astrios

2.3.2.3. *Single cells sorting*

Single cell analysis was started with the isolation of the single cells into 96 well plates (Eppendorf twin.tec PCR Plate 96 full skirt). The cells were sorted to the plates based on the design in Figure 2-1 that includes forty-five single cell samples, two 0 cells as a negative control, and one bulk (300 – 1,000 cells) as positive control sample in one

96 well plate. This design provides the plate with only 48 total samples to have a good time allocation for distribution of the whole genome amplification (WGA) reagents.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0	0	0	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0
C	0	bulk ctrl	sample	sample	0 (neg ctrl)	0						
D	0	sample	sample	sample	sample	sample	sample	sample	sample	sample	sample	0
E	0	sample	sample	sample	sample	sample	sample	sample	sample	sample	sample	0
F	0	sample	sample	sample	sample	sample	sample	sample	sample	sample	sample	0
G	0	sample	sample	sample	sample	sample	sample	sample	0 (neg ctrl)	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0

Figure 2-1 Plate design for single cell sorting. The design included two 0 cells as negative controls, one bulk (300 – 1,000 cells) as a positive control, and forty-five single cell samples.

Before the sorting, the plates were filled with 3.5 µl PBS/well and sealed with tape pad (it is recommended to prepare extra plates per cell type). The cells were prepared from as many as 10⁵ cells. They were centrifuged and re-suspended in 500 µl filtered 0.2% PBSA.

When the preparations were finished, the plate to be sorted was centrifuged at 2,000 rpm for 1 min. The sorting was performed by the Flow Cytometry Core Facility team using FACSaria Fusion Cell Sorter. During sorting, plates not actively on the cell sorter were kept on ice. Following the sorting, the plate was resealed with new tape pad and centrifuged at 2,000 rpm for 1 min to ensure single cells were not attached to the no-media surface of the well. Plates were then stored at -20°C until further application.

2.4. Molecular biology methods

2.4.1. RNA isolation

RNA was isolated using QIAGEN RNA isolation kits, based on the number of cells. These include the isolation from less than and more than 10^5 cells that were done using QIAGEN RNeasy Plus Micro Kit and RNeasy Mini Kit, respectively.

2.4.1.1. RNA isolation from small number of cells ($<10^5$ cells)

RNA from a small number of cells ($<10^5$ cells; e.g. cell clones) were isolated using RNeasy Plus Micro Kit according to the manufacturer's protocol. Cells were centrifuged at 500 g for 5 min, and then the supernatant was aspirated. They were lysed and homogenised by adding 350 μ l RLT Plus containing 1% β -mercaptoethanol and vortexing for 1 min. The lysate was transferred into a gDNA Eliminator spin column and centrifuged at 8,000 g for 30 s to eliminate gDNA. The flow-through was mixed with 350 μ l of 70% ethanol and transferred to the RNeasy MinElute spin column. The RNA bound to the column by centrifugation at 8,000 g for 30 s. The column with the RNA was washed three times with (1) 700 μ l Buffer RW1, (2) 500 μ l Buffer RPE (each was at 8,000 g for 30 s centrifugation), and (3) 500 μ l of 80% ethanol (8,000 g for 2 min centrifugation). The column was placed in a new collection tube, the lid was opened, and centrifuged at full speed for 5 min to remove residual ethanol. After that, the column was transferred to a 1.5 ml collection tube and the RNA was eluted from the column by adding 12.5 μ l RNase-free water, incubated at room temperature for 5 min, and then centrifuged at full speed for 1 min. The RNA was stored on the ice when the cDNA synthesis was performed immediately, otherwise RNA was stored at -20° C.

2.4.1.2. RNA isolation from larger number of cells ($>10^5$ cells)

General RNA isolation ($>10^5$ cells) was done using QIAGEN RNeasy Mini Kit. Up to 5×10^6 cells were collected and washed once with PBS (300 g for 5 min). The cell pellet was lysed by adding 350 μ l RLT buffer containing 1% β -mercaptoethanol. The lysed cells were transferred to Qiashredder column and centrifuged at full speed for 2 min to homogenise the sample. The homogenised sample was mixed with 350 μ l of 70% ethanol and applied to RNeasy spin column. The RNA bound to the column by centrifugation at 8,000 g for 30 s. The column with the RNA was washed three times that included once with 700 μ l RW1 buffer, and twice with 500 μ l RPE buffer (all of the washing was done by centrifugation at 8,000 g for 30 s). The column was dried by centrifugation at full speed for 2 min to remove residual ethanol. After that, the

column was transferred to a 1.5 ml collection tube and the RNA was eluted from the column by adding 50 µl RNase-free water, incubated at room temperature for 5 min, and then centrifuged at full speed for 1 min. The RNA was stored on the ice when the cDNA synthesis could be done soon afterwards, or otherwise stored at -20° C.

2.4.2. cDNA synthesis

The cDNA was synthesised from RNA using RevertAid H Minus First-Strand cDNA Synthesis Kit (Thermo Fisher Scientific). Up to 1 µg RNA was used as the template and mixed with 1 µl random hexamer primers (0.2 µg/µl), and the volume was adjusted to 12 µl with H₂O. Subsequently, the mixture was incubated at 70°C for 5 min to heat-denature the secondary structure of the RNA, followed by cooling at 4°C and adding 8 µl reverse transcription master mix components. The master mix includes 4 µl of 5X reaction buffer, 1 µl Ribolock RNase inhibitor (20 U/µl), 2 µl of 10 mM dNTP, and 1 µl RevertAid H minus reverse transcriptase (200 U/µl). The reaction was started with incubation at 25°C for 10 min to extend the random hexamer primers, continued with the reverse transcriptase by incubating at 42°C for 60 min, and heat-inactivated at 70°C for 10 min. The product was diluted with H₂O, and the added volume depends on the starting template amount (Table 2-23).

Template	V_{H₂O} (µl)
Small amount of cells (<10 ⁵ cells)	0
<500 ng RNA	30
500 – 1,000 ng RNA	80

Table 2-23 H₂O volume for cDNA dilution

2.4.3. DNA isolation

The DNA was isolated using QIAGEN DNeasy Blood & Tissue Kit. Up to 5 x 10⁶ cells were harvested by centrifugation at 300 g for 5 min. The cells were resuspended in 200 µl PBS and added with 20 µl proteinase K (enzymatic activity 600 mAU/ml solution) to digest protein, in particular, DNases. The cells were lysed by adding 200 µl Buffer AL, vortexed, and incubated at 56°C for 10 min, and then mixed with 200 µl ethanol. The mixture was transferred into the DNeasy Mini spin column and centrifuged at 6,000 g for 1 min to allow the DNA bind to the column. The column with the DNA was transferred to another collection tube. It was washed twice, firstly by adding 500 µl Buffer AW1 and centrifuged at 6,000 g for 1 min, and secondly by adding 500 µl Buffer AW2 and centrifuged at 20,000 g for 2 min (new collection tube

was used for every washing step). The flow-through was removed and centrifuged again at 20,000 g for 1 min to remove any residual liquid. The column was transferred to a 1.5 ml microfuge tube and the DNA was eluted by adding 100 – 200 µl Buffer AE (depending on the necessity of concentrated DNA), incubated for ~5 min, and centrifuge at 17,900 g for 1 min. The DNA was stored at -20°C.

2.4.4. Simultaneous DNA and RNA extraction

DNA and RNA from a small number of cells (<10⁵ cells) from precious samples, e.g. the sorted haematopoietic populations, were extracted simultaneously using QIAGEN AllPrep DNA/RNA Micro Kit. The cells were centrifuged at 500 g for 5 min, and the supernatant was aspirated. They were lysed and homogenised by adding 350 µl RLT Plus containing 1% β-mercaptoethanol and vortexing for 1 min. The lysate was transferred to an AllPrep DNA spin column and centrifuged at 8,000 g for 30 s to bind the DNA, while the RNA would pass the membrane and was in the flow-through. The column was stored at 4°C, and the RNA was isolated first.

The flow-through containing RNA was augmented with 350 µl of 70% ethanol and transferred into the RNeasy MinElute spin column. It was centrifuged at 8,000 g for 30 s to allow the column to bind the RNA. The column was washed three times by adding 700 µl Buffer RW1 and 500 µl Buffer RPE (centrifuged 8,000 g for 30 s) and 500 µl of 80% ethanol (centrifuged 8,000 g for 2 min). After the washing steps, the column was placed in a new 2 ml collection tube, the lid was opened, and centrifuged at full speed for 5 min to dry any residual ethanol. Subsequently, it was placed in a new 1.5 ml collection tube, and the RNA was eluted by adding 12.5 µl RNase-free water, incubated at room temperature for 5 min, and then centrifuged at full speed for 1 min. The RNA was stored on the ice when the cDNA synthesis could be done soon afterwards, or otherwise stored at -20° C.

The process was continued by isolating the DNA that had bound to the AllPrep DNA spin column. The DNA and column were washed twice by adding 500 µl Buffer AW1 and AW2, centrifuged at 8,000 g for 30 and 20,000 g for 2 min, respectively. The column was placed in a new 1.5 ml collection tube, and the DNA was eluted by adding 30 µl preheated to 70°C Buffer EB. It was incubated at room temperature for 5 min, and then centrifuged at full speed for 1 min. After that, a further 30 µl pre-heated Buffer EB was added to elute any remaining DNA. It was incubated at room temperature for 5 min, and then centrifuged at full speed for 1 min. The DNA was stored at -20°C.

2.4.5. PCR

Different PCR components and reactions were performed depending on the length of amplicons, proofreading activity, and difficulty of the templates.

2.4.5.1. General PCR – short, non-complex amplicon

PCR on short (<750 bp), non-complex, and no proofreading required amplicons were performed using HotStarTaq DNA Polymerase (QIAGEN). It was done in 25 µl reaction with annealing temperature ranged from 0 – 5°C below melting temperature of the primer sets. DNA template was used in a range 20 – 200 ng, while cDNA was used 1 - 2 µl of the diluted product (refer to Section 2.4.2). The amount of template added was dependent on the availability of the material.

The pipetting scheme and thermal cycle are described in Table 2-24 and Table 2-25, respectively.

Components	Final concentration/amount
10x PCR buffer	1 x
dNTP mix (10 mM of each)	200 µM of each dNTP
10 µM primer mix	400 nM
DNA/cDNA	20 - 200 ng DNA or 1 - 2 µl cDNA
H ₂ O	up to 25 µl total volume
HotStarTaq DNA Polymerase (5U/µl)	0.625 U or 0.025 U/µl

Table 2-24 PCR components using HotStarTaq DNA Polymerase (QIAGEN)

Step	Temperature (°C)	Time	Cycles
Initial activation	95	15 min	1
Denaturation	94	30 s	
Annealing	T _m -(0 to 5)	30 s	18 – 35
Extension	72	1 min/kb	
Final extension	72	10 min	1

Table 2-25 Thermal cycle of PCR reaction using HotStarTaq DNA Polymerase (QIAGEN)

2.4.5.2. PCR longer or complex amplicon

Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB) was used for the reaction with (1) longer than 750 bp amplicon, (2) complex template (e.g. GC-rich, containing high impurities, a small amount of template, etc.), or (3) where proofreading was required. The annealing temperature was determined using the manufacturer's specific program (tmcalculator.neb.com/). DNA template was used in a range of 20 – 200 ng, while cDNA was used in the range of 1 – 5 µl, either from the diluted (refer to Section 2.4.2) or non-diluted product, dependent on the availability of the materials.

The pipetting scheme and thermal cycle are described in Table 2-26 and Table 2-27, respectively.

Components	Amount
2x Phusion master mix	12.5 µl
10 µM Primer mix	200 nM
DNA/cDNA	20 - 200 ng DNA or 1 - 5 µl cDNA
H ₂ O	Up to 25 µl total volume

Table 2-26 PCR components using Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB)

Step	Temperature (°C)	Time	Cycles
Initial activation	98	2 min	1
Denaturation	98	10 s	
Annealing	T _a	30 s	18 – 35
Extension	72	30 s/kb	
Final extension	72	10 min	1

Table 2-27 Thermal cycle of PCR reaction using Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB)

2.4.5.3. PCR identification transcript breakpoint MLL/AF4 and MLL/AF9

Common primers were designed to identify transcript breakpoint *MLL/AF4* and *MLL/AF9*. The locations of the primers were determined from the study cases by Cerveira *et al.*[192] and De Braekeleer *et al.*[193]. These studies showed exon 8 as the earliest location of the *MLL* breakpoint, and exon 6 and 9 as the latest location of *AF4* and *AF9* breakpoints, respectively. Based on these criteria, the combinations of the common primers were listed in Table 2-28.

Primer mix	Primer component
MLL/AF4 common 1	MLL exon 8
	AF4 spanning exon 8 and 9
MLL/AF4 common 2	MLL exon 9
	AF4 spanning exon 8 and 9
MLL/AF9 common 1	MLL exon 8
	AF9 spanning exon 10 and 11_A
MLL/AF9 common 2	MLL exon 9
	AF9 spanning exon 10 and 11_B

Table 2-28 Primer components of the common primer *MLL/AF4* and *MLL/AF9* transcript breakpoints. The sequences of them were listed in Table 2-15.

PCR reaction using Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB) Section 2.4.5.2 was used for the identification, with annealing temperature 65°C. The *MLL/AF4* and *MLL/AF9* products might have <1,000 bp and <1,800 bp sizes, respectively.

2.4.6. Multiplex PCR

Multiplex PCR was performed to amplify numerous candidate driver genes in order to preserve the valuable material. Before performing the multiplex PCR, single reaction PCR (Section 2.4.5.2) was conducted to validate the primers and the locations of the target bands. In our case, the multiplex PCR products would be tagged with a barcode for next generation sequencing (Illumina MiSeq). Therefore, high-fidelity polymerase (Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB)) was used to avoid A-tailing on the amplicons. Furthermore, Illumina MiSeq will read lengths up to 300 bp from both ends. Accordingly, the primers were designed to contain small amplicon sizes, ranging from 100 bp to 300 bp, or up to 600 bp if necessary (however, longer than 300 bp would only be covered by either one of the sequencing ends). The primers were also designed to have the same melting temperature (66°C) for all of the analysed genes.

It was recommended that each amplicon had different size of at least 25 bp to be able to be differentiated by gel electrophoresis. Up to six amplicons were designed in a single multiplex reaction. The primers were designed using Primer Express® Software v2.0 (Applied Biosystems).

After the validation by single reaction PCR, multiplex PCR was conducted with the pipetting scheme and thermal cycle shown in Table 2-29 and Table 2-30.

Components	Amount
2x Phusion master mix	12.5 µl
10 µM Primer mix	80 nM each primer set
DNA/cDNA	Up to 5 µl
H ₂ O	Up to 25 µl total volume

Table 2-29 Multiplex PCR components

Step	Temperature (°C)	Time	Cycles
Initial activation	98	2 min	1
Denaturation	98	10 s	
Annealing	T _m (66°C)	30 s	35
Extension	72	30 s	
Final extension	72	10 min	1

Table 2-30 Thermal cycle of the multiplex PCR

In the case when a particular amplicon had much lower signal than the others on the electrophoresis, the primer amount of the appropriate amplicon was increased.

2.4.7. Barcoding PCR

Multiplex PCR products required the addition of barcodes for Illumina MiSeq. Barcodes tagging was produced by barcoding PCR using the primers Access Array Barcode Library for Illumina Sequencers-384, Single Direction (Fluidigm). Multiplex PCR products per sample were pooled for one reaction, *i.e.* one barcode/sample. Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB) was used for the amplification. The pipetting scheme and thermal cycle are described in Table 2-31 and Table 2-32, respectively.

Components	Amount
2x Phusion master mix	10 µl
Fluidigm barcoded primer (2 µM)	4 µl
Multiplex PCR products	0.4 µl/multiplex product
H ₂ O	Up to 20 µl

Table 2-31 PCR components for barcoding PCR using the Access Array Barcode Library for Illumina Sequencers-384, Single Direction (Fluidigm)

Step	Temperature (°C)	Time	Cycles
Initial activation	98	2 min	1
Denaturation	98	10 s	
Annealing	60	30 s	6
Extension	72	1 min	
Final extension	72	10 min	1

Table 2-32 Thermal cycle of barcoding PCR using the Access Array Barcode Library for Illumina Sequencers-384, Single Direction (Fluidigm)

The products were run on gel electrophoresis and purified subsequently, to remove any impurities, and eluted in H₂O. The concentration was measured using Qubit 3.0 Fluorometer (Thermo Scientific). It was recommended to have a concentration of >10 nM or the minimum 5 nM in 5µl for the sequencing.

2.4.8. Nested PCR

Sample with limited amounts of available DNA were amplified using nested PCR method. Due to undergoing many cycles, high-fidelity polymerase (Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB)) was used. The first PCR reaction followed the protocol in Section 2.4.5.2, but the template was up to 5 µl and 30 PCR cycles used. Following the first PCR reaction, the product was diluted 100 fold in H₂O and used as the template for the second PCR. The pipetting scheme and thermal cycles are described in Table 2-33 and 2-34, respectively.

Components	Amount
2x Phusion master mix	12.5 µl
10 µM Primer mix	0.5 µl
Diluted PCR product	1 µl
H₂O	11 µl

Table 2-33 PCR components second reaction of nested PCR

Step	Temperature (°C)	Time	Cycles
Initial activation	98	2 min	1
Denaturation	98	10 s	
Annealing	Ta	30 s	20 – 35
Extension	72	30 s/kb	
Final extension	72	10 min	1

Table 2-34 Thermal cycle second PCR reaction of nested PCR

2.4.9. Quantitative PCR (qPCR)

Quantitative analysis of specific mRNA was performed using qPCR ABI Prism 7900HT Sequence Detection System (Applied Biosystems) or ViiA Real-Time PCR System (Applied Biosystems). SYBR Green (Platinum® SYBR® Green qPCR SuperMix-UDG with ROX, 2X, Invitrogen) which intercalates with double-stranded DNA and emits a fluorescent signal upon excitation, was used as the detection reagent. The primers were designed using Primer Express® Software v2.0 or v3.0 (Applied Biosystems). The reactions were performed in 384 well PCR plates, triplicates for each analysed gene, with components are described in Table 2-35.

Components	Amount _{general}	Amount _{MLL/AF4}
2X SYBR Green master mix	1X	1X
10 µM Primer mix	300 nM	50 nM
cDNA	2 µl	2 µl
H ₂ O	up to 10 µl	up to 10 µl

Table 2-35 qPCR components. Amount_{general} column is the common components for routine reaction. Amount_{MLL/AF4} column is exclusively used for MLL/AF4 L826 and SEM.

The component mix was prepared on ice, and the plate was sealed (MicroAmp® Optical Adhesive Film, Applied Biosystems) and centrifuged for 1 min at 1,000 g.

The thermal profile of the PCR is described in Table 2-36.

Temperature (°C)	Time	Cycles
50	2 min	1
95	10 min	
95	15 s	40
60	1 min	
95	15 s	1
60	15 s	
95	15 s	

Table 2-36 Thermal cycle of qPCR. The initial temperature setting, 50°C, activated the UDG to remove cross-contamination from uracil-containing DNA (e.g. previous PCR products). The reaction was inactivated at the next temperature setting, 95°C. This temperature also functioned to activate the polymerase. The second segment was the amplification reaction, where the signal was recorded during the extension step (i.e. 60°C). The third part was the amplicon dissociation phase where the melting curve was recorded.

The data were analysed using SDS 2.3 System Software (Applied Biosystems) or QuantStudio Real-Time PCR System (Applied Biosystems), and the comparative Ct method was used. Each gene of interest was normalised to the housekeeping gene (*GAPDH*), yielding Δ Ct value. This normalisation is required as the internal control for gene expression. Subsequently, the Δ Ct value was compared with reference Ct value (i.e. between reference control and experimental samples), resulting in $\Delta\Delta$ Ct value. Since Ct value was represented in a two-fold expression change (i.e. log2), the $\Delta\Delta$ Ct value was linearised by using antilog to the base of 2 to obtain the relative RNA expression.

2.4.10. Agarose gel electrophoresis

DNA fragments were visualised using agarose gel electrophoresis. The agarose (UltraPure™ Agarose; Invitrogen) was made in 1X TAE buffer with different concentration depending on the nucleic acid size, listed as follow:

Agarose concentration (w/v)	Nucleic acid size (bp)
1%	>1,000
2%	100 – 1,500
3%	<150

The mixture was heated using the microwave until all the agarose was dissolved. It was then swirled gently to mix the solution. GelRed (Biotium) was added to 1X concentration (stock is 10,000X), and the solution was poured into the constructed horizontal gel tray. The gel was left at room temperature for approximately 15 minutes to allow it to completely polymerise, before being transferred into the gel electrophoresis tank. The 1X TAE buffer was used as the running buffer. The DNA was mixed with 5X loading dye and loaded onto the gel. The electrophoresis was run at 3.5 – 4 V/cm.

The gel was visualised either by GelDoc (Biorad) or G:BOX (Syngene).

2.4.11. DNA polyacrylamide gel electrophoresis (DNA-PAGE)

Multiple DNA fragments that were shorter than 300 bp, e.g. multiplex PCR products, were visualised using DNA-PAGE to provide appropriate separation of the bands. The electrophoresis was performed on 8% PAGE gel in 1X TBE buffer as denoted in Table 2-37.

Composition	Volume	Final concentration
1 x TBE	8 ml	~1X
40% acrylamide/bis-acrylamide	2 ml	8%
10% ammonium persulfate	50 µl	0.05%
TEMED	10 µl	0.1%

Table 2-37 DNA-PAGE compositions

The gel was run on vertical electrophoresis apparatus using 1X TBE buffer as running medium. The sample was mixed with 5X loading dye. The electrophoresis was run at 50 V. Afterwards, the gel was stained with GelRed (Biotium) (Section 2.4.12) and visualised using Geldoc or G:BOX.

2.4.12. Staining polyacrylamide gel

The GelRed solution was used for staining the DNA-PAGE. 15 µl of GelRed (10,000X stock concentration) was added to 50 ml of 1X TBE. The gel was immersed in the solution and incubated at room temperature for 15 – 30 min with gentle shaking.

2.4.13. Whole genome amplification (WGA) single cells

DNA from single cell samples was amplified using REPLI-g Single Cell Kit (QIAGEN). The protocol was modified and adapted from Weizmann Institute. The sorting was

done as described in Section 2.3.2.3 and then followed with the amplification procedures. The WGA work was performed in a laminar-flow hood to reduce the risk of foreign DNA contaminations. Outside the laminar-flow hood, the plate was sealed with tape pad.

The WGA procedures consisted of lysis of the cell, DNA denaturation, terminating the denaturation, and the amplification itself. The plate was centrifuged at 2,000 rpm for 1 min, and the samples were added with 1.5 µl Buffer D2 to lyse the cells and denature the DNA. The reaction was incubated at 65°C for 10 min in a thermal cycler (HYBAID PCRExpress). Subsequently, the plate was transferred to the ice and centrifuged briefly. The reaction was terminated by adding 1.5 µl Stop Solution, centrifuged briefly, and stored on the ice. Meanwhile, the polymerase master mix was prepared with components described in Table 2-38.

Component	V/reaction (µl)
H ₂ O	4.5
REPLI-g sc Reaction Buffer	14.5
REPLI-g sc DNA Polymerase	1

Table 2-38 WGA polymerase master mix

The 20 µl polymerase master mix was added to each sample, centrifuged briefly, and incubated at 30°C for 8 h in the thermal cycler (without lid heating setting). Following the amplification, the reaction was heat-inactivated at 65°C for 3 min. A small volume of the amplified products was taken and diluted 100 fold with TE buffer for further applications. They were stored at -20°C.

2.4.14. Production of competent bacteria

Competent bacteria for cloning was made based on the Inoue method[194]. All of the containers, buffers, and solutions were prepared sterile.

The bacterial source was streaked on antibiotic-free agar plate and incubated for 16 – 20 h at 37°C. A single colony was picked and grown in 25 ml SOB medium in a 250 ml flask, incubated for 6 – 8 h at 37°C with vigorous shaking (250 rpm). This culture was used for inoculation into three different flasks containing 250 ml SOB media with different inoculated culture volumes 2 ml, 4 ml, and 10 ml, respectively. They were incubated at 18°C overnight with moderate shaking (200 rpm). On the following morning, the OD₆₀₀ of the cultures was read, continuously monitored every 45 min until one of the cultures reached 0.55 (*i.e.* the exponential phase). The culture was

transferred to an ice bath and incubated for 10 min (the others two cultures were discarded), and then the cells were harvested by centrifugation at 2,500 g for 10 min at 4°C. The medium was discarded and dried on a paper towel for 2 min.

Subsequently, the cells were resuspended in 80 ml ice-cold Inoue transformation buffer by gentle swirling. They were harvested by centrifugation at 2,500 g for 10 min at 4°C. The cells were resuspended in 20 ml ice-cold Inoue transformation buffer by gentle swirling. DMSO as the stabiliser of the frozen cells was added at a concentration of 7% (*i.e.* 1.5 ml) as the optimum concentration. The cell suspension was incubated on the ice for 10 min and aliquoted 200 µl/vial. They were cold shocked in liquid nitrogen to enhance the transformation efficiency and stored at -80°C afterwards.

2.4.15. Site-directed mutagenesis

Site-directed mutagenesis method was used for mutating genes of interest in a plasmid vector. KOD Hot Start DNA Polymerase (Merck Millipore) was used for the reaction. The process consists of annealing oligonucleotides that contain the mutated sequences to the plasmid DNA and amplifying it several times. The amplified sequences would then carry the mutation base pair(s).

The mutant oligonucleotides were designed with several characteristics, including 25 – 45 nt with melting temperature around 78°C, and the mutation located in the middle of the sequences. The reaction and thermal cycle are described in Table 2-39 and Table 2-40, respectively.

Component	V (µl)	Final conc. or amount
10X KOD buffer	2.5	1X
25 mM MgSO ₄	1.5	1.5 mM
2 mM dNTP mix	2.5	0.2 mM
10 µM primer forward	1	0.4 µM
10 µM primer reverse	1	0.4 µM
Plasmid DNA template (20 ng/µl)	1	20 ng
H ₂ O	15	
KOD polymerase (1 U/µl)	0.5	0.5 U or 0.02 U/µl

Table 2-39 Site-directed mutagenesis using KOD Hot Start DNA Polymerase (Merck Millipore)

Step	Temperature (°C)	Time	Cycles
Initial activation	95	15 min	1
Denaturation	95	30 s	
Annealing	55	30 s	12
Extension	72	1 min/kb	
Final extension	72	15 min	1

Table 2-40 Thermal cycle of site-directed mutagenesis using KOD Hot Start DNA Polymerase (Merck Millipore)

Following the amplification, the product was diluted by adding 25 µl H₂O. The parental DNA that did not contain the mutation was cleaved by the addition of 1 µl DpnI restriction enzyme (10 U/µl; the enzyme recognised methylated and hemi-methylated DNA that are present on the parental DNA but not on the amplification products) and incubated at 37°C for 1 h. The final product was transformed into XL1-Blue cells (Section 2.4.22).

Following the plasmid DNA extraction, the product was verified by Sanger sequencing.

2.4.16. Hybridisation and phosphorylation pair of oligonucleotides

The pairs of shRNA oligonucleotides (sense and antisense) were synthesised by Sigma-Aldrich. They were hybridised and phosphorylated on the 5' termini to create oligo duplexes and able to ligate into the vectors, respectively. The ATP-containing reaction buffer was used to provide the condition for phosphorylation. The reaction compositions are described in Table 2-41.

Component	V (µl)
Forward oligo (100 µM)	1
Reverse oligo (100 µM)	1
10X T4 DNA ligase buffer	1
H ₂ O	6.5
T4 Polynucleotide kinase (10 U/µl)	0.5 µl

Table 2-41 Hybridisation and phosphorylation reaction of oligonucleotides

Firstly, the mixture was incubated at 37°C for 30 min to allow the phosphorylation reaction. After that, it was transferred to the thermal block (Eppendorf Thermomixer® Comfort) that was set at 95°C and incubated for 4 min to denature the DNA and heat-

inactivate the kinase. Subsequently, the hybridisation was carried by ramping down the temperature 5°C/min to 70°C. When the thermal block reached 70°C, the reaction was left for 10 min, and continued the slow cooling to 22°C. The complete hybridisation was achieved by incubating the reaction at 22°C between 4 h and overnight. Finally, the annealed oligo duplex was diluted 1:200 with H₂O for further applications.

2.4.17. Ligation

Ligation reactions included ligating the DNA or gene of interest and oligo duplex into their appropriate vectors.

For ligating the DNA or gene of interest into the vector, a molar ratio from 1:3 to 1:5 was used. T4 DNA Ligase (5 U/μl; Thermo Scientific) was used for the reaction. The pipetting scheme is described in Table 2-42.

Component	Amount
Vector	20 ng
Insert	3 x or 5 x of vector molarity
10X T4 DNA ligase buffer	2 μl
T4 DNA ligase	1 U for sticky end ligation 5 U for blunt end ligation 4 U for sticky and blunt end ligation
H ₂ O	up to 20 μl

Table 2-42 Ligation reaction of DNA or gene of interest into the vector

For oligo duplex ligation, a different pipetting scheme was prepared, as described in Table 2-43.

Component	Amount
Vector	50 ng
Diluted oligo duplex (refers to Section 2.4.16)	1 μl
10X T4 DNA ligase buffer	1 μl
T4 DNA ligase	2.5 U
H ₂ O	up to 10 μl

Table 2-43 Ligation reaction of oligo duplex into the vector

The reactions were performed at room temperature for ≥ 3 h or overnight for both conditions. The final products were either directly transformed into bacteria (Section 2.4.22) or stored at -20°C .

2.4.18. Restriction endonuclease digestion

An appropriate amount of plasmids (300 – 1,000 ng) were digested using specific restriction endonucleases (Thermo Fisher Scientific). Digestions were performed by adding 2 – 3 U of the enzyme (stock is 10 U/ μl) in their particular buffers. H_2O was added to the reactions to make the volume 20 μl , followed by incubation at the appropriate temperature of the enzymes (mostly 37°C ; referred to the datasheet of the enzyme) for 1 h.

2.4.19. Fill-in/removal DNA overhangs

The fill-in of 5'-overhangs or removal of 3'-overhangs DNA was performed using Klenow Fragment (10 U/ μl ; Thermo Fisher Scientific). The enzyme works in various buffers, including restriction enzyme buffers. The DNA (500 – 1000 ng) in those buffers were supplemented with 0.5 μl of 2 mM each dNTP mix and 0.5 μl Klenow Fragment. The reaction was incubated for 10 min at 37°C and inactivated by heating at 75°C for 10 min.

2.4.20. DNA dephosphorylation

Vector plasmids with the blunt end on both or one of their termini were dephosphorylated to reduce the formation of re-ligation. The dephosphorylation was performed using FastAP Thermosensitive Alkaline Phosphatase (1 U/ μl ; Thermo Fisher Scientific). The enzyme worked in a large variety of restriction enzyme buffers. One microliter of the enzyme was added to up to 1 μg of the digested vector and incubated at 37°C for 10 min. The reaction was inactivated by heating at 75°C for 5 min.

2.4.21. Gateway recombination

DNA sequences in a vector containing attL sites (e.g. pDONR221 and pENTR1A) can be recombined to the destination vector that contains attR sites (e.g. pSIEW). The reaction was catalysed by Gateway® LR Clonase® II enzyme mix (Thermo Fisher Scientific). The pSIEW destination vector contains *ccdB* death cassette between the attR sites so that non-recombined vector clones do not grow.

The recombination reaction is described in Table 2-44.

Component	Amount
Entry clone	75 ng
Destination vector	75 ng
TE buffer	to 4 μ l
5X enzyme mix	1 μ l

Table 2-44 Gateway LR clonase reaction

They were incubated for at least 1 h at 25°C, and the reaction was stopped by adding 0.5 μ l Proteinase K, incubated at 37°C for 10 min. One microliter of the recombination reaction was used for transformation into Stbl3 (Section 2.4.22).

2.4.22. Bacterial transformation of plasmid ligation and mutagenesis

Plasmids from ligation and mutagenesis products were amplified by transformation into Stbl3 and XL-1 blue bacteria, respectively (for the genotype of the strains refer to Section 2.1.4). As many as 4 μ l ligation or 1 μ l mutagenesis product was placed in 1.5 ml microfuge on the ice, followed by transferring 40 – 50 μ l competent bacteria into the tube. They were incubated for 30 min on ice. Cells were heat shocked at 42°C for 40 s, allowing the uptake of the plasmid to the cells. They were re-incubated at 4°C at least for 2 min before adding 400 μ l SOC medium and continued with 37°C, 200 – 250 rpm incubation for 1 h. Subsequently, 100 – 200 μ l of the bacteria were plated on the LB agar plate which contained the plasmid-specific antibiotic. They were incubated at 37°C for 12 – 16 h. The bacterial colonies were either processed for further applications or stored at 4°C for up to one and half weeks.

2.4.23. Bacterial transformation using supercoiled DNA

Plasmid DNA that was isolated directly from bacteria (*i.e.* supercoiled DNA) was amplified by bacterial transformation. As many as 1 – 2 μ l that contain 300 – 1000 ng DNA was added into 25 μ l bacteria and directly plated onto an agar plate with the plasmid-specific antibiotic. They were incubated at 37°C for 12 – 16 h. The bacterial colonies were either processed for further applications or stored at 4°C for up to one and half weeks.

2.4.24. Miniprep plasmid isolation

QIAprep Miniprep Kit (QIAGEN) was used to isolate a small plasmid quantity (depending on the plasmids, ranging from 5 μ g to 25 μ g). The procedures consist of

alkaline lysis (NaOH/SDS) with the presence of RNase A, neutralisation and binding of the plasmid to the silica column, washing for removing endonucleases and salts, and elution of the DNA.

A single colony was inoculated in either 3 ml TB or 5 ml LB media with appropriate selection antibiotic and grown at 37°C with vigorous shaking (200 – 250 rpm) for 8 – 12 h or 12 – 18 h, respectively. The bacterial cells were harvested by centrifugation 3,000 g for 10 min at 4°C. They were resuspended in 250 µl resuspension buffer (Buffer P1) containing LyseBlue (0.1% v/v) and RNase A (100 µg/ml). After that, the cells were lysed by adding 250 µl the alkaline lysis buffer (Buffer P2), followed by mixing by gently inverting the tube six times and incubating at room temperature for 5 min. The lysis was terminated by adding 350 µl Buffer N3 and mixed by gently inverting the tube six times. The buffer also contained high-salt conditions that would precipitate denatured proteins, chromosomal DNA, cellular debris, and SDS, but allowed the plasmid DNA stayed soluble. The sample was centrifuged at 17,900 g for 10 min to separate the plasmid and the precipitations. The supernatant was transferred to the QIAprep 2.0 spin column and centrifuged at 17,900 g for 1 min to allow the DNA bind to the column. The sample was washed twice, firstly with 500 µl Buffer PB followed by 750 µl Buffer PE to remove the endonucleases and salts, respectively. The washings were done by centrifugation at 17,900 for 1 min. Residual washing buffer was removed by additional centrifugation 17,900 g for 1 min. After that, the plasmid was eluted by adding 50 µl Buffer EB, incubated for 1 min, and centrifuged at 17,900 g for 1 min. It was stored at 4°C or at -20°C for long-term storage.

2.4.25. *Maxiprep plasmid isolation*

EndoFree Plasmid Maxi kit (QIAGEN) was used to isolate larger plasmid quantity (up to 500 µg). The procedures consist of alkaline lysis (NaOH/SDS), neutralisation, endotoxin removal, binding of the plasmid DNA to the QIAGEN resin, removing the impurities, plasmid elution with high-salt buffer, and concentrated and desalted of the plasmid.

A single colony was inoculated in 5 ml LB media with appropriate selection antibiotic, grown as a starter culture by incubating at 37°C for approximately 8 h with vigorous shaking (225 – 300 rpm). Following the incubation, 500 µl of the culture was inoculated into 250 ml LB media with the appropriate selection antibiotic, grown at 37°C for 12 – 16 h with vigorous shaking. The bacterial cells were harvested by

centrifugation at 4,000 g for 15 min at 4°C. They were resuspended in 10 ml Buffer P1 that was supplemented with LyseBlue (0.1% v/v) and RNase A (100 µg/ml) and then lysed by adding 10 ml the alkaline lysis buffer (Buffer P2). It was mixed by gently inverting the tube six times or until the cell suspension turned homogeneously blue, and incubated for 5 min at room temperature. The reaction was terminated by adding 10 ml chilled Buffer P3, gently inverted six times or until the suspension turned homogeneously colourless. It was immediately transferred into QIAfilter Cartridge and incubated for 10 min at room temperature to float the precipitation containing proteins, genomic DNA, and detergent. The lysate containing plasmid was filtered through the cartridge, and 2.5 ml of the endotoxin removal reagent (Buffer ER) was added, mixed by inverting the tubes for ten times, and incubated on ice for 30 min.

QIAGEN-tip 500 was equilibrated by applying 10 ml Buffer QBT, left to drain completely by gravity flow. The lysate was applied to the QIAGEN-tip and allowed to empty by gravity flow. It was then washed twice with 30 ml Buffer QC to remove the contaminants. The plasmid was eluted by adding 15 ml high-salt buffer (Buffer QN), followed by the addition of 10.5 ml room-temperature isopropanol to the eluted DNA to concentrate and desalt the plasmid. The solution was centrifuged at 16,000 g for 30 min at 4°C. The supernatant was discarded, and the DNA pellet was rewashed with endotoxin-free 70% ethanol to remove any precipitated salt. It was centrifuged at 16,000 g for 10 min at 4°C. The supernatant was discarded, and the DNA pellet was dissolved in 400 – 500 µl endotoxin-free Buffer TE.

2.5. Proteomic methods

2.5.1. Protein isolation and quantification – western blotting

Protein was isolated from $10^6 - 10^7$ cells. The cells were washed once with PBS and lysed using RIPA buffer at a concentration of $30 \mu\text{l}/10^6$ cells. The lysis reaction was conducted on ice for 20 min with regular vortexing. Afterwards, the samples were centrifuged at 14,000 g for 10 min at 4°C to separate the cell debris. The supernatants contained proteins and transferred into a new vial.

For protein quantification, $2 \mu\text{l}$ of the protein was taken and diluted 10 fold with PBS. The diluted protein was quantified using BCA Protein Assay Kit (Santa Cruz Biotechnology) and read using plate reader FLUOstar Omega (BMG Labtech). The concentration was adjusted to 1 mg/ml with RIPA buffer. Before the immunoblotting, the protein was added with sample buffer (5X stock concentration) and denatured by heating at 100°C for 7 min.

2.5.2. Protein electrophoresis

The proteins that have been denatured and provided with a negative charge due to heat inactivation and SDS treatment in sample buffer, respectively (Section 2.5.1), were separated using SDS-polyacrylamide gel electrophoresis (SDS-PAGE). An appropriate concentration of separating gel was prepared based on the size of the protein of interest. The gel was poured into three-quarters of 1.5 mm casting chamber (BioRad), and 70% ethanol was pipetted onto the surface to set a flat surface of the gel. The gel was left until completely polymerised, and then the ethanol was discarded. After that, the stacking gel was poured on top of the separating gel surface. A comb was placed on the gel for providing the wells. After complete polymerisation, the gel was transferred into a vertical electrophoresis tank, and the appropriate amount of protein samples and ladder were loaded into the wells. The electrophoresis buffer was added into the tank, and the electrophoresis was run at 50 V until the samples passed through stacking gel, continuing with 150 V for the remainder of the electrophoresis.

2.5.3. Immunoblotting

The proteins from SDS-PAGE were transferred onto PVDF membrane (Millipore) and then detected by specific antibodies for the protein of interest. Before the assembly, the membrane was submerged into methanol for 30 s and then soaked in the blotting buffer for 5 min to equilibrate it. After the equilibration the SDS-PAGE gel and the membrane were assembled into blotting cassette, in order: the black side of cassette

holder, foam pad, filter paper, gel, membrane, filter paper, foam pad, and red side of the cassette holder. The transfer was performed at 100 V for 1 h in the blotting buffer. Following transfer, the membrane was blocked with 5% milk-TST in a gentle shaking for 60 min. Subsequently, the membrane was incubated with the primary antibody (the antibody concentrations are in Table 2-12 and Table 2-13) in 5% milk-TST on a rotating platform for 1 h at room temperature or overnight at 4°C. The membrane was washed three times with TST buffer for 10 min/washing. Following the washing, the membrane was incubated with the secondary antibody in 5% milk-TST on the rotating platform for 1 h at room temperature. It was rewashed three times with TST buffer to remove the unbound antibodies, followed with immunodetection.

2.5.4. Immunodetection

Immunodetection was visualised by the addition of luminol-based detection using Immobilon Western HRP Substrate (Millipore) on the membrane, followed by exposure to the X-ray film. Luminol and peroxidase solutions were mixed in a ratio of 1:1 and pipetted onto the membrane. It was incubated at room temperature for 5 min. After that, the reagent was drained off, and the membrane was placed on the X-ray film cassette. The protein was screened by exposing it to the imaging film (Kodak) for an appropriate time.

2.5.5. Co-immunoprecipitation (Co-IP)

The Co-IP was conducted from overexpression samples by transfecting 293T cells. The transfection was performed on two 100 mm tissue culture dishes and was collected 24 – 36 h afterwards. The cells were washed once with PBS, lysed by adding 500 µl Co-IP lysis buffer/plate (*i.e.* 1 ml/construct), and put on the ice. In addition to the buffer, the lysis reaction was also assisted by sonication twice for 10s at 6 A. The cell debris was separated by centrifugation at 13,000 g for 30 min at 4°C, and the supernatant containing protein was transferred into a new 2 ml microfuge tube. Meanwhile, 70 µl of the supernatant was taken, added with 17.5 µl of 5X sample buffer, and stored at -20°C as the total cell lysate sample (control). The supernatant was added with 40 µl diluted agarose G beads (dilution 50:50 agarose G beads stock with Co-IP lysis buffer) and optimised amount of the antibody (5 µg for twin-strep-tag antibody). The mixture was further added with 760 µl lysis buffer and incubated at 4°C for 3 h with slow rotation (approximately 4 rpm) to allow the binding of the antibodies to the beads and antigens. After that, the complex bead-antibody-antigen was pelleted by centrifugation at 3,000 rpm for 3 min at 4°C. The supernatant

was removed and washed three times. Each washing was done by adding 1000 μ l lysis buffer and centrifugation at 3,000 rpm for 3 min at 4°C. Subsequently, it was resuspended in 200 μ l Co-IP lysis buffer, mixed with 50 μ l of 5X sample buffer. The complex interaction and the protein structure were denatured by boiling at 100°C for 10 min and further analysed by western blotting (Section 2.5.2-4) or stored at -20°C.

Chapter 3 Pre-leukaemic populations of L826 and other MLLr leukaemia

Fundamental to understanding the process of lineage commitment in a lineage switch situation is to know what the origin of the relapse is. Despite the preliminary data presented in the introduction, it remains undetermined if the AML relapse directly changed from the ALL blasts or if there a common pre-leukaemic progenitor. If it is the latter, will it be possible to identify where in the differentiation pathway that common ancestor is located? And then, is it feasible to trace the candidate driver genes that are present in that population?

This chapter aims to understand:

1. the origin of the AML relapse L826
2. the pre-leukaemic population of L826
3. evaluate and compare the pre-leukaemic populations on other MLLr leukaemia cases.

Immunoglobulin rearrangement is commonly found in B-ALL cases, but not in AML. Looking at the rearrangement pattern between the ALL presentation and the AML relapse was the first approach in this identification study. This investigation would identify whether the common point of origin lies within a population which has already rearranged its B cell receptor, *i.e.* an early B cell progenitor, or is more primitive than that. Secondly, the presentation and relapse cells were sorted into different haematopoietic subpopulations. The pre-leukaemic mutation *MLL/AF4* was evaluated on each of the subpopulations to identify the location of the common clonal origin. Lastly, the cell of origin in MLLr infant ALL was further explored by studying the location of the transforming translocation in other MLLr cases.

3.1. Clonal V(D)J rearrangement in ALL and AML L826

Assessment on Ig and TCR gene rearrangements was commonly performed because of the principle that the cells of the lymphoid malignancy share a common clonal origin[195]. The Ig and TCR gene loci consist of a large number of different variable (V), diversity (D), and joining (J) gene segments that will undergo recombination during early lymphoid differentiation[195]. The Ig loci consist of the heavy chain (IGH) and the kappa (IGK) or lambda (IGL) light chain. The heavy chain gene segment rearrangement starts with D to J, followed with V to DJ recombination. On the other hand, the light chains consist of only V to J recombination. The rearrangement process follows an order started with (1) IGH recombination, and then (2) IGK recombination yields to IGH/k expression, or (3), if IGK rearrangement is unsuccessful, followed by IGK deletion and IGL rearrangement that yields IGH/λ expression[195]. Here, we performed the Ig rearrangement analyses, conducted by Dr Paul Evans, Leeds Institute of Oncology, to study the clonal similarity between the ALL presentation and the AML relapse. These analyses would show if the AML carried a common clonal origin at the B cell differentiation, *i.e.* at the Ig recombination level. This study was conducted based on the standard protocol BIOMED-2[195].

The VH (variable domain of heavy chain) consists of 46 – 52 functional segments, depending on the individual haplotypes and divided into 6 to 7 subgroups (VH1, VH2, VH3, VH4, VH5, VH6, and VH7). These subgroups can be detected by three different multiplex amplifications where different primers are located in the region called framework region (FR) as the forward primers and on a single JH as the reverse primer. These multiplex amplifications are named as FR1, FR2, and FR3, Figure 3-1a and b. Assessment of this rearrangement can determine whether the cells have the complete IGH gene rearrangement.

The DH consists of 27 functional segments and is divided into 7 families (DH1, DH2, DH3, DH4, DH5, DH6, and DH7), Figure 3-1a and c. Similar to the VH analysis, multiplex amplification, named Tube D, is used to detect DH1 to DH6 and a single amplification, called Tube E, to assess the DH7 if these regions are included in the clonal rearrangement. This evaluation can determine if the cells have incomplete IGH gene rearrangement.

In addition to the VH-JH and the DH-JH assessments, we also performed the immunoglobulin light chain IGK analysis. The IGK variable region (*i.e.* VK-JK) consists of 5 JK gene segments and many distinct VK gene segments which are

grouped into 7 VK families. Multiplex primers were designed to cover these gene segments and concluded into two different reactions, named Tube A and Tube B.

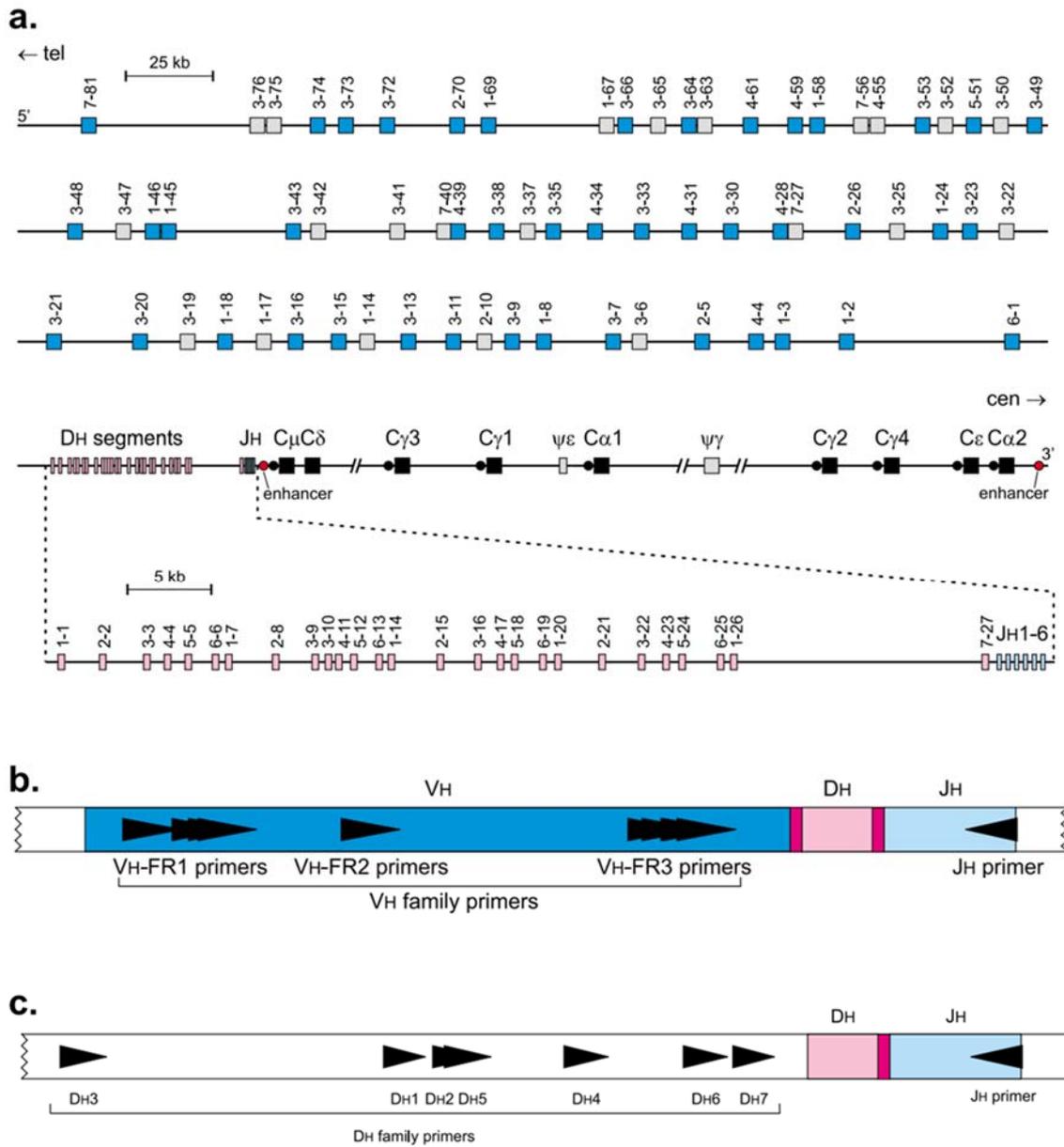


Figure 3-1 The schematic diagrams of IGH gene complex and the primer locations. a) IGH gene complex consists of VH, DH, JH segments and the constant (or C) region. The VH segment is composed of the functional (blue) and rearrangeable pseudogenes (grey). The two digits number on VH and DH segments represent the subgroup and the member, respectively. b) Schematic diagram of IGH VH-JH rearrangement indicates the location of the three framework regions (FR1, FR2, FR3). Each of the FR regions consists of multiplex forward primers. The single reverse primer is located and indicated as in 'JH primer'. c) Schematic diagram of IGH DH-JH rearrangement shows the location of the seven DH subgroups and the consensus JH reverse primer. DH7 primer (Tube E) is separated from the other six family primers (Tube D). The relative position of the DH and JH primers is given according to their most 5' nucleotide upstream or downstream of the involved recombination signal sequences. The schemes are adapted from BIOMED-2 guidelines[195].

We analysed whether the ALL presentation carried the clonal rearrangement and if that rearrangement was also present in the AML relapse. The results showed an oligoclonal pattern of rearrangements in the ALL, but no rearrangement on the AML depicted on Figure 3-2.

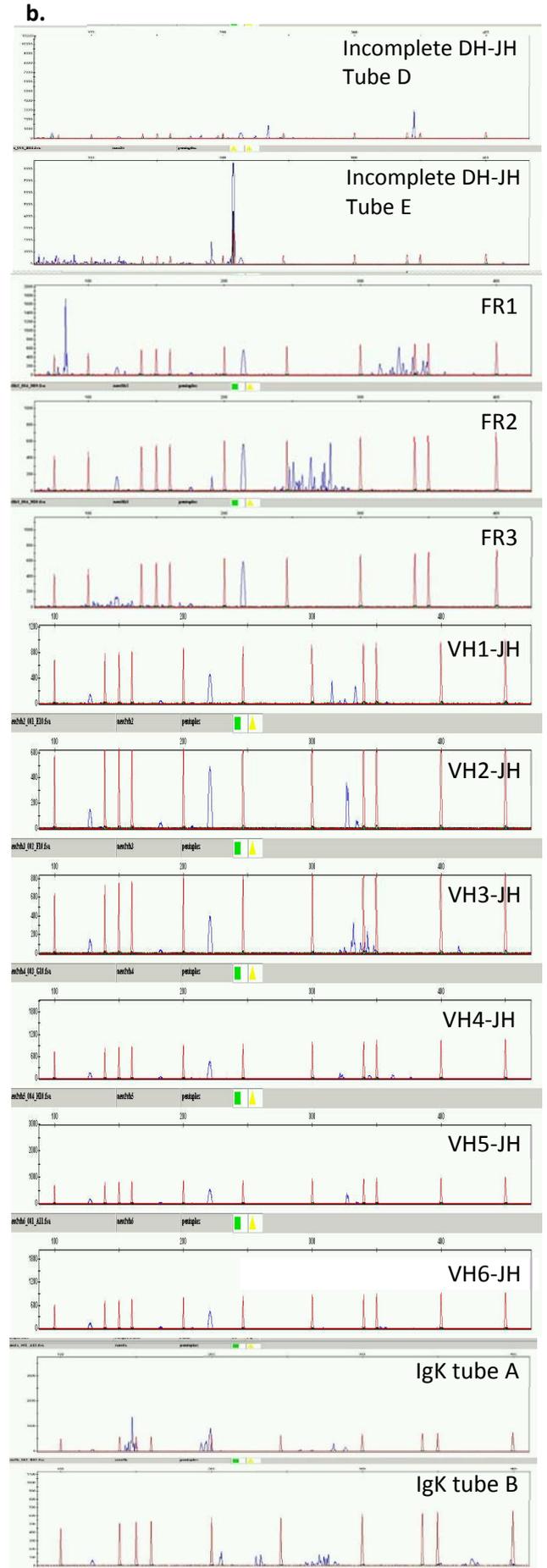
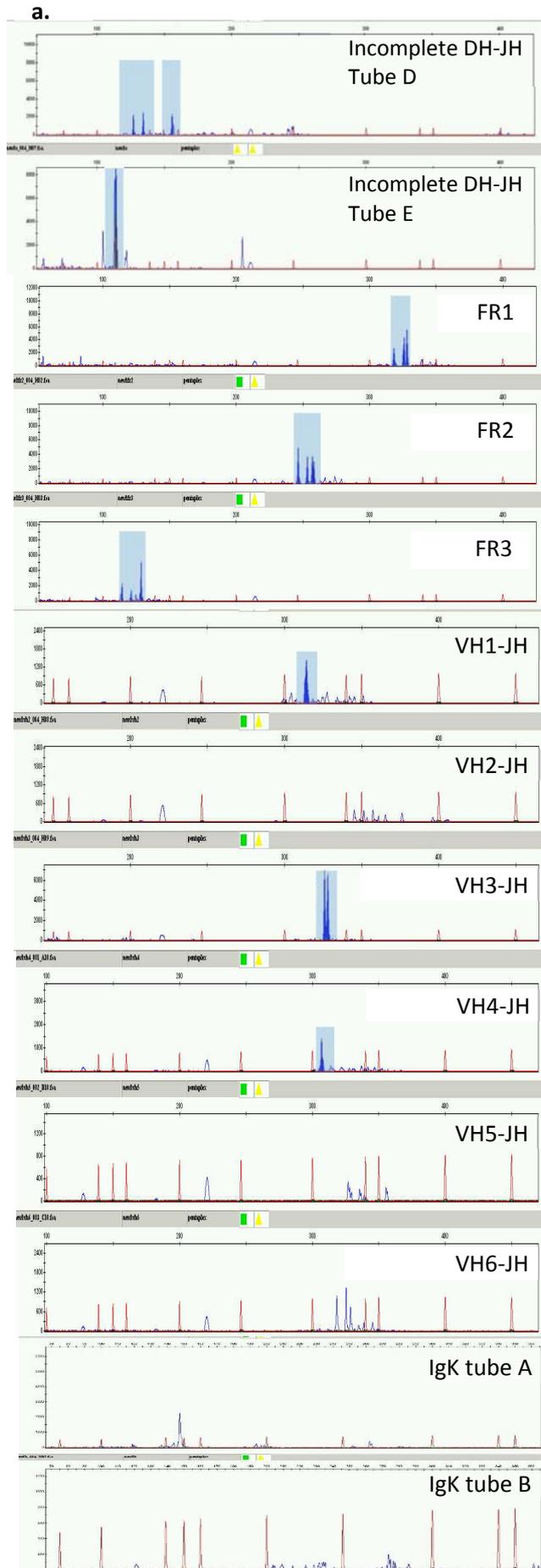


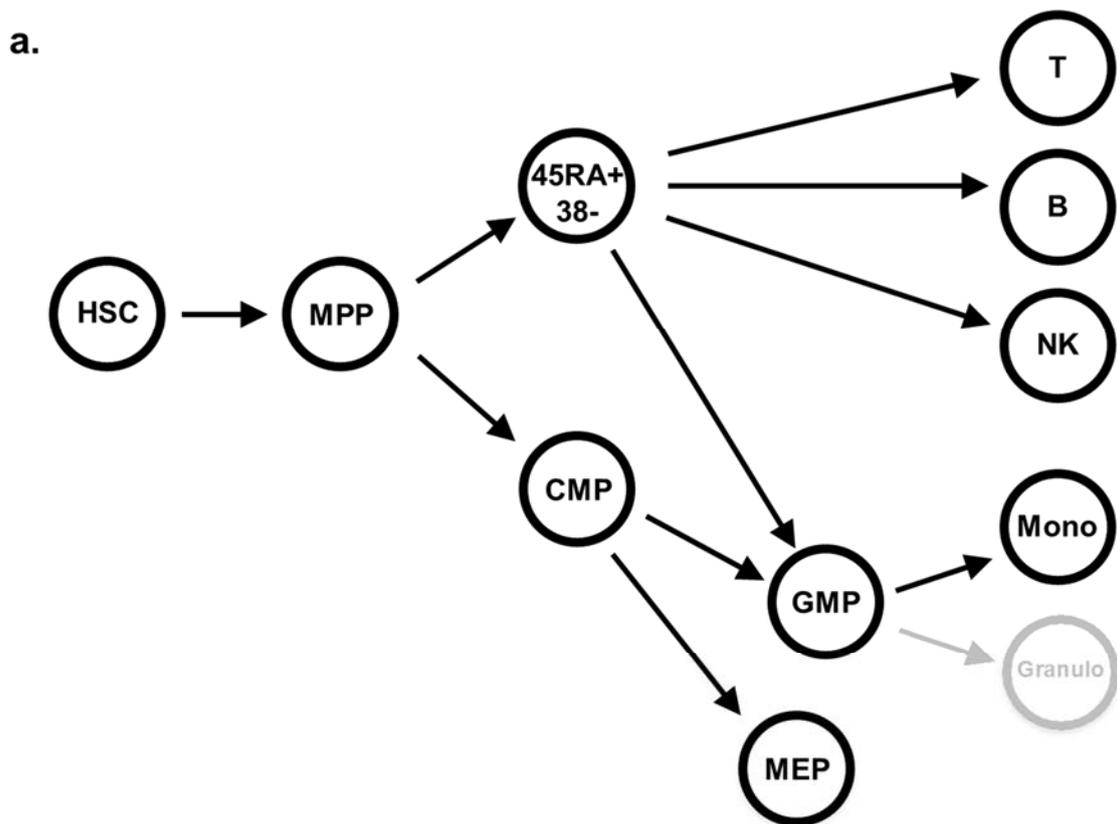
Figure 3-2 Immunoglobulin rearrangement on presentation (a) and relapse (b) L826, referring to the standard protocol BIOMED-2[195]. Incomplete DH-JH tube D consisted of DH1, DH2, DH3, DH4, DH5, and DH6 segment primer sets, while DH-JH tube E included DH7 segment primer set. Tube E also detected the germline DH-JH band at location 221, as shown on the relapse graph (b; incomplete DH-JH tube E). The electropherogram showed the incomplete DH-JH peaks in the presentation, but not in the relapse. Similarly, multiplex amplification FR1, FR2, and FR3 of the VH segments showed the clonal products only in the presentation. The result was confirmed by single amplification using VH1-JH, VH2-JH, VH3-JH, VH4-JH, VH5-JH, and VH6-JH primer sets, which further showed two clonal VH3, one clonal VH1, and one clonal VH4 rearrangement. Moreover, no IgK recombination was observed in both samples. The amplification signal is denoted by blue signal and box, and the marker signal by red. The experiment was performed by Dr Paul Evans, Leeds Institute of Oncology.

The incomplete DH-JH rearrangement was assessed by the amplifications on Tube D and Tube E. The incomplete rearrangement signals were seen in the presentation, but no clonal rearrangement on the relapse (Figure 3-2 blue box on the presentation Tube D and E, but not on the relapse column). Furthermore, the VH-JH recombination was assessed by multiplex FR1, FR2, and FR3 VH segments. Similarly, the prominent clonal peaks (blue box in the Figure 3-2) were observed only in the presentation. This analysis was further explored by single amplification of the segment VH1, VH2, VH3, VH4, VH5, and VH6, which showed two VH3, one VH1, and one VH4 rearrangement. These recombinations mean that the B cells of L826 presentation had undergone B cell receptor heavy chain constructions, although had not been completed with light chain partner. These data demonstrate an oligoclonal (>3 clones) pattern of rearrangement in the ALL which is a feature typically found in the MLLr ALL due to the continued expression of *TdT* and *RAG1* genes[195, 196]. Because the clonal patterns were not identified in the AML, it indicated that the relapse did not arise directly from the ALL blasts, but that any common cell of origin must originate within a precursor population more primitive than those undergoing early V(D)J rearrangement.

3.2. Pre-leukaemic populations L826 presentation and relapse

The V(D)J clonal analysis indicated that the AML arose from a common cell that was primitive to early V(D)J rearrangement. The next step was to identify where the origin of the relapse is within the haematopoietic stem/progenitor cells.

Our method to address this was by taking advantage of the fusion gene *MLL/AF4*. The L826 presentation and relapse bone marrow cells were separated into the differential haematopoietic subpopulations, including HSC, MPP, CD38-CD45RA+, CMP, GMP, MEP, T cells, B cells, NK cells, and monocytes, and then the presence of *MLL/AF4* was evaluated on each of the subpopulations. These major subpopulations and their specific markers are shown in Figure 3-3.



b.

Haematopoietic subset	Markers
HSC	CD34+ CD38- CD90+ CD45RA-
MPP	CD34+ CD38- CD90- CD45RA-
CD38-CD45RA+	CD34+ CD38- CD45RA+
MLP	CD34+ CD38- CD90- CD45RA+ CD10+
LMPP	CD34+ CD38- CD90- CD45RA+ CD10-
GMP	CD34+ CD38+ CD10- CD123+ CD45RA+
CMP	CD34+ CD38+ CD10- CD123+ CD45RA-
MEP	CD34+ CD38+ CD10- CD123- CD45RA-
T cells	CD3+
B cells	HLA DR+ CD19+
NK cells	CD3- CD56+
Monocytes	CD45+ HLA DR+ CD14+ CD16-

Figure 3-3 Haematopoiesis scheme (a) and the markers of the subsets (b). The scheme shows the construction of the blood cell subsets. It consists of the progenitor cells (HSC, MPP, CD38-CD45RA+, CMP, GMP, and MEP), lymphocytes (T cells, B cells, and NK cells), and monocytes. The CD38-CD45RA+ population is a mixed population of MLP and LMPP since both populations alone have too few cells for analysis (Figure 3-4). The progenitor cells are indicated by the positivity of CD34, while the more mature cells do not express CD34.

The HSC and progenitor cells are characterised by the expression of CD34. While HSC is further defined by the presence of cell surface antigens CD34+CD38-CD45RA-CD90+, the MPP has lost the CD90 expression, *i.e.* it is described as CD34+CD38-CD45RA-CD90-. Both HSC and MPP are CD38^{low/-}, and this marker expression gradually increases upon more mature differentiation (refers to Figure 1-1)[2, 197].

Based on the markers listed in Figure 3-3b, the cells were stained and sorted using FACS (Section 2.3.2.2). The DNA was isolated from each subset, and the *MLL/AF4* was evaluated by nested PCR. *ATP10A* gene detection was used as a positive control to show each sample had sufficient template amount for the amplification. The distribution of the haematopoietic subpopulations and the agarose gel electrophoresis of the PCR products are shown in Figure 3-4.

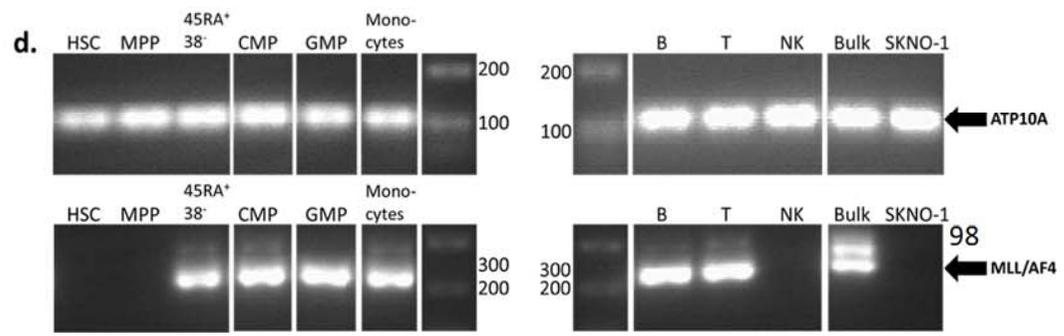
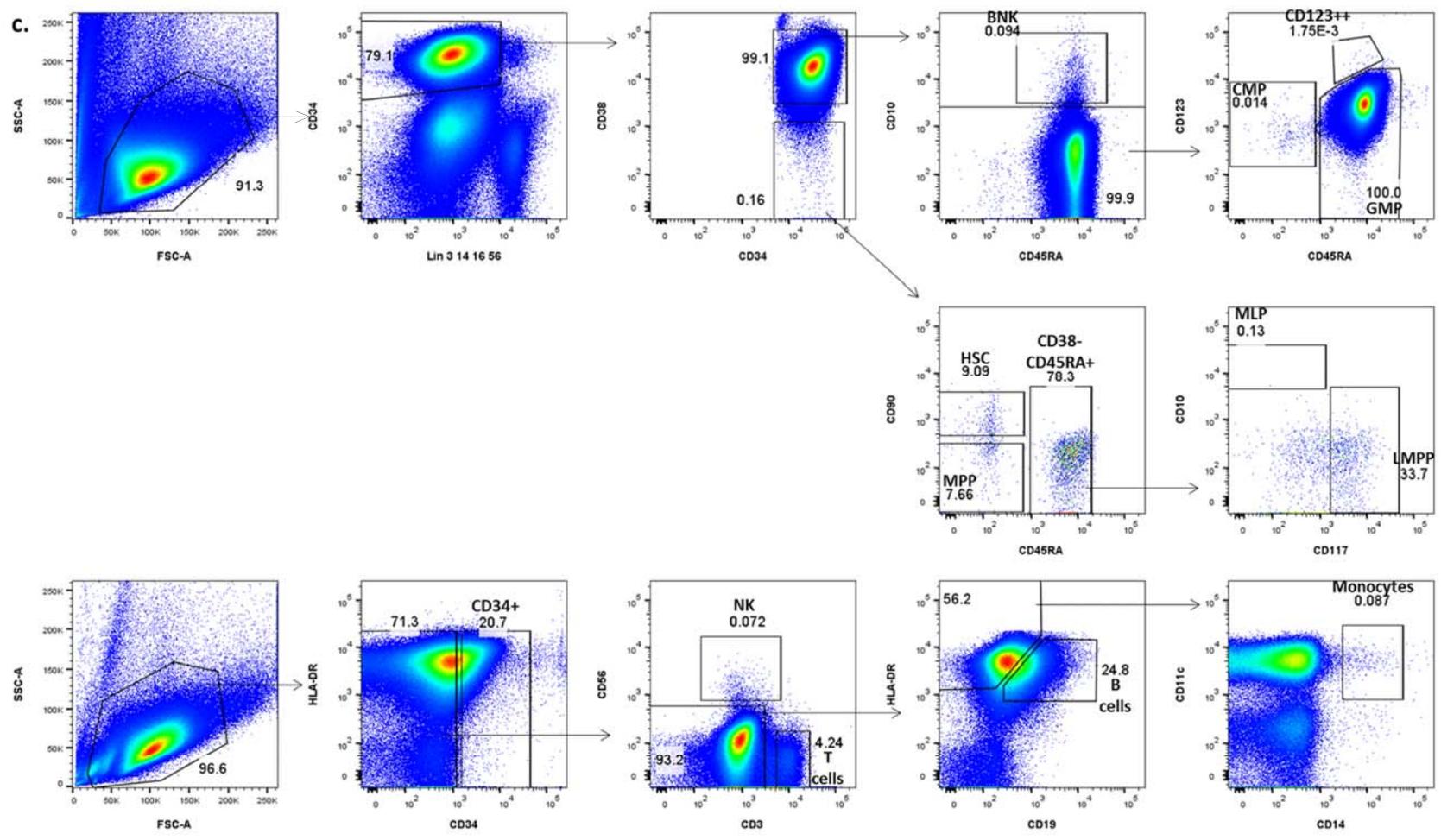


Figure 3-4 Distribution of haematopoietic populations and evaluation of *MLL/AF4* on L826 presentation (a, b) and relapse (c, d). The haematopoietic subset analyses were divided into two panels: progenitor (a, c, top panels) and lymphocytes/monocytes (a, c, bottom panels) populations. The distinct progenitor subpopulations of HSC, MPP, CMP, GMP, and LMPP were seen. On the lymphocytes/monocyte panel, the flow analysis showed B cells, T cells, NK, and monocytes subpopulations. Following the cell sorting, too few cells were obtained from MLP (CD34+CD38-CD90-CD45RA+CD10+) and LMPP (CD34+CD38-CD90-CD45RA+CD10-) for the analysis; therefore these populations were represented by CD38-CD45RA+ population (mixed of MLP and LMPP; CD34+CD38-CD90-CD45RA+) that were also known for differentiation potential into GMP and lymphoid lineage[6]. Each population was evaluated for the presence of the pre-leukaemic mutation *MLL/AF4* by PCR. The gel electrophoresis was shown on the presentation (b) and relapse (d). *ATP10A* was used as positive control PCR to show each population had sufficient template for the amplification. SKNO-1 is a t(8;21) AML cell line as a negative control of *MLL/AF4*. *ATP10* amplicon size is 100 bp. *MLL/AF4* amplicon size is 208 bp. All of the samples were analysed from the same experiment, but the locations were rearranged for these figures. Original images are shown in Appendix A. Flow axis Lin 3 14 16 56 represented combination Lin CD3 CD14 CD16 CD56 antibodies that were used for selecting the lineage negative progenitor cells.

Haematopoietic stem and progenitor populations included HSC, MPP, CD38-CD45RA+, GMP, and CMP were observed at both presentation and relapse L826. These populations were relatively conserved between the two states although the separation of MPP and HSC from CD38-CD45RA+ was much sharper in the AML case. The flow data showed that GMP population was the major CD34+ population. Considering the differentiation potential of GMP into myeloid fate (granulocytes and monocytes), it was intriguing to see the large population within the ALL. The staining for lymphocytes/monocytes showed that both samples also had the B cells, T cells, NK, and monocytes subsets. However, the analysis excluded MLP and LMPP due to too low cell numbers (less than 200 cells), but the population was represented by CD38-CD45RA+ that is similar to the LMPP, known to have the potential to differentiate into both lymphoid and myeloid lineages[6].

The evaluation of *MLL/AF4* within each population was performed three times (separate PCRs from the same sample) in order to review the consistency of the data. Figure 3-4b,d showed one of the three replicates. Consistent results were achieved in the fusion gene negativity within the HSC population, and positivity within the CD38-CD45RA+, CMP, and GMP at both L826 presentation and relapse. A similarly consistent finding was also obtained for the presence of the mutation within the MPP at presentation. However, it was not apparent in the MPP L826 relapse.

Among the three replicates, there was one replicate that showed the positivity, but not the others. It was difficult to conclude if it was because of the insensitivity of the assay due to the minimum number of pre-leukaemic cells or it was a contamination. However, by looking at the positive fusion gene result within the CMP and CD38-CD45RA+ at relapse, it is likely that there are pre-leukaemic cells within the MPP population that were not detected in this current assay. In order to provide more conclusive results, the analysis was extended by evaluating the secondary mutations that accompanied the *MLL/AF4*, described in Chapter 4.3.

Secondly, the data were further analysed by looking at the expression of *MLL/AF4* by reverse transcriptase nested PCR method. A similar result was obtained, where the *MLL/AF4* fusion gene was identified in MPP and CD38-CD45RA+/CMP in both the presentation and relapse, summarised in Figure 3-5. It is interesting because both CD38-CD45RA+ and MPP have the potential to differentiate into lymphoid and myeloid lineages. This result suggests the origin of the AML relapse may come from either one of these populations.

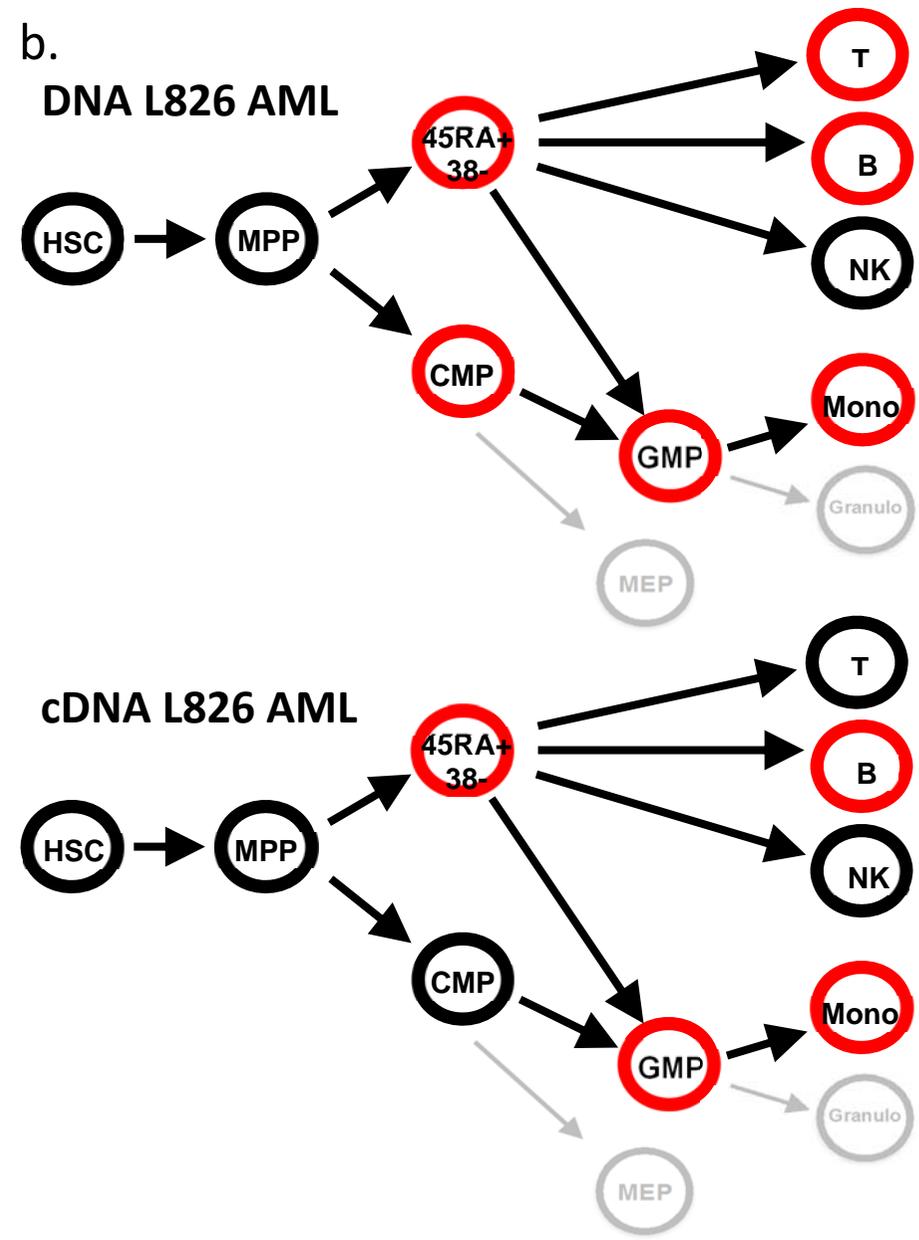
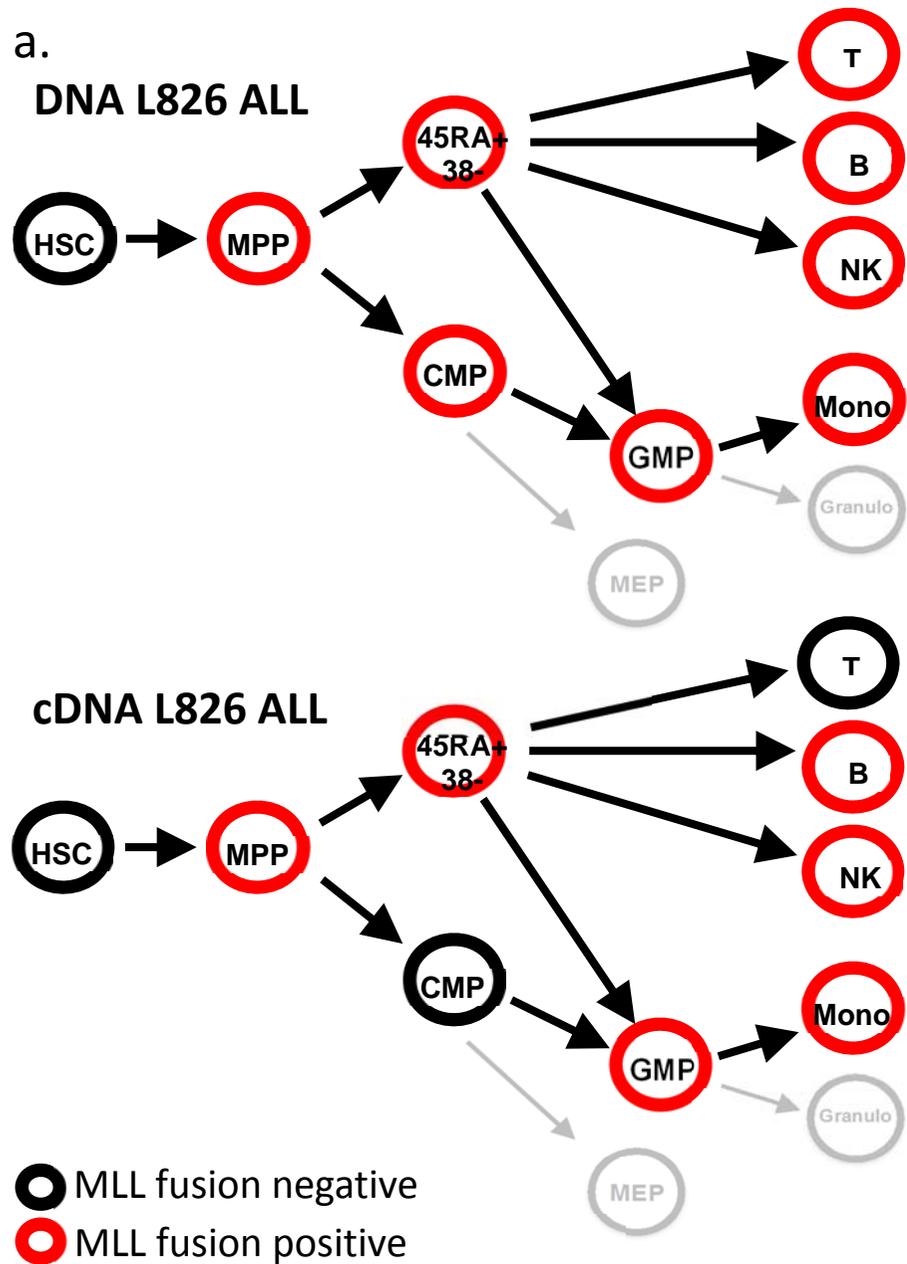


Figure 3-5 Progenitor leukaemic populations in L826 presentation (a) and relapse (b). The progenitor leukaemic populations were determined by PCR *MLL/AF4* on the DNA (top graph) and the cDNA (bottom graph). (a) *MLL/AF4* positivity was present starting from the MPP subpopulation on the ALL. The mutation was also expressed in the MPP. Additionally, the data showed that although the mutation was observed in CMP and T cell populations, it was not shown on the transcripts of these populations. (b) *MLL/AF4* was identified on the CD38-CD45RA+ onwards on the AML. The mutation also showed on the CMP, but not on the MPP (although one out of three PCR replications showed the positivity of *MLL/AF4* in MPP which may indicate a sensitivity issue). The reason remained unclear, but might be due to few pre-leukaemic cells on the MPP. Similarly, CMP and T cell populations also had the mutation on the DNA analysis, but not expressed as transcript. The AML also showed negativity *MLL/AF4* on the NK population.

Lastly, these data also showed that the positivity of *MLL/AF4* on CMP and T cells on the DNA, but not on the transcript. Interestingly, *MLL* wt analysis on another *MLL/AF4* patient case, LK228, showed that the wild-type gene was expressed on all of the subpopulations, including the CMP and T cells (n = 2; Section 3.5.3, Figure 3-11; both *MLL* fusion and *MLL* wt primers have been optimised to have a high sensitivity prior to the analysis, although *MLL* wt was in favour, probably because *MLL* fusion primer was restricted to a specific region within a certain breakpoint amplicon). It remains unknown how this differently regulated *MLL* wt and *MLLr* expression can occur on these different subsets, and if this phenomenon has any effect on the leukaemia propagation.

3.3. Pre-leukaemic populations on the primograft L826

Having identified the L826 pre-leukaemic population, the study was continued by evaluating the development of the leukaemia within primograft patient-derived xenograft models (primary transplant). The ALL sample was transplanted into NSG mice (non-irradiated) via intrafemoral injection under isoflurane anaesthetic with analgesia Carprofen 5 mg/kg subcutaneous injection. The engraftments were present on spleen of the mice, collected when the mice became ill (weight loss, pale skin, piloerection, and an enlarged spleen that can be seen through the skin) after approximately 4 months after the injection. The transplantation and leukaemic cells collection were conducted by Dr Helen Blair and Dr Alex Elder.

The harvested cells were stained for different haematopoietic markers to see the distribution of the haematopoietic subsets. The analysis of the populations showed the blast cells as the majority population, but also showed the presence of GMP and a small proportion of HSC, as shown in Figure 3-6a. These populations were collected, and the *MLL/AF4* was evaluated. Interestingly, the translocation was found within all of the collected populations including the HSC, shown in Figure 3-6b. This is in contrast to the primary ALL cells analysis (Figure 3-4b), in which the HSC population were not found to be *MLL/AF4* positive.

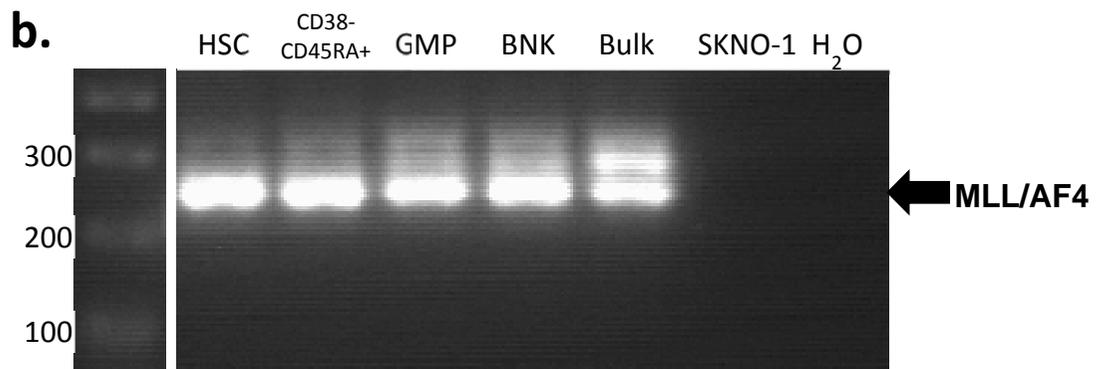
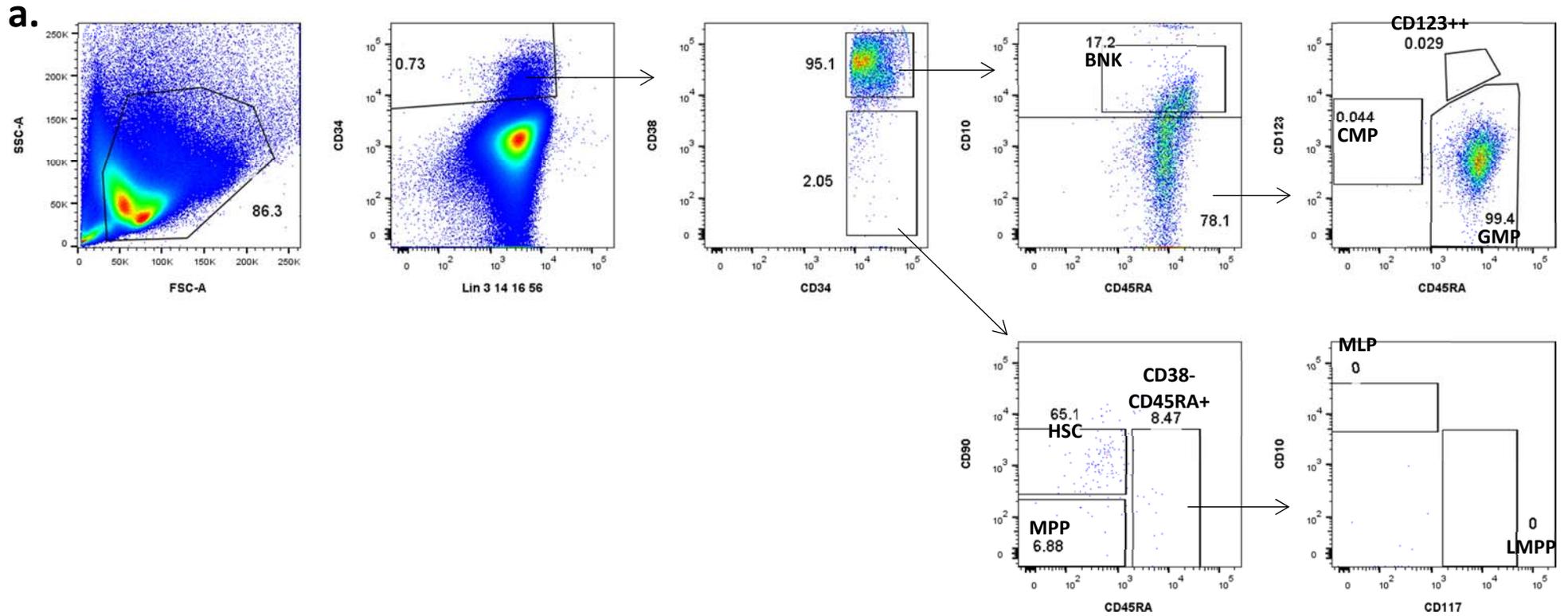


Figure 3-6 L826 ALL primograft progenitor population distribution (a) and the fusion gene (b).

a) Flow analysis on the primograft sample showed the GMP population as the majority of the progenitor population. In contrast to the primary cells, the CMP population was not seen on the primograft. Also, there was a substantial reduction in the proportion of MPP and CD38-CD45RA+ populations. Interestingly, the HSC population seemed enriched on the primograft. b) The HSC, CD38-CD45RA+, GMP, and BNK populations were collected. The pre-leukaemic mutation *MLL/AF4* was evaluated in these populations by PCR, and it was shown that all of them contained the mutation. Bulk: bulk, unsorted population; SKNO-1: an AML cell line, not carrying the *MLL/AF4* sequences.

The flow data demonstrated the distribution of different haematopoietic populations between the primary and primograft cells. It showed a substantial reduction in the MPP, CMP, and CD38-CD45RA+ populations. Interestingly, the GMP that dominated the progenitor population on the primary cells was also the major population on the primograft. By considering the potential of GMP to differentiate into granulocytes and monocytes, it remained unknown if the pre-leukaemic GMP is related to the characteristics of MLLr leukaemia with its occasional co-expression of both lymphoid and myeloid markers.

Another finding in this study was the enrichment of the HSC population among the CD34+CD38- (HSC, MPP, and CD38-CD45RA+) subpopulations. While the HSC was the minor of the three subpopulations in the primary cells, it became the only subpopulation in the primograft. It raised the question of whether there was pre-leukaemic potential in the HSC, which was enriched by the xenograft environment. Therefore, the *MLL/AF4* was evaluated in this population, together with the CD38-CD45RA+, GMP, and BNK subsets. Intriguingly, despite the fact that no fusion gene had been identified in the HSC primary cells, the mutation was found within the primograft HSC compartment (Figure 3-6b). This might indicate (1) an enrichment of the HSC pre-leukaemic population by the primograft environment and (2) a possibility of a minimal number of HSC pre-leukaemic cells on the primary cells that were not able to be detected by the assay.

To further validate the HSC enrichment, flow analysis on more L826 ALL primograft, including the primary, secondary, and quaternary transplant samples were performed. The secondary and quaternary transplant were injected from 10^6 leukaemic cells from the spleen of the previous transplant generation. In addition, in this case, each population was also gated for CD19- to prevent recruiting the leukaemic/blast/CD19+ cells, Figure 3-7.

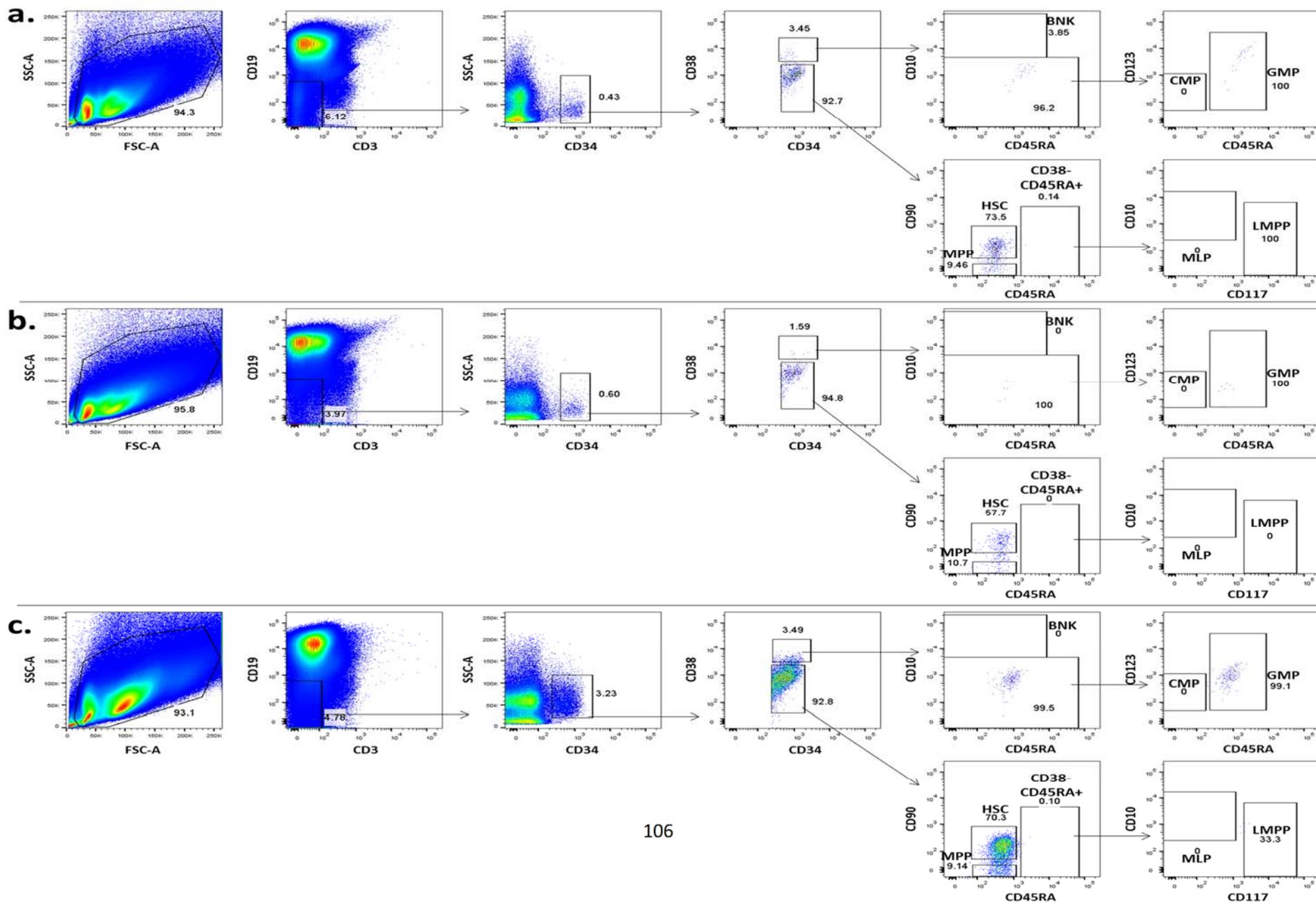


Figure 3-7 The distribution of haematopoietic subpopulations L826 primograft on the primary (a), secondary (b), and quaternary (c) transplants. The progenitor populations were isolated for CD19- expression. A distinct HSC and small GMP subset was observed on all of the samples. The other populations were markedly depleted.

The three different L826 primograft samples showed the HSC as the majority of the population with percentages: 0.018% (513 HSC/2,884,223 total live cells), 0.013% (275 HSC/2,093,476 total live cells), and 0.101% (4,058 HSC/4,026,989 total cells) on the primary, secondary, and quaternary transplants, respectively (it must be underlined that these are rough proportions because the antibodies are specific to human, but the number of total live cells includes the mouse cells, too). A small GMP subset was the only other precursor population identified. These data indicated the enrichment of HSC population in the xenograft.

This result was further confirmed by the presence of *MLL/AF4* sequences on the gDNA of the isolated HSC from the quaternary sample primograft L826 (Figure 3-8).

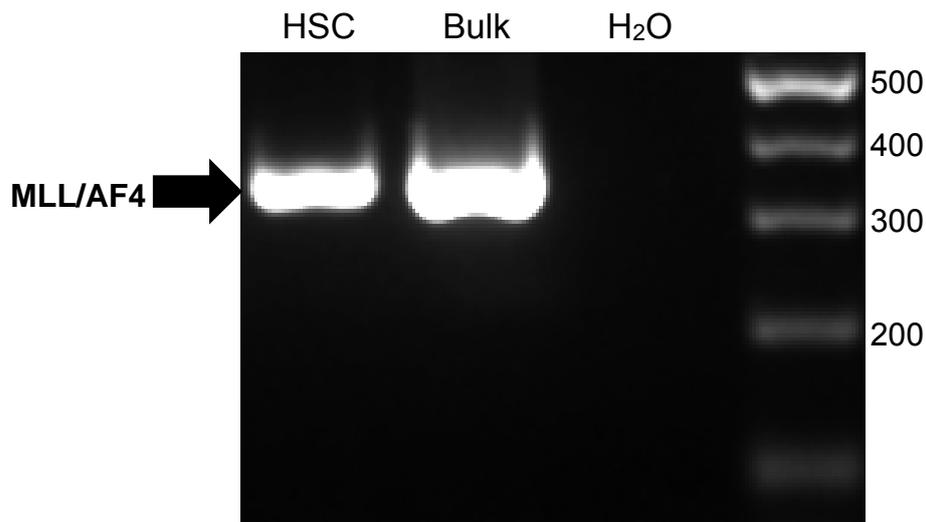


Figure 3-8 *MLL/AF4* evaluation on HSC primograft L826 quaternary transplant. *MLL/AF4* positivity was observed in the isolated DNA from quaternary transplant mouse L826 presentation. The sample was obtained from the spleen. HSC was the only progenitor population able to be collected (the GMP was too few cells). Amplicon size: 343 bp.

3.4. Pre-leukaemic populations in other primografts, t(4;11) and t(9;11)

Having shown the potential of pre-leukaemic population maintenance on L826 primograft, the experimental approach was extended to other MLLr leukaemia cases. Therefore, the same flow analysis was performed on two other ALL samples including a secondary transplant t(4;11) and a primary transplant t(9;11) sample, named P929, and A789, respectively. This analysis was also gated for the CD34+CD19- progenitor population. It showed the presence of a major progenitor population, but in the CD38-CD45RA+ population, rather than in the HSC, as illustrated in Figure 3-9.

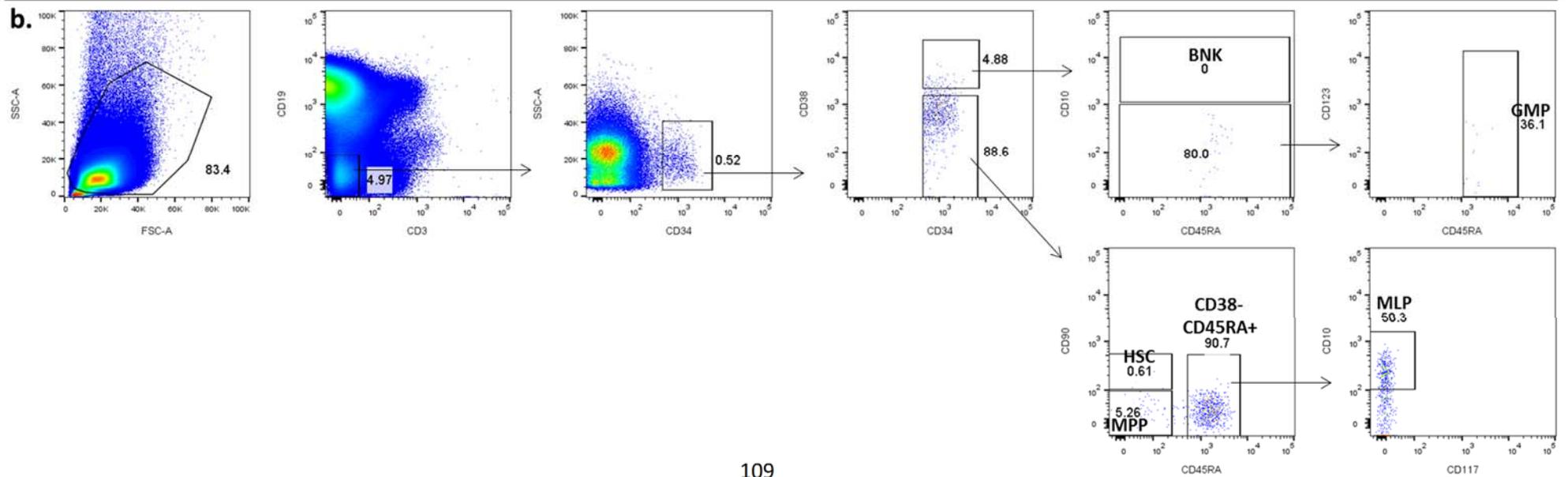
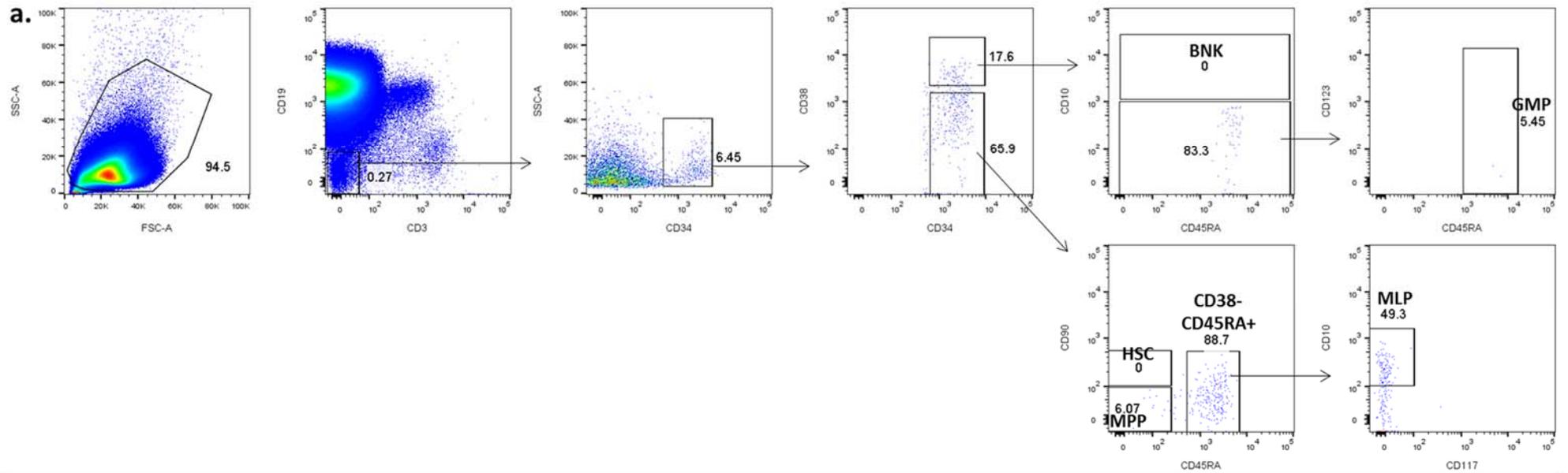


Figure 3-9 The distribution of haematopoietic subpopulations on the secondary transplant t(4;11) patient P929 (a) and primary transplant t(9;11) patient A789 (b) primograft. Patient P929 and patient A789 were diagnosed with ALL at the age of 3.6 months and 22 months, respectively. These primograft samples were isolated from the mouse spleen engraftments. Both samples were selected for CD34+CD19- expression. CD38-CD45RA+ population appeared as the only distinct progenitor population. Although the data showed the MLP (CD34+CD38-CD90-CD45RA+CD10+) subset, this population was not present as a distinct CD10+ population, but rather a continuum from CD10+ to CD10-, which made it unclear if that was a true MLP population or an aberrant leukaemic population.

In contrast to the primograft L826 that showed the HSC population, P929 and A789 showed the CD38-CD45RA+ as the majority progenitor population. Of note, the P929 sample came from a secondary transplant mouse which might suggest that the population was maintained, rather than carried over from the primary cells. It remains unanswered regarding (1) the reason of different subpopulation maintenance between L826 and P929 and A789, and (2) the meaning of HSC or CD38-CD45RA+ presence within the primograft cells.

3.5. Pre-leukaemic population in other MLLr iALL

The current data on L826 showed the presence of MLL fusion gene within progenitor populations as early as MPP during the presentation. This study was extended to examine the pattern of fusion gene presence in other MLLr iALL.

3.5.1. Clinical characteristics MLLr iALL

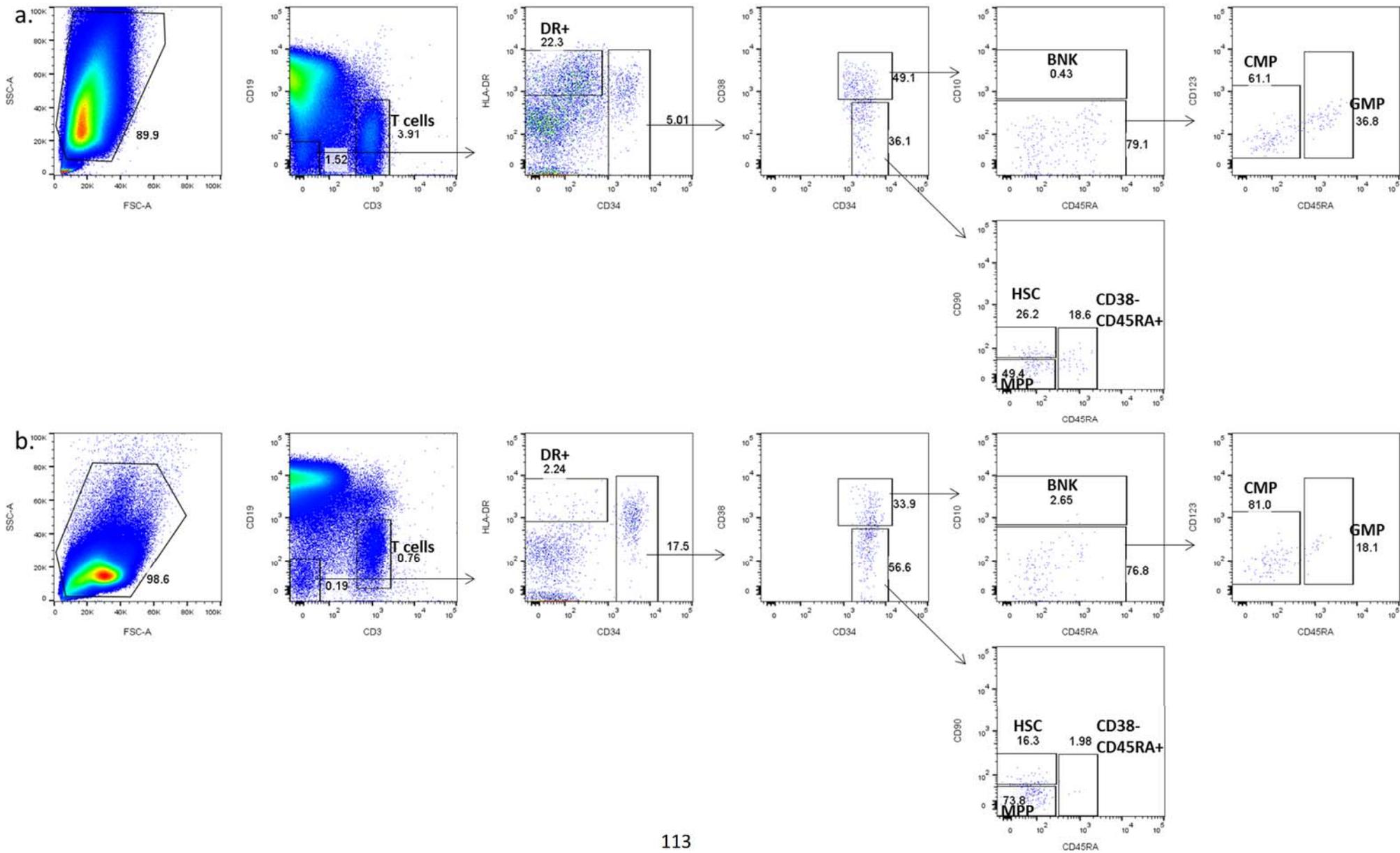
Four t(4;11) and two t(9;11) ALL cases were included in the study. The details of the patients were summarised in Table 3-1. The analyses were performed on the peripheral blood (L876, L880, and LK124) or bone marrow (LK228, LK230, and LK271) samples.

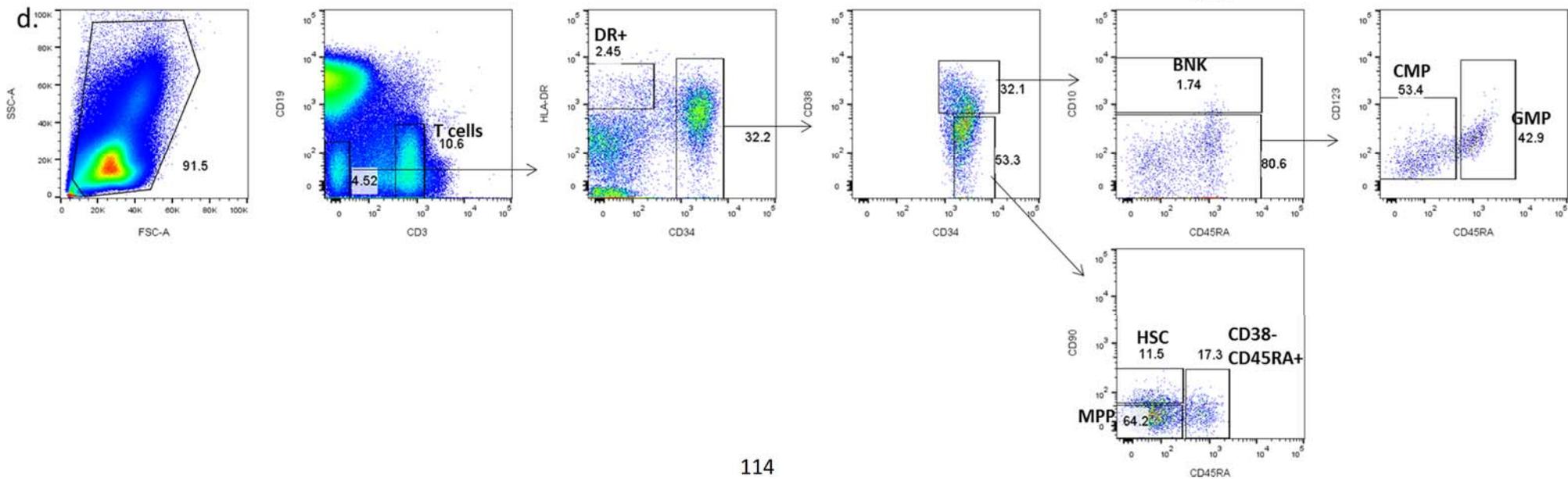
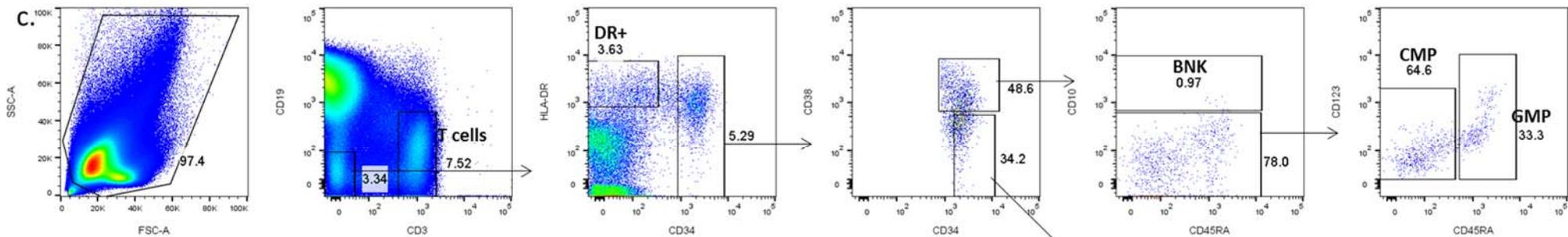
Patient ID	Gender	Cytogenetic phenotype	Age at presentation	WBC (x 10 ⁹ cells/l)	Additional notes
L876	Female	t(4;11)	4.5 months	334	Immunophenotype - CD19 89%, CD34 39%, CD15 56%, cyCD79a 21%, Tdt 10%, HLA-DR 90%.
L880	Female	t(9;11)	2.3 months	147	Immunophenotype - CD19 88%, CD34 10%, CD10 13%, CD15 10%, cyCD79a 76%, Tdt N/A, HLA-DR 90%. Karyotype - 49, XX, +X, t(9;11)(p22;q23), +18, +21.
LK124	Female	t(4;11)	3.3 months	333	Morphology - The film is overwhelmed by abundant small and medium sized blasts with a high N:C ratio and basophilic cytoplasm. Some larger more pleomorphic blasts are also present. Many smear cells and bare nuclei seen. No cytoplasmic granulation or Auer rods seen on high power inspection. Severe anaemia and thrombocytopenia in addition. Likely acute leukaemia. Immunophenotype - CD19 97%, CD34 1%, CD10<1%, CD15 32%, CD79a 92%, HLA-DR 97%, Tdt (90%).
LK228	Male	t(4;11)	1.5 months	95	Bone marrow morphology - Markedly hypercellular sample. Normal haematopoiesis largely replaced by medium sized lymphoblasts, blast count >90%. Immunophenotype - CD45 87%. CD34 variable +/-neg, CD19+, CD10 neg, HLA DR+, CD15+, slg neg, nTDT+, cCD79a+, CD20 neg, CD22+. Karyotype 46,XY,t(4;11)(q21;q23)[8]/46,XY[2].
LK230	Female	t(4;11)	4 months	130	Peripheral blood morphology – Many small-to-medium sized blasts with fairly high N:C ratio and large nucleoli. There is no granulation or Auer rods. Immunophenotype - CD45 77%. CD34 +, CD19+, CD10 neg, HLA DR+, CD15 variable +/-neg, slg neg, nTDT+, cCD79a+, CD20 neg, CD22+. Karyotype - 46,XX,t(4;11)(q21;q23)[5]/46,XX[1].
LK271	Male	t(9;11)	7.5 months	718	Immunophenotype - CD34-, Tdt-, CD19+, CD10+, CD7 -/weak, HLA-DR+, CD22+, CD20 variable, slg-, cCD79a+.

Table 3-1 Characteristics of patient infant ALL MLLr samples.

3.5.2. Haematopoietic populations in the MLLr iALL

The six MLLr iALL were stained for the haematopoietic population immunophenotypes. The distribution of the population was shown in Figure 3-10.





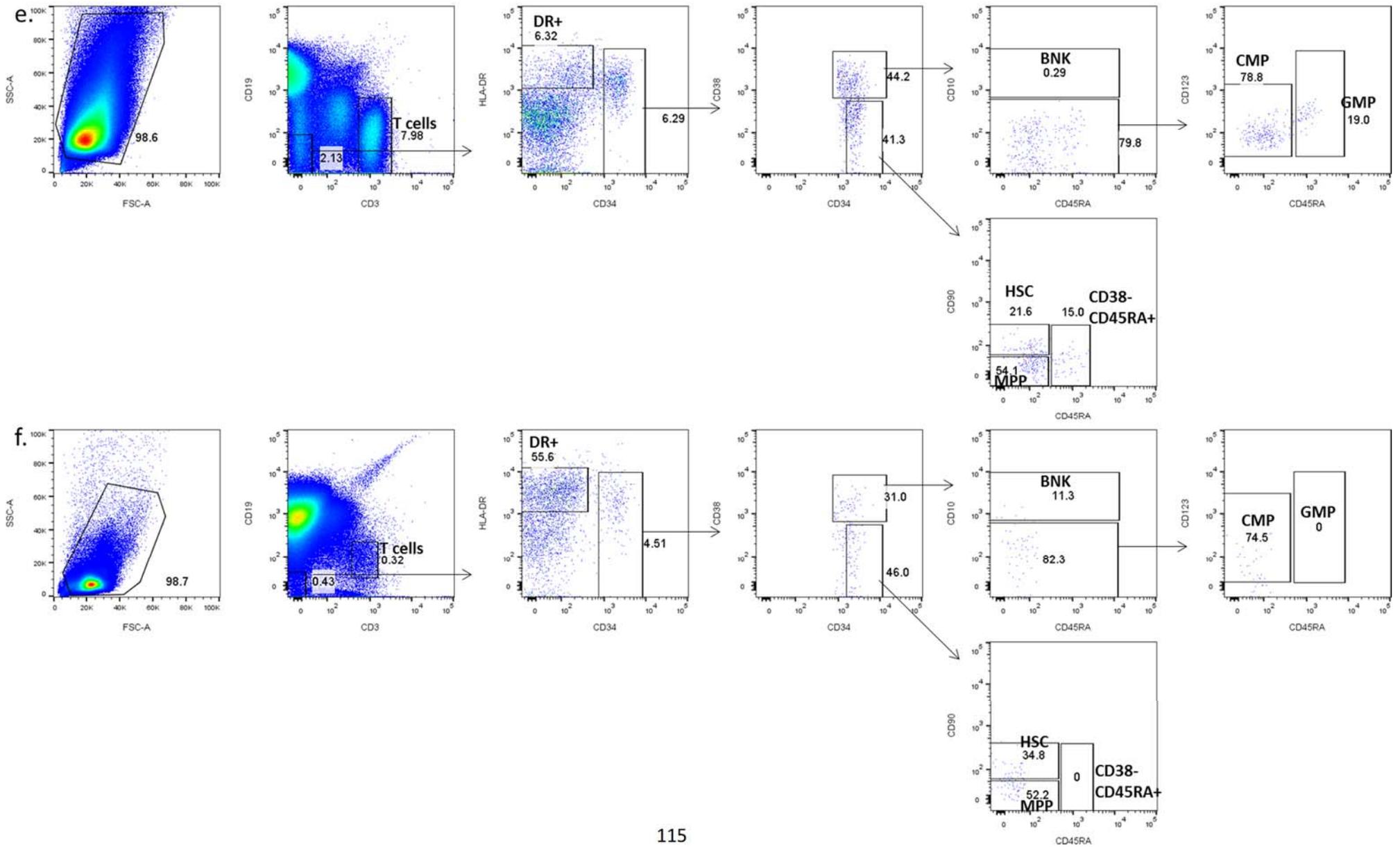


Figure 3-10 Haematopoietic progenitor populations on L876 (a), LK124 (b), LK228 (c), LK230 (d), L880 (e), and LK271 (f). a, b, c, and d were t(4;11) iALL cases, whereas e and f were t(9;11) iALL cases. The populations were gated for CD19- expression, as shown in the second graph in each instance. Despite having different progenitor population proportions, the samples L876, LK124, LK228, LK230, and L880 contained the entire progenitor populations. However, the sample LK271 had almost no CD19-CD34+ fractions, dominated only by the blast. It included no CD38-CD45RA+ and GMP subsets. The DR+ population is the mixture of monocytes and dendritic cells.

The progenitor populations were shown by initially selecting by absence of CD19 expression, as indicated on the second pseudocolour plot in each case. The flow data showed the presence of HSC, MPP, CD38-CD45RA+, CMP, GMP, and T cells on all of the samples, apart from the LK271, which did not show the GMP and CD38-CD45RA+ populations. LK271 also showed an unusual 'BNK' population that had a negative expression of CD45RA or a population with CD34+CD38+CD45RA-CD10^{negative-to-high}. Therefore, sample LK271 was sorted only for HSC, MPP, and the population named CD34+CD38+. It was still unknown why LK271 had different population distribution, despite being a bone marrow sample, but it might correlate with the substantially high number of blast cells present in peripheral blood at diagnosis (718×10^9 white blood cells/l).

3.5.3. Pre-leukaemic populations on the *MLLr* iALL

The *MLL* fusion as the primary mutation was evaluated on the sorted progenitor populations described in section 3.5.2 (Figure 3-10) in both the DNA and transcript. The fusion DNA sequences were determined in collaboration with Dr Claus Meyer (DCAL, Frankfurt)[198]. Based on these DNA sequences, the primers for DNA and cDNA fusion genes were designed using Primer Express 2.0. Detection of the mutations by PCR identified the pre-leukaemic populations, summarised in Figure 3-11a.

a.	t(4;11)								t(9;11)		LK271	DNA
	L876		LK124		LK228		LK230		L880			
	DNA	cDNA	DNA	cDNA	DNA	cDNA	DNA	cDNA	DNA	cDNA		
HSC												HSC
MPP												MPP
45RA+38-												34+38+
CMP												Blast
GMP												
Blast												
B cells												
T cells												
Mono + DC												

MLL fusion negative
 MLL fusion positive
 No cells

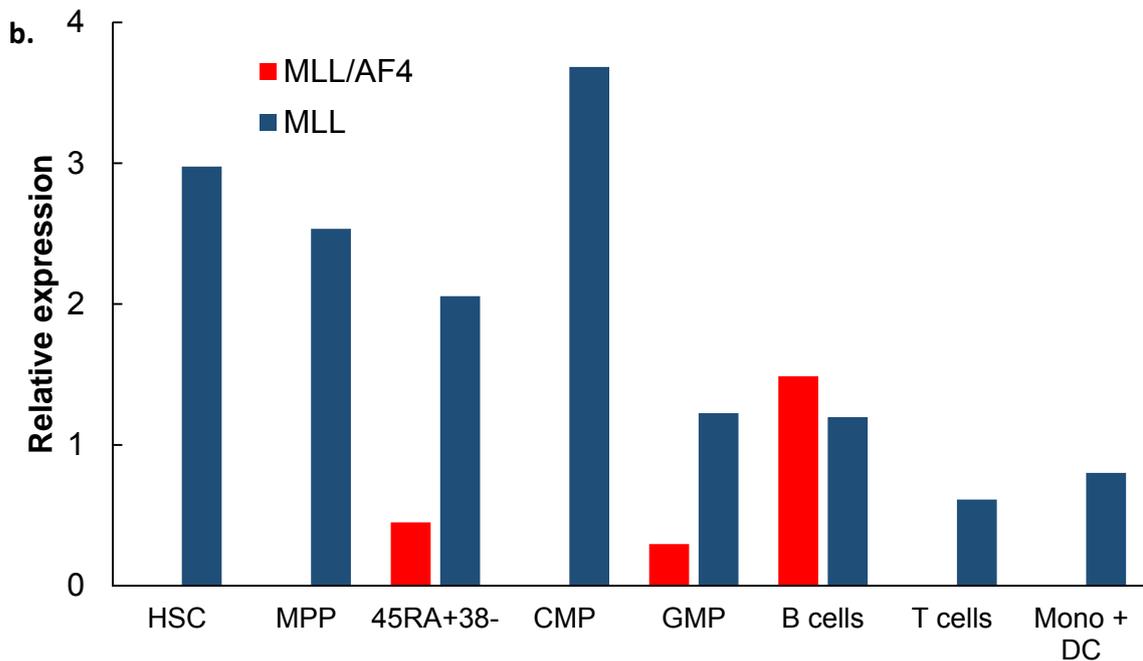


Figure 3-11 The pre-leukaemic populations on t(4;11) and t(9;11) iALL. a) The *MLL* fusion was evaluated on the different haematopoietic subpopulations on the DNA and cDNA. The variation in positivity between DNA and cDNA might be either the gene was not expressed on the transcript, or the assay was not sensitive enough because of the very low number of cells. *MLL/AF4* was present as early as in CD38-CD45RA+ on two of the cases and GMP population on the other two of the cases. *MLL/AF9* was only detected in the blast cells (B cells might be biased with blast cells due to co-expression of CD19) on patient L880. Patient LK271 had different population distribution, where the distinct and isolated populations were HSC, MPP, and CD34+CD38+. The *MLL/AF9* was observed on the CD34+38+ and blast populations. b) Nested qRT-PCR *MLL/AF4* and *MLL* wt on patient LK228 showed that the *MLL* gene was expressed on all of the populations (n=2). Although the fusion gene was found in the HSC and monocytes + DC DNA, the transcript was not detected. The analysis was relative to Blast sample. Due to the saturation that might have been reached on the first PCR of the nested reaction, the value on the graph might not be fully quantitative.

The analysis showed a difference progenitor harbouring fusion gene between t(4;11) and t(9;11) cases. In t(4;11) cases, two samples (LK228 and LK230) contained *MLL/AF4* in CD38-CD45RA+ population, and the other two cases (L876 and LK124) had the fusion as early as in GMP population. Case LK228 also showed the fusion gene on the HSC on the DNA analysis, but not on the transcript. The reason might be due to the low number of cells that became the limitation of the assay or because the gene was not transcribed into mRNA. Moreover, patient LK228 also showed the *MLL/AF4* positivity on the CD38-CD45RA+ and the downstream population, GMP, but not on the CMP. These data may indicate the GMP *MLL/AF4* positive population was derived from CD38-CD45RA+, but not from CMP.

In contrast to the t(4;11), the analysis of t(9;11) on case L880 showed the fusion gene only appeared in the blast and B cells. The mutation was not observed in any of the progenitor populations, which might suggest that the primary mutation happened in the more mature cells. This result was further supplemented with the case LK271. Interestingly, LK271 showed a different population distribution, where there were nearly no CD19- precursor populations, *i.e.* dominated by the blast cells. These CD19- progenitor populations were isolated for HSC, MPP, and the more mature CD34+CD38+ populations. The latter was not able to be differentiated into the commonly known populations. Nevertheless, the analysis of this population showed the presence of *MLL/AF9* sequences. Ultimately, it might indicate that *MLL/AF9* mutation occurred in the more mature population or the population CD34+CD38+ was the result of incomplete differentiation due to the effect of the fusion gene.

In addition to the *MLL* fusion gene evaluation on DNA and transcript of the subpopulations, the *MLL* wt expression was also assessed in order to understand whether the wild-type gene itself was expressed. Interestingly, all of the subpopulations showed the transcript, including the HSC and monocytes + DC (n = 2), exemplified by case LK228, Figure 3-11b. Although it does not conclusively exclude the limitation, it may suggest the possibility of different transcription regulation between the *MLL* and *MLL* fusion gene.

3.6. Discussion

The major focus of these investigations was to identify the origin of leukaemia in the lineage switch L826 and MLLr iALL. The study was started with the analysis of the Ig rearrangement within L826. The B cells and T cells rearrange their BCR and TCR loci, respectively, during their development. These receptors retain their germline configuration and are not expressed by myeloid cells. The analysis of the recombination in both presentation and relapse L826 indicated that the relapse AML cells were not derived directly from the pre-B lineage ALL cells, but arose as a result of a differential lineage fate from a common, non-lineage committed precursor, primitive to B cell differentiation and V(D)J rearrangement. Nevertheless, this result may not represent all of the lineage switch cases because some reports have identified the same IgH or TCR rearrangements in both the presentation and relapse[100, 106, 108, 110]. This phenomenon is intriguing because it implies that the lineage switch driver(s) can even convert the more mature blast cells.

The next question is where the common precursor cell in L826 switch is located. Based on the studies on neonatal blood spots[32, 199], it was suggested that the t(4;11) ALL has arisen during the foetal haematopoiesis. Furthermore, several backtracking studies in non-MLLr AML[200-202] have shown evidence of mutations in non-leukaemic progenitor cells. These studies suggest the possibility of finding the pre-leukaemic population L826 by identifying the primary mutation, *MLL/AF4*.

The progenitor populations were isolated from Lin-CD3-CD14-CD16-CD56-CD34+ cell haematopoietic subsets. The fusion gene was consistently found as early as the MPP and the CD38-CD45RA+ populations in both the L826 presentation and relapse samples, respectively. However, the result also showed an inconsistency on the MPP relapse case. The fusion gene was detected once among the three repetitions which might indicate the possibility of an assay sensitivity issue, but may equally result from a contamination problem. Nevertheless, this finding is encouraging because both MPP and CD38-CD45RA+ population[6] have potential for lymphoid and myeloid differentiation. It indicates the origin of the AML relapse may come from either one of these populations. This origin of the relapse population may provide a hint of what are the driver gene(s) that create the different lineage fate in L826 by detailed examination of the mutation profiles of these populations. This approach was investigated in Section 4.3 and 4.4.

The presence of a progenitor population harbouring MLLr in the MLLr ALL cases was supported by including the additional four t(4;11) and two t(9;11) samples. The *MLL/AF4* was identified in CD38-CD45RA+ in two of the cases and GMP in the other two cases. Additionally, the mutation was found in fully differentiated populations, e.g. B cells, T cells, monocytes and, in some cases, DC cells. These data demonstrate that the presence of *MLL/AF4* may still allow the normal differentiation of a pre-leukaemic precursor population, although the functional role of the fusion in the specific subpopulations has not been assessed here. Interestingly, the CD38-CD45RA+ and GMP populations were also shown to have leukaemic stem cell activity in most AML cases (in a study on 100 primary AML samples)[6]. Furthermore, this study included the expression analysis of specific genes on normal HSC, CD38-CD45RA+, GMP, and MEP populations that revealed *HOXA9*, *CD79A*, *IGHM*, and *FLT3* only expressed predominantly in the CD38-CD45RA+ population[6]. The *HOXA9* expression was accompanied by *MEIS1* that was expressed in CD38-CD45RA+ and HSC, but substantially lower on the other populations. This finding is fascinating because the correlation between the upregulation of *HOXA9* and its cofactor *MEIS1* has been previously shown in t(4;11)[203] and t(9;11) leukaemias[204, 205]. Moreover, the immunophenotyping of P929, a t(4;11) primograft from infant ALL and A789, a t(9;11) primograft from mixed phenotypic acute leukaemia, showed that both samples had the CD38-CD45RA+ as the majority of their progenitor populations (Figure 3-9). This evidence raises a question if there is any correlation between the subpopulation characteristics and the MLLr, and in different MLL fusion partner genes. This question could be partially addressed by further studying the transcriptomic profiles of each haematopoietic precursor population harbouring the MLLr.

In contrast to the t(4;11), the analysis of iALL with t(9;11), L880, identified the fusion gene only in the CD19+ blast and B cell populations. This finding is consistent with the identification of more mature Ig gene rearrangements on MLL/AF9-positive iALL[30]. This study[30] showed the incomplete IgH rearrangement in the t(4;11), but a complete IgH rearrangement in the t(9;11), also accompanied by significantly higher frequencies of patients with IgK, IgL, TCRB, and V δ 2-J α rearrangements. These data may suggest the difference between the two leukaemias, with the t(9;11) ALL coming from a more mature clone. Nevertheless, further assessments need to be performed to provide a conclusion.

In addition to L880, the case LK271 also represented the t(9;11) iALL. Interestingly, patient LK271 showed a different haematopoietic subpopulations pattern compared with the other samples. This sample showed a population of CD34+CD38+ that was not presenting a distinct population of either BNK, GMP, or CMP. Whilst this sample was derived from the patient's bone marrow, it is possible that this finding relates to the extremely high peripheral white blood count seen in this patient at presentation. In contrast to other primary mutations such as the *MLL/AF4* in this study, *AML1/ETO*[201], or *DNMT3A*[202] that appear to allow normal differentiation, it remains unknown if *MLL/AF9* prevents the canonical differentiation fate of the progenitor cells. It is also unclear whether the potential of *MLL/AF9* to create ALL, AML, or MPAL is related to secondary mutations, microenvironment, or it is also influenced by the location of the cell of origin.

The incidence of recurrence and relapse ALL remain an important issue in the current therapy. The identification of the pre-leukaemic population described in this study provides the possibility of exploring the cellular components and the system to test drug efficiency.

Chapter 4 Order of accumulations of mutations in candidate driver genes L826

Following understanding the origin of the AML relapse L826, the next question is to identify the driver of the lineage switch and study its mechanism; how does it allow the progenitor cells to transform into the myeloid rather than the lymphoid lineage, despite the strong lymphoid lineage specification provided by *MLL/AF4*.

The first approach was to investigate the genomic landscape of the switch by performing WGS, WES, and RNAseq on the presentation, relapse, and remission L826, as described in the preliminary data (Section 1.4.5). Here we found 1 and 12 non-synonymous mutations in the presentation and relapse, respectively, which were also predicted to have a deleterious effect based on Condel scoring. Moreover, the mutated amino acids of the AML candidate genes are found to be conserved throughout the different species, analysed using Clustal Omega tool[206-208], indicating that they might be essential for the structure of the protein. The conservation is shown in Table 4-1.

Based on these data, we hypothesised that understanding the order of acquisition of mutations within the AML relapse would help determine the candidates that had the role as the driver of the switch. The selected candidates from these results would be carried further for the functional studies.

This chapter aims to characterise the mutation acquisition of the candidate driver genes.

There were three different approaches performed in this study. The first was an analysis based on the combination of mutations within individual clones. The AML cells were cultured in a semi-solid medium colony formation assay, and then the candidate mutation profiles from each colony were compared. The second was the analysis based on single cells. The AML cells were sorted into single cells, and similarly, the mutated candidate genes were examined within an individual cell. Lastly, the mutation profiles of candidate genes were evaluated in the sorted haematopoietic stem/progenitor populations. These different methods were expected to complement each other or to develop the previous analysis.

	Species	UniProt Entry	Amino acid alignment	Mutation
ACAP1	Human	Q15027	620 LLACEFLLQNGANVNQADSTGRGPLHHATILGHTGLACLFLKRGADLGARDSEGRDPLTI 679	R662P
	Rhesus macaque	F7HLN4	620 LLACEFLLQNGANVNQADSTGRGPLHHATILGHTGLACLFLKRGADLGARDSEGRDPLTI 679	
	Dog	F1PZP3	620 LLACEFLLQNGANVNQADNHRGRPLHHATILGHTGLACLFLKRGADLGARDSEDKDPLTI 679	
	Mouse	Q8K2H4	620 LLACEFLLQNGANVNQADSTGRGPLHHATILGHTGLACLFLKRGADLGARDTEGRDPLTI 679	
	-	-	-	
	Western clawed frog	F6ZX71	584 -----	
Channel catfish	A0A2D0RBA0	635 LAACEFLLQNGANVNQADSNRGRPLHHATILGHTGLVCLFLKRGADYNNKDISQKDPITI 694		
-	-	-		
CHD4	Human	Q14839	1033 PKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGGHRLVLFISQMTKMLDLEDLDFLEHEGYK 1092	R1068H
	Rhesus macaque	F6ZS61	1030 PKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGGHRLVLFISQMTKMLDLEDLDFLEHEGYK 1089	
	Dog	J9NW81	1053 PKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGGHRLVLFISQMTKMLDLEDLDFLEHEGYK 1112	
	Mouse	Q6PDQ2	1026 PKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGGHRLVLFISQMTKMLDLEDLDFLEHEGYK 1085	
	Chicken	F1NH79	1027 PKMPNGMYDGSALIRASGKLLLLQKMLKNLKEGGHRLVLFISQMTKMLDLEDLDFLEHEGYK 1086	
	Western clawed frog	A3KN93	1023 PKMPNGMYDGSALIKAGKLLLLQKMLRKLKDDGHRVLFISQMTKMLDLEDLDFLEHEGYK 1082	
	Channel catfish	A0A2D0S5T1	1037 PKMPNGMYDGSALTKSSGKLLLLHKMLRKLKEGGHRLVLFISQMTKMLDLEDLDFLEHEGYK 1096	
	Fruit fly	E1JI46	1035 TTAAGGLYEINSLTKAAGKLVLLSKMLKQLKAQNHRLVLFISQMTKMLDILEDLDFLEGEQYK 1094	
-	-	-		
CEP164	Human	Q9UPV0	942 LALLE-----VQEEETARREKQQLLDVQRQVALKSEEA---T---ATHQOLEEAAQKEHT 988	R953Q
	Rhesus macaque	F7CQ41	903 LALLE-----VQEEETARREKQQLLDVQRQVALKSEEA---T---ATHQOLEEAAQKEHT 949	
	Dog	F1PB71	929 LAQLD-----LQEEETARREKQQLLDVQRQVVLKSSQEA---T---ANHQHLEDEAKKEHT 975	
	Mouse	Q5DU05	933 LAQLN-----VQEEENIRKEKQLLDQAQRQAALEREAA---T---ATHQHLEEAQKEHT 979	
	Chicken	E1C8Q1	905 SVQLL-----SQEESLRKKQQLLDEDRRTERERDEAA--L---ASQLRLEENRKEHS 952	
	-	-	-	
	Channel catfish	A0A2D0ST88	946 ESRLL-----THAADLKKRRKQLDEEEDVEVESGLLETLPRLMKERERLRADLERARQESD 999	
Fopius arisanus	A0A0C9Q384	507 YLKLKHEVRVAVERRSKRRRESNT-----TASETERSAS-TKT--RTEKNESMGQKTPS 556		
-	-	-		
CHTF8	Human	P0CG13	13 GLAEWVLMELQGEIEARYSTGLAGNLLGDLHYTTEGIPVLIIVGHILYGKIIHLEKPFV 72	E27Q
	Rhesus macaque	H9EX86	13 GLAEWVLMELQGEIEARYSTGLAGNLLGDLHYTTEGIPVLIIVGHILYGKIIHLEKPFV 72	
	-	-	-	
	Mouse	P0CG15	13 GLAEWVLMELQGEIEARYSTGLAGNLLGDLHYTTEGIPVLIIVGHILYGKIIHLEKPFV 72	
	Chicken	A0A1L1RT85	61 GLERWVLELQGEVEPRGGGALPGSLLDGLHYTREGIPVLIIVGHILYGVVQLEKPFV 120	
	Western clawed frog	Q28J63	13 QPEDWVLMELQGEIEARKQDGLAGKMMGDLHYTKEGVPLLVGHILYGKMLDKPFV 72	
	Channel catfish	E3TD55	13 SPGEWVLELQGETVSRHDTGLAGNLLGDLHYTKEGVPIVLIIVGHILYGKQVLEKPFV 72	
Malaysian fruit fly	A0A0K8W6V6	12 MLEEYAVLELQGDLEIRSEENIHLNFIGDLYNNKYGQPIIIGHHILQGREQKLEKPFV 71		
-	-	-		
DHX36	Human	Q9H2U1	513 MFKSD---KFLIIPHLSLMPTVNQTVQVFRKTPPGVGRKVIATNIAETISITIDVVVVVID 568	S557G
	Rhesus macaque	F6U0W0	513 MFKSD---KFLIIPHLSLMPTVNQTVQVFRKTPPGVGRKVIATNIAETISITIDVVVVVID 568	
	Dog	F6V8H1	627 MFKSD---RFLIIPHLSLMPTVNQTVQVFRKTPPGVGRKVIATNIAETISITIDVVVVVID 682	
	Mouse	Q8VHK9	506 MFKSD---KFLIIPHLSLMPTVNQTVQVFRKTPPGVGRKVIATNIAETISITIDVVVVVID 561	
	Chicken	A0A1D5PHB6	492 MFKSD---RFIIPHLSLMPTVNQTVQVFRKTPPGVGRKVIATNIAETISITIDVVVVVID 547	
	Western clawed frog	F7ETR2	501 MFKSD---KFLIIPHLSLMPTVNQTVQVFRKTPPGVGRKVIATNIAETISITIDVVVVVID 556	
	Channel catfish	A0A2D0QSL4	531 MFKSD---KFVIIPHLSLMPTVNQTVQVFRKTPPGVGRKVIATNIAETISITIDVVVVVID 586	
Fruit fly	A0A1W4UVM3	443 TPKGQRWRNQLTVFPLHSLMQSAEQAVFRPPSGQRKVIISITIAETISITIDVVVVVID 502		
-	-	-		
NCOA2	Human	Q15596	600 AESSCHPGEQKETNDPNLPPAVSSERADGQSRHDSKGGTKLLQLLTTKSDQMEPSPLAS 659	K640Q
	Rhesus macaque	F6XNZ4	600 AESSCHPGEQKETNDPNLPPAVSSERADGQSRHDSKGGTKLLQLLTTKSDQMEPSPLAS 659	
	Dog	E2QXC0	600 AESSCHPGEQKETNDPNLPPAVSSERADGQSRHDSKGGTKLLQLLTTKSDQMEPSPLAS 659	
	Mouse	Q61026	600 AEASCHPEEQKPNDSMPPAASGDRAEGHSRLHDSKGGTKLLQLLTTKSDQMEPSPLAS 659	
	Chicken	A0A1D5PK20	395 -----	
	Western clawed frog	B5DE09	593 GESGCHSNEQKDCSEN---LSSVGDQTEGQSRLLDSDKGGTKLLQLLTTKSDQMEPSPPLPS 649	
	Channel catfish	A0A2D0PS41	545 ENDNIRHQDEKGN-----PGQFNSSDENKIHESKGGTKLLQLLTTKNENTESSSPS 597	
-	-	-		
PHF3	Human	Q92576	1117 -----NCKICTIGRMAPPVD-DL-----SPKKVKKVVGVGVA 1144	K1119I
	Rhesus macaque	F7H0T5	1118 -----NCKICTIGRMAPPVD-DL-----SPKKVKKVVGVGVT 1145	
	Dog	E2R727	1119 -----NCKICTIGRMAPPVD-DL-----SPKKVKKVVGVGS 1146	
	Mouse	B2RQ62	1096 -----NCKICTIGRMAPPID-DL-----SPKTKVKKVVGVGGA 1123	
	Chicken	E1BVY1	1036 -----NCKICTIGRMAPPDDDL-----SAKKVKKVVGVGVA 1064	
	Western clawed frog	F7D558	1026 -----NCKICTIGRMAPPVE-DI-----SPTKVTNSTGLII 1053	
	Channel catfish	A0A2D0QH5	856 -----HCKICTIGRMAPPVE-EA-----TTKAVKVVSTVI 883	
Mediterranean fruit fly	W8B032	1776 KPAIKNSPANINLNKSSVMSSSKKCKLKKPIAPKPIEVVSLIDQILESTKTVEEAAANLI 1835		
-	-	-		
PPP17	Human	Q15435	180 ENLSNLHQLQMLELGSNRIENIDTLTNLESFLGKKNITKLNLDALTNLTVLSMQS 239	R199L
	Rhesus macaque	F7GK00	137 ENLSNLHQLQMLELGSNRIENIDTLTNLESFLGKKNITKLNLDALTNLTVLSMQS 196	
	-	-	-	
	Mouse	Q3UM45	181 ENLSNLHQLQMLELGSNRIENIDTLTNLESFLGKKNITKLNLDALTNLTVLSVQS 240	
	Chicken	A0A1D5PH53	174 ENLSNLHQLQMLELGSNRIENIDTLTNLESFLGKKNITKLNLDALTNLTVLSIQS 233	
	Western clawed frog	Q6DIQ3	166 ENFGTLQLRLELGSNRIENIDTLTNLESFLGKKNITKLNLDALTNLTVLSVQS 225	
Channel catfish	E3TGG0	165 SNLEHLTGLQMLELGSNRIENIDTLTNLESFLGKKNITKLNLDALTNLTVLSIQS 224		
Fruit fly	Q9U3W5	143 ENLDMLTNLTMLELGNLKKIENIEMLVNLRQLFLGKKNIAKINLDTLVNLELTLVLSQA 202		
-	-	-		

Table 4-1 Conservation of mutated residues of candidate genes across different species. The protein sequences from different species are aligned. It includes chimpanzee, rhesus macaque, mouse, dog, and cat. The mutated amino acid is indicated in the orange box. The analysis was performed using Clustal Omega tool. Asterisk ‘*’ indicates fully conserved residue. Colon ‘:’ indicates conservation between groups of strongly similar properties, scored >0.5 in the Gonnet PAM 250 matrix. Period ‘.’ indicates conservation between groups of weakly similar properties, scored ≤0.5 in the Gonnet PAM 250 matrix[206-208].

4.1. Mutation acquisition analysis at the clonal level

The relapse L826 bone marrow (BM) cells were grown in semi-solid medium at a low concentration, allowing the formation of individual colonies. The colonies were picked, and their RNA was isolated. Eight out of the twelve candidate genes were selected for multiplex PCR on the colonies' RNA, based on their functions. The eight genes were *CHD4*, *CEP164*, *PPP1R7*, *DHX36*, *NCOA2*, *CHTF8*, *PHF3*, and *ACAP1*. *CACNB4* and *SLC4A8* that act as calcium channels and sodium and bicarbonate channels, respectively, were excluded in this primary study due to the assumption that their functions may not closely correlate with the cell lineage determination. *ZNF136*, that has zinc finger protein and a Kruppel-associated box (KRAB) A-box domain, might have roles in transcription regulation, but was excluded initially to simplify the assay. Also, *BACE2* was omitted due to low transcript expression in the primary samples. Lastly, *MLL/AF4* was included in the multiplex PCR to confirm that the colonies were from the leukaemic cells, rather than any remaining normal BM cells.

Following the multiplex PCR, 99 colonies that showed the clear candidate gene and *MLL/AF4* bands on the agarose gel electrophoresis were selected. In addition, three samples that did not show the *MLL/AF4* band and one bulk SKNO-1 (a t(8;21) AML cell line) sample were also included as negative controls that were used to determine the cut-off values. Subsequently, the amplicons were barcoded (based on Fluidigm barcodes) and sequenced using Next Generation Sequencing MiSeq with coverage minimum 1,000 reads per amplicon.

The analysis was made by calculating the percentage of each mutation per sample (*i.e.* %mutant = $\text{read_count}_{\text{mutant}} / (\text{read_count}_{\text{mutant}} + \text{read_count}_{\text{wt}}) \times 100\%$). However, it is important to note that there is a risk of contamination when using this technique, including the scattering floating RNA from dead cells in the culture media and the possibility that more than one colony was picked at the same time. Therefore, a higher cut-off value was used to conclude the results, which then ended up with only 79 samples out of the total 99 samples. The results are represented in two forms, firstly the whole 79 samples (Figure 4-1a) and secondly by including only the samples with higher sequencing reads and %mutant close to 100% (Figure 4-1b).

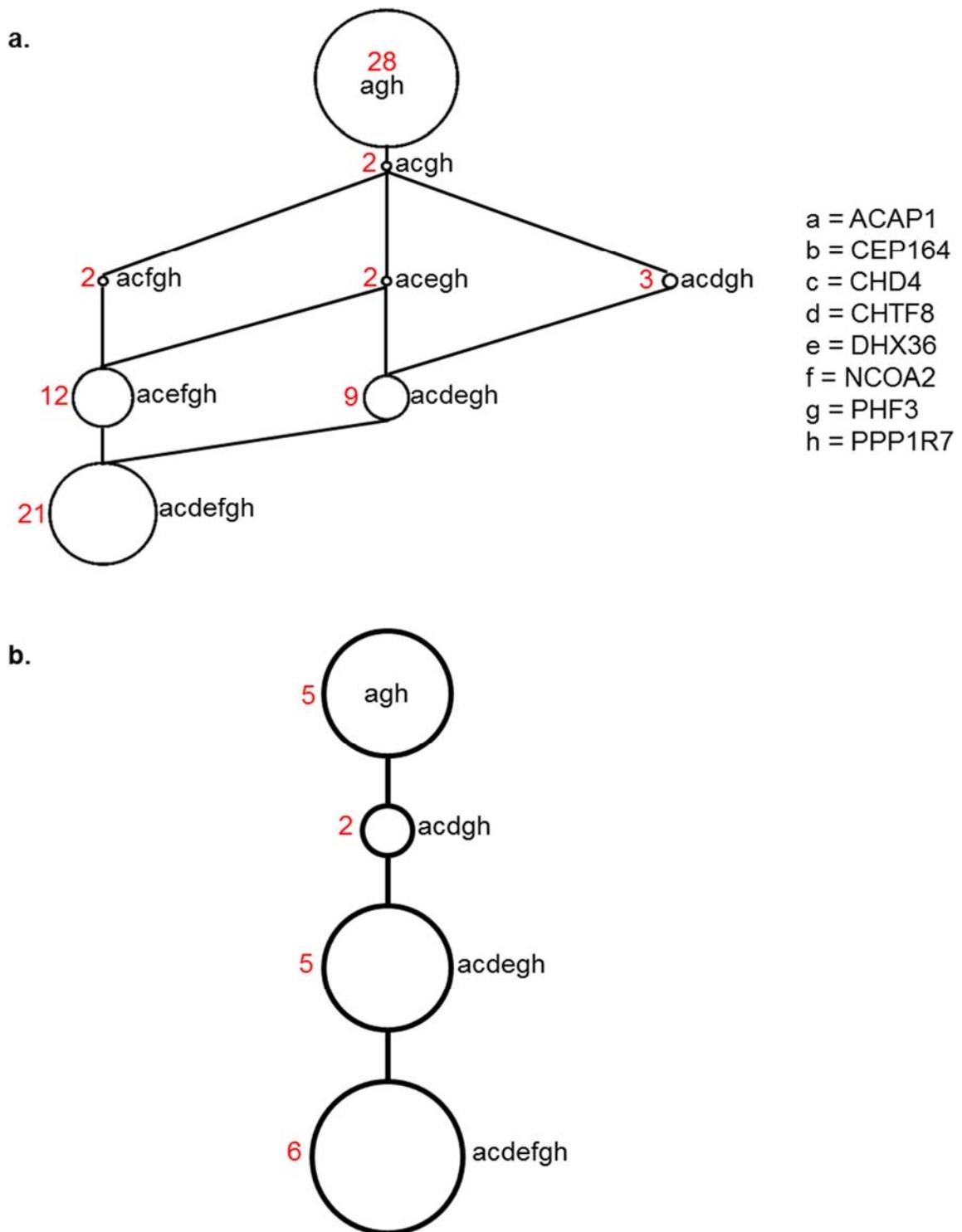


Figure 4-1 The order of mutation acquisition based on the individual clones derived from L826 relapse. a) Multiplex sequencing of 8 candidate genes from 79 different colonies showed concurrent *PPP1R7*, *PHF3*, and *ACAP1* (symbolised with *agh*) mutations are the first mutations. *CHD4* was acquired subsequently before any others mutations. b) The analysis by including only the samples with high sequencing reads and mutation percentage close to 100%. This more stringent analysis identifies simultaneous *CHD4* and *CHTF8* mutations following *PPP1R7*, *PHF3*, and *ACAP1*. *DHX36* and then *NCOA2* mutations occurred after the others mutation. *CEP164* was excluded in the analyses due to read counts not meeting the cut-off criteria. Red numbers indicate the number of representative clones.

These results suggest that concurrent mutations of *PPP1R7*, *PHF3*, and *ACAP1* are the initial mutations. Looking in more detail at these genes, *PHF3* is the candidate with the highest number of degrees relating to the difference between ALL and AML transcriptional changes in an ARACNe analysis (performed by Dr Dan Williamson, Section 1.5.2). The mutations of *ACAP1* and *PPP1R7* are located on their active domain. R662P on *ACAP1* is located on its ANK repeat (Section 1.5.1), a region that may interact with Arf-GAP domain[133, 146]. It may be significant since Arf-GAP regulates Arf6, a molecular switch that acts in a variety of targets, including metabolism (GLUT4 trafficking)[143] and phagocytosis in macrophages[144]. Mutation R199L occurred on LLR6 of *PPP1R7* (Section 1.5.4). These repeats are also known to interact with PP1 and regulate cell proliferation through counteracting Aurora B kinase activity[134, 135].

In addition to the order of mutations, these results also suggest the concurrent mutations may have an advantage in the AML cells clonogenicity, indicated by their presence in all of the colonies. It also means that this analysis is subject to the risk that a mutation or a particular group of mutations may confer either a clonal advantage or disadvantage, which then became the reason for their dominance or absence, respectively. Therefore, additional approaches, such as the analysis on the single cell level needs to be performed to validate the results.

4.2. Mutational acquisition analysis at the single cell level

Analysis of mutational evolution at the single cell level is required to validate the analysis at the clonal level for two reasons. Firstly, to reduce the risk of contamination and secondly to prevent clonal dominance within the samples where cells with a dominant mutation will out compete and form the majority of the colonies.

Furthermore, a phylogenetic tree that shows chronology of cell divisions has also been developed from single cell samples using microsatellite (MS) sequence analysis[209-212]. Thus, it is possible to overlay candidate driver gene acquisition with the microsatellite analysis of the single cells to confirm the evolutionary development of the relapse and the order of acquisition of mutations. These experiments were conducted in collaboration with the Shapiro lab; Weizmann Institute, Israel[212].

Single cells were obtained using single cell flow cytometry sorting in 96 well plate format. The cells were lysed and the DNA was directly amplified using the multiple displacement amplification (MDA) whole genome amplification (WGA) method. Due to high risk of extraneous DNA contamination that could be amplified by the WGA, two negative control (0 cells) were included per amplification run, in addition to the forty-five single cell samples.

The amplified products were validated using a multiplex PCR assay with four MS PCR primers. From these, at least two amplified products were required as a quality control. The validated products were divided into two pathways. Firstly, the mutational analysis was performed in Newcastle, whilst in parallel, MS sequences were amplified and sequenced at the Weizmann Institute.

The investigation of mutational analysis was performed by multiplex PCR of the candidate genes, followed by next generation sequencing. The amplification of the target genes was achieved, represented by the DNA electrophoresis in Figure 4-2. The electrophoreses showed the amplifications on the majority of the samples, but not 100% on some of the samples, e.g. sample 4 and 10 in Figure 4-2. It was probably because the efficiency issue of the WGA or the multiplex PCR reaction.

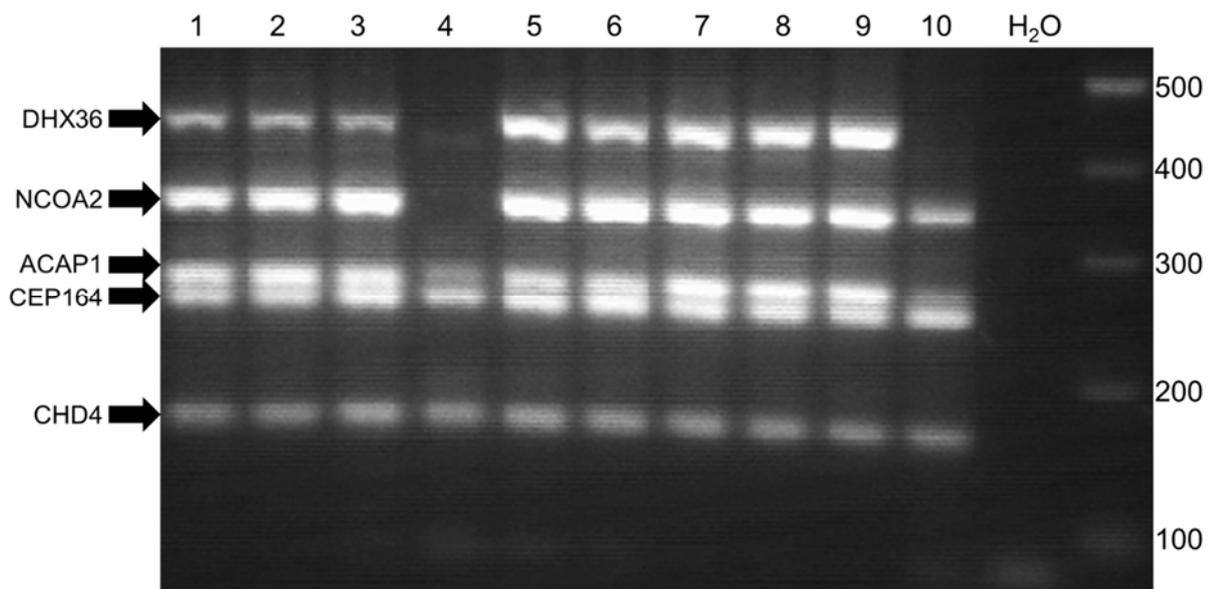


Figure 4-2 Multiplex PCR candidate driver genes from amplified DNA single cell L826 relapse.

The electrophoresis represented one of the three different multiplex PCRs for the 12 candidate driver genes. This multiplex reaction included *DHX36* (445 bp), *NCOA2* (345 bp), *ACAP1* (270 bp), *CEP164* (245 bp), and *CHD4* (155 bp). It showed that the majority of the genes could be amplified even though the coverage was not 100% on some of the samples (sample 4 and 10).

Unfortunately, the order of mutations from the sequencings have not been successfully constructed. The major problem was the presence of various mutation percentages, such as 0%, 3%, 7%, 10%, to any other percentages up to 100%, while we expect to have mutation percentage either ~0 or ~100% (homozygous) or ~50% (heterozygous) per gene, exemplified in Table 4-2. A possible explanation for this effect might be allelic drop-out (ADO), a condition where one of the alleles outcompetes the other during amplification. This remains a primary difficulty with WGA from the genomic DNA as template[213-215]. Consequently, the analysis showed disordered sequences of mutations.

In addition to our data, the MS analysis by Shapiro's lab was still undergoing at the time of this report writing.

		Cell_1	Cell_2	Cell_3	Remission
CHD4	wt	31912	74	432	18779
	mut	9722	28532	28717	54
	mut percentage (%)	23	100	98	0
NCOA2	wt	2192	4488	3603	6407
	mut	4109	1375	3186	103
	mut percentage (%)	65	23	47	2
PHF3	wt	4579	143	317	7826
	mut	26	9680	8104	39
	mut percentage (%)	1	98	96	0
ACAP1	wt	15902	11742	7498	11415
	mut	18	14	494	6
	mut percentage (%)	0	0	6	0
CEP164	wt	76	7028	12713	12994
	mut	18917	7021	2165	7
	mut percentage (%)	99	50	14	0
PPP1R7	wt	15471	13201	27449	17052
	mut	14756	790	50	43
	mut percentage (%)	49	37	0	0
CHTF8	wt	3297	12	53	13400
	mut	8816	7493	17554	11
	mut percentage (%)	71	97	97	0
CACNB4	wt	2028	3318	3160	4666
	mut	1417	1713	42	7
	mut percentage (%)	40	33	1	0
SLC4A8	wt	28	22860	6620	16330
	mut	23423	2358	13933	50
	mut percentage (%)	100	9	68	0
ZNF136	wt	34	21344	2901	19047
	mut	27517	1617	27087	37
	mut percentage (%)	100	7	90	0

Table 4-2 Examples of read count and mutation percentage sequencing L826 single cells.

Column Cell_1, _2, and _3 represent each single cell sample. Remission is a negative control carrying no mutation but underwent the whole genome amplification. The mutation percentages vary from 0 to 100% in the single cell samples, rather than showing heterozygous (~50%) or homozygous wild-type or mutant (~0 or ~100%). Further, a contrary mutation acquisition pattern was shown in the result. For instance, mutant PHF3 was shown on Cell_3, but not on Cell_1 which indicated Cell_1 was located at a higher hierarchy. However, CACNB4 mutation was found in Cell_1, not in Cell_3, which implicated a contrary result. wt = wild-type, mut = mutant.

Although at the moment a conclusive result has not been obtained, we expect that a study using single cell RNA sequencing might provide better results because of the availability of the initial templates (*i.e.* higher copies on the RNA).

4.3. Mutational acquisition analysis based on the haematopoietic hierarchy

Haematopoietic hierarchy depicts the order of blood cells formation. Backtracking the fusion gene allowed us to identify its presence in a progenitor population (Chapter 3). Based on this principle, we hypothesised that the method could also enable us to construct the sequential order of acquisition of the secondary mutations.

The DNA from L826 relapse haematopoietic sorted populations (Section 3.2.1) was used as the template for the analysis. Multiplex PCRs containing the 12 candidate genes and *MLL/AF4* were performed, followed by next generation sequencing of the products. The results of the sequencing were shown in Table 4-3 and summarised in Figure 4-3.

Concurrent *PHF3* and *CHD4* mutations were identified as the first abnormalities within the hierarchy, as early as in the MPP, followed with the others mutations in the progeny populations. Intriguingly, in contrast to the previous attempts to identify the *MLL/AF4* fusion in the AML related MPP, Section 3.2, in this experiment the fusion gene was detected along with candidate mutations of *CHD4* and *PHF3*. This may reflect a greater sensitivity with the next generation sequencing approach taken here. Taken together, these results indicate that MPP is the first mutated population, and *PHF3* and *CHD4* are the first secondary mutations.

These data also suggest the control of lineage commitment was affected by the mutated genes. The mutated genes were only identified within the myeloid lineage, despite the ongoing identification of the *MLL/AF4* fusion gene within lymphoid restricted (B, T and NK) populations. It indicates that the mutated genes may restrict the leukaemic or pre-leukaemic cells to myeloid differentiation.

		HSC	MPP	45RA+38-	CMP	GMP	T cells	B cells	NK cells	Mono	AML bulk	Kasumi-1
CHD4	wt	55446	23424	3191	81	4721	54326	62485	37900	11553	6072	85497
	mut	65	13857	63998	87	57983	5710	6353	118	63514	40925	105
	mut pct (%)	0.1	37.2	95.3	51.8	92.5	9.5	9.2	0.3	84.6	87.1	0.1
CEP164	wt	26008	16959	18915	55	12967	24247	26975	21848	21213	16547	45459
	mut	20	7	20133	24	12965	608	233	59	15975	13107	52
	mut pct (%)	0.1	0.0	51.6	30.4	50.0	2.4	0.9	0.3	43.0	44.2	0.1
PPP1R7	wt	19278	17068	11502	54379	18875	25431	8869	14940	7016	22334	45667
	mut	17	124	12928	12469	19003	1144	10	5	10568	20002	28
	mut pct (%)	0.1	0.7	52.9	18.7	50.2	4.3	0.1	0.0	60.1	47.2	0.1
CHTF8	wt	27877	22591	13242	29432	13544	30440	19410	24697	19912	26416	27794
	mut	17	2400	18618	8977	16323	1892	1553	25	15944	21771	14
	mut pct (%)	0.1	9.6	58.4	23.4	54.7	5.9	7.4	0.1	44.5	45.2	0.1
DHX36	wt	1465	818	1443	1	534	768	3103	1565	1239	110	480
	mut	1	1	1242	1	572	45	2	3	1340	319	0
	mut pct (%)	0.1	0.1	46.3	50.0	51.7	5.5	0.1	0.2	52.0	74.4	0.0
ACAP1	wt	1432	1566	146	3325	861	2020	1367	1814	1486	547	1684
	mut	1	2	1218	5	779	235	0	0	1165	533	2
	mut pct (%)	0.1	0.1	89.3	0.2	47.5	10.4	0.0	0.0	43.9	49.4	0.1
ZNF136	wt	25895	15000	3729	8952	14225	23286	11922	17100	13749	34055	15409
	mut	9	6	14543	7997	14913	668	1551	260	11709	29692	9
	mut pct (%)	0.0	0.0	79.6	47.2	51.2	2.8	11.5	1.5	46.0	46.6	0.1
SLC4A8	wt	36434	38918	19949	32653	12855	43565	29882	31166	19207	16130	29204
	mut	28	20	13964	53	10674	1128	14	76	9977	12891	23
	mut pct (%)	0.1	0.1	41.2	0.2	45.4	2.5	0.0	0.2	34.2	44.4	0.1
PHF3	wt	11923	12305	4236	33343	4495	16308	12448	17003	5041	8335	11600
	mut	10	5537	12806	23	5591	1522	1187	22	6574	2754	5
	mut pct (%)	0.1	31.0	75.1	0.1	55.4	8.5	8.7	0.1	56.6	24.8	0.0
NCOA2	wt	16544	10987	20694	19	7529	15526	17604	17523	8938	6344	20186
	mut	34	24	1585	11	7658	500	37	41	13512	8072	48
	mut pct (%)	0.2	0.2	7.1	36.7	50.4	3.1	0.2	0.2	60.2	56.0	0.2
CACNB4	wt	9004	11489	7158	4767	3778	11046	8986	4148	3747	4437	7196
	mut	9	7	4921	5627	3103	293	11	1896	4453	3067	15
	mut pct (%)	0.1	0.1	40.7	54.1	45.1	2.6	0.1	31.4	54.3	40.9	0.2
BACE2	wt	2611	2357	1509	2504	1071	2780	2509	1266	1220	791	2500
	mut	3	164	1368	2563	1670	31	0	2	1757	594	1
	mut pct (%)	0.1	6.5	47.5	50.6	60.9	1.1	0.0	0.2	59.0	42.9	0.0
MLL/AF4		6	1905	4964	762	9232	1546	1477	773	4987	7920	8

Table 4-3 Sequencing reads candidate genes of L826 haematopoietic subsets. Reads are based on next generation sequencing multiplex candidate genes and *MLL/AF4*. A cut-off of 10% was used to determine a mutation. The red values indicate the mutated genes. The green values indicate very low read counts. 'AML bulk' sample is unsort AML relapse L826 as a positive control containing all of the mutations. Kasumi-1 is an AML cell line as the negative control.

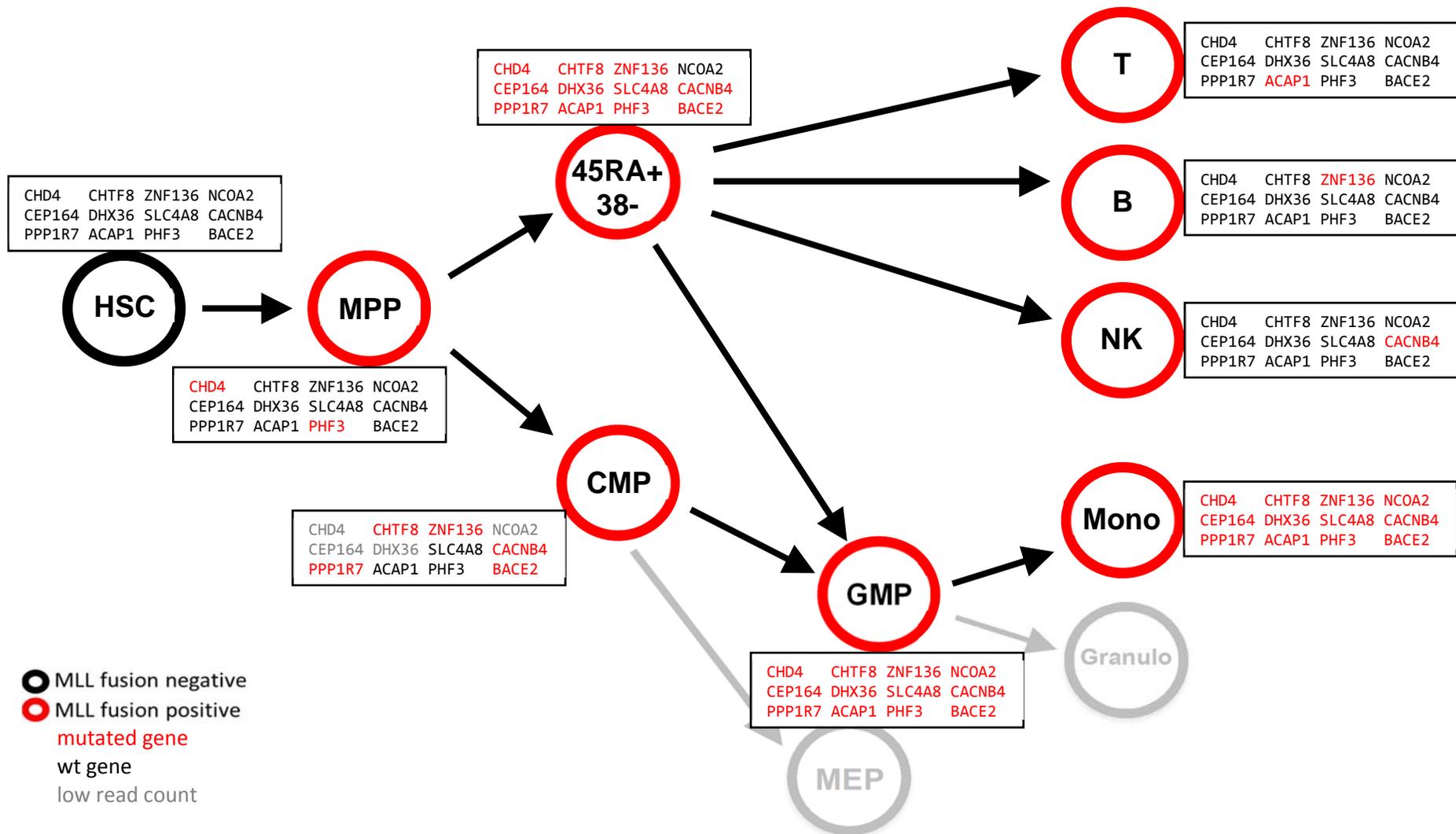


Figure 4-3 The schematic order of mutations L826 relapse based on the haematopoietic hierarchy. Concurrent *PHF3* and *CHD4* mutations were found in the MPP, in which *MLL/AF4* as the transforming mutation also presented. Other mutations were identified in the progeny populations, with minimum mutations in the lymphoid restricted populations.

In addition to these results, it also has to be noted that there is a contrary outcome with the analysis of mutation acquisition based on the clonal level (Section 4.1), where the latter showed the concurrent mutation of *PHF3*, *PPP1R7*, and *ACAP1* as the initial mutations, and *CHD4* as the subsequent mutation. The absence of *CHD4* mutation on the initial colonies may relate with the clonogenicity disadvantage affected by the mutant. The study by Zhao *et al.*[216] demonstrated that *Chd4* was required for the maintenance of ESC self-renewal in the mouse model, while knockdown of *Chd4* disrupted self-renewal. It might be possible that the mutation of *CHD4* disrupts its activity, affecting colony formation.

This hypothesis is supported by the colony formation studies on depleted *CHD4* within (1) AML cells by Sperlazza *et al.*[177] and (2) SEM cells in my MRes project. *CHD4* knockdown in different AML cell lines (U937, AML3, and MV4-11) resulted in a dramatic reduction of the colony formation[177]. Similarly, the knockdown in SEM cells showed ~50% reduction of colony formation (Figure 4-4), although it is important to mention that the functional effect of the *CHD4* mutation seen in L826 has not been addressed.

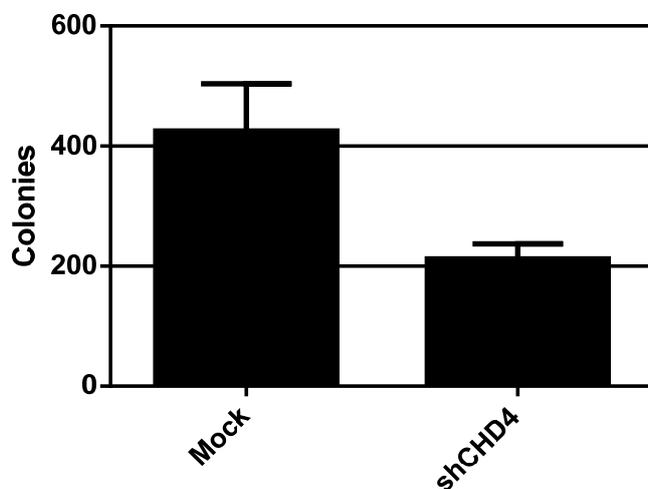


Figure 4-4 Colony formation assay on CHD4-depleted SEM cells. The cells were grown in 24 well plate at a concentration of 5,000 cells/ml (or 2,500 cells/well). The colonies were counted at day 8 after the plating. Error bars showed standard error from 4 replicates of a single experiment (pilot data from the MRes project).

4.4. Secondary mutations on the L826 presentation, relapse, and remission

After observing the presence of *PHF3* and *CHD4* mutations in the MPP L826 relapse, it led to a further question of whether these mutations have occurred on the presentation haematopoietic subpopulations that were not detected by the bulk sequencing (WGS, WES, and RNAseq). Therefore, the evaluation study of the relapse secondary mutations was conducted on the haematopoietic subsets of the presentation. This study also included the genes with secondary mutations that were detected only on the ALL and the mutations that were shared between the ALL and AML to provide additional information. Of note, these other genes have intronic mutation, rather than being non-synonymous coding mutations. They are summarised:

- a. mutations only on the presentation: *IL1RAPL2* and *MAGED1*
- b. mutations only on the relapse: *PHF3*, *CHD4*, and *NCOA2*
- c. mutations on the presentation and relapse: *PIN4* and *OSTF1*
- d. germline mutation: *PHF5A*, as a control

These different groups were expected to complement each other to construct the result. The mutations were analysed by multiplex PCR followed by MiSeq sequencing, summarised in Table 4-4.

Leukaemia	Subpopulation	ALL		AML			ALL and AML		Germline
		IL1RAPL2	MAGED1	CHD4	NCOA2	PHF3	PIN4	OSTF1	PHF5A
ALL Presentation	HSC	wt gene	wt gene	wt gene	no cell or no read	wt gene	mutated gene	wt gene	wt gene
	MPP	mutated gene	mutated gene	wt gene	wt gene	wt gene	low mutated percentage (≤8%)	mutated gene	mutated gene
	45RA+38-	mutated gene	wt gene	wt gene	wt gene	wt gene	mutated gene	wt gene	low mutated percentage (≤8%)
	CMP	mutated gene	wt gene	wt gene	wt gene	wt gene	mutated gene	low mutated percentage (≤8%)	wt gene
	GMP	mutated gene	mutated gene	wt gene	wt gene	wt gene	mutated gene	mutated gene	mutated gene
	Monocytes	mutated gene	mutated gene	wt gene	wt gene	wt gene	mutated gene	mutated gene	wt gene
	BNK	mutated gene	mutated gene	wt gene	wt gene	wt gene	mutated gene	mutated gene	mutated gene
	B cells	mutated gene	mutated gene	wt gene	wt gene	wt gene	mutated gene	mutated gene	wt gene
	T cells	wt gene	mutated gene	wt gene	wt gene	wt gene	mutated gene	wt gene	mutated gene
	NK cells	wt gene	wt gene	wt gene	wt gene	wt gene	low mutated percentage (≤8%)	wt gene	mutated gene
	DR+	mutated gene	mutated gene	wt gene	wt gene	wt gene	mutated gene	mutated gene	wt gene
	Bulk	mutated gene	mutated gene	wt gene	wt gene	wt gene	mutated gene	mutated gene	mutated gene
AML relapse	HSC	no cell or no read	no cell or no read	no cell or no read	no cell or no read	no cell or no read	no cell or no read	no cell or no read	no cell or no read
	MPP	wt gene	no cell or no read	mutated gene	wt gene	mutated gene	low mutated percentage (≤8%)	wt gene	mutated gene
	45RA+38-	wt gene	wt gene	mutated gene	mutated gene	mutated gene	mutated gene	mutated gene	wt gene
	CMP	no cell or no read	no cell or no read	no cell or no read	no cell or no read	no cell or no read	no cell or no read	no cell or no read	no cell or no read
	GMP	wt gene	wt gene	mutated gene	mutated gene	mutated gene	mutated gene	mutated gene	mutated gene
	Monocytes	wt gene	wt gene	mutated gene	mutated gene	mutated gene	mutated gene	mutated gene	wt gene
	BNK	wt gene	wt gene	mutated gene	mutated gene	mutated gene	mutated gene	mutated gene	wt gene
	B cells	wt gene	wt gene	wt gene	wt gene	low mutated percentage (≤8%)	low mutated percentage (≤8%)	mutated gene	wt gene
	T cells	wt gene	wt gene	low mutated percentage (≤8%)	wt gene				
	NK cells	wt gene	wt gene	wt gene	wt gene	wt gene	mutated gene	wt gene	mutated gene
	DR+	wt gene	wt gene	mutated gene	mutated gene	mutated gene	mutated gene	mutated gene	wt gene
	Bulk	wt gene	wt gene	mutated gene	mutated gene	mutated gene	mutated gene	mutated gene	wt gene
Remission	Remission	wt gene	wt gene	wt gene	wt gene	wt gene	wt gene	wt gene	wt gene

wt gene
 mutated gene
 low mutated percentage (≤8%)
 no cell or no read

Table 4-4 Secondary mutations of L826 in different haematopoietic subpopulations. Mutations that present only in the ALL (*IL1RAPL2* and *MAGED1*), only in the AML (*CHD4*, *NCOA2*, and *PHF3*), and shared between ALL and AML (*PIN4* and *OSTF1*) were evaluated on DNA of the haematopoietic subpopulations of L826 presentation and relapse samples by multiplex sequencing. *PHF5A* was a control mutated gene that was found in the germline cells.

These results demonstrated that the mutations which were found only in the AML based on WGS, WES, and RNAseq were absent in any haematopoietic subpopulations of ALL or remission, and vice versa. Additionally, the *PIN4* mutation was found as early as on the HSC of the ALL and was also found in the MPP of the AML. Although disappointingly HSC of the AML could not be analysed due to the availability of the material, these data again corroborated that the origin of the AML mutations occurred no later than in the MPP.

Lastly, this separate experiment showed the repeated finding of *PHF3* and *CHD4* mutation, but not *NCOA2*, in the MPP, indicating the techniques were operating at the limit of detection.

4.5. Discussion

These results showed the suggested order of acquisition obtained from different approaches. The first method was by constructing the mutation tree from the transcripts of individual AML colonies that showed *PHF3*, *PPP1R7*, and *ACAP1* as the first mutations, where the others mutations always carried these three mutations together with them. Interestingly, the colonies with these concurrent mutations were also the most prevalent colonies (28/79 colonies), despite approximately 50% of bulk sequencing reads showing the mutant allele of all mutations. This result might be arguable due to the potential of the mutations having a stemness advantage so that they become the dominant clones. Nevertheless, it also indicates that the mutations support AML clonogenicity.

The second method by using single cells of the relapse samples for the mutation tree construction aimed to validate the clonal mutation order. However, the results have not been conclusive, probably because of the impact of ADO, as has been reported in some studies[212-215]. Nonetheless, these data provide a pilot study, in which at least two aspects can be suggested. Firstly, the using of single cell transcript, rather than the genome may reduce the possibility of ADO, even though the possibility of one allele being transcribed more than the other needs to be considered. Secondly, selecting the progenitor cells rather than the bulk cells should be applied. The results of mutation acquisition based on haematopoietic hierarchy (Section 4.3) indicated two key findings: (1) MPP is the initial mutated population, and (2) more mature cells (blast cells) had all of the mutations. Therefore, it will potentially be better to isolate and sort the earlier populations, such as MPP, CD38-CD45RA+, and CMP as the samples, rather than the bulk cells since the order of mutation acquisition might be more accurately derived within these populations.

Even though a conclusive result has not been achieved by the single cell method, the study was able to be developed from the haematopoietic hierarchy. This approach interestingly showed *PHF3* as the initial mutation, accompanied with *CHD4* mutation, while *CHD4* mutation was also identified as the subsequent mutation after the concurrent *PHF3*, *PPP1R7*, and *ACAP1* mutations on the clonal analysis. The absence of *CHD4* mutation in the initial colonies is hypothesised to be related to the

disadvantage of clonogenicity potential affected by the mutant. It is supported by colony formation study on depleted CHD4 on AML cells[177], assuming the mutation generated the loss of function. These functional studies were explored further in the subsequent chapters.

Lastly, mutations in *PHF3*, *CHD4*, and *PIN4* further support the MPP as the originating population of the AML relapse. The next investigation is to identify if PHF3 and CHD4 co-mutations are required and adequate for the lineage fate regulation.

Chapter 5 Functional evaluation of candidate driver genes

The analysis of the 12 acquired mutations identified PHF3, PPP1R7, ACAP1, and CHD4 as the most likely founder mutations. The next part of the study was to investigate the function of the candidate genes with respect to their control of lineage commitment in MLL/AF4 bearing cells. The initial approach to this was over-expression of the mutant versus wild-type candidates in SEM cells, an ALL cell line with t(4;11). The impact of over-expression of mutant candidates was assessed by analysis of the expression of genes associated either with ALL or AML. Furthermore, genes known to be targets of the candidate genes, which have also been shown to be differentially expressed in L826 presentation and relapse RNAseq data, were included. They are summarised in Table 5-1.

Gene	Read counts		Fold change (AML _{read} /ALL _{read})
	ALL	AML	
HOXA5	2458.918	2.577109	0.001048
HOXA6	2322.468	3.163938	0.001362
HOXA7	3037.711	5.096784	0.001678
PRSS12	2834.465	9.222607	0.003254
EBF1	8244.554	69.50365	0.008430
LEF1	11173.05	169.5515	0.015175
ZFHX3	1705.756	116.4877	0.068291
PAX5	28182.33	4559.287	0.161778
MDK	2197.134	519.7407	0.236554
HOXA9	28664.86	32039.23	1.117718
CDKN2A	831.075	1089.459	1.310904
ZHX3	607.5331	1512.568	2.489688
MT2A	137.8099	343.4829	2.492440
HIPK1	7664.73	19933.04	2.600619
IL6R	1729.571	5853.621	3.384435
HOXC5	36.67685	913.2538	24.90000

Table 5-1 RNAseq reads L826 on the assessed genes. The genes were selected based on the known related function with ALL and AML cases or known target genes of the candidate driver genes, also differentially expressed between L826 ALL and AML.

Regarding CHD4 target genes, a chromatin immunoprecipitation sequencing (ChIP-seq) CHD4 dataset in human glioblastoma has been deposited by Chudnovsky *et al.*[217] in the NCBI Gene Expression Omnibus under accession number GSE52419. This dataset provides the information of CHD4 binding to the target DNA. In addition to this, we have L826 presentation and relapse DNase hypersensitivity sequencing datasets (generated by Constanze Bonifer, University of Birmingham) that show the DNA (chromatin) locations that are accessible for the protein to bind. Lastly, we also have L826 paired RNAseq that indicate the genes that are differentially expressed. Analysis of these three combined datasets identifies three important points: (1) CHD4 binds to several haematopoietic-associated genes, (2) the binding patterns are different between presentation and relapse, and (3) those genes are also differentially expressed between presentation and relapse. This intersection analysis was performed by Dr Sirintra Nakjang, and can be accessed in https://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=sirintra&hgS_otherUserSessionName=hg19_ricky_olaf. Several of these mentioned binding regions are shown in Appendix B, including HOXA5, HOXA6, HOXA7, EBF1, LEF1, and PAX5. These genes are also the target genes that were evaluated on the CHD4 section in this chapter.

Furthermore, the function of two of our lead candidate genes, CHD4 and PHF3, was explored in more depth in a model of MLL/AF4 leukaemia using CD34+ cord blood cells carrying a chimeric MLL/Af4 fusion gene[82]. The lymphoid and myeloid immunophenotypes were assessed following knockdown of CHD4 and/or PHF3.

5.1. Over-expression of mutant candidate driver genes

The over-expression experiments were performed on *PHF3*, *CHD4*, *PPP1R7*, and *ACAP1* genes. Mutant expression constructs were prepared in different vectors with different fluorescent protein tags, which allowed analysis of combinations of expression. An empty vector control was provided on each of the different vector backbones. The experiment was started with single transduction, and then continued with the combination transduction, summarised in Figure 5-1.

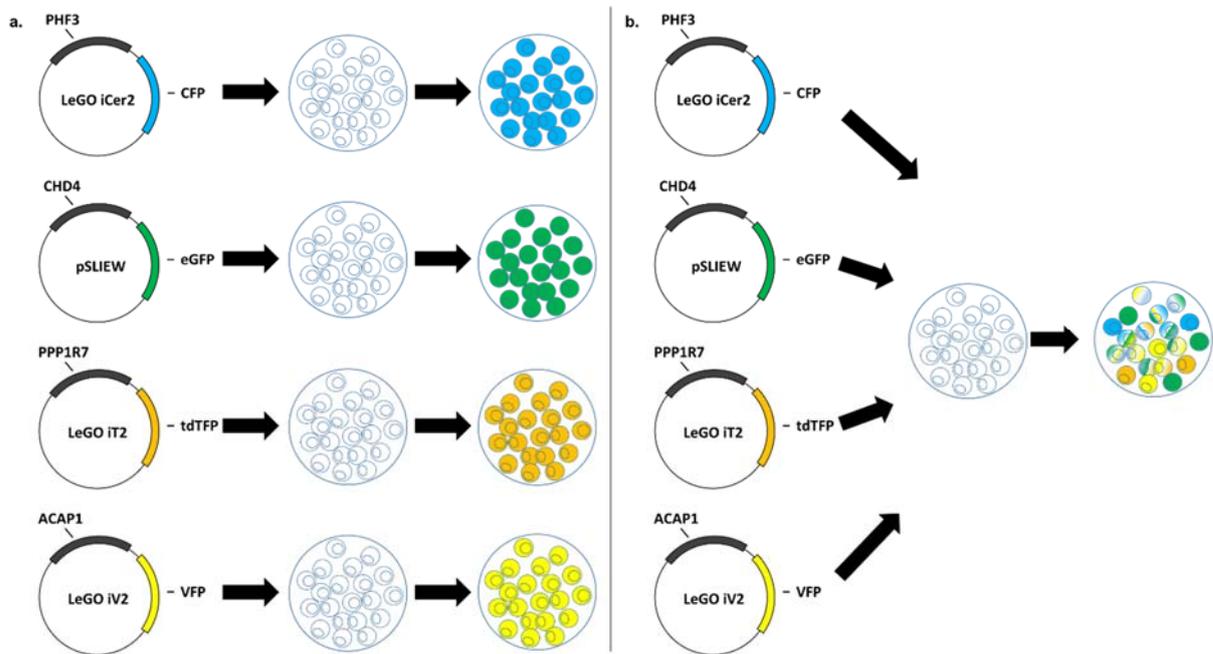


Figure 5-1 Workflow over-expression of candidate genes analysis. Candidate genes were constructed in different lentivirus vectors each with different fluorescent proteins. The fluorescent proteins included were as follows: cerulean fluorescent protein (CFP), enhanced green fluorescent protein (eGFP), tandem tomato fluorescent protein (tdTFP), and venus fluorescent protein (VFP) in *PHF3*, *CHD4*, *PPP1R7*, and *ACAP1*, respectively. The study was initiated with individual gene transduction of the SEM cell line (a) and following this a combination transduction was performed (b). The presence of different fluorescent protein markers allows tracking the cells that contain that specific gene.

5.1.1. *PHF3* over-expression

A 3382A>T (accession number NM_015153.3) mutation was generated for *PHF3* (6120 bp), resulting in the K1119I mutation, using site-directed mutagenesis of the pCR-XL-TOPO *PHF3* vector. The cDNA was added with a twin-strep-tag on its N-terminal and then transferred to LeGO iCer2 lentiviral vector using the following method:

- a. Isolation of the *PHF3* cDNA by PCR with additional *SacII* and *XhoI* restriction site sequences on its forward and reverse primers, respectively
- b. Creating the restriction sites *SacII* and *XhoI* on the *PHF3* PCR product using the restriction enzymes
- c. Digestion of pEXPR-IBA105 twin-strep-tag vector using *SacII* and *XhoI* to provide the cloning sites
- d. Ligation of *PHF3* into pEXPR-IBA105 vector to obtain twin-strep-tag on N-terminal of *PHF3*
- e. Isolation of the *PHF3*-twin-strep-tag in pEXPR-IBA105 using *NotI* and *XbaI*, where *XbaI* was made blunt-ended (Section 2.4.19)
- f. Digestion of LeGO iCer2 vector using *NotI* and *EcoRI*, where *EcoRI* was made blunt-ended, to provide the cloning sites
- g. Ligation of *PHF3*-twin-strep-tag into the LeGO iCer2 vector

The *PHF3*-twin-strep-tag LeGO iCer2 lentiviral vector also had cerulean fluorescent protein (CFP), linked by an internal ribosome entry site (IRES) to the spleen focus-forming virus (SFFV) promoter. The final construct was confirmed by Sanger sequencing. The vector map and mutated nucleotide sequencing are depicted in Appendix C.1.

The lentivirus was generated and then transduced into SEM cells (Section 2.2.6-8). The RNA was collected on day 8 after transduction. Genes that are associated with lymphoid and myeloid haematopoiesis, and that were differentially expressed between presentation and relapse L826, were evaluated by qRT-PCR. These results are shown in Figure 5-2.

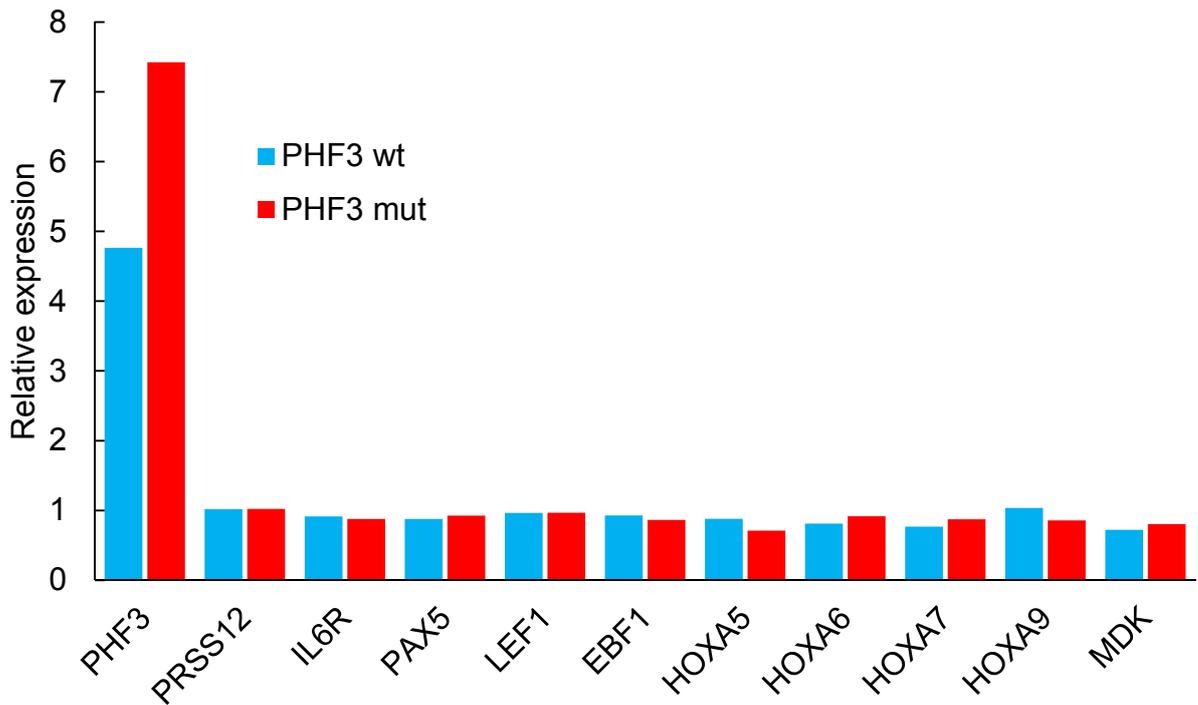


Figure 5-2 *PHF3* over-expression effects on lymphoid- or myeloid-associated genes on SEM cells. The analysis was performed on day 8 after transduction, relative to the empty vector control. *PHF3* wt and mutant were over-expressed 4.7 fold and 7.4 fold, respectively. No effect on the lymphoid- and myeloid-associated genes were observed.

Although higher expression of *PHF3* wt and mutant were seen, none of the haematopoietic-associated gene transcripts was changed. This result suggests several possibilities, including (1) no effect of the mutation, (2) no effect of *PHF3* on these genes, or (3) the over-expression does not affect the equilibrium condition, especially in a cell line. This experiment may need to be continued by firstly knocking down the endogenous *PHF3*, followed by rescue by over-expression.

5.1.2. *CHD4* over-expression

A 3415G>A (accession number NM_001273.3) mutation was generated on *CHD4* (5814 bp) that created the R1068H mutation using site-directed mutagenesis on the pDONR221 *CHD4* vector. The cDNA was added with twin-strep-tag on either its N- or C-terminal end and then transferred to pSIEW lentiviral vector using the following method:

- a. Isolation of the *CHD4* cDNA by PCR with additional *SacII* and *KpnI* restriction site sequences on its forward and reverse primers, respectively
- b. Creating the restriction sites *SacII* and *KpnI* on the *CHD4* PCR product using the restriction enzymes
- c. Digestion of pEXPR-IBA105 (N-terminal twin-strep-tag) and pEXPR-IBA103 (C-terminal twin-strep-tag) vectors using *SacII* and *KpnI* to provide the cloning sites
- d. Ligation of *CHD4* to the pEXPR-IBA105 or pEXPR-IBA103 vector to obtain the twin-strep-tag on N- or C-terminal of *CHD4*, respectively
- e. Isolation of the *CHD4*-twin-strep-tag (N- and C-terminal had the same methods) using *NotI* and *XbaI*, where *XbaI* was made blunt-ended
- f. Digestion of pENTR1A using *NotI* and *Sall*, where *Sall* was made blunt-ended, to provide the cloning sites
- g. Ligation of *CHD4*-twin-strep-tag into the pENTR1A vector
- h. Gateway cloning to transfer *CHD4*-twin-strep-tag from pENTR1A vector to pSIEW vector (method is described in Section 2.4.21)

The pSIEW *CHD4*-twin-strep-tag vector also had enhanced green fluorescent protein (eGFP), linked by IRES with SFFV promoter. The final construct was confirmed by Sanger sequencing. Further analysis was using the C-terminal tag construct unless stated otherwise. The vector map and mutated nucleotide sequencing are depicted in Appendix C.2.

The *CHD4* wt and mutant vectors were transduced into SEM cells. The transduction level was evaluated by FACS on the eGFP expression, as shown in Figure 5-3.

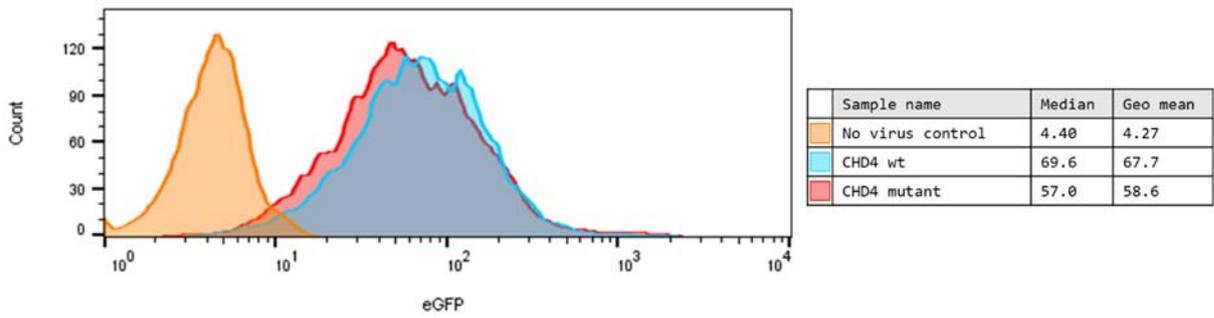


Figure 5-3 *CHD4* over-expression evaluation by FACS on eGFP expression. The wild-type and mutant *CHD4* over-expressed SEM cells showed the positive eGFP expression.

5.1.2.1. Analysis of transcript target genes

The transcripts of *CHD4* target genes on the over-expressed mutant versus wild-type samples were assessed by qRT-PCR. The analysis was performed on day 9 and 10 after the transduction, summarised in Figure 5-4.

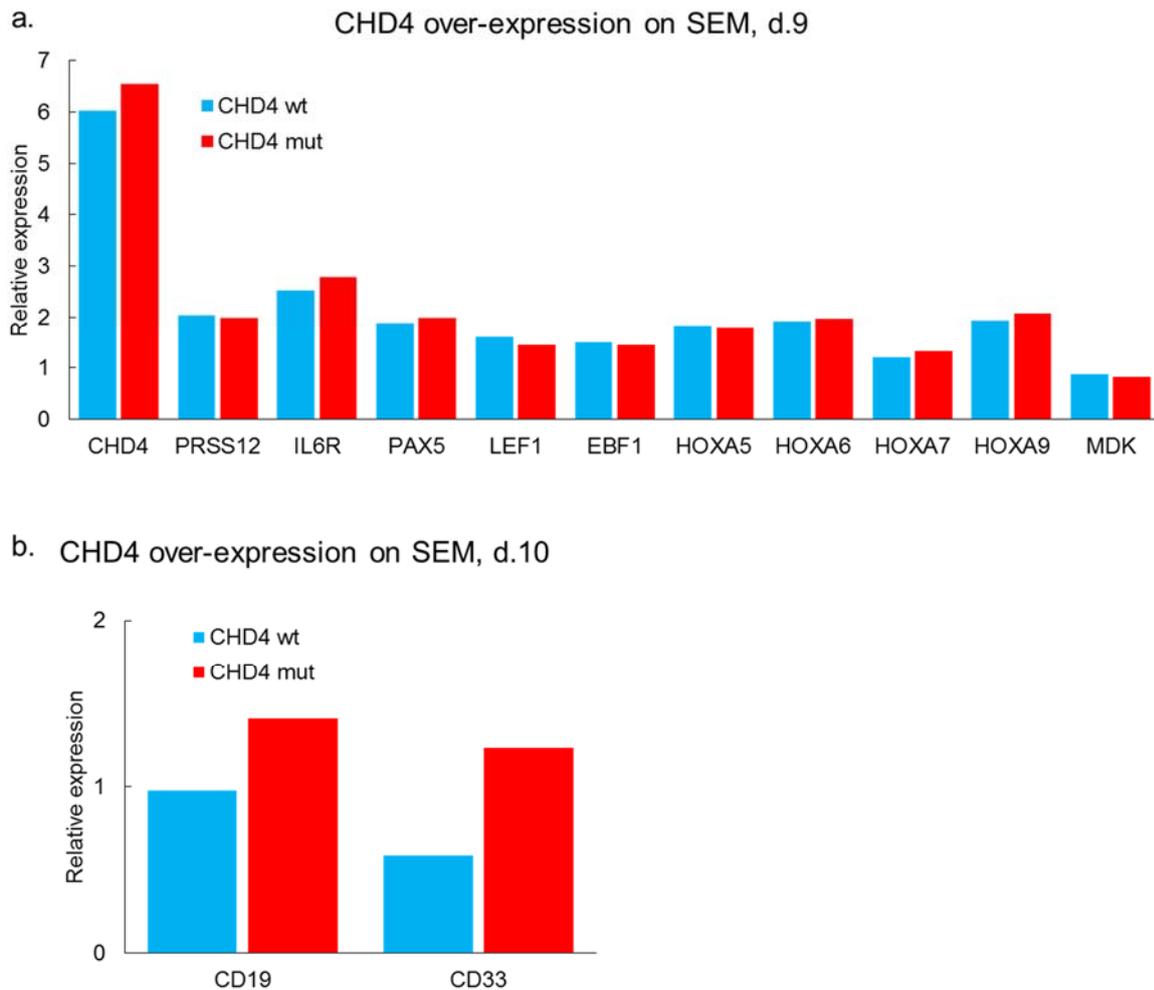


Figure 5-4 CHD4 over-expression effect on SEM cells transcripts. a) Assessment of *CHD4* wt and mutant on day 9 showed the increase by 6.0 fold and 6.5 fold, respectively. The over-expression was accompanied by higher expression of the other genes, apart from *MDK*, but no differential expression pattern was seen between the wild-type and mutant. b) The transcript of the lymphoid marker (*CD19*) and myeloid marker (*CD33*) showed a difference between the wild-type and mutant *CHD4*. *CD19* had an increase up to 1.4 fold on the mutant over-expression. A 40% decrease of *CD33* expression was observed upon *CHD4* wt, but not mutant over-expression. All data were relative to the empty vector control. The result was generated from a single experiment.

The analysis of the target gene transcripts showed their increase upon *CHD4* over-expression compared to the empty vector control. However, the increase had the same pattern between wild-type and mutant. It was continued by evaluating the lymphoid marker (*CD19*) and myeloid marker (*CD33*) expressions. A reduction by 40% was seen in the wild-type sample, but not on the mutant. Unexpectedly, *CD19* was expressed higher in the mutant sample. These data are preliminary results that need to be confirmed by repetitions and further validation with others methods (e.g. protein expression).

5.1.2.2. Co-IP CHD4 and EZH2

Although no consistent difference was observed in the CHD4 target gene expression, it was possible that the point mutation disturbed the interaction of CHD4 and its protein partners, rather than the interaction with DNA. One of primary protein binding partners of CHD4 is EZH2. Ezh2 is known to be required for early B cell development, studied in mouse model[218]. Additionally, EZH2 is also shown to be recruited by CHD4 during DNA-damage response[127]. In order to assess this hypothesis, over-expression and immunoprecipitation of CHD4 using its twin-strep-tag were performed on 293T cells that also included PPP1R7 wt (also has the twin-strep-tag) as a control. EZH2 binding by Co-IP was evaluated, shown in Figure 5-5.

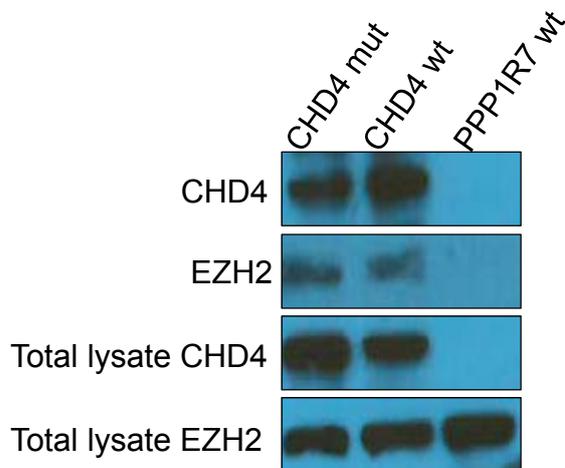


Figure 5-5 Co-IP EZH2 in CHD4 over-expression on 293T cells. CHD4 mutant, wt, and PPP1R7 wt (control) were over-expressed in 293T cells. All of the constructs contained the twin-strep-tag that was used for the immunoprecipitation. No difference of EZH2 Co-IP product was observed between CHD4 mutant and wt. CHD4 and EZH2 total lysate were used as positive control.

This result showed the binding of CHD4 and EZH2, but no difference between mutant and wild-type CHD4.

5.1.3. *PPP1R7* over-expression

Wild-type and mutant *PPP1R7* cDNA (1083 bp) were isolated by RT-PCR, using L826 relapse as the template. It contains the 1434G>T (accession number NM_002712.2) that created the R119L mutation. A twin-strep-tag was added to the N-terminal of the cDNA. The initial plan was to use LeGO iT2 lentiviral vector that contained tandem tomato fluorescent protein (tdTFP) as described in Figure 5-1. However, the transduction was only stable for few days (~8 days) in this vector, indicated by the expression of the fluorescent protein. Therefore, the gene was inserted into LeGO iV2 that contained venus fluorescent protein (VFP), instead, by using the following method:

- a. Digestion of the *PPP1R7* PCR product using *SacII* and *XhoI* to provide the restriction sites (the primers contain these additional sequences)
- b. Digestion of pEXPR-IBA105 twin-strep-tag vector using *SacII* and *XhoI* to provide the cloning sites
- c. Ligation of *PPP1R7* into pEXPR-IBA105 to obtain twin-strep-tag on N-terminal of *PPP1R7*
- d. Isolation of the *PPP1R7*-twin-strep-tag in pEXPR-IBA105 using *NotI* and *SacI*, where *SacI* was made blunt-ended
- e. Digestion of LeGO iV2 vector using *NotI* and *EcoRI*, where *EcoRI* was made blunt-ended to provide the cloning sites
- f. Ligation of *PPP1R7*-twin-strep-tag into the LeGO iV2 vector

The final construct was confirmed by Sanger sequencing. The vector map and mutated nucleotide sequencing are depicted in Appendix C.3.

The *PPP1R7* wt and mutant vectors were transduced into SEM cells. The transduction level was evaluated by FACS based on the VFP expression, as shown in Figure 5-6.

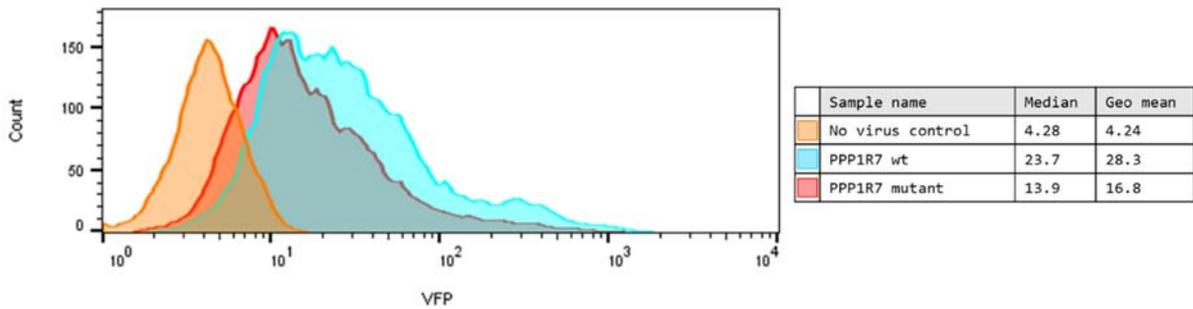


Figure 5-6 PPP1R7 over-expression evaluation by FACS on VFP expression. The wild-type and mutant PPP1R7 over-expressed SEM cells showed the positive VFP expression.

The genes that were differentially expressed between L826 presentation and relapse were assessed on the over-expressed cells, four days post-transduction by qRT-PCR, as shown in Figure 5-7a. Although higher expression of *PPP1R7* wt and mutant were seen on the graph, no differential expression was observed in the associated genes. The analysis was continued with a second experiment by looking at the second set of differentially expressed genes, including *ZHX3*, *ZFH3*, *HOXC5*, *HIPK1*, and *p16*, shown in Figure 5-7b. However, no different expression pattern between wild-type and mutant target genes were seen.

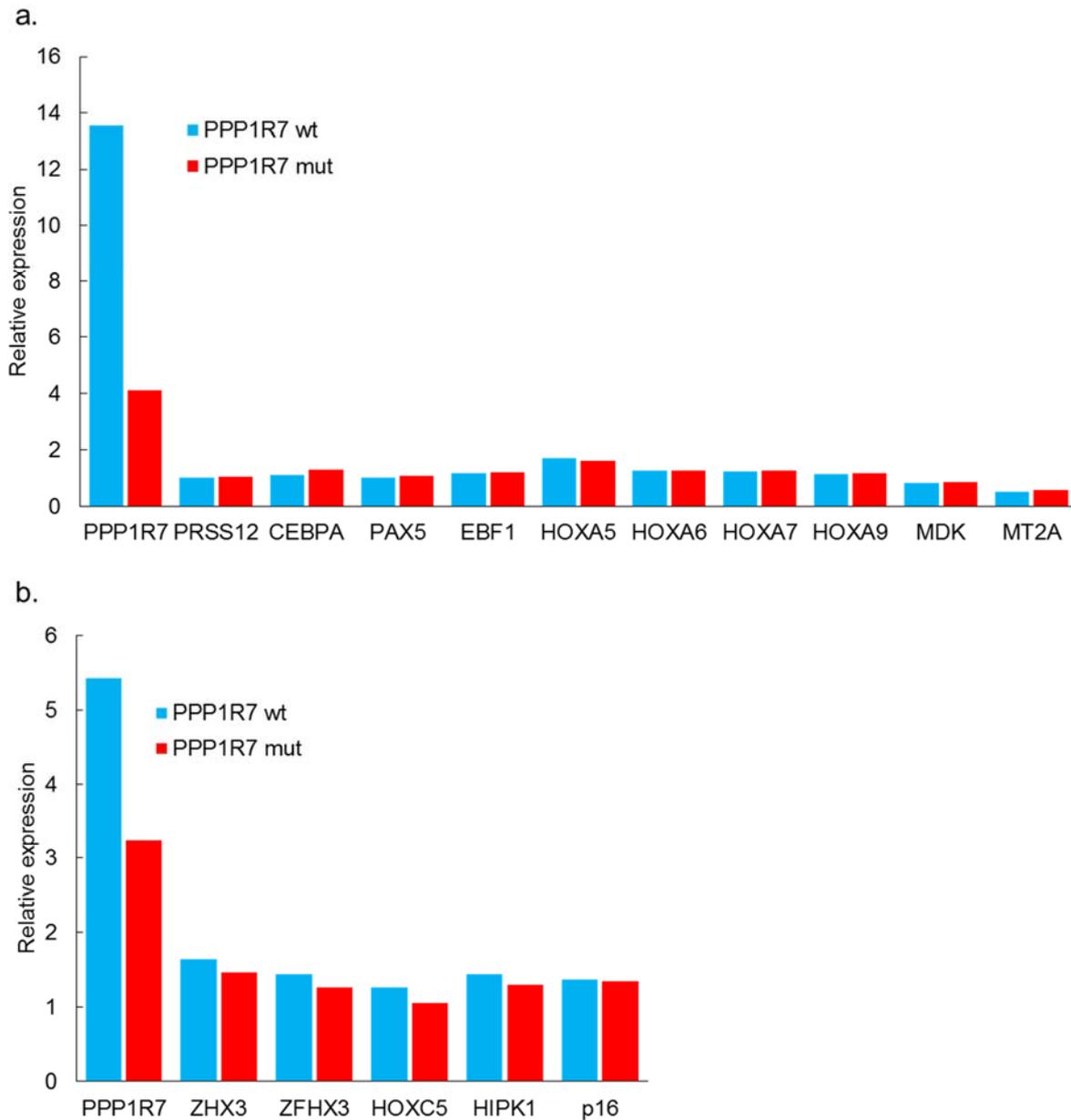


Figure 5-7 PPP1R7 over-expression effect on SEM cells transcripts. The genes that were differentially expressed between L826 presentation and relapse were evaluated after four days of transduction. Graphs (a) and (b) show two different experiments with different target genes.

The over-expression experiment was also performed on *ACAP1* that was inserted into LeGO iV2 lentivirus vector, again followed by qRT-PCR assessment of the same target genes. However, it again showed no difference between the mutant and wild-type gene. In addition, no growth curve difference was observed on any of these over-expressed genes. Therefore, the strategy was revised to take the approach of knocking-down the candidate genes, under the assumption that the mutation generates the loss of gene function. This approach was taken to address the

possibility that the SEM cells have reached the saturation or equilibrium of the candidate gene expression, so that over-expression may not produce an effect.

5.2. Knockdown candidate driver genes on SEM cells

The knockdown experiment was performed on seven of the candidate genes, including *PHF3*, *PPP1R7*, *ACAP1*, *CHD4*, *NCOA2*, *DHX36*, and *CEP164* by shRNA approach. Three constructs per gene were selected from The RNAi Consortium (TRC) shRNA library database. These sequences were inserted into lentivirus vector pLKO5d.SFFV.eGFP.miR30n (later the name is simplified to pLKO5d eGFP). Also, a non-targeting shRNA (shNTC) in pLKO5d.SFFV.RFP657.miR30n (simplified to pLKO5d RFP657) was provided as a control. All of the constructs were validated by Sanger sequencing. The sequences are listed in Table 2-19.

5.2.1. Knockdown levels

The analysis was performed on SEM cells. As the first step, the knockdown efficiency was evaluated by qRT-PCR, shown in Figure 5-8.

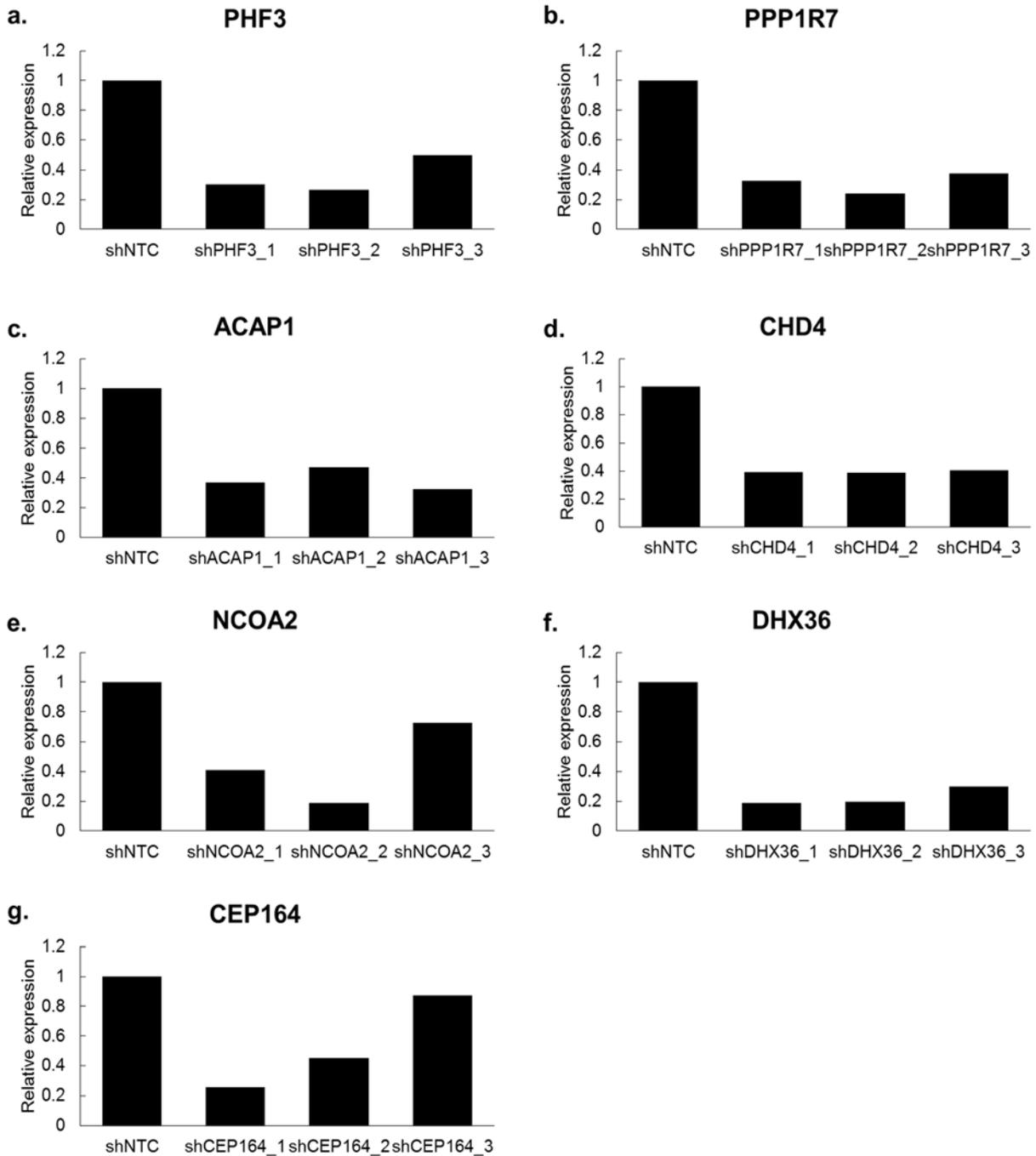


Figure 5-8 Candidate gene expression level after shRNA knockdown. Three different shRNA per gene were transduced into SEM cells. The transcript expressions were evaluated four days post-transduction. The y-axis was the relative expression of each particular gene according to the specific shRNA. The majority of the constructs reached $\geq 50\%$ knockdown. The expressions were relative to shNTC control.

The majority of the constructs yielded more than 50% knockdown of the target gene. It is also important to note that there was no effect on the growth curve observed on any of these genes compared with the shNTC. One construct/gene was selected for further analysis. They were:

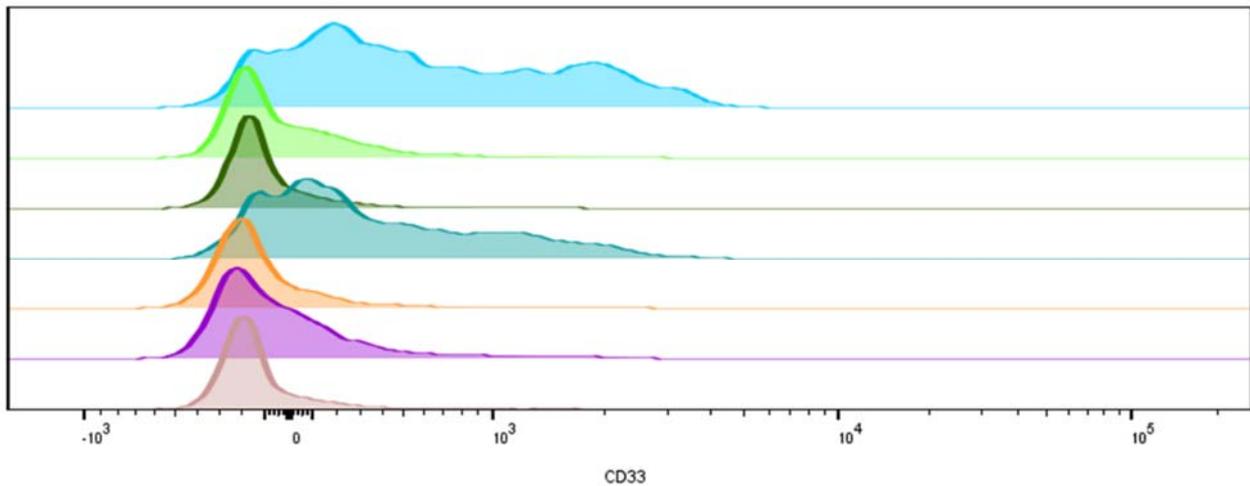
- shPHF3_2 with 73% knockdown
- shPPP1R7_2 with 76% knockdown
- shACAP1_3 with 67% knockdown
- shCHD4_3 with 59% knockdown
- shNCOA2_2 with 81% knockdown
- shDHX36_1 with 81% knockdown
- shCEP164_1 with 75% knockdown

These constructs were continued with immunophenotyping analysis.

5.2.2. Immunophenotyping of the knockdown cells

The most efficient shRNA constructs were selected and carried forward for evaluation of the lymphoid and myeloid markers (CD19 and CD33, respectively) by FACS.

Using this approach, higher expression of the myeloid surface marker CD33 was observed most convincingly in two samples, shPHF3 and shCHD4, summarised in Figure 5-9.



	Sample Name	Median : CD33	Geometric mean : CD33
	shPHF3_2	480	614
	shPPP1R7_2	-100	3.22
	shACAP1_3	-139	-82.6
	shCHD4_3	241	386
	shNCOA2_2	-168	-99.8
	shDHX36_1	-97.7	-6.76
	shCEP164_1	-170	-117

Figure 5-9 Flow analysis CD33 on knockdown candidate genes on SEM cells. The histogram showed CD33 expression that were gated for fluorescent protein-positive cells. The shPHF3 and shCHD4 samples had higher CD33 expression compared with the other samples.

This result demonstrated the expression of myeloid immunophenotype CD33 on ALL cells following knockdown of *PHF3* or *CHD4*. Confirmation by qRT-PCR using three different targeting constructs showed the increase of CD33 expression, despite stable expression of CD19, Figure 5-10.

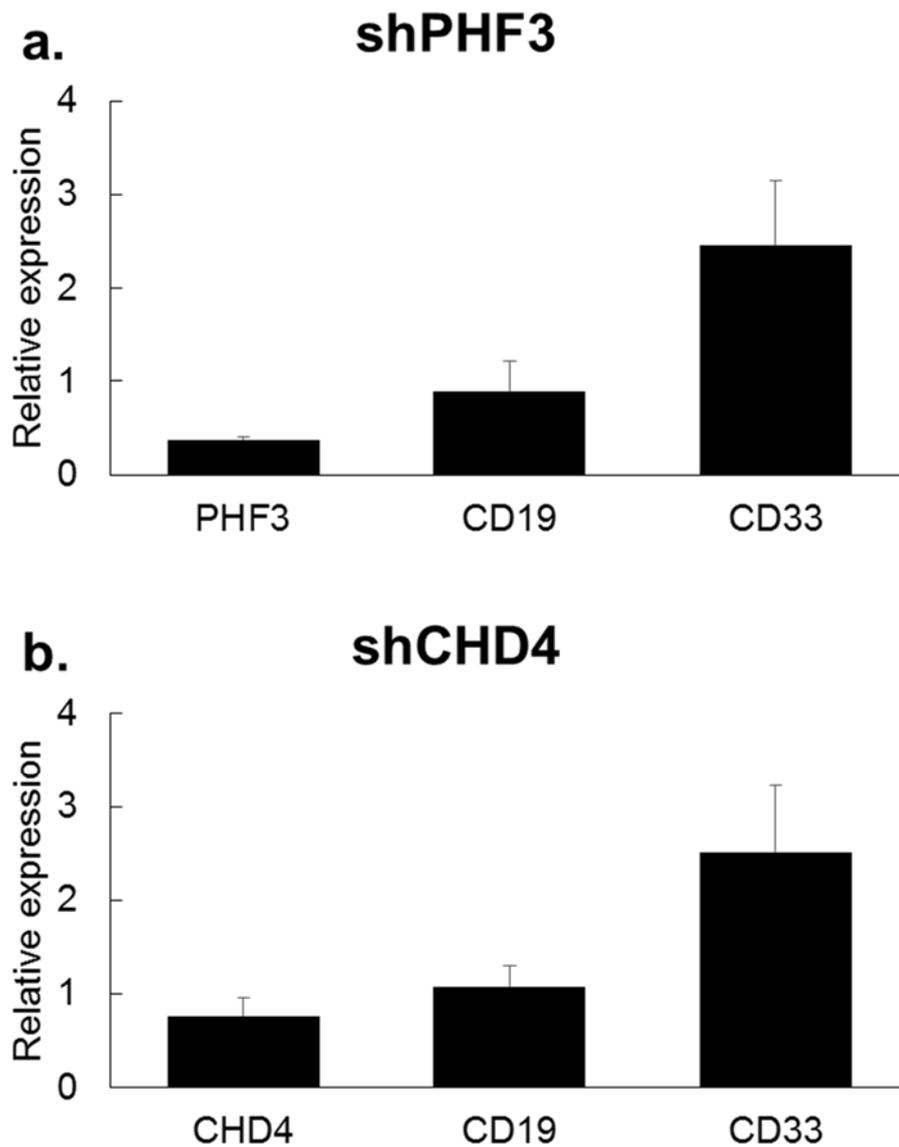


Figure 5-10 qRT-PCR following *PHF3* (a) and *CHD4* (b) knockdown. Knockdown of either *PHF3* or *CHD4* did not affect *CD19* expression. However, the myeloid marker *CD33* showed an increase of ~2.5 fold. Expression is relative to shNTC. The error bar shows the standard deviation from the three different shRNA constructs. Due to a technical issue, the three shCHD4 samples and one shPHF3 sample were analysed on day 7 post-transduction, and the other two shPHF3 constructs were analysed on day 8 post-transduction. The expressions were relative to shNTC.

Collectively, these results indicate a possible function for both *PHF3* and *CHD4* in controlling the haematopoietic lineage fate. These findings were further explored by studying the combined knockdown of both genes.

5.2.3. Combination *PHF3* and *CHD4* knockdown on SEM cells

After observing the increase of the myeloid marker by single knockdown *PHF3* and *CHD4*, the impact of simultaneous knockdown was investigated since both genes are concurrently mutated in L826 relapse. Therefore, the combination study of shPHF3, shCHD4, and shNTC was conducted. The shPHF3 constructs were transferred from pLKO5d eGFP to pLKO5d Tomato, the same backbone vector but containing the tomato fluorescent protein (TFP), instead of eGFP. These different fluorescent proteins allowed a combination transduction and differentiating each different knockdown population.

CD19 and CD33 surface markers were assessed on the combination knockdowns. Again, no change was seen in CD19 expression in this cell line model. However, as expected, the simultaneous knockdown showed even higher CD33 increase than the single knockdown, shown in Figure 5-11.

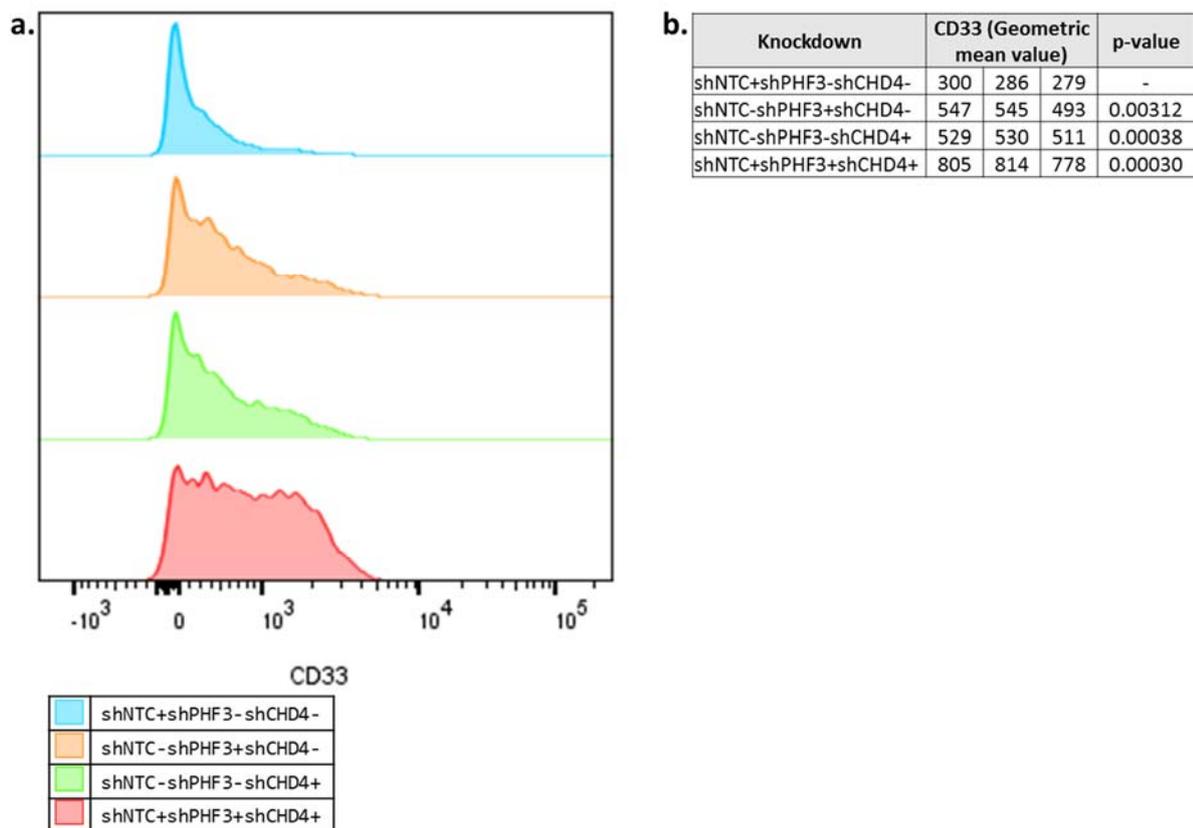


Figure 5-11 *PHF3* and *CHD4* combination knockdown effect on CD33 expression. a) The graphs show the single shNTC (blue), shPHF3 (orange), shCHD4 (green), and the combination of the shRNA (red). b) The average values of CD33 geometric means from three experiments. It shows more than two-fold increase in CD33 expression following the combination knockdown compared with the single shNTC+ sample. The p-values are generated from three different experiments, compared to the single shNTC+ sample. The data were collected 10 days after the transduction.

These results showed more than two-fold increase in CD33 expression following the combined knockdown of *PHF3* and *CHD4*, compared to the shNTC sample. It may indicate a co-activity of PHF3 and CHD4 in regulating the lineage commitment. There are at least two hypotheses can be proposed: (1) CHD4 and PHF3 interact with each other, since both of them have been described to be involved in DNA-damage response[127, 137] and (2) disruption of CHD4 changes the chromatin accessibility of haematopoiesis associated genes[132], where those genes are also controlled by PHF3 by its transcription elongation activity[136].

5.3. Knockdown PHF3 and CHD4 on CD34+ cord blood MLL/Af4 cells

PHF3 and *CHD4* knockdowns showed the increase of myeloid immunophenotype on the pro-B ALL cell line, SEM. However, no change was seen in the expression of CD19, a result which may reflect the intrinsic properties of this cell line model. To further investigate the role of these genes in lineage commitment, the knockdown of *PHF3* and *CHD4* was performed in a more complex model of haematopoietic progenitor cell development. In collaboration with Mulloy lab (Cincinnati Children's Hospital Medical Center), we received samples of a human CD34+ cord blood model, transduced with a chimeric *MLL/Af4* fusion gene[82]. The fusion gene consists of N-terminal *MLL* from human and C-terminal *Af4* from murine that were introduced by retrovirus vector. This model initially has a myeloid immunophenotype (CD33+ and CD19-). However, when cultured in lymphoid conditions (Section 2.2.2), the cells differentiate to express a lymphoid signature and surface immunophenotype, *i.e.* CD19+CD33-[82].

Initially, the ability to differentiate the model cells down both myeloid and lymphoid lineages was replicated. This experiment obtained the expected differentiation in the lymphoid culture, shown in Figure 5-12b.

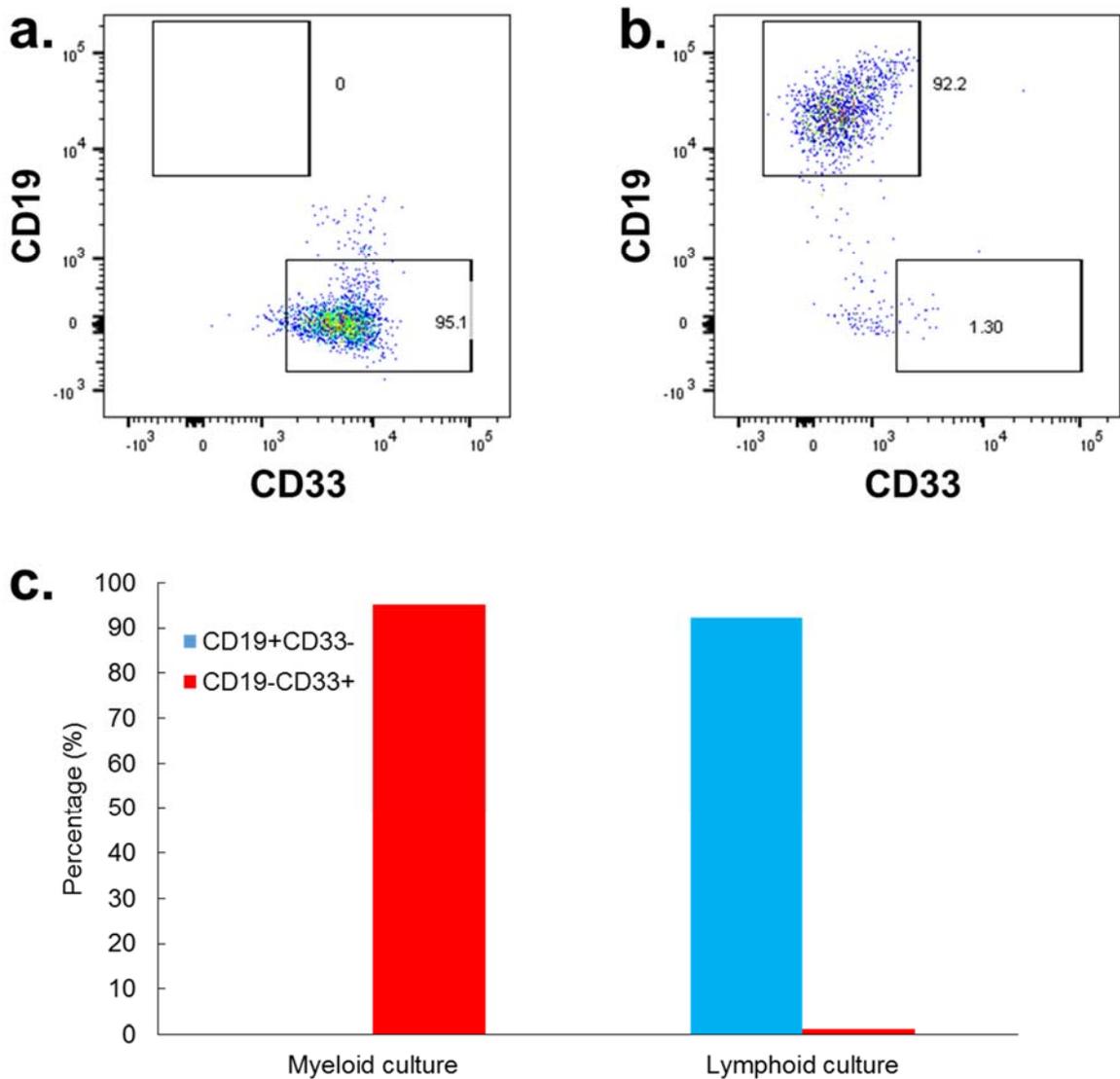


Figure 5-12 Flow analysis CD34⁺ cord blood MLL/Af4 cells in myeloid culture (a) and lymphoid culture (b), summarised in the graph (c). The cells in myeloid culture showed the CD19⁻CD33⁺ immunophenotype, while the cells in lymphoid culture had the CD19⁺CD33⁻ on almost all of the cells. The data were collected after 41 days in their specific culture conditions.

Following confirmation that the model was reproducible within our lab, the single and combination knockdowns were performed alongside the two differentiation conditions. Lymphoid and myeloid markers were assessed on day 24 following transduction. This analysis showed a reduced potential for lymphoid differentiation following single *PHF3* and *CHD4* knockdown compared with the shNTC. Once again, the lineage restriction was seen to be more pronounced following combination knockdown, with only a minimal lymphoid population identified, Figure 5-13.

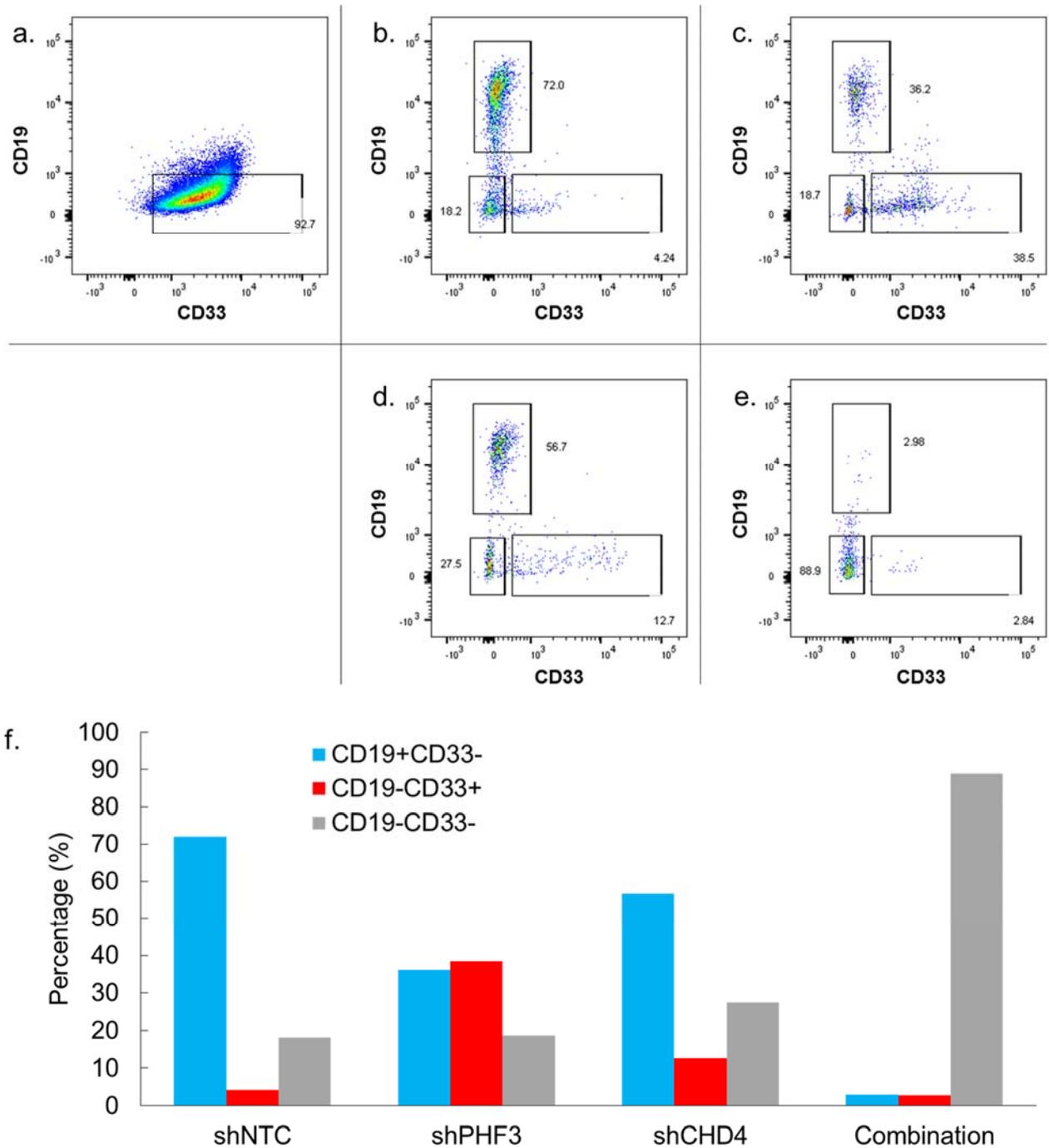


Figure 5-13 Flow analysis of CD34+ cord blood MLL/Af4 cells in myeloid culture control (a), and lymphoid culture with shNTC (b), shPHF3 (c), shCHD4 (d), and combination shNTC, shPHF3, and shCHD4 (e), and summarised in the graph (f). While the myeloid culture maintained the myeloid immunophenotype (CD19-CD33+) as shown in (a), the lymphoid culture lost the myeloid marker (CD33). The cells with shNTC (b) showed the lymphoid population (CD19+CD33-) as the majority (72%) of the population. This proportion was less in cells transduced with shPHF3 (c) and shCHD4 (d). Following the combination knockdown (e), almost no differentiation was demonstrated. The distribution of lymphoid and myeloid populations in lymphoid culture are summarised in the graph (f).

This result showed the majority of the cells (72%) differentiated into lymphoid population on the shNTC control sample. The shPHF3- and shCHD4-treated cells also presented the lymphoid population (36.2% and 56.7%, respectively), although in a lower proportion of cells than the shNTC. In keeping with this, the proportion of cells demonstrating myeloid differentiation was increased, despite the lymphoid culture conditions. The graphs also showed the population of CD19-CD33- which might be the transition phase from CD19-CD33+ to CD19+CD33-. Interestingly, the combination NTC, PHF3, and CHD4 knockdown resulted in the CD19-CD33- becoming the majority population (88.9%), with nearly no distinct populations of CD19-CD33+ (2.84%) and CD19+CD33- (2.98%). This implies an intolerable effect of combined knockdown of *PHF3* and *CHD4*, as opposed to heterozygous mutation, on lineage determination in this model

In order to obtain more information in the study, the flow analysis was recollected after extended culture period, d.32. This assessment was only successfully performed on shNTC and shPHF3 samples because the other two samples had lost their IRES-driven fluorescent protein expressions. Nevertheless, the analysis confirmed an even more pronounced myeloid differentiation following transduction with shPHF3, Figure 5-14.

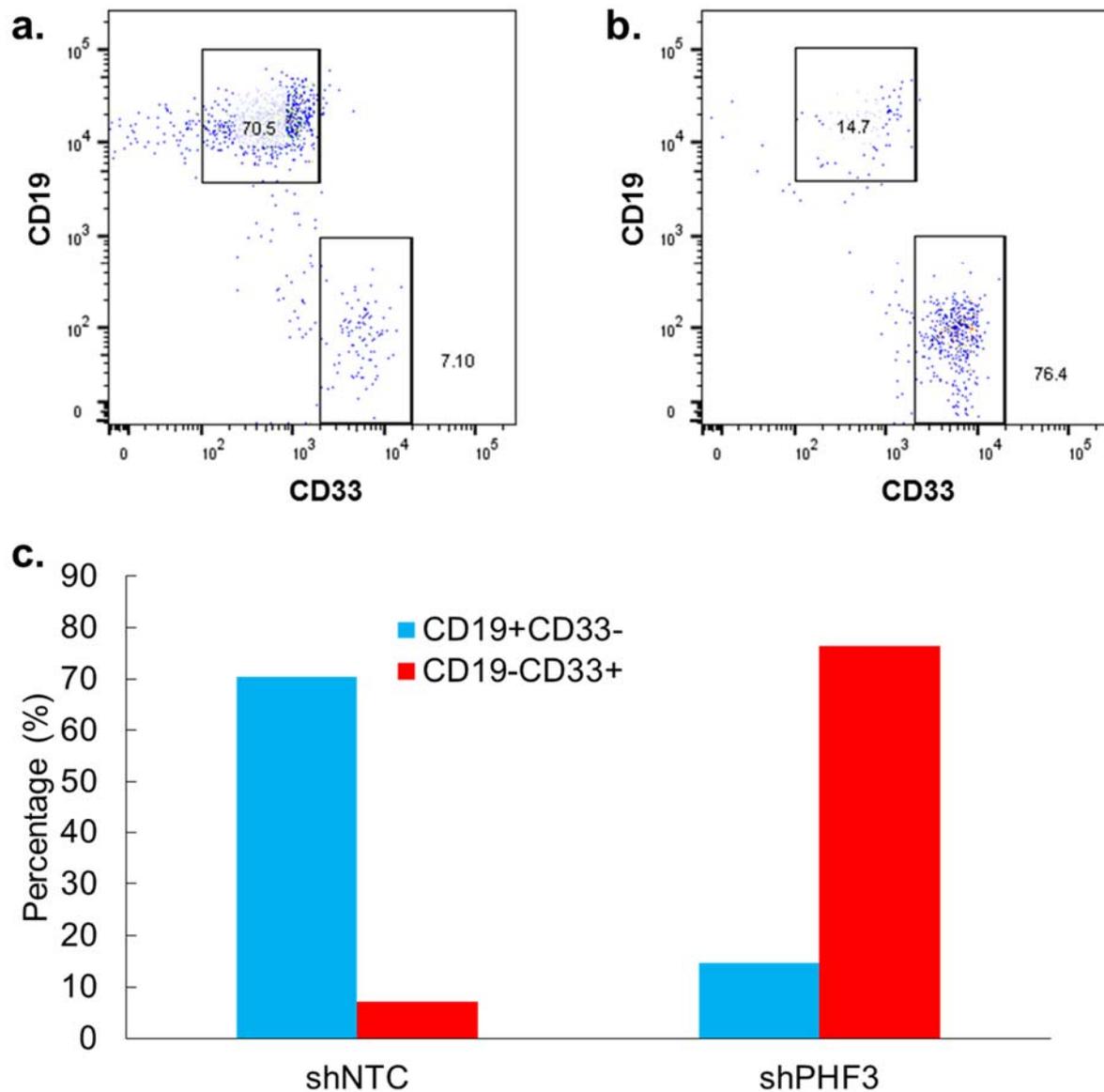


Figure 5-14 Flow analysis CD34⁺ cord blood MLL/Af4 cells with shNTC (a) and shPHF3 (b), 32 days post-transduction. a) Following transduction with shNTC, 70.5% of cells show a CD19⁺CD33⁻ lymphoid immunophenotype. In contrast, b) cells with PHF3 knockdown showed only a small proportion of lymphoid differentiation and were dominated by the myeloid population (76.4%).

These data demonstrated the two distinct populations, the lymphoid and myeloid populations, without a residual undifferentiated CD19⁻CD33⁻ population. This may support the hypothesis that the CD19⁻CD33⁻ population is a transient transition phase, in which *PHF3* knockdown diverts differentiation down the myeloid lineage.

In conclusion, down-regulation of *PHF3* skews differentiation potential away from the lymphoid lineage, towards the myeloid lineage. This effect was more pronounced in combination with *CHD4* knockdown.

5.4. Discussion

Firstly, it is necessary to note the interpretation difficulty in Section 5.1.1 (*PHF3* over-expression) data. The qRT-PCR on day 8 after transduction showed 4.7 fold and 7.4 fold increase of *PHF3* wt and mutant, respectively, but no CFP expression was seen by FACS analysis (n = 2). This might be because the large size of *PHF3* cDNA that reaches 6.2 kb inhibits the IRES-mediated translation process, as has been described previously[219]. Nevertheless, it should be able to be solved by replacing the IRES with 2A peptide that should translate both proteins equally[220].

Knocking down *CHD4* and *PHF3* showed an increase of the myeloid marker CD33, both in the transcript and protein on the SEM cell line, which was then confirmed on the CD34+ cord blood cells model. This model contains a hybrid of human *MLL* and mouse *Af4* that creates pro-B ALL in the transplanted mice, validated by the surface marker phenotypes, DNA binding sites, and gene expression signatures[82]. The results on these models indicate the role of both genes in the control of lineage commitment. More interestingly, the combination knockdown showed even greater impact.

CHD4 is known to have roles in the DNA-damage response. It is recruited to the damage site upon DNA DSB, SSB, and oxidative damage[122-127]. A study by Shanbhag *et al.*[221] demonstrated the inhibition of RNAP II elongation on the region of DSB damage site. Additionally, *PHF3* has been shown to interact with RNAP II upon UV irradiation[137], although its role remains unknown. These data may indicate a correlation between *CHD4* and *PHF3* at the DNA damage sites.

In addition to the RNAP II, the SPOC domain of *PHF3* may also support its association with *CHD4*. This domain was found in SHARP and has been shown to bind directly to HDAC1[222], a member of NuRD complex, of which *CHD4* is also an intrinsic subunit. It is possible that *PHF3* and *CHD4* may cooperate through their function with NuRD.

In regard to the hypothesis of *PHF3* and *CHD4* co-activity, perhaps several further investigations can be suggested. These would include Co-IP to assess their interaction, *e.g.* within NuRD, and finding the target DNA of *PHF3* (due to its TFIIS domain). Since *CHD4* acts a nucleosome remodeller, it may also be interesting to perform micrococcal nuclease assay to study the different nucleosome locations under different *CHD4* conditions (knockdown, rescue mutant and wild-type). Also,

very interestingly, CHD4 and PHF3 are known for being autoantigens in dermatomyositis[149, 150] and glioblastoma[136], respectively. It means that it may be possible to target it using engineered T cells. When the working mechanisms of both genes in acute leukaemia lineage commitment have been completely explored, it will be interesting to consider the possibility of coupling it with a haematopoietic marker for dual CAR T cells[223] or something more developed.

Chapter 6 PHF3 haematopoietic lineage control

The impact of PHF3 knockdown on the regulation of lineage specific differentiation in SEM and CD34+ cord blood MLL/Af4 cells was investigated in the previous chapter. This chapter aims to extend understand of the working mechanism of PHF3.

Different PHF3 transcript variants were explored from Ensembl, NCBI RefSeq, and the collaborative consensus coding sequence (CCDS) databases. This chapter will describe that the knockdown of specific *PHF3* variants increased expression of CD33, and subsequent knockdown of the other variants brought back the expression to baseline levels.

Two different cell lines in various settings were used in this study, however the results presented so far represent a single experimental replicate. Nevertheless, by understanding the working mechanism of PHF3, it is expected to provide the materials to benefit PHF3 for the therapy, as the final aim of this study.

6.1. PHF3 variants

There are at least nine *PHF3* transcript variants described to generate proteins (ENSG00000118482), and three of them are recognised in the CCDS, *i.e.* consistently represented on the NCBI, Ensembl, and UCSC Genome Browsers[224]. These three variants include NM_001290259 (ENST00000262043.7), NM_015153 (ENST00000393387.5), and NM_001290260 (ENST00000509330.5). In addition to this, the protein from the first two variants (*i.e.* NM_001290259 and NM_015153) has also been reviewed by UniProtKB/Swiss-Prot, as entry Q92576. Although the other variants have not been reviewed, all of them are aligned and depicted in Figure 6-1 to get more insight into PHF3. In this figure, each variant is also given an identifier name, to help identification in the subsequent text.

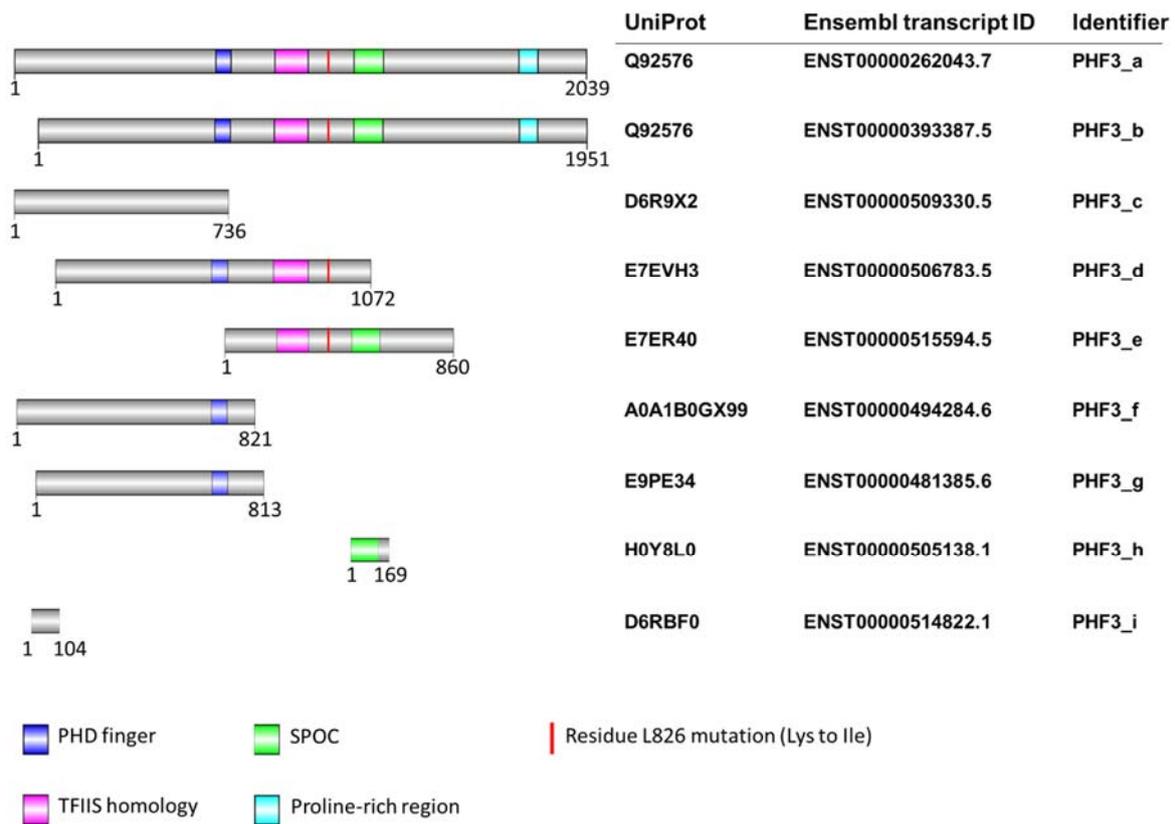
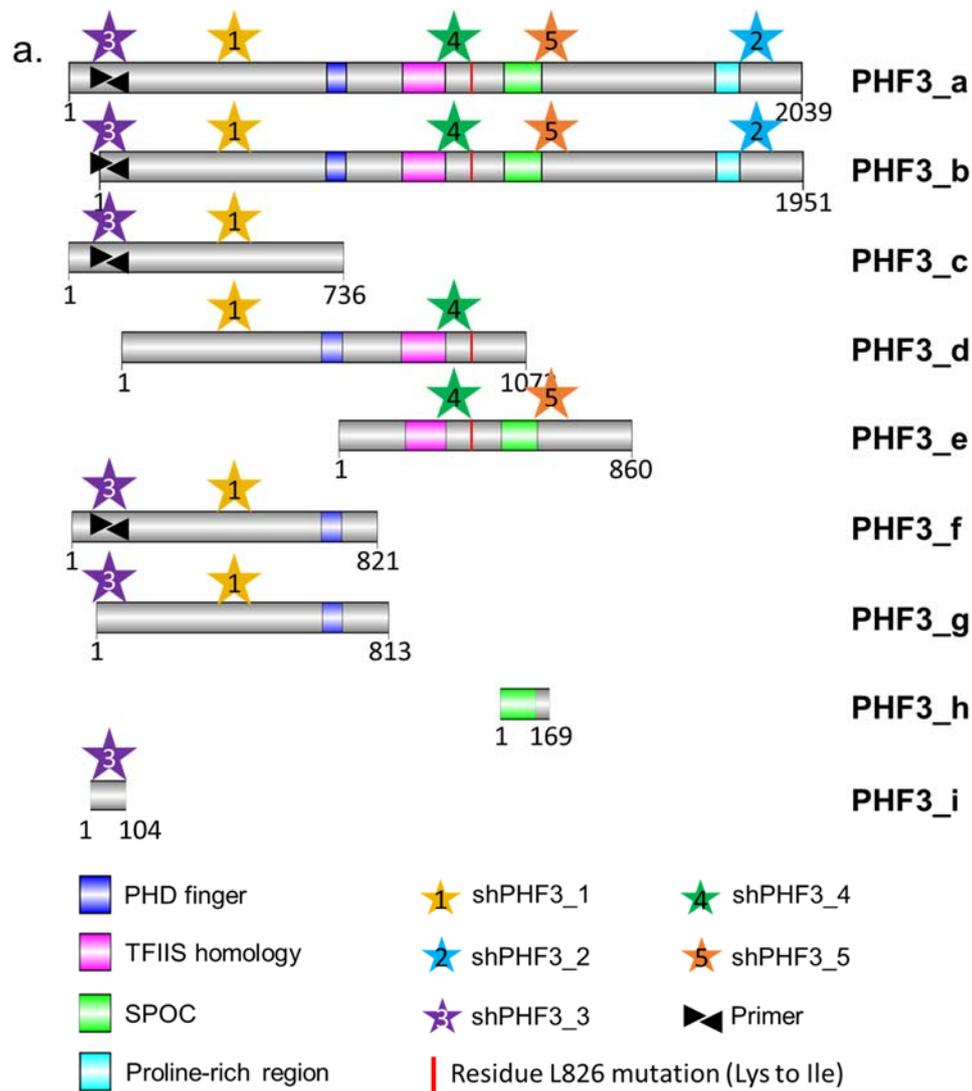


Figure 6-1 PHF3 variants. The protein schemes are constructed based on UniProt protein sequences. The 'identifier' column is generated for identification within the text. PHF3_a and _b are members of UniProtKB/Swiss-Prot. PHF3_a, _b, and _c are members of CCDS. PHF3_c and PHF3_d have an incomplete PHD finger domain and an incomplete SPOC domain, respectively, on their C-terminal end. The mutation in L826 is annotated with the red line, present only in PHF3_a, _b, _d, and _e.

The lysine to isoleucine mutation on L826 is not found in all of the variants, but only in PHF3_a, _b, _d, and _e. All of these mutated variants contain the transcription elongation factor, TFIIIS domain. Interestingly, no other variants have the domain. Three of these mutated variants also contain the SPOC domain, a domain that is known to interact with SMRT, NCoR, and HDAC1[222, 225]. Apart from PHF3_h, no other non-mutated variant have the SPOC domain.

Despite the complexity of the variant lengths and positions, several different shRNA constructs were designed (the sequences are listed in Table 2-19) to include different target combination. These constructs and their targets are depicted in Figure 6-2. It includes shPHF3_1 and shPHF3_3 that target nearly all of the variants, also shPHF3_2, shPHF3_4, and shPHF3_5 that differentially target the mutated variants alone (*i.e.* without any non-mutated variants).



b.

	shPHF3_1	shPHF3_2	shPHF3_3	shPHF3_4	shPHF3_5
PHF3_a	✓	✓	✓	✓	✓
PHF3_b	✓	✓	✓	✓	✓
PHF3_c	✓	–	✓	–	–
PHF3_d	✓	–	–	✓	–
PHF3_e	–	–	–	✓	✓
PHF3_f	✓	–	✓	–	–
PHF3_g	✓	–	✓	–	–
PHF3_h	–	–	–	–	–
PHF3_i	–	–	✓	–	–

Figure 6-2 shRNA constructs target different PHF3 transcript variants. a) Domain maps of *PHF3* isoforms showing position of shRNA construct target sequences. shPHF3_1 and shPHF3_3 target the majority of variants. Constructs shPHF3_2, shPHF3_4, and shPHF3_5 target only the L826 mutated variants. shPHF3_2 binds only the PHF3 isoforms with all domains present. shPHF3_4 targets all of the mutated variants, and shPHF3_5 targets three of the four mutated variants. The primer for qRT PCR, spanning an exon-exon junction, is indicated with the double-arrow symbol. PHF3_g and PHF3_i are not detected due to alternative splicing at this junction. b) Summary of target specificity.

These combinations were investigated started from SEM cells.

6.2. The effect of different PHF3 variants on myeloid marker expression

The five shRNA and one shNTC constructs were inserted into the pLKO5d lentiviral vector and transduced into SEM cells. The RNA was collected for analysis of the myeloid marker *CD33* on day 7 post-transduction. This demonstrated that the three constructs targeting the L826 mutated variants (*i.e.* shPHF3_2, _4, and _5) resulted in a higher *CD33* expression, whilst the other two constructs did not, Figure 6-3.

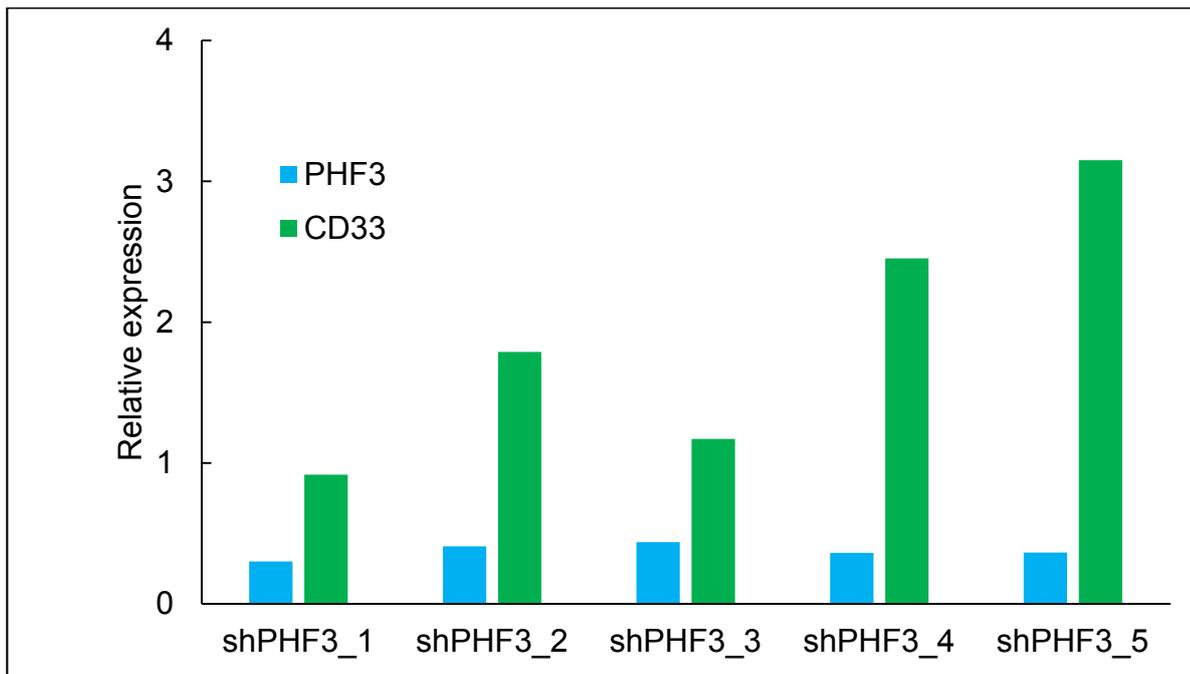


Figure 6-3 qRT-PCR knockdown different PHF3 variant on CD33 on SEM cells, day 7. The constructs shPHF3_2, shPHF3_4, and shPHF3_5 that target the L826 mutated variant showed the increase of *CD33*. The rise was not observed on the shPHF3_1 and shPHF3_3 constructs that target the whole variants. The expression levels are relative to shNTC.

The graph showed the increase of *CD33* only on shPHF3_2, shPHF3_4, and shPHF3_5 (the constructs that target L826 mutated variants), but no effect on the shPHF3_1 and shPHF3_3 (the constructs that also target the non-mutated variants). This result provides a hint that the PHF3 variants that contain TFIIS and/or SPOC are required for the *CD33* regulation.

6.3. Downregulating CD33 in SEM CD33+ cells

Firstly, the knockdown of *PHF3* L826 mutated variants showed an increase in *CD33*, whilst knockdown of all variants did not. Secondly, the flow analysis in Section 5.2.2 proved the increased *CD33* surface protein expression on SEM cells. These results raised the question of what would happen on the *CD33* surface protein expression if

the knockdown of all *PHF3* variants was performed on the CD33+ SEM cells? Was it related to the balance of the different *PHF3* variants? Therefore, a sequential second transduction (*i.e.* not a co-transduction) was conducted on those cells that included combinations:

- shNTC only as the control for CD33- cells
- shPHF3_2 + shNTC
- shPHF3_2 + shPHF3_1

Interestingly, the flow analysis on day 15 after the second sequential transduction showed a substantial reduction of CD33 expression on the shPHF3_1 treated cells, summarised in Figure 6-4.

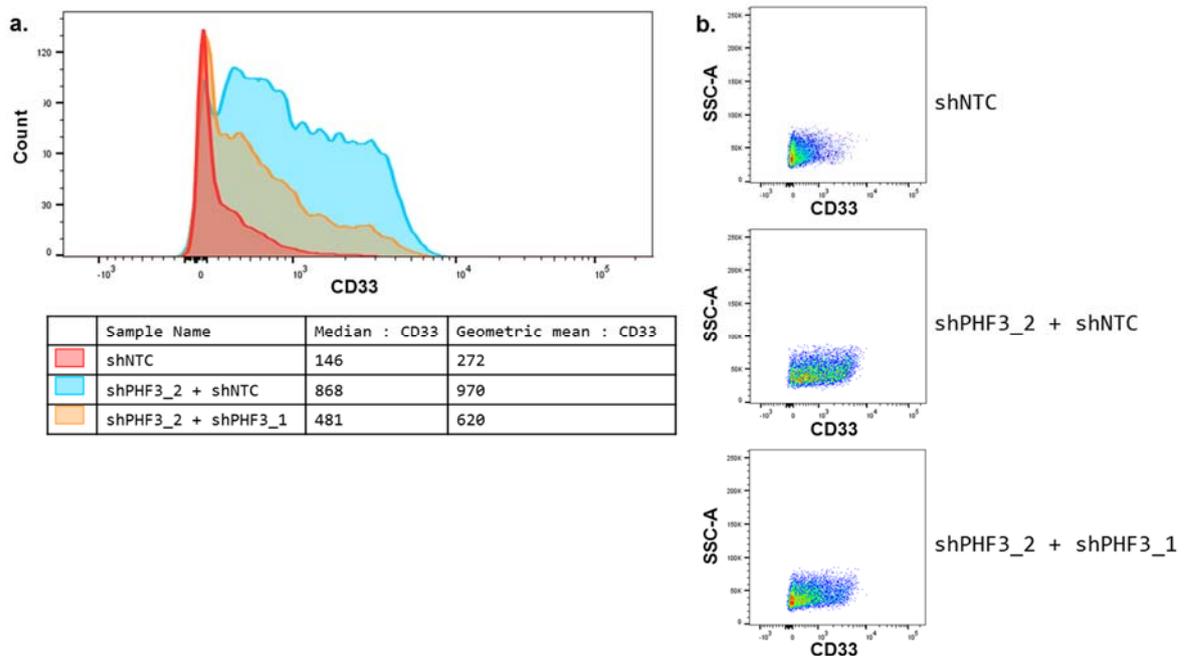


Figure 6-4 CD33 expression following differential *PHF3* variant knockdown in SEM cells. a) The single transduction shNTC had minimal CD33 expression (red curve). The expression increased substantially with the introduction of shPHF3_2 (blue curve). Sequential transduction shPHF3_1 54 days after the first transduction showed a reduction ~50% CD33 expression (orange curve). The analysis was collected 15 days after the second transduction. b) Pseudocolor plot of CD33 expression by flow cytometry.

This result indicates the requirement for balanced *PHF3* isoforms in the lineage maintenance. Downregulating the *PHF3* variants containing TFIIS and SPOC domains drives the myeloid lineage commitment. In contrast, downregulating *PHF3* variants not containing the two domains either inhibits the myeloid differentiation or drives the lymphoid lineage commitment.

6.4. *PHF3* knockdown on the AML cell line

A different study model may provide more information regarding the general understanding of *PHF3*. Knockdown of different *PHF3* variants was performed in Kasumi-1, a t(8;21) AML cell line to study if there any effect in a myeloid environment. The constructs shPHF3_1, shPHF3_4, and shPHF3_5 were transduced into the cells, and their CD33 expression was measured by FACS, shown in Figure 6-5.

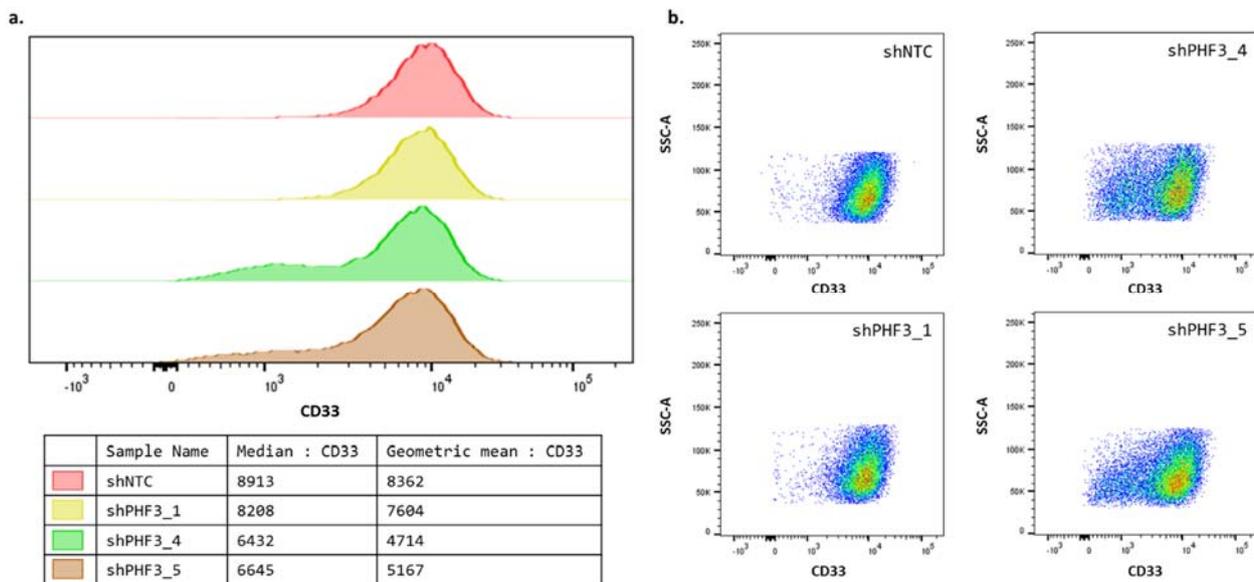


Figure 6-5 CD33 expression following knockdown of different *PHF3* variants in Kasumi-1 cells.

a) Histogram showing shNTC and shPHF3_1 produced no change in CD33 expression. However, shPHF3_4 and shPHF3_5 constructs result in reduced CD33 expression. b) Pseudocolour plots show reduced CD33 expression following shPHF3_4 and shPHF3_5 transduction, with a distinct CD33-population identified (right plots).

Unexpectedly, shPHF3_1 that reduced CD33 expression on SEM did not show any changes in Kasumi-1, when compared with shNTC. However, the constructs that increased CD33 in SEM (shPHF3_4 and shPHF3_5) reduced CD33 in Kasumi-1 cells. Although these data are coming from a single experiment, they indicate the role of *PHF3* in haematopoietic cell identity. They identify the impact of isoform on myeloid lineage commitment, highlighting the importance of TFIIS- and SPOC-containing *PHF3* variants.

6.5. Discussion

The findings in this chapter are summarised (1) the point mutation found in L826 is present only within *PHF3* variants containing TFIIS and SPOC domains, (2) downregulating these variants increased CD33 on SEM but decreased CD33 on

Kasumi-1, and (3) downregulation of all variants in the experimental CD33+ SEM population reduced the CD33 expression back towards baseline. These preliminary data suggest that an imbalance of different PHF3 isoforms drives the lineage infidelity or PHF3 is required for myeloid cell identity.

The functional mechanism of PHF3 remains unknown. However, these data support a role for the TFIIIS and/or SPOC domains.

Kinkelin *et al.*[181] studied in detail the structure of Bye1, a homologue of PHF3 from *Saccharomyces cerevisiae*. They demonstrated the TFIIIS-like domain of Bye1 bound directly to the RNAP II, also showed that the surface forming the interface between the species was conserved as in human RNAP II and PHF3[181]. This interaction is supported by Boeing *et al.*[137] who showed the binding of PHF3 and RNAP II upon UV radiation. In addition to these two studies, the opposing effect of PHF3 knockdown on CD33 expression in SEM and Kasumi-1 cells may result from the interaction of PHF3 with super-enhancers that define the cell identity[226]. One of the predictions is that PHF3 may be a member of the super-enhancers via its TFIIIS – RNAP II interaction. Or, as discussed in Section 5.4, there is a possibility of PHF3-HDAC1 interaction via its SPOC domain[222], with a consequent impact on NuRD complex function and subsequent lineage specific transcription via super-enhancers[226].

The other interesting finding in this study is the decrease of CD33 expression on CD33+ SEM cells upon knockdown of all PHF3 variants. It may indicate there is an inhibitor version among the various PHF3 variants. It may be possible since this functional protein-inhibitor-single gene setting has also been identified at least in Tn5 transposase (Tnp) and inhibitor (Inh), although it is in prokaryotes[227-230]. The *Tn5* consists of the full-length *Tnp* transcript and another transcript that loses 55 amino acids on the N-terminal, named as *Inh*. While the Tnp acts as a transposase by binding to the target transposed DNA via its N-terminal 55 amino acids, the Inh inactivates the Tnp by forming mixed oligomers with the Tnp[227]. It is suspected a similar mechanism of full-length PHF3 that is counteracted by the other isoforms. Perhaps mechanisms such as binding competition to the target DNA, the formation of direct interaction inhibition among the isoforms, or competition with some host factors that are necessary for the PHF3 function can be proposed for further investigation of this phenomenon.

Lastly, these data remain preliminary. Broader and deeper approaches are required to generate robust conclusions. It may be started with over-expression of certain transcripts, certain domain, or DNA region deletion (*e.g.* by CRISPR) experiment, verification of the protein target binding (*e.g.* co-immunoprecipitation PHF3, RNAP II, HDAC1, SMRT, NCoR), identification of any DNA target binding, if present (*e.g.* CHIP-seq), and then overlaying the information with L826 DHS and RNAseq data. Also, the role of MLL/AF4 should not be excluded since it may also be possible that the increase of CD33 expression was related with this mutation.

Chapter 7 Concluding discussion and future works

Different MLL fusion partner genes strongly specify a certain lineage of leukaemia. MLL/AF4 leads to ALL in more than 90% of the cases regardless of age[67]. Other fusion partners such as ELL, AF1Q, SEPT6, and SEPT9 result in AML in majority of the cases[67]. In addition to this, the age of diagnosis also influences the lineage of the leukaemia, but only in certain fusion partners, especially MLL/AF9. ALL has been shown to occur in 71% of infant MLL/AF9 leukaemia, but the trend changes to 69% and 90% of AML MLL/AF9 cases in paediatric and adult, respectively[67]. These data suggest the potential of MLL and its fusion partner as a master regulator of lineage instruction. Here in the lineage switch patient L826 it was shown that the ALL MLL/AF4 was skewed by the presence of the mutated driver genes.

This study showed the presence of a common cell of origin between ALL presentation and AML relapse in the MPP population. In addition to harbouring the MLL/AF4, the MPP population also carried PHF3 and CHD4 mutations. The functional studies showed that both genes could change the lymphoid differentiation towards the myeloid lineage.

The immunoglobulin rearrangement analysis suggested that the relapse L826 arose from a common origin prior to the B cell development stage. The haematopoietic stem/progenitor populations sorting followed by evaluating the presence of *MLL/AF4* indicated the CD34+CD38-CD90- (MPP) as the initial progenitor population harbouring the fusion gene, both in the presentation and relapse. However, the analysis on primograft ALL-derived xenograft L826 showed not only the presence of HSC as the majority of the haematopoietic stem/progenitor populations but also the positivity of fusion gene in this population. These results suggested (1) the enrichment of the MLL/AF4 harbouring population in primograft model, (2) possibility that the fusion gene was present on HSC of L826 primary cells that were not able to be detected by the assay, or (3) reversion to HSC.

Although differentiated haematopoietic cells were initially determined to be incapable of cell reprogramming, this paradigm had now been overturned by several experimental manipulations to suggest that cells can dedifferentiate or trans-differentiate the committed haematopoietic cells[231]. These ectopic conditions included introducing cytokine receptors[232-234] and different transcription factors,

such as PAX5[235, 236] and CEBPB[237] which allow the committed cells to traverse backwards and forwards in the differentiation hierarchy. The reprogramming of the committed cells was further studied by Bonifer et al who observed the epigenetic profile of the haematopoietic precursors[231]. They suggested that (1) cell fate determination occurred gradually at the epigenetic level and (2) different early progenitors still showed epigenetic similarities[231]. The presence of lineage-specific activators (*e.g.* PU.1, PAX5) changed the chromatin structure of certain lineage-specific genes which then allowed the cells to differentiate[231]. In other words, their model suggested that instead of being hierarchical, the early haematopoietic progenitors constituted a pool of epigenetically dynamic cells[231]. This stem cell plasticity characteristic is in-line with the well-studied plasticity of epithelial stem cells (reviewed in [238]). These studies infer that it might be possible that the pre-leukaemic population in HSC primograft L826 presentation was due to reversion of the pre-leukaemic progenitor cells, possibly due to the microenvironment of the mouse.

The pre-leukaemic population study was extended by including four t(4;11) and two t(9;11) cases. The *MLL/AF4* was found in CD38-CD45RA+ in two and GMP on the other two of t(4;11) cases. Interestingly, in t(9;11), *MLL/AF9* was only found in the lymphoblast cells in L880 case and no classical haematopoietic progenitor cells on the other case, LK271. It suggests *MLL/AF9* is located in the more mature population, or it may disrupt the canonical haematopoietic differentiation pathway if it is found on earlier progenitor populations. Future work to evaluate the different cellular profiles between *MLLr* and *MLL wt* progenitor cells may provide a new system to test drug efficiency in order to eliminate both leukaemic and pre-leukaemic cells.

Further studies on the relapse progenitor cells showed co-mutation PHF3, CHD4, and the fusion gene. These secondary mutations were presented on MPP, CD38-CD45RA+, CMP, GMP, and monocytes, but not on T cells, B cells, and NK cells, *i.e.* only on the population containing myeloid lineage differentiation potentials, but not on the lymphoid lineage. The roles of these genes were confirmed by knockdown experiment on t(4;11) ALL, SEM cells. The myeloid marker CD33 increased in the transcript and protein upon the individual gene knockdown and even greater impact on the combined knockdown of both genes. Moreover, the results were further validated in the CD34+ cord blood cells harbouring chimaeric *MLL/AF4*. The

knockdown of PHF3 and CHD4 prevents the lymphoid differentiation, *i.e.* maintain the myeloid immunophenotypes on this model, and once again, the lineage restriction was more pronounced following the combination knockdown.

CHD4 is known as a major subunit of NuRD complex that plays a role in nucleosome mobilisation through its ATPase domain[239, 240]. The mutation in L826 (R1068H) is located in this domain. Further study by looking at the ATPase function of mutated CHD4 may suggest its role in the lineage switch mechanism. Recently, Morra *et al* (2016) described the methods of ATPase assay specifically in CHD4[240] that may need to be reproduced in the L826 case. Also, micrococcal nuclease assay, in which the micrococcal nuclease cut the DNA within nucleosome linker regions, *i.e.* at the accessible region between nucleosomes[241], may provide further insight of mutation CHD4 role. Since CHD4 induces nucleosome movements (*i.e.* nucleosome mobilisation) to provide the accessible (euchromatin) or inaccessible (heterochromatin) DNA[239, 240], the micrococcal nuclease assay is expected to provide the information chromatin pattern between mutant and wild-type CHD4.

Although PHF3 has not been widely studied, its protein domain has been defined[136]. The different variant study in Chapter 6 showed particular splicing variants of PHF3 had a different effect on CD33 expression, while that effect was compromised by the other variant. However, its regulation mechanism remains a question. Domain studies may need to be explored. Homologous recombination CRISPR to delete certain domain (PHD, TFIIS, SPOC) and/or add a tag protein may provide the initial step of the study. The TFIIS and domain may suggest its role in transcription elongation[181]. Also, its SPOC domain may interact with HDAC1[222] which then relates it to nucleosome remodelling complex.

Chd4 deletion in mouse bone marrow has been established[132]. This study showed early loss of lymphoid and myeloid cells, compensated by erythroid cells accumulation. However, *in vivo* PHF3 deletion study has not been found. Investigation on the PHF3 knockout model may provide an essential understanding of PHF3 function.

Since the lineage restriction was shown to be more pronounced in combination *PHF3* and *CHD4* deregulation on SEM and cord blood cell models, it may suggest their cooperation. A proposed hypothesis is CHD4 converts certain lymphoid- or myeloid-essential regulatory regions from heterochromatin to euchromatin, and that area

becomes accessible for transcription initiation and elongation, regulated by PHF3 (Figure 7-1).

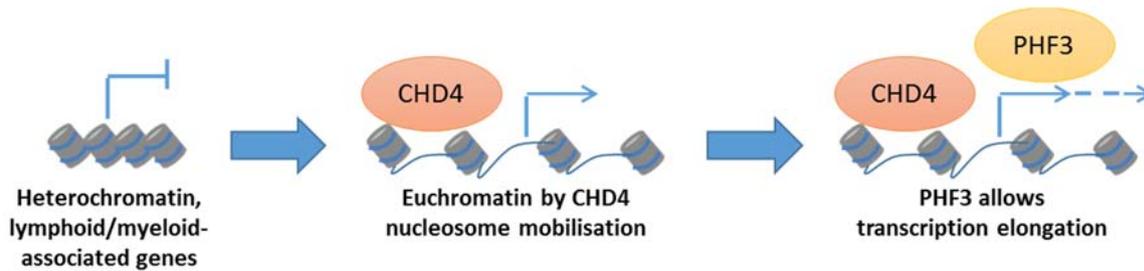


Figure 7-1 Model of CHD4 and PHF3 cooperation. CHD4 induces nucleosome movement that changes the chromatin density from heterochromatin (inaccessible) to euchromatin (accessible) form. The accessible form allows the transcription initiation complex to transcribe the DNA until reaching the pause site. At this site, the presence of PHF3 that contains TFIIIS domain may allow the RNA polymerase II to continue the transcription, *i.e.* elongation process.

Based on this hypothesis, it will be interesting to have ChIP-seq data of CHD4 and PHF3 on the haematopoietic cells in order to observe their DNA binding region, whether they have the same binding target related to the haematopoiesis or leukaemogenesis.

Another hypothesis for cooperation of PHF3-CHD4 is based on their interaction with HDAC1, another member of NuRD complex. CHD4 is widely known to directly interact with HDAC1 as in the NuRD complex[242]. And, as mentioned earlier, PHF3 may also interact with HDAC1 through its SPOC domain. In addition to this interaction, CHD4 has also been shown to interact with MLL[243], which raises the question if there is any correlation between CHD4 mutation and MLL or MLL/AF4 in lineage switch. It might be interesting to perform an immunoprecipitation of CHD4, PHF3, and MLL to observe the interaction among them, which then may suggest their mutual work in lineage commitment.

Furthermore, both PHF3 and CHD4 have been shown to be recruited into the DNA-damage lesion, which suggests their roles in DNA-damage response[122, 125, 127, 137]. It encourages a further study of drug testing using DNA-damaging agents, *e.g.* PARP or ATR inhibitor. It is expected that the mutations will increase the sensitivity of the drugs if those mutations impair the repair mechanisms.

Lastly, both CHD4 and PHF3 have been found as autoantigens in dermatomyositis[149, 150] and glioblastoma[136], respectively. The presence of

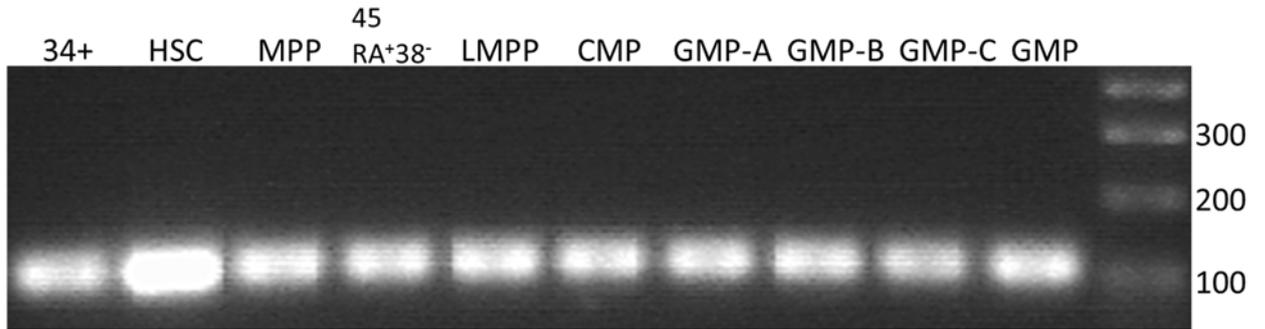
PHF3 autoantibodies in the glioblastoma cases had higher median survival time 14.5 months compared with 7.2 months in the cases without the autoantibodies. These data may indicate (1) the potential of the immune system to target CHD4 and PHF3 in the transformed cells, and (2) that function yields a higher survival rate. It will be interesting to isolate and/or expand the tumour-specific lymphocytes to target the tumour/leukaemic cells, such as via cancer vaccines[244].

Appendix

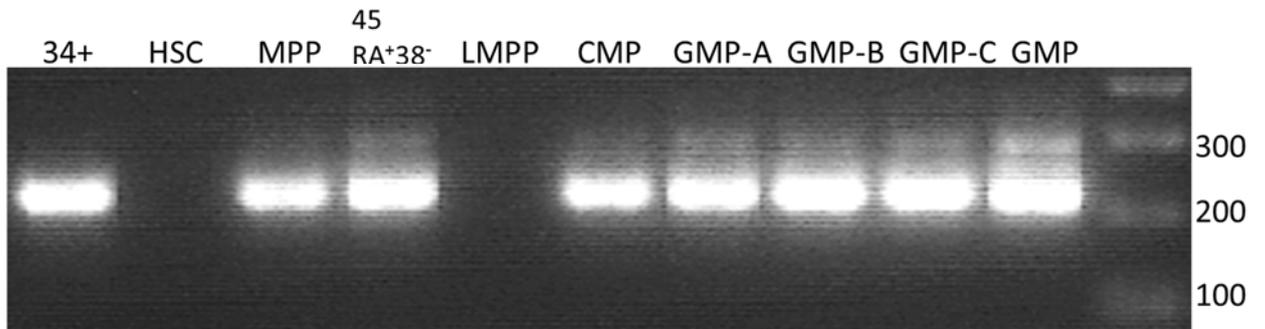
Appendix A PCR MLL/AF4 L826 on different haematopoietic populations

A.1 L826 Presentation

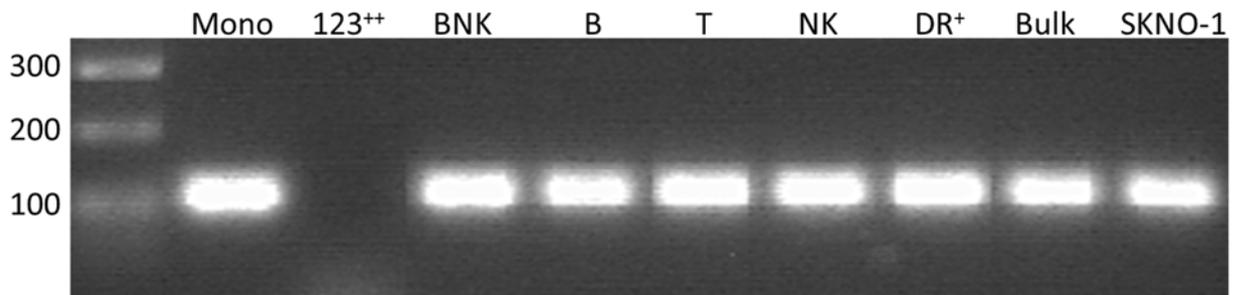
Nested PCR ATP10A



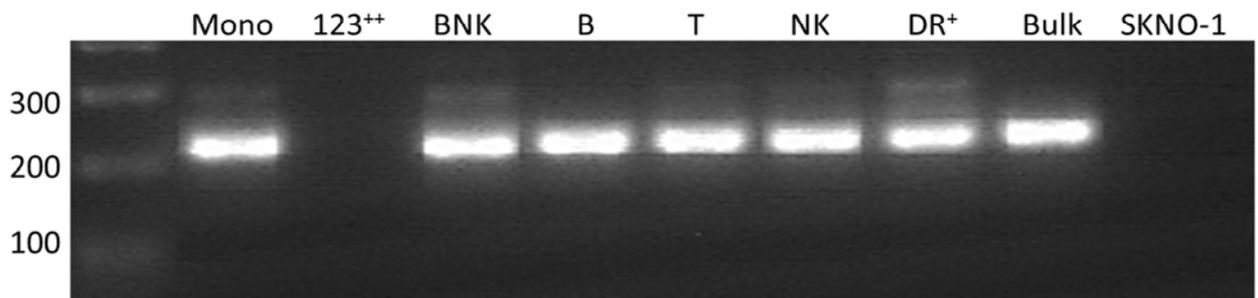
Nested PCR MLL/AF4



Nested PCR ATP10A

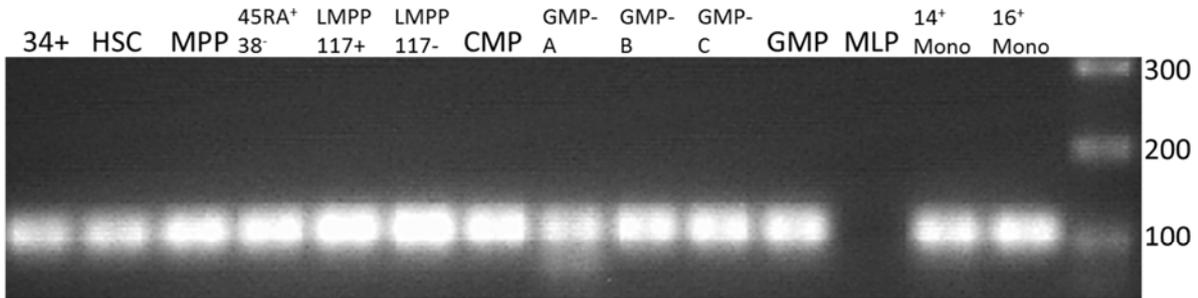


Nested PCR MLL/AF4

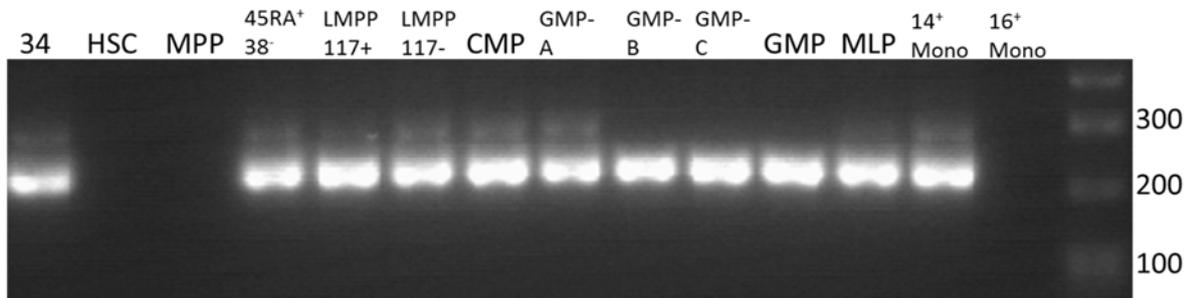


A.2 L826 Relapse and L826 ALL primograft

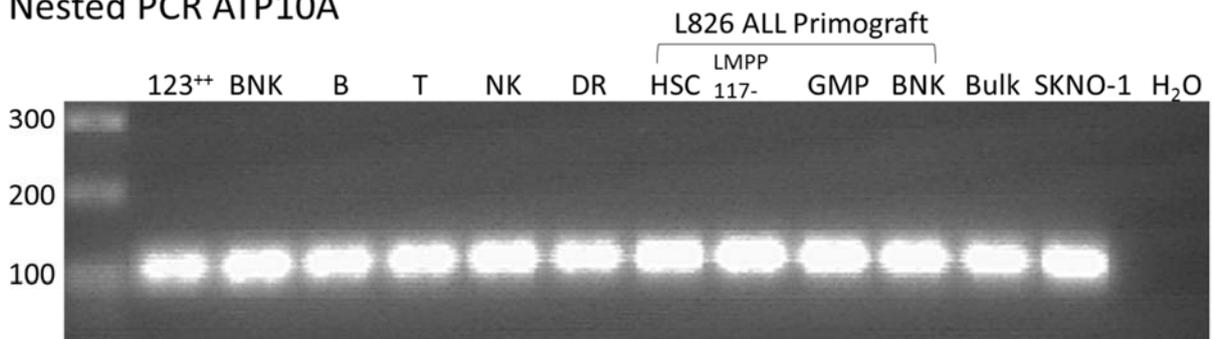
Nested PCR ATP10A



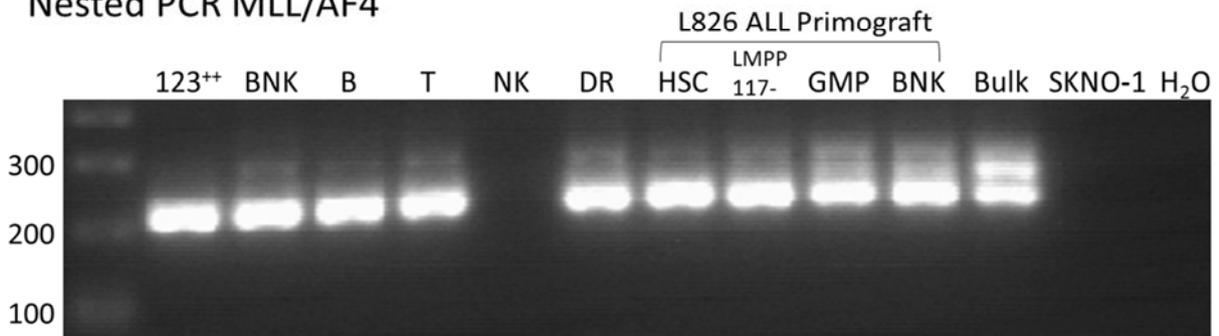
Nested PCR MLL/AF4



Nested PCR ATP10A



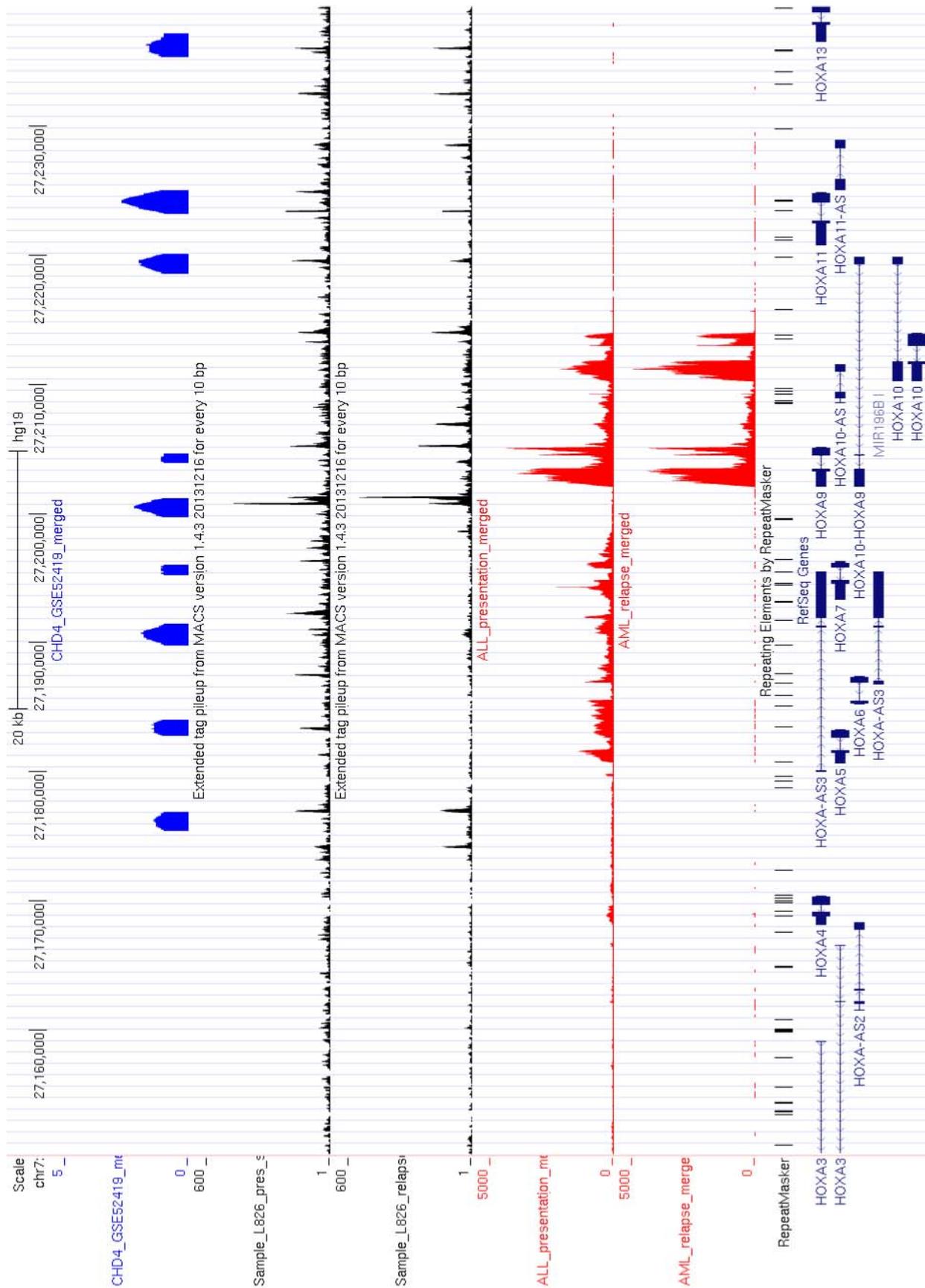
Nested PCR MLL/AF4



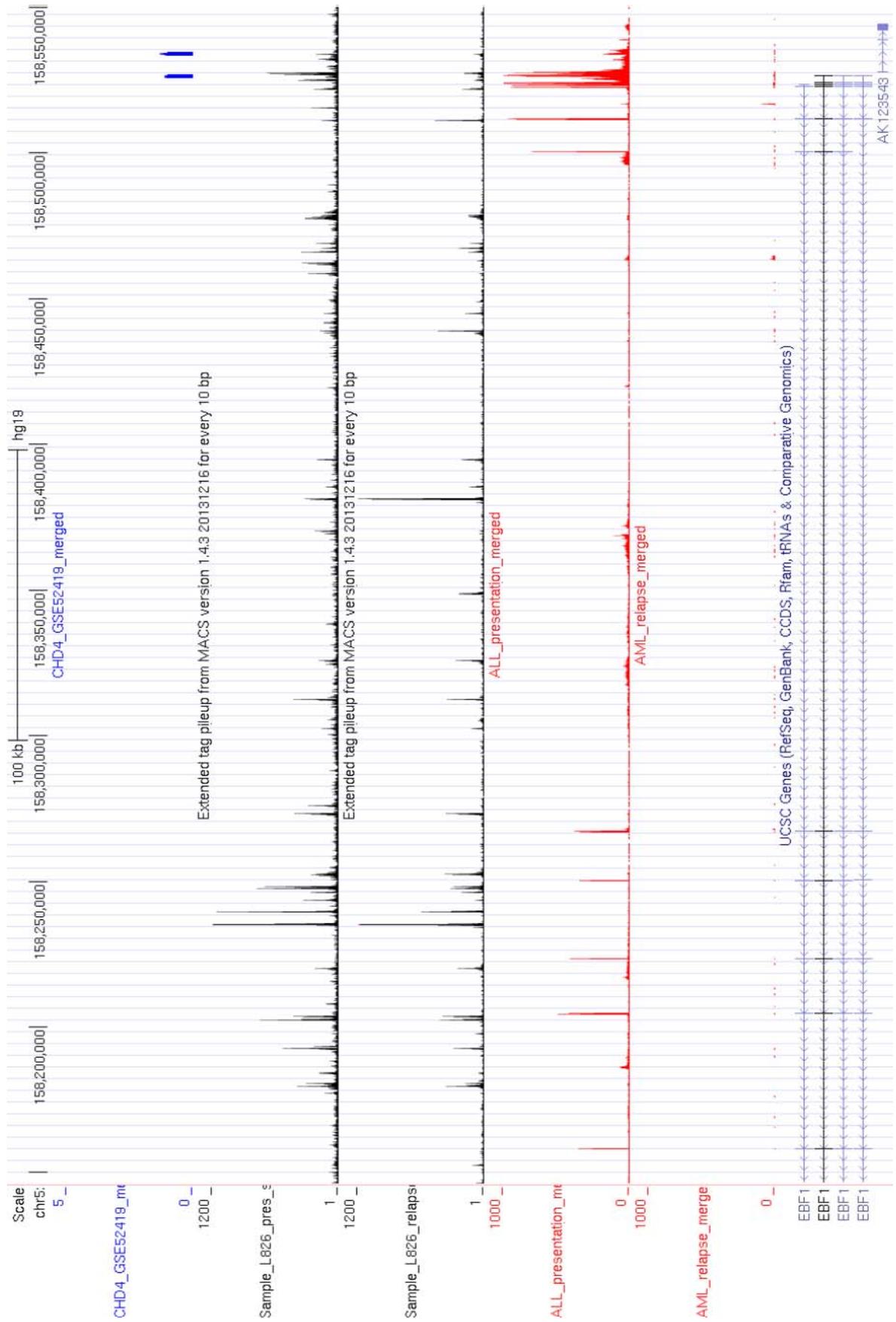
Gel electrophoresis image of nested PCR *MLL/AF4* and ATP10 (positive control to show sufficient template amount for PCR reaction). Sample 34+ is a mixed of any haematopoietic stem and progenitor expressing CD34. LMPP is originally characterised by CD34+CD38-CD90-CD45RA+CD10-CD117+. However, the cell sorting showed similar population, but did not express CD117. This population might be due to aberrant population in L826 sample or technically limitation of fluorochromes spill-over. GMP is characterised by CD34+CD38+CD10-CD123+CD45RA+. GMP-A, -B, and -C were separated based on the level of CD123 expression, from lower to higher expression, respectively. Sample 123++ is GMP that highly expressed CD123. Monocyte is defined by CD45+DR+Lin-CD14+CD16-marker expression. Monocyte CD16+ is a non-classical monocyte that has not fully characterised. SKNO-1 is t(8;21) AML cell line as a negative control of *MLL/AF4* expression. MLP population L826 relapse showed positivity of *MLL/AF4*, but not on the positive control ATP10A. The ambiguity was probably because the minimum template for the reaction that came only from 16 cells.

Appendix B Overlaying DNase hypersensitivity, RNA-seq L826 and CHD4 ChIP-seq

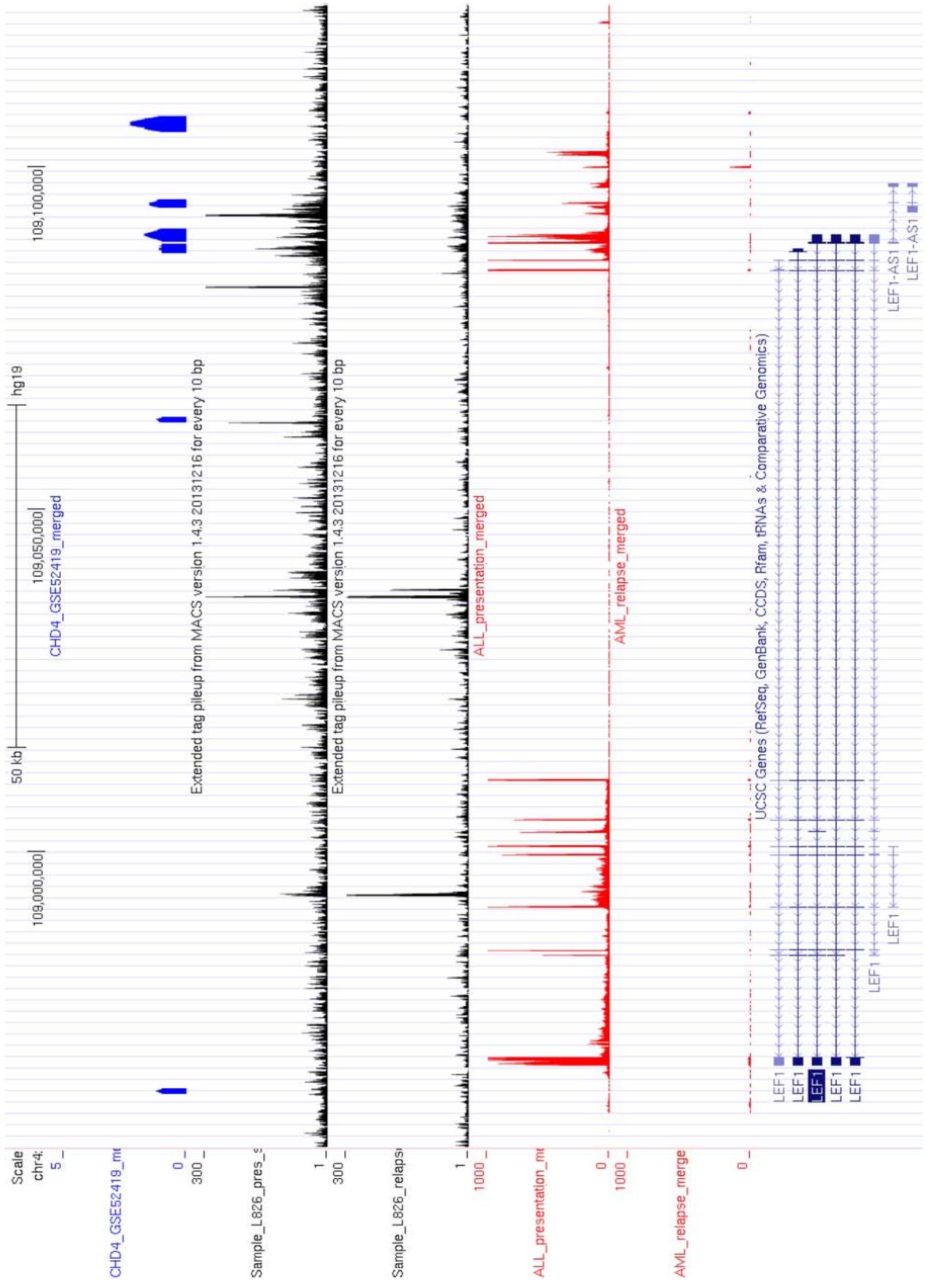
B.1 Hox cluster



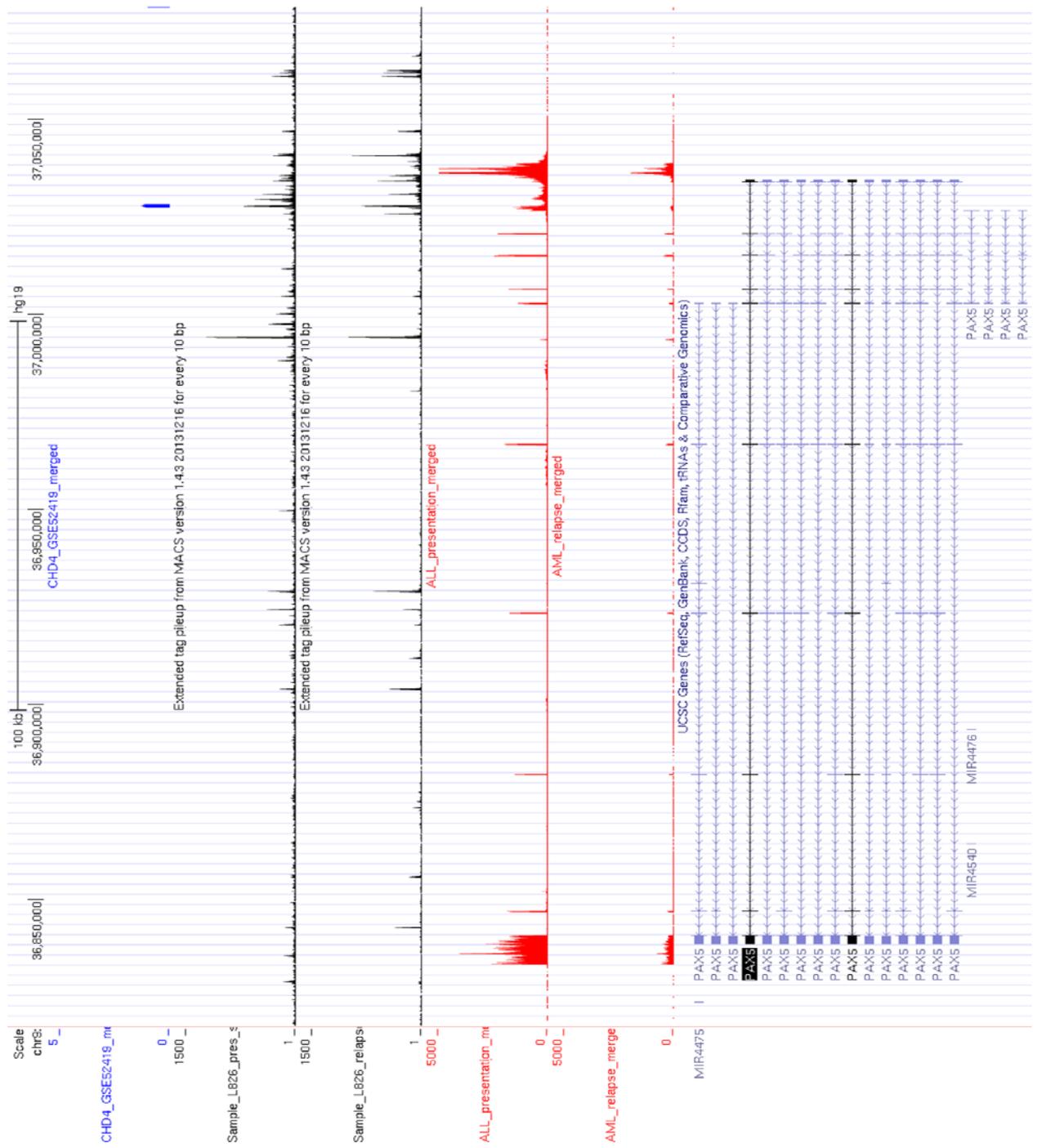
B.2 EBF1



B.3 LEF1



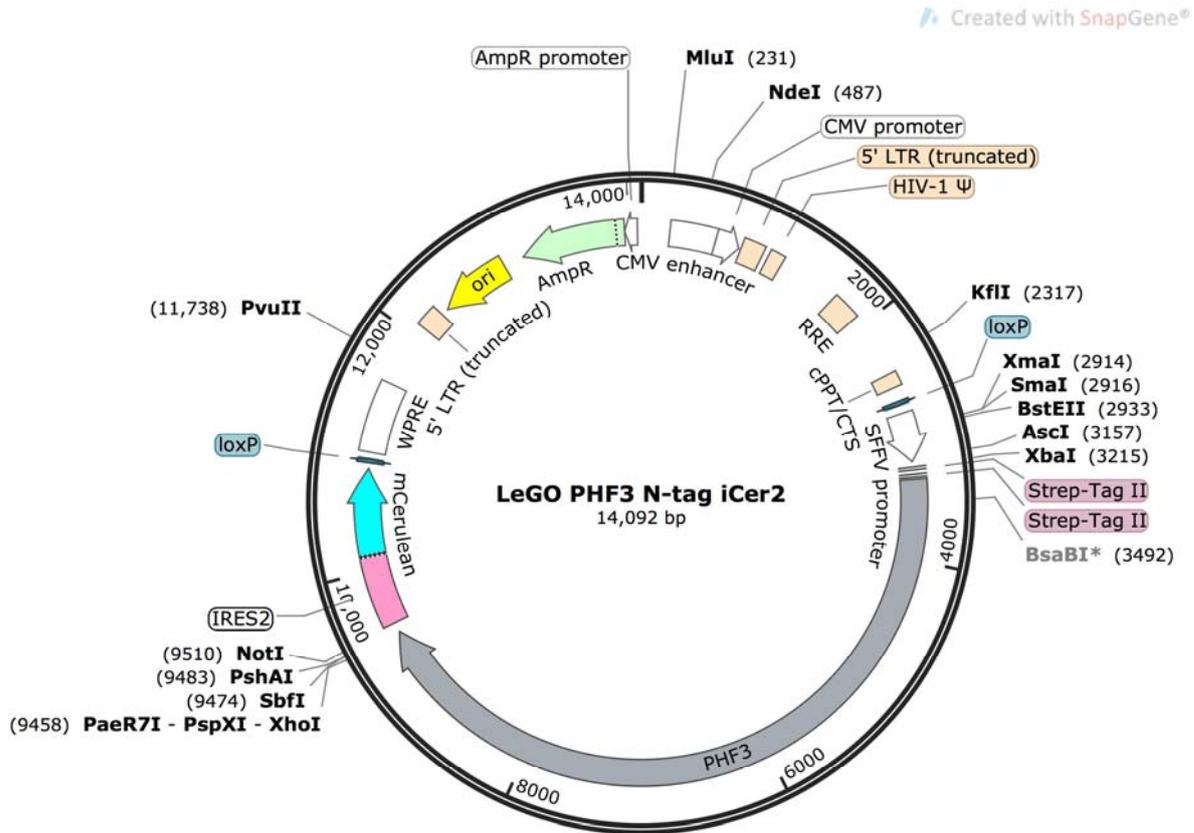
B.4 PAX5



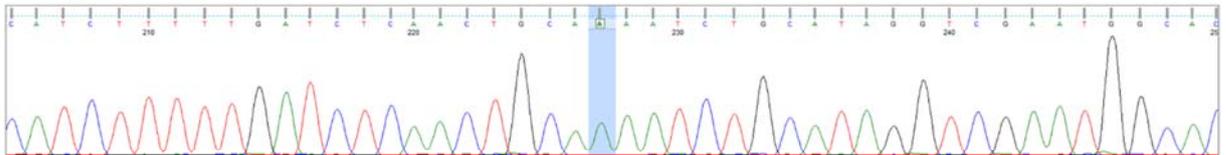
The graphs consist of CHD4 binding region based on CHIP-seq human glioblastoma (GSE52419) on the top panel, DNase hypersensitivity L826 presentation and relapse (second and third panels, respectively), and RNAseq L826 presentation and relapse (fourth and fifth panels, respectively). Multiple CHD4 binding sites are seen on HOXA cluster and LEF1. The binding sites are also observed on the area around promoter in EBF1 and PAX5.

Appendix C Vector maps

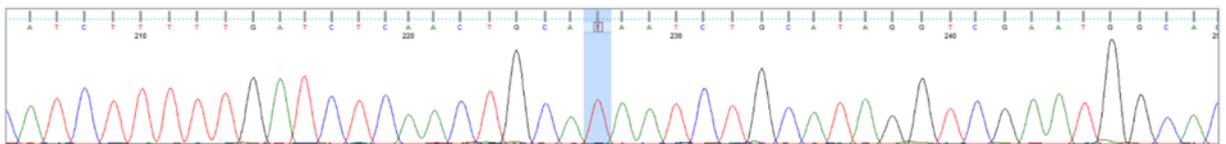
C.1 LeGO PHF3 N-tag iCer2



Wild-type



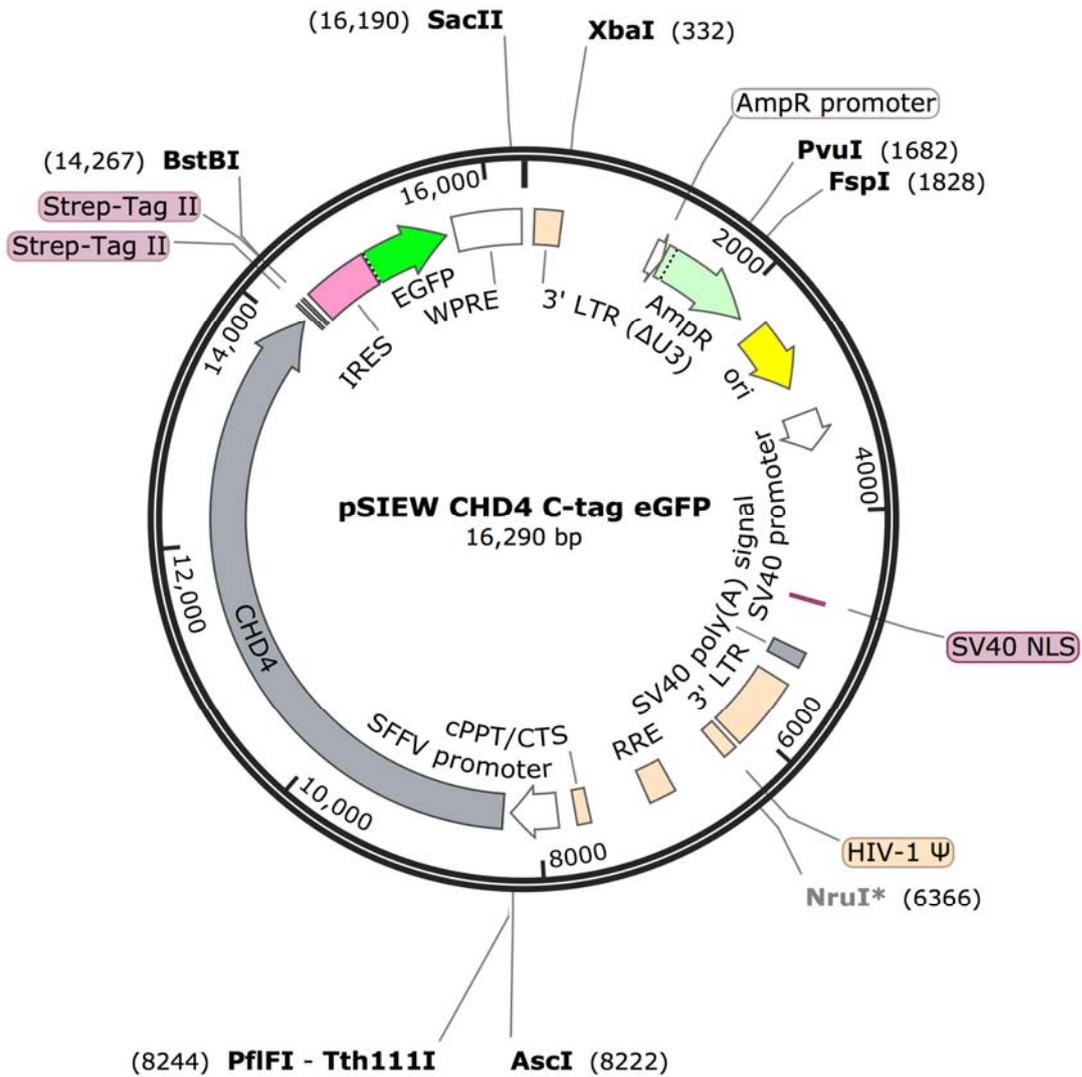
Mutant



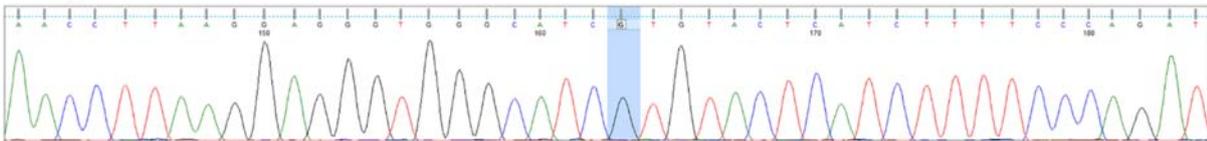
PHF3 mutation 3382A>T (accession number NM_015153.3) created K1119I was generated in LeGO iCer2 lentiviral vector.

C.2 pSIEW CHD4 C-tag eGFP

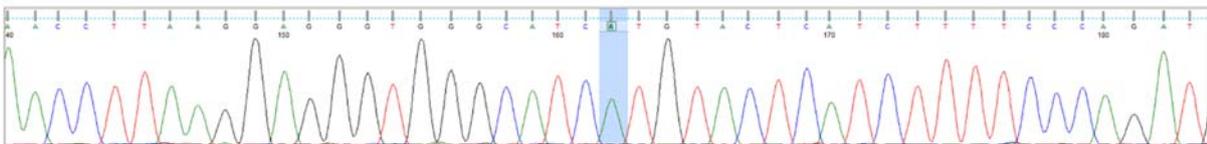
Created with SnapGene®



Wild-type



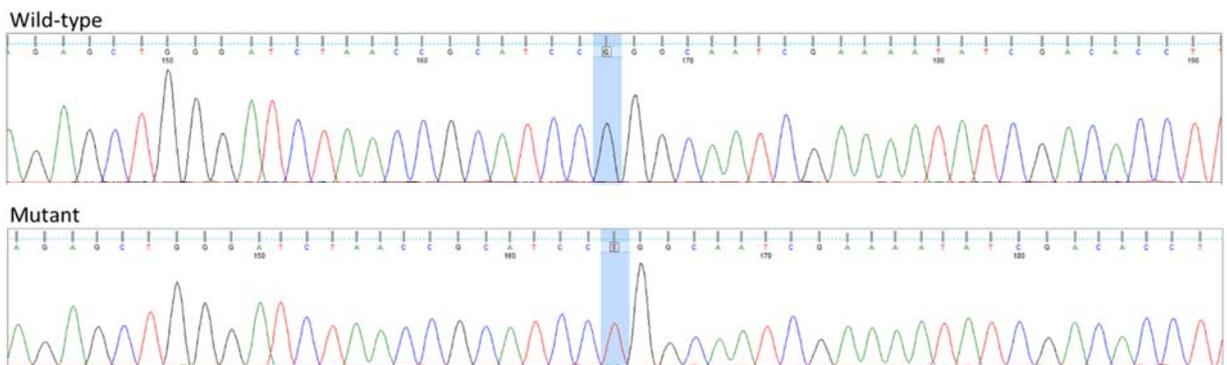
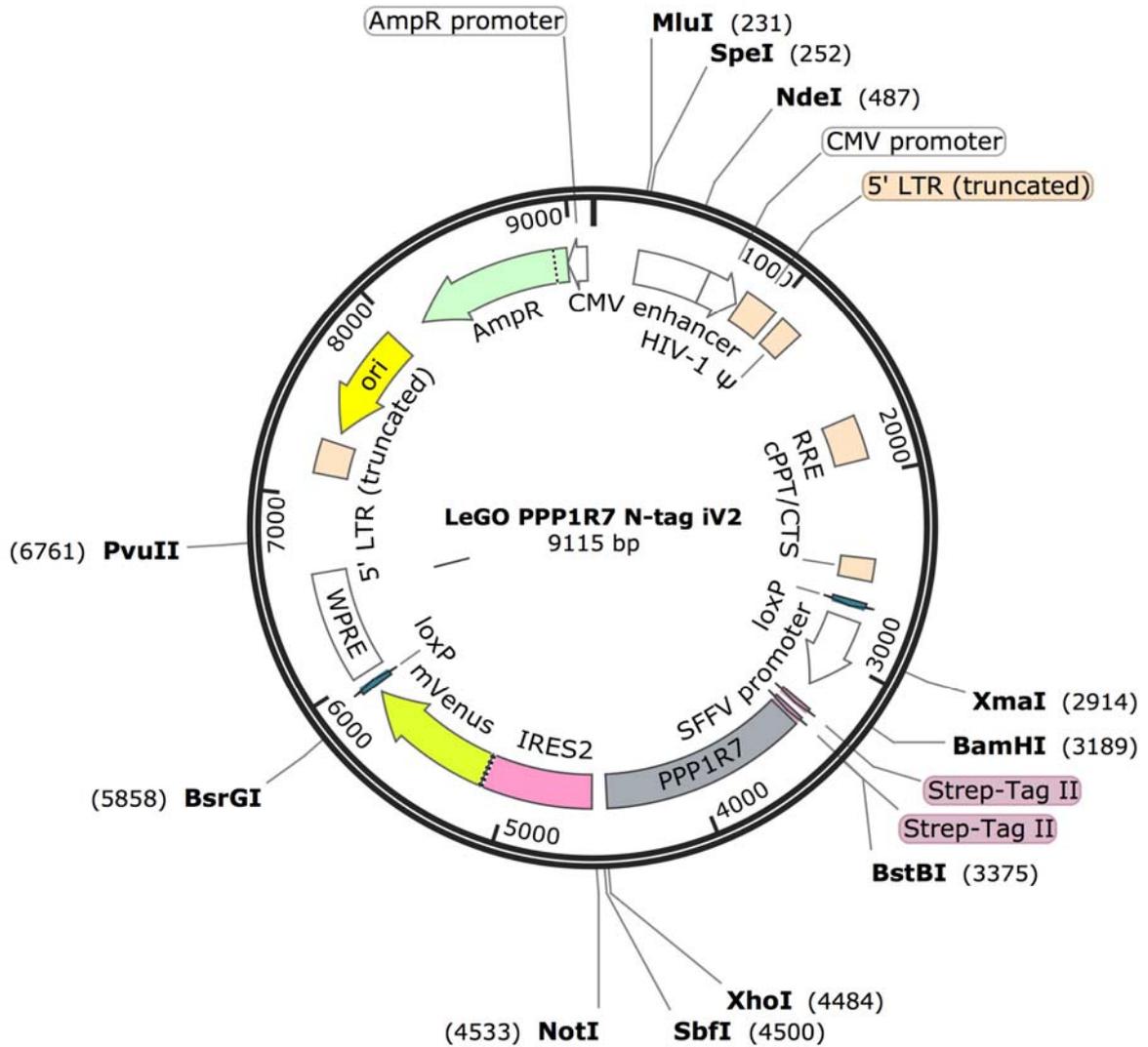
Mutant



CHD4 mutation 3415G>A (accession number NM_001273.3) created R1068H was generated in pSIEW lentiviral vector.

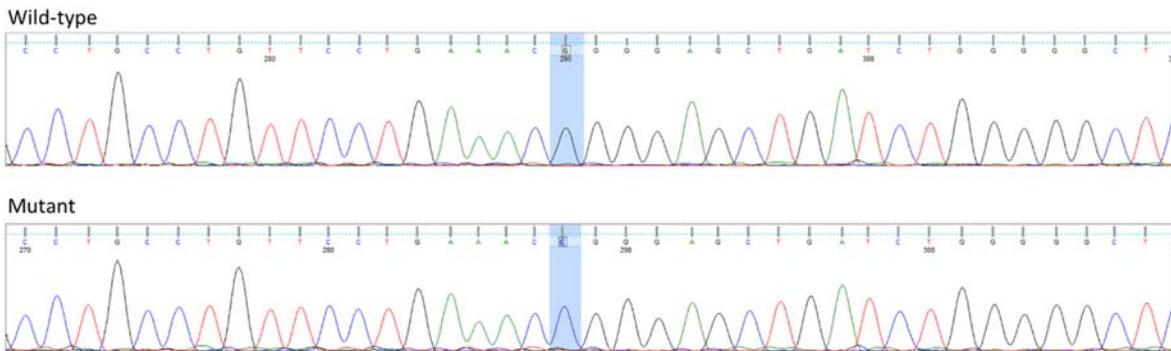
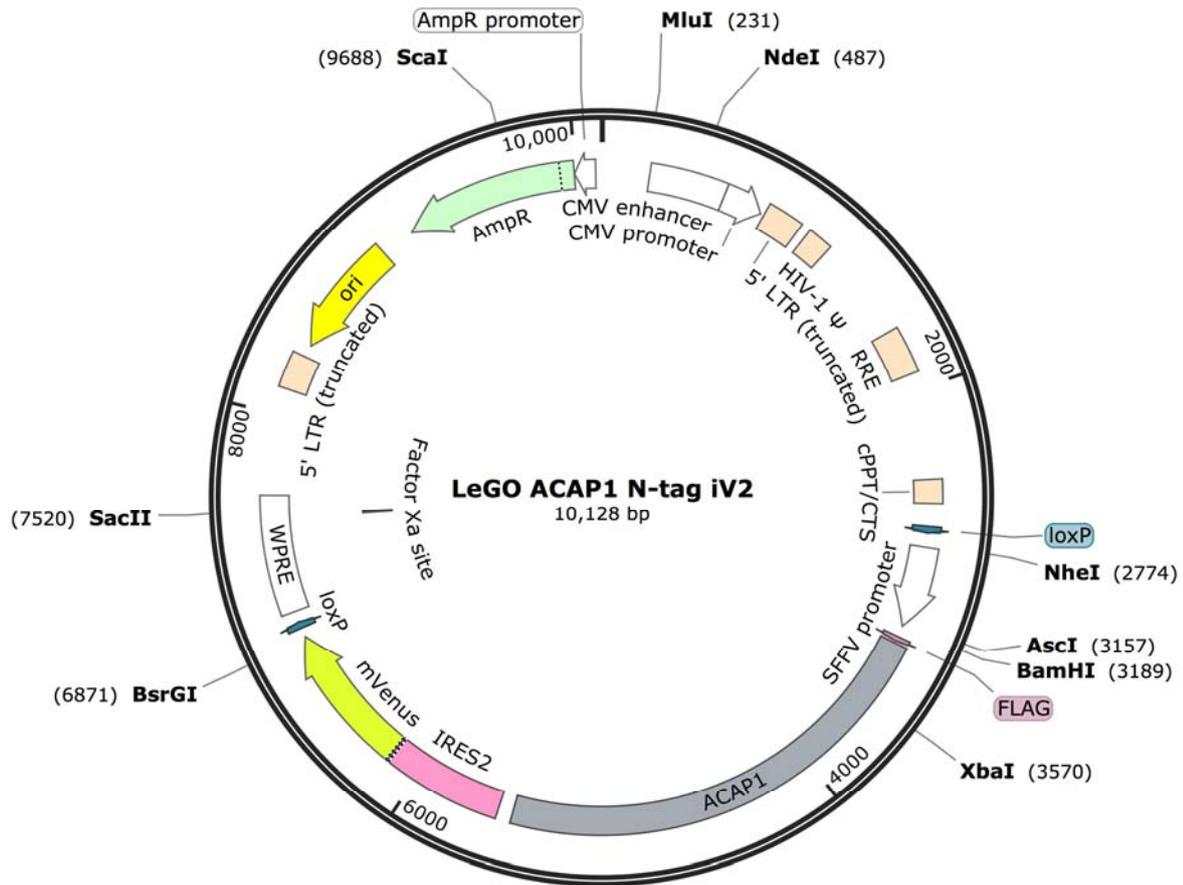
C.3 LeGO PPP1R7 N-tag iV2

Created with SnapGene®



C.4 LeGO ACAP1 N-tag iV2

Created with SnapGene®



ACAP1 mutation 2191G>C (accession number NM_014716.3) created R662P was generated in LeGO iV2 lentiviral vector.

References

1. Doulatov, S., et al., *Hematopoiesis: a human perspective*. Cell Stem Cell, 2012. **10**(2): p. 120-36.
2. Milne, P., *Dendritic cell development in haematological malignancies and neoplasia*, in *Institute of Cellular Medicine*. 2015, Newcastle University: Newcastle upon Tyne. p. 324.
3. Abkowitz, J.L., et al., *Evidence for the maintenance of hematopoiesis in a large animal by the sequential activation of stem-cell clones*. Proc Natl Acad Sci U S A, 1990. **87**(22): p. 9062-6.
4. Manz, M.G., et al., *Prospective isolation of human clonogenic common myeloid progenitors*. Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11872-7.
5. Doulatov, S., et al., *Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development*. Nat Immunol, 2010. **11**(7): p. 585-93.
6. Goardon, N., et al., *Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia*. Cancer Cell, 2011. **19**(1): p. 138-52.
7. Ceredig, R., A.G. Rolink, and G. Brown, *Models of haematopoiesis: seeing the wood for the trees*. Nat Rev Immunol, 2009. **9**(4): p. 293-300.
8. Graf, T., *Differentiation plasticity of hematopoietic cells*. Blood, 2002. **99**(9): p. 3089-101.
9. Friedman, A.D., *Transcriptional control of granulocyte and monocyte development*. Oncogene, 2007. **26**(47): p. 6816-28.
10. Wang, W., et al., *C/EBPalpha and G-CSF receptor signals cooperate to induce the myeloperoxidase and neutrophil elastase genes*. Leukemia, 2001. **15**(5): p. 779-86.
11. Pongubala, J.M.R., et al., *Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5*. Nat Immunol, 2008. **9**(2): p. 203-215.
12. Jamieson, C.H.M., I.L. Weissman, and E. Passegué, *Chronic versus acute myelogenous leukemia: A question of self-renewal*. Cancer Cell, 2004. **6**(6): p. 531-533.
13. Pinkerton, C.R., P.N. Plowman, and R. Pieters, *Paediatric oncology*. 3rd ed. ed. 2004, London: Arnold. 728.
14. Service, M.C.S.a.N.C.R.a.A. *Cancer Prevalence UK Data Tables*. 2015 [cited 2017 27 October]; Available from: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-all/incidence#heading-Five>.
15. Pui, C.H., et al., *Biology, risk stratification, and therapy of pediatric acute leukemias: an update*. J Clin Oncol, 2011. **29**(5): p. 551-65.
16. Szczepanski, T., V.H. van der Velden, and J.J. van Dongen, *Classification systems for acute and chronic leukaemias*. Best Pract Res Clin Haematol, 2003. **16**(4): p. 561-82.
17. Mughal, T.I., *Understanding leukaemias, lymphomas and myelomas*, ed. J. Goldman. 2006, London: London : Taylor & Francis.
18. Deschler, B. and M. Lubbert, *Acute myeloid leukemia: epidemiology and etiology*. Cancer, 2006. **107**(9): p. 2099-107.
19. Behm, F.G. and D. Campana, *Childhood Leukaemias*. Immunophenotyping, ed. C.H. Pui. 1999, Cambridge: Cambridge University Press. 567.
20. Koh, K., et al., *Early use of allogeneic hematopoietic stem cell transplantation for infants with MLL gene-rearrangement-positive acute lymphoblastic leukemia*. Leukemia, 2015. **29**(2): p. 290-6.
21. ZoAnn, E.D., et al., *Analysis of the Role of Hematopoietic Stem-Cell Transplantation in Infants With Acute Lymphoblastic Leukemia in First Remission and MLL Gene Rearrangements: A Report From the Children's Oncology Group*. Journal of Clinical Oncology, 2011. **29**(2): p. 214-222.
22. Dreyer, Z.E., et al., *Intensified chemotherapy without SCT in infant ALL: results from COG P9407 (Cohort 3)*. Pediatr Blood Cancer, 2015. **62**(3): p. 419-26.
23. Nagayama, J., et al., *Infants with acute lymphoblastic leukemia and a germline *MLL* gene are highly curable with use of chemotherapy alone: results from the Japan Infant Leukemia Study Group*. Blood, 2006. **107**(12): p. 4663-4665.

24. Patrick, K. and A. Vora, *Update on biology and treatment of T-cell acute lymphoblastic leukaemia*. *Curr Opin Pediatr*, 2015. **27**(1): p. 44-9.
25. Pieters, R., et al., *A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial*. *Lancet*, 2007. **370**(9583): p. 240-50.
26. Hilden, J.M., et al., *Analysis of prognostic factors of acute lymphoblastic leukemia in infants: report on CCG 1953 from the Children's Oncology Group*. *Blood*, 2006. **108**(2): p. 441-51.
27. van der Linden, M.H., et al., *Clinical and molecular genetic characterization of wild-type MLL infant acute lymphoblastic leukemia identifies few recurrent abnormalities*. *Haematologica*, 2016. **101**(3): p. e95-8.
28. De Lorenzo, P., et al., *Cytogenetics and outcome of infants with acute lymphoblastic leukemia and absence of MLL rearrangements*. *Leukemia*, 2014. **28**(2): p. 428-30.
29. Brown, P., *Treatment of infant leukemias: challenge and promise*. Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program, 2013. **2013**: p. 596-600.
30. Jansen, M.W.J.C., et al., *Immunobiological diversity in infant acute lymphoblastic leukemia is related to the occurrence and type of MLL gene rearrangement*. *Leukemia*, 2007. **21**(4): p. 633-641.
31. Gale, K.B., et al., *Backtracking leukemia to birth: Identification of clonotypic gene fusion sequences in neonatal blood spots*. *Proceedings of the National Academy of Sciences of the United States of America*, 1997. **94**(25): p. 13950-13954.
32. Fasching, K., et al., *Presence of clone-specific antigen receptor gene rearrangements at birth indicates an in utero origin of diverse types of early childhood acute lymphoblastic leukemia*. *Blood*, 2000. **95**(8): p. 2722-4.
33. Kotecha, R.S., et al., *The evolution of clinical trials for infant acute lymphoblastic leukemia*. *Blood Cancer Journal*, 2014. **4**: p. e200.
34. Pui, C.H., et al., *11q23/MLL rearrangement confers a poor prognosis in infants with acute lymphoblastic leukemia*. *J Clin Oncol*, 1994. **12**(5): p. 909-15.
35. Silverman, L.B., et al., *Intensified therapy for infants with acute lymphoblastic leukemia: results from the Dana-Farber Cancer Institute Consortium*. *Cancer*, 1997. **80**(12): p. 2285-95.
36. Dordelmann, M., et al., *Prednisone response is the strongest predictor of treatment outcome in infant acute lymphoblastic leukemia*. *Blood*, 1999. **94**(4): p. 1209-17.
37. Pieters, R., et al., *A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial*. *The Lancet*. **370**(9583): p. 240-250.
38. Nederland, S.K., *INTERFANT-06*, in *INTERNATIONAL COLLABORATIVE TREATMENT PROTOCOL FOR INFANTS UNDER ONE YEAR WITH ACUTE LYMPHOBLASTIC OR BIPHENOTYPIC LEUKEMIA*. 2006: The Hague.
39. Tomizawa, D., *Recent progress in the treatment of infant acute lymphoblastic leukemia*. *Pediatr Int*, 2015. **57**(5): p. 811-9.
40. Mann, G., et al., *Improved outcome with hematopoietic stem cell transplantation in a poor prognostic subgroup of infants with mixed-lineage-leukemia (MLL)-rearranged acute lymphoblastic leukemia: results from the Interfant-99 Study*. *Blood*, 2010. **116**(15): p. 2644-2650.
41. de Rooij, J.D.E., C.M. Zwaan, and M. van den Heuvel-Eibrink, *Pediatric AML: From Biology to Clinical Management*. *Journal of Clinical Medicine*, 2015. **4**(1): p. 127-149.
42. Dores, G.M., et al., *Acute leukemia incidence and patient survival among children and adults in the United States, 2001-2007*. *Blood*, 2012. **119**(1): p. 34-43.
43. Grimwade, D., A. Ivey, and B.J.P. Huntly, *Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance*. *Blood*, 2016. **127**(1): p. 29-41.
44. Rubnitz, J.E., *Current Management of Childhood Acute Myeloid Leukemia*. *Paediatr Drugs*, 2017. **19**(1): p. 1-10.

45. Christine, J.H., et al., *Cytogenetics of Childhood Acute Myeloid Leukemia: United Kingdom Medical Research Council Treatment Trials AML 10 and 12*. Journal of Clinical Oncology, 2010. **28**(16): p. 2674-2681.
46. Gamis, A.S., et al., *Gemtuzumab ozogamicin in children and adolescents with de novo acute myeloid leukemia improves event-free survival by reducing relapse risk: results from the randomized phase III Children's Oncology Group trial AAML0531*. J Clin Oncol, 2014. **32**(27): p. 3021-32.
47. Gibson, B.E., et al., *Results of a randomized trial in children with Acute Myeloid Leukaemia: medical research council AML12 trial*. Br J Haematol, 2011. **155**(3): p. 366-76.
48. Rubnitz, J.E., et al., *Minimal residual disease-directed therapy for childhood acute myeloid leukaemia: results of the AML02 multicentre trial*. Lancet Oncol, 2010. **11**(6): p. 543-52.
49. Imamura, T., et al., *Outcome in 146 patients with paediatric acute myeloid leukaemia treated according to the AML99 protocol in the period 2003-06 from the Japan Association of Childhood Leukaemia Study*. Br J Haematol, 2012. **159**(2): p. 204-10.
50. Creutzig, U., et al., *Randomized trial comparing liposomal daunorubicin with idarubicin as induction for pediatric acute myeloid leukemia: results from Study AML-BFM 2004*. Blood, 2013. **122**(1): p. 37-43.
51. Pession, A., et al., *Results of the AIEOP AML 2002/01 multicenter prospective trial for the treatment of children with acute myeloid leukemia*. Blood, 2013. **122**(2): p. 170-8.
52. Abrahamsson, J., et al., *Response-guided induction therapy in pediatric acute myeloid leukemia with excellent remission rate*. J Clin Oncol, 2011. **29**(3): p. 310-5.
53. Walter, R.B., et al., *Significance of FAB subclassification of "acute myeloid leukemia, NOS" in the 2008 WHO classification: analysis of 5848 newly diagnosed patients*. Blood, 2013. **121**(13): p. 2424-2431.
54. Arber, D.A., et al., *The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia*. Blood, 2016. **127**(20): p. 2391-2405.
55. Godley, L.A. and R.A. Larson, *Therapy-related Myeloid Leukemia*. Seminars in oncology, 2008. **35**(4): p. 418-429.
56. Larson, R.A., *Therapy-related myeloid neoplasms*. Haematologica, 2009. **94**(4): p. 454-459.
57. Smith, S.M., et al., *Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: the University of Chicago series*. Blood, 2003. **102**(1): p. 43-52.
58. Lene Sofie Granfeldt, Ø., et al., *Epidemiology and Clinical Significance of Secondary and Therapy-Related Acute Myeloid Leukemia: A National Population-Based Cohort Study*. Journal of Clinical Oncology, 2015. **33**(31): p. 3641-3649.
59. Bhatia, S., et al., *Therapy-related myelodysplasia and acute myeloid leukemia after Ewing sarcoma and primitive neuroectodermal tumor of bone: A report from the Children's Oncology Group*. Blood, 2007. **109**(1): p. 46-51.
60. Travis, L.B., et al., *Treatment-associated leukemia following testicular cancer*. J Natl Cancer Inst, 2000. **92**(14): p. 1165-71.
61. Vardiman, J.W., N.L. Harris, and R.D. Brunning, *The World Health Organization (WHO) classification of the myeloid neoplasms*. Blood, 2002. **100**(7): p. 2292-2302.
62. Bhatia, S., *Therapy-related myelodysplasia and acute myeloid leukemia*. Seminars in oncology, 2013. **40**(6): p. 10.1053/j.seminoncol.2013.09.013.
63. Sanderson, B.J. and A.J. Shield, *Mutagenic damage to mammalian cells by therapeutic alkylating agents*. Mutat Res, 1996. **355**(1-2): p. 41-57.
64. Zhang, Y., et al., *Genomic DNA breakpoints in AML1/RUNX1 and ETO cluster with topoisomerase II DNA cleavage and DNase I hypersensitive sites in t(8;21) leukemia*. Proceedings of the National Academy of Sciences, 2002. **99**(5): p. 3070-3075.
65. Felix, C.A., et al., *Chromosome band 11q23 translocation breakpoints are DNA topoisomerase II cleavage sites*. Cancer Res, 1995. **55**(19): p. 4287-92.

66. Ahuja, H.G., C.A. Felix, and P.D. Aplan, *Potential role for DNA topoisomerase II poisons in the generation of t(11;20)(p15;q11) translocations*. *Genes Chromosomes Cancer*, 2000. **29**(2): p. 96-105.
67. Meyer, C., et al., *The MLL recombinome of acute leukemias in 2017*. *Leukemia*, 2017.
68. Daser, A. and T.H. Rabbitts, *The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis*. *Semin Cancer Biol*, 2005. **15**(3): p. 175-88.
69. Fair, K., et al., *Protein Interactions of the MLL PHD Fingers Modulate MLL Target Gene Regulation in Human Cells*. *Molecular and Cellular Biology*, 2001. **21**(10): p. 3589-3597.
70. Guenther, M.G., et al., *Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia*. *Genes Dev*, 2008. **22**(24): p. 3403-8.
71. Marschalek, R., *Mechanisms of leukemogenesis by MLL fusion proteins*. *Br J Haematol*, 2011. **152**(2): p. 141-54.
72. Nakamura, T., et al., *ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation*. *Mol Cell*, 2002. **10**(5): p. 1119-28.
73. Bueno, C., et al., *A human ESC model for MLL-AF4 leukemic fusion gene reveals an impaired early hematopoietic-endothelial specification*. *Cell Res*, 2012. **22**(6): p. 986-1002.
74. Osaki, H., et al., *The AAA+ ATPase RUVBL2 is a critical mediator of MLL-AF9 oncogenesis*. *Leukemia*, 2013. **27**(7): p. 1461-8.
75. Wei, J., et al., *Microenvironment determines lineage fate in a human model of MLL-AF9 leukemia*. *Cancer Cell*, 2008. **13**(6): p. 483-95.
76. Stavropoulou, V., et al., *MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome*. *Cancer Cell*, 2016. **30**(1): p. 43-58.
77. Tamai, H. and K. Inokuchi, *Establishment of MLL/AF4 transgenic mice with the phenotype of lymphoblastic leukemia or lymphoma*. *J Nippon Med Sch*, 2013. **80**(5): p. 326-7.
78. Chen, W., et al., *A murine Mll-AF4 knock-in model results in lymphoid and myeloid deregulation and hematologic malignancy*. *Blood*, 2006. **108**(2): p. 669-677.
79. Metzler, M., et al., *A conditional model of MLL-AF4 B-cell tumorigenesis using invertebrate technology*. *Oncogene*, 2006. **25**(22): p. 3093-103.
80. Krivtsov, A.V., et al., *H3K79 methylation profiles define murine and human MLL-AF4 leukemias*. *Cancer Cell*, 2008. **14**(5): p. 355-68.
81. Bursen, A., et al., *The AF4-MLL fusion protein is capable of inducing ALL in mice without requirement of MLL-AF4*. *Blood*, 2010. **115**(17): p. 3570-3579.
82. Lin, S., et al., *Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia*. *Cancer Cell*, 2016. **30**(5): p. 737-749.
83. Gagnon, G.A., et al., *Molecular heterogeneity in acute leukemia lineage switch*. *Blood*, 1989. **74**: p. 2088 - 2095.
84. Stass, S., et al., *Lineage switch in acute leukemia*. *Blood*, 1984. **64**(3): p. 701-6.
85. Gupta, S.K., et al., *Secondary pure erythroid leukaemia in relapsed acute lymphoblastic leukaemia: lineage switch or chemotherapy effect?* *BMJ Case Rep*, 2014. **2014**.
86. Hossain B., et al., *Lineage switch in acute lymphoblastic leukaemia*, in 2005; **21** (1) : 29-32, D.C.H. J, Editor. 2005.
87. Ikarashi, Y., et al., *Double leukemias simultaneously showing lymphoblastic leukemia of the bone marrow and monocytic leukemia of the central nervous system*. *Am J Hematol*, 2004. **75**(3): p. 164-7.
88. Kobayashi, S., et al., *Double Lineage Switch from Acute Megakaryoblastic Leukemia (AML-M7) to Acute Lymphoblastic Leukemia (ALL) and Back Again: A Case Report*. *Journal of Blood Disorders & Transfusion*, 2014.
89. Stasik, C., et al., *Infant acute lymphoblastic leukemia with t(11;16)(q23;p13.3) and lineage switch into acute monoblastic leukemia*. *Cancer Genet Cytogenet*, 2006. **168**(2): p. 146-9.
90. Boeckx, N., et al., *An inv(16)(p13q22) positive acute myeloid leukaemia relapsing as acute precursor B-cell lymphoblastic leukaemia*. *Haematologica*, 2004. **89**(8): p. ECR28.

91. Dorantes-Acosta, E., et al., *Acute myelogenous leukemia switch lineage upon relapse to acute lymphoblastic leukemia: a case report*. Cases Journal, 2009. **2**(1): p. 154.
92. Grammatico, S., et al., *Lineage switch from pro-B acute lymphoid leukemia to acute myeloid leukemia in a case with t(12;17)(p13;q11)/TAF15-ZNF384 rearrangement*. Leuk Lymphoma, 2013. **54**(8): p. 1802-5.
93. Ivanov Iuliu, C., et al., *Infant acute leukemia with lineage switch at relapse expressing a novel t(4;11)(q21;q23) MLL-AF4 fusion transcript*, in *Romanian Review of Laboratory Medicine*. 2013. p. 47.
94. Jiang, J.G., et al., *Congenital MLL-positive B-cell acute lymphoblastic leukemia (B-ALL) switched lineage at relapse to acute myelocytic leukemia (AML) with persistent t(4;11) and t(1;6) translocations and JH gene rearrangement*. Leuk Lymphoma, 2005. **46**(8): p. 1223-7.
95. Krawczuk-Rybak, M., J. Zak, and B. Jaworowska, *A lineage switch from AML to ALL with persistent translocation t(4;11) in congenital leukemia*. Med Pediatr Oncol, 2003. **41**(1): p. 95-6.
96. Lounici, A., et al., *Lineage switch from acute myeloid leukemia to acute lymphoblastic leukemia: report of an adult case and review of the literature*. Am J Hematol, 2000. **65**(4): p. 319-21.
97. Park, B.G., et al., *Erythroleukemia relapsing as precursor B-cell lymphoblastic leukemia*. Korean J Lab Med, 2011. **31**(2): p. 81-5.
98. Higuchi, Y., et al., *Lineage switch with t(6;11)(q27;q23) from T-cell lymphoblastic lymphoma to acute monoblastic leukemia at relapse*. Cancer Genet, 2016. **209**(6): p. 267-71.
99. Ittel, A., et al., *First description of the t(10;11)(q22;q23)/MLL-TET1 translocation in a T-cell lymphoblastic lymphoma, with subsequent lineage switch to acute myelomonocytic myeloid leukemia*. Haematologica, 2013. **98**(12): p. e166-e168.
100. Rossi, J.G., et al., *Lineage switch in childhood acute leukemia: an unusual event with poor outcome*. Am J Hematol, 2012. **87**(9): p. 890-7.
101. Moschiano, E., et al., *Congenital B-lymphoblastic leukemia with a cryptic MLL rearrangement and post-treatment evolution to mixed phenotype acute leukemia*. Leukemia Research Reports, 2016. **6**: p. 29-32.
102. Park, M., et al., *Lineage switch at relapse of childhood acute leukemia: a report of four cases*. J Korean Med Sci, 2011. **26**(6): p. 829-31.
103. Germano, G., et al., *Two consecutive immunophenotypic switches in a child with MLL-rearranged acute lymphoblastic leukemia*. Haematologica, 2006. **91**(5 Suppl): p. Ecr09.
104. Blutters-Sawatzki, R., et al., *Secondary acute myeloid leukemia with translocation (4;11) and MLL/AF4 rearrangement in a 15-year-old boy treated for common acute lymphoblastic leukemia 11 years earlier*. Ann Hematol, 1995. **70**(1): p. 31-5.
105. Paganin, M., et al., *A Case of T-cell Acute Lymphoblastic Leukemia Relapsed As Myeloid Acute Leukemia*. Pediatr Blood Cancer, 2016. **63**(9): p. 1660-3.
106. Ford, A.M., et al., *Protracted dormancy of pre-leukemic stem cells*. Leukemia, 2015. **29**(11): p. 2202-7.
107. van den Ancker, W., et al., *Uncommon lineage switch warrants immunophenotyping even in relapsing leukemia*. Leuk Res, 2009. **33**(7): p. e77-80.
108. Carulli, G., et al., *B-cell acute lymphoblastic leukemia with t(4;11)(q21;q23) in a young woman: evolution into mixed phenotype acute leukemia with additional chromosomal aberrations in the course of therapy*. Hematology Reports, 2012. **4**(3): p. e15.
109. Matern, S., et al., *Successful second allogeneic stem cell transplantation in a patient with T-lymphoblastic leukemia (T-ALL) relapsed as myeloid sarcoma*. Ann Hematol, 2017. **96**(7): p. 1201-1203.
110. Della Starza, I., et al., *A case of lineage switch from B-cell acute lymphoblastic leukaemia to acute myeloid leukaemia. Role of subclonal/clonal gene mutations*. British Journal of Haematology, 2016. **174**(4): p. 648-651.

111. Hershfield, M.S., et al., *Conversion of a stem cell leukemia from a T-lymphoid to a myeloid phenotype induced by the adenosine deaminase inhibitor 2'-deoxycoformycin*. Proc Natl Acad Sci U S A, 1984. **81**(1): p. 253-7.
112. Ben-Yehuda, D., et al., *Microsatellite instability and p53 mutations in therapy-related leukemia suggest mutator phenotype*. Blood, 1996. **88**(11): p. 4296-303.
113. Das-Gupta, E.P., C.H. Seedhouse, and N.H. Russell, *Microsatellite instability occurs in defined subsets of patients with acute myeloblastic leukaemia*. Br J Haematol, 2001. **114**(2): p. 307-12.
114. Horiike, S., et al., *Distinct genetic involvement of the TP53 gene in therapy-related leukemia and myelodysplasia with chromosomal losses of Nos 5 and/or 7 and its possible relationship to replication error phenotype*. Leukemia, 1999. **13**(8): p. 1235-42.
115. Baarcenai, A., et al., *Tracing the Expression of CD7 and other Antigens during T- and Myeloid-cell Differentiation in the Human Fetal Liver and Thymus*. Leukemia & Lymphoma, 1995. **17**(1-2): p. 1-11.
116. Kita, K., et al., *Clinical importance of CD7 expression in acute myelocytic leukemia. The Japan Cooperative Group of Leukemia/Lymphoma*. Blood, 1993. **81**(9): p. 2399-405.
117. Casorelli, I., C. Bossa, and M. Bignami, *DNA Damage and Repair in Human Cancer: Molecular Mechanisms and Contribution to Therapy-Related Leukemias*. International Journal of Environmental Research and Public Health, 2012. **9**(8): p. 2636-2657.
118. Bacher, J.W., et al., *Development of a fluorescent multiplex assay for detection of MSI-High tumors*. Dis Markers, 2004. **20**(4-5): p. 237-50.
119. Umar, A., et al., *Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability*. J Natl Cancer Inst, 2004. **96**(4): p. 261-8.
120. González-Pérez, A. and N. López-Bigas, *Improving the Assessment of the Outcome of Nonsynonymous SNVs with a Consensus Deleteriousness Score, Condel*. American Journal of Human Genetics, 2011. **88**(4): p. 440-449.
121. Robinson, J.T., et al., *Integrative genomics viewer*. Nature Biotechnology, 2011. **29**: p. 24.
122. Larsen, D.H., et al., *The chromatin-remodeling factor CHD4 coordinates signaling and repair after DNA damage*. J Cell Biol, 2010. **190**(5): p. 731-40.
123. Luijsterburg, M.S., et al., *A new non-catalytic role for ubiquitin ligase RNF8 in unfolding higher-order chromatin structure*. Embo j, 2012. **31**(11): p. 2511-27.
124. Pan, M.R., et al., *Chromodomain helicase DNA-binding protein 4 (CHD4) regulates homologous recombination DNA repair, and its deficiency sensitizes cells to poly(ADP-ribose) polymerase (PARP) inhibitor treatment*. J Biol Chem, 2012. **287**(9): p. 6764-72.
125. Polo, S.E., et al., *Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4*. Embo j, 2010. **29**(18): p. 3130-9.
126. Urquhart, A.J., et al., *ATM mediated phosphorylation of CHD4 contributes to genome maintenance*. Genome Integr, 2011. **2**(1): p. 1.
127. Xia, L., et al., *CHD4 Has Oncogenic Functions in Initiating and Maintaining Epigenetic Suppression of Multiple Tumor Suppressor Genes*. Cancer Cell, 2017. **31**(5): p. 653-668.e7.
128. Johnson, D.R., et al., *NuRD complex component Mi-26 binds to and represses RORγ-mediated transcriptional activation*. Biochemical and Biophysical Research Communications, 2004. **318**(3): p. 714-718.
129. Ramirez, J., et al., *MBD2 and multiple domains of CHD4 are required for transcriptional repression by Mi-2/NuRD complexes*. Mol Cell Biol, 2012. **32**(24): p. 5078-88.
130. Ramirez, J., K. Lukin, and J. Hagman, *From hematopoietic progenitors to B cells: mechanisms of lineage restriction and commitment*. Curr Opin Immunol, 2010. **22**(2): p. 177-84.
131. Williams, C.J., et al., *The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development*. Immunity, 2004. **20**(6): p. 719-33.
132. Yoshida, T., et al., *The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation*. Genes Dev, 2008. **22**(9): p. 1174-89.
133. Jackson, T.R., et al., *Acaps Are Arf6 Gtpase-Activating Proteins That Function in the Cell Periphery*. The Journal of Cell Biology, 2000. **151**(3): p. 627-638.

134. Posch, M., et al., *Sds22 regulates aurora B activity and microtubule-kinetochore interactions at mitosis*. J Cell Biol, 2010. **191**(1): p. 61-74.
135. Wurzenberger, C., et al., *Sds22 and Repo-Man stabilize chromosome segregation by counteracting Aurora B on anaphase kinetochores*. The Journal of Cell Biology, 2012. **198**(2): p. 173-183.
136. Struss, A.K., et al., *PHF3-specific antibody responses in over 60% of patients with glioblastoma multiforme*. Oncogene, 2001. **20**(31): p. 4107-14.
137. Boeing, S., et al., *Multiomic Analysis of the UV-Induced DNA Damage Response*. Cell Reports, 2016. **15**(7): p. 1597-1610.
138. Tran, H., et al., *Facilitation of mRNA Deadenylation and Decay by the Exosome-Bound, DExH Protein RHAU*. Molecular Cell, 2004. **13**(1): p. 101-111.
139. Voegel, J.J., et al., *TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors*. Embo j, 1996. **15**(14): p. 3667-75.
140. Graser, S., et al., *Cep164, a novel centriole appendage protein required for primary cilium formation*. J Cell Biol, 2007. **179**(2): p. 321-30.
141. Chaki, M., et al., *Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling*. Cell, 2012. **150**(3): p. 533-548.
142. Pan, Y.R. and E.Y. Lee, *UV-dependent interaction between Cep164 and XPA mediates localization of Cep164 at sites of DNA damage and UV sensitivity*. Cell Cycle, 2009. **8**(4): p. 655-64.
143. Yang, C.Z. and M. Mueckler, *ADP-ribosylation factor 6 (ARF6) defines two insulin-regulated secretory pathways in adipocytes*. J Biol Chem, 1999. **274**(36): p. 25297-300.
144. Zhang, Q., et al., *A requirement for ARF6 in Fcγ receptor-mediated phagocytosis in macrophages*. J Biol Chem, 1998. **273**(32): p. 19977-81.
145. Radhakrishna, H., R.D. Klausner, and J.G. Donaldson, *Aluminum fluoride stimulates surface protrusions in cells overexpressing the ARF6 GTPase*. J Cell Biol, 1996. **134**(4): p. 935-47.
146. Mandiyan, V., et al., *Crystal structure of the ARF-GAP domain and ankyrin repeats of PYK2-associated protein beta*. The EMBO Journal, 1999. **18**(24): p. 6890-6898.
147. Randazzo, P.A. and D.S. Hirsch, *Arf GAPs: multifunctional proteins that regulate membrane traffic and actin remodelling*. Cellular Signalling, 2004. **16**(4): p. 401-413.
148. Liu, W., et al., *IBS: an illustrator for the presentation and visualization of biological sequences*. Bioinformatics, 2015. **31**(20): p. 3359-61.
149. Ge, Q., et al., *Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2 autoantigen*. J Clin Invest, 1995. **96**(4): p. 1730-7.
150. Seelig, H.P., et al., *The major dermatomyositis-specific mi-2 autoantigen is a presumed helicase involved in transcriptional activation*. Arthritis & Rheumatism, 1995. **38**(10): p. 1389-1399.
151. Merlo, G., et al., *Specific autoantibodies in dermatomyositis: a helpful tool to classify different clinical subsets*. Archives of Dermatological Research, 2017. **309**(2): p. 87-95.
152. Ghirardello, A., et al., *Autoantibodies in polymyositis and dermatomyositis*. Curr Rheumatol Rep, 2013. **15**(6): p. 335.
153. Wang, H.-B. and Y. Zhang, *Mi2, an auto-antigen for dermatomyositis, is an ATP-dependent nucleosome remodeling factor*. Nucleic Acids Research, 2001. **29**(12): p. 2517-2521.
154. Hill, C.L., et al., *Frequency of specific cancer types in dermatomyositis and polymyositis: a population-based study*. Lancet, 2001. **357**(9250): p. 96-100.
155. Denslow, S.A. and P.A. Wade, *The human Mi-2/NuRD complex and gene regulation*. Oncogene, 2007. **26**(37): p. 5433-8.
156. Shen, J.K., et al., *Polymyositis/dermatomyositis associated with acute myelocytic leukemia*. Rheumatol Int, 2008. **28**(12): p. 1265-7.
157. Millot, F., et al., *Acute lymphoblastic leukemia associated with dermatomyositis in a child*. Pediatr Dermatol, 1993. **10**(2): p. 199-200.
158. Osako, T., et al., *Flare-up of dermatomyositis along with recurrence of breast cancer*. Breast J, 2007. **13**(2): p. 200-2.

159. Bonnetblanc, J.M., P. Bernard, and J. Fayol, *Dermatomyositis and malignancy. A multicenter cooperative study*. *Dermatologica*, 1990. **180**(4): p. 212-6.
160. Stockton, D., V.R. Doherty, and D.H. Brewster, *Risk of cancer in patients with dermatomyositis or polymyositis, and follow-up implications: a Scottish population-based cohort study*. *Br J Cancer*, 2001. **85**(1): p. 41-5.
161. Wang, Y., et al., *LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer*. *Cell*, 2009. **138**(4): p. 660-72.
162. Gao, H., et al., *Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5*. *Proc Natl Acad Sci U S A*, 2009. **106**(27): p. 11258-63.
163. Mansfield, R.E., et al., *Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9*. *J Biol Chem*, 2011. **286**(13): p. 11779-91.
164. Morra, R., et al., *Concerted action of the PHD, chromo and motor domains regulates the human chromatin remodelling ATPase CHD4*. *FEBS Lett*, 2012. **586**(16): p. 2513-21.
165. Lai, A.Y. and P.A. Wade, *Cancer biology and NuRD: a multifaceted chromatin remodelling complex*. *Nat Rev Cancer*, 2011. **11**(8): p. 588-96.
166. Brehm, A., et al., *dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties*. *Embo j*, 2000. **19**(16): p. 4332-41.
167. Le Gallo, M., et al., *Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes*. *Nat Genet*, 2012. **44**(12): p. 1310-5.
168. Sifrim, A., et al., *Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing*. *Nat Genet*, 2016. **48**(9): p. 1060-5.
169. Deciphering Developmental Disorders, S., *Prevalence and architecture of de novo mutations in developmental disorders*. *Nature*, 2017. **542**: p. 433.
170. Amaya, M., et al., *Mi2beta-mediated silencing of the fetal gamma-globin gene in adult erythroid cells*. *Blood*, 2013. **121**(17): p. 3493-501.
171. Curtis, C.D. and C.T. Griffin, *The chromatin-remodeling enzymes BRG1 and CHD4 antagonistically regulate vascular Wnt signaling*. *Mol Cell Biol*, 2012. **32**(7): p. 1312-20.
172. Zhao, H., et al., *The chromatin remodeler Chd4 maintains embryonic stem cell identity by controlling pluripotency- and differentiation-associated genes*. *J Biol Chem*, 2017. **292**(20): p. 8507-8519.
173. Minegishi, Y., et al., *Mutations in Igalpha (CD79a) result in a complete block in B-cell development*. *J Clin Invest*, 1999. **104**(8): p. 1115-21.
174. Schwickert, T.A., et al., *Stage-specific control of early B cell development by the transcription factor Ikaros*. *Nat Immunol*, 2014. **15**(3): p. 283-293.
175. Ng, S.Y.-M., et al., *Genome-wide Lineage-Specific Transcriptional Networks Underscore Ikaros-Dependent Lymphoid Priming in Hematopoietic Stem Cells*. *Immunity*, 2009. **30**(4): p. 493-507.
176. Reynaud, D., et al., *Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros*. *Nat Immunol*, 2008. **9**(8): p. 927-936.
177. Sperlazza, J., et al., *Depletion of the chromatin remodeler CHD4 sensitizes AML blasts to genotoxic agents and reduces tumor formation*. *Blood*, 2015. **126**(12): p. 1462-72.
178. Pallasch, C.P., et al., *Autoantibodies against GLEA2 and PHF3 in glioblastoma: tumor-associated autoantibodies correlated with prolonged survival*. *Int J Cancer*, 2005. **117**(3): p. 456-9.
179. Fischer, U., et al., *PHF3 expression is frequently reduced in glioma*. *Cytogenet Cell Genet*, 2001. **94**(3-4): p. 131-6.
180. Sroczynska, P., et al., *shRNA screening identifies JMJD1C as being required for leukemia maintenance*. *Blood*, 2014. **123**(12): p. 1870-1882.
181. Kinkelin, K., et al., *Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue*. *Proc Natl Acad Sci U S A*, 2013. **110**(38): p. 15277-82.

182. Gatchalian, J., et al., *Dido3 PHD modulates cell differentiation and division*. Cell Rep, 2013. **4**(1): p. 148-58.
183. Margolin, A.A., et al., *ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context*. BMC Bioinformatics, 2006. **7 Suppl 1**: p. S7.
184. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. Genome Res, 2003. **13**(11): p. 2498-504.
185. Shannon, P., et al., *Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks*. Genome Research, 2003. **13**(11): p. 2498-2504.
186. Ali, S.F., et al., *Translocation (6;15)(q12;q15): A Novel Mutation in a Patient with Therapy-Related Myelodysplastic Syndrome*. Case Reports in Hematology, 2015. **2015**: p. 5.
187. Renouf, S., et al., *Molecular cloning of a human polypeptide related to yeast sds22, a regulator of protein phosphatase-1*. FEBS Letters, 1995. **375**(1-2): p. 75-78.
188. Hisamoto, N., et al., *The EGP1 gene may be a positive regulator of protein phosphatase type 1 in the growth control of Saccharomyces cerevisiae*. Molecular and Cellular Biology, 1995. **15**(7): p. 3767-76.
189. Zencir, S., et al., *Identification of transcriptional and phosphatase regulators as interaction partners of human ADA3, a component of histone acetyltransferase complexes*. Biochemical Journal, 2013. **450**(2): p. 311-320.
190. Wang, T., et al., *hADA3 is required for p53 activity*. The EMBO Journal, 2001. **20**(22): p. 6404-6413.
191. Jiang, Y., et al., *Sds22/PP1 links epithelial integrity and tumor suppression via regulation of myosin II and JNK signaling*. Oncogene, 2011. **30**(29): p. 3248-60.
192. Cerveira, N., et al., *Genetic and clinical characterization of 45 acute leukemia patients with MLL gene rearrangements from a single institution*. Mol Oncol, 2012. **6**(5): p. 553-64.
193. De Braekeleer, E., et al., *Identification of MLL partner genes in 27 patients with acute leukemia from a single cytogenetic laboratory*. Mol Oncol, 2011. **5**(6): p. 555-63.
194. Sambrook, J. and D.W. Russell, *The inoue method for preparation and transformation of competent e. Coli: "ultra-competent" cells*. CSH Protoc, 2006. **2006**(1).
195. van Dongen, J.J., et al., *Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936*. Leukemia, 2003. **17**(12): p. 2257-317.
196. Peham, M., et al., *Low frequency of clonotypic Ig and T-cell receptor gene rearrangements in t(4;11) infant acute lymphoblastic leukaemia and its implication for the detection of minimal residual disease*. British Journal of Haematology, 2002. **117**(2): p. 315-321.
197. Huang, S. and L.W. Terstappen, *Lymphoid and myeloid differentiation of single human CD34+, HLA-DR+, CD38- hematopoietic stem cells*. Blood, 1994. **83**(6): p. 1515-26.
198. Meyer, C., et al., *Diagnostic tool for the identification of MLL rearrangements including unknown partner genes*. Proc Natl Acad Sci U S A, 2005. **102**(2): p. 449-54.
199. Gale, K.B., et al., *Backtracking leukemia to birth: Identification of clonotypic gene fusion sequences in neonatal blood spots*. Proceedings of the National Academy of Sciences, 1997. **94**(25): p. 13950-13954.
200. Jan, M., et al., *Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia*. Sci Transl Med, 2012. **4**(149): p. 149ra118.
201. Miyamoto, T., I.L. Weissman, and K. Akashi, *AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation*. Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7521-6.
202. Shlush, L.I., et al., *Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia*. Nature, 2014. **506**(7488): p. 328-33.
203. Rozovskaia, T., et al., *Upregulation of Meis1 and HoxA9 in acute lymphocytic leukemias with the t(4 : 11) abnormality*. Oncogene, 2001. **20**(7): p. 874-8.
204. Krivtsov, A.V., et al., *Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9*. Nature, 2006. **442**(7104): p. 818-822.

205. Horton, S.J., et al., *MLL-AF9-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny*. *Leukemia*, 2013. **27**(5): p. 1116-1126.
206. Li, W., et al., *The EMBL-EBI bioinformatics web and programmatic tools framework*. *Nucleic acids research*, 2015. **43**(W1): p. W580-4.
207. McWilliam, H., et al., *Analysis Tool Web Services from the EMBL-EBI*. *Nucleic acids research*, 2013. **41**(Web Server issue): p. W597-600.
208. Sievers, F., et al. *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. *Molecular Systems Biology*, 2011. **7**, 539 DOI: 10.1038/msb.2011.75.
209. Reizel, Y., et al., *Cell lineage analysis of the mammalian female germline*. *PLoS Genet*, 2012. **8**(2): p. e1002477.
210. Wasserstrom, A., et al., *Reconstruction of Cell Lineage Trees in Mice*. *PLoS ONE*, 2008. **3**(4): p. e1939.
211. Shlush, L.I., et al., *Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability*. *Blood*, 2012. **120**(3): p. 603-12.
212. Frumkin, D., et al., *Genomic Variability within an Organism Exposes Its Cell Lineage Tree*. *PLoS Computational Biology*, 2005. **1**(5): p. e50.
213. Zhang, C.Z., et al., *Calibrating genomic and allelic coverage bias in single-cell sequencing*. *Nat Commun*, 2015. **6**: p. 6822.
214. Nawy, T., *Single-cell sequencing*. *Nat Meth*, 2014. **11**(1): p. 18-18.
215. Grün, D. and A. van Oudenaarden, *Design and Analysis of Single-Cell Sequencing Experiments*. *Cell*. **163**(4): p. 799-810.
216. Zhao, H., et al., *The chromatin remodeler protein Chd4 maintains embryonic stem cell identity by controlling pluripotency- and differentiation-associated genes*. *Journal of Biological Chemistry*, 2017.
217. Chudnovsky, Y., et al., *ZFH4 Interacts with the NuRD Core Member CHD4 and Regulates the Glioblastoma Tumor-Initiating Cell State*. *Cell Reports*, 2014. **6**(2): p. 313-324.
218. Danis, E., et al., *Ezh2 Controls an Early Hematopoietic Program and Growth and Survival Signaling in Early T Cell Precursor Acute Lymphoblastic Leukemia*. *Cell reports*, 2016. **14**(8): p. 1953-1965.
219. Hennecke, M., et al., *Composition and arrangement of genes define the strength of IRES-driven translation in bicistronic mRNAs*. *Nucleic Acids Research*, 2001. **29**(16): p. 3327-3334.
220. de Felipe, P., et al., *E unum pluribus: multiple proteins from a self-processing polyprotein*. *Trends Biotechnol*, 2006. **24**(2): p. 68-75.
221. Shanbhag, N.M., et al., *ATM-Dependent Chromatin Changes Silence Transcription In cis to DNA Double-Strand Breaks*. *Cell*, 2010. **141**(6): p. 970-981.
222. Shi, Y., et al., *Sharp, an inducible cofactor that integrates nuclear receptor repression and activation*. *Genes & Development*, 2001. **15**(9): p. 1140-1151.
223. Ruella, M., et al., *Dual CD19 and CD123 targeting prevents antigen-loss relapses after CD19-directed immunotherapies*. *The Journal of Clinical Investigation*, 2016. **126**(10): p. 3814-3826.
224. Pruitt, K.D., et al., *The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes*. *Genome Research*, 2009. **19**(7): p. 1316-1323.
225. Ariyoshi, M. and J.W.R. Schwabe, *A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling*. *Genes & Development*, 2003. **17**(15): p. 1909-1920.
226. Hnisz, D., et al., *Super-Enhancers in the Control of Cell Identity and Disease*. *Cell*. **155**(4): p. 934-947.
227. de la Cruz, N.B., et al., *Characterization of the Tn5 transposase and inhibitor proteins: a model for the inhibition of transposition*. *J Bacteriol*, 1993. **175**(21): p. 6932-8.
228. Isberg, R.R., A.L. Lazaar, and M. Syvanen, *Regulation of Tn5 by the right-repeat proteins: control at the level of the transposition reaction?* *Cell*, 1982. **30**(3): p. 883-92.
229. Johnson, R.C., J.C. Yin, and W.S. Reznikoff, *Control of Tn5 transposition in Escherichia coli is mediated by protein from the right repeat*. *Cell*, 1982. **30**(3): p. 873-82.

230. Yin, J.C. and W.S. Reznikoff, *p2 and inhibition of Tn5 transposition*. Journal of Bacteriology, 1988. **170**(7): p. 3008-3015.
231. Bonifer, C., *Epigenetic Plasticity of Hematopoietic Cells*. Cell Cycle, 2005. **4**(2): p. 214-217.
232. Kondo, M., et al., *Cell-fate conversion of lymphoid-committed progenitors by instructive actions of cytokines*. Nature, 2000. **407**(6802): p. 383-6.
233. Iwasaki-Arai, J., et al., *Enforced granulocyte/macrophage colony-stimulating factor signals do not support lymphopoiesis, but instruct lymphoid to myelomonocytic lineage conversion*. J Exp Med, 2003. **197**(10): p. 1311-22.
234. Borzillo, G.V., R.A. Ashmun, and C.J. Sherr, *Macrophage lineage switching of murine early pre-B lymphoid cells expressing transduced fms genes*. Molecular and Cellular Biology, 1990. **10**(6): p. 2703-2714.
235. Nutt, S.L., et al., *Commitment to the B-lymphoid lineage depends on the transcription factor Pax5*. Nature, 1999. **401**(6753): p. 556-62.
236. Horcher, M., A. Souabni, and M. Busslinger, *Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis*. Immunity, 2001. **14**(6): p. 779-90.
237. Xie, H., et al., *Stepwise reprogramming of B cells into macrophages*. Cell, 2004. **117**(5): p. 663-76.
238. Blanpain, C. and E. Fuchs, *Plasticity of epithelial stem cells in tissue regeneration*. Science, 2014. **344**(6189).
239. Watson, A.A., et al., *The PHD and Chromo Domains Regulate the ATPase Activity of the Human Chromatin Remodeler CHD4*. Journal of Molecular Biology, 2012. **422**(1-2): p. 3-17.
240. Morra, R., et al., *Biophysical Characterization of Chromatin Remodeling Protein CHD4*. Methods Mol Biol, 2016. **1431**: p. 175-93.
241. Clark, D.J., *Nucleosome Positioning, Nucleosome Spacing and the Nucleosome Code*. Journal of biomolecular structure & dynamics, 2010. **27**(6): p. 781-793.
242. Zhang, Y., et al., *The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities*. Cell, 1998. **95**(2): p. 279-89.
243. Nakamura, T., et al., *ALL-1 Is a Histone Methyltransferase that Assembles a Supercomplex of Proteins Involved in Transcriptional Regulation*. Molecular Cell, 2002. **10**(5): p. 1119-1128.
244. Butterfield, L.H., *Cancer vaccines*. Bmj, 2015. **350**: p. h988.