

**PHF6 as a novel interactor of the transcriptional co-regulator LMO2  
in T-Cell acute lymphoblastic leukaemia (T-ALL) and myeloid  
differentiation**

By

Sarah Saeed Binhassan

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences  
Birmingham Centre for Genome Biology  
College of Medical and Dental Sciences  
University of Birmingham  
September 2019

UNIVERSITY OF  
BIRMINGHAM

**University of Birmingham Research Archive**

**e-theses repository**

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

## Abstract

The LIM-Only protein 2 (LMO2) is a transcriptional co-regulator that forms a multi-protein complex with TAL1, GATA, and LDB1. This complex regulates transcription from the onset of haematopoietic development and differentiation. However, LMO2 is normally downregulated during T-cell development. Genetic abnormalities that trigger the aberrant expression of LMO2 and other members of the complex lead to T-cell acute lymphoblastic leukaemia (T-ALL). The components of the LMO2 complex are not widely explored in T-ALL and its functions are not fully characterised in this type of leukaemia.

Through proteomic analysis we identified the Plant homeodomain zinc finger protein 6 (PHF6) as a novel interactor of LMO2. Mutations of *PHF6* have been found to occur in several types of leukaemia including T-ALL. We show a functional interaction between PHF6 and the LMO2 complex members by genome-wide approaches in T-ALL. These outcomes indicated that the LMO2 complex recruits PHF6 to the DNA and regulates gene expression. Importantly, we identified many genes of transcription factors, which are involved in haematopoietic stem cell (HSC) self-renewal and development of early T-cell progenitors, as putative targets of LMO2/PHF6 complex, providing compelling indications of an oncogenic mechanism of the complex.

Additionally, *Phf6* gene knockdown in myeloid cells suggested a role for PHF6 in the differentiation of this lineage. Finally, loss of PHF6 and LMO2 caused chromosomal instability and signs of DNA damage in these cells, implying a possible function in chromosome and genome stability.

## **Acknowledgments**

Firstly, I express my sincere gratitude to my supervisors Dr. Maarten Hoogenkamp and Dr. Vesna Stanulović. They have been great teachers and offered tremendous amount of knowledge, mentorship, and support over the past years. They welcomed me with their generosity and kindness since I started and have become friends to me. Thanks to the rest of the Hoogenkamp group who were helpful and fun to work with. I also thank my co-supervisor Prof. Constanze Bonifer for her support and encouragement.

I am beyond grateful to my mother who was my rock and backbone providing unconditional love, morals, support, and empowerment. I am also extremely thankful to my late father who has taught me invaluable lessons in life and had always believed in me. I would like to extend a special thanks to Mohammad for helping me find the strength within me. I also thank my wonderful sons, Faisal and Bader who went through this journey with me giving me continuous joy and happiness. They offered me the greatest blessings and always looked at me with pride. Thanks to my siblings who showed me care, love, and encouragement.

Finally, I thank my friends Farah, Malak, Rana, Ruba, and Zain for uplifting me when I was down and being great listeners. They were always there for me when I needed them the most and helped me overcome the obstacles to achieve my goals.

This work was supported by the Saudi Cultural Bauer and King Saud University.



## Table of contents

<b>1</b>	<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1	<b>Haematopoiesis.....</b>	<b>1</b>
1.1.1	Embryonic haematopoiesis.....	1
1.1.2	Adult haematopoiesis .....	2
1.1.3	The classical model of haematopoiesis .....	3
1.1.4	The alternative models of haematopoiesis.....	5
1.2	<b>T-cell development.....</b>	<b>7</b>
1.2.1	T-cell identity genes.....	13
1.3	<b>Transcription factors role in T-cell development .....</b>	<b>14</b>
1.3.1	Phase 1.....	15
1.3.2	Phase 2.....	18
1.3.3	Phase 3.....	21
1.4	<b>T-cell Acute Lymphoblastic Leukaemia .....</b>	<b>21</b>
1.4.1	The TAL/LMO subgroup of T-ALL.....	25
1.5	<b>LIM domain only 2 (LMO2).....</b>	<b>28</b>
1.5.1	History of LMO2.....	28
1.5.2	LMO2 structure and DNA-binding complexes.....	31
1.5.3	LMO2 function in normal haematopoiesis and vascular development ...	33
1.5.4	LMO2 oncogenic role in T-ALL .....	35
1.5.5	LMO2 complexes in T-ALL .....	39
1.6	<b>Plant homeodomain zinc finger protein 6 (PHF6).....</b>	<b>41</b>
1.6.1	History of PHF6 .....	41
1.6.2	PHF6 structure.....	42
1.6.3	Functional interactions of PHF6.....	43
1.6.4	PHF6 functions in haematopoiesis and leukaemia .....	46
1.6.5	<i>PHF6</i> mutations in leukaemia .....	49
1.7	<b>Aims and objectives.....</b>	<b>51</b>
<b>2</b>	<b>Chapter 2: Materials and methods .....</b>	<b>53</b>
2.1	<b>Cell culture.....</b>	<b>53</b>
2.2	<b>Cloning steps for cDNA .....</b>	<b>54</b>
2.2.1	RNA isolation .....	54
2.2.2	cDNA synthesis .....	54
2.2.3	Amplification of cDNA .....	54
2.2.4	Cloning of cDNA .....	55

<b>2.3</b>	<b>shRNA mediated knockdown</b>	<b>59</b>
2.3.1	Cloning of shRNA expression vectors	59
2.3.2	Retroviral production and transduction of PUER cells	61
<b>2.4</b>	<b>Morphological analysis</b>	<b>64</b>
<b>2.5</b>	<b>Flow cytometry</b>	<b>64</b>
<b>2.6</b>	<b>Nuclear extract</b>	<b>65</b>
<b>2.7</b>	<b>Immunoprecipitation (IP)</b>	<b>66</b>
<b>2.8</b>	<b>Western blotting</b>	<b>66</b>
<b>2.9</b>	<b>Mass Spectrometry (MS)</b>	<b>67</b>
<b>2.10</b>	<b>Immunostaining</b>	<b>68</b>
<b>2.11</b>	<b>Cell proliferation assay</b>	<b>70</b>
<b>2.12</b>	<b>Proximity ligation assay (PLA)</b>	<b>71</b>
<b>2.13</b>	<b>Cytogenetics</b>	<b>73</b>
<b>2.14</b>	<b>RNA sequencing (RNA-seq)</b>	<b>74</b>
2.14.1	RNA-seq library preparation	74
2.14.2	Library quantification by qPCR	75
2.14.3	RNA-seq data analysis	76
<b>2.15</b>	<b>Chromatin Immunoprecipitation (ChIP)</b>	<b>77</b>
2.15.1	Chromatin preparation	77
2.15.2	ChIP	78
2.15.3	ChIP sample validation by qPCR	79
2.15.4	ChIP-seq library preparation	80
2.15.5	ChIP-seq data analysis	81
<b>2.16</b>	<b>DNaseI hypersensitive site (DHS) mapping</b>	<b>82</b>
2.16.1	DNaseI-seq assay and library preparation	82
2.16.2	DHS data analysis	83
<b>2.17</b>	<b>Data visualisation</b>	<b>83</b>
<b>3</b>	<b>Chapter 3: PHF6 as a new interactor of LMO2 in T-ALL</b>	<b>85</b>
<b>3.1</b>	<b>Introduction</b>	<b>85</b>
<b>3.2</b>	<b>Results</b>	<b>87</b>
3.2.1	LMO2, TAL1 and LDB1 bind regulatory elements	87
3.2.2	The LMO2/TAL1 complex binds TFs motifs in SIL/TAL1 and ARR cells	90
3.2.3	LMO2/TAL1 complex targets distinct genes in different cell lines	95
3.2.4	LMO2/TAL1 binding sites are associated with differentially expressed genes in T-ALLs	97

3.2.5	Identification of PHF6 as a potential interactor of LMO2 in T-ALL cell lines .....	107
3.2.6	<i>PHF6</i> is expressed in T-ALL cells.....	111
3.2.7	PHF6 as a novel interactor of LMO2.....	113
3.2.8	PHF6 binds with LMO2 at <i>RUNX1</i> enhancer .....	115
3.2.9	PHF6 co-occupies genomic sites and regulatory elements with LMO2, TAL1, and LDB1 .....	116
3.2.10	PHF6 peaks are co-localised by LMO2, TAL1 and LDB1 .....	121
3.2.11	DHS mapping identifies regions accessible by PHF6 and LMO2 complex members.....	123
3.2.12	A high percentage of PHF6/LMO2/TAL1/LDB1 common peaks bind to genomic regions distant from TSSs .....	125
3.2.13	PHF6 peaks occupy haematopoietic TF motifs.....	126
3.2.14	PHF6/LMO2/TAL1/LDB1 co-localised regions are linked to genes involved in transcriptional regulation .....	130
3.2.15	PHF6/LMO2/TAL1/LDB1 binding sites are associated with differentially expressed genes in T-ALLs .....	135
3.2.16	PHF6 is associated with active enhancers in DU.528 cells.....	147
3.2.17	Identification of PHF6 interacting partners by mass spectrometry .....	150
<b>3.3</b>	<b>Discussion .....</b>	<b>154</b>
3.3.1	ChIP experiments confirmed LMO2/TAL1 binding to regulatory elements .....	154
3.3.2	Commonalities and differences between gene expression in ARR and SIL/TAL1 cell lines .....	156
3.3.3	PHF6 as a novel interactor of LMO2.....	159
3.3.4	PHF6, LMO2, TAL1, and LDB1 co-localise at regulatory elements and TF motifs.....	160
3.3.5	PHF6, LMO2, TAL1, and LDB1 common peaks are associated with haematopoietic TF genes .....	162
3.3.6	PHF6 has multiple interacting partners .....	168
<b>4</b>	<b>Chapter 4: The role of PHF6 in myeloid cells .....</b>	<b>170</b>
<b>4.1</b>	<b>Introduction .....</b>	<b>170</b>
<b>4.2</b>	<b>Results .....</b>	<b>171</b>
4.2.1	Identification of new LMO2 Protein–Protein interactions in myeloid progenitors.....	171
4.2.2	Retroviral mediated <i>Phf6</i> knockdown in PUER cells.....	171
4.2.3	Knockdown of <i>Phf6</i> slows myeloid differentiation .....	174
4.2.4	<i>Phf6</i> deprivation affects gene expression .....	178

4.2.5	LMO2, TAL1 and LDB1 binding is redistributed in the absence of PHF6 .....	185
4.2.6	LMO2 binds to novel sites after <i>Phf6</i> depletion.....	188
4.2.7	shPhf6 cells feature distinctive binding motifs.....	190
4.2.8	<i>Phf6</i> knockdown affects genome and chromosome stability.....	194
4.2.9	<i>Lmo2</i> knockdown and overexpression shows a similar phenotype to shPhf6 .....	198
<b>4.3</b>	<b>Discussion .....</b>	<b>202</b>
4.3.1	LMO2 physically interacts with PHF6 in common myeloid progenitors	202
4.3.2	Knockdown of <i>Phf6</i> shows signs of morphological abnormalities and delayed differentiation.....	203
4.3.3	<i>Phf6</i> depletion affects genes involved in haematopoiesis and chromatin assembly.....	204
4.3.4	<i>Phf6</i> depletion repositions LMO2, TAL1, and LDB1 binding sites.....	206
4.3.5	Genome and chromosome stability are affected by <i>Phf6</i> and <i>Lmo2</i> expression levels .....	207
<b>5</b>	<b>Chapter 5: Overall discussion .....</b>	<b>211</b>
5.1	Role of the LMO2/TAL1 complex in T-ALL.....	211
5.2	Identifying PHF6 as a novel interactor of LMO2 .....	212
5.3	The transcriptional role of PHF6 .....	214
5.4	Limitations .....	216
5.5	Final conclusions .....	218
5.6	Future perspectives .....	219
	References .....	221

## list of figures

Figure 1.1 The classical model of haematopoiesis. ....	5
Figure 1.2 The alternative model of haematopoiesis. ....	6
Figure 1.3 The CLOUD-HSPCs model of haematopoiesis. ....	7
Figure 1.4 Overview of T-cell development in mice. ....	9
Figure 1.5 Model of T-cell development in human. ....	13
Figure 1.6 Stage specific patterns of gene expression for important transcriptional regulators during T-cell development. ....	16
Figure 1.7 T-ALL genetic subgroups characterised according to cluster analysis of T-ALL patient samples. ....	25
Figure 1.8 Schematic representation of LMO2 protein structure. ....	31
Figure 1.9 LMO2 protein complexes. ....	33
Figure 1.10 Schematic representation of PHF6 protein structure. ....	43
Figure 1.11 PHF6 interaction with NuRD complex. ....	45
Figure 2.1 cDNA cloning in pDrive plasmid. ....	56
Figure 2.2 cDNA cloning in MIGR1 plasmid. ....	58
Figure 2.3 shRNA cloning in pMSCVhyg plasmid. ....	61
Figure 2.4 shRNA cloning in pMSCV-IRES-GFP plasmid. ....	63
Figure 2.5 Proximity ligation assay (PLA). ....	72
Figure 2.6 Representation of RNA-seq libraries run on 3% gel. ....	75
Figure 2.7 Representation of ChIP-seq libraries run on 3% gel. ....	81
Figure 2.8 Representation of DNaseI digestion validation gel. ....	83
Figure 3.1 Immunophenotype of T-ALL cell lines. ....	87
Figure 3.2 ChIP qPCR analyses of LMO2, TAL1 and LDB1 binding in T-ALL cells. ....	88
Figure 3.3 Binding profiles of LMO2, TAL1 and LDB1 in T-ALL cells. ....	89
Figure 3.4 ChIP-seq analyses of LMO2 and TAL1 in SIL/TAL1 cells. ....	91
Figure 3.5 GREAT Region-Gene Association graphs of genomic regions bound by LMO2/TAL1 shared peaks. ....	92
Figure 3.6 ChIP-seq analyses of LMO2 and TAL1 comparing ARR and SIL/TAL1 cells. ....	94
Figure 3.7 Intersection of LMO2/TAL1 peaks and differentially expressed genes that are potential targets of the LMO2/TAL1 complex in ARR and SIL/TAL1 cells. ....	96
Figure 3.8 RNA-seq analysis of differentially expressed genes associated with LMO2/TAL1 unique peaks in ARR and SIL/TAL1 cell lines. ....	98
Figure 3.9 GO analysis on differentially expressed genes associated with LMO2/TAL1 unique peaks in ARR cells show genes involved in RNA processes and nucleosome assembly. ....	100
Figure 3.10 GO analysis on differentially expressed genes associated with LMO2/TAL1 unique peaks in SIL/TAL1 cells show genes involved in cellular and metabolic processes. ....	101
Figure 3.11 RNA-seq analysis of differentially expressed genes associated with LMO2/TAL1 common and distinct peaks in ARR and SIL/TAL1 cell lines. ....	102

Figure 3.12 GO analysis on differentially expressed genes associated with LMO2/TAL1 common peaks in ARR and SIL/TAL1 cells show genes involved in regulation of translation, regulation of gene expression, and T and B cell signalling pathways. ....	104
Figure 3.13 GO analysis on differentially expressed genes associated with LMO2/TAL1 distinct peaks in ARR and SIL/TAL1 cells show genes involved in apoptotic processes and regulation of GTPase activity. ....	105
Figure 3.14 GO analysis on differentially expressed genes associated with LMO2/TAL1 distinct peaks in ARR and SIL/TAL1 cells show genes involved in leukocytes differentiation and development, and regulation of transcription. ....	106
Figure 3.15 Mass Spectrometry assay, T-ALL cell lines immunoprecipitation with LMO2 and IgG antibodies. ....	109
Figure 3.16 Mass Spectrometry analyses show PHF6 as a potential interactor of LMO2. ....	110
Figure 3.17 PHF6 expression in T-ALL cell lines. ....	112
Figure 3.18 PHF6 is a novel interacting partner of LMO2. ....	113
Figure 3.19 <i>In situ</i> Proximity Ligation Assay (PLA) showing LMO2 and PHF6 interaction in HSB2 and CCRF-CEM cells. ....	115
Figure 3.20 ChIP qPCR analyses of PHF6 in T-ALL cells. ....	116
Figure 3.21 PHF6 co-occupies genomic sites with LMO2, TAL1, and LDB1 in T-ALL cells. ....	117
Figure 3.22 PHF6 peaks overlap with LMO2, TAL1, and LDB1 peaks in T-ALL cells at the <i>RUNX1</i> gene. ....	119
Figure 3.23 PHF6 co-localises with LMO2, TAL1, and LDB1 in T-ALL cells at the <i>NFE2</i> regulatory element. ....	120
Figure 3.24 PHF6 peaks are co-localised with LMO2, TAL1 and LDB1. ....	122
Figure 3.25 PHF6, LMO2, TAL1, and LDB1 bind at DHSs in T-ALL cells. ....	124
Figure 3.26 GREAT Region-Gene Association graphs of genomic regions bound by PHF6, LMO2, TAL1, and LDB1 common peaks. ....	126
Figure 3.27 Motifs enriched at PHF6 binding sites in T-ALL cells. ....	127
Figure 3.28 Motifs enriched at PHF6, LMO2, TAL1, and LDB1 common binding sites in T-ALL cells. ....	128
Figure 3.29 Motifs enriched at PHF6, LMO2, TAL1, and LDB1 common binding sites that are within 50 kb of TSSs in T-ALL cells. ....	130
Figure 3.30 GO analysis show genes associated with cellular processes of lymphocytes and regulation of transcription in ARR cells. ....	132
Figure 3.31 GO analysis show genes associated with regulation of MAP kinase activity, vascular endothelial growth factor receptor signalling pathway, and regulation of transcription in DU.528 cells. ....	133
Figure 3.32 GO analysis show genes associated with JAK-STAT cascade, regulation of transcription, and germ cell development in HSB2 cells. ....	134

Figure 3.33 GO analysis show genes associated with T-helper 17 cell lineage commitment, and regulation of H3-K9 methylation in CCRF-CEM cells.....	135
Figure 3.34 PHF6/LMO2/TAL1/LDB1 common peaks in T-ALL cells are associated with genes involved in HSC proliferation. ....	136
Figure 3.35 RNA-seq analyses of ARR line.....	137
Figure 3.36 GO analysis on differentially expressed genes associated with PHF6/LMO2/TAL1/LDB1 shared peaks in ARR cells show genes involved in regulation of transcription and haematopoiesis. ....	139
Figure 3.37 RNA-seq analyses of DU.528 cell line.....	140
Figure 3.38 GO analysis on differentially expressed genes associated with PHF6/LMO2/TAL1/LDB1 shared peaks in DU.528 cells show genes involved in lymphocyte processes, regulation of MAPK cascade, and chromatin remodelling. .	142
Figure 3.39 RNA-seq analysis of HSB2 cell line.....	143
Figure 3.40 GO analysis on differentially expressed genes associated with PHF6/LMO2/TAL1/LDB1 shared peaks in HSB2 cells show genes involved in B-cell processes, haematopoiesis, and stem cell differentiation.....	144
Figure 3.41 RNA-seq analysis of CCRF-CEM cell line.....	145
Figure 3.42 GO analysis on differentially expressed genes associated with PHF6/LMO2/TAL1/LDB1 shared peaks in CCRF-CEM cells show genes involved in regulation of GTPase activity, vascular endothelial growth factor signalling pathway, and mitotic cell cycle.....	146
Figure 3.43 PHF6 peaks bind H3K27ac enriched regions in DU5.528 cells.....	148
Figure 3.44 Binding profiles of PHF6, LMO2, TAL1, LDB1, and H3K27ac in DU.528 cells. ....	149
Figure 3.45 Mass Spectrometry assay, T-ALL cell lines immunoprecipitation with PHF6 and IgG antibodies. ....	150
Figure 3.46 Mass Spectrometry analyses show coprecipitation of ESWR1 with PHF6 in all four cell lines. ....	152
Figure 3.47 Venn diagrams showing overlap of proteins pulled down with LMO2 (yellow circle) versus PHF6 (blue circle) in T-ALL cells. ....	153
Figure 4.1 Flow cytometry analysis of transfected PlatE cells. ....	173
Figure 4.2 <i>Phf6</i> knockdown in PUER cells. ....	174
Figure 4.3 Kwik-Diff staining of shCtrl and shPhf6 cells. ....	175
Figure 4.4 Flow cytometry analyses of shCtrl and shPhf6 cell differentiation. ....	177
Figure 4.5 RNA-seq analyses of shCtrl and shPhf6 at D0, D1 and D3 of differentiation.....	179
Figure 4.6 Intersection of differentially expressed genes in shCtrl and shPhf6 PUER cells. ....	181
Figure 4.7 <i>Phf6</i> deprivation causes changes in gene expression levels.....	182
Figure 4.8 Gene ontology analysis performed on all differentially expressed genes between shCtrl and shPhf6 cells. ....	183

Figure 4.9 Gene ontology enrichment analyses for biological process performed on (A) cluster 3, (B) cluster 4 as identified in Figure 4.7. ....	185
Figure 4.10 Intersections of LMO2, TAL1 and LDB1 ChIP-seq data in shCtrl and shPhf6 cells. ....	187
Figure 4.11 Screenshot from the UCSC browser showing LMO2, TAL1 and LDB1 binding profiles in shCtrl and shPhf6 cells at <i>Gnb1l</i> locus. ....	188
Figure 4.12 Heat map plots of ChIP-seq data for LMO2, PHF6, TAL1 and LDB1 in shCtrl and shPhf6 cells. ....	190
Figure 4.13 <i>De novo</i> motif enrichment analyses within LMO2 peaks in shCtrl and shPhf6 cells. ....	191
Figure 4.14 <i>De novo</i> motif enrichment analyses within TAL1 peaks in shCtrl and shPhf6 cells. ....	192
Figure 4.15 <i>De novo</i> motif enrichment analyses within LDB1 peaks in shCtrl and shPhf6 cells. ....	193
Figure 4.16 <i>De novo</i> motif enrichment analyses within common peaks for LMO2, TAL1 and LDB1 in shCtrl and shPhf6 cells. ....	194
Figure 4.17 Chromosomal spread analyses in shCtrl and shPhf6 cells. ....	195
Figure 4.18 Flow cytometry analysis of ethidium bromide (EtBr) stained cells labelled with ethynyldeoxyuridine (EdU). ....	196
Figure 4.19 <i>Phf6</i> depletion shows signs of DNA damage. ....	197
Figure 4.20 Kwik-Diff staining of shLmo2 and HA-LMO2 cells. ....	199
Figure 4.21 Chromosome spreads of shLmo2 and HA-LMO2 cells. ....	200
Figure 4.22 Chromosome counts of shCtrl, shPhf6, shLmo2, and HA-LMO2 cells. ....	201



## List of tables

Table 1.1 List of chromosomal rearrangements involving the T-cell receptor gene...	24
Table 2.1 cDNA amplification primers. ....	55
Table 2.2 Thermocycling conditions for cDNA amplification by PCR. ....	55
Table 2.3 pDrive sequencing primers. ....	57
Table 2.4 shRNA oligonucleotides. ....	59
Table 2.5 pSM2C universal primers. ....	60
Table 2.6 Thermocycling conditions for shRNA amplification by PCR. ....	60
Table 2.7 Flow cytometry antibodies. ....	65
Table 2.8 Primary antibodies. ....	69
Table 2.9 Secondary antibodies. ....	70
Table 2.10 qPCR conditions. ....	76
Table 2.11 ChIP antibodies. ....	78
Table 2.12 ChIP qPCR primers. ....	80

## List of abbreviations

AAV	Adeno-associated virus
AML	Acute myeloid leukaemia
ATAC-seq	Assay for transposase-accessible chromatin sequencing
ATM	Ataxia telangiectasia mutated
B-ALL	B-cell acute lymphoblastic leukaemia
BCL11B	BAF Chromatin Remodeling Complex Subunit
BFLS	Börjesson-Forssman-Lehman syndrome
bHLH	basic Helix loop helix
CASTing	Cyclic amplification and selection of targets
CD	Cluster of differentiation
CDK	Cyclin-dependent kinase
CDKN2	Cyclin-dependent kinase 2
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CLOUD-HSPCs	Continuum of low-primed undifferentiated haematopoietic stem and progenitor cells
CLP	Common lymphoid progenitors
CML	Chronic myeloid leukaemia
CMP	Common myeloid progenitors
CRE	cis-regulatory element
CRISPR	Clustered regularly interspaced short palindromic repeats
DAPI	4',6-diamidino-2-phenylindole
DHS	DNaseI hypersensitivity site
DN	Double negative
DNA	Deoxyribonucleic acid
DNaseI-seq	Deoxyribonuclease I sequencing
DP	Double positive
dsDNA	double stranded Deoxyribonucleic acid
EdU	Ethynyldeoxyuridine
EKLF	Erythroid Krüppel-like factor
EMSA	Electrophoretic mobility shift assay
ER	Estrogen receptor
ES cells	Embryonic stem cells
EtBr	Ethidium bromide
ETP	Early T-cell progenitor
ETP-ALL	Early T-cell precursors acute lymphoblastic leukaemia
ETS	Erythroblast Transformation Specific
EWSR1	EWS RNA Binding Protein 1
FACS	Fluorescence-activated cell sorting
FBS	Foetal bovine serum
FLT	Fms Related Receptor Tyrosine Kinase 1
FOG1	Friend of GATA1

FPKM	Fragments per kilobase of transcript per million mapped reads
GATA	GATA binding protein
GFP	Green fluorescent protein
GMP	Granulocyte/monocyte progenitors
GO	Gene ontology
Gtf	Gene transfer files
H2AX	H2A.X Variant Histone
H3K27ac	Acetylated lysine 27 of histone H3
H3K4me3	Trimethylated lysine 4 of histone H3
HDAC	Histone deacetylase
HHEX	Hematopoietically Expressed Homeobox
HSC	Hematopoietic stem cell
IL	Interleukin
IL2RG	Interleukin 2 receptor-γ
IM	Immunoprecipitation
IP	Immunoprecipitation
IRES	Internal ribosome entry site
ISP	Immature single positive
KD	Knockdown
KIT	KIT Proto-Oncogene, Receptor Tyrosine Kinase
KSL	Kaurene synthase-like protein
LB	Lysogeny Broth
LDB1	LIM domain binding1
LMO	LIM-domain only
LMPP	Lymphoid-primed multipotent progenitor
LSK	Lin/Sca-1/c-Kit
LT-HSC	Long-term hematopoietic stem cells
LYL1	Lymphoblastic leukaemia associated haematopoiesis regulator 1
MEP	Megakaryocyte/Erythroid progenitor
mES cells	mouse Embryonic stem cells
MHC	Major histocompatibility complex
MPP	Multipotent progenitor
MS	Mass spectrometry
NFE2	Nuclear Factor, Erythroid 2
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NK	Natural killer
NLS	Nuclear localization sequences
NMR	Nuclear magnetic resonance
NoLS	Nucleolar localization sequence
NOTCH1	NOTCH Receptor 1
NuRD	Nucleosome Remodelling and Deacetylation
OHT	4-hydroxytamoxifen

ORF	Open reading frame
PAF1	RNA Polymerase II Associated Factor 1
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PHD	Plant homeodomain
PHF6	Plant homeodomain zinc finger protein 6
PLA	Proximity ligation assays
RAG	Recombination activating gene
RCA	Rolling-circle amplification
RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
rRNA	ribosomal RNA
RT	Room temperature
RT-PCR	real-time PCR
RUNX1	RUNX Family Transcription Factor 1
SCID-X1	X-linked severe combined immunodeficiency
sgRNA	single-guide RNA
shRNA	short hairpin RNA
SIL	SCL interrupting locus
SP	Single positive
SPI1	Spi-1 Proto-Oncogene
ST-HSCs	Short-term hematopoietic stem cells
TAL1	T-cell acute lymphoblastic leukaemia
T-ALL	T-cell Acute lymphoblastic leukaemia
TCF-1	Transcription factor 1
TCR	T-cell receptor
TF	Transcription factor
TNF $\alpha$	Tumour necrosis factor alpha
TSS	Transcription start site
UBF	Upstream binding factor
UTR	Untranslated region
WB	Western Blot
ZK	Zinc knuckles

# **1 Chapter 1: Introduction**

## **1.1 Haematopoiesis**

Haematopoiesis is the process by which all blood cell types are formed. It starts during early embryonic development and throughout adulthood to produce the blood system. It includes processes of formation, development, and differentiation of blood cells (Jagannathan-Bogdan and Zon, 2013). The haematopoietic system is comprised of more than ten different blood cell types with very distinct functions. Despite their vast functional diversity, all mature blood cells originate from a common cell type, the hematopoietic stem cell (HSCs), which has the potential to promote all blood lineages. Red blood cells (Erythrocytes) control oxygen and carbon dioxide transport between tissue cells and the lungs. White blood cells (Leukocytes) embody many cell types essential for innate and adaptive immunity. Megakaryocytes produce platelets (Thrombocytes) essential for blood clotting and wound healing (Rieger and Schroeder, 2009, Morrison et al., 1995).

During embryonic development, haematopoiesis occurs gradually in different anatomical locations, from the yolk sac to the liver, to the spleen and to the bone marrow (Dzierzak and Bigas, 2018). During adulthood, haematopoiesis takes place in the bone marrow, spleen and thymus (Jones and Wagers, 2008).

### **1.1.1 Embryonic haematopoiesis**

In mammals, the hematopoietic system develops from two distinct waves of precursors. The first wave starts in the yolk sac where nucleated primitive erythrocytes are produced. Myeloid cells also emerge from the yolk sac, which migrate to the central nervous system and to the skin to form tissue specific macrophages (Palis et al., 1999,

Schulz et al., 2012). The second wave happens in the aorta-gonad-mesonephros region of the developing embryo. In particular the dorsal aorta contains specialised endothelial cells, called haemogenic endothelium, which undergo endothelial-to-hematopoietic cell transition. This process produces multipotential hematopoietic progenitors, including erythroid, myeloid, and lymphoid cells, as well as the first definitive HSCs. These HSCs are produced once in the lifetime and migrate to the foetal liver and placenta, where they expand. They further colonise the spleen where differentiation to the myeloid lineage and particular lymphoid subsets is favoured, before they finally move to the bone marrow, where they will reside for the rest of life (Chen et al., 2011, Yoshimoto et al., 2012, Medvinsky and Dzierzak, 1996, Muller et al., 1994, Golub and Cumano, 2013).

### **1.1.2 Adult haematopoiesis**

Adult HSCs are at the apex of the multilineage haematopoietic hierarchy, which features multiple progenitor cells with increasingly lower lineage potentials that give rise to all blood cell types (Rieger and Schroeder, 2009). HSCs have the innate capacity to self-renew and to differentiate. The ability to self-renew ensures sustained production of comparatively short-lived mature blood cells throughout life. Functional understanding of haematopoiesis emerged in the 1960s when the identification of rare cells in the bone marrow took place. These cells were characterised by the ability to clonally repopulate *in vivo*, forming macroscopic colonies on recipient mice spleens (Till and Mc, 1961, Becker et al., 1963).

With the development of fluorescence-activated cell sorting (FACS) and the identification of cell surface markers it became possible to characterise cells by their immunophenotype (Herzenberg et al., 1976, Muller-Sieburg et al., 1986). The

immature multipotent cells were contained in the LSK (Lin/Sca-1/Kit) cell population. These cells express stem cell antigen-1 (Sca-1) and CD117 (Kit) but lack the lineage (Lin) markers, which are expressed on lineage committed cells. The LSK cells are responsible for producing variable types of mature blood cell lineages (Spangrude et al., 1988, Uchida and Weissman, 1992). Additional markers were identified to separate the LSK population into long-term HSCs (LT-HSCs) and short-term HSCs (ST-HSCs). The LT-HSCs are characterised by CD34<sup>-</sup> / FLT3<sup>-</sup> / CD150<sup>+</sup> expression markers and have long term repopulating activity. The ST-HSCs are identified by CD34<sup>+</sup> / FLT3<sup>-</sup> and have limited self-renewal activity (Kiel et al., 2005, Christensen and Weissman, 2001, Yang et al., 2005, Adolfsson et al., 2005).

HSCs go through a multistep differentiation process from progenitor cells with a decreasing multilineage potential to committed single lineage cells (Kondo et al., 1997). Adult haematopoiesis is a hierarchical process and the described models of this hierarchy are constantly progressing.

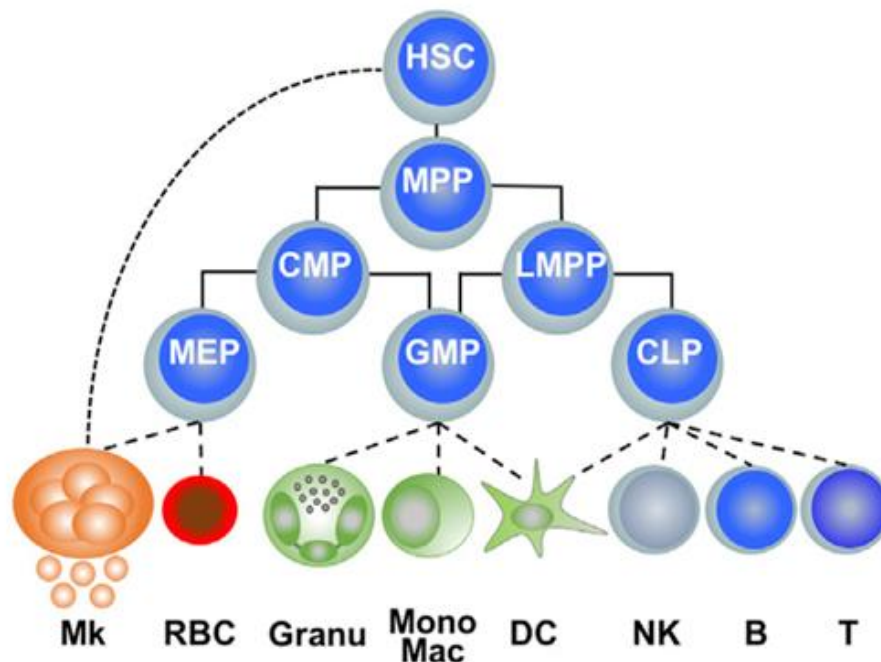
### **1.1.3 The classical model of haematopoiesis**

At the top of the hierarchy of haematopoiesis are self-renewing, multipotent HSCs (Figure 1.1) which give rise to all mature blood cell types via ordered fate restriction events (Orkin, 2000). The self-renewal capacity is lost when the cells enter the multipotent progenitor (MPP) compartment and the cells obtain greater cell proliferation ability (Figure 1.1). These MPPs retain potency to differentiate into any mature lineage despite being already developmentally committed and losing the ability to re-enter the HSC compartment (Morrison et al., 1997).

Following the MPP stage the cells become committed to either lymphoid or myeloid/erythroid lineages, which are termed the common lymphoid progenitor (CLP) and common myeloid progenitor (CMP) compartments, respectively. The CMPs retain potency to produce either myeloid or erythroid cells following partitioning into one of two more specified compartments: the megakaryocyte/erythroid progenitors (MEPs) and granulocyte/monocyte progenitors (GMPs). Eventually, all mature myeloid lineages emerge from GMPs (Figure 1.1) (Akashi et al., 2000). On the other hand, CLPs generate all the lymphoid cells, including T-cells, B-cells, natural killer (NK) cells and others, comprising the adaptive immune system. Lymphoid development happens in the bone marrow, where immature lymphocytes emerge which enter the circulation to mature at distal sites, such as the thymus (T-cells) and germinal centres (B-cells) (Ciofani and Zuniga-Pflucker, 2007, Zhang et al., 2016).

Current research efforts have reconsidered key aspects of the hierarchical haematopoiesis model. An extra compartment of progenitors was identified, featuring the expression of lymphoid marker genes and FMS-like tyrosine kinase (Flt3). These cells were named lymphoid-primed multipotent progenitors (LMPPs) due to their ability to produce both lymphoid and myeloid cells, but not erythroid lineages (Figure 1.1) (Adolfsson et al., 2001). It was also suggested that strict lymphoid or myeloid branching is not the earliest lineage restriction step because some MEPs skipped the intermediate CMP state and were derived directly from MPPs (Figure 1.1) (Adolfsson et al., 2005).





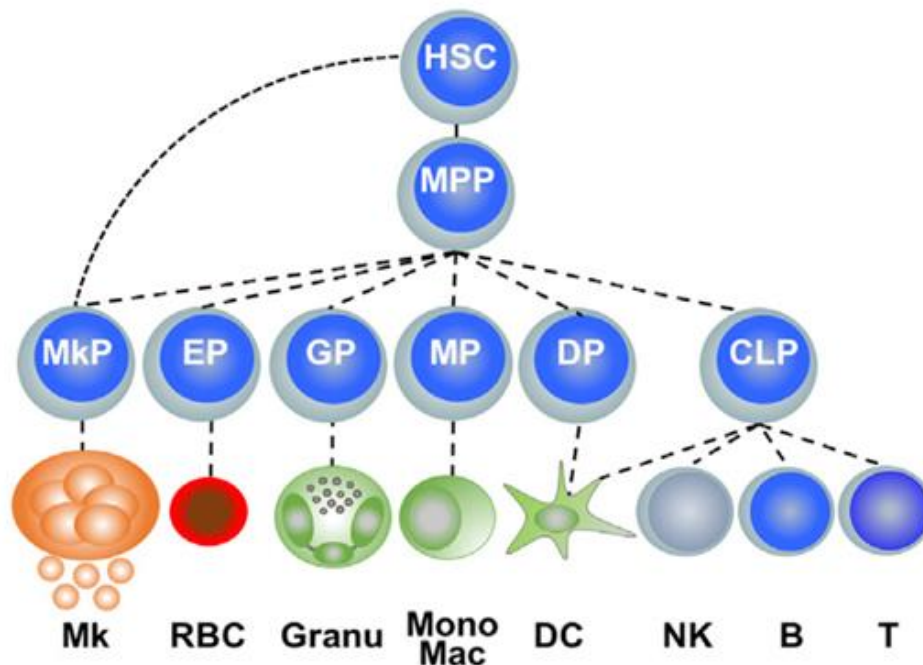
**Figure 1.1 The classical model of haematopoiesis.**

HSCs undergo defined lineage commitment in a stepwise transition into intermediate progenitor stages. The lineage decisions occur at the MPP stage, where cells commit to either the myeloid/erythroid or lymphoid lineages, through CMP or CLP precursor stages. HSCs may skip the MPP stage and directly commit to erythroid/megakaryocyte lineages. Adapted from (Haas et al., 2018).

#### 1.1.4 The alternative models of haematopoiesis

Chromatin profiling and gene expression methods of single cells prompted further revision of the classical model of haematopoiesis. Recent studies suggest that lineage commitment starts at a much earlier stage than previously described, *i.e.* in the phenotypic HSC (CD150<sup>+</sup>CD34<sup>+</sup>Flt3<sup>+</sup>KSL) compartment. Traceable barcoded cells from this population demonstrated the ability to produce unipotent progenitors without need for an oligopotent intermediate (Yamamoto et al., 2013, Naik et al., 2013)

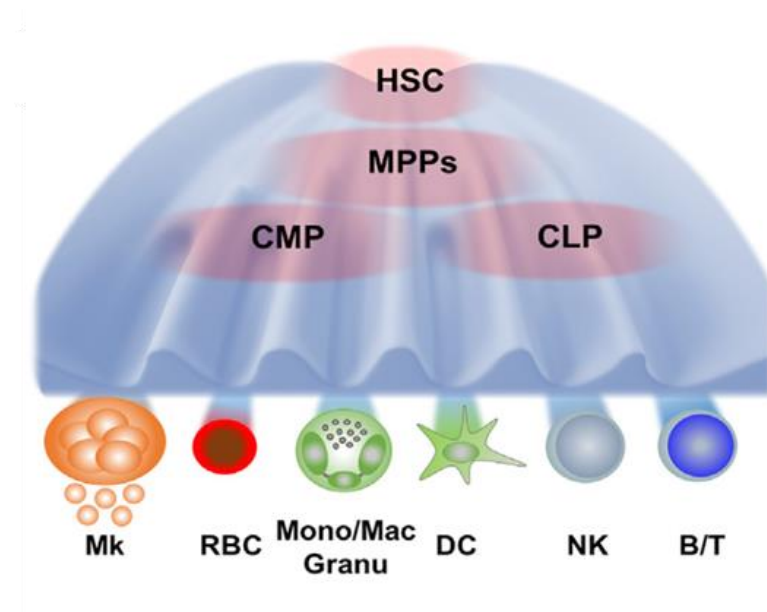
consistent with previous reports regarding the origin of MEPs (Figure 1.2) (Adolfsson et al., 2005).



**Figure 1.2 The alternative model of haematopoiesis.**

The lineage restriction earlier than the previous models. The MPP directly produce unipotent progenitors without transitioning into intermediate oligopotent progenitors. Adapted from (Haas et al., 2018).

A more recent investigation suggested that previously defined compartments such as MPPs and LMPPs are not discrete cell types. Instead, uni-lineage restricted cells emerge directly from a 'continuum of low-primed undifferentiated haematopoietic stem and progenitor cells' (CLOUD-HSPCs). These findings also suggested that in adult haematopoiesis, oligopotent lineage priming is rare, with most of the cells being either multipotent or committed to a single lineage (Figure 1.3) (Velten et al., 2017).



**Figure 1.3 The CLOUD-HSPCs model of haematopoiesis.**

HSCs have some degree of predetermined lineage bias, where they do not undergo discrete intermediate stages and uni-lineage cells are produced directly from HSPC continuum. Adapted from (Haas et al., 2018)

## 1.2 T-cell development

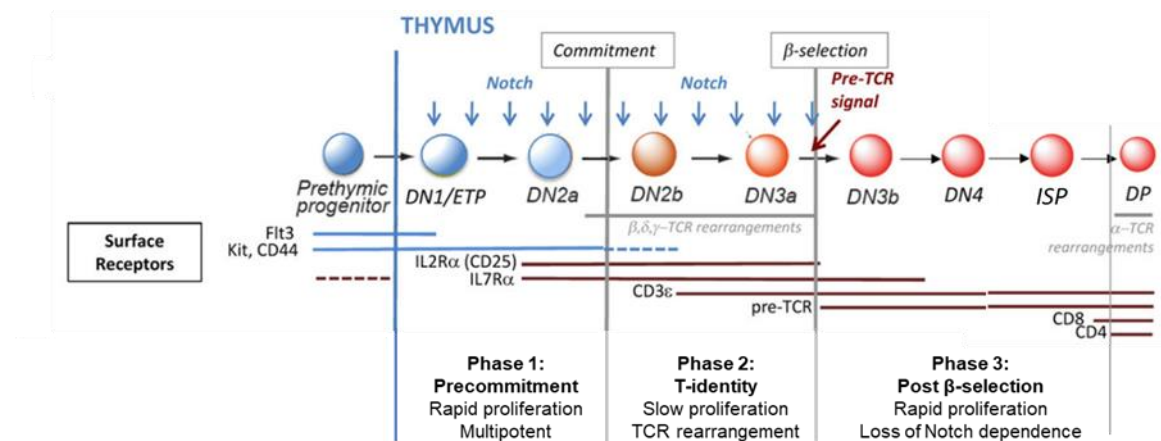
Haematopoietic progenitors with T-cell potential originate in the bone marrow then migrate to the thymus where they go through multiple rounds of expansion and differentiation (Aifantis et al., 2008, Yui and Rothenberg, 2014). The thymus is the ultimate site for T-lineage commitment and T-cell development, where strong environmental signals guide the cells into the T-cell pathway. The Pre-thymic progenitors, such as CLPs, and the LMPPs can enter the thymus and give rise to T-cells through a common pathway (Saran et al., 2010, Serwold et al., 2009). When the lymphoid progenitors arrive in the thymus, they lack expression of T-cell receptors (TCRs) and go through phenotypically distinct stages by expression of specific surface markers and different versions of TCR (Germain, 2002). The first stages of development are termed double negative (DN) to signify that they do not express the

two T-cell coreceptors, CD4 and CD8. Subsequently, these stages are subdivided in mice according to the expression of other surface markers, the growth factor receptors Flt3 and Kit, the adhesion molecule CD44, and the cytokines IL2R $\alpha$  (CD25) and IL7R $\alpha$  (Rothenberg et al., 2016, Staal et al., 2001).

T-cell development stages in the mouse thymus start by early T-cell progenitors (ETPs or DN1), followed by DN2a, DN2b, DN3a, DN3b, transitional DN4, immature single-positive (ISP), and finally CD4/CD8 double positive (DP) (Figure 1.4) (Rothenberg et al., 2016). The somatic assembly of TCR genes (*Tcra*, *Tcrb*, *Tcr $\gamma$* , *Tcr $\delta$* ) is crucial for T-cell development and instructs cell fate decisions. The four TCR genes are assembled according to developmental programs directed by the lymphoid-specific recombinase complex (RAG, composed of RAG1 and RAG2) (Bassing et al., 2002). The RAG1/RAG2 complexes are critical to catalyse the recombination of TCR variable (V), diversity (D), and joining (J) segments to assemble *TCR* coding sequences (Chen et al., 2001). RAG enzymes create double-strand breaks at recombination signal sequences that flank TCR V(D)J gene segments followed by nonhomologous end joining which promotes the assembly of  $\alpha\beta$  or  $\gamma\delta$  TCR. The outcome of these TCR gene recombination events instructs the thymocytes commitment to the either  $\alpha\beta$  or  $\gamma\delta$  lineages (Hayday and Pennington, 2007).

The early T-cells proliferate and differentiate in a TCR-independent, Notch-dependent manner. When TCR gene rearrangement successfully occurs, cell survival and differentiation become TCR-dependent, and the influence of Notch is decreased (Rothenberg et al., 2016). Yui and Rothenberg divided the Notch-dependent period of T-cell development into two phases (phase 1: pre-commitment and phase 2: T-identity) (Figure 1.4) (Yui and Rothenberg, 2014). Pre-committed early T-cell progenitors in the

ETP/DN1 stage maintain Kit and CD44 expression but start to downregulate Flt3 under the influence of Notch pathway signalling. After multiple rounds of cell division, IL2R $\alpha$  (CD25) and IL7R $\alpha$  expression is activated by the Notch signals and the entry to the DN2a stage is initiated (Figure 1.4) (Yui and Rothenberg, 2014, Rothenberg et al., 2016). During Notch-dependent DN2a to DN2b transition, T-lineage commitment occurs and marks the final loss of factors and gene regulatory networks that allow alternative lineage fate choices. This transition causes both the loss of progenitor-specific genes including, *CD44* and *Kit* and substantial upregulation of many T-cell specific genes (Figure 1.4) (Radtke et al., 2013).



**Figure 1.4 Overview of T-cell development in mice.**

Cells begin differentiation when entering the thymic environment, which provides Notch signals (blue arrows). Thymocytes start as Early T-cell progenitors (ETPs), and transit sequentially through the double negative (DN) stages and then via the immature single-positive (ISP) stage to the double positive (DP) stage, which give rise to the functional CD4<sup>+</sup> and CD8<sup>+</sup> single positive T-cells. The presence of cell surface receptors is depicted in lines, where blue lines represent receptors expressed on cells seeding the thymus and red lines represent receptors whose expression is gained within the thymus. Dotted lines indicate lower expression levels. The stages during which TCR rearrangements occur are marked with grey lines. Development is categorised by the commitment and  $\beta$ -selection checkpoints into three regulatory phases, *i.e.* Phase1: Precommitment, Phase 2: T-identity, Phase 3: post  $\beta$ -selection. Adapted from (Yui and Rothenberg, 2014).

Committed T-cells in phase 2, slow their proliferation to the point of G1 arrest at the DN3a stage (Yui and Rothenberg, 2014). At this phase most T-cell identity genes are turned on and RAG enzymes are activated and recombination of *Tcrd*, *Tcrg* and *Tcrb* occurs (Figure 1.4) (Hayday and Pennington, 2007, Mingueneau et al., 2013). The successful recombination of *Tcrd* and *Tcrg* promotes assembly of a  $\gamma\delta$  TCR, while successful recombination of *Tcrb* promotes assembly of TCR $\beta$  (Krangel, 2009). If TCR $\gamma\delta$  is successfully rearranged, the cells diverge from the  $\alpha\beta$  T-cell development path. However, the unsuccessful expression of either TCR $\beta$  or TCR $\gamma\delta$  chains leads to cellular apoptosis (Hoffman et al., 1996, Taghon et al., 2006). Most of T-cells express TCR $\alpha\beta$ , while approximately 4% of T-cells express TCR $\gamma\delta$  (Gaulard et al., 1990). Cells with successfully rearranged TCR $\beta$  chain and non-rearranging pre-T $\alpha$  chain start to express the pre-TCR (Figure 1.4). During  $\beta$ -selection, the TCR $\beta$  protein assembles at the cell membrane with pre-TCR complex members that are already expressed, including pre-TCR $\alpha$ , CD3 $\gamma$ , CD3 $\delta$ , CD3 $\epsilon$  and TCR $\zeta$  (Krangel, 2009, Hoffman et al., 1996, Taghon et al., 2006).

Subsequently, the expression of a functional TCR allows growth and differentiation when the cells transition from DN3a to DN3b under the pre-TCR signalling which activates the phase 3 gene network (Figure 1.4) (Germain, 2002, Yui and Rothenberg, 2014). Although the cells require Notch signalling during the passage through  $\beta$ -selection checkpoint, they become Notch-independent at phase 3 and Notch target genes are turned off (Taghon et al., 2009, Kreslavsky et al., 2012, Yui and Rothenberg, 2014). Post the  $\beta$ -selection checkpoint the cells enter DN3b and then DN4 stages, where they rapidly proliferate and undergo many changes (Figure 1.4). Such changes include, downregulation of RAG expression, cell enlargement, progression into rapid

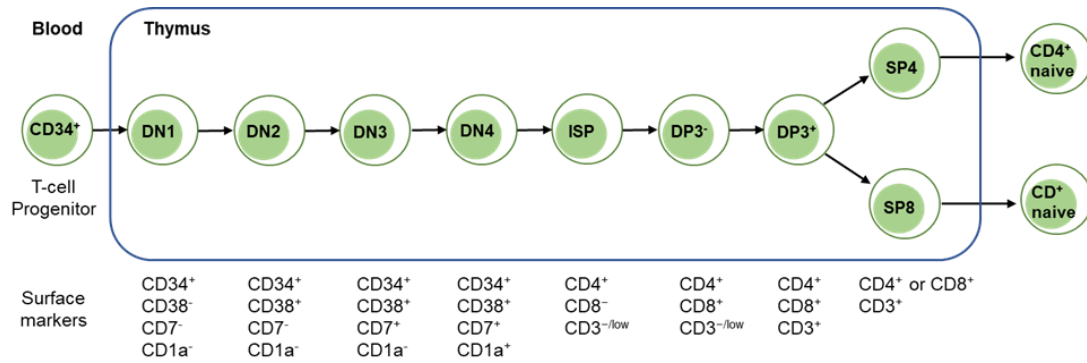
cell-cycling, termination of IL2R $\alpha$  expression, and increased surface expression of CD27 and CD28 (Teague et al., 2010, Yui and Rothenberg, 2014, Krangel, 2009). DN4 cells proceed through a CD4<sup>-</sup> CD8<sup>+</sup> TCR $\beta$ <sup>-</sup> ISP stage as they proliferate on their way to transform into a CD4<sup>+</sup> CD8<sup>+</sup> DP phenotype (Figure 1.4). The resulting DP cells acquire a major change in cellular transcriptional and regulatory state, and they finally stop proliferating and accumulate in a G0 state to rearrange their TCR $\alpha$  genes (Figure 1.4) when *Rag* genes are re-expressed (Germain, 2002, Rothenberg et al., 2016, Krangel, 2009). DP cells live for a few days to assemble a TCR $\alpha\beta$  complex that can interact with major histocompatibility complex (MHC) molecules on the surrounding stroma. This interaction is termed as positive selection. T-cells that express TCRs that bind MHC class I complexes become single positive (SP) CD8<sup>+</sup> T-cells, whereas those that express TCRs that bind MHC class II ligands become single positive CD4<sup>+</sup> T-cells (Teh et al., 1988, Swain, 1983). These cells mature in the thymic medulla and go through further negative selection against strongly self-reactive cells, then finally become ready for export from the medulla to peripheral lymphoid sites (Germain, 2002).

In human, equivalent stages of T-cell development are made according to the expression of different surface markers such as, CD34, CD38, CD7, and CD1a (Figure 1.5). For example, CD34<sup>+</sup> precursors seed the human thymus which reflect Flt<sup>+</sup> pre-thymic progenitors in mice (Vicente et al., 2010). CD34<sup>+</sup> cells can differentiate into multiple lineages, giving rise to T, B and NK cells of the lymphoid compartment, and to myeloid cells. Immature thymocytes that express CD34 but lack CD38 and CD1a expression resemble the mouse ETP/D1 population (Figure 1.5) (Dik et al., 2005). The next stage of differentiation is marked by the expression of both CD34 and CD38

resembling the mouse DN2 stage which reflects T-cell commitment. Thymocytes differentiation proceed by CD7 expression in DN3 stage, followed by the final DN4 stage, which is characterised by the expression of CD34, CD38, CD7 as well as CD1a (Figure 1.5) (Blom and Spits, 2006, Spits, 2002, van der Weerd et al., 2013).

As in mouse, the first round of *TCR* genes rearrangement occurs at the DN2 and DN3 stages. While  $TCR\gamma\delta$  cells take a different developmental path than  $TCR\alpha\beta$ , the  $TCR\beta$  cells go through  $\beta$ -selection. These cells express  $TCR\beta$  chain on their surface coupled with pre- $T\alpha$  receptor in the pre-TCR complex (Staal et al., 2001, Staal et al., 2016). After  $\beta$ -selection, the cells express CD4 with low levels of CD3, in the absence CD8, and they are referred to as  $CD4^+$  (ISP) cells. Subsequently, they start to express CD3 in addition to both CD4 and CD8 and they are referred to as DP cells (Figure 1.5). During the DP stages,  $TCR\alpha$  gene rearrangements are initiated, followed by expression of  $TCR\alpha\beta$  on the cell surface (Ramiro et al., 1996, Blom et al., 1999, Spits, 2002). The  $TCR\alpha\beta$  expressing cells are then tested for the interaction with MHC molecules (positive selection), and the absence of self-reactivity (negative selection) (Conroy and Alexander, 1996, Punt et al., 1994). Eventually, the cells that pass positive and negative selection commit to either the  $CD4^+$  T-helper lineage or the  $CD8^+$  cytotoxic T-lineage and travel to the periphery as SP cells (Figure 1.5) (Blom and Spits, 2006).





**Figure 1.5 Model of T-cell development in human.**

Illustrated are the different phenotypic stages of T-cell development as commonly discerned in the human thymus, with the corresponding surface marker expression. CD34<sup>+</sup> T-cell progenitors enter the thymus where they go through sequential differentiate stages marked as double negative (DN), immature single-positive (ISP), double positive (DP) stage, and single positive (SP). T-cells leave the thymus into the periphery either as CD4<sup>+</sup> T-helper cells or CD8<sup>+</sup> cytotoxic T-cells. Adapted from (Staal et al., 2016).

### 1.2.1 T-cell identity genes

T-lineage programming includes tight regulation of T-cell identity genes, as well as the transcription factors (TFs) that regulate the expression of these genes. Initially, the cells experience shifts in gene expression of growth factors and cytokines. For example, the growth factor receptor *KIT* is important in the earliest stages of thymocytes and is gradually repressed as they differentiate (Tabrizifard et al., 2004, Massa et al., 2006). By contrast, the cytokine *IL7Rα* is essential for intermediate stages and starts being expressed at the DN2 and DN3 stages (Kang and Der, 2004). At these stages many of precursor cell genes are downregulated and genes that define the T-cell lineage are upregulated. T-cell Identity genes encode proteins that are essential for TCR rearrangement, e.g. RAG1 and RAG2, TCR complex assembly, i.e. CD3 chains, and signalling components, including kinases, phosphatases and adaptor proteins, such as LCK, ZAP70, and LAT (Igarashi et al., 2002, Gounari et al., 2002).

Most of the T-cell specification genes, such as *Cd3g*, *Cd3e*, *Zap70* and *Il7ra* begin to be expressed at the DN2 stage (Taghon et al., 2006, Masuda et al., 2007, Rothenberg et al., 2008). T-cell identity is set by the DN3 stage where they must express *RAG1* and *RAG2* (Chen et al., 2001). Some genes that are sharply upregulated between multipotent progenitors and Pro-T cell stages are not differentiation genes, but those encoding transcription factors (Tydell et al., 2007). Thus, T-cell differentiation program includes precisely timed silencing and transient up- and down-regulation activities as well as the steady increase in T-cell identity gene expression controlled by transcriptional regulatory networks (Rothenberg et al., 2016).

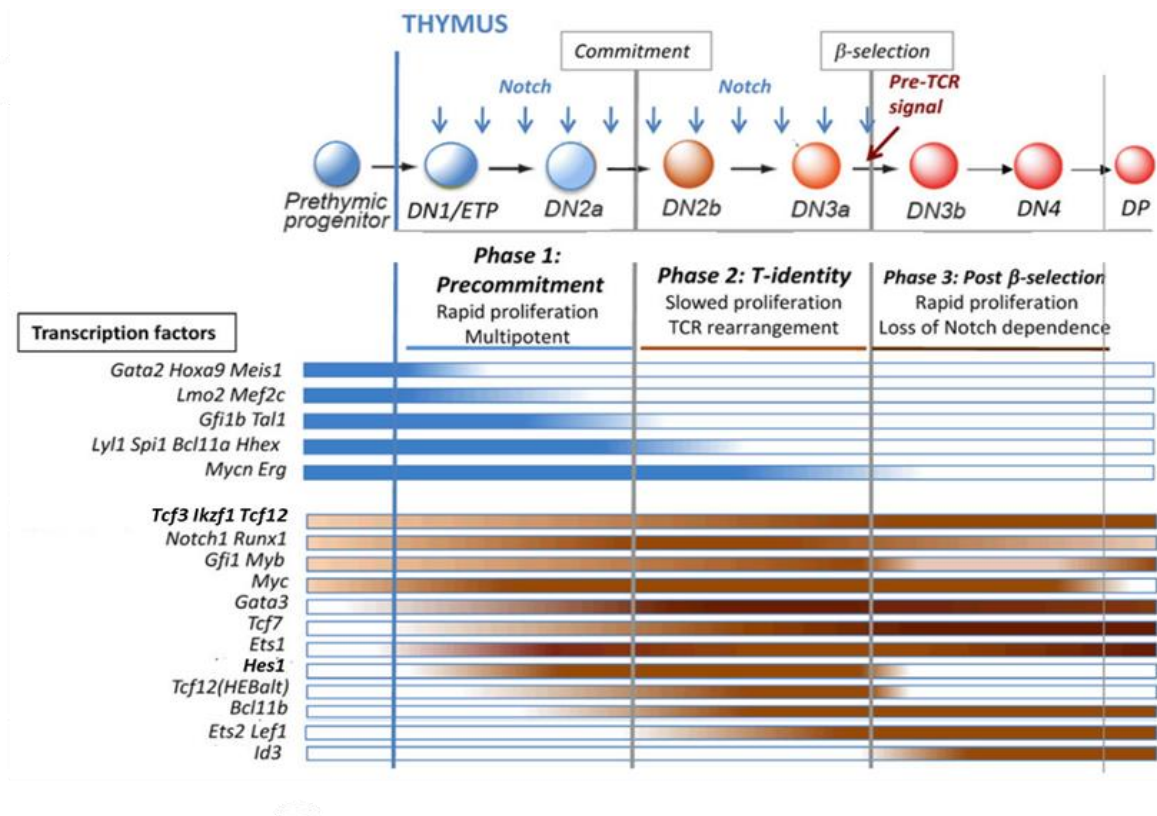
### **1.3 Transcription factors role in T-cell development**

#### **Notch signalling**

Notch signalling is crucial to initiate T-cell development, and the right amount of signal strength ensures the sequential activation of T-lineage genes in order for the cells to gradually go through the differentiation stages in the thymus (Weerkamp et al., 2006). T-cells start their developmental process when the Notch pathway is stimulated through the interaction of NOTCH1 transmembrane molecules on the lymphoid precursors with Delta-like 4 ligands on thymic epithelial cells (Koch et al., 2008). During the initial stages of T-cell development Notch signalling interacts with the stem and progenitor-cell gene regulatory network that is passed on from multipotent precursors. Some of these genes are progenitor-specific genes with roles confined to the earliest stages of T-cell development and others play ongoing roles in T-cells. Nonetheless, they all participate in the network of pre-commitment stage (ETP/DN1-DN2a) that was termed “Phase 1” by the Rothenberg group (Figure 1.6) (Yui and Rothenberg, 2014).

### 1.3.1 Phase 1

From the thymic entry the cells either continue or silence the expression of a set of TFs used in stem and progenitor cells. A number of TF genes that are instrumental to regulatory networks in Flt3<sup>+</sup> pre-thymic progenitors continue to be expressed only in ETP cells. For example, genes including *Gata2*, *Hoxa9*, *Hlf*, and *Meis1* are turned off before the cells proceed to the DN2 stage. Progenitor cell regulatory factors genes, including *Lmo2* and *Mef2c* are silenced after the DN2a stage (Figure 1.6) (Gwin et al., 2013, Huang et al., 2012, Yui and Rothenberg, 2014). In comparison, PU.1 (*Spi1*), *Hhex*, *Gfi1b*, *Lyl1*, *Tal1*, *Bcl11a*, and N-Myc (*Mycn*) persist longer, *i.e.* into DN2a and DN2b stages (Figure 1.6) (Yui and Rothenberg, 2014). The final silencing of these factor genes occurs during  $\beta$ -selection and failure to silence them can lead to T-cell malignancies (Yui and Rothenberg, 2014, Rothenberg et al., 2016). Other progenitor cell factor genes such as, *Myb*, *Gfi1*, *Runx1*, *Ikzf1* (IKAROS) and *Tcf3* (E2A) continue to be expressed or even upregulated during T-cell commitment stages (Figure 1.6) (Heng et al., 2008, Mingueneau et al., 2013, Rothenberg et al., 2016). Additionally, T-cell specific genes such as *Gata3*, *Tcf7*, and *Bcl11b* are upregulated through Notch signalling. These genes encode TFs that play essential roles in most subsets of T-cells (Figure 1.6) (Yui and Rothenberg, 2014).



**Figure 1.6 Stage specific patterns of gene expression for important transcriptional regulators during T-cell development.**

Blue bars represent genes of proteins involved in transcriptional regulation in stem and progenitor cells. The red bars represent the expression of T-cell specification and commitment factors. The colour intensity shows the dynamic changes in expression of the genes (dark=high, light=low). Adapted from (Yui and Rothenberg, 2014).

## TCF-1

TCF-1 transcription factor, encoded by the *Tcf7* gene, is directly activated by NOTCH from the early stages of T-cell development (Figure 1.6). Once *Tcf7* gets activated it is no longer primarily affected by NOTCH inhibition (Germar et al., 2011, Del Real and Rothenberg, 2013). In ETPs, TCF-1 activates genes that are essential for cell survival and expansion. TCF-1 collaborates with NOTCH to positively regulate T-cell specification genes in a feedforward network circuit (Germar et al., 2011). High levels

of TCF-1 can drive the activation of T-cell genes, including *Gata3*, *Bcl11b*, *Il2ra* (CD25), *Cd3g*, *Lat*, *Lck* and endogenous *Tcf7*, even in the absence of Notch signals (Weber et al., 2011). *Tcf7* can be expressed at later stages of T-cell development (phase 3 and beyond), in a Notch independent manner and in a way correlated to GATA3 levels (Yui and Rothenberg, 2014).

### **GATA3**

Like TCF-1, the expression of the GATA3 transcription factor begins within the ETP population (Figure 1.6) (Germar et al., 2011, Weber et al., 2011, Wakabayashi et al., 2003). In T-cells, GATA3 is activated by the progenitor cell factor MYB, as well as by TCF-1 (Del Real and Rothenberg, 2013, Gimferrer et al., 2011, Maurice et al., 2007). GATA3 is essential for T-cell development from an early stage, and insufficient doses of GATA3 can be lethal to pro-T cells at ETP, DN3, and advanced stages (Rothenberg et al., 2016). Lack of GATA3 in DN2 cells leads to downregulation of *Kit*, *Ets1*, *Bcl11b*, *Itk*, *Tcf7*, *Zfp101*, *Cd3e*, and *Rag1* and upregulation of *SpiB*, *Bcl11a*, *Dtx1*, *Spi1*, and possibly also *Lmo2* and *Nfil3*. Some of these abnormal gene expressions alter the normal developmental sequence in T-cells and may lead to leukemogenesis. The deficiency of GATA3 in later developmental stages causes failure of  $\beta$ -selection as well as failure to generate CD4 cells (Hernandez-Hoyos et al., 2003, Pai et al., 2003).

### **BCL11B**

TCF-1, GATA3, NOTCH, and RUNX/CBF $\beta$  are positive regulators for *Bcl11b* expression (Weber et al., 2011, Li et al., 2013a, Garcia-Ojeda et al., 2013). The BCL11B transcription factor is not expressed until the cells are in the mid-DN2a stage and the timing of *Bcl11b* activation is crucial for lineage commitment (Figure 1.6)

(Rothenberg et al., 2008, Tydell et al., 2007, Li et al., 2010b). BCL11B functions as a repressor during T-cell development and there is little evidence for it to act as a positive regulator of T-lineage genes (Rothenberg et al., 2010, Yui and Rothenberg, 2014). BCL11B deficient thymocytes fail to repress genes that are normally downregulated by the DN2 stage including *Tal1*, *Bcl11a*, *Spi1*, *Erg*, and *Flt3*. However, T-lineage specification genes such as *Tcf7*, *Gata3*, *CD3g*, *Ptcra*, and *Rag1* remain expressed in the absence of BCL11B (Li et al., 2010b). Thus, BCL11B is required to shift the cells from a program of progenitor-cell gene expression, to a commitment program of limited growth, progenitor-cell gene silencing, and continued differentiation to the DN3a stage (Ikawa et al., 2010, Li et al., 2010b). The expression of BCL11B together with GATA3 and TCF-1 prepares the cells for the transition to phase 2 (Yui and Rothenberg, 2014).

### **1.3.2 Phase 2**

Phase 2 refers to T-cell commitment stages (DN2b, DN3a), where the cells slow their proliferation and undergo TCR rearrangement. The expression of GATA3 and TCF-1 in the previous phase allow the cells to exert T-lineage promoting functions, because knockout of either gene leads to reduced survival and differentiation of ETPs as well as their descendants (Yui and Rothenberg, 2014, Hosoya et al., 2009, Germar et al., 2011). The role of BCL11B in silencing many of the stem and progenitor-cell genes contributes to the T-cell commitment transition from DN2a to DN2b, terminating the ability to differentiate into alternative lineages under normal conditions (Li et al., 2010b, Rothenberg et al., 2016). The subsequent committed T-cell entry into the DN3 stage depends on the RUNX/CBF $\beta$  complexes (Naito et al., 2011). Thymocytes that lacked RUNX1 experienced a developmental block at the DN2 stages (Growney et al., 2005). In addition to the RUNX/CBF $\beta$  complex, transcription factors including MYB, GATA3,

and BCL11B are needed for full TCR $\beta$  gene rearrangement and competence. Many of these factors occupy regulatory elements of *Tcrb* and *Rag1/2* genes through which they directly regulate the expression of these genes (Anderson, 2006, Rothenberg et al., 2008).

Other transcription factors are also required during T-cell commitment, such as the ETS family TFs. The expression of ETS1 starts from the early DN1 stage, whereas ETS2 is later expressed at the DN2 stage (Figure 1.6). Both are upregulated from DN2 to DN3 when another ETS factor, PU.1, is suppressed, as PU.1 specifically inhibits their expression (David-Fung et al., 2009). On one hand, the ETS family member (ETV5) is upregulated in the TCR $\gamma\delta$  lineage cells and plays an important role in the activation and maturation of  $\gamma\delta$  T effector cell subset (Jojic et al., 2013). On the other hand, expression of E-proteins is required in TCR $\beta$  thymocytes. E-proteins are particularly important to maintain the integrity of the  $\beta$ -checkpoint, and also involved in the regulation of expansion of thymocytes that have passed  $\beta$ -selection (Engel and Murre, 2004, Wojciechowski et al., 2007).

### **E-Proteins**

The basic helix loop helix (bHLH) transcription factors E2A and HEB are the main components of the “E-protein” heterodimers in early T progenitors. Both are expressed throughout T-cell development stages and E2A (*Tcf3*) continues to be expressed at constant levels from pre-thymic precursors, whereas HEB (*Tcf12*) is strongly upregulated as the cells differentiate (Figure 1.6) (Rothenberg et al., 2016). The E2A-HEB dimer regulates essential lineage- and stage-specific target genes. The complex can bind regulatory elements of genes important for lymphocyte development, including the TCR $\alpha$  and  $\beta$  enhancers, the Pre-T $\alpha$  promoter, and various CD4

regulatory elements (Greenbaum and Zhuang, 2002, Sawada and Littman, 1993, Takeuchi et al., 2001). The simultaneous deletion of E2A and HEB in developing thymocytes led to a severe developmental block in DN to DP transition stages and a dramatic reduction of Pre-T $\alpha$  expression. These T-cell progenitors showed increased proliferation *in vivo* and remarkable expansion *ex vivo*. These results implicated that E2A and HEB may play a role in maintaining thymocytes in cell cycle arrest before pre-TCR expression in addition to their roles in T-cell differentiation (Wojciechowski et al., 2007).

There are two major isoforms of HEB that are expressed in T-cells. The first canonical isoform “HEBcan” is transcribed from the conventional transcriptional start site in (*Tcf12*) exon 1. The second, alternative, isoform “HEBalt” is formed by an internal promoter, under the control of NOTCH, that encodes a truncated protein with an alternative N-terminal sequence. HEBcan is expressed in early T-cell progenitors and peaks at the DP stage of thymocyte development, while HEBalt is induced in ETPs and downregulated after the DN3a during  $\beta$ -selection (Figure 1.6). HEBalt may be more supportive of cell proliferation than canonical HEB or E2A (Wang et al., 2006, Braunstein and Anderson, 2012).

E-proteins play a major role in regulating gene expression in DN3a cells. They collaborate with NOTCH to regulate *Hes1*, *Notch3*, and *Ptcra* genes (Ikawa et al., 2006). They induce the expression of *Rag* genes in cooperation with GATA3 and upregulate the *Cd3* gene cluster expression with the aid of GATA3 and TCF-1 (Zhang et al., 2012b, Xu et al., 2013, Oosterwegel et al., 1991). E-proteins also promote and sustain *Notch1* expression at the DN3 stage and Notch signalling can be antagonised by E-protein antagonist ID2 (Del Real and Rothenberg, 2013, Yashiro-Ohtani et al.,



2009). Therefore, BCL11B may support the DN3a state by suppressing *Id2* (Li et al., 2010b).

### **1.3.3 Phase 3**

After T-cells pass the  $\beta$ -selection they undergo rapid proliferation as they transition through DN3b to DN4 stages. Thereafter, the thymocytes no longer depend on Notch signalling and rely on pre-TCR signals which allow IKAROS transcription factor to suppress NOTCH-activated genes (Geimer Le Lay et al., 2014, Yui and Rothenberg, 2014). The expression of *Id3* is transiently turned on upon pre-TCR signalling which temporarily inhibits E-protein activity, however E-proteins regain their function afterwards (Figure 1.6) (Engel et al., 2001). RUNX1 is also known to be required to regulate the transitions of developing thymocytes from the DN to DP stage and from the DP stage to the mature SP stage (Egawa et al., 2007). The expression of *Ets1*, *Ets2*, *lef1*, *Tcf7*, and *Tcf12* is upregulated at the DP stage (Figure 1.6), whereas the histone tail modification (H3K27me3) repressive marks accumulate at the promoters of pre-thymic cell genes and Notch targets (Yui and Rothenberg, 2014, Zhang et al., 2012b, Vigano et al., 2014). The tight regulation of regulatory networks during  $\beta$ -selection and post  $\beta$ -selection, as well as the silencing of Notch target genes are crucial to prevent malignant transformations of T-cells (Yui and Rothenberg, 2014).

## **1.4 T-cell Acute Lymphoblastic Leukaemia**

In normal haematopoiesis, gene expression is determined by the effect of chromatin structure, gene regulatory networks, and ultimately binding of TFs (Gottgens, 2015). Dysregulation of these networks by mutations or oncogenic fusions of TF genes can lead to malignant transformations in haematopoietic cells. Disturbances in the

developmental stages of blood cells by chromosomal aberrations, epigenetic changes and sequential mutations can cause a block in differentiation of precursor cells and in uncontrolled proliferation (Sive and Gottgens, 2014). The malignant progression of proliferating blood cells is called leukaemia. Different types of leukaemia are characterised according to the lineage of origin; myeloid or lymphoid. Acute myeloid leukaemia (AML) is a rapid, uncontrolled accumulation of abnormal myeloblast cells (Dohner et al., 2010). Acute lymphoblastic leukaemia (ALL) is continuous proliferation of immature lymphoid progenitor. The ALL is subdivided into B-cell (B-ALL) and T-cell (T-ALL) (Inaba et al., 2013).

T-cell acute lymphoblastic leukaemia (T-ALL) is an aggressive haematological malignancy characterised by accumulation of genomic lesions that affect the development of thymocytes. T-ALL represents about 25% of acute lymphoblastic leukaemia (ALL) cases (Hunger and Mullighan, 2015, Aifantis et al., 2008). Treatment of T-ALL has advanced over the years; however, it still remains a challenge due to the aggressive presentation of this type of leukaemia (Pieters and Carroll, 2008, Pui and Evans, 2006, Litzow and Ferrando, 2015). Therefore, it is important to understand the pathogenesis and biology of T-ALL to create better and more targeted therapies. The pathogenesis of T-ALL involves numerous genetic lesions that drive normal T-cells into uncontrolled proliferation and expansion. Van Vlierberghe *et al.* suggested a classification of these genetic aberrations in paediatric T-ALL into type A and type B abnormalities (Van Vlierberghe et al., 2008).

Type A abnormalities may delineate specific molecular-cytogenetic T-ALL subgroups (Van Vlierberghe et al., 2008). This type acquires large chromosomal deletions or amplifications with maturational arrest at a particular stage of thymocyte differentiation

(van Grotel et al., 2006). In type A, the genetic aberrations prime abnormal expression of genes in certain developmental stages of T-cells, thus they function as oncogenes. Normally developing T-cells exhibit rearrangements of the T-cell receptor (TCR) in order to acquire a wide diversity of antigen recognition. The TCR enhancers are located in the TCR $\beta$  (7q34) or TCR $\alpha$  – TCR $\delta$  (14q11) loci. Chromosomal rearrangements at these regions involving, breakage and re-ligation of the double stranded DNA, can trigger the aberrant expression of leukemogenic TFs in T-cell progenitors. These TFs are involved in T-cell differentiation and cause deregulation of haematopoiesis when abnormally expressed (Van Vlierberghe and Ferrando, 2012, Van Vlierberghe et al., 2008). Among the aberrantly expressed genes in type A are *TLX1*, *TLX3*, *HOXA*, *TAL1*, *LMO1*, and *LMO2* (Table 1.1) (Soulier et al., 2005, Ferrando et al., 2002). Although these oncogenes function predominantly by blocking differentiation, the exact mechanism by which they contribute to leukemogenesis remains largely unknown.

**Table 1.1 List of chromosomal rearrangements involving the T-cell receptor gene.**

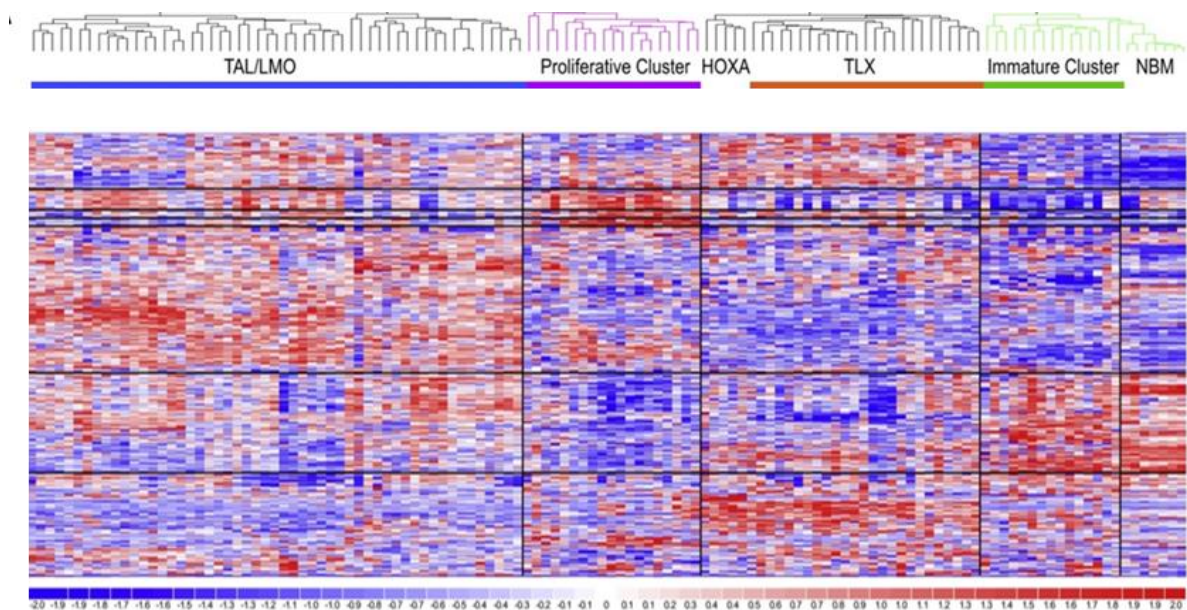
Adapted from (Chiaretti and Foa, 2009).

Translocation	Involved gene	Fusion gene function	Frequency
t(7;10)(q34;q24) and t(10;14)(q24;q11)	<i>TLX1 (HOX11)</i>	Transcription factor	7% Children 31% Adults
t(5;14)(q35;q32)	<i>TLX3 (HOX11L2)</i>	Transcription factor	20% Children 13% Adults
inv(7)(p15q34), t(7;7)	<i>HOXA genes</i>	Transcription factor	5%
t(1;14)(p32;q11) and t(1;7)(p32;q34)	<i>TAL1</i>	Transcription factor	3%
t(11;14)(p15;q11)	<i>LMO1</i>	Protein-protein interaction	2%
t(11;14)(p13;q11) and t(7;11)(q35;p13)	<i>LMO2</i>	Protein-protein interaction	3%

Type B abnormalities usually consist of point mutations, small insertions or deletions. These lesions are found in all major T-ALL subgroups and synergise with type A mutations during T-cell pathogenesis. Type B lesions occur in genes involved in diverse cellular processes such as cell cycle control, self-renewal, TCR signalling, T-cell differentiation or tyrosine kinase activation (Van Vlierberghe et al., 2008). These defects can be activating and inactivating mutations of genes such as *CDKN2A/B*, *ETV6*, *NOTCH1* and *RAS* (Zhang et al., 2012a, Weng et al., 2004, von Lintig et al., 2000).

Homminga *et al.* classified T-ALL into four groups based on hierarchical clustering of microarray gene expression data of paediatric T-ALL samples (Figure 1.7). The largest cluster corresponds to elevated expression due to genetic abnormalities of the TAL/LMO genes, whereas, the second largest cluster was typified by high expression

of TLX/HOXA transcription factors (Figure 1.7). The third cluster termed as “proliferative cluster” comprised most TLX1-translocated cases with high expression of CD1. The fourth cluster was named the “immature cluster” (Figure 1.7) and included immunophenotypic immature CD4/CD8 double negative cases, with expression of a number of early progenitor genes (Homminga et al., 2011).



**Figure 1.7 T-ALL genetic subgroups characterised according to cluster analysis of T-ALL patient samples.**

Unsupervised hierarchical cluster analysis based on microarray expression data of 117 diagnostic T-ALL samples and seven normal bone marrow controls. Four groups were identified, as depicted directly below the dendrogram. Adapted from (Homminga et al., 2011)

#### 1.4.1 The TAL/LMO subgroup of T-ALL

The transcription factor T-cell acute lymphoblastic leukaemia 1 (TAL1) belongs to the bHLH family; it is an essential factor for the development of haematopoietic cells (Porcher et al., 1996). The bHLH proteins form a large superfamily of transcriptional

regulators that are found in almost all eukaryotes (Jones, 2004). Members of the bHLH superfamily can bind to DNA at a consensus hexanucleotide sequence known as the E-box. Different families of bHLH proteins recognise different E-box consensus sequence (Jones, 2004). Murre and colleagues first identified the bHLH motif in the binding preference of two murine TFs known as E12 and E47, which are both encoded by the *E2A* gene (Murre et al., 1989). With the subsequent identification of many other bHLH proteins they were classified into six groups according to their tissue distributions, DNA-binding specificities and dimerisation potential (Murre et al., 1994). Another used classification is four major groups (A-D) on the basis of E-box binding, the amino acid patterns in the other parts of the motif, and the presence or absence of additional domains (Atchley and Fitch, 1997).

Class A bHLH proteins, including E2A (E12/E47) and HEB, regulate the expression of various genes that are associated with V(D)J recombination of the TCR genes. Class B bHLH proteins, such as TAL1, TAL2, LYL1, and bHLHB1 function as transcriptional co-factors that heterodimerise with E-proteins, e.g. E2A or HEB (Jones, 2004). The heterodimer can bind DNA at the E-box motifs (CANNTG) and plays a transcriptional regulatory function through a multiprotein complex that includes members of GATA family of transcription factors, LMOs, and LDB1 (Lecuyer and Hoang, 2004, Ellenberger et al., 1994, Ma et al., 1994, Matthews et al., 2013).

Class B bHLH transcription factors are prone to being aberrantly expressed in T-ALL due to chromosomal rearrangement that target TCR enhancers. TAL1 is required for the generation of all hematopoietic lineages in primitive and definitive haematopoiesis. Although, it is necessary for the generation of HSCs, it is not required for their long-term maintenance and once they are established, they can maintain their stemness

characteristics in the absence of TAL1 (Mikkola et al., 2003). Normal thymocyte maturation from DN to DP stages involves suppression of TAL1 specifically before the DN3 stage. At the same time other TFs upregulation is required to ensure proper differentiation of T-cells (Herblot et al., 2000). Thus, TAL1 activation in T-ALL causes a block of differentiation in thymocyte. *TAL1* overexpression in T-ALL often results from chromosomal translocations including t(1;14)(p32;q11) or variant t(1;7)(p32;q35) (Baer, 1993, Aifantis et al., 2008). Furthermore, an interstitial deletion (1p32), placing the protein coding domains of *TAL1* under the control of the promoter region of SCL interrupting locus (*SIL*) forms the *SIL/TAL1* fusion gene. The *SIL* gene is a cell cycle regulator which is expressed throughout T-cell development; therefore *SIL/TAL1* fusion activates *TAL1* expression even when TAL1 should be downregulated (Cave et al., 2004, Janssen et al., 1993).

Studies using transgenic mice demonstrated the oncogenic role of *Tal1* activation in the induction of Leukaemia *in vivo*. Mice with induced TAL1 protein expression in their lymphoid tissue showed lower survival rates than the control mice and died from T-ALL with long latency, indicating a leukaemogenic role of TAL1 expression in T-cells (Condorelli et al., 1996).

When TAL1 activation is combined with another oncogene it leads to an aggressive type of T-ALL. For example, TAL1 collaboration with the LIM-only proteins LMO1 or LMO2 induce aggressive T-cell tumours *in vivo* (Grutz et al., 1998). Transgenic mice that were engineered to overexpress both *Lmo2* and *Tal1* in their T-cells, developed T-cell cancers with accumulation of immature thymocytes, proving the role of TAL1/LMO2 in altering T-cell development and potentiating oncogenesis (Larson et

al., 1996). Furthermore, TAL1 and LMO1 collaboration induced T-cell malignancies in mice, which were similar to T-ALL recognised in human patients (Aplan et al., 1997).

A potential mechanism by which TAL1/LMO2 induce leukaemia is through interfering with the transcriptional activity of E-proteins during T-cell leukemogenesis. The interaction between TAL1 with E-proteins inhibits E-proteins from activating their target genes resulting in impaired T-cell development (Chervinsky et al., 1999). Mice overexpressing *Tal1* showed repression of E47/HEB target genes which led to T-ALL acceleration and perturbed thymocyte development (O'Neil et al., 2004). Similar characteristics in T-cell development have been described for TAL1/LMO transgenic mice and E-proteins null mice. In both cases, thymocytes were blocked before CD4<sup>+</sup>/CD8<sup>+</sup> DP stage of T-cell differentiation (Herblot et al., 2000, Aplan et al., 1997).

## **1.5 LIM domain only 2 (LMO2)**

### **1.5.1 History of LMO2**

The LIM domain only 2 (*LMO2*) gene was discovered more than two decades ago as a recurrent chromosomal translocation partner of TCR loci in patients with T-ALL (Boehm et al., 1991, Royer-Pokora et al., 1991). *LMO2* belongs to the LIM-domain-only proteins family which consists of four genes *LMO1*, 2, 3 and 4 (Foroni et al., 1992, Kenny et al., 1998).

This family name was derived from their tertiary structure that is composed of two tandemly arranged regions called LIM-domains (LIM1 and LIM2). The LIM domain protein motif was named after the initials of three proteins containing such domain; Lin-1, Isl-1, and Mec-3 (Way and Chalfie, 1988). Generally, the structure of LIM domains is composed of 2 zinc fingers (Zheng and Zhao, 2007). The classical consensus



sequence of LIM domains is a cysteine-rich motif, defined as CX<sub>2</sub>CX<sub>16</sub>–23HX<sub>2</sub>CX<sub>2</sub>CX<sub>2</sub>CX<sub>16</sub>–21CX<sub>2</sub> (C/H/D) (where X denotes any amino acid) (Schmeichel and Beckerle, 1994). Although many zinc fingers are known DNA binding structures, there is no evidence that LIM domains bind DNA and are considered to be involved in protein-protein interactions (Sanchez-Garcia and Rabbitts, 1993, Bridwell et al., 2001, Yaden et al., 2005).

The LIM proteins acquire complex functions due to their sequence diversity and structural nature. For instance, some of the LIM proteins are composed of LIM domains only, while others carry variable functional domains including, homeodomains, cytoskeleton binding domains or catalytic domains (Zheng and Zhao, 2007). Thus, they might function as biosensors that mediate communication between the cytosolic and the nuclear compartments, or through protein-protein interactions in complexes that regulate gene expression (Kadrmas and Beckerle, 2004, Matthews et al., 2013). The LMO proteins are considered as transcriptional cofactors that participate in multiprotein regulatory complexes, and play essential roles in regulation of tissue-specific gene expression and control of cell fate (Matthews et al., 2013, Dawid et al., 1998, Zheng and Zhao, 2007).

All four members of the *LMO* family are implicated in the onset and progression of cancers. *LMO1* overexpression has been reported in T-ALL as a result of the chromosomal translocation t(11;14) (p15;q11), which brings *TCR* gene in proximity of *LMO1* gene (Boehm et al., 1988, McGuire et al., 1989). Chromosomal lesions that target the *LMO2* locus on chromosome 11, band p13 have also been demonstrated in T-ALL (Boehm et al., 1991, Royer-Pokora et al., 1991, Van Vlierberghe et al., 2006). Even in the absence of chromosomal rearrangements near *LMO* genes,

overexpression of LMO1 or LMO2 is found in ~50% of human T-ALLs, most often in association with elevated expression of the related bHLH transcription factors as described in section 1.4.1.

The *LMO3* gene is normally not expressed in T-cells but can be abnormally activated due to translocation t(7;12)(q35;p12.3) in T-ALLs (Simonis et al., 2009). Both LMO1 and LMO3 have also been implicated in neuroblastoma, a type of cancer that affects nerve tissues, and it has been suggested that they regulate similar transcriptional pathways in neuroblastoma (Matthews et al., 2013). LMO3 participates in a transcriptional regulatory complex that targets genes crucial for sympathetic neuron development and its overexpression is indicative of poor prognosis in neuroblastoma (Aoyama et al., 2005, Isogai et al., 2011).

On the other hand, *LMO4* plays multiple roles in breast cancer either by affecting transcriptional programmes of proteins that directly control cell cycle progression, or through interaction with such proteins (Sang et al., 2014). It has been demonstrated that *LMO4* depletion in breast cancer cells induces G2/M arrest, accompanied by increased cell death, amplification of centrosomes and the formation of abnormal mitotic spindles. Conversely, upregulation of *LMO4* in a breast cancer cell line resulted in abnormal mitotic spindle formation and supernumerary centrosomes, indicating the importance of LMO4 in maintaining the fidelity of chromosome segregation during mitosis in these cells (Montanez-Wiscovich et al., 2010). LMO4 also interacts with multiple proteins in complexes that play a transcriptional regulatory function in breast cancer cells (Sang et al., 2014). LMO members exhibit transcriptional regulation functions in normal and cancer cells. However, the exact mechanisms of how they

induce tumorigenesis and their oncogenic roles in cancer largely remain to be elucidated.

### 1.5.2 LMO2 structure and DNA-binding complexes

The *LMO2* gene encodes a 156 amino acid protein. As mentioned above, it comprises two zinc-binding LIM domains, each with two LIM fingers (Figure 1.8) (Appert et al., 2009). The LIM domains do not possess DNA-binding activity, but rather function in protein–protein interactions by building DNA-binding complexes (Wadman et al., 1997).



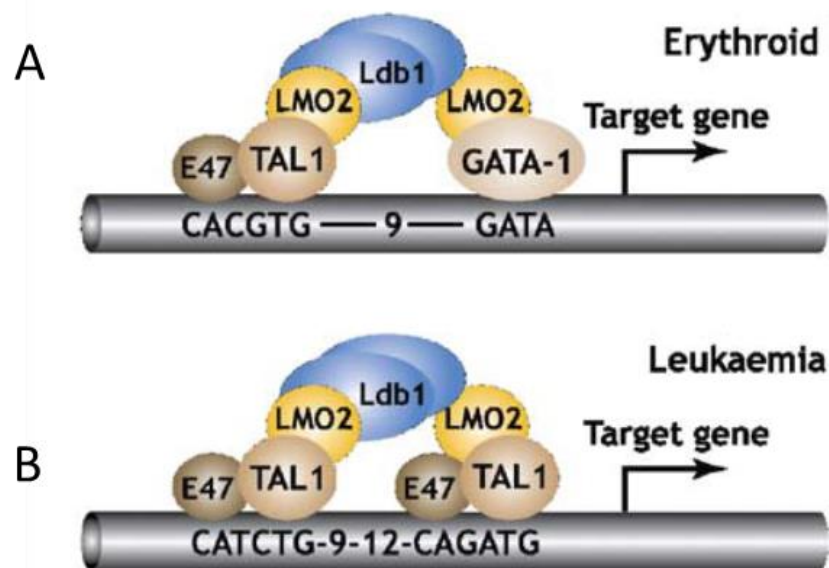
**Figure 1.8 Schematic representation of LMO2 protein structure.**

The first LIM domain (LIM1) and the second LIM domain (LIM2) of LMO2 protein are indicated.

LMO2 forms multimeric complexes and most commonly interacts with two classes of DNA binding proteins; the zinc finger GATA transcription factors (GATA1-3), and an E-box bHLH heterodimer (Bach, 2000). The LMO2 complex that is important for normal erythrocyte development has been most extensively described and is composed of GATA1, TAL1, E2A proteins (E47 or E12), and LIM-domain binding 1 (LDB1) (Wilkinson-White et al., 2011). These multiprotein complexes bind to a bipartite DNA motif, comprising an E-box and GATA site, separated by ~9 bp, which is just less than one turn of the DNA helix. TAL1/E2A heterodimers target E-box sites (CANNTG) N denotes any nucleic acid, whereas GATA1 binds to the DNA sequence (WGATAR) W indicates A/T and R indicates A/G (Figure 1.9A). Another participant of the complex is

Friend of GATA (FOG1) which directly binds GATA1 and stimulates its transcriptional activity (Wilkinson-White et al., 2011). Alternatively, GATA2 or GATA3 and Lymphoblastic leukaemia associated haematopoiesis regulator 1 (LYL1) may replace GATA1 and TAL1 in the LMO2 complex. Deletions of LMO2 complex components cause embryonic lethality and loss of definitive haematopoiesis (Tsai et al., 1994, Smith et al., 2014, Mukhopadhyay et al., 2003).

A distinct LMO2 complex from the one found in erythroid cells has been identified in T-cell malignancies. Isolation of DNA-binding sites by CASTing (cyclic amplification and selection of targets) and band shift assays demonstrated an oligomeric LMO2 complex that can bind to a bipartite DNA motif comprising two E-box sequences approximately 10 bp apart rather than a single E-box and GATA (Figure 1.9B). This complex occurred in immature CD4<sup>+</sup>/CD8<sup>-</sup> DN thymocytes (Grutz et al., 1998). The same complex has been found at the *KIT* promoter in erythroblasts (Vitelli et al., 2000). Another variant of the complex that binds DNA through a motif consisting of two GATA sites flanking an E-box has been shown in erythrocytes at the Erythroid Krüppel-like factor (EKLF) locus (Anderson et al., 1998). This shows the flexibility of the LMO2 complex and its ability to have different partners that can associate to variable DNA motifs.



**Figure 1.9 LMO2 protein complexes.**

LMO2 protein-protein interaction with TAL1, E47, and LDB1 (A) The complex binds a bipartite DNA motif comprising an E-box and GATA site in erythroid cells. (B) An alternative complex that bind bipartite DNA motif comprising of double E-box in T-cell leukaemia. Adapted from (Nam and Rabbitts, 2006).

### 1.5.3 LMO2 function in normal haematopoiesis and vascular development

LMO2 plays a crucial role in haematopoiesis and angiogenesis. It is a key regulator of erythroid cell development. Several studies have demonstrated the pivotal role of LMO2 in lineage specification, particularly at specific stages of erythroid differentiation. A study of homozygous *Lmo2*-null mice showed failure of yolk sac erythropoiesis and embryonic lethality around embryonic day 9-10 (Warren et al., 1994). Due to the limitation in the previous study to assess the role of LMO2 in different stages of haematopoiesis, the Rabbitts group investigated the haematopoietic contribution of homozygous mutant *Lmo2*<sup>-/-</sup> mES cells in chimeric mice. *Lmo2*<sup>-/-</sup> cells did not contribute to any hematopoietic lineage in adult mice. However, reintroduction of *Lmo2* rescued the ability of *Lmo2*<sup>-/-</sup> mES cells to facilitate *in vitro* erythroid differentiation and restored

the macrophage/monocyte developmental lineage *in vivo* (Yamada et al., 1998). These studies showed the essential role of *Lmo2* in early stages of haematopoiesis at the level of the hematopoietic stem cell of the immediate multipotential progeny and probably earlier in development when ventral mesoderm gives rise to these precursors (Rabbitts, 1998).

Furthermore, studying the fate of *Lmo2*<sup>-/-</sup> mES cells in mouse chimeras revealed a transcriptional regulatory function of LMO2 complexes in angiogenesis. *Lmo2*<sup>-/-</sup> mES cells contributed to the capillary network formation until around embryonic day 9, thus LMO2 is not required during early embryogenesis for *de novo* vessel formation. Following embryonic day 10 the mice exhibited a disorganised vascular system indicating the requirement of LMO2 for angiogenic remodelling of vascular networks (Yamada et al., 2000).

Studies have also shown that either *Gata1*, *Gata2*, *Tal1*, or *Ldb1* deletions are fatal to mice. The mice suffered of severe anaemia by mid gestation and died because these proteins are crucial in regulating both primitive and definitive erythropoiesis (Pevny et al., 1991, Perkins et al., 1995, Love et al., 2014). *Lmo2* deficiency resembled loss of the erythroid transcription factor Gata1 or Tal1 (Shivdasani et al., 1995). Moreover, *Ldb1*<sup>-/-</sup> mice showed similar defects in erythropoiesis as the ones described in *Lmo2*<sup>-/-</sup> mice (Warren et al., 1994). The *Ldb1*<sup>-/-</sup> embryos uncovered *Ldb1* function in regulating both definitive erythropoiesis and megakaryopoiesis. *Ldb1* depletion downregulated the expression of multiple erythroid-specific genes, indicating the role of *Ldb1* in erythropoiesis and regulation of the erythroid/megakaryocyte transcriptional program (Li et al., 2010a).

Recently, Stanulović *et al.* showed that *Lmo2*<sup>-/-</sup> mES cells were able to differentiate to Flk-1<sup>+</sup> haemangioblast stage, using an *in vitro* differentiation system. However, these *Lmo2* deficient mES cells had less efficient development of haemogenic endothelium and only generated primitive haematopoietic progenitors. Genome-wide analysis revealed that LMO2 is required at the haemangioblast stage to position the TAL1/LMO2/LDB1 complex at target regulatory elements of genes necessary for the establishment of the haematopoietic developmental program. The depletion of *Lmo2* impaired the ability of TAL1 to recognise target sites and Flk-1<sup>+</sup> cells revealed that TAL1 was required to initiate or sustain *Lmo2* expression (Stanulovic *et al.*, 2017). Altogether, these studies demonstrated the importance of the LMO2 complex involving GATA1, TAL1 and LDB1 in regulating haematopoiesis.

#### **1.5.4 LMO2 oncogenic role in T-ALL**

High levels of *LMO2* expression are normally observed in HSCs, multipotent progenitor cells, and in ETPs. Subsequently, LMO2 gets downregulated during the development of T-cells at the DN2 immature CD4<sup>-</sup>/CD8<sup>-</sup> thymocyte stage and null expression is observed in the following T-cell progenitor cells and mature T-cells (Figure 1.6) (Cleveland *et al.*, 2014, Yui and Rothenberg, 2014).

Human *LMO2* is involved in several different human T-ALL chromosomal rearrangements (Boehm *et al.*, 1990, Garcia *et al.*, 1991). The chromosomal abnormalities that cause the aberrant expression of LMO2 in T-ALL involve translocations: t(11;14)(p13;q11) and t(7;11)(q35;p13), as well as a cryptic deletion (del(11)(p12p13)) (Boehm *et al.*, 1991, Royer-Pokora *et al.*, 1991, Van Vlierberghe *et al.*, 2006). These translocations account for approximately 5% of primary paediatric T-ALL patients (Wu *et al.*, 2015), while the cryptic deletion is implicated in 4% of

paediatric T-ALL cases tested by karyotype analyses (Van Vlierberghe et al., 2006). Although >50% of T-ALL patients have aberrant expression of *LMO2*, only approximately 10% of patients have detectable cytogenetic abnormalities that drive *LMO2* expression (Ferrando et al., 2002). Mutation screening of T-ALL patient samples and T-ALL cell lines identified activating intronic mutations of *LMO2*. These mutations were found in 3.7% of paediatric and 5.5% of adult T-ALL samples (Rahman et al., 2017). The mutations create a neomorphic promoter that drives *LMO2* overexpression and harbours putative *de novo* MYB, ETS1, or RUNX1 binding sites. Taken that MYB, ETS1, or RUNX1 can interact with the *LMO2*/TAL1 complex, this indicates that *LMO2* overexpression is caused by an autoregulatory positive feedback loop in these T-ALL (Rahman et al., 2017).

The upregulation of *LMO2* also occurred as a side effect of trials for the correction of X-linked severe combined immunodeficiency (SCID-X1). The gene therapy was conducted by retroviral gene transfer of the replacement interleukin-2 receptor- $\gamma$  (*IL2RG*) gene into bone marrow cells, which led to aberrant transcription and expression of *LMO2*. The unexpected exponential proliferation of T-cells was thought to be derived from retrovirus enhancer activity on the *LMO2* gene promoter (Hacein-Bey-Abina et al., 2003). Other genetic mutations of genes that have a regulatory effect on T-cell progenitors could also participate to the oncogenesis in these cases. Some of the lesions that have been observed included activating mutations of *NOTCH1* and deletion of the tumour suppressor gene locus *CDKN2A* (Howe et al., 2008, Hacein-Bey-Abina et al., 2008). In the SCID-X1 gene therapy trials *Lmo2* expression was maintained throughout the progeny of the T-cell (Hacein-Bey-Abina et al., 2008, Hacein-Bey-Abina et al., 2003, Howe et al., 2008). Similarly, a mouse model where



*Lmo2* expression was initiated in HSCs and maintained in all hematopoietic cells showed development of T-ALL in 96.7% of the cases. T-ALL in this model exhibited an immature CD8 or CD4 SP/DP (CD8<sup>+</sup>CD4<sup>+/-</sup>) phenotype. Also, clonal immature TCR rearrangement and somatic mutations in cancer genes including *Notch1* were observed (Garcia-Ramirez et al., 2018). Moreover, placing *Lmo2* cDNA under the control of the stem cell specific (Sca1) promoter showed a mosaic of *Lmo2* expression in the mice thymi. These mice developed T-ALL with a clonally immature CD8<sup>+</sup>CD4<sup>+/-</sup> population. A high percentage of the studied Sca1-*Lmo2* T-ALL cases lacked *Lmo2* expression, indicating that early expression of the *Lmo2* in HSC and progenitor cells has the potential to initiate T-ALL without any need for its continuous expression to develop T-ALL (Garcia-Ramirez et al., 2018). This emphasises the role of LMO2 in reprogramming cellular identity of HSCs and progenitor cells into a leukemic one before entering the thymus. T-ALL cells in thymus-deficient mice with Sca1-*Lmo2* overexpression showed enrichment in ETP-ALL genes, stemness, and pluripotency (Garcia-Ramirez et al., 2018).

Additionally, a study of *Lmo2* knockout mice proved that LMO2 is not required for lymphoid development. Efficient deletion of *Lmo2* was accomplished using conditional knockout of mouse *Lmo2* with loxP-flanked *Lmo2* and Cre recombinase alleles driven by the promoters of the lymphoid-specific genes *Rag1*, *CD19*, and *Lck*. Lymphopoiesis in T- and B-cell lineages were not disturbed in the absence of *Lmo2* and normal distributions of CD4<sup>+</sup>/CD8<sup>+</sup> thymocytes were observed. This suggested that the role of LMO2 in T-cell oncogenesis occurs as a result of reprogramming of gene expression as a consequence of its imposed expression in T-cell precursors (McCormack et al., 2003).

A research group generated clonal cell lines from T-ALLs derived from *CD2-Lmo2* transgenic mice. This cell line represents three different stages of T-cell differentiation, DN, DP, and ISP. These cells demonstrated different levels of CD4 expression. One population was presented with high expression of CD4 and the other with no expression at all. Subsequent cell cultures of each population separately revealed that CD4<sup>-</sup> and CD4<sup>+</sup> populations interconvert and have dissimilar rates of apoptosis (Cleveland et al., 2014). This indicated that T-ALLs may arise from different stages of T-cell differentiation because LMO2 overexpression blocks apoptosis of T-cells at DN3 stage, while they still highly express CD4 and CD8 (Cleveland et al., 2014, Chambers and Rabbitts, 2015). Also enforcing the expression of *Lmo2* led to a differentiation block at the CD4<sup>-</sup>/CD8<sup>-</sup> DN thymocyte stage and clonal T-cell malignancy in transgenic mice (Larson et al., 1995).

A study using a humanised mouse model, which was transplanted with *LMO2*-transduced human hematopoietic stem/progenitor cells, suggested three potential cellular mechanisms that lead to T-ALL development. First, *LMO2* overexpression led to delayed T-cell development with a block in the DN/ISP compartment. A small subset of *LMO2*-overexpressing cells in these mice retained CD34 expression. Also, a reduced number of DP cells was observed with a relative increase in ISP and CD3<sup>-</sup> CD4<sup>-</sup> CD8<sup>-</sup> triple negative cells. Second, *LMO2* overexpression caused accumulation of DP CD3<sup>-</sup> cells. A high percentage of double positive CD3<sup>-</sup> cells was observed in mice transplanted with *LMO2*-overexpressing CD34<sup>+</sup> cells, compared to GFP-transplanted mice. Third, the ratio of CD8 over CD4 SP cells was different between GFP and *LMO2*-transduced cells. This altered CD8/CD4 ratio became more prominent with higher expression levels of *LMO2* and was also found in peripheral blood

(Wiekmeijer et al., 2016). Moreover, they compared the genes that were differentially expressed between LMO2 and GFP DP cells with the human T-ALL microarray gene expression data sets described by Homminga *et al.* Many upregulated genes in these DP cells clustered with the “proliferative” and “TAL/LMO” subtypes in human T-ALLs seen in Figure 1.7, indicating that the effects and mechanisms that were observed after transplantation of *LMO2*-overexpressing cells in mice are also at work in human T-ALL (Homminga et al., 2011, Wiekmeijer et al., 2016). Although these studies have highlighted potential mechanisms through which LMO2 induce T-cell malignancy, it is still unknown what are the exact roles it plays in leukaemia. It also remains unclear what are the transcriptional regulatory functions that LMO2 complexes exert in T-ALL.

#### **1.5.5 LMO2 complexes in T-ALL**

In a study by Draheim *et al.* *Lmo2* transgenic mouse lines were mated with transgenic mice expressing wild type *Tal1* or a *Tal1* harbouring a mutation that alters its DNA binding domain (Draheim et al., 2011). This mutant form of TAL1 heterodimerises with E47 or HEB proteins but fails to stably bind E-box motif (Hsu et al., 1994, O'Neil et al., 2001). The comparison between these transgenic mice showed that T-ALL was accelerated in both lines, suggesting that LMO2 does not contribute to leukaemia by enhancing or altering the transcriptional activity of the TAL1/E47 or HEB heterodimer (Draheim et al., 2011). Perturbed thymocyte development was also demonstrated due to repression of E47/HEB target genes which are important for thymocytes differentiation. These data revealed that LMO2 binding to TAL1 or mutTAL1/E47 or HEB heterodimers may stabilise the TAL1/E47 or HEB complex which results in reductions in E47/HEB transcriptional activity (Draheim et al., 2011).

Furthermore, *Lmo2*-transgenic mice with depleted *Tal1* or *Ly11* were studied to address which of these factors participates in *Lmo2* driven thymocyte self-renewal and Leukaemogenesis (McCormack et al., 2013). The outcomes demonstrated that TAL1 is dispensable for *Lmo2*-driven leukaemia but LYL1 cooperation with LMO2 plays an important role in leukemogenesis. The oncogenic functions of LMO2/LYL1 complex include upregulation of a stem cell-like gene signature, aberrant self-renewal of thymocytes, and subsequent generation of T-cell leukaemia. These mice models expressed *Ly11* in pre-leukemic and leukemic stem cells which described stage-specific expression of the *Lmo2*-induced gene expression program. Similarly, LYL1 and LMO2 are both expressed in ETP-ALL patient samples implementing the need of LYL1 for the development of ETP-ALL cells (McCormack et al., 2013). Overall, the contributions of LMO2 to T-ALL probably arise from a combination of transcriptional reprogramming functions of LMO2 complexes. The full composition of LMO2 complex in T-ALLs has not been fully studied. Hence, LMO2 interacting partners remain to be elucidated which will help better understand its oncogenic function in T-ALL.

The first aim of this PhD study was to identify new LMO2 interacting partners beyond the known LMO2 complex members. Our investigations through proteomic analysis revealed the Plant homeodomain zinc finger protein 6 (PHF6) as a new interactor of LMO2. Several studies have identified *PHF6* mutations in different types of leukaemia including T-ALL (Van Vlierberghe et al., 2010, Van Vlierberghe et al., 2011, Wang et al., 2011).

## 1.6 Plant homeodomain zinc finger protein 6 (PHF6)

### 1.6.1 History of PHF6

PHF6 is the plant homeodomain zinc finger protein 6, which was discovered as the underlying cause of the X-linked intellectual disability disorder Börjesson-Forssman-Lehman syndrome (BFLS) (Lower et al., 2002). Several groups have identified mutations of *PHF6* associated with BFLS cases (Baumstark et al., 2003, Crawford et al., 2006, Di Donato et al., 2014). The *PHF6* gene is located on the non-autosomal region of the X-chromosome and spans a 90-kb region of genomic DNA. The human *PHF6* gene consists of 11 exons, that are transcribed into an alternatively spliced 4.5-kb transcript. Exons 1 and 11 comprise the 5' and 3' untranslated regions (UTRs), respectively, while exons 2-10 encode for a 1.1-kb open reading frame (ORF), which encodes a 365 amino acid protein (41 kDa). Another mRNA isoform exist which incorporates intron 10 to increase the size of the 3' UTR (Lower et al., 2002). The PHF6 protein is highly conserved among vertebrates; for example ~97.5% amino acid identity occurs between humans and mice (Lower et al., 2002, Perry, 2006).

The *Phf6* gene and protein are widely expressed in various tissues in mice, including the developing central nervous system, the anterior pituitary gland, the primordia of facial features, the limb buds and hematopoietic cells (Van Vlierberghe et al., 2011, Voss et al., 2007). A comparison between *Phf6* expression in HSCs, myeloid and lymphoid progenitors in mouse revealed higher levels of *Phf6* in lymphoid cells. The highest levels expressed in the T-lineage were observed in double positive CD4<sup>+</sup>/CD8<sup>+</sup> and single positive CD4<sup>+</sup> or CD8<sup>+</sup>. However, in the B-lineage the pre-B cells demonstrated the highest levels of *Phf6* expression of all cells that were analysed (Van Vlierberghe et al., 2011). In humans, high levels of PHF6 expression were detected in

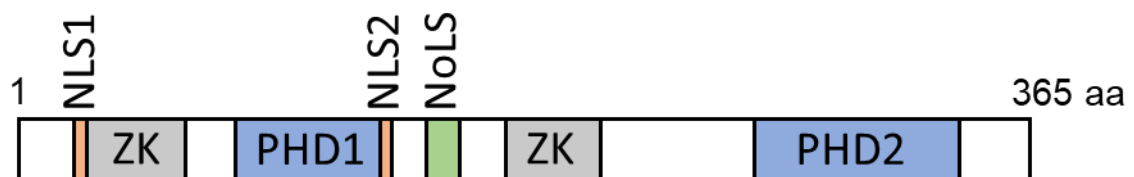
the thymus, thyroid, and ovary, and moderate levels in the testes, adipose tissue and spleen (Van Vlierberghe et al., 2010). At different stages of thymocyte development, *PHF6* transcripts showed a marked upregulation in CD4<sup>+</sup>/CD8<sup>+</sup> double positive cells (Van Vlierberghe et al., 2010), consistent with what was observed in mouse T-lymphocytes (Van Vlierberghe et al., 2011). These observations suggested a putative role for PHF6 in haematopoiesis.

### **1.6.2 PHF6 structure**

The *PHF6* gene is a member of the plant homeodomain (PHD)-like finger (PHF) family and encodes a protein with two zinc finger domains (ZaP1: aa 14–134; ZaP2: aa 209–332) that are derived from a PZP motif (Perry, 2006). The PZP motifs consist of elongated plant homeodomain-type zinc fingers (ePHD): PHD1 at position 79-131 (Cys4-His-Cys2-His), followed by a zinc knuckle, and an atypical PHD2 at position 277-329 (Cys4-His-Cys2-His) (Figure 1.10) (Todd and Picketts, 2012). Such zinc-binding motifs can have a neutral or negative charge (Liu et al., 2012). PHD domains are most commonly found as part of proteins involved in transcriptional regulation and/or modification of chromatin structure. They are functionally versatile and act as epigenome readers that control gene expression through recruitment of multiprotein complexes of chromatin regulators and transcription factors (Sanchez and Zhou, 2011). A subset of PHD fingers has been shown to bind trimethylated lysine 4 of histone H3 (H3K4me3) (Wysocka et al., 2006). However, the atypical PHD domains present in PHF6 cannot interact with histones and have a net positively charged surface, suggestive of interactions with negatively charged molecules like DNA (Liu et al., 2012). Studies based on Nuclear magnetic resonance (NMR) solution and crystal structures indicate that PHF6 ZaP2 can bind double-stranded DNA templates in a

sequence-independent manner but not histones (Liu et al., 2012, Liu et al., 2014). Using high throughput screen analysis, PHF6 has also demonstrated putative ability to interact with mRNA (Castello et al., 2012).

Moreover, the unique structure of PHF6 consist of two nuclear localisation sequences (NLS1; NLS2) and a nucleolar localisation sequence (NoLS) (Figure 1.10) (Lower et al., 2002, Todd and Picketts, 2012). PHF6 localisation to the nucleus and nucleolus have been confirmed in several studies by immunocytochemistry, subcellular fractionation, and mass spectrometry analysis (Lower et al., 2002, Todd and Picketts, 2012, Wang et al., 2013, Zhang et al., 2013).



**Figure 1.10 Schematic representation of PHF6 protein structure.**

PHF6 protein consists of 365 amino acids, comprising two zinc knuckles (ZK) in grey, two atypical plant homeodomain (PHD) zinc-finger domains in blue, two nuclear localisation sequences (NLS) in orange and a nucleolar localisation sequence (NoLS) in green.

### 1.6.3 Functional interactions of PHF6

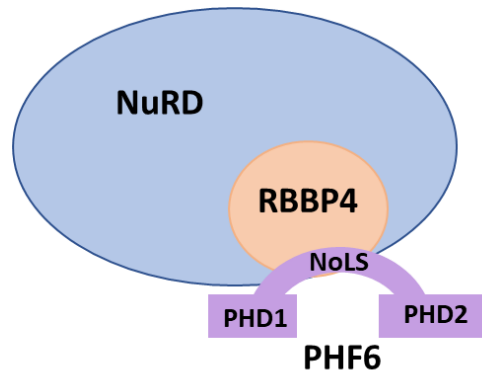
Research efforts have been conducted to investigate the molecular mechanisms of PHF6 that may contribute to the developmental processes of neurogenesis and haematopoiesis. Investigations of PHF6 containing complexes, structural characterisation of the PHF6 zinc finger domain, nuclear and nucleolar localisation studies, have revealed some of the potential roles that PHF6 play. However, there is

not much known about the cellular function(s) and underlying mechanisms controlled by PHF6.

### **PHF6 interaction with the Nucleosome Remodelling and Deacetylation (NuRD)**

Todd and Picketts performed immunoprecipitation followed by mass spectrometry (IP/MS) analysis to identify the interacting partners of PHF6. These experiments revealed that PHF6 co-purified with multiple constituents of the Nucleosome Remodelling and histone Deacetylation (NuRD) complex (CHD3/4, HDAC1, and RBBP4/7) (Todd and Picketts, 2012). To identify which of the NuRD proteins directly interact with PHF6, Liu *et al.* showed that it binds the RBBP4 subunit through a conserved amino acid sequence and the crystal structure demonstrated that the PHF6 peptide embeds with a binding pocket located upon the surface of the RBBP4  $\beta$ -propeller (Liu et al., 2014, Liu et al., 2015). The binding domain of RBBP4 overlaps with the PHF6 NoLS sequence (Figure 1.11), suggesting that these two activities may be mutually exclusive. Hence, PHF6 interaction with NuRD complex components was specifically confined to the nucleoplasm, but not found within the nucleolus (Todd and Picketts, 2012). The NuRD complex is a multifunctional transcriptional regulator with various target genes that are involved in embryogenesis, oncogenesis, neurogenesis, and haematopoiesis (Lai and Wade, 2011, Yoshida et al., 2008, Yamada et al., 2014, Reynolds et al., 2012). However, the NuRD complex exerts either a repressive or activating transcriptional function on its target genes depending on its interacting partners (Williams et al., 2004, Wang et al., 2009). Interacting proteins of NuRD that possess zinc finger domains can direct the complex to its target genes. Likewise, PHF6 could potentially lead the NuRD association to a certain set of genes via its binding affinity to dsDNA (Basta and Rauchman, 2015, Todd and Picketts, 2012).





**Figure 1.11 PHF6 interaction with NuRD complex.**

PHF6 binds the NuRD complex through a direct interaction with RBBP4, where the blue oval represents the multiprotein NuRD complex.

### **PHF6 interaction with RNA Polymerase II Associated Factor 1 (PAF1)**

In a study of the function of PHF6 in neurogenesis, Zhang *et al.* discovered a novel function for PHF6 in the development of the cerebral cortex *in vivo*. Knockdown of *Phf6* in mouse cerebral cortices caused impaired neuronal migration (Zhang et al., 2013). To investigate whether PHF6 associated with factors that regulate transcription in neuronal development, they conducted IP/MS experiments. All four core components of the PAF1 transcription elongation complex, PAF1, LEO1, CDC73 and CTR9, were found as interacting partners of PHF6 (Zhang et al., 2013). Because the PAF1 complex promotes transcription (Marton and Desiderio, 2008), the authors studied downstream transcriptional changes that promote neuronal migration using gene expression arrays. They identified *Neuroglycan C/Chondroitin Sulfate Proteoglycan 5 (NGC/CSPG5)* as a commonly regulated gene target. Furthermore, when the NGC/CSPG5 transcript was electroporated into cortical tissue alongside *Phf6* shRNA, the neuronal migration phenotype was rescued (Zhang et al., 2013). Consistently, similar results were demonstrated by another research group where expression of *Phf6* rescued the

neuronal migration defects caused by miR-128 expression on neuronal migration, growth, and intrinsic physiological properties of these cells (Franzoni et al., 2015).

### **PHF6 interaction with Upstream binding factor (UBF)**

An interaction between PHF6 and Upstream binding factor (UBF) has been demonstrated in two separate studies, implying a nucleolar function of PHF6 (Wang et al., 2013, Zhang et al., 2013). UBF is a ribosomal DNA (rDNA) transcriptional activation factor that associates with the RNA Pol I pre-initiation complex (Learned et al., 1986). Wang *et al.* reported that PHF6 binds to the nucleolus and associates with ribosomal RNA (rRNA) promoter. The interplay of PHF6 with UBF, through its PHD1 domain, downregulated rRNA transcription, by affecting the protein level of UBF (Wang et al., 2013). Additionally, they revealed that *PHF6* depletion impairs cell proliferation and arrests cells at the G2/M phase, leading to accumulation of DNA damage in the cell with high levels of phosphorylated H2AX ( $\gamma$ H2AX). The effect of DNA damage at the rDNA locus in *PHF6* deprived cells was reversed by knocking down UBF or overexpressing RNASEH1, indicating that there is a functional link between rRNA synthesis and genomic stability at the rDNA locus. These data uncovered a regulatory function of PHF6 that is involved in rRNA synthesis, which may contribute to its roles in cell cycle control, genomic maintenance, and tumour suppression (Wang et al., 2013).

#### **1.6.4 PHF6 functions in haematopoiesis and leukaemia**

The role of genes involved in epigenetic gene regulation has been demonstrated in genomic studies on patients with haematological cancers. These genes are known, or suspected, to play a role in modifying the chromatin structure and usually acquire

lesions, *i.e.* mutations, that lead to malignancies (Bowman et al., 2018, Van Vlierberghe et al., 2010, Van Vlierberghe et al., 2011). Research efforts have been conducted to understand how these genes regulate normal haematopoiesis and how their mutations contribute to tumorigenic transformation and/or progression, however their functions are still debatable. Since inactivating mutations of PHF6 have been identified in T-ALL, AML and myelodysplastic syndrome (Mori et al., 2016, Van Vlierberghe et al., 2010, Van Vlierberghe et al., 2011), several groups have been studying its role in haematopoietic development and regulation of leukaemia.

In one study, *Phf6* was identified through a genome-scale shRNA screen, as a “context-specific” regulator of leukaemia development (Meacham et al., 2015). Depletion of *Phf6* negatively impaired pre B-ALL tumour growth in all of the hematopoietic organs they tested, including bone marrow, spleen, and blood. Additionally, the absence of *Phf6* in AML and T-cell lymphoma had a neutral effect in T-cell lymphomas but promoted significant AML cells growth *in vivo*. These data suggest distinct differential requirements for *Phf6* gene function in B-cell, T-cell, and myeloid cell malignancies. Moreover, they performed genome wide data analysis on T-ALL cell line (Jurkat), which showed that PHF6 associated with promoters of active genes. Some of these PHF6 target genes play essential roles in hematopoietic tumours including *RUNX1*, *DMNT3A*, *NOTCH1*, and *JAG1*. These findings propose that PHF6 may participate in transcriptional complexes that influences tumour progression through the regulation of leukaemia associated genes (Meacham et al., 2015).

A subsequent study by the same group demonstrated the role of PHF6 in regulating chromatin accessibility to lineage-specific TFs (Soto-Feliciano et al., 2017). Through

integrated genomics and *in vivo* studies, they showed that, upon loss of *Phf6*, chromatin structure is modulated around B-cell- and T-cell specific genes, resulting in focal genomic plasticity and acquisition of T-cell lineage markers in B-ALL cells. They revealed a potential role of PHF6 in regulating transcriptional programs, by binding nucleosomes at specific genomic regions and remodelling the chromatin architecture. Furthermore, *Phf6* depletion resulted in altered chromatin accessibility and nucleosome positioning which caused lineage disruption. These outcomes imply a possible mechanism through which PHF6 functions in lymphoid malignancies (Soto-Feliciano et al., 2017).

More recent, independent groups have investigated the role of PHF6 in HSCs. McRae *et al.* showed that *Phf6* depletion in mice altered the hematopoietic stem and progenitor cell homeostasis (McRae et al., 2019). They observed a reduced number of HSCs, an increased number of hematopoietic progenitor cells, and an increased proportion of cycling stem and progenitor cells, upon loss of *Phf6*. Although *Phf6* did not affect HSC self-renewal, the absence of *Phf6* in HSCs enhanced differentiation to white blood cells, suggesting that these cells have enhanced competitive repopulating ability. Alterations in thymus cellularity or on ETP number was not observed in the absence of PHF6. In addition, they demonstrated that PHF6 controlled haematopoietic cell homeostasis partly through the regulation of interferon signalling. They also showed that PHF6 is a putative haematopoietic tumour suppressor, as *TLX3* overexpression in the absence of PHF6 caused fully penetrant early-onset leukaemia, while *TLX3* expression alone caused partially penetrant leukaemia (McRae et al., 2019).

In contrast, two other studies reported that *Phf6* depletion enhances self-renewal of HSCs in a setting of serial transplantation in mice (Wendorff et al., 2019, Miyagi et al.,

2019). Miyagi *et al.* demonstrated the effect of *Phf6* deletion in hematopoietic cells at various developmental stages. In embryos, loss of *Phf6* increased the ability of HSCs to proliferate in cultures and to reconstitute haematopoiesis in recipient mice. In neonates, *Phf6* inactivation immediately gave cycling HSCs a competitive repopulation advantage. In adults, *Phf6* depleted HSCs required a longer time and/or haematopoietic stress to activate dormant HSCs and exert an obvious competitive advantage. Furthermore, *Phf6* depletion downregulated the expression of genes involved in tumour necrosis factor alpha (TNF $\alpha$ ) signalling, which counteracted the TNF $\alpha$ -mediated growth inhibition on HSCs. However, loss of *Phf6* did not induce haematopoietic tumorigenesis (Miyagi *et al.*, 2019). Conversely, data from Wendorff *et al.* supports the tumour suppressor role of *Phf6* in hematopoietic cells. They reported that *Phf6* inactivation in leukaemia lymphoblasts induced a leukaemia stem cell transcriptional program and enhanced T-ALL leukaemia-initiating cell activity. They also observed that development of NOTCH1-induced T-ALL was reduced upon *Phf6* depletion (Wendorff *et al.*, 2019). Altogether, these studies highlight PHF6 functions as a potential transcriptional regulator and tumour suppressor in haematopoietic cells.

#### **1.6.5 PHF6 mutations in leukaemia**

Somatic inactivating mutations of *PHF6* have been described in T-ALL patients for the first time by Van Vlierberghe and colleagues (Van Vlierberghe *et al.*, 2010). Initially, they analysed data from primary T-ALL patient samples, which demonstrated a recurrent deletion in X chromosome band 26q. The *PHF6* gene is located on the Xq26 locus and deletions in this region are somatically acquired leukaemia-associated genetic events (Van Vlierberghe *et al.*, 2010, Baumstark *et al.*, 2003). Subsequent mutational analyses of *PHF6* in a panel of paediatric and adult T-ALL primary samples

revealed nonsense or missense mutations in *PHF6* at a prevalence of 16% in paediatric and 38% in adult patients (Van Vlierberghe et al., 2010). The identified *PHF6* missense mutations target zinc ion stabilising residues at the ZaP domain on the second PHD finger (PHD2). The rest of reported *PHF6* lesions involve deletions, frameshifts and nonsense mutations that span throughout the gene (Van Vlierberghe et al., 2010). Other research groups have also reported *PHF6* mutations in T-ALL (Grossmann et al., 2013, Wang et al., 2011, Yoo et al., 2012, Huh et al., 2013, Jang et al., 2019). Even though Vlierberghe *et al.* showed that the majority of *PHF6* lesions are found in male subjects with T-ALL, the other studies did not report any gender differences (Grossmann et al., 2013, Wang et al., 2011, Yoo et al., 2012, Huh et al., 2013, Van Vlierberghe et al., 2010, Jang et al., 2019).

Additionally, inactivating mutations of *PHF6* have been observed in cancer that overexpress *TLX1* and *TLX3* oncogenes (Van Vlierberghe et al., 2010). This highlights the role of *PHF6* as a tumour suppressor, since *TLX1* induced T-cell leukaemia commonly undergo developmental arrest at DP stage where *PHF6* is normally highly expressed (Van Vlierberghe et al., 2010, Ferrando et al., 2002). *PHF6* lesions have also been observed in combination with *JAK1* mutations, *SET-NUP214* translocations, *CDKN2A/B* deletions and activating *NOTCH1* mutations, which account for more than half of the genetic abnormalities reported in T-ALL patients (Huh et al., 2013, Weng et al., 2004, Wang et al., 2011). On the other hand, cases of B-cell acute lymphoblastic leukaemia (B-ALL) that were screened for genetic lesions did not demonstrate any *PHF6* mutations implicating that inactivation of *PHF6* might be confined to the T-cell lineage tumours (Van Vlierberghe et al., 2010, Yoo et al., 2012).

To investigate whether inactivating mutations of *PHF6* are involved in the pathogenesis of AML, Ferrando and colleagues sequenced all coding exons of *PHF6* in 353 samples of AML subjects. Moreover, 41 AML patient samples were tested for deletions in the *PHF6* gene using real-time quantitative PCR (RT-PCR). This study documented genetic lesions of *PHF6* consisting of deletions, frameshifts, nonsense mutations and missense mutations, particularly enriched in the PHD2 domain (Van Vlierberghe et al., 2011). AML cases harbouring *PHF6* mutations have also been reported to be accompanied by *RUNX1* mutation (Cancer Genome Atlas Research et al., 2013, Patel et al., 2012). It has been demonstrated that *PHF6* is normally expressed at lower levels in CMP or granulocyte/macrophage precursors than HSCs or erythroid precursors, but *PHF6* alterations have not been correlated to a specific developmental stage of myeloid cells (Van Vlierberghe et al., 2011). Furthermore, a study of 81 patients of chronic myeloid leukaemia (CML) identified two cases with inactivating *PHF6* mutations, however no further studies reported *PHF6* lesions in this type of cancer (Li et al., 2013b).

## **1.7 Aims and objectives**

The crucial function of the LMO2/TAL1 complex in haematopoiesis and particularly erythropoiesis has been defined in many studies. During early T-cell development, the expression of the TAL1 transcription factor and LMO2 transcription co-regulator become suppressed in T-cells and the aberrant expression of these factors leads to T-ALL. The composition of the LMO2 complex has not been thoroughly studied in T-ALL and the oncogenic role of LMO2 is not fully understood.

The main questions we addressed in this project were:

- What are the other members of the LMO2 complex beyond those already known in T-ALL?
- Does the complex exert a transcriptional regulatory function in these cells?
- What are the potential targets of the complex?
- What is the mechanism of the LMO2 complex in Leukemogenesis?

Therefore, the first aim of this project was to perform proteomic analyses to identify LMO2 interacting partners in human T-ALL cells. The second aim was to perform an integrative analyses of accessible chromatin, genomic occupancy, and gene expression in T-ALL cells to gain better understanding of which transcriptional and/or epigenetic regulators are recruited by LMO2 to its target sites, explore which genes this complex regulates, what is the effect on gene expression and how they influence proliferation, differentiation and survival of leukemic cells.

Lastly, we aimed to explore the LMO2 complex with its binding partners in an *in vitro* myeloid differentiation system. We identified PHF6 protein as an interacting partner of the LMO2 complex in both T-ALL and myeloid cells. Thus, the experimental methods and objectives of T-ALL cells were applied on wild type and *Phf6* knockdown myeloid progenitors that can be induced to differentiate. Additionally, the effect of PHF6 and LMO2 loss on chromosome and genome stability in these myeloid cells was studied through cytogenetic, cell cycle, and DNA damage experiments.



## **2 Chapter 2: Materials and methods**

### **2.1 Cell culture**

T-ALL human cell lines (ARR, DU.528, HSB2, and CCRF-CEM) were cultured in RPMI 1640 medium (Gibco), supplemented with 10% heat inactivated foetal bovine serum (HI-FBS, Life Technologies), 2 mM GlutaMax (Gibco), 5 U/ml Penicillin/Streptomycin (Life Technologies) and 0.075 mM Monothioglycerol (MTG, Sigma). T-ALL cells were maintained at a concentration of  $0.4 - 2 \times 10^6/\text{ml}$  and were spun down and resuspended in fresh medium every other day.

Murine PUER cell lines were maintained in phenol red-free Iscove's Modified Dulbecco's Medium (Gibco), supplemented with 10% HI-FBS, 2 mM GlutaMax, 5 U/ml Penicillin/Streptomycin, 0.075 mM MTG and 10  $\mu\text{g}/\text{ml}$  IL-3 (PeproTECH). PUER cells and derivative cell lines were maintained at a concentration of  $0.2 - 2 \times 10^6/\text{ml}$  and were spun down and resuspended in fresh medium every other day. Addition of 100 nM of 4-hydroxytamoxifen (OHT) to PUER cells was used to trigger their terminal differentiation into macrophages (Walsh et al., 2002). Differentiation status was analysed by the presence of adherent cells by light microscopy, morphologic changes in Kwik-Diff stained (Kwik-Diff™ Stains, ThermoFisher) cytospin preparations respectively and flow cytometry for macrophage surface marker expression.

PlatE cells were grown in DMEM Dulbecco's Modified Eagle Medium (Gibco), supplemented with 10% HI-FBS, 2 mM GlutaMax, 5 U/ml Penicillin/Streptomycin and 0.075 mM MTG. Cells were passed when reaching 70-80% confluence by removing media from the plate, washing with PBS (Sigma-Aldrich) and treating with trypsin (Gibco), followed by resuspending the cells in growth media. All cells were incubated at 37° C in a humidified incubator with 5% CO<sub>2</sub>.

## **2.2 Cloning steps for cDNA**

### **2.2.1 RNA isolation**

RNA isolation was carried out using a NucleoSpin® RNA kit (MACHEREY-NAGEL), according to the manufacturer's protocol. In brief, cells were harvested and resuspended in lysis buffer containing 143 mM  $\beta$ -mercaptoethanol (Gibco). The lysate was filtered, bound to an extraction column, and DNA was digested. The column was washed and dried and RNA was eluted in RNase-free H<sub>2</sub>O. RNA was quantified by 260 nm absorbance using a NanoDrop™ 2000 UV-Vis Spectrophotometer (ThermoFisher). RNA samples were stored at -80° C, after the addition of RNase inhibitor (Invitrogen).

### **2.2.2 cDNA synthesis**

The mRNA was reverse-transcribed into its complementary DNA (cDNA) copy. For first strand synthesis, a reaction mix of 0.5-1  $\mu$ g RNA, 1  $\mu$ l oligo d(T)20 primer (0.5  $\mu$ g/ml, Life Technologies) and 10  $\mu$ l nuclease-free H<sub>2</sub>O were incubated at 70° C for 5 min. The samples were then placed on ice and a second reaction mix was added, consisting of 1x Reverse Transcription (RT) buffer (Invitrogen), 0.5  $\mu$ l 20 mM dNTP (Bioline), 1  $\mu$ l 0.2 mM Moloney Murine Leukemia Virus (MMLV) reverse transcriptase (Promega), 0.5  $\mu$ l 0.4 mM Recombinant RNase inhibitor. The samples were incubated at 42° C for 45 min and then at 95° C for 5 min to inactivate the enzyme. The cDNA samples were stored at -20° C.

### **2.2.3 Amplification of cDNA**

The *Phf6* gene was amplified by polymerase chain reaction (PCR) from 10 ng of mouse cDNA using Phusion® High-Fidelity PCR Master mix (ThermoFisher), 1  $\mu$ M of *Phf6*

cloning primers, the primer sequences (Table 2.1) were modified to have BglII and XhoI restriction sites compatible for cloning into the MIGR1 vector, in 50 µl reaction volume and run in a Tprofessional TRIO thermocycler (Biometra, Analytik Jena). Reactions were carried out using the thermocycler parameters in Table 2.2.

**Table 2.1 cDNA amplification primers.**

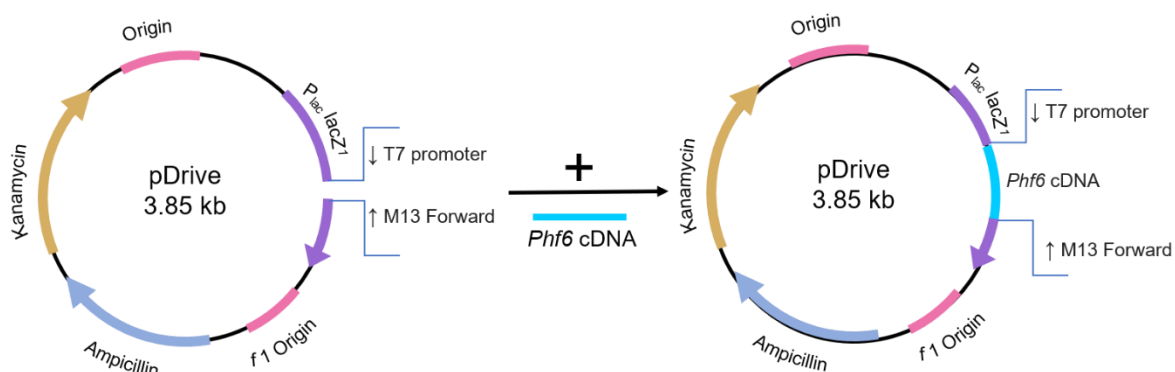
Name	Primer sequence
Phf6_BglII_F	AGATCTATGTCAAGCTCAATTGAACA
Phf6_XhoI_R	CTCGAGTTATCTTACTTGTAATTCCT

**Table 2.2 Thermocycling conditions for cDNA amplification by PCR.**

Step	Temperature	Time	Number of cycles
Denaturation	92° C	1 min	X35
Annealing	95° C	30 sec	
Extension	65° C	30 sec	
Elongation	72° C	1 min	
Pause	4° C	10 min	

#### 2.2.4 Cloning of cDNA

The *Phf6* PCR product was cloned into pDrive plasmid (QIAGEN, PCR cloning kit), according to the manufacturer's protocol. Briefly, the PCR product, pDrive cloning vector and ligation master mix were mixed and incubated for 30 min-2 h at 37° C (Figure 2.1).



**Figure 2.1 cDNA cloning in pDrive plasmid.**

The cDNA of *Phf6* gene was amplified by PCR using *Phf6* sequencing primers that have BglII and XhoI restriction sites and cloned in pDrive plasmid. Abbreviation: *P lac lacZ*<sup>1</sup>, LacZ gene promoter.

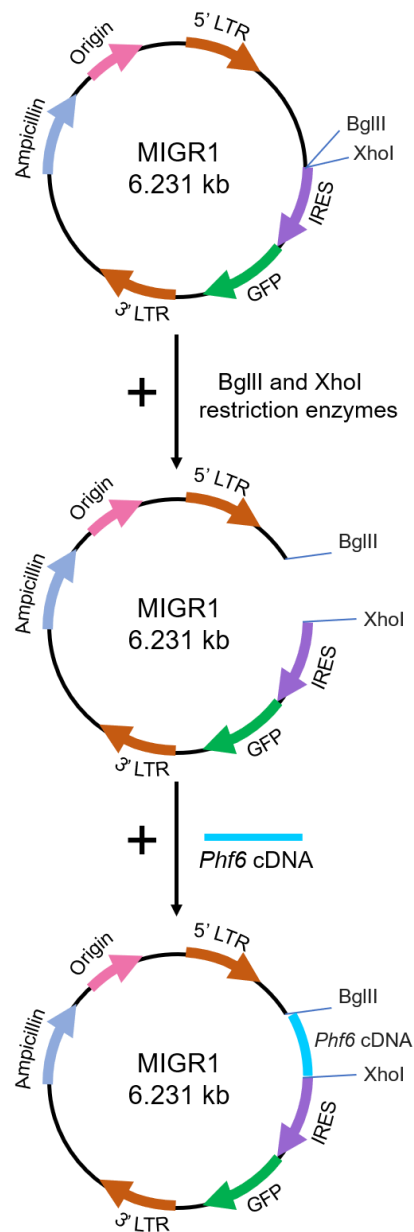
The ligation product was transformed into chemically-competent JM109 *E. coli* cells (Promega) by incubation on ice for 2 min, heat shock at 42° C for 30 sec, recovery on ice for 2 min and incubation with 250 µl of LB medium (Lysogeny Broth- Sigma) for 5 min. Transformed cells were plated onto LB agar plates, containing 100 µg/ml ampicillin (Sigma-Aldrich), 10x of X-gal and 10x of IPDG (both Bioline) and placed in a 37° C incubator overnight. X-gal and IPDG are used for blue-white screening, a technique for the identification of recombinant bacteria. The enzyme  $\beta$ -galactosidase which cleaves lactose into glucose and galactose occur in *E. coli*. If  $\beta$ -galactosidase is produced, X-gal produce an insoluble blue pigment which is an indication of colonies formed by non-recombinant cells. The plasmid vector pDrive is constructed in a way that this  $\alpha$ -complementation process serves as a marker for recombination where functional  $\beta$ -galactosidase enzyme is not produced, and recombinant colonies appear white. The white colonies were picked to inoculate 2 ml cultures of LB medium with 100 µg/ml ampicillin, which were placed in a shaking incubator at 37° C overnight.

The following morning, plasmid mini-preps were made from the cultures by pelleting the cells and resuspending them in TENS buffer (10 mM Tris pH8, 1 mM EDTA (TE), 10 M NaOH, 10% SDS, dd H<sub>2</sub>O), after which 3 M sodium acetate was added. Samples were mixed by inversion and spun down at 11,000 g for 3 min. The supernatant was transferred to a clean tube, after which, DNA was precipitated with 100% ethanol and then washed twice with 70% ethanol and left to airdry. DNA was eluted with 0.1x TE mixed with 20 µg/ml RNase A and was quantified by 260 nm absorbance using a NanoDrop™ 2000 UV-Vis Spectrophotometer. Colonies were validated for correct insert sequence and orientation by Sanger sequencing (Source Bioscience), using the T7 promoter and M13 forward sequencing primers (Table 2.3).

**Table 2.3 pDrive sequencing primers.**

Name	Primer sequence
<b>T7 promoter</b>	GTAATACGACTCACTATAG
<b>M13 forward</b>	GTAAAACGACGGCCAGT

Subsequently, 10 µg of extracted DNA was digested with 1 µl BglII and 1 µl XhoI restriction enzymes (ThermoFisher) and 1x fast digest buffer (ThermoFisher), which was then incubated at 37° C for 1 h. Samples were run on a 1% agarose gel (Sigma) and the correct size *Phf6* fragment was excised and eluted from the gel using a MinElute® Gel Extraction Kit (QIAGEN), according to the manufacturer's protocol. MIGR1 plasmid (Addgene) (Figure 2.2) was digested with BglII and XhoI under the same conditions, run on 1% agarose gel, excised and purified as described above. The *Phf6* fragment was cloned into the digested MIGR1 (Figure 2.2), following the same protocol as described above for pDrive cloning.



**Figure 2.2 cDNA cloning in MIGR1 plasmid.**

MIGR1 plasmid was digested with BglII and XhoI restriction enzymes and the *Phf6* fragment, digested from pDrive plasmid, was cloned into the digested MIGR1. Abbreviations: LTR, Long terminal repeat; IRES, Internal ribosome entry site; GFP, Green fluorescent protein.

## 2.3 shRNA mediated knockdown

### 2.3.1 Cloning of shRNA expression vectors

The sequences of shRNA (Table 2.4) targeting the mouse *Phf6* transcript were designed using ([http://cancan.cshl.edu/RNAi\\_central/RNAi.cgi?type=shRNA](http://cancan.cshl.edu/RNAi_central/RNAi.cgi?type=shRNA)) and purchased from Sigma-Aldrich. Each shRNA oligonucleotide was amplified by PCR using pSM2C universal primers (Table 2.5) that contain XhoI and EcoRI sites to facilitate cloning into pMSCV backbones. The 50 µl PCR reaction was composed of 1x Reddy Mix PCR Master Mix (ThermoFisher), 0.4 µM shRNA oligonucleotide, 1 µM of each pSM2C forward and reverse primers and performed in a Tprofessional TRIO thermocycler.

**Table 2.4 shRNA oligonucleotides.**

shRNA Name	Sequence
<b>shPhf6-1</b>	TGCTGTTGACAGTGAGCGAACTGTGCATTGCATGATAAAGTAGTGAAGCCACA GATGTACTTTATCATGCAATGCACAGTGTGCCTACTGCCTCGGA
<b>shPhf6-2</b>	TGCTGTTGACAGTGAGCGAACAAGGAATGTGGACAGTTACTAGTGAACCACAG ATGTAGTAACTGTCCACATTCCTTGTCTGCCTACTGCCTCGGA
<b>shPhf6-3</b>	TGCTGTTGACAGTGAGCGCCACATCCTCCCATGGAACAGTAGTGAAGCCACAG ATGTACTGTTCCATGGGAGGATGTGGTGCCTACTGCCTCGGA
<b>shPhf6-4</b>	TGCTGTTGACAGTGAGCGCACTCGGAAGCTGATTTAGAAGTAGTGAAGCCACAG ATGTACTTCTAAATCAGCTTCCGAGTTTGCCTACTGCCTCGGA
<b>shPhf6-5</b>	TGCTGTTGACAGTGAGCGCGGACAGTTACTGATATCTGAATAGTGAAGCCACAG ATGTATTCAGATATCAGTAACTGTCCATGCCTACTGCCTCGGA
<b>shPhf6-6</b>	TGCTGTTGACAGTGAGCGCAGAGGGAAATTGCATATATTTTAGTGAAGCCACAG ATGTAAAATATATGCAATTTCCCTCTTTGCCTACTGCCTCGGA

**Table 2.5 pSM2C universal primers.**

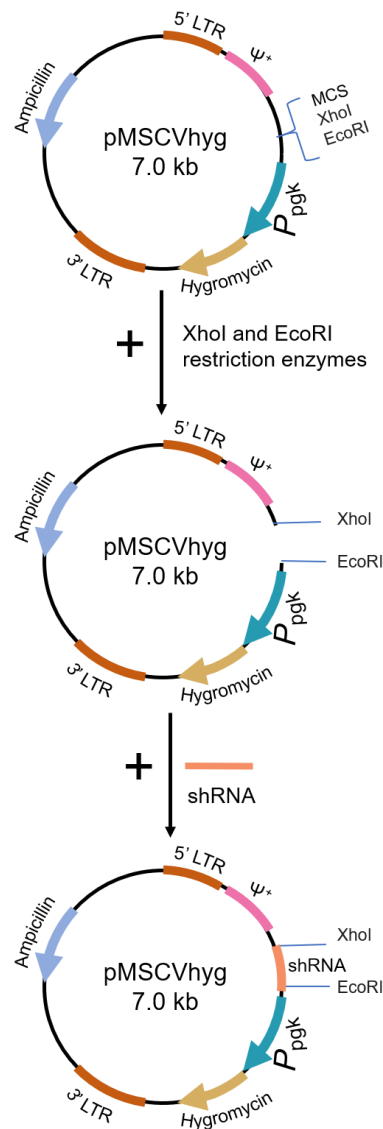
Name	Primer sequence
<b>pSM2C_F</b>	GATGGCTGCTCGAGAAGGTATATTGCTGTTGACAGTGAGCG
<b>pSM2C_R</b>	GTCTAGAGGAATTCCGAGGCAGTAGGCA

Reactions were carried out using the thermocycler parameters described in (Table 2.6). Each shRNA fragment was cloned into pDrive (QIAGEN) and digested with XhoI and EcoRI restriction enzymes (NEB). The shRNA fragment was subsequently cloned into XhoI and EcoRI digested pMSCVhyg plasmid (CloneTech) (Figure 2.3), following the same digestion and cloning protocols described for cloning of cDNA in section 2.2.4.

**Table 2.6 Thermocycling conditions for shRNA amplification by PCR.**

Step	Temperature	Time	Number of cycles
Denaturation	94° C	5 min	X12
Annealing	94° C	30 sec	
Extension	54° C	30 sec	
Elongation	75° C	30 min	
	75° C	2 min	
Pause	4° C	10 min	





**Figure 2.3 shRNA cloning in pMSCVhyg plasmid.**

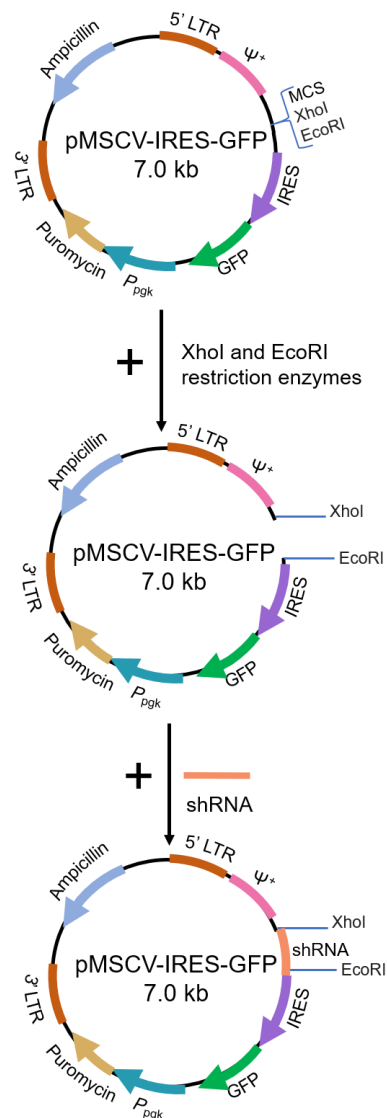
pMSCVhyg plasmid was digested with XhoI and EcoRI restriction enzymes and the shPhf6 fragment, digested from pDrive, was cloned into the digested pMSCVhyg. Abbreviations: LTR, Long terminal repeat; P<sub>pgk</sub>: *Pgk* gene promoter; MCS, multiple cloning site.

### 2.3.2 Retroviral production and transduction of PUER cells

PlatE packaging cells were used to produce ecotropic retrovirus, which can only readily infect mouse or rat cells. PlatE cells were re-plated 24 h prior to transfection, to reach a confluency of 60-80% at time of transfection. On the day of transfection, cells were

provided with fresh DMEM. For each well (9 cm<sup>2</sup>) of a 6 well-plate to be transfected, 2 µg DNA of either shRNA or cDNA expression vectors, 6 µl JetPei (3 µl/µg DNA) transfection reagent (Polyplus-transfection®) and 200 µl DMEM were mixed and incubated for 30 min at RT. The transfection mixture was then added dropwise over the PlatE cells and incubated for 24 h at 37° C. The media was changed the next day, after which virus containing media was collected at 48 h and 72 h and stored at 4° C until used.

In order to test the efficiency of the shRNA constructs, PlatE cells were co-transfected with both *Phf6* MigR1 and pMSCVhyg shPhf6 plasmids, either containing a control sequence or one of the six shPhf6 sequences. The two controls used were: pMSCVhyg shFF3 which targets the firefly luciferase gene (negative control) and pMSCVhyg shGFP which targets the GFP gene (positive control) (Paddison et al., 2004, Dow et al., 2012, Morita et al., 2000). The shRNAs cloned in pMSCVhyg that were co-transfected with *Phf6* MigR1 were examined for successful knockdown of the gene by flow cytometry analysis which will be further described in chapter 4 section 4.2.2. Following this, the shRNA that showed successful knockdown of the gene were subcloned into pMSCV-IRES-GFP vector (Figure 2.4) as described for cloning of cDNA in section 2.2.4.



**Figure 2.4 shRNA cloning in pMSCV-IRES-GFP plasmid.**

pMSCV-IRES-GFP plasmid was digested with XhoI and EcoRI restriction enzymes and the shPhf6 fragment, digested from pMSCVhyg, was cloned into the digested pMSCV-IRES-GFP. Abbreviations: LTR, Long terminal repeat; IRES, Internal ribosome entry site; GFP, Green fluorescent protein;  $P_{pgk}$ : *Pgk* gene promoter; MCS, multiple cloning site.

To infect PUER cells by spin infection, the virus generated by PlatE cells, was pre-coated on wells (9 cm<sup>2</sup>) of 6 well-plate plates by centrifugation at 850 g at RT for 15 min. The virus containing media was recollected from the plate and mixed with  $0.8 \times 10^6$  PUER cells and 12  $\mu$ g/ml polybrene (Sigma-Aldrich). The plate was then centrifuged

at 850 g at 37° C for 2 h. Following spin infection, the plate was returned to the 37° C incubator and the media was changed after 15 min. Cells transduced with pMSCVhyg were selected by addition of 300 µg/ml of Hygromycin B (Invitrogen). GFP expressing cells were isolated by FACS sorting. To generate single cell clones, FACS sorted GFP positive cells were diluted to 0.5 cells/100 µl and spread over a 96-wells plate. Wells containing single cells were identified by following the early growth through a microscope.

## **2.4 Morphological analysis**

For morphological analysis of PUER cells and derivative cell lines,  $10^6$  cells were washed with PBS spun onto glass slides at 700 rpm for 2 min using Cytospin III centrifuge (Shandon). For staining, the cells were fixed in methanol for 30 sec, stained with eosinophilic and basophilic stains 30 and 20 sec respectively (Shandon™ Kwik-Diff™ Stains, ThermoFisher), and excess stain was washed away with H<sub>2</sub>O. The cells were photographed, using a light microscope (Leica DM600) at 100x with oil immersion lenses.

## **2.5 Flow cytometry**

Flow cytometry was performed on  $5 \times 10^5$  cells per analysis. Cells were collected by centrifugation at 300 g at RT for 5 min and washed twice with PBS. The pellet was resuspended in 100 µl PBS with 0.5% FBS and 2 mM EDTA and incubated for 30 min on ice with the relevant antibodies, as indicated in Table 2.7 . Cell suspensions were washed once with PBS prior to analysis with a CyAn™ ADP flow cytometer (DakoCytomation-Beckman Coulter).

**Table 2.7 Flow cytometry antibodies.**

<b>Antibody</b>	<b>Source</b>	<b>Application</b>
<b>Anti-Mouse F4/80 Antigen APC</b>	eBioscience, 17-4801	FACS (2 µg)
<b>Anti-Mouse CD11b eFlour450</b> <b>Antigen</b>	eBioscience, 48-0112	FACS (2 µg)
<b>Anti-Mouse CD117 (c-kit) Antigen</b> <b>PE-Cyanine7</b>	eBioscience, 25-1171	FACS (2 µg)
<b>Rat IgG2a Isotype Control APC</b>	eBioscience, 17-4321	FACS (2 µg)
<b>Rat IgG2a Isotype Control</b> <b>PE-Cyanine7</b>	eBioscience, 25-4321	FACS (2 µg)
<b>Rat IgG2a Isotype Control</b> <b>FITC</b>	eBioscience, 11-4031	FACS (2 µg)
<b>Rat IgG2a Isotype Control</b> <b>eFluor450</b>	eBioscience, 48-4031	FACS (2 µg)

## **2.6 Nuclear extract**

Cells were washed in PBS once, followed by a wash with hypotonic buffer (10 mM HEPES pH7.6, 10 mM KCL, 1.5 mM MgCl<sub>2</sub>). Cells were then resuspended in 1 ml hypotonic buffer per 2x10<sup>7</sup> cells and incubated for 30 min on ice. Nuclei were centrifuged at 300 g at 4° C for 5 min and resuspended in 100 µl per 3x10<sup>7</sup> cells hypertonic buffer (20 mM HEPES pH7.6, 420 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.5% NP40, 20% Glycerol). After 20 min incubation on ice, the supernatant was collected after centrifugation at 21100 g at 4° C for 2 min. The supernatant was diluted with 1.8 volume of no-salt buffer (20 mM HEPES pH7.6, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA,

0.5% NP40, 20% Glycerol) to bring the salt concentration to near physiological concentrations. All buffers were supplemented with 1:1000 Phosphatase Protease Inhibitor (PPI, Roche) and samples were kept on ice throughout. Protein concentrations of the nuclear extracts were determined using a BCA Protein Assay Kit (Pierce Biotechnology).

## **2.7 Immunoprecipitation (IP)**

Protein G coated Dynabeads (30 µg/µl, Life Technologies) were washed with PBS (Sigma-Aldrich) and resuspended in PBS, containing 3% Bovine serum albumin (BSA, Sigma) and 6 µg antibody of interest or IgG (Table 2.8) per 60 µl of beads. After 30 min incubation on a rotator at 4° C, the beads were washed with hypertonic: hypotonic buffer at a ratio of 1:1.8, (hypertonic and hypotonic buffers described in section 2.6), after which they were mixed with 600 µg of nuclear extract and incubated for 2 h at 4° C. This was followed by 2 washes in 1:1.8 hypertonic: hypotonic buffer. For input controls, 100 µg of the nuclear extract was taken. The beads and inputs were boiled at 95° C for 5-10 min with 10x sample reducing agent and 4x Novex loading buffer (both Invitrogen).

## **2.8 Western blotting**

Proteins were resolved by SDS-PAGE on 4-12% Bis-Tris Novex gels (Invitrogen), alongside a PageRuler™ Prestained Protein Ladder (ThermoFisher). Gels were electrophoresed in a vertical tank filled with 1x 2(N-morpholino)-Gethanesulfonic acid 1x MES buffer (Invitrogen) at 200 Volts for 30-40 min, depending on the size of the protein of interest. Proteins separated by SDS-PAGE were transferred onto

polyvinylidene fluoride (PVDF) membrane (Invitrogen) using a dry blotting system (iBlot™, Invitrogen).

Membranes were stained with 0.1% Ponceau S (Sigma) to visualise the total amounts of protein and confirm equal loading, after which they were blocked with 5% milk in PBS for 1 h at RT on a rocker. The blocked membranes were incubated with primary antibody (Table 2.8) overnight in 5 ml PBS on a roller at 4° C. Following three PBS washes of 10 min each, the membranes were incubated for 30 min at RT with secondary antibody (Table 2.9) in 5 ml PBS. After a further three washes in PBS, the membranes were exposed, using Odyssey CLx. Infrared Imaging System (Li-Cor) for visualisation.

## **2.9 Mass Spectrometry (MS)**

Immunoprecipitation was performed as described in section 2.7, using 100 µl of beads with 10 µg of antibody of interest or IgG (Table 2.8), and 1 mg nuclear extract. The IPs were resolved on 4-12% SDS-PAGE gels and stained with Coomassie blue (ThermoFisher) on a rocking platform for 2 h at RT and then de-stained overnight in 1% acetic acid.

Gel lanes were manually excised and subdivided into 12 slices containing the precipitated proteins. Gel slices went through a series of 30 min incubations at RT on a shaker in 300 µl of solutions as follows; solution 1 (50% acetonitrile, 50 mM ammonium bicarbonate), solution 2 (100 mM iodoacetamide in 50% acetonitrile, 50 mM ammonium bicarbonate) in the dark, solution 3 (10% acetonitrile, 50 mM ammonium bicarbonate).

The slices were dried overnight in a vacuum centrifuge. Trypsin (Promega) was reconstituted in 10% acetonitrile, 50 mM ammonium bicarbonate, of which 40 µl was added to each gel slice and incubated for 2 h at RT to digest the proteins into peptides. The resulting peptides were eluted by incubation with 100 µl 1% formic acid in 10% acetonitrile for 1 h, followed by 100 µl 2% formic acid in 60% acetonitrile for 30 min at RT. The two fractions were pooled for each gel slice and peptides were lyophilised in a vacuum centrifuge overnight.

Prior to mass spectrometry, the peptides were resuspended in 1% formic acid and subsequently run on an Impact ESIQTOF machine (Bruker Daltonics) controlled by Otof Control and Hystar software packages. The machine was kindly operated by Dr. Doug Ward, Institute of Cancer and Genomic Sciences, University of Birmingham.

For mass spectrometry analysis, the peptides were searched against a protein database, containing random false positives, and results were sorted according to the highest Mascot score. All proteins of a score <25 were excluded, as this removed the vast majority of false positive hits from the dataset.

## **2.10 Immunostaining**

Cells were harvested and fixed for 10 min at RT in 4% formaldehyde (Sigma), then washed with PBS. They were permeabilised with 0.5% Triton X-100 (Sigma-Aldrich) and blocked with 1% FBS (Life technologies) for 30 min at RT. Fixed cells were incubated with the primary antibody (1 µg/ml) for 1 h (Table 2.8), followed by the secondary antibody (0.6 µg/ml) for 30 min at RT (Table 2.9).

Cells were washed twice with PBS, then centrifuged at 700 rpm for 2 min using Cytospin III centrifuge (Shandon) to prepare glass microscope slides. Nuclei were



counterstained with 4',6-diamidino-2-phenylindole stain (DAPI) (Invitrogen). Confocal images were obtained using a Carl Zeiss LSM780 Meta Confocal microscope (Carl Zeiss, Germany).

**Table 2.8 Primary antibodies.**

<b>Antibody</b>	<b>Source</b>	<b>Applications</b>
<b>PHF6 (rabbit)</b>	Bethyl laboratories, A301-451A	WB (1 µg/ml) IP (6 µg) MS (10 µg) Immunostaining (1 µg) PLA (1 µg)
<b>PHF6 (rabbit)</b>	Abcam ab173304	WB (1 µg/ml) Immunostaining (1 µg)
<b>PHF6 (mouse)</b>	Santa Cruz, sc-365237	WB (1 µg/ml)
<b>TAL1 C-21 (goat)</b>	Santa Cruz, sc-12984	WB (1 µg/ml)
<b>LMO2 (goat)</b>	R&D systems, AF2726	WB (1 µg/ml) IP (6 µg) MS (10 µg) Immunostaining (1 µg)
<b>53BP1(rabbit)</b>	Abcam ab36823	WB (1 µg/ml)
<b>γH2AX (mouse)</b>	Abcam ab26350	WB (1 µg/ml)
<b>Normal Goat IgG</b>	Santa Cruz, sc-2028	IP (6 µg) MS (10 µg)
<b>Normal Rabbit IgG</b>	Santa Cruz, sc-2027	IP (6 µg) MS (10 µg)

**Table 2.9 Secondary antibodies.**

<b>Antibody</b>	<b>Source</b>	<b>Applications</b>
<b>Alexa Fluor® 790 IgG Fraction Monoclonal Mouse Anti-Rabbit IgG, light chain specific</b>	Jackson ImmunoResearch Laboratories, 211-652-171	WB (1:20000)
<b>Alexa Fluor® 790 IgG Fraction Monoclonal Mouse Anti-Goat IgG, light chain specific</b>	Jackson ImmunoResearch Laboratories, 205-652-176	WB (1:20000)
<b>IRDye 800CW donkey anti-mouse</b>	Li-Cor, 926-32212	WB (1:2000)
<b>Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488</b>	ThermoFisher, A11055	Immunostaining (1:1000)
<b>Chicken anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 647</b>	ThermoFisher, A-21463	Immunostaining (1:1000)
<b>Donkey anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 568</b>	ThermoFisher, A10042	Immunostaining (1:1000)
<b>Donkey anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 594</b>	ThermoFisher, A 21207	Immunostaining (1:1000)

## **2.11 Cell proliferation assay**

The cell proliferation assay was performed by EdU incorporation using Click-iT® Plus EdU Alexa Fluor® imaging kit (ThermoFisher). According to the manufacturer's recommendations, the cells were resuspended in fresh media containing 1:1000 EdU (10 µM) and incubated for 2 h at 37° C. Subsequently, the cells were harvested and washed with 1% FBS in PBS. Next, the cells were fixed with 4% formaldehyde, permeabilised with 0.5% Triton X-100, and blocked with 1% FBS as described in section 2.10.

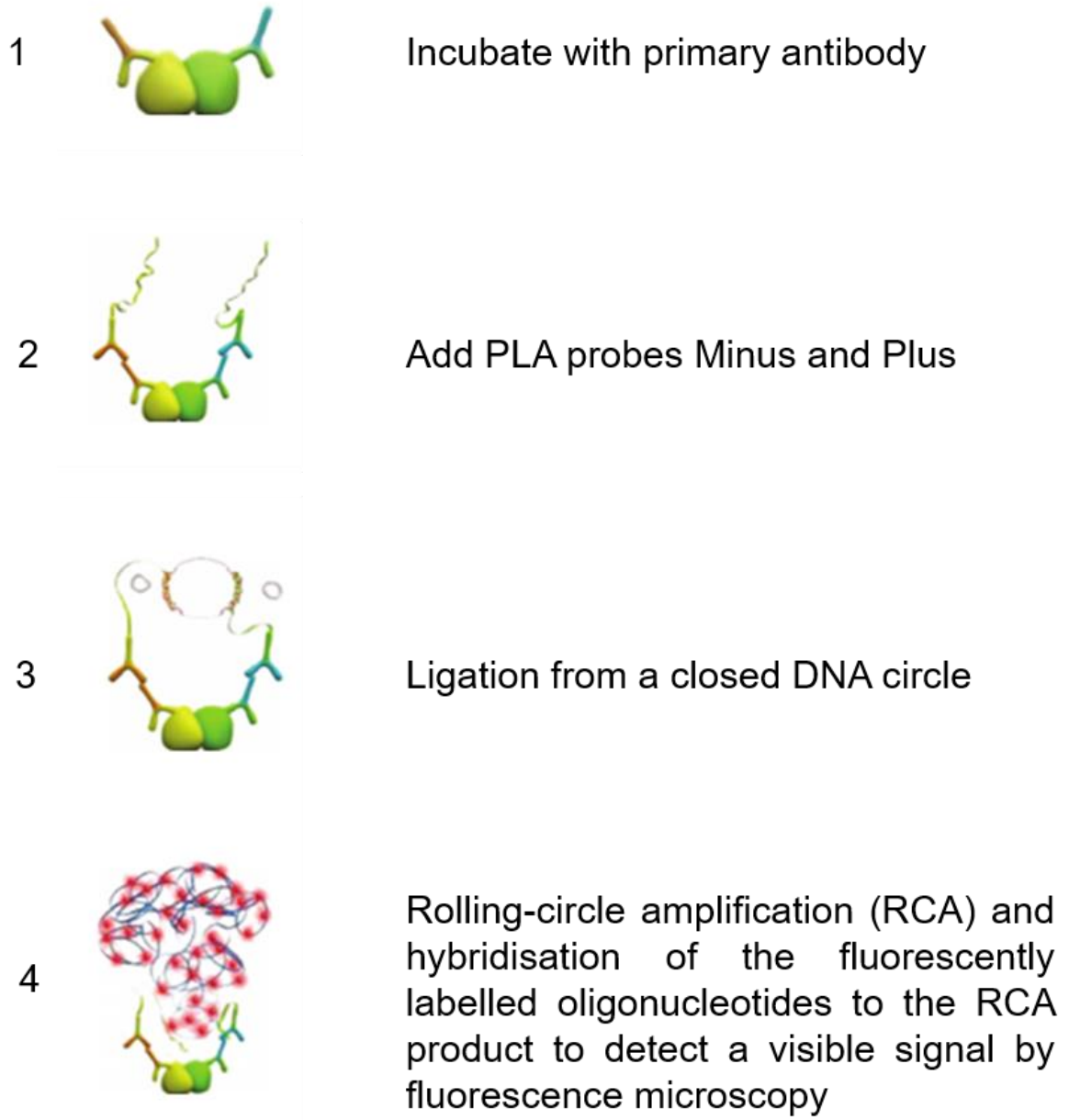
The Click-iT™ Plus reaction cocktail was prepared according to the manufacturer's protocol, and fixed cells were incubated with the reaction cocktail for 30 min at RT, protected from light. The cells were then washed with 1% FBS in PBS and incubated with ethidium bromide (Invitrogen) (1 µg/µl) to detect the DNA content. Cell suspensions were washed once with PBS prior to analysis with a CyAn™ ADP flow cytometer (DakoCytomation-Beckman Coulter).

## **2.12 Proximity ligation assay (PLA)**

PLA was performed in a similar manner as immunostaining in section 2.10 until the primary antibody incubation step. The consecutive steps were performed following the manufacturer's protocol (Duolink® In Situ – Fluorescence by Sigma-Aldrich), as follows and shown in Figure 2.5; Cells were incubated for 1 h at 37° C in antibody diluent containing PLA secondary probes conjugated with oligonucleotides (anti-rabbit PLUS and anti-goat MINUS). In the Ligation and hybridisation step, ligase diluted in a ligation solution consisting of two oligonucleotides was incubated with cells for 30 min at 37° C. The oligonucleotides hybridise to the PLA probes and form a closed circle if they are in close proximity.

Cells were then fixed on slides using a Cytospin III centrifuge (Shandon). In the amplification step, slides were incubated for 2 h at 37° C with amplification solution, consisting of nucleotides and fluorescently labelled oligonucleotides and Polymerase. A rolling-circle amplification (RCA) reaction is generated using the ligated circle as a template, producing a repeated sequence. The fluorescently labelled oligonucleotides hybridise to the RCA product and generate a visible signal as a distinct fluorescent spot. After final washing steps, cells were mounted with DAPI and analysed by

fluorescence microscopy, using a Carl Zeiss LSM780 Meta Confocal microscope (Carl Zeiss, Germany).



**Figure 2.5 Proximity ligation assay (PLA).**

Figure adopted from Duolink® PLA Technology by Sigma-Aldrich.

### 2.13 Cytogenetics

Mitotic cells were arrested at metaphase 24 h after passaging, by treating them with 20 ng/ml Colcemid (KryoMax Colcemid, Gibco) for 15 min-2 h at 37° C. They were then pelleted at 300 g for 8 min and swelled by hypotonic treatment with 0.075 M KCl for 20 min at 37° C. After this, cold Carnoys fixative (3:1 methanol/ glacial acetic acid) was added and they were pelleted at 300 g for 8 min. Cells were fixed for 15-20 min at RT, then washed twice in cold Carnoys fixative.

The fixed cells were resuspended in cold Carnoys solution then dropped onto humidified, chilled glass slides. The mitotic index, quality of metaphase spread, presence of cytoplasm, and overlaps were evaluated under a phase contrast microscope. The slides were aged overnight on a heating block at 60° C. Chromosomes during metaphase spreading were stained by Giemsa banding (G-banding) technique as follows; slides were immersed in 2.5% trypsin working solution at 37° C, then 0.9% NaCl solution at 37° C to stop the digestion. Finally, they were stained in Giemsa (karyoMax, Gibco) at 37° C and washed in H<sub>2</sub>O at RT. The cytogenetics protocol was kindly provided by Dr. Manar Samman and Mr. Khelad AlSaidi from King Fahd medical city, Riyadh, Saudi Arabia.

Chromosomes were visualised and counted under a light microscope (Leica DM600) at 100x using an oil immersion lens. Typically, 20 mitotic cells per cell line from three independent experiments were randomly selected and chromosomes were counted and imaged. Statistical analysis for the chromosome counts were performed using two-way ANOVA test using prism 8 software, version 8.4.1 (460) from GraphPad company.

## **2.14 RNA sequencing (RNA-seq)**

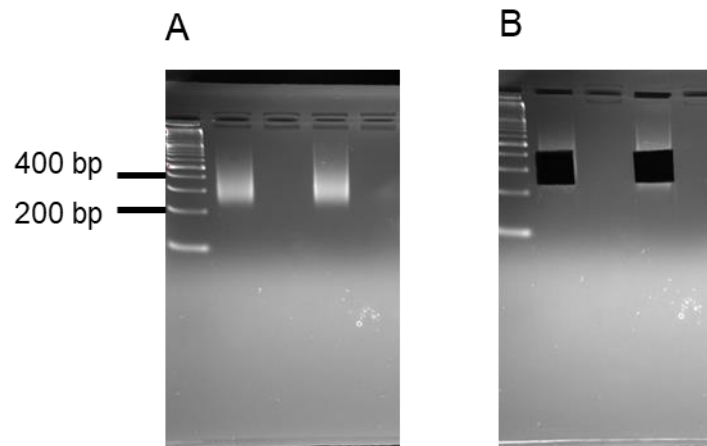
### **2.14.1 RNA-seq library preparation**

RNA isolation was carried out using a NucleoSpin® RNA kit (MACHEREY-NAGEL), according to the manufacturer's protocol as described in RNA isolation section 2.2.1. RNA-seq libraries were generated from 4 µg total RNA per sample, using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina). RNA was mixed with RNA purification beads and mRNA was denatured in a PCR machine at 65° C for 5 min. Beads were washed with Bead washing buffer and mRNA was eluted with Elution buffer in PCR machine at 80° C for 2 min. mRNA was rebound to the beads by Bead binding buffer followed by washing with Bead washing buffer. Beads were then mixed with Fragment Prime finish mix (random hexamers) and mRNA was unbound in the PCR machine at 94° C for 8 min. The supernatant was incubated at 16° C for 1 h with Super script II and First Strand Marking Master Mix.

Double stranded (ds) cDNA was bound to Agencourt AMPure XP beads (Beckman Coulter) for 15 min at RT. The beads were washed twice with 80% ethanol, air dried and resuspended with Resuspension buffer. The ds cDNA was collected and mixed with Resuspension buffer and A-Tailing Mix for end repair in a PCR machine at 37° C for 30 min then at 70° C for 5 min.

Illumina adapters were ligated to the ends of the DNA fragments using Ligation mix for 10 min at 30° C. Ligation was terminated by Stop ligation mix then AMPure® XP beads were added, washed twice with 80% ethanol, air dried and resuspended with Resuspension buffer two times. The cDNA was eluted from the beads and mixed with PCR primer cocktail and PCR primer mix then amplified in 15 enrichment PCR cycles.

Eluted cDNA was size selected between 200-400 bp by gel electrophoresis on 3% agarose gels (Figure 2.6).



**Figure 2.6 Representation of RNA-seq libraries run on 3% gel.**

(A) prior to size selection, (B) After gel excision.

#### **2.14.2 Library quantification by qPCR**

Libraries were quantified by qPCR using a KAPA Library Quantification Kit (Illumina). Following the manufacturer's guidelines, dilution series were made for all libraries and run in parallel with the provided standards (a series of pre-made dilutions of a 452 nt product). A set of three dilutions (1:5000; 1:10,000; 1:15,000) was made for all libraries and compared to six standards (respective concentrations of 20, 2, 0.2, 0.02, 0.002, 0.0002 pM). All samples and standards were run in triplicates and subjected to PCR cycles as indicated in Table 2.10. The manufacturer's recommended standard curve was used to calculate the PCR efficiency and subsequently the molarity of the libraries after correction, and to scale the quantification based on average fragment size. Three independent libraries were prepared for each cell line. A minimum of 30 million 100 nt

paired-end reads were acquired per library. The libraries were run on an Illumina Nextseq 500 sequencer.

**Table 2.10 qPCR conditions.**

Step	Temperature	Time	Number of cycles
Denaturation	95° C	10 min	
Cycling stage	95° C	15 sec	X40
	60° C	1 min	
Melt curve stage	95° C	15 sec	
	60° C	1 min	
	95° C	15 sec	
Pause	4-10° C	10 min	

### 2.14.3 RNA-seq data analysis

RNA-Seq reads were mapped to the hg38 human reference genome or to the mouse genome mm10 (GRCm38), using TopHat 2.0.9 and HISAT2 2.1.0 (Trapnell et al., 2012). Cufflinks 2.0.0 was used to assemble Gene transcripts based on the reference genome with quartile normalisation and effective length correction. Cuffmerge 2.1.1 was used to produce a combined gene transfer format (gtf) file to determine gene expression through Cuffdiff 2.0.1. and Cuffdiff 2.0.1.5.

Gene IDs were selected using gene differential expression and FPKM (fragments per kilobase of transcript per million mapped reads) files. A gene was considered significantly differentially expressed if it had  $p < 0.05$ , longer than 200 bp and FPKM  $> 10$  in at least one of the samples.

To perform Hierarchical clustering, the RNA-seq data was computed by MultiExperimentViewer v4.9.0. according to Pearson correlation with complete linkage.



The number of clusters was conducted by Self Organising Tree Analyses. Gene Ontology (GO) analysis was performed using DAVID 6.8 (Huang et al., 2009) and The Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) on lists of up and down regulated genes. GO terms for biological processes were considered significant with p-value <0.05, and the first 10 categories were presented after filtering out redundant terms.

## **2.15 Chromatin Immunoprecipitation (ChIP)**

### **2.15.1 Chromatin preparation**

Cells were crosslinked with 1% formaldehyde at RT for 12 min, after which crosslinking reactions were quenched by addition of glycine to a concentration of 450 mM. The cells were then centrifuged at 300 g at RT for 5 min and pellets were washed in cold PBS prior to resuspension with ice-cold buffer A (10 mM HEPES pH8, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, 1/1000 PPI, 0.1 mM PMSF). Cells were lysed on a rotating wheel at 4° C for 10 min and by centrifugation at 500 g at 4° C for 5 min. The pellets were resuspended in ice-cold buffer B (10 mM HEPES pH8, 200 mM NaCl, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, 1/1000 PPI, 0.1 mM PMSF) and incubated on a rotating wheel at 4° C for 10 min. Nuclear pellets were resuspended at  $1 \times 10^7$  cells/300  $\mu$ l in ice-cold IP buffer I (25 mM Tris-HCl pH 8, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.25% SDS, 1/1000 PPI, 0.1 mM PMSF).

The 300  $\mu$ l aliquots of chromatin were transferred to 1.5 ml Eppendorf tubes for sonication using a Bioruptor® Pico Sonication System (Diagenode), 25 cycles of sonication pulses and rest intervals of 30 sec each. Sonicated chromatin was cleared by centrifugation at 15000 g at 4° C for 10 min and the collected supernatant was

diluted with 2 volumes of IP buffer II (IP buffer I with 7.5% glycerol but without SDS) and aliquots were frozen at -80° C until usage.

### 2.15.2 ChIP

Prior to ChIP the antibody was bound to Protein G Dynabeads (Life Technologies). Using a magnetic rack, 20 µl of beads were immobilised and washed twice with PBS. The beads were then resuspended in PBS, 0.5% BSA, and antibody of interest (Table 2.11) and incubated at 4° C for 2 h on a rotating wheel. Cross-linked chromatin representing  $1.5 \times 10^7$  cells was added to the antibody conjugated beads and incubated on a rotating wheel at 4° C for 2 h. At this point, 10% of chromatin was taken aside as an input control and 200 µg/ml RNaseA was added. Following 1 h at 37° C, 1 mg/ml proteinase K and 0.15% SDS were added and the samples were left overnight at 65° C to digest proteins and to reverse crosslinking.

**Table 2.11 ChIP antibodies.**

Antibody	Source	Applications
PHF6 (rabbit)	Abcam ab173304	ChIP (2 µg/10 µL of beads)
PHF6 (mouse)	Santa Cruz, sc-365237	ChIP (2 µg/10 µL of beads)
TAL1 C-21 (goat)	Santa Cruz, sc-12984	ChIP (2 µg/10 µL of beads)
LMO2 (goat)	R&D systems, AF2726	ChIP (2 µg/10 µL of beads)
LDB1 (rabbit)	Abcam ab96799	ChIP (2 µg/10 µL of beads)
Normal Goat IgG	Santa Cruz, sc-2028	ChIP (2 µg/10 µL of beads)
Normal Rabbit IgG	Santa Cruz, sc-2027	ChIP (2 µg/10 µL of beads)
Normal Mouse IgG	Santa Cruz, sc-2025	ChIP (2 µg/10 µL of beads)

ChIP reactions were washed on a magnetic rack using 700 µl volumes of four buffers in the following order: once with wash buffer I (20 mM Tris-HCl pH 8, 2 mM EDTA, 1%

Triton X-100, 0.1% SDS, 150 mM NaCl), twice with wash buffer II (20 mM Tris-HCl pH 8, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 500 mM NaCl), once with wash buffer III (20 mM Tris-HCl pH 8, 2 mM EDTA, 250 mM LiCl, 0.5% NP40, 0.5% Na deoxycholate), and finally twice with TE/NaCl (10 mM Tris-HCl pH 8, 1 mM EDTA, 50 mM NaCl). A 10 min incubation at 4° C on a rotating wheel was carried out between the washing steps. Beads were then resuspended in elution buffer (0.1 M NaCo<sub>3</sub>, 1% SDS, 5 M NaCl, 5 mg/ml proteinase K).

ChIP samples were then incubated at 65° C overnight to digest proteins and to reverse crosslinking. Following overnight incubation, ChIP supernatants were collected from the beads by magnetic separation. ChIP and input DNA were purified, using 1.8x volumes of Agencourt AMPure XP beads (Beckman Coulter). Two washes were performed with 80% ethanol. The beads were then air dried and DNA was eluted in 0.2x TE buffer.

### **2.15.3 ChIP sample validation by qPCR**

The efficiency of ChIP experiments was validated by qPCR using specific primers against positive and negative control regions (Table 2.12). Samples were validated by qPCR in 10 µl reactions using 1 µM of each forward and reverse primers (Table 2.12) and SYBR® Green master mix (Life Technologies), according to the manufacturer's protocol and subjected to PCR cycles indicated in Table 2.10. Analysis was performed on an ABI 7500 real-time PCR system using StepOne™ Plus software. The data was normalised to input DNA and measured against a standard curve of 5-fold serially diluted genomic DNA. The fold-enrichment was calculated from the relative quantities against negative control regions.

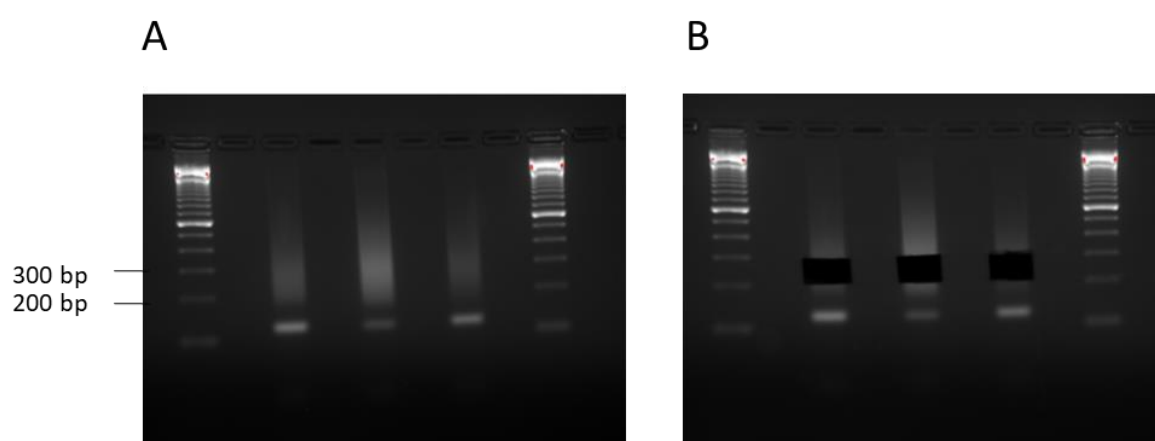
**Table 2.12 ChIP qPCR primers.**

Human	Forward	Reverse
<b>Chr.18</b>	ACTCCCCTTTCATGCTTCTGATATCCATT	AGGTCCCAGGACATATCCATT
<b>SPI1-14kb</b>	AACAGGAAGCGCCCAGTCA	TGTGCGGTGCCTGTGGTAAT
<b>RUNX1+23kb</b>	AACTGCCGGTTTATTTTTTCG	TCTCTGGAAGCCTCTTGAC
<b>RUNX1promoter</b>	GTGCCTGGAAATGAACGTGC	CCGGAAGGCCTGTGATTGG
<b>JAG1</b>	CTCGCGCTCCCCTTCTTTTA	GTCATTGTGTTACCTGCGGC
<b>DNMT3A</b>	TGGAGGGACGGAGAATGAGAT	CAAGACCCCGGTGAAGCAAC
<b>rDNA</b>	AATCCTGCTCAGTACGAGAG	GACAAACCCTTGTGTGAGG
<b>1_rDNA</b>	AGAGGGGCTGCGTTTTCGGCC	CGAGACAGATCCGGCTGGCAG
<b>2_rDNA</b>	GGTATATCTTTCGCTCCGAG	GACGACAGGTCGCCAGAGGA
<b>3-rDNA</b>	CCCTCGCTCGTTTTCTCTCT	AGCGGCTAAGTCTCAAGAGC
<b>4-rDNA</b>	CGGAACTCCCTCTCTCACATT	AGAGAGAGAGAGGGCGAGAG
Mouse	Forward	Reverse
<b>CHR.2</b>	AGGGATGCCCATGCAGTCT	CCTGTCATCAGTCCATTCTCC
<b>TBP</b>	TGCAGTCAAGAGCGCAACTG	CACCGCTACCGGACTCGAT
<b>PU1. 3H</b>	GCTGTTGGCGTTTTGCAAT	GGCCGGTGCCTGAGAAA

#### 2.15.4 ChIP-seq library preparation

Following the qPCR validation, ChIP-seq libraries were produced using the Illumina protocol. Briefly, DNA with damaged ends (due to sonication) was converted to blunt-ended DNA by End-Repair, using Klenow DNA Polymerase, T4 polynucleotide kinase and T4 DNA polymerase at 20° C for 30 min. An A base was added to the 3' of the

blunt-ended DNA fragments to enable adaptor ligation by adding Klenow exo<sup>-</sup> DNA polymerase for 30 min at 37° C. Illumina adapters, containing unique oligonucleotide tag sequences, were ligated to the ends of the DNA fragments, using T4 DNA ligase at 20° C for 30 min. DNA was amplified in 18 PCR cycles with Phusion hot start polymerase (ThermoFisher). After each step DNA was purified using a MinElute PCR purification kit (QIAGEN). Subsequently, DNA was size selected between 200-300 bp by gel electrophoresis on a 3% agarose gel (Figure 2.7). After quantification by a Kappa Library Quantification Kit (Illumina) as described in section 2.14.2 and Bioanalyzer (Agilent) analysis for fragment size distribution, the library was sequenced on a NextSeq 500 (Illumina).



**Figure 2.7 Representation of ChIP-seq libraries run on 3% gel.**

A) prior to size selection. (B) After gel excision.

### 2.15.5 ChIP-seq data analysis

Global analysis of the genome wide data was performed on the University of Birmingham High Performance Computing cluster and usegalaxy.org (Giardine et al., 2005, Blankenberg et al., 2010, Goecks et al., 2010). Reads from ChIP-seq

experiments were aligned to the hg38 revision of the human reference genome or to the mm10 (GRCm38) revision of the mouse genome. Alignment was performed using Bowtie 1.1.2. Duplicates were removed using Picard MarkDuplicates 1.56.0 and peaks across the reference genome were created using MACS 1.0.1 (Zhang et al., 2008) with tag size 28 and band width 200. The resulting files were visualised in the UCSC genome browser.

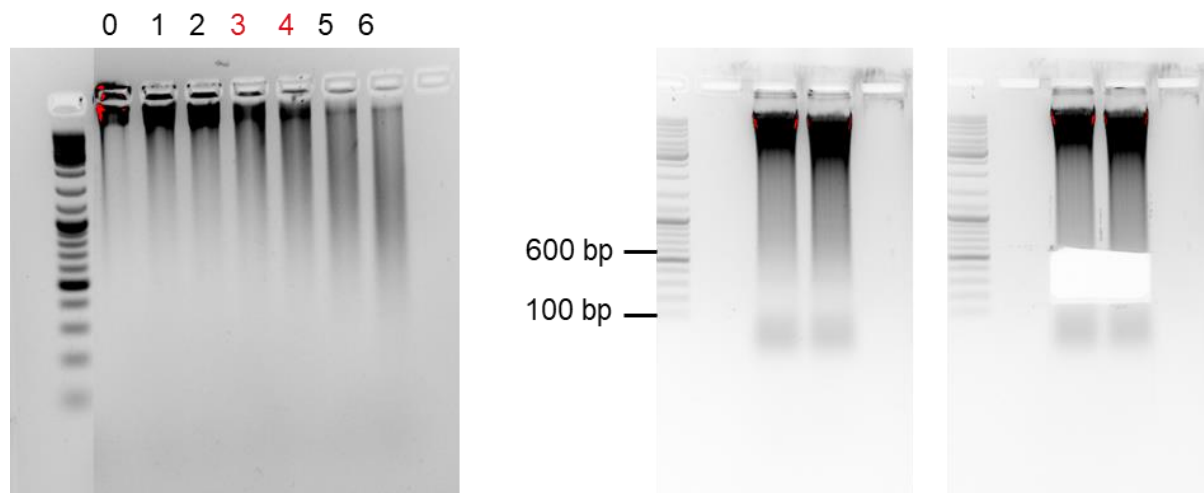
## **2.16 DNaseI hypersensitive site (DHS) mapping**

### **2.16.1 DNaseI-seq assay and library preparation**

Cells were washed with  $\psi$  buffer (11 mM KPO<sub>4</sub> pH 7.4, 108 mM KCl, 22 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 1 mM DTT), and resuspended at 10<sup>8</sup> cells/ml in  $\psi$  buffer containing 1 mM ATP. To every 100  $\mu$ l of this cell suspension, 104  $\mu$ l ice-cold buffer (80  $\mu$ l  $\psi$  buffer containing 1 mM ATP, 4  $\mu$ l 10% NP40 and 20  $\mu$ l H<sub>2</sub>O, that was supplemented with increasing amounts of DNase I (DPFF, Worthington) was added and incubated for exactly 6 min at RT. The reaction was terminated by the addition of 200  $\mu$ l lysis buffer (100 mM Tris pH 8.0, 40 mM EDTA, 2% SDS, 200  $\mu$ g/ml proteinase K).

To facilitate protein digestion by proteinase K, the samples were incubated overnight at 55° C. The samples were treated with 50  $\mu$ g/ml RNase A at 37° C for 1 h, then extracted twice with phenol/chloroform/isoamylalcohol (PCI; 25:24:1) and once with chloroform. The DNA was then precipitated by addition of 0.5 M NaCl and 1 volume of propanol. After a wash with 70% EtOH the resulting DNA pellet was dissolved in 100  $\mu$ l 0.1x TE. After size fractionation of 10  $\mu$ g of suitable samples on a 0.8% agarose gels (Figure 2.8), DNA fragments in the range of 100-600 bp were excised and extracted,

using a QIAGEN mini-elute gel extraction kit, according to the manufacturer's protocol. Library preparation was performed according to the Illumina library preparation protocol as described in ChIP-seq library preparation section 2.15.4. Sequencing was performed on an Illumina 500 Nextseq.



**Figure 2.8 Representation of DNaseI digestion validation gel.**

DNaseI digestions from the DHS assay were run on 0.8% agarose gel to determine the degree of digestion. Samples 3 and 4 (highlighted in red) showed the optimal level of digestion with DNaseI (left gel). Samples 3 and 4 were further run on 0.8% agarose gel and excised (right gel).

### **2.16.2 DHS data analysis**

DHS reads were mapped to the human genome hg38 or mouse genome mm10 (GRCm38) and analysed as described above for ChIP-seq data analysis section 2.15.5.

### **2.17 Data visualisation**

ChIP-seq and DHS heat maps were generated using the EaSeq analysis software (Lerdrup et al., 2016). Venn diagrams were constructed using application Venny

version 2.1 (available from <http://bioinfogp.cnb.csic.es/tools/venny/>). *De novo* motif search including Tomtom and CentriMo was performed online with MEME-ChIP version 4.12.0 (available from <http://memesuite.org/tools/meme-chip>) (Bailey et al., 2009). Association graphs of the input genomic regions, bound by a protein of interest, with their putative target genes were conducted using The Genomic Regions Enrichment of Annotations Tool (GREAT) version 4.0.4 (available from <http://great.stanford.edu/public/html/>) (McLean et al., 2010).



### 3 Chapter 3: PHF6 as a new interactor of LMO2 in T-ALL

#### 3.1 Introduction

T-cell acute lymphoblastic leukaemia (T-ALL) is an aggressive type of haematological cancer arising from malignant transformation of thymocytes and occurs more frequently in males than females (Pui et al., 2012, Pui et al., 2014). T-ALL is a genetically heterogeneous malignancy in which multiple oncogenes and tumour suppressors cooperate to alter the normal T-cell progenitor development (Pui et al., 2008, Ferrando et al., 2002).

The pathogenesis of T-ALL is acquired by molecular alterations in addition to the cytogenetic anomalies which cause deregulation in specific cellular processes, including cell cycle signalling, cell growth and proliferation, chromatin remodelling, T-cell differentiation, and self-renewal. For example, alterations of haematopoietic transcription factor genes, including *LMO2*, *TAL1*, *GATA3*, and *RUNX1* were noted in T-ALL cases (Zhang et al., 2012a, Van Vlierberghe et al., 2008). It is known that the LMO2 complex regulates hematopoietic cell differentiation and is expressed in immature thymocytes, at the DN1 and DN2 stage of their development. Normally, thymocyte maturation through the DN3, DN4 and double-positive (DP) CD4/CD8 stages is associated with a downregulation of *LMO2* and *TAL1*, particularly prior to the DN3 stage (Herblot et al., 2000). The aberrant expression of LMO2 leads to block of T-cell differentiation and is associated with a large subgroup of T-ALL named TAL/LMO (Homminga et al., 2011).

Focusing on LMO2, we wanted to understand the function of LMO2 complex in T-ALL and how it contributes to T-cell oncogenesis. In particular, we aimed to find new interacting partners of LMO2 and explore how these complexes interact with the

chromatin, influence gene regulation, leading to malignancies. To address this, we studied the LMO2 complex in four human T-ALL cell lines by utilising proteomic analysis in combination with genome-wide approaches. These cell lines (ARR, DU.528, HSB2 and CEM) overexpress *LMO2*, and are blocked at different stages of early T-cell differentiation (Figure 3.1) (Sandberg et al., 2007). ARR, DU.528, and HSB2 cells originate from CD3<sup>-</sup> T-ALLs, whereas CCRF-CEM cells are CD3<sup>+</sup>. All the cell lines express the T-cell marker CD7, additionally HSB2 and CCRF-CEM cells express CD5 and cytoplasmic CD3 (CyCD3). CCRF-CEM cell line is the only one that originates from T-ALL with rearranged  $\alpha\beta$  TCR and express additional surface markers such as, CD1a, CD4b, and TdT (Figure 3.1). Immunophenotypical data of these cell lines were collected from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ; [www.dsmz.de](http://www.dsmz.de)) GmbH (Braunschweig, Germany) supplemented with results from Y Sandberg *et al.* (Sandberg et al., 2007). These cell lines also overexpress *TAL1*, which for three of them (DU.528, HSB2 and CCRF-CEM) is associated with an interstitial deletion (1p32). This alteration brings the complete *TAL1* ORF under the control of the ubiquitously expressed *SIL* gene which results in the *SIL/TAL1* gene that causes the continuous expression of *TAL1* (Cave et al., 2004, Janssen et al., 1993).

Cell line	T-cell malignancy of origin	CD7	CD5	CD1 <sup>a</sup>	CyCD3	CD3	TCR	CD4 <sup>b</sup>	CD8	TdT	CD2
ARR	CD3- T-ALL	+	–	–	ND	–	–	–	–	ND	–
DU.528	CD3- T-ALL	+	ND	–	ND	–	–	–	–	–	–
H-SB2	CD3- T-ALL	+	+	–	+	–	–	–	–	–	–
CCRF-CEM	TCR $\alpha\beta$ + T-ALL	+	+	+	+	+	$\alpha\beta$	+	–	+	–

**Figure 3.1 Immunophenotype of T-ALL cell lines.**

Table showing T-ALL cell lines; names, origin, and surface markers expression. Abbreviations: CD, Cluster of differentiation; TCR, T-cell receptor; TdT, Terminal deoxynucleotidyl transferase; ND, Not done. Different densities of grey colour are used to emphasise the expression pattern of the tested markers. Adapted from (Sandberg et al., 2007).

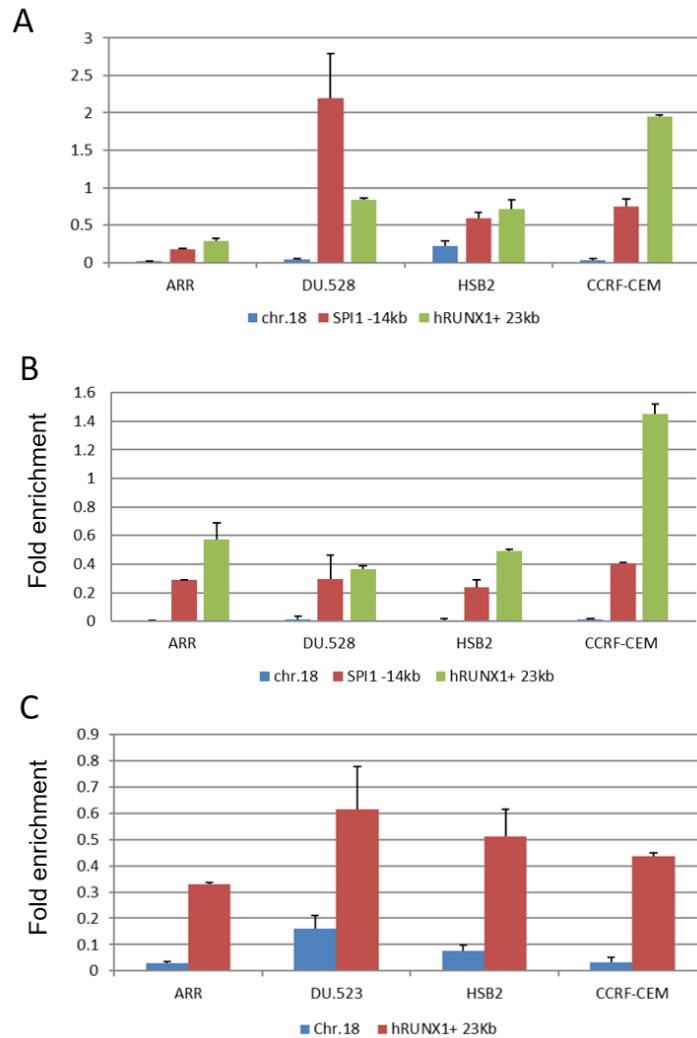
## 3.2 Results

### 3.2.1 LMO2, TAL1 and LDB1 bind regulatory elements

Given that the T-ALL cell lines we used abnormally express *LMO2* and *TAL1*, we started by performing ChIP experiments to determine the LMO2 complex binding sites and possible target genes, using antibodies recognising components of the LMO2 complex, such as LMO2, TAL1 and LDB1 (Wadman et al., 1997, Grutz et al., 1998). TAL1 antibody recognises the wild type TAL1 protein in all cell lines, as the *SIL/TAL1* translocation does not affect the TAL1 amino acid sequence.

ChIP experiments followed by quantitative PCR (qPCR) showed enrichment of LMO2, TAL1 and LDB1 in all four T-ALL cell lines at regulatory elements of haematopoietic transcription factor genes *SPI1* and *RUNX1*, as compared to the negative control amplicon, which is a heterochromatic region on chromosome 18 without any nearby genes (Figure 3.2). *SPI1* and *RUNX1* both encode transcription factors that are critical for haematopoiesis and the development of HSC and have roles in blood lineage determination during differentiation (Wilson et al., 2010). PU.1 (*SPI1*) is observed at

the early stages of T-cell differentiation, whereas RUNX1 persists throughout T-cell development (Yui and Rothenberg, 2014).

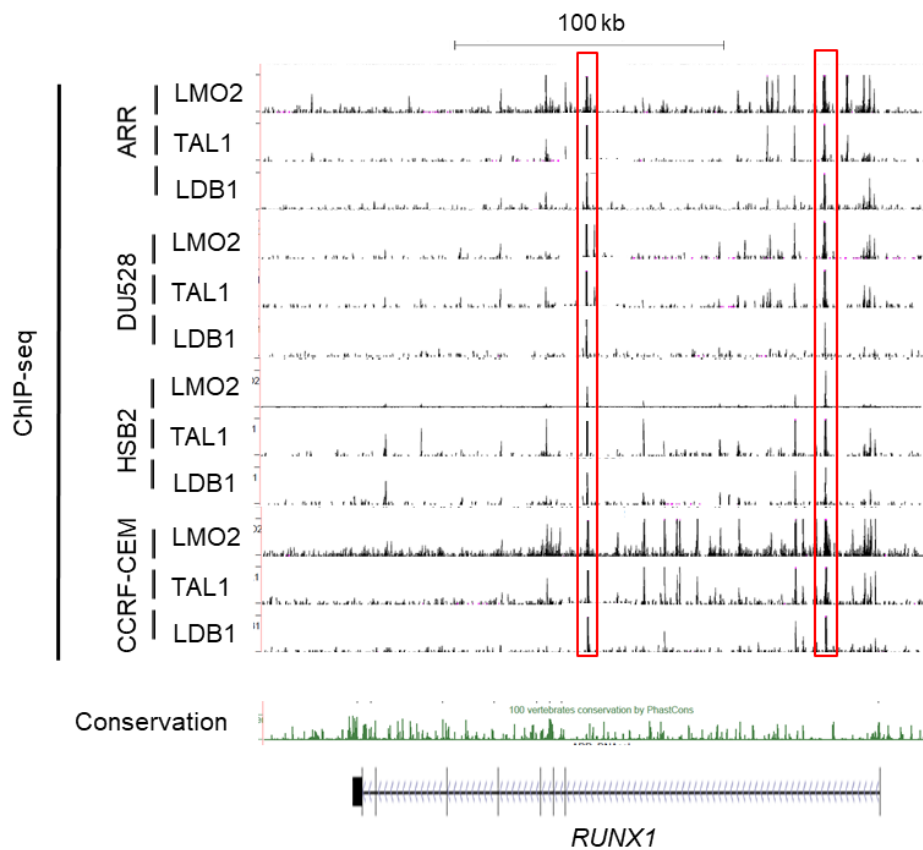


**Figure 3.2 ChIP qPCR analyses of LMO2, TAL1 and LDB1 binding in T-ALL cells.**

Bar graphs are showing (A) LMO2 (B) TAL1 (C) LDB1 binding to different DNA regions as the fold enrichment over the input control levels. Error bars represent SEM (n=2).

On basis of the qPCR results we prepared genome-wide libraries for these ChIP samples followed by sequencing. ChIP sequencing (ChIP-seq) analyses were successful and resulted in peaks for all samples. These ChIP-seq analyses revealed

19277, 13586 and 16254 LMO2, TAL1 and LDB1 peaks, respectively in ARR. In DU.528 more peaks were detected comprising of 26087 for LMO2, 24720 for TAL1 and 12042 for LDB1. The HSB2 cell line showed 29073 of LMO2, 24860 of TAL1 and 10668 of LDB1 peaks. CCRF-CEM cells had the highest number of peaks which were 41647 for LMO2, 28317 for TAL1 and 12844 for LDB1. The UCSC genome browser showed the overlap of LMO2, TAL1 and LDB1 peaks in ARR, DU.528, HSB2 and CCRF-CEM. One example of a target element where all three factors bind together is the enhancer of the *RUNX1* gene (Figure 3.3), in line with the qPCR results shown in Figure 3.2.



**Figure 3.3 Binding profiles of LMO2, TAL1 and LDB1 in T-ALL cells.**







Screenshot from the UCSC browser showing LMO2, TAL1 and LDB1 ChIP-seq binding profiles of ARR, DU.528, HSB2, and CCRF-CEM at *RUNX1* gene.

### 3.2.2 The LMO2/TAL1 complex binds TFs motifs in SIL/TAL1 and ARR cells

Three of the cell lines we utilised in our experiments overexpress *TAL1* as a result of an interstitial deletion (1p32), which results the *SIL/TAL1* gene (Cave et al., 2004, Janssen et al., 1993). DU.528, HSB2 and CCRF-CEM cells exhibit *SIL/TAL1* gene making *TAL1* constitutively active, unlike ARR cells which overexpress *TAL1* without having *SIL/TAL1* mutation. To understand how TAL1 cooperates with LMO2 and whether *SIL/TAL1* mutation leads to different binding patterns, we first looked at the ChIP-seq data and binding profiles of the SIL/TAL1 cell lines. The intersection of the data sets showed that the two way overlap between LMO2 and TAL1 peaks was 17061 for DU.528, 18465 for HSB2 and 23692 for CCRF-CEM peaks. This means approximately 70-80% of TAL1 peaks overlap with LMO2. However, 55-65% of LMO2 binding sites are co-occupied by TAL1 (Figure 3.4). This may indicate that LMO2 forms complexes with other interacting partners that do not involve TAL1, or it can be partially due to peaks that do or do not reach the threshold to be called a peak. To identify the enrichment of sequence motifs, MEME suite online tool (<http://meme-suite.org/>) was used to analyse the ChIP-seq FASTA format sequences of LMO2/TAL1 peaks (Bailey et al., 2009). These *de novo* motif analyses revealed strong enrichment for ETS, GATA, and RUNX motifs and to a lesser extent E-box containing motifs (e.g. TAL1) in the LMO2/TAL1 common peaks in all SIL/TAL1 cells (Figure 3.4). The RUNX1 and ETS family of TFs often co-occupy the same regulatory elements and co-ordinately regulate gene expression with the LMO2/TAL1 complex (Sanda et al., 2012, Tan et al., 2019).






## DU.528



Motif	Known or similar motifs	E-value
	RUNX	6.8e-080
	GATA	9.2e-076
	Ets	1.5e-055
	Unknown	5.4e-004
	Unknown	6.5e-003
	Atoh1 TAL1::TCF3 ZNF238	9.1e-003





## HSB2



Motif	Known or similar motifs	E-value
	Ets	2.5e-076
	GATA	1.5e-066
	RUNX	2.1e-054
	Mtf1 FOXp1 Srf	1.4e-007
	ZNF328 TAL1::TCF3	6.4e-004

## CCRF-CEM

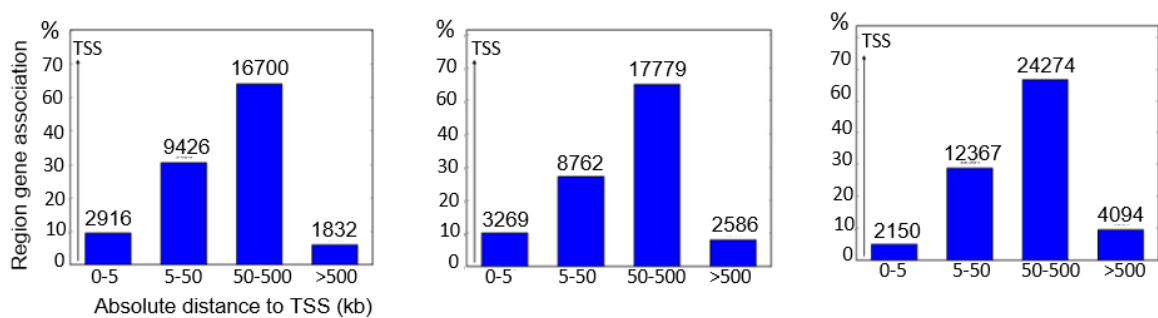


Motif	Known or similar motifs	E-value
	GATA	1.0e-062
	RUNX	1.6e-039
	Ets	8.5e-016
	MESP1 Tcf3 Tcf4	1.2e-014

**Figure 3.4 ChIP-seq analyses of LMO2 and TAL1 in SIL/TAL1 cells.**

Venn diagrams showing the intersection of LMO2 and TAL1 ChIP-seq data (left side). *De novo* motif analysis tables showing enriched motifs for LMO2/TAL1 overlap (right side), the results are presented as sequence of the motif, name of the motif, and the E-value.

To investigate the locations of these peaks and their distance from transcription start sites (TSSs), analyses were performed using The Genomic Regions Enrichment of Annotations Tool (GREAT) which associates the input genomic regions, bound by a protein of interest, with their putative target genes (McLean et al., 2010). The analyses were performed by submitting the ChIP-seq bed files to (<http://great.stanford.edu/public/html/>) online tool. GREAT analyses in these three cell lines revealed that between 5-10% of the LMO2/TAL1 peaks were binding within 5 kb of TSSs, 20-30% were 5-50 kb away, 50-60% were 50-500 kb farther and approximately 5% were >500 kb distant (Figure 3.5).



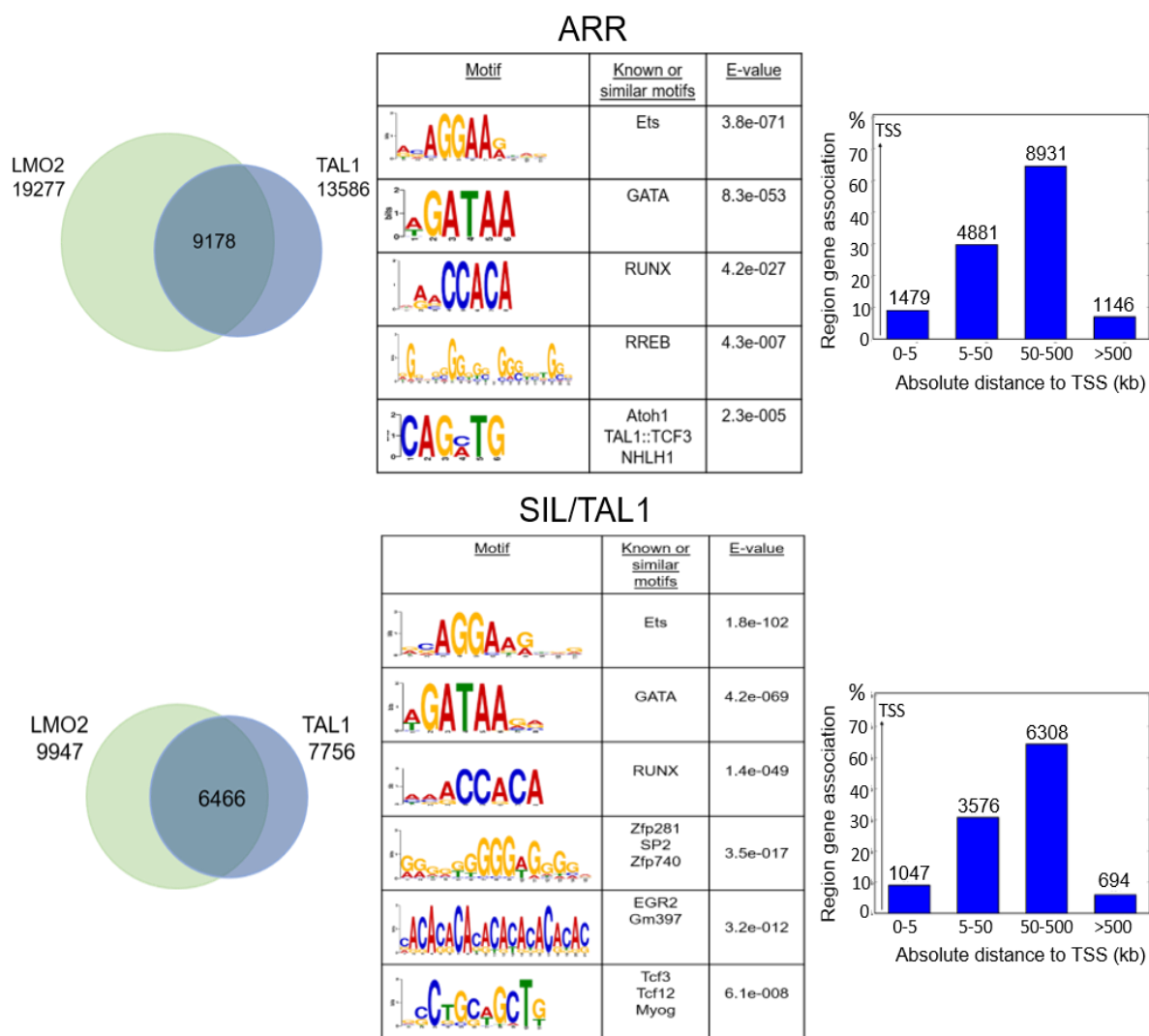
**Figure 3.5 GREAT Region-Gene Association graphs of genomic regions bound by LMO2/TAL1 shared peaks.**

The y-axis is given in percentages, the absolute number of genes being counted is listed above each bar in the graph.

To explore whether the SIL/TAL1 mutation affects site-specific binding of the LMO2/TAL1 complex we combined the ChIP-seq data of the three SIL/TAL1 positive cell lines and compared their common LMO2 and TAL1 binding sites with the ARR cell line. Data analyses of these ChIP-seq experiments revealed that 48% and 65% of LMO2 target sites are co-occupied by TAL1 in ARR and SIL/TAL1 cells respectively (Figure 3.6). Notably, the number of LMO2 peaks shared with TAL1 is higher in SIL/TAL1 cells compared to ARR cells.



*De novo* motif analyses of LMO2/TAL1 complex binding sites revealed that the ETS motif was the most significantly overrepresented in addition to GATA, RUNX and E-box for all cell lines. The SIL/TAL1 cells also showed enrichment for motifs similar to SP2 and EGR2 (Figure 3.6). The association between LMO2/TAL1 peaks with their putative target genes showed a similar peak distribution in ARR and SIL/TAL1, where the majority of peaks were located 50-500 kb from TSSs (Figure 3.6).



**Figure 3.6 ChIP-seq analyses of LMO2 and TAL1 comparing ARR and SIL/TAL1 cells.**

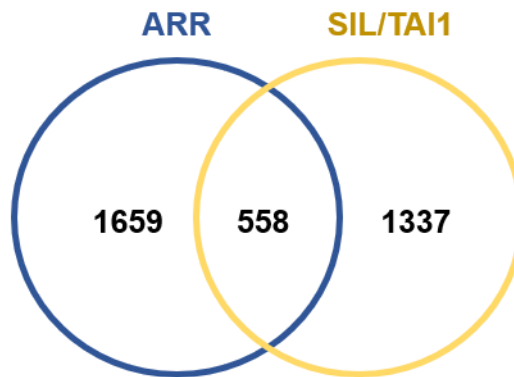
Venn diagrams showing the intersection of LMO2 and TAL1 ChIP-seq data (left side). *De novo* motif analysis tables showing enriched motifs for LMO2/TAL1 overlap (middle), the results are presented as sequence of the motif, name of the motif, and the E-value. GREAT Region-Gene Association graphs of genomic regions bound by LMO2/TAL1 shared peaks (right side), The y-axis is given in percentages, the absolute number of genes being counted is listed above each bar in the graph.

### **3.2.3 LMO2/TAL1 complex targets distinct genes in different cell lines**

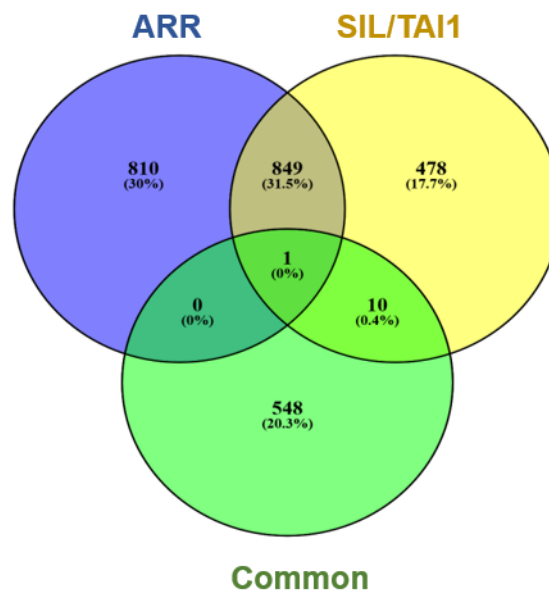
To assess how LMO2 and TAL1 binding sites associate with lower or higher levels of gene expression the above ChIP-seq data was integrated with RNA-seq expression data, which the Hoogenkamp lab had already obtained from these cell lines. For the RNA-seq data the gene IDs were selected using differential gene expression and sorted according to their FPKM values. The genes that had P value < 0.05 and longer than 200 bp with FPKM >10 in at least one of the samples were considered significantly differentially expressed.

Peak coordinates of joint LMO2/TAL1 binding sites that identified the nearest genes up and down stream, combined with the gene expression level of these genes in the respective cell lines, demonstrated 2217 differentially expressed genes associated with LMO2/TAL1 peaks in ARR and 1895 in SIL/TAL1 cells. To distinguish between shared and unique peaks between the cell lines, the peak coordinates of ARR and SIL/TAL1 cells were compared. The number of unique peaks was 1659 in ARR and 1337 in SIL/TAL1, while 558 peaks were common between them (Figure 3.7A). To explore whether the unique and common peaks associated with the same genes, we selected the genes that are linked to ARR peaks, SIL/TAL1 peaks and common peaks then overlapped them. Interestingly, 849 genes were associated with the LMO2/TAL1 complex in all cells, however, the complex bound distinct genomic regions in different cell lines. There were 810 genes exclusively linked to unique peaks found only in ARR and 478 to unique peaks found in SIL/TAL1 cells. The common 558 genes were associated with the common peaks between ARR and SIL/TAL1 cells (Figure 3.7B).

A



B



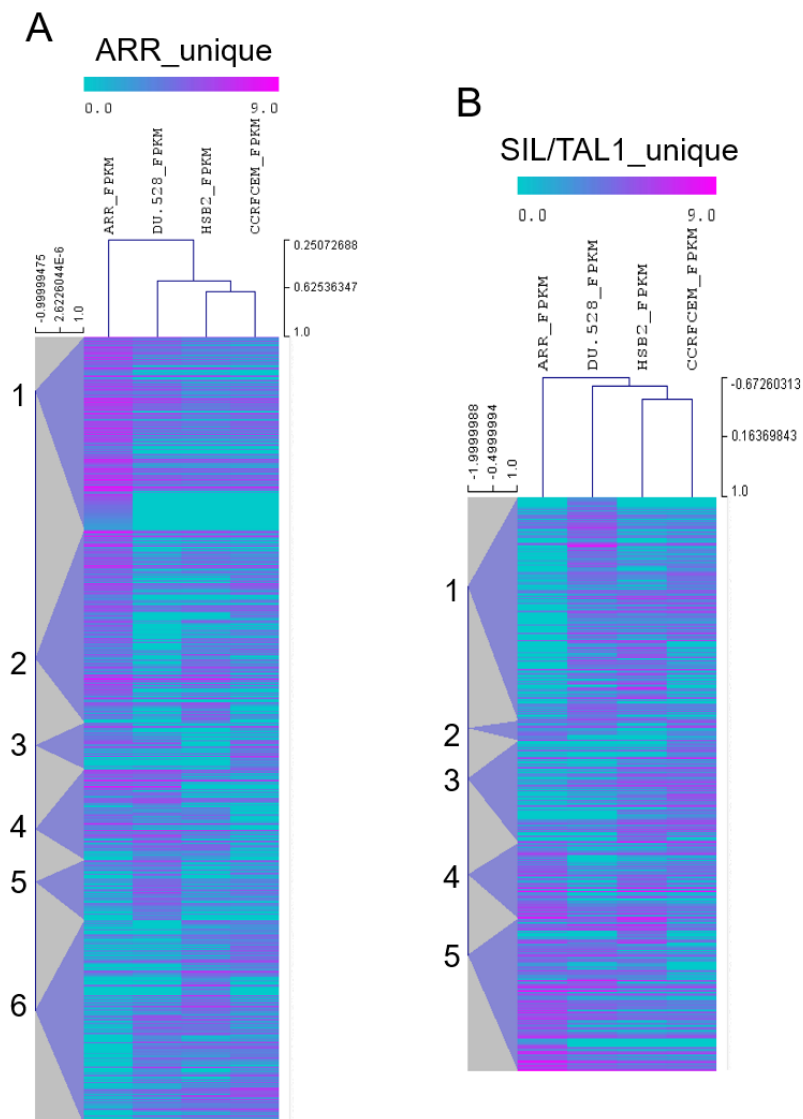
**Figure 3.7 Intersection of LMO2/TAL1 peaks and differentially expressed genes that are potential targets of the LMO2/TAL1 complex in ARR and SIL/TAL1 cells.**

Venn diagrams showing (A) overlap between LMO2/TAL1 peaks in ARR and SIL/TAL1 cells. (B) Three way overlap between genes associated with ARR unique peaks (1659), SIL/TAL1 unique peaks (1337) and common peaks (558).

### **3.2.4 LMO2/TAL1 binding sites are associated with differentially expressed genes in T-ALLs**

To investigate whether the genes that correlate to LMO2/TAL1 have variable levels of gene expression between the cell lines, clustering of differentially expressed genes was carried out on log2 FPKM values. Initially, peak coordinates of joint LMO2/TAL1 binding sites that identified the nearest genes up and down stream per cell line, together with gene expression files were combined as described in section 3.2.3 and matrices were generated. The matrices were analysed using MultiExperimentViewer v4.9.0. software where pairwise Pearson correlation with complete linkage was calculated between the log2 FPKM values to produce hierarchical clustering.

The self-organising tree clustering produced six clusters for the 810 genes associated with unique peaks only found in ARR (Figure 3.8A). Comparably, the hierarchical clustering of the peaks only found in SIL/TAL1 that were associated with 478 unique genes generated five clusters (Figure 3.8B). The clusters that showed distinctive ARR gene expression compared to the other cell lines were chosen to be analysed to explore the biological processes related to these genes. Gene ontology (GO) analyses of biological processes based on the genomic coordinates were performed using DAVID 6.8 (Remke et al., 2009). GO terms for biological processes were ordered according to their Modified Fisher Exact P-value and only terms with  $P < 0.05$  were considered significant, and the first 10 categories were presented after filtering out redundant terms.



**Figure 3.8 RNA-seq analysis of differentially expressed genes associated with LMO2/TAL1 unique peaks in ARR and SIL/TAL1 cell lines.**

Heat maps showing hierarchical clustering of differentially expressed genes comparing (A) ARR to (B) SIL/TAL1 cell lines. These genes are either the nearest 5' or 3' gene or contain the peak within the gene body. Scale bars represent colour index for the log2 FPKM values. Self-organising tree analyses identified 6 clusters in ARR, and 5 clusters in SIL/TAL1 cells.

In Figure 3.8A, cluster 1 showed genes with higher expression levels and cluster 6 showed genes with lower expression levels in ARR compared to the other cell lines. GO terms revealed involvement of the genes in cluster 1 with protein folding, RNA

phosphodiester bond hydrolysis, rRNA processes, translation, viral transcription, nuclear-transcribed mRNA catabolic process, glycogen biosynthetic process, and cation transport (Figure 3.9A). However, the genes in cluster 6 were associated with regulation of Rho protein signal transduction, protein K48-linked de-ubiquitination, inositol biosynthetic process, blood vessel development and nucleosome assembly (Figure 3.9B).

On the other hand, the genes with higher expression levels in SIL/TAL1 cells in cluster 1 (Figure 3.8B) were involved in fatty acid beta-oxidation using acyl-CoA oxidase, lysosome organisation and granulocyte chemotaxis (Figure 3.10A). And the genes with lower expression levels in cluster 5 (Figure 3.8B) were linked to regulation of endosome size, carbohydrate metabolic processes, cellular response to amino acid stimulus, and mRNA 5' splice site recognition (Figure 3.10B).

A

GO term BP	%	P-value	Genes
Protein folding	4.572	0.0027	<i>ATF6, ERP44, CCT4, CWC27, PDIA6, HSP90AA2P, CANX, CLPX</i>
RNA phosphodiester bond hydrolysis, exonucleolytic	2.286	0.0030	<i>PAN2, REXO2, CNOT2, ISG20</i>
righting reflex	1.715	0.0037	<i>AUTS2, PCDH15, FOXP2</i>
rRNA processing	4.572	0.0070	<i>RPL27A, WDR12, RPS20, RPS4X, RPS6, UBA52, TEX10, ISG20</i>
anterior/posterior pattern specification	2.858	0.0096	<i>HOXC10, HOXC6, HOXC9, HOXC4, HOXC5</i>
translation	4.572	0.0163	<i>MRPS9, SLC25A5, RPL27A, SLC25A3, RPS20, RPS4X, RPS6, UBA52</i>
viral transcription	2.858	0.0292	<i>RPL27A, RPS20, RPS4X, RPS6, UBA52</i>
glycogen biosynthetic process	1.715	0.0295	<i>PTGES3, GBE1, UBA52</i>
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	2.858	0.0353	<i>RPL27A, RPS20, RPS4X, RPS6, UBA52</i>
cation transport	1.715	0.0384	<i>TRPM3, TMEM63C, ATP13A4</i>

B

GO term BP	%	P-value	Genes
regulation of Rho protein signal transduction	2.674	0.0125	<i>MCF2L2, PLEKHG4, ARHGEF18, ARHGEF17, ARHGEF40</i>
tissue regeneration	1.605	0.0184	<i>TMEM110-MUSTN1, CPQ, MUSTN1</i>
labyrinthine layer blood vessel development	1.605	0.0178	<i>HES1, HS6ST1, VASH2</i>
protein K48-linked deubiquitination	1.605	0.0287	<i>OTUB1, OTUD7B, USP33</i>
nucleosome assembly	2.674	0.0434	<i>HIST1H2BF, HIST1H1B, HIST1H2BH, KAT6B, HIST1H3I</i>
inositol biosynthetic process	1.07	0.0438	<i>ISYNA1, IMPA2</i>
positive regulation of t-circle formation	1.07	0.0441	<i>SLX4, ERCC1</i>
midbrain-hindbrain boundary morphogenesis	1.07	0.0436	<i>HES1, KDM2B</i>

**Figure 3.9 GO analysis on differentially expressed genes associated with LMO2/TAL1 unique peaks in ARR cells show genes involved in RNA processes and nucleosome assembly.**

GO analyses for biological processes performed on (A) cluster 1, (B) cluster 6 as identified in Figure 3.8A. Table of GO term biological processes, percentage of genes, P-value and genes.



A

GO term BP	%	P-value	Genes
positive regulation of substrate adhesion-dependent cell spreading	3.226	2.00E-04	<i>CDC42, DMTN, ARHGEF7, DOCK5, NET1</i>
fatty acid beta-oxidation using acyl-CoA oxidase	1.936	0.0061	<i>AMACR, HSD17B4, CROT</i>
focal adhesion assembly	1.936	0.0203	<i>ARHGEF7, FERMT2, ITGA2</i>
Golgi organization	2.581	0.0306	<i>CDC42, ARHGEF7, STX17, ATP8B1</i>
ovulation cycle process	1.291	0.0362	<i>PAM, MAP2K6</i>
positive regulation of epithelial cell migration	1.936	0.0367	<i>EPB41L5, ITGA2, DOCK5</i>
negative regulation of epidermal growth factor receptor signalling pathway	1.936	0.0439	<i>CDC42, ARHGEF7, ITGA1</i>
lysosome organization	1.936	0.0435	<i>CLN3, ZKSCAN3, CLN6</i>
granulocyte chemotaxis	1.291	0.0441	<i>ANXA1, IL17RA</i>
floor plate development	1.291	0.0451	<i>COBL, FOXB1</i>

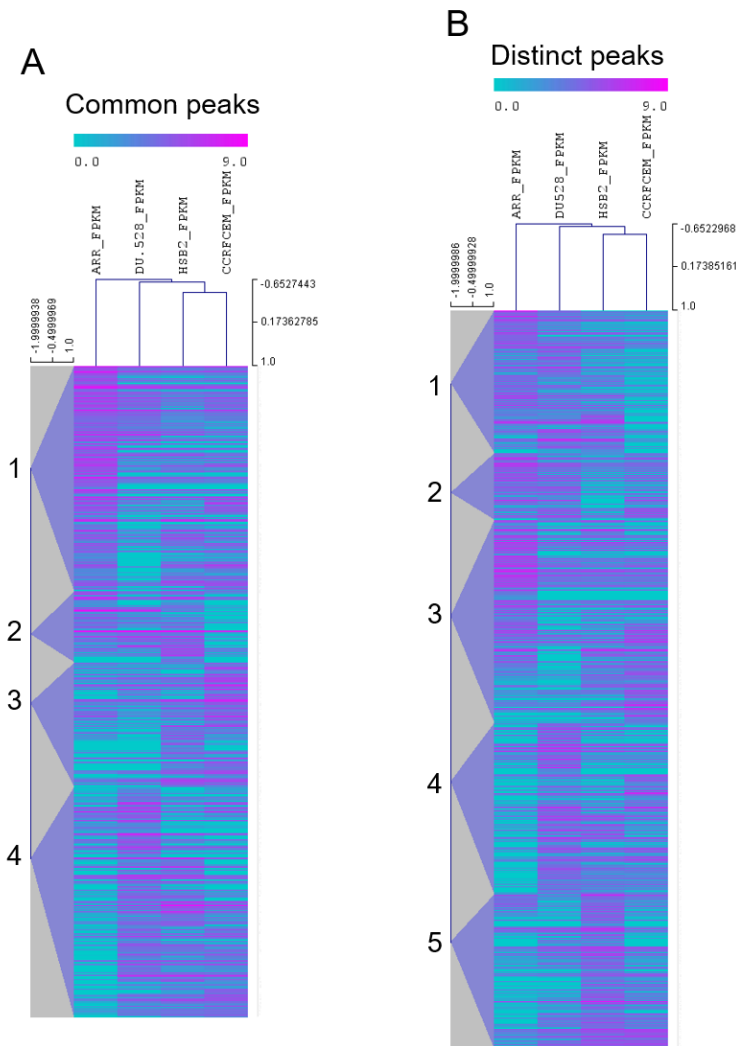
B

GO term BP	%	P-value	Genes
regulation of endosome size	1.835	0.0128	<i>ALS2, RAB5A</i>
retrograde vesicle-mediated transport, Golgi to ER	3.67	0.0159	<i>ATP9B, PITPNB, KIF3C, ERGIC2</i>
carbohydrate metabolic process	4.588	0.0260	<i>MGAT1, SLC2A3, POC1B-GALNT4, GALNT4, GLO1</i>
cell-cell adhesion	5.505	0.0307	<i>PDLIM5, RARS, PPFIBP1, OLA1, RAB10, ADGRL3</i>
cellular response to amino acid stimulus	2.753	0.0367	<i>SLC38A9, XBP1, CAPN2</i>
mRNA 5'-splice site recognition	1.835	0.0380	<i>SRSF1, PSIP1</i>
protein transport	6.423	0.0422	<i>RHBDF2, TRAM2, DNAJC13, XBP1, ZMAT3, RAB5A, KIF20A</i>

**Figure 3.10 GO analysis on differentially expressed genes associated with LMO2/TAL1 unique peaks in SIL/TAL1 cells show genes involved in cellular and metabolic processes.**

GO analyses for biological processes performed on (A) cluster 1, (B) cluster 5 as identified in Figure 3.8B. Table of GO term biological processes, percentage of genes, P-value and genes.

Next, we explored the expression of genes that were associated with the same peaks or different peaks across the cell lines (Figure 3.11). The genes associated with common peaks that had the highest gene expression in ARR in cluster 1 showed lower expression in the other cell lines especially in DU.528 cells (Figure 3.11A).



**Figure 3.11 RNA-seq analysis of differentially expressed genes associated with LMO2/TAL1 common and distinct peaks in ARR and SIL/TAL1 cell lines.**

Heat maps showing hierarchical clustering of differentially expressed genes associated with (A) common peaks and (B) distinct peaks. These genes are either the nearest 5' or 3' gene or contain the peak within the gene body. Scale bars represent colour index for the log2 FPKM values. Self-organising tree analysis identified 4 clusters in the common peaks group and 5 clusters in the distinct peaks group.

These genes are required for the maintenance of cellular and molecular functions such as negative regulation of apoptotic process, negative regulation of translation, regulation of gene expression and endocytosis (Figure 3.12A). By contrast, the genes in cluster 4 (Figure 3.11A) that had low levels of expression in ARR showed higher expression in the other cell lines and were mainly involved in cell signalling pathways, e.g. B and T-cell receptor signalling pathways. Genes in this cluster were also associated with regulation of NFκB activity, and apoptotic processes (Figure 3.12B). Importantly, *NKX3-1* gene has been identified in this cluster (Figure 3.12B). This gene encodes an NKL homeobox transcription factor that is necessary for T-cell proliferation, and has been reported as a downstream target of TAL1 in T-ALL cells including CCRF-CEM (Kusy et al., 2010, Sanda et al., 2012).

Lastly, we assessed the 849 genes that were differentially expressed in ARR and SIL/TAL1 but associated with different peaks. Hierarchical clustering showed that ARR cells had higher levels of expression of genes in clusters 1, 2 and 3 with a slight variation in the expression levels between the other cell lines (Figure 3.11B). In comparison, the genes that were observed in clusters 4 and 5 were expressed at a higher levels in SIL/TAL1 cells than ARR cells (Figure 3.11B). GO terms revealed biological processes that are important for cellular functions. For example, the genes collected from cluster 1 were responsible for apoptotic processes, cell-cell adhesion, endocytic recycling and positive regulation of GTPase activity (Figure 3.13). This cluster showed the *GATA2* gene which encodes for a member of the LMO2/TAL1 complex in HSCs. Normally *GATA2* is transcriptionally silenced during lymphocyte development (Seita et al., 2012, Yui and Rothenberg, 2014).

A

GO term BP	%	P-value	Genes
negative regulation of apoptotic process	8.876	1.98E-04	<i>PPARD, NAA15, SOD2, MAP4K4, DDX3X, RPS6KA1, UCP2, BCL2, CHST11, EEF2K, PRNP, CTSH, SLC40A1, CAMK1D, IL2</i>
negative regulation of translation	2.959	0.0026	<i>DDX3X, CPEB3, SYNCRIP, GIGYF2, CAPRIN2</i>
positive regulation of peptidyl-serine phosphorylation	2.959	0.0053	<i>OSM, PRKD2, BCL2, LATS1, CAPRIN2</i>
post-embryonic development	2.959	0.0061	<i>BCL2, CHST11, PYGO2, GIGYF2, SOD2</i>
oxidation-reduction process	8.285	0.0066	<i>GPD2, FADS1, EGLN3, FADS2, CYP2G1P, EGLN2, MTHFD1L, SOD2, GLRX2, MTHFD2, FAM213A, DNAJC24, KDSR, SARDH</i>
response to hypoxia	4.143	0.0077	<i>CCL2, UCP2, EGLN3, EGLN2, XRCC1, ALKBH5, SOD2</i>
negative regulation of gene expression	3.551	0.0122	<i>ANXA7, PICALM, LDLR, NDFIP2, MKKS, LGALS9</i>
endocytosis	3.551	0.0129	<i>STON2, PICALM, LDLR, EPS15L1, TNK2, CSNK1G3</i>
regulation of protein localization	2.367	0.0151	<i>PICALM, BCL2, PRNP, DNAJB6</i>
regulation of neuron apoptotic process	1.776	0.0152	<i>EGLN3, EGLN2, NSMF</i>

B

GO term BP	%	P-value	Genes
positive regulation of NFkB transcription factor activity	4.375	0.0017	<i>CFLAR, CARD11, NOD1, NTRK1, IL1RAP, FER, TRAF5</i>
B cell receptor signalling pathway	3.125	0.0017	<i>MEF2C, LAT2, PIK3CD, LCK, IGLL1</i>
thrombin receptor signalling pathway	1.875	0.0031	<i>F2RL3, F2RL1, IQGAP2</i>
activation of cysteine-type endopeptidase activity involved in apoptotic process	3.125	0.0080	<i>CASP3, NOD1, LCK, NKX3-1, PMAIP1</i>
T cell receptor signalling pathway	3.75	0.0134	<i>CARD11, PIK3CD, CD247, LCK, SKAP1, LCP2</i>
intracellular signal transduction	6.25	0.0147	<i>CORO2A, LAT2, RASSF5, NOD1, ADCY7, SPSB1, PREX2, FER, RGS9, LCP2</i>
regulation of GTPase activity	2.5	0.0237	<i>EPHA4, PLXNB2, IQGAP2, ADAP1</i>
platelet activation	3.125	0.0238	<i>F2RL3, LCK, PIK3R5, RHOG, LCP2</i>
phosphatidylinositol 3-kinase signalling	1.875	0.0268	<i>PIK3CD, PREX2, PIK3R5</i>
diapedesis	1.25	0.0281	<i>FER, ITGA4</i>

**Figure 3.12 GO analysis on differentially expressed genes associated with LMO2/TAL1 common peaks in ARR and SIL/TAL1 cells show genes involved in regulation of translation, regulation of gene expression, and T and B cell signalling pathways.**

GO analyses for biological processes performed on (A) cluster 1, (B) cluster 4 as identified in Figure 3.11A. Table of GO term biological processes, percentage of genes, P-value and genes.

GO term BP	%	P-value	Genes
apoptotic process	9.678	7.36E-04	<i>PRF1, HTT, PIM1, DUSP22, MAL, ZBTB16, BCL2L11, TAX1BP1, PLAGL1, ZFP36L1, ARRB1, MAP3K8, RALB, PDCD5, CUL1</i>
cell-cell adhesion	6.452	8.99E-04	<i>EPS15, SNX9, CHMP5, BAIAP2, USO1, S100A11, ARHGAP18, PDLIM1, TMEM2, EHD4</i>
endocytic recycling	2.581	1.33E-03	<i>EPS15, DENND1A, DENND1B, EHD4</i>
positive regulation of GTPase activity	9.033	2.13E-03	<i>SNX9, DENND1A, AKAP13, ARHGAP18, ITGB1, FARP1, DENND1B, ARHGAP25, ANKRD27, SH2D3C, ARRB1, AXIN2, ARHGAP10, FGD4</i>
heart development	4.517	6.86E-03	<i>ZFP36L1, SH3PXD2B, GATA2, RBM20, AKAP13, FOXC1, ADAM19</i>
proteasome-mediated ubiquitin-dependent protein catabolic process	4.517	1.11E-02	<i>PSMC6, PSMD12, ARRB1, RNF38, BIRC2, CUL1, USP44</i>
stimulatory C-type lectin receptor signalling pathway	3.226	1.59E-02	<i>RPS6KA5, PSMC6, PSMD12, PLCG2, CUL1</i>
positive regulation of receptor internalization	1.936	2.02E-02	<i>ARRB1, PLCG2, AHI1</i>
NIK/NFκB signalling	2.581	2.27E-02	<i>PSMC6, PSMD12, BIRC2, CUL1</i>
cardiac muscle cell differentiation	1.936	2.52E-02	<i>AKAP13, ITGB1, FOXP1</i>

**Figure 3.13 GO analysis on differentially expressed genes associated with LMO2/TAL1 distinct peaks in ARR and SIL/TAL1 cells show genes involved in apoptotic processes and regulation of GTPase activity.**

GO analyses for biological processes performed on cluster 1 as identified in Figure 3.11B. Table of GO term biological processes, percentage of genes, P-value and genes.

The genes in clusters 4 and 5 play a role in leukocytes differentiation and migration in addition to regulation of transcription (Figure 3.14A, B). The GO terms in these clusters highlighted genes that have previously been reported as TAL1 targets in T-ALL cells. Some of these genes encode DNA binding proteins such as *ETV6*, which belongs to the ETS family genes, and *ARID5B*, which is a member of the ARID family of transcription factors (Sanda et al., 2012, Leong et al., 2017, Tan et al., 2019). *CD84*, which encodes a cell surface marker required for T-cell activation (Sanda et al., 2012) was also found in cluster 4 (Figure 3.14A). Additionally, *NOTCH1* which plays a role in determination of lymphoid cell fate and is found mutated in many T-ALL cases (Mansour et al., 2007), was identified in cluster 4 (Figure 3.14A).

A

GO term BP	%	P-value	Genes
negative regulation of myeloid leukocyte differentiation	1.911	4.95E-04	<i>NME2, NME1-NME2, NME1</i>
angiogenesis	5.096	0.0045	<i>PTK2B, PECAM1, WASF2, ERAP1, BCAS3, PDCD6, RAMP1, SYK</i>
positive regulation of protein ubiquitination involved in ubiquitin-dependent protein catabolic process	1.911	0.0081	<i>PTK2B, CDK5RAP3, FBXO22</i>
regulation of establishment of cell polarity	1.911	0.0104	<i>PTK2B, BCAS3, KANK1</i>
positive regulation of GTPase activity	7.644	0.0151	<i>F11R, LIMS1, RINL, RGS3, SRGAP3, CYTH4, BCAS3, DENND2D, DENND3, ARHGEF12, CDC42EP3, RIN3</i>
positive regulation of transcription from RNA polymerase II promoter	10.829	0.0181	<i>MAVS, IKZF2, ARID3A, ARID3B, WT1, ZBTB38, NOTCH1, NME2, AHRR, IRF5, NME1-NME2, GLMP, CDK5RAP3, ETV6, BCAS3, NFATC3, CD28</i>
regulation of neutrophil degranulation	1.274	0.0183	<i>PRAM1, SYK</i>
leukocyte migration	3.185	0.0260	<i>CD84, F11R, PTPN6, CD244, PECAM1</i>
negative regulation of B cell receptor signalling pathway	1.274	0.0362	<i>PTPN6, LPXN</i>
endocytosis	3.185	0.0391	<i>RINL, NME1, RAB22A, WIPF1, RIN3</i>

B

GO term BP	%	P-value	Genes
intracellular signal transduction	8.333	1.23E-04	<i>ARHGEF3, CARHSP1, STK17B, WNK1, PRKCH, DGKH, GSG2, LATS2, PLCL2, TYK2, MAST4, SH2B3, DEDD2, NFATC1</i>
protein phosphorylation	8.333	4.07E-04	<i>IRAK1, STK10, TGFBR2, WNK1, STK17B, PRKCH, GSG2, LATS2, CDC25B, TYK2, ST3GAL1, PRKAR1B, RASSF2, MATK</i>
negative regulation of transforming growth factor beta receptor signalling pathway	2.976	0.0032	<i>SMAD7, TGFBR2, SMAD3, BCL9L, SKI</i>
negative regulation of osteoblast differentiation	2.381	0.0059	<i>SMAD3, SKI, SEMA4D, HDAC7</i>
negative regulation of inflammatory response	2.976	0.0067	<i>PTGIS, ADORA2A, ETS1, SMAD3, NFKB1</i>
signal transduction	12.500	0.0068	<i>IRAK1, SIT1, PKIG, PRKCH, NFKB1, DPYSL2, IL17RD, RALGDS, SIGIRR, LNPEP, ARHGAP21, P2RX4, CD38, LGALS3BP, IL10RB, RASGRP1, RASAL3, SH2B3, HBEGF, INPP4A, FAS</i>
negative regulation of transcription from RNA polymerase II promoter	8.929	0.0085	<i>TCF7, SMAD7, ARID5B, PKIG, WWC1, SMAD3, NFKB1, SKI, CBX2, FOXO3, HIC1, ATN1, ZNF683, SEMA4D, HDAC7</i>
positive regulation of transcription, DNA-templated	7.143	0.0100	<i>IRAK1, KLF6, CD38, TESC, ETS1, IRF1, SMAD3, NFKB1, FOXO3, SEC14L2, NFATC1, PBX4</i>
transmembrane receptor protein serine/threonine kinase signalling pathway	1.786	0.0110	<i>IRAK1, PALM2-AKAP2, AKAP2</i>
transforming growth factor beta receptor signalling pathway	2.976	0.0114	<i>SMAD7, TGFBR2, SMAD3, SKI, FURIN</i>

**Figure 3.14 GO analysis on differentially expressed genes associated with LMO2/TAL1 distinct peaks in ARR and SIL/TAL1 cells show genes involved in leukocytes differentiation and development, and regulation of transcription.**

GO analyses for biological processes performed on (A) cluster 4, (B) cluster 5 as identified in Figure 3.11B. Table of GO term biological processes, percentage of genes, P-value and genes.

Integration of the ChIP-seq and RNA-seq data indicated that the LMO2/TAL1 complex associated to genomic regions that could be involved in regulating genes that are crucial for cellular functions of T-cell and genes involved in transcriptional regulation in these T-ALL cell lines. This highlights the importance of controlling these genes for cell survival irrespective of the complex binding sites on the DNA. In addition, the LMO2/TAL1 complex binding might also regulate genes that are differentially expressed between individual T-cell lines. For instance, in ARR cells the potential target genes were necessary for RNA processing and cellular biosynthetic processes, while in SIL/TAL1 cells the associated genes were mainly required for metabolic processes. Notably, in all cell lines several genes that were associated with LMO2/TAL1 common peaks such as, *NKX3-1*, *ETV6*, *ARID5B*, and *CD84* are known targets of TAL1 in T-ALL (Kusy et al., 2010, Sanda et al., 2012, Tan et al., 2019).

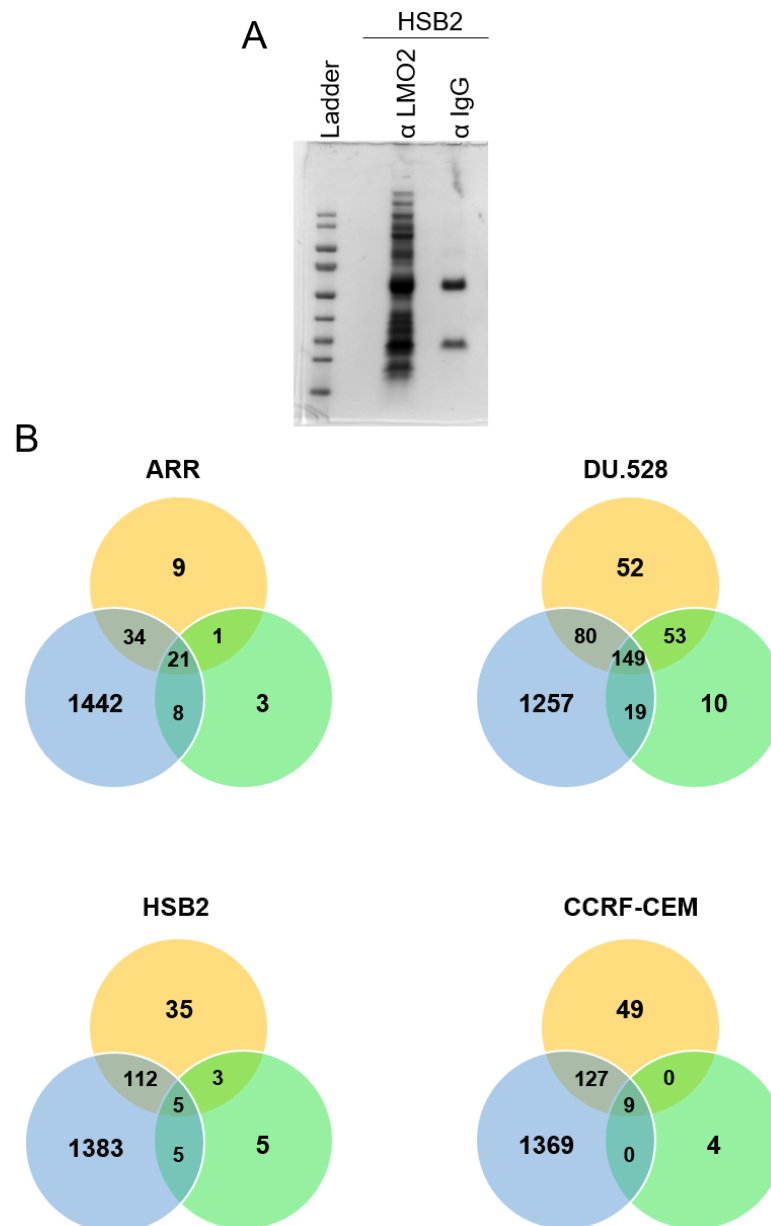
### **3.2.5 Identification of PHF6 as a potential interactor of LMO2 in T-ALL cell lines**

The ChIP-seq data showed that LMO2 and TAL1 mutually bind target elements in T-ALL cells. It also showed that LMO2/TAL1 bound regions were associated with differentially expressed genes and many of these genes are known targets of the complex. However, these experiments revealed that there are many genomic sites occupied by LMO2 but not by TAL1. We hypothesised that LMO2 participates in different complexes in T-ALL, in addition to LMO2/TAL1 complex which may also recruit proteins that are unknown yet. To test this hypothesis, we performed protein pull-down experiments with subsequent mass spectrometry to identify interacting partners of LMO2. Antibodies against LMO2 and IgG (negative control) were used for the immunoprecipitations in all four T-ALL cell lines (Figure 3.15A).

The peptides of proteins pulled down by LMO2 and IgG antibodies were searched against a protein database. These proteins were sorted in a descending order according to the highest Mascot score. All proteins of a score less than 25 and random false positives were excluded. To identify peptides that represent *bona fide* interactors versus those that are background contaminants, lists of gene IDs of common protein contaminants for *Homo sapiens* in T-ALL cell lines including Jurkat and CEM were downloaded from [www.crapome.org](http://www.crapome.org) online database (Mellacheruvu et al., 2013).

To exclude non-specific protein interactions, intersections of the LMO2, IgG, and contaminants datasets per cell line were carried out. Peptides that were common between LMO2, IgG, and contaminants were eliminated and only proteins specifically coprecipitated with LMO2 were selected, as depicted by Venn diagrams (Figure 3.15B). These Mass spectrometric analyses identified 9, 52, 35 and 49 proteins exclusively associated with LMO2 in ARR, DU.528, HSB2 and CCRF-CEM cells, respectively (Figure 3.15B). In addition to known LMO2 interacting proteins, such as TAL1, and LDB1 we identified potential interactions with two components of the NuRD complex, namely, PHF6 and HDAC1 (Todd and Picketts, 2012).



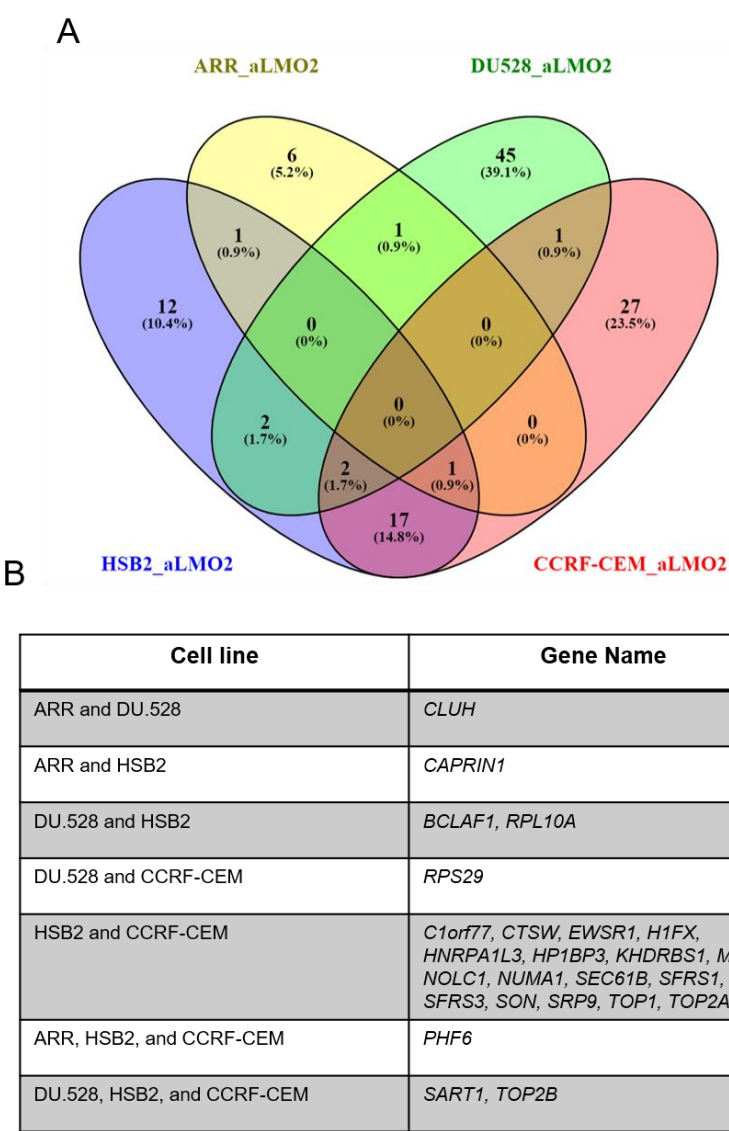


**Figure 3.15 Mass Spectrometry assay, T-ALL cell lines immunoprecipitation with LMO2 and IgG antibodies.**

(A) Representation of gel stained with Coomassie blue. (B) Venn diagrams showing overlap of proteins pulled down with LMO2 (yellow circle) versus IgG (green circle) and common contaminants (blue circle) in T-ALL cells.

To check for the proteins that were immunoprecipitated with LMO2 in more than one cell line, the unique proteins pulled down in each cell line were selected and intersected

(Figure 3.16A). Subsequently, the proteins that were in common between two or more cell lines were investigated (Figure 3.16B). Interestingly, PHF6 was identified as a potential protein-protein interactor with LMO2 in three of these T-ALL cell lines namely, ARR, HSB2, and CCRF-CEM.



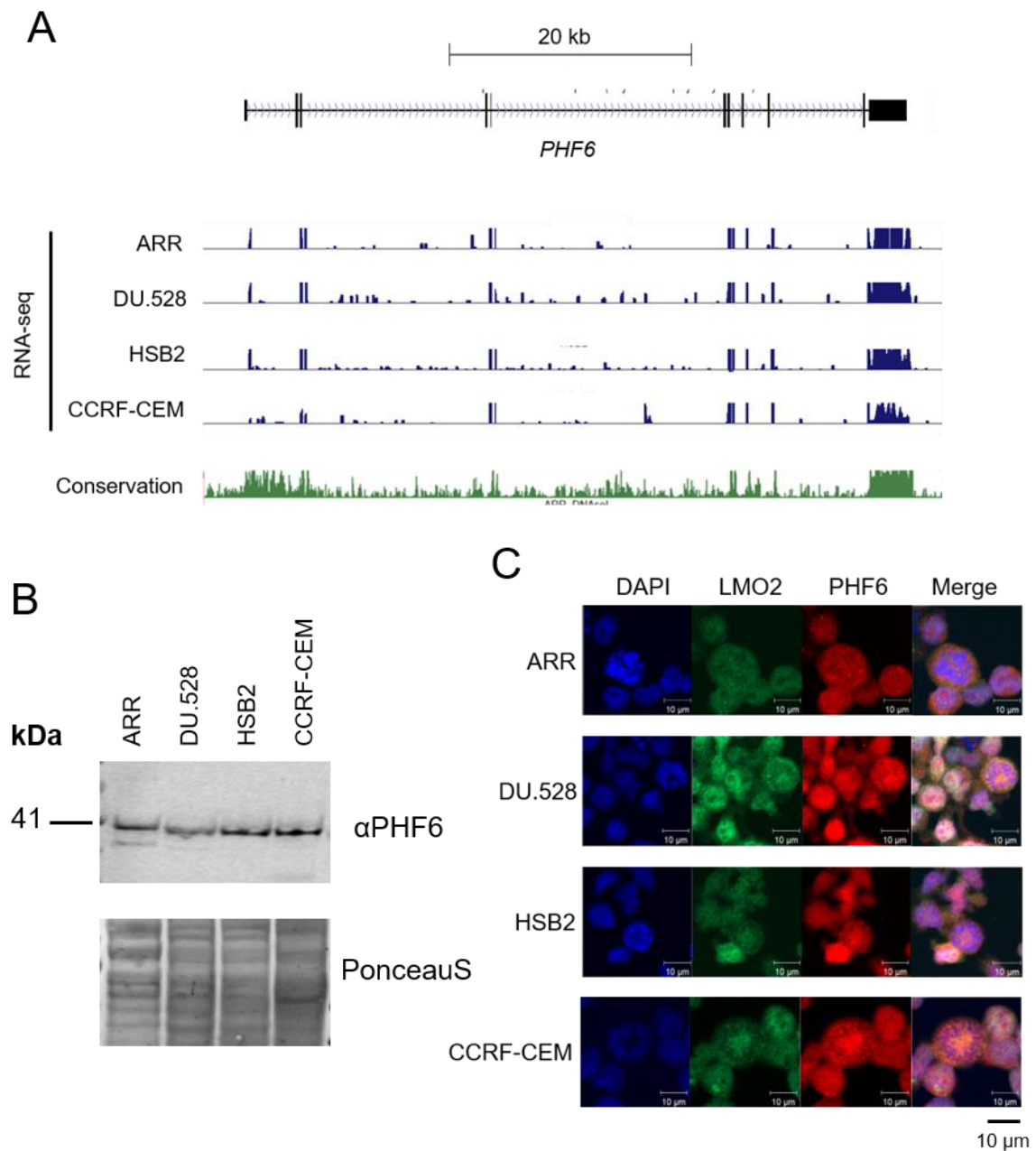
**Figure 3.16 Mass Spectrometry analyses show PHF6 as a potential interactor of LMO2.**

(A) Venn diagram showing Gene ID overlaps of proteins exclusively pulled down with LMO2 in T-ALL cells. (B) Table of gene names of proteins common in two cell lines or more.

### 3.2.6 *PHF6* is expressed in T-ALL cells

The unbiased screen for LMO2 protein interactors revealed PHF6 as a potential interacting partner by mass spectrometry. We investigated *PHF6* expression in our T-ALL cell models, because several studies have shown inactivating mutations of *PHF6* in T-ALL cases (Van Vlierberghe et al., 2010, Van Vlierberghe et al., 2011, Li et al., 2016, Wang et al., 2011). However, Wang *et al.* showed the absence of *PHF6* lesions in T-ALL cell lines including CCRF-CEM cells (Wang et al., 2011). Another study also reported a negative association between LMO2/TAL1 T-ALL subgroup and *PHF6* mutations, meaning they were largely mutually exclusive (Vicente et al., 2015). The RNA-seq data in all four T-ALL cell lines showed *PHF6* expression on the mRNA level in all of them as demonstrated by RNA-seq UCSC screenshot (Figure 3.17A).

To test whether PHF6 protein is present in a detectable level in these cells, western blot analyses were performed by running nuclear extracts of all four cell lines. Membranes with separated nuclear extracts were incubated with  $\alpha$ PHF6 antibody followed by incubation with a secondary antibody. The western blots showed PHF6 protein expression in all the cell lines (Figure 3.17B). Finally, we visualised the cellular localisation of LMO2 and PHF6 in these cell lines. The immunofluorescence staining using  $\alpha$ PHF6 and  $\alpha$ LMO2 antibodies showed that PHF6 was localised in the nucleus with LMO2 in all cells. Additionally, these cells showed minimal cytoplasmic staining of both LMO2 and PHF6 (Figure 3.17C).

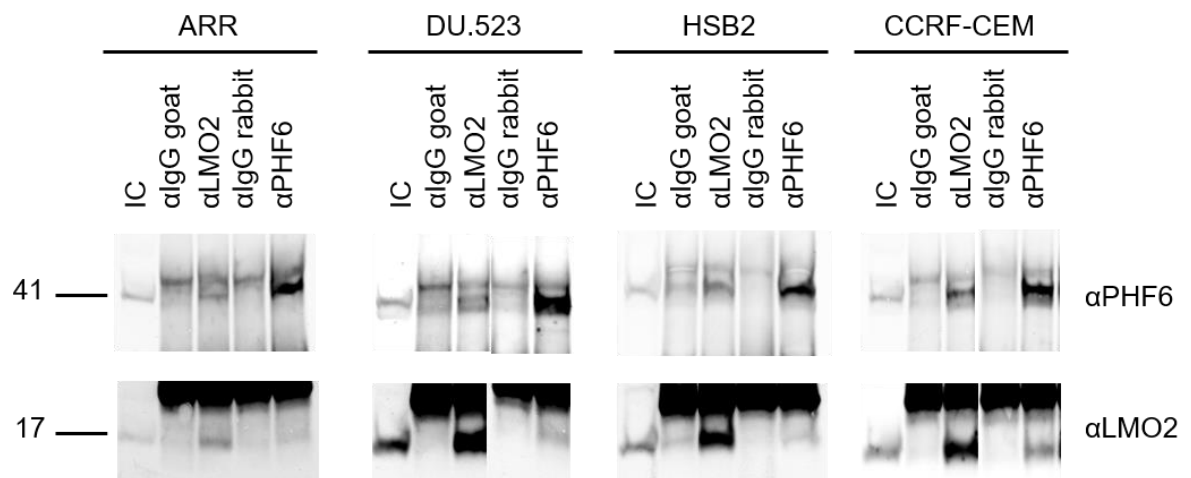


**Figure 3.17 PHF6 expression in T-ALL cell lines.**

(A) Screenshot from UCSC illustrating the distribution of reads over the *PHF6* loci. Uniform y-axis scales were used for each genomic region and positive values correspond to the direction of the transcript. Data represent the mean of three independent samples. (B) PHF6 protein levels detected by western blotting using nuclear extracts from T-ALL cells, PonceauS was used to show equal loading. (C) Immunofluorescent stained cells with antibodies against LMO2 (green), PHF6 (red) and counterstained with DAPI to visualise the nucleus, Scale bar, 10  $\mu$ m. Images were taken using Carl Zeiss LSM510 Meta Confocal microscope (Carl Zeiss, Germany).

### 3.2.7 PHF6 as a novel interactor of LMO2

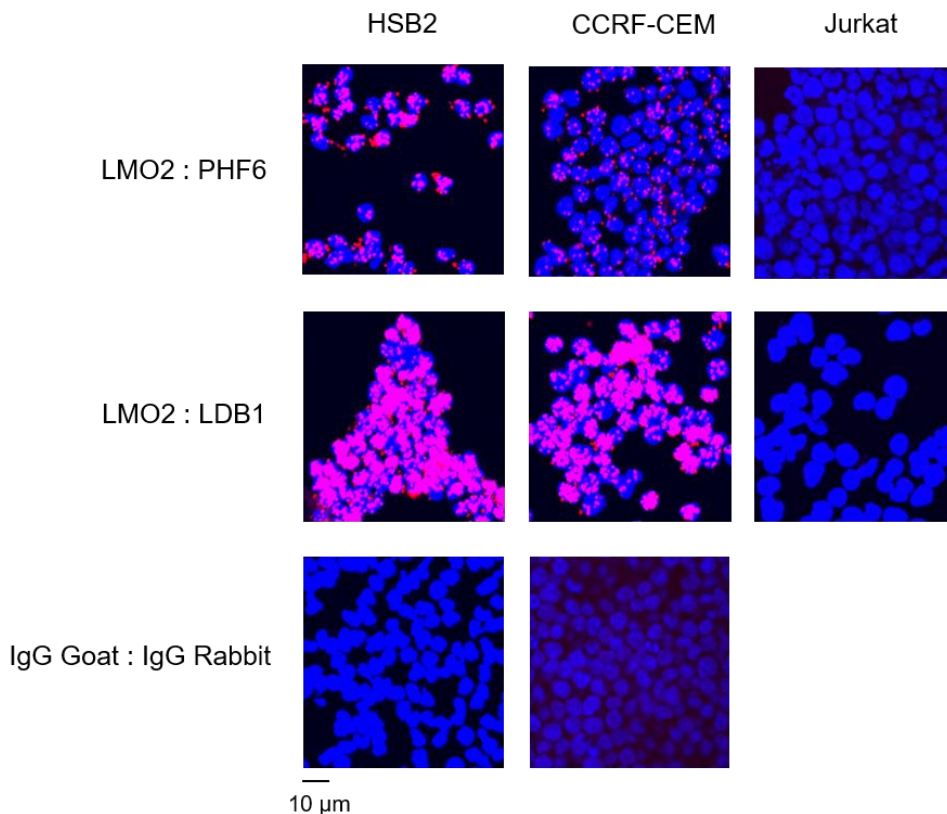
Given that PHF6 is expressed in T-ALL cells and localised with LMO2 in the nucleus, we sought to further validate their interaction which was initially detected by mass spectrometry analyses. Reciprocal immunoprecipitations were performed, using antibodies against the endogenous  $\alpha$ LMO2 and  $\alpha$ PHF6, followed by western blot analyses. The western blot analysis showed that PHF6 was present on a detectable level in samples immunoprecipitated with LMO2. The immunoprecipitation of LMO2 was also validated in each sample by detecting LMO2 itself. The same was applied to samples that were pulled down with PHF6. This revealed that LMO2 coprecipitated with PHF6 and vice versa indicating their interaction in T-ALL cells (Figure 3.18).



**Figure 3.18 PHF6 is a novel interacting partner of LMO2.**

Western blot analyses using LMO2 and PHF6 antibodies following immunoprecipitating nuclear extracts with goat IgG, LMO2, rabbit IgG, and PHF6 antibodies. Goat IgG antibody was used as a negative control for LMO2, rabbit IgG antibody was used as a negative control for PHF6. Nuclear extracts were used as positive input control (IC).

To further investigate the interaction between PHF6 and LMO2, *in-situ* proximity ligation assay (PLA) was performed. The immunoassay PLA allows the detection of endogenous protein-protein interactions via the production of fluorescent signal when proteins of interest are in extremely close proximity as described in chapter 2, section 2.12. The PLA experiments were conducted in HSB2 and CCRF-CEM cells, which verified the interaction between LMO2 and PHF6, appearing as localised dots in the nuclei (Figure 3.19). As positive control, cells were probed with antibodies raised against  $\alpha$ LMO2 and  $\alpha$ LDB1, as they form a well-established interaction (Wadman et al., 1997). As technical negative control, cells were probed with  $\alpha$ Goat and  $\alpha$ Rabbit IgGs and as biological negative control Jurkat cells were used (Figure 3.19), as these cells are known to express LMO1 instead of LMO2 (Palomero et al., 2006).



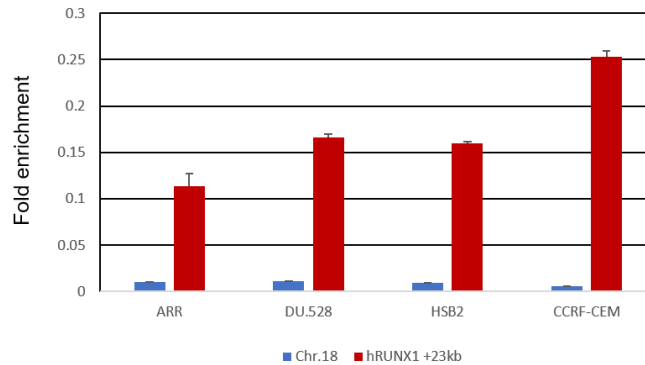
**Figure 3.19 *In situ* Proximity Ligation Assay (PLA) showing LMO2 and PHF6 interaction in HSB2 and CCRF-CEM cells.**

LMO2 and PHF6 interactions in HSB2; and CCRF-CEM cells were visualised using PLA. Pink staining indicates protein–protein interactions and DAPI (blue) is used to visualise the nucleus, Scale bar, 10  $\mu$ m. LMO2 and LDB1 interaction (positive control), IgG Goat and IgG Rabbit interaction (technical negative control), Jurkat cells (biological negative control). Confocal images were analysed using a Carl Zeiss LSM510 Meta Confocal microscope (Carl Zeiss, Germany).

### 3.2.8 PHF6 binds with LMO2 at *RUNX1* enhancer

The identification of an interaction between LMO2 and PHF6 suggested that PHF6 may act as a co-regulator for the LMO2-mediated oncogenic program by regulating gene expression in T-ALL. To test this hypothesis, we employed PHF6 ChIP assays followed by qPCR to determine at which regulatory elements it could be found. Initial ChIP experiments showed LMO2, TAL1 and LDB1 binding to the enhancer of *RUNX1* gene (Figure 3.2). Using  $\alpha$ PHF6 antibody co-localisation of PHF6 was tested against

this site. The qPCR results showed enrichment of PHF6 binding at this enhancer (Figure 3.20). Having substantiated the binding of PHF6 with LMO2 at *RUNX1* enhancer, we proceeded with PHF6 ChIP sequencing.



**Figure 3.20 ChIP qPCR analyses of PHF6 in T-ALL cells.**

Bar graphs are showing PHF6 binding to different DNA regions as the fold enrichment over the input control levels. Error bars represent SEM (n=2).

### 3.2.9 PHF6 co-occupies genomic sites and regulatory elements with LMO2, TAL1, and LDB1

The PHF6 ChIP-seq data revealed 5262 peaks for ARR, 2928 peaks for DU.528, 9543 peaks for HSB2 and 250 peaks for CCRF-CEM (Figure 3.21). In order to understand whether PHF6 participates in the LMO2/TAL1 complex at the DNA, we intersected the ChIP-seq data sets of PHF6 with LMO2, TAL1 and LDB1 per cell line (Figure 3.21). ARR cells revealed 5262 sites associated with PHF6 where 2389 of them were bound by LMO2, 2212 by TAL1 and 3828 by LDB1. Interestingly, PHF6 had a greater overlap with LDB1 than LMO2 and TAL1, however, the four factors bound a total of 1738 genomic regions in ARR cell lines (Figure 3.21). Comparably, PHF6 occupied 2928 sites in DU.528 cells whilst approximately 1200 of these sites were shared with each factor separately and 982 were common between all four factors (Figure 3.21). In



HSB2 cell line, 9543 regions were bound by PHF6. Similar to DU.528 cells, roughly 1300 peaks were shared individually with LMO2, TAL1, and LDB1 and 970 with the three factors (Figure 3.21). By contrast, a low number of reads was obtained in CCRF-CEM cells where only few peaks were called, but the majority of them overlapped with LMO2, TAL1, and LDB1 (Figure 3.21).

### ARR

Intersect (2 factors)	PHF6	LMO2	TAL1	LDB1	4 factors
Total peaks	5262	19277	13586	16254	1738
PHF6		2389	2212	3828	
LMO2			9077	4701	
TAL1				4091	

### DU.528

Intersect (2 factors)	PHF6	LMO2	TAL1	LDB1	4 factors
Total peaks	2928	26087	24720	12042	982
PHF6		1211	1218	1174	
LMO2			17011	5645	
TAL1				5578	

### HSB2

Intersect (2 factors)	PHF6	LMO2	TAL1	LDB1	4 factors
Total peaks	9543	29073	24860	10668	970
PHF6		1383	1341	1291	
LMO2			18389	5605	
TAL1				5346	

### CCRF-CEM

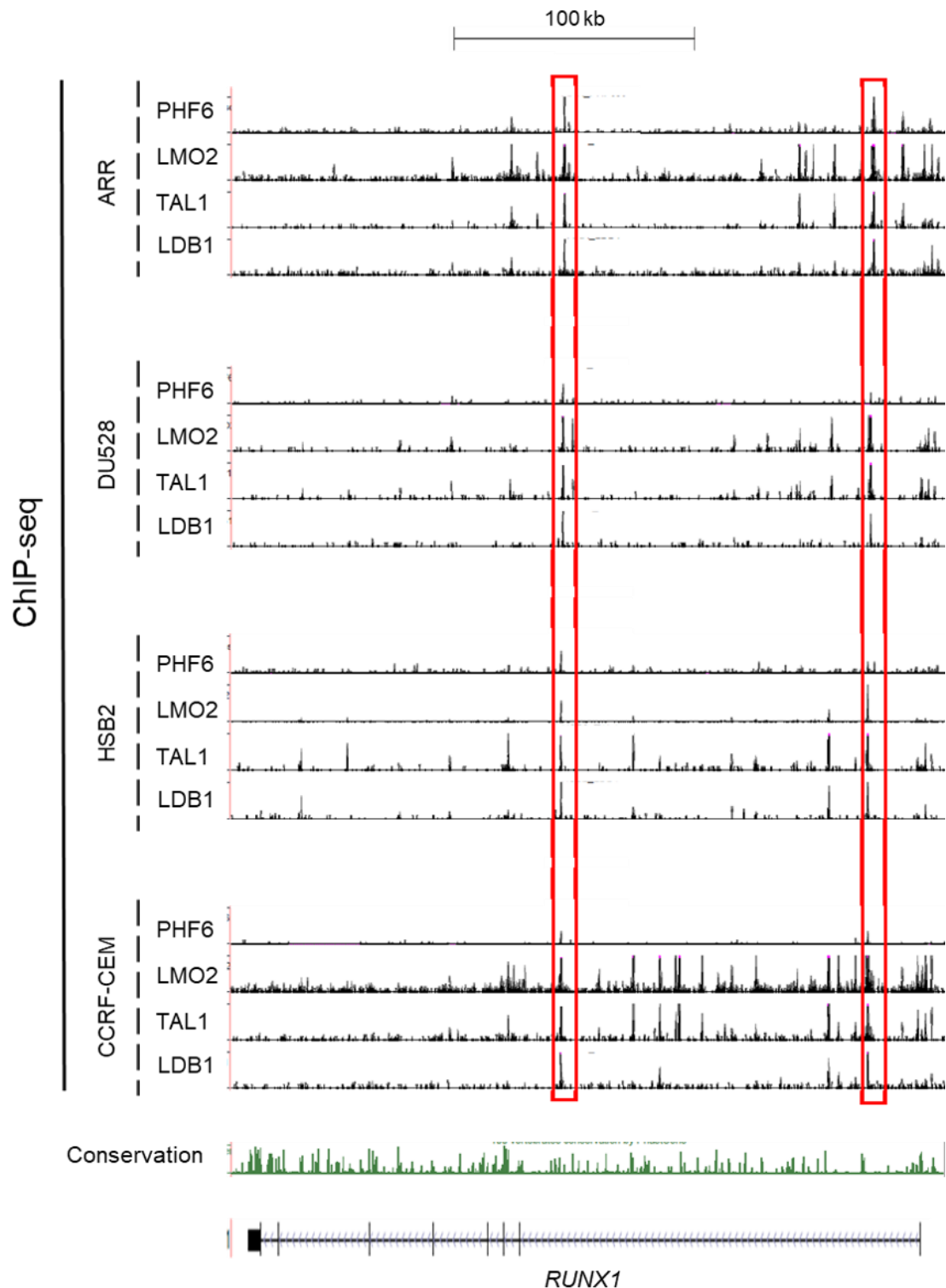
Intersect (2 factors)	PHF6	LMO2	TAL1	LDB1	4 factors
Total peaks	250	41647	28317	12844	228
PHF6		233	233	239	
LMO2			23546	9295	
TAL1				8648	

**Figure 3.21 PHF6 co-occupies genomic sites with LMO2, TAL1, and LDB1 in T-ALL cells.**

Tables showing intersection of PHF6, LMO2, TAL1, and LDB1 ChIP-seq peaks in T-ALL cell lines.

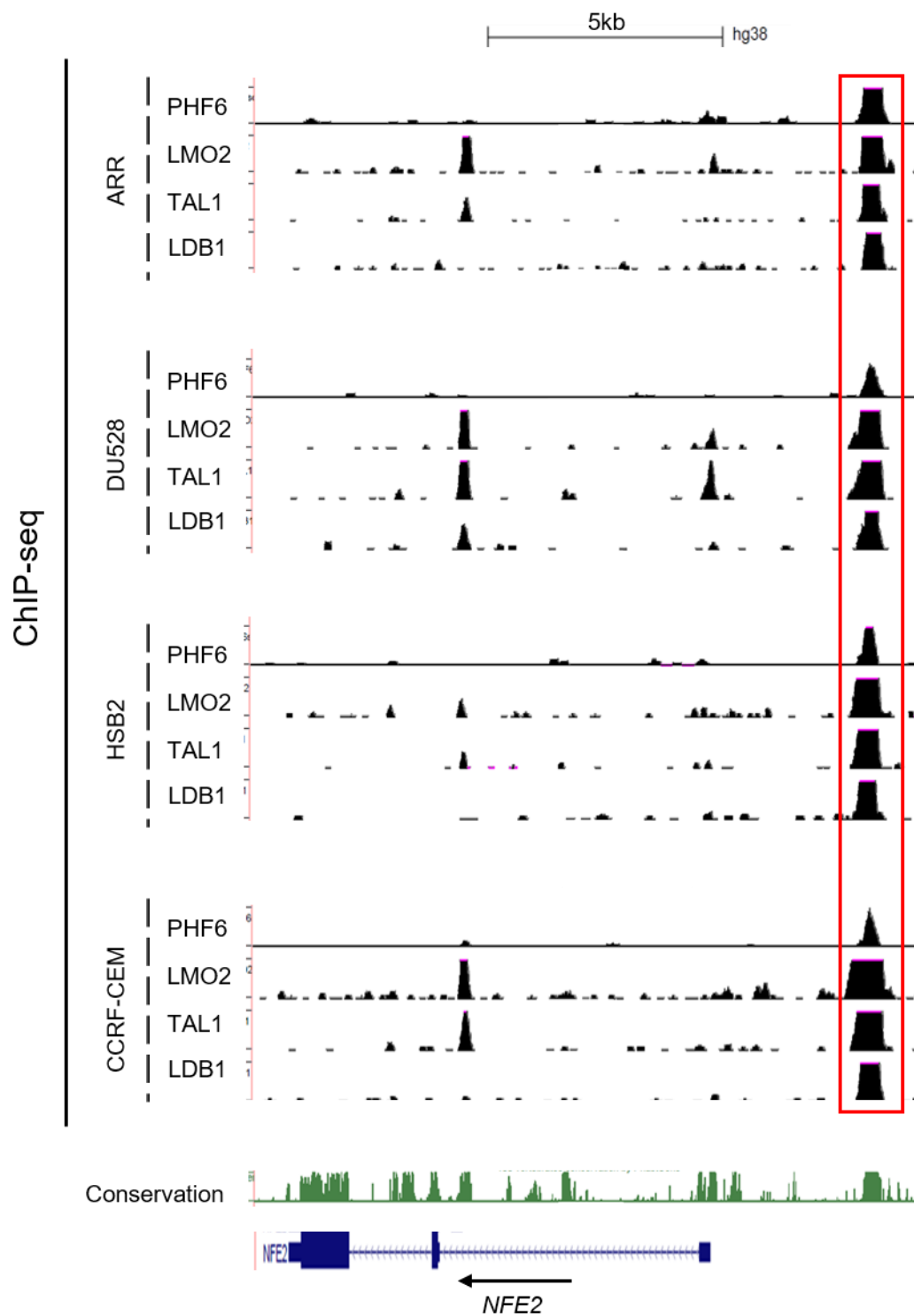
To assess target elements where PHF6 and the LMO2 complex components bind together, we analysed their ChIP-seq tracks on the UCSC genome browser. Visualisation using UCSC screenshots demonstrated overlapping peaks of PHF6 with the LMO2 complex within the gene body of *RUNX1*, at the *RUNX1* +23 kb enhancer, in all four T-ALL cell lines (Figure 3.22). Notably, this enhancer is known to be occupied by the LMO2 complex in normal haematopoietic cells and T-ALL (Nottingham et al., 2007, Sanda et al., 2012).

PHF6 also co-localised with LMO2/TAL1/LDB1 at the upstream regulatory element of *NFE2* gene as shown by the UCSC screenshot (Figure 3.23). The *NFE2* gene encodes a TF required for expression of several erythroid-specific genes and plays a role in regulating erythroid and megakaryocytic maturation and differentiation (Gavva et al., 1997). These results not only support our identification of PHF6 as an interacting partner of LMO2, but also imply that PHF6 may be a participating protein in LMO2/TAL1/LDB1 complex regulatory function.



**Figure 3.22 PHF6 peaks overlap with LMO2, TAL1, and LDB1 peaks in T-ALL cells at the *RUNX1* gene.**

Screenshot from the UCSC browser showing PHF6, LMO2, TAL1 and LDB1 ChIP-seq binding profiles within *RUNX1* gene body (red boxed panel left) including the *RUNX1* +23 kb enhancer (red boxed panel right).

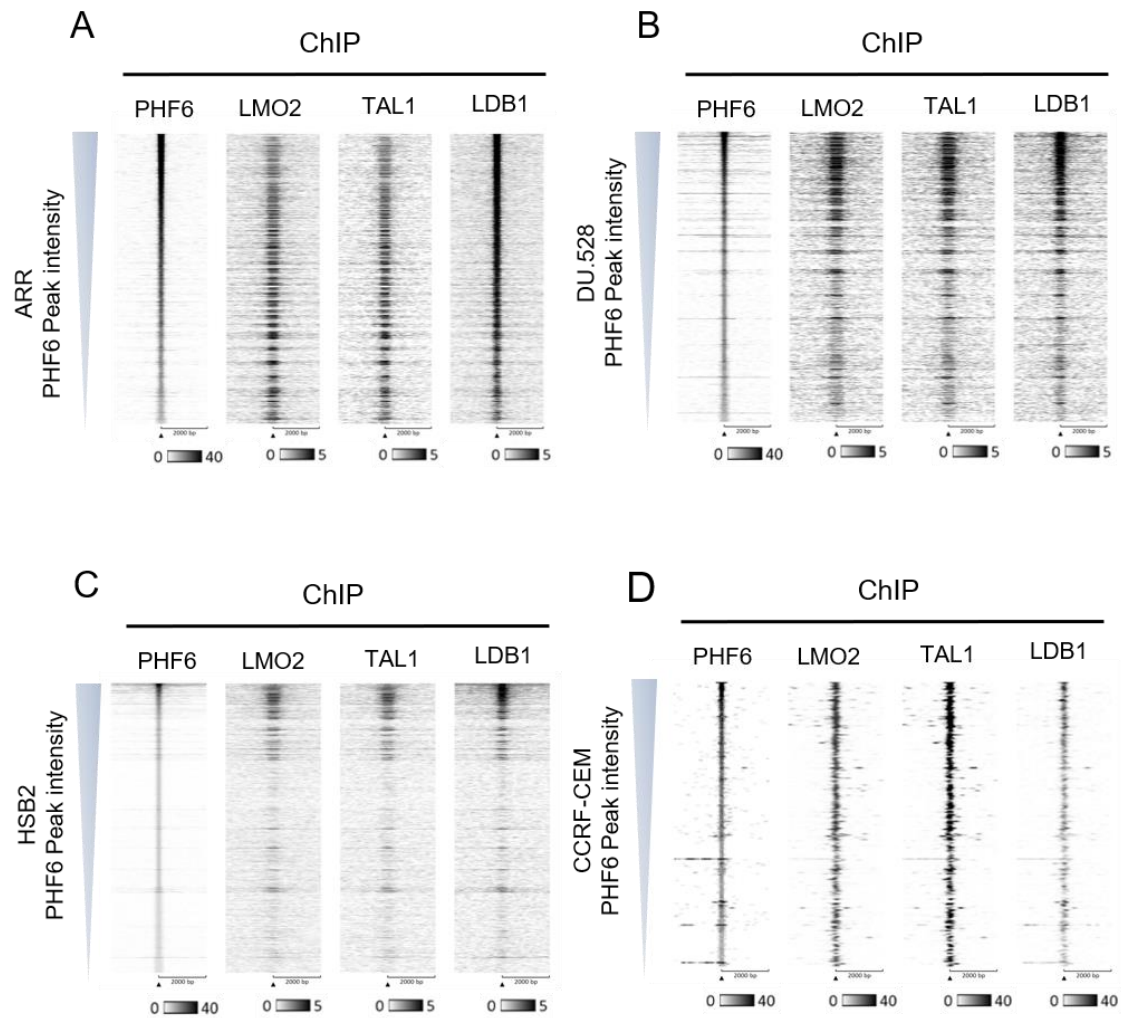


**Figure 3.23 PHF6 co-localises with LMO2, TAL1, and LDB1 in T-ALL cells at the *NFE2* regulatory element.**

Screenshot from the UCSC browser showing PHF6, LMO2, TAL1 and LDB1 ChIP-seq binding profiles at the *NFE2* regulatory element (red boxed panel).

### **3.2.10 PHF6 peaks are co-localised by LMO2, TAL1 and LDB1**

To visualise the genome wide association of PHF6 localisation with the LMO2 complex members, we generated a matrix for each cell line ranked according to the PHF6 peaks intensity scores and sorted them in a descending order. The matrices were overlaid with the Bam files, which represent aligned sequences, of the ChIP-seq data of PHF6, LMO2, TAL1 and LDB1 to generate heat maps, using EaSeq analysis software (Lerdrup et al., 2016). Comparing the PHF6 ChIP signal to the LMO2, TAL1 and LDB1 peaks demonstrated that in ARR cells LMO2 and TAL1 peaks were distributed throughout the heat map showing that they co-occupied region sets with PHF6 peaks of variable signal intensities. LDB1 also occupied these regions with LMO2 and TAL1 in addition to PHF6 binding sites with the highest intensities. This indicated that in the ARR cell line PHF6 might be interacting not only with the full LMO2/TAL1/LDB1 complex, but also with LDB1 in distinct complex, particularly at the sites with the strongest PHF6 binding (Figure 3.24A). On the other hand, the strongest PHF6 peaks were mirrored by the strongest LMO2, TAL1 and LDB1 peaks specifically in DU.528 and HSB2 cells. Looking at less strong peaks in these two cell lines, there was less overlap (Figure 3.24B, C). In CCRF-CEM cells almost all PHF6 associated regions were bound by LMO2, TAL1 and LDB1 regardless of the PHF6 peaks signal intensity (Figure 3.24D). This overlap reflected the intersection data (Figure 3.21) where CCRF-CEM ChIP-seq identified a low number of reads and only the strongest sites were recognised as peaks and these strongest PHF6 binding sites are co-localised by LMO2, TAL1 and LDB1.



**Figure 3.24 PHF6 peaks are co-localised with LMO2, TAL1 and LDB1.**

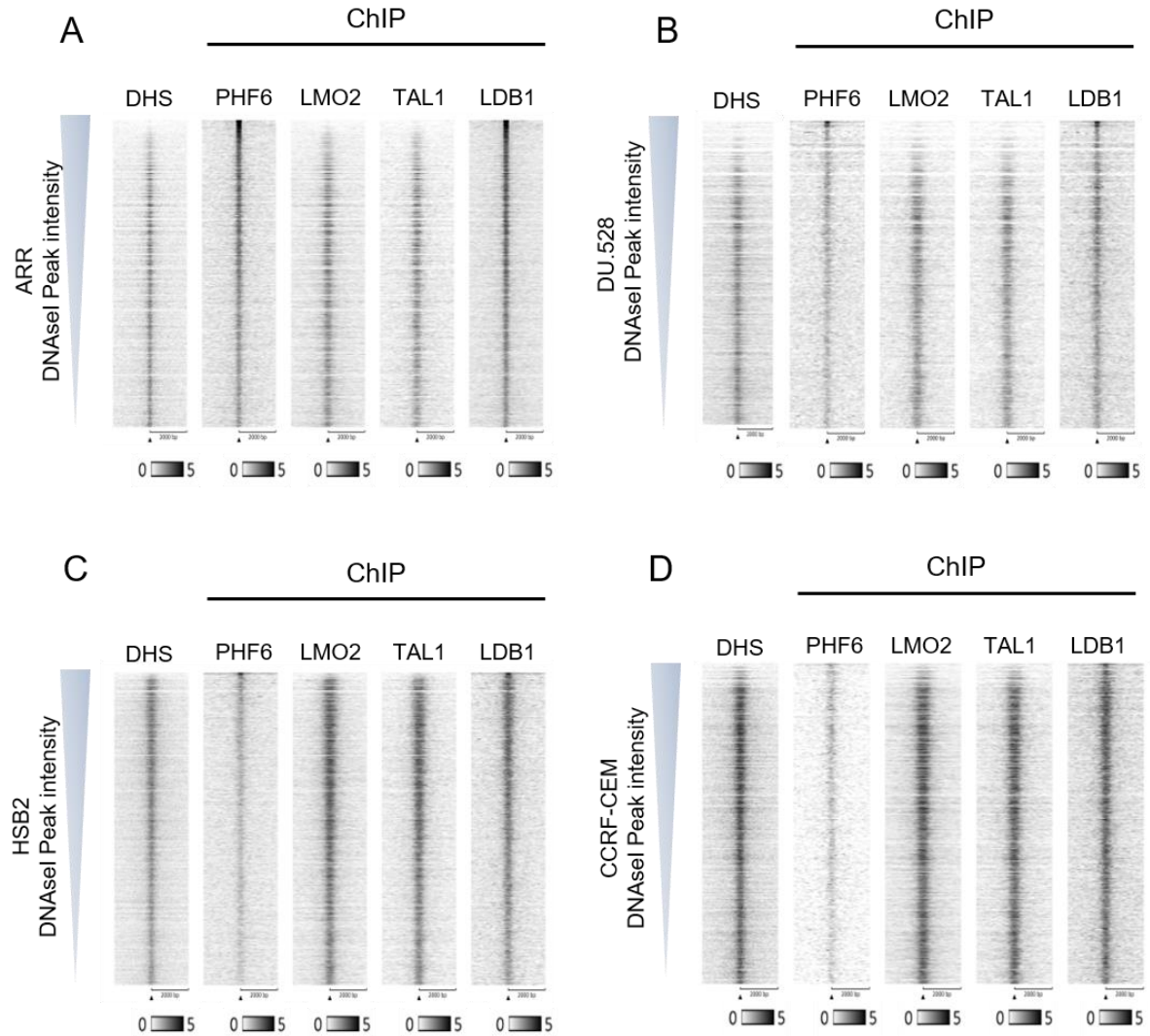
Heat map plots of ChIP-seq data for PHF6, LMO2, TAL1, and LDB1 in (A) ARR, (B) DU.528, (C) HSB2, (D) CCRF-CEM. The PHF6 ChIP-seq peaks (region sets) ranked according to peak intensity score in descending order. The heat map plots showing ChIP-seq data of PHF6, LMO2, TAL1, and LDB1 overlaid on PHF6 ChIP-seq region-sets. The Y-axis represents individual positions of the regions, and the X-axis represents -2 kb to +2 kb window centred on the summits of PHF6 peaks. At the bottom of each plot a bar showing the relationship between colouring and signal intensity.

### **3.2.11 DHS mapping identifies regions accessible by PHF6 and LMO2 complex members**

To investigate the accessibility of chromatin at binding sites of PHF6, LMO2, TAL1 and LDB1, we performed DNaseI hypersensitive site (DHS) mapping on each cell line. The DNaseI-seq experiments identified 15136 peaks for ARR, 9463 peaks for DU.528, 14261 peaks for HSB2, and 7831 peaks for CCRF-CEM. To generate heat map plots, DNaseI-seq matrices were overlaid with Bam files of the DHSs and ChIP-seq of PHF6, LMO2, TAL1 and LDB1 as described in section 3.2.10.

PHF6 ChIP signals showed that the strongest PHF6 binding occurred at DHSs with the top score in ARR cells. Moreover, LDB1 ChIP showed the highest binding intensity at the top with PHF6 (Figure 3.25A). On the other hand, LMO2 and TAL1 bound to open chromatin regions with lower peak intensity, where PHF6 and LDB1 also co-localised with them (Figure 3.25A). The DHS profiles of the other cell lines looked remarkably similar to those for ARR with PHF6 and LDB1 overlapping at highly accessible regions (Figure 3.25B, C, D). However, LMO2, TAL1 and LDB1 revealed more accessibility to open chromatin regions and showed higher signal intensities throughout the DHSs profiles of HSB2 and CCRF-CEM cells compared to ARR and DU.528 (Figure 3.25C, D). Surprisingly, the plots of DHSs showed that the regions with highest DNaseI scores had low signal intensity specially in ARR and DU528 cells (Figure 3.25A, B). According to Methylated DNA immunoprecipitation (MeDIP) experiments (data not shown) that were performed on these cell lines by another PhD student (Shoroog AlOmair), these top DHSs were methylated but we could not explain the demonstrated low intensity. Nonetheless, these analyses indicated that PHF6,

LMO2, TAL1, and LDB1 can access open chromatin regions where gene regulatory elements may reside.



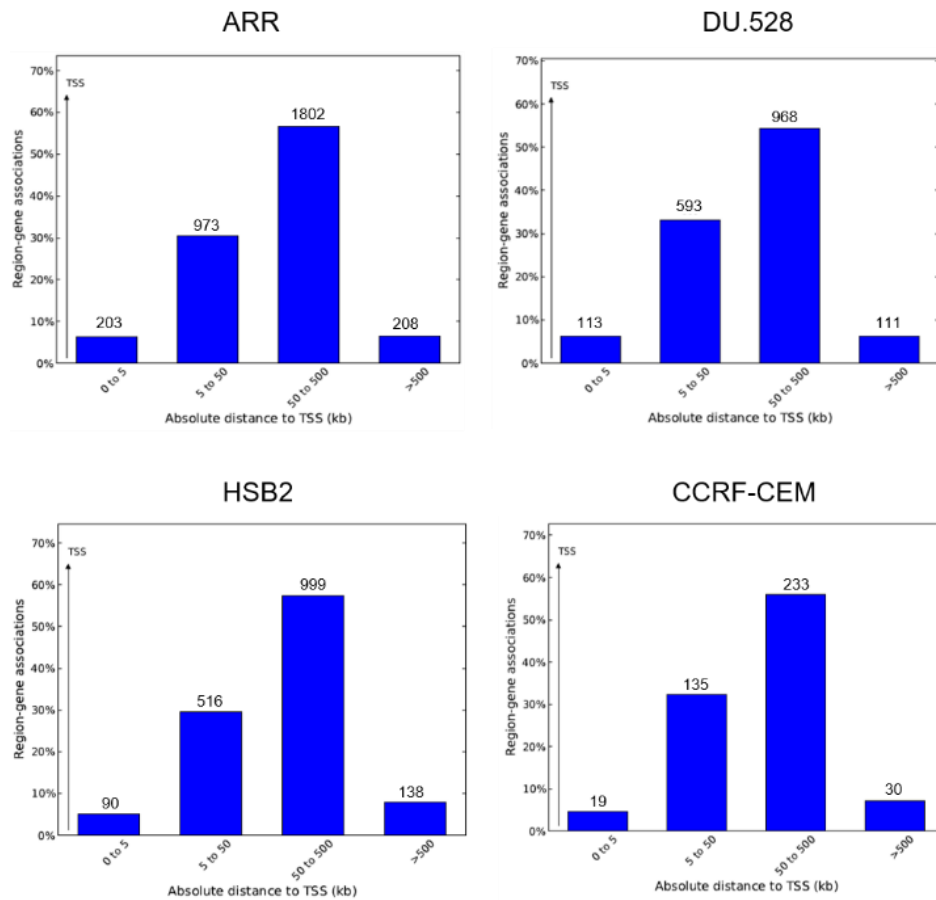
**Figure 3.25 PHF6, LMO2, TAL1, and LDB1 bind at DHSs in T-ALL cells.**

Heat map plots of DNaseI-seq in (A) ARR, (B) DU.528, (C) HSB2, (D) CCRF-CEM. The DNaseI peaks (region sets) ranked according to their peak intensity score in descending order. The heat map plots showing DHSs (left plot) and ChIP-seq data of PHF6, LMO2, TAL1, and LDB1 (right plots) overlaid on DNaseI region-sets. The Y-axis represents individual positions of the regions, and the X-axis represents from -2 kb to +2 kb window centred on the summits of DHSs. At the bottom of each plot a bar showing the relationship between colouring and signal intensity.



### **3.2.12 A high percentage of PHF6/LMO2/TAL1/LDB1 common peaks bind to genomic regions distant from TSSs**

Next, we examined the association between PHF6/LMO2/TAL1/LDB1 mutually bound genomic regions with their putative target genes. The analyses were performed using GREAT online tool as described in section 3.2.2. GREAT analysis showed that approximately 5% of the shared peaks were found at gene TSSs. Between 30-40% of the bound regions were located 5-50 kb away, whereas 50-60% were found 50-500 kb far from TSSs (Figure 3.26). The percentage of sites that were >500 kb distant was around 5%. These percentages were similar in all four cell lines (Figure 3.26). Notably, most of the genomic sites which were bound by PHF6, LMO2, TAL1 and LDB1 common peaks were localised remote from transcriptional start sites and were allocated to a high number of genes. These results were comparable to LMO2/TAL1 binding where the majority of peaks were distant from TSSs (Figure 3.5 & Figure 3.6).



**Figure 3.26 GREAT Region-Gene Association graphs of genomic regions bound by PHF6, LMO2, TAL1, and LDB1 common peaks.**

The y-axis is given in percentages, the absolute number of genes being counted is listed above each bar in the graph.

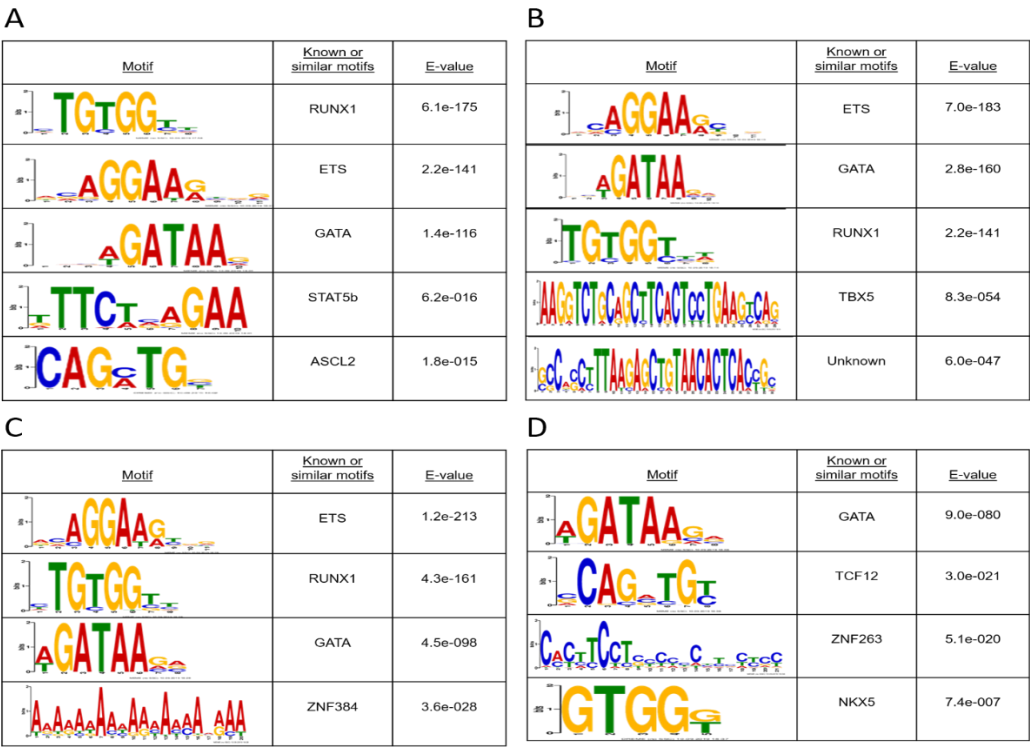
### 3.2.13 PHF6 peaks occupy haematopoietic TF motifs

The DNase-seq experiments showed that PHF6 binds at DHSs, either separately or mutually with LMO2 complex members. DHSs are usually enriched for gene regulatory elements such as enhancers and promoters (Cockerill, 2011). Because a previous study has proposed that PHF6 can bind DNA directly (Liu et al., 2014), we wanted to investigate whether PHF6 has any DNA sequence-specific properties in T-ALL cells. Thus, we analysed the enriched DNA motifs within PHF6 ChIP-seq binding sites using MEME (Bailey et al., 2009). These *de novo* motif analysis identified stretches of long

DNA sequences in ARR, DU.528, and HSB2 cells (Figure 3.27A, B, C). In ARR cells the enriched motifs were unknown (Figure 3.27A). Significantly, a GATA motif was enriched in DU.528, HSB2, and CCRF-CEM cells (Figure 3.27B, C, D). HSB2 cells showed enrichment of unknown motifs except for GATA (Figure 3.27C). All the enriched motifs in CCRF-CEM cells belonged to haematopoietic TFs including GATA, E-box (TCF12), SPI1, and RUNX1 (Figure 3.27D). These analyses did not give any indication of a PHF6 specific DNA binding motif.

**Figure 3.27 Motifs enriched at PHF6 binding sites in T-ALL cells.**

Having determined which peaks are in common between PHF6, LMO2, TAL1, and LDB1 we explored the enriched motifs within these shared peaks. *De novo* motif analyses revealed significantly enriched motifs of haematopoietic TFs such as GATA, RUNX1, and ETS (Figure 3.28A, B, C). ARR cells revealed additional motifs which are STAT5a and ASCL2 (Figure 3.28A). The results from DU.528 cells showed additional two long DNA sequences, where one was unknown and the other contained a TBX5 motif (Figure 3.28B). HSB2 and CCRF-CEM cells showed long DNA sequences, where zinc finger motifs including ZNF384 and ZNF263 are likely to reside (Figure 3.28C, D). CCRF-CEM was the only cell line that had an enriched E-box motif, *i.e.* TCF12, in addition to NKX3 sequence motif (Figure 3.28D).

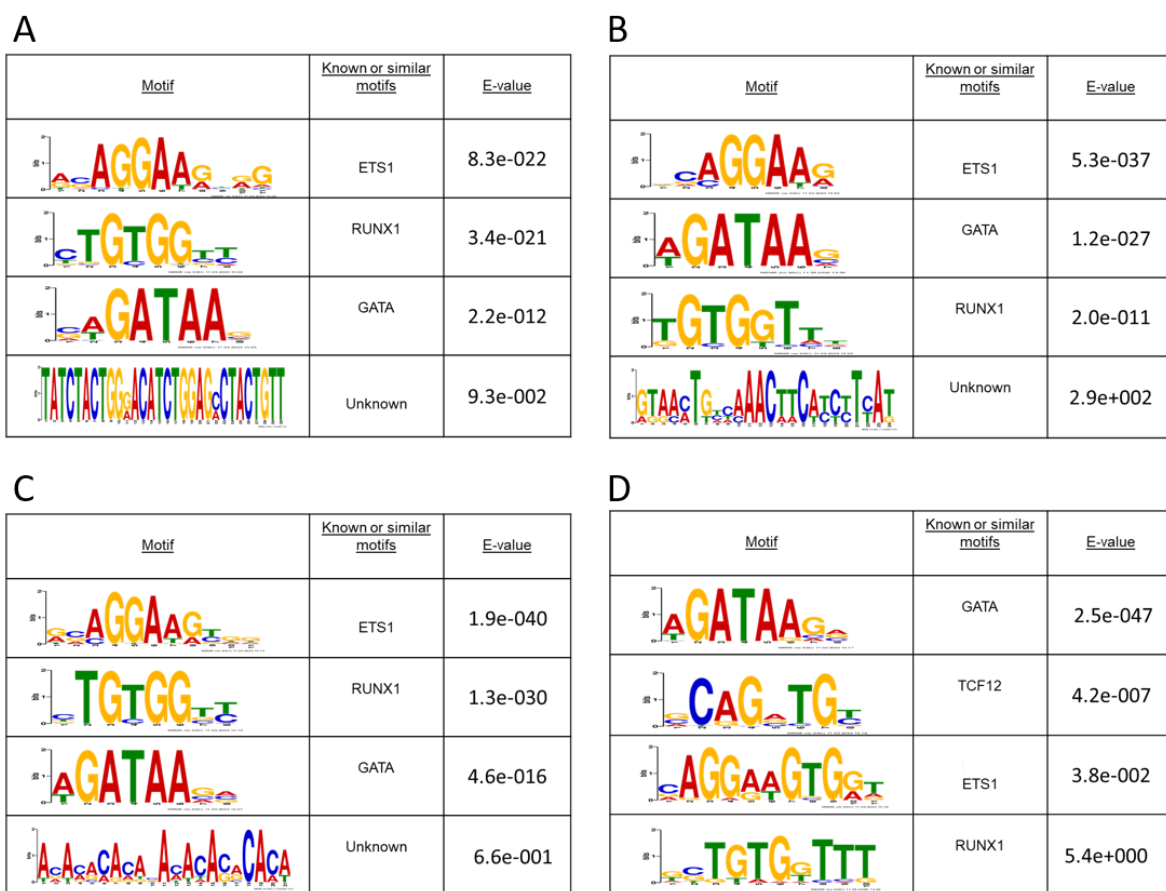


**Figure 3.28 Motifs enriched at PHF6, LMO2, TAL1, and LDB1 common binding sites in T-ALL cells.**

*De novo* motif analysis showing enriched motifs in (A) ARR, (B) DU.528, (C) HSB2, and (D) CCRF-CEM cells. The results are presented as sequence of the motif, name of the motif, and the E-value.

We further assessed the motifs that were specifically enriched by common peaks that were close to gene TSSs. We selected the PHF6/LMO2/TAL1/LDB1 peaks that were located at genomic regions 0-50 kb away from TSSs seen in Figure 3.26, followed by MEME motif analyses (Bailey et al., 2009). These analyses revealed that in all four cell lines ETS, GATA, and RUNX1 motifs were highly enriched in these shared binding sites (Figure 3.29). Additionally, ARR cells analysis showed an E-box motif (CATCTG) that resided within an enriched long DNA motif (Figure 3.29A). Comparably, in DU.528 and HSB2 cells the identified stretched DNA motifs did not harbour any known motifs (Figure 3.29B, C). While in CCRF-CEM cells no unknown DNA motifs were enriched and TCF12 E-box motif was identified (Figure 3.29D).

Taken together, these analyses identified *de novo* motifs which belong to hematopoietic TFs that are commonly associated with LMO2. These results are in line with PHF6 interaction with LMO2, indicating that PHF6 is recruited to DNA through DNA binding TFs which are part of the LMO2 complex.



**Figure 3.29 Motifs enriched at PHF6, LMO2, TAL1, and LDB1 common binding sites that are within 50 kb of TSSs in T-ALL cells.**

*De novo* motif analysis showing enriched motifs for peaks 0-50 kb away from TSSs in (A) ARR, (B) DU.528, (C) HSB2, and (D) CCRF-CEM cells. The results are presented as sequence of the motif, name of the motif, and the E-value.

### 3.2.14 PHF6/LMO2/TAL1/LDB1 co-localised regions are linked to genes involved in transcriptional regulation

We next wanted to examine the probable target genes of PHF6/LMO2/TAL1/LDB1 common peaks which were proximal to TSSs. To achieve this, the shared peaks that were located within 50 kb of TSSs, as seen in Figure 3.26, were selected for each cell line and the lists of their putative target genes were retrieved from GREAT analyses

data. These genes were further analysed using DAVID 6.8 (Huang et al., 2009), to assess their roles in biological processes.

The gene ontology analysis in ARR cells revealed genes linked to cellular processes of lymphocytes such as, T-cell activation, leukocyte migration, and immune response. Genes involved in regulation of transcription, regulation of blood vessel endothelial cell migration and regulation of I- $\kappa$ B kinase/NF $\kappa$ B signalling were also identified (Figure 3.30). This analysis featured genes encoding hematopoietic TFs such as, *TAL1*, *LYL1*, *RUNX1*, *ETS1*, *MYB*, and *NFE2* (Wilson et al., 2011, Palis et al., 1999, Gavva et al., 1997). Also known targets of TAL1 including *SATA5A*, *CD84*, and *CD87* (Figure 3.30) were identified by this analysis (Sanda et al., 2012).

GO Term BP	%	PValue	Genes
<b>T cell activation</b>	1.254	6.69E-06	<i>LAT, NLRC3, CASP8, CD276, SMAD3, TNFSF14, TREML2, IRF4, AZI2, ADA, KIF13B, HSH2D, CD7</i>
<b>positive regulation of blood vessel endothelial cell migration</b>	0.772	4.37E-05	<i>PRKD2, PDGFB, PLCG1, STAT5A, MAPK14, VEGFA, HSPB1, TGFB1</i>
<b>positive regulation of transcription, DNA-templated</b>	4.919	8.05E-05	<i>BMP10, GDF2, PDGFB, ELF4, SPI1, FOXO3, RORA, IL10, TGFB1, RARA, MYB, PCBD2, ARID1A, ESR2, PIM2, TRERF1, CD38, BPTF, SIX1, TADA3, UBE2V1, SOX7, ZBTB16, ZBTB17, WT1, TASP1, TAL1, LYL1, RUNX1, RUNX3, KLF6, IL5, NFE2, TRIP4, TGFB1, TAF8, SMAD3, SMAD2, CREB5, STAT1, FZD7, STAT3, ETS1, HIVEP3, IRF1, KDM8, PTCH1, IRF4, KLF2, WNT7A, F2R</i>
<b>leukocyte migration</b>	1.833	1.31E-04	<i>GLG1, ITGAL, CD244, C5AR1, SLC7A10, FPR3, DBH, CD84, CD47, ITGA6, PLCG1, CD34, PECAM1, MSN, TREM1, SELE, SELPLG, PIK3R1, SPN</i>
<b>immune response</b>	4.147	1.63E-04	<i>CSF2, TNFRSF21, IL19, JCHAIN, OAS3, TNFSF15, TNFSF14, PNP, IL10, MBP, CCL25, IL17A, TNFRSF1B, IL10RB, HRH2, SEMA7A, S1PR4, IL1RAP, IGLL1, IL2RG, NFIL3, SPN, CD7, SECTM1, TCF7, C5AR1, IL5, IL2RA, TNFSF4, NCF4, CD276, SMAD3, IL21, VAV1, OSM, LAT, CCR8, CCR7, TNFSF10, TNFSF11, ETS1, IL2, LCP2</i>
<b>signal transduction</b>	9.065	1.78E-04	<i>IMPA2, PDIA3, IL19, SORL1, TNFSF15, IQGAP2, TNFSF14, PRKG2, ARL2BP, ACVR1B, MAP3K6, ARHGAP22, WISP1, DTHD1, CHRNA9, ZFYVE16, CHRNA5, CSF2RB, RARA, STAM, RAPGEF1, PAG1, SPN, KIF13B, AVP, C5AR1, BCR, NFAM1, ESR2, IL21, IRS1, ARHGAP26, ARHGAP25, CD38, CCND3, ATP2C1, CD34, GTF2I, CAMK1, TGFB1, GRASP, MAVS, ITGAL, TNFRSF21, CD244, SAV1, NEDD9, FPR3, CXXC5, SFN, VDR, NPHP4, ANXA9, PDE1C, IL10RB, CD69, DGKD, RASAL3, NFAT5, LIMD1, IL2RG, RUNX1, PIK3R1, SECTM1, TNFSF4, TGFB1, NR4A1, NDFIP2, SLAMF1, STAT3, ITPR2, SH3BP5, TRIM55, CORO1C, ATF6, LAT, RASSF4, TNFSF10, PDE2A, P2RX1, PLCG1, MAPK14, PECAM1, CRH, HIVEP3, HBEGF, RAP1A, LRP8, GDF15, XCL2, ARAP1, PLA2, SH3BP2, BCAR3</i>
<b>negative regulation of I-<math>\kappa</math>B kinase/NF<math>\kappa</math>B signalling</b>	0.965	2.72E-04	<i>DAB2IP, NLRC3, CASP8, CARD19, ZC3H12A, RORA, STAT1, ADIPOQ, GSTP1, RHOH</i>
<b>negative regulation of inflammatory response</b>	1.351	3.72E-04	<i>ZFP36, TNFRSF1B, BCR, IL2RA, ETS1, ADORA2A, OTULIN, CD276, SMAD3, GHRL, RORA, ADIPOQ, ADA, IL2</i>
<b>positive regulation of cytosolic calcium ion concentration</b>	1.833	4.28E-04	<i>AVP, GNA15, C5AR1, SWAP70, CD52, FPR3, CCR8, GATA2, CD38, CCR7, GPR35, CHRNA9, S1PR4, PTGDR, AVPR1B, PLA2G6, GHRL, F2R, IL2</i>
<b>pathway-restricted SMAD protein phosphorylation</b>	0.579	4.57E-04	<i>BMP10, GDF2, SMAD7, TGFB1, USP15, TGFB1</i>

**Figure 3.30 GO analysis show genes associated with cellular processes of lymphocytes and regulation of transcription in ARR cells.**

GO analysis on putative target genes of PHF6/LMO2/TAL1/LDB1 shared peaks that are within 50 kb of TSSs in ARR cells. Table show the top 10 GO term biological processes, percentage of genes, P-value and gene names.



In DU.528 cells, the recognised genes were linked to regulation of MAP kinase activity, vascular endothelial growth factor receptor signalling pathway, and regulation of transcription including *HHEX*, *LYL1*, and *NFE2* (Figure 3.31).

GO Term BP	%	PValue	Genes
signal transduction	8.074	1.79E-04	<i>PTGES3, F2RL3, SORL1, FASLG, ACVR1B, CHRNA9, PTGES, RALA, STAM, RANBP1, MAP2K6, GTPBP1, LYN, STK24, ARHGAP25, SAG, GABRR2, CD34, IPO7, PDCL, RIN1, PDE9A, MAPK7, EXT2, GRASP, RIN3, OPN1LW, LITAF, PPFIA1, MAPKAPK3, CD70, CXXC5, POMC, PXN, ADRA2A, SH2B3, LIMD1, CALML5, INPP5D, PIK3R1, RASA2, WNT8B, IL18R1, IL2RB, IL1RL1, SPHK1, OR1F1, TNFSF9, SLAMF1, ITPR2, SH3BP5, LSP1, RASSF6, CXCL14, HIVEP3, BCAR3, SH3BP2</i>
positive regulation of signal transduction	1.275	3.97E-04	<i>SORBS1, LASP1, SH2B3, STAM, SKAP1, CRK, SH3BP2, SLA, BCAR3</i>
phosphatidylinositol biosynthetic process	1.134	0.001502	<i>PIK3C2B, PLCG2, SYNJ2, PIK3R5, INPP5D, PIK3R6, PIP4K2A, PIK3R1</i>
platelet activation	1.559	0.002097	<i>F2RL3, LYN, PLCG2, ADRA2A, PIK3R5, PIK3R6, PIK3R1, LCP2, ITPR2, GP9, F2R</i>
positive regulation of cell migration	1.984	0.003173	<i>COL18A1, LYN, MYO1C, DIAPH1, SPHK1, SNAI2, SYNE2, ITGA6, XBP1, SEMATA, ADRA2A, SEMA4A, PIK3R1, F2R</i>
negative regulation of MAP kinase activity	0.85	0.003833	<i>CBLC, SPRY1, LYN, SORL1, MAPK7, DUSP7</i>
response to cytokine	0.992	0.004069	<i>TYMS, REL, PTGES, JUN, MAPKAPK3, BCL2L1, STAT1</i>
vascular endothelial growth factor receptor signalling pathway	1.134	0.005179	<i>XBP1, MAPKAPK3, CYFIP2, HSPB1, BRK1, CRK, PIK3R1, PXN</i>
positive regulation of transcription, DNA-templated	3.97	0.005501	<i>E2F3, RAI1, SOX7, ZBTB16, HHEX, SKAP1, IL10, CTNNB1, EPC1, FUBP3, LYL1, HNRNPD, TESC, TRIP4, NFE2, IL5, VHL, MECP2, PIM2, STAT1, TNNT2, JUN, HIVEP3, KDM8, IRF1, PTCH1, ZFH3, F2R</i>
negative regulation of neuron death	0.85	0.006078	<i>UCN, REL, CD34, SNCA, SORL1, CTNNB1</i>

**Figure 3.31 GO analysis show genes associated with regulation of MAP kinase activity, vascular endothelial growth factor receptor signalling pathway, and regulation of transcription in DU.528 cells.**

GO analysis on putative target genes of PHF6/LMO2/TAL1/LDB1 shared peaks that are within 50 kb of TSSs in DU.528 cells. Table show the top 10 GO term biological processes, percentage of genes, P-value and gene names.

The GO terms also highlighted genes associated with JAK-STAT cascade, regulation of transcription, germ cell development, and regulation of anoikis in HSB2 cells, where *STAT5A* and *RUNX1* were also identified (Figure 3.32).

GO Term BP	%	PValue	Genes
JAK-STAT cascade	0.991	0.001224	<i>IFNL2, SOCS3, STAT5A, IFNL3, CTR9, IL31RA</i>
negative regulation of transcription from RNA polymerase II promoter	5.446	0.001794	<i>BACH2, WFS1, CTCF, CXXC5, WT1, CTNNB1, EPC1, EZR, REL, DRAP1, ZNF683, JUND, RARA, BHLHE40, MAF, TCF7, RCOR3, IKZF1, VHL, DMRT1, SMAD3, CELA1, SNAI2, CTR9, NOTCH1, SNAI3, ACVR2B, BPTF, RIPPLY1, PTCH1, PRDM1, RERE, HDAC7</i>
leukocyte cell-cell adhesion	0.826	0.003501	<i>VCAM1, ITGAL, EZR, CD40LG, ITGB1</i>
negative regulation of nitric oxide biosynthetic process	0.661	0.003939	<i>PTGIS, ATP2B4, CD34, IL10</i>
positive regulation of Rho protein signal transduction	0.826	0.004055	<i>GPR18, LPAR6, ARR1, GPR17, ADGRG1</i>
positive regulation of signal transduction	1.156	0.004607	<i>LAT, SH2B3, STAM, GRAP2, SH3BP2, SLA, BCAR3</i>
signal transduction	7.426	0.00594	<i>ITGAL, OXT, SORL1, FPR3, CXXC5, POMC, VIPR2, ARHGAP5, RAC2, CHRNA9, SH2B3, RALA, RARA, STAM, RANBP1, RUNX1, RASA3, KIF13B, AVP, EPAS1, STK24, NDFIP2, ARHGAP25, SAG, TNFSF8, SH3BP5, TRIM55, LAT, ACVR2B, MAPK6, ARR1, CD34, HIVEP3, CHRN3, GRK7, RAP1A, RIN1, PDE9A, VOPP1, IGFBP2, XCL2, CD226, PLA2, BCAR3, SH3BP2</i>
germ cell development	0.826	0.006859	<i>RARA, PRDM1, BCL2L1, LIN28A, WT1</i>
cell-matrix adhesion	1.321	0.008299	<i>VCAM1, ITGAL, ITGA6, CD34, HPSE, ITGB1, CTNNB1, MUC4</i>
negative regulation of anoikis	0.661	0.00868	<i>NOTCH1, BCL2L1, SNAI2, ITGB1</i>

**Figure 3.32 GO analysis show genes associated with JAK-STAT cascade, regulation of transcription, and germ cell development in HSB2 cells.**

GO analysis on putative target genes of PHF6/LMO2/TAL1/LDB1 shared peaks that are within 50 kb of TSSs in HSB2 cells. Table show the top 10 GO term biological processes, percentage of genes, P-value and gene names.

The analysis in CCRF-CEM cells showed genes such as *CD84* and *CD47* involved in leukocyte migration. Biological processes also included transport and T-helper 17 cell lineage commitment. Genes including *JARID2* and *MYB* were linked to regulation of H3-K9 methylation (Figure 3.33). Interestingly, many genes that are associated with hematopoietic transcriptional regulation involving *TAL1*, *LYL1*, *RUNX1*, *MYB*, *SPI1*, *NOTCH1*, *TCF7* and *ETS1* (Hoang et al., 2016, Baer, 1993, Imperato et al., 2015, Palis et al., 1999, Ciau-Uitz et al., 2013, Wilson et al., 2011, Weng et al., 2006, Rothenberg et al., 2016) were potential targets of the PHF6/LMO2/TAL1/LDB1 complex in these T-ALL cell lines. These findings are in line with the ChIP qPCR results which showed

binding of PHF6, LMO2, TAL1, and LDB1 at the *RUNX1* enhancer, and reflected the UCSC analyses which showed peaks of these factors at the regulatory elements of *RUNX1* and *NFE2*.

GO Term BP	%	PValue	Genes
leukocyte migration	3.922	0.002341	<i>CD84, CD47, CD244, ITGA6, CD34, DBH</i>
transport	5.229	0.016649	<i>GET4, SLC35B1, PITPNM2, LAPTM5, ABCB10, CLCNKB, MFSD2B, ITPR2</i>
T-helper 17 cell lineage commitment	1.308	0.037252	<i>SLAMF6, LY9</i>
positive regulation of histone H3-K9 methylation	1.308	0.044536	<i>JARID2, MYB</i>

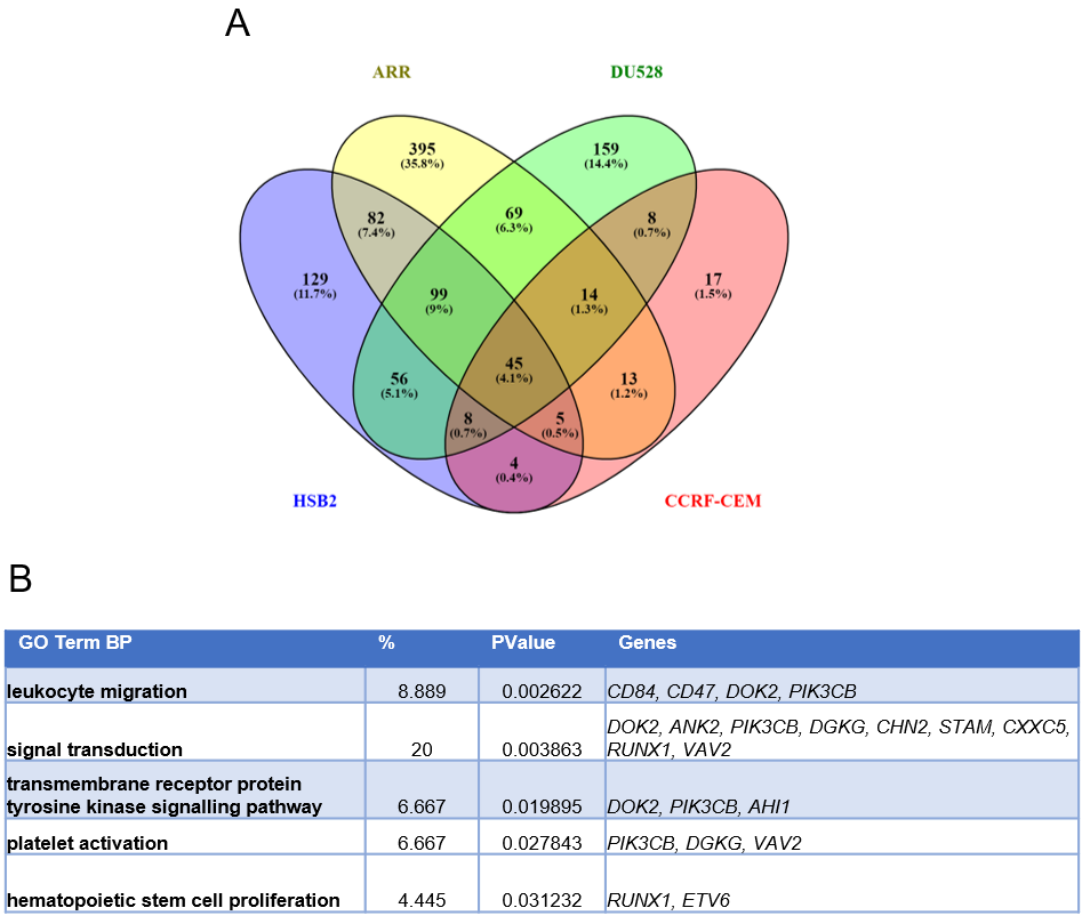
**Figure 3.33 GO analysis show genes associated with T-helper 17 cell lineage commitment, and regulation of H3-K9 methylation in CCRF-CEM cells.**

GO analysis on putative target genes of PHF6/LMO2/TAL1/LDB1 shared peaks that are within 50 kb of TSSs in CCRF-CEM cells. Table show the GO term biological processes, percentage of genes, P-value and gene names.

### 3.2.15 PHF6/LMO2/TAL1/LDB1 binding sites are associated with differentially expressed genes in T-ALLs

To investigate whether PHF6/LMO2/TAL1/LDB1 common binding sites may potentially regulate gene expression, we integrated the ChIP-seq data of these bound regions with RNA-seq data in each cell line as described in section 3.2.3. These data showed 722 differentially expressed genes associated with the common peaks in ARR. Lower numbers of differentially expressed genes were identified in the other cell lines with 458 genes in DU.528, 428 genes in HSB2 and 114 genes in CCRF-CEM. To explore the genes that were possibly regulated by PHF6/LMO2/TAL1/LDB1 shared peaks in all four cell lines, the gene names of the targeted differentially expressed genes per cell line were intersected. The Venn diagram showed 45 differentially expressed genes common between the four cell lines which are potential targets of

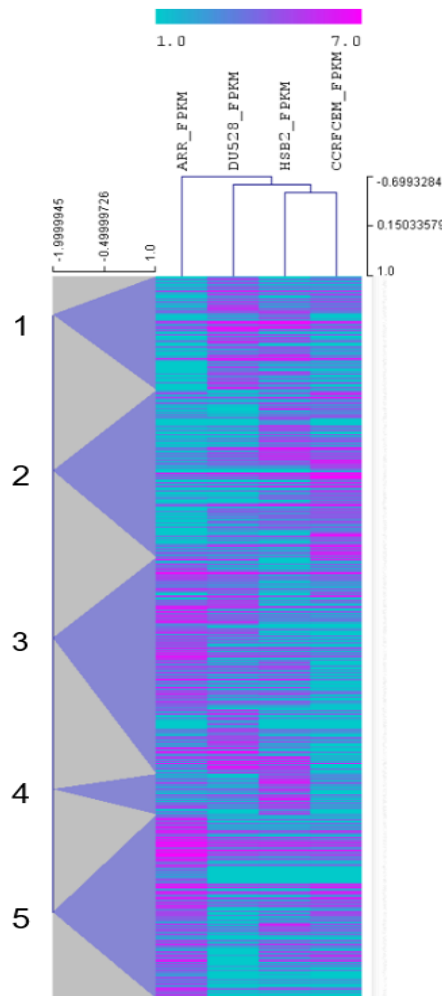
PHF6/LMO2/TAL1/LDB1 binding (Figure 3.34A). The gene ontology analysis of these common genes revealed involvement in leukocyte migration through CD84 and CD47. GO terms also showed signal transduction, transmembrane receptor protein tyrosine kinase signalling pathway and platelet activation. It also highlighted haematopoietic stem cell proliferation through RUNX1 and ETV6 (Figure 3.34B). Strikingly, *CD84*, *CD47*, *RUNX1* and *ETV6* have previously been found to be regulated by TAL1 in T-ALL (Sanda et al., 2012, Tan et al., 2019).



**Figure 3.34 PHF6/LMO2/TAL1/LDB1 common peaks in T-ALL cells are associated with genes involved in HSC proliferation.**

(A) Venn diagram show Gene ID overlaps of differentially expressed genes associated to PHF6/LMO2/TAL1/LDB1 shared peaks in T-ALL cell lines. (B) GO analysis on the 45 genes common in all T-ALL four cell lines. Table shows GO term biological processes, percentage of genes, P-value and gene names.

To illustrate the variation in gene expression between the cell lines a pairwise Pearson Correlation with complete linkage was calculated between the log2 FPKM values as described in section 3.2.4. Hierarchical clustering of these correlations in ARR identified five clusters (Figure 3.35). Some clusters such as 1 and 2 showed genes with lower expression in ARR compared to the other cell lines, while clusters 3, 4, and 5 showed genes with higher expression in ARR (Figure 3.35).



**Figure 3.35 RNA-seq analyses of ARR line.**

Heat maps showing hierarchical clustering of differentially expressed genes. These genes are either the nearest 5' or 3' gene or contain the PHF6/LMO2/TAL1/LDB1 peaks within the gene body. Scale bars represent colour index for the log2 FPKM values. Self-organising tree analysis identified 5 clusters.

The gene ontology analysis on genes with low levels of expression in clusters 2 showed their involvement in T-cell receptor signalling pathway, regulation of I- $\kappa$ B kinase/NF $\kappa$ B signalling, organ development, intracellular signal transduction and regulation of transcription (Figure 3.36A). The genes with elevated expression in cluster 3 were associated with haematopoiesis, cell proliferation, protein phosphorylation, and regulation of autophagy (Figure 3.36B).

**A**

GO term BP	%	Pvalue	Genes
T cell receptor signalling pathway	4.055	0.00975	<i>TNFRSF21, CARD11, RBCK1, GRAP2, PAG1, CD28</i>
positive regulation of I- $\kappa$ B kinase/NF $\kappa$ B signalling	4.055	0.013701	<i>MAVS, CARD11, CCR7, TRIM8, RBCK1, TRIM13</i>
pancreas development	2.028	0.014334	<i>ALDH1A2, WFS1, ILDR2</i>
O-glycan processing	2.703	0.015636	<i>XXYL1, GALNT6, POFUT1, C1GALT1</i>
intracellular signal transduction	6.082	0.025596	<i>PLCL1, MAGI3, BCR, SPSB1, DGKG, STK17B, HSPB1, PAG1, NFATC1</i>
protein O-linked glycosylation	2.028	0.031805	<i>TET3, GALNT6, POFUT1</i>
heart morphogenesis	2.028	0.031805	<i>ALDH1A2, ASXL1, PTCH1</i>
positive regulation of transcription, DNA-templated	6.757	0.037442	<i>NCOA3, ETS1, TAF8, KDM8, IRF1, PTCH1, RUNX1, TRERF1, WT1, NFATC1</i>
kidney development	2.703	0.039727	<i>ALDH1A2, WFS1, WT1, C1GALT1</i>
negative regulation of osteoblast differentiation	2.028	0.045676	<i>GDF10, PTCH1, HDAC7</i>

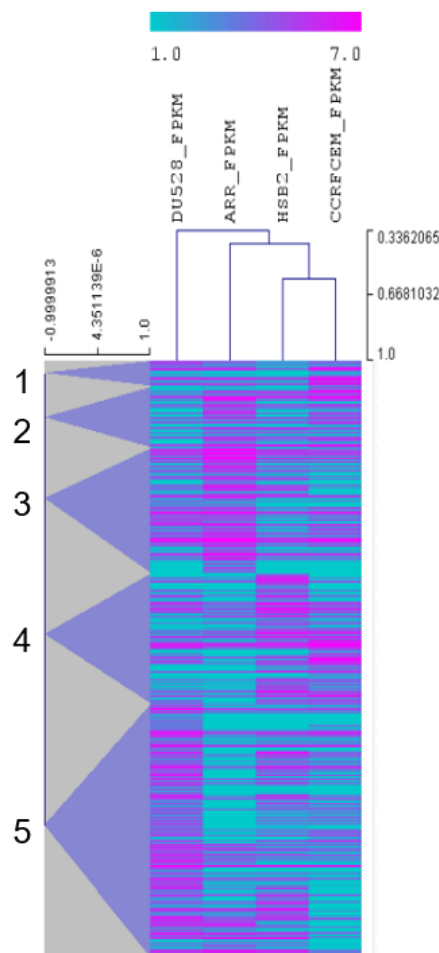
**B**

GO term BP	%	Pvalue	Genes
haematopoiesis	3.744	4.72E-05	<i>TAL1, GATA2, GFI1B, ZBTB16, CD164, RUNX2, ADD2</i>
positive regulation of receptor internalization	2.674	1.30E-04	<i>PLCG2, TBC1D5, AHI1, CD63, SYK</i>
signal transduction	14.974	1.97E-04	<i>LITAF, PDIA3, NEDD9, IGF1R, ANK2, CD69, NR2F6, CSF2RB, STAM, NDRG2, NR2F2, ERG, IRS2, PIK3CB, NDFIP2, PDE4D, CD164, STK3, CORO1C, GRB10, DOK2, P2RX1, CD33, ZDHHC13, BRE, CHN2, INPP4B, ARHGAP10</i>
cell proliferation	6.952	8.01E-04	<i>ERG, IRS2, PIM1, DUSP22, PIM2, ISG20, GFI1B, FAM83A, BCL2, FOXC1, RAP1B, AXIN2, SYK</i>
lymph vessel development	1.605	0.00327	<i>FOXC1, NR2F2, SYK</i>
transmembrane receptor protein tyrosine kinase signalling pathway	3.209	0.004284	<i>IGF1R, MTSS1, DOK2, PIK3CB, AHI1, SYK</i>
protein phosphorylation	6.952	0.004936	<i>ERG, TNK1, MMD, PIM1, AKAP13, MKNK1, TRIO, PIM2, STK3, RPS6KA5, APP, ERN1, SYK</i>
positive regulation of I- $\kappa$ B kinase/NF $\kappa$ B signalling	3.744	0.009135	<i>LITAF, ZDHHC13, F2RL1, TGM2, NDFIP2, AKAP13, PIM2</i>
positive regulation of autophagy	2.14	0.00978	<i>PIK3CB, ATG7, TBC1D5, PIM2</i>
cellular protein localization	2.14	0.011925	<i>ANK2, AHI1, CD63, AXIN2</i>

**Figure 3.36 GO analysis on differentially expressed genes associated with PHF6/LMO2/TAL1/LDB1 shared peaks in ARR cells show genes involved in regulation of transcription and haematopoiesis.**

GO analyses for biological processes performed on (A) cluster 2, (B) cluster 3 as identified in Figure 3.35. Table of GO term biological processes, percentage of genes, P-value and genes.

Hierarchical clustering in DU.528 cells showed five clusters (Figure 3.37). The genes in clusters 1, 2, and 3 showed similar levels of expression between DU.528, HSB2, and CCRF-CEM. Most of the genes with the higher levels of gene expression in DU.528 compared to the other cell lines, were grouped in cluster 5 (Figure 3.37).



**Figure 3.37 RNA-seq analyses of DU.528 cell line.**

Heat maps showing hierarchical clustering of differentially expressed genes. These genes are either the nearest 5' or 3' gene or contain the PHF6/LMO2/TAL1/LDB1 peaks within the gene body. Scale bars represent colour index for the log2 FPKM values. Self-organising tree analysis identified 5 clusters.



GO analysis of clusters 1 and 2 combined identified genes that are associated with regulation of I- $\kappa$ B kinase/NF $\kappa$ B signalling, activated T-cell proliferation, cellular response to UV light, regulation of MAPK cascade, and chromatin remodelling (Figure 3.38A). Many genes involved in regulating the activity and signalling pathways of enzymes essential for fundamental cellular processes such as GTPase and kinases were found in cluster 5. Also, genes linked to apoptotic process, leukocyte migration, and platelet formation were identified in cluster 5 (Figure 3.38B).

**A**

GO term BP	%	Pvalue	Genes
positive regulation of I- $\kappa$ B kinase/NF- $\kappa$ B signalling	6.452	0.020755	<i>FYN, PELI2, PRKCE, LGALS9</i>
activated T cell proliferation	3.226	0.025158	<i>SATB1, FYN</i>
cellular response to UV-B	3.226	0.028701	<i>MME, MFAP4</i>
positive regulation of ceramide biosynthetic process	3.226	0.032231	<i>PLA2G6, SMPD3</i>
positive regulation of MAPK cascade	4.839	0.035009	<i>PELI2, PRKCE, KSR1</i>
positive regulation of glial cell proliferation	3.226	0.035748	<i>MYC, ETV5</i>
chromatin remodelling	4.839	0.039042	<i>SATB1, SATB2, MYC</i>
regulation of release of sequestered calcium ion into cytosol	3.226	0.039253	<i>ANK2, PRKCE</i>

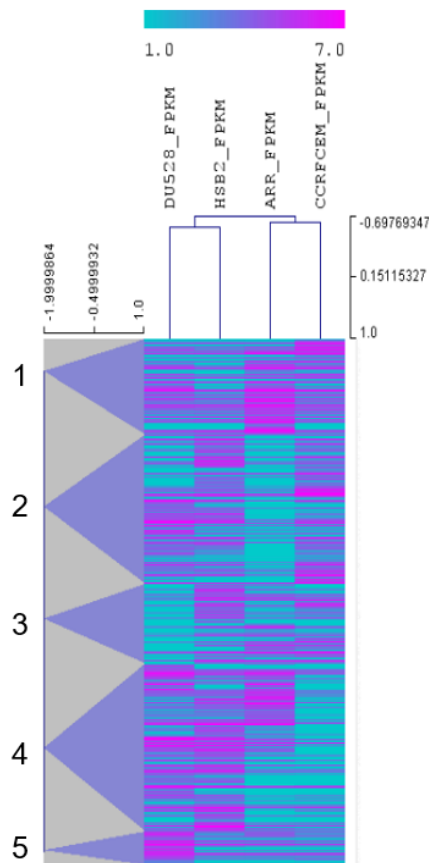
**B**

GO term BP	%	Pvalue	Genes
phosphatidylinositol 3-kinase signalling	3.247	9.88E-05	<i>ZFP36L1, IGF1R, PIK3CB, XBP1, PIK3R5, LIMS1, CYTH1, PREX1, CYTH4, DENND2D, ARHGEF12, ITGB1, FARP1, ARHGAP25, ANKRD27, NTRK1, CHN2, RGS9, NET1, FGD4, RIN3</i>
positive regulation of GTPase activity	10.39	2.05E-04	
transmembrane receptor protein tyrosine kinase signalling pathway	3.897	0.001847	<i>IGF1R, DOK2, PIK3CB, NTRK1, AHI1, LCP2, MEF2C, PRF1, CFLAR, PTPN6, DFFB, RMDN3, PIM1, ZBTB16, PIM2, BCL2L11, PLAGL1, ZFP36L1, CASP3, DAD1</i>
apoptotic process	9.091	0.002074	
leukocyte migration	3.897	0.005191	<i>CD84, PTPN6, DOK2, PIK3CB, F2RL1, ITGB1</i>
anterior/posterior pattern specification	3.247	0.006131	<i>HOXB3, HOXB4, HOXB6, ZBTB16, NR2F2</i>
regulation of Rho protein signal transduction	3.247	0.006405	<i>PREX1, ARHGEF12, FARP1, FGD4, NET1</i>
intracellular signal transduction	6.494	0.01165	<i>TRAF3IP2, MAST4, PTPN6, PREX1, CHN2, FER, RGS9, ARHGEF12, LCP2, NET1</i>
platelet formation	1.949	0.012741	<i>MEF2C, PTPN6, CASP3</i>
cellular response to peptide hormone stimulus	1.949	0.014073	<i>ZFP36L1, MAX, XBP1</i>

**Figure 3.38** GO analysis on differentially expressed genes associated with PHF6/LMO2/TAL1/LDB1 shared peaks in DU.528 cells show genes involved in lymphocyte processes, regulation of MAPK cascade, and chromatin remodelling.

GO analyses for biological processes performed on (A) cluster 1 and 2, (B) cluster 5 as identified in Figure 3.37. Table of GO term biological processes, percentage of genes, P-value and genes.

The five clusters identified in HSB2 cells showed a variety of high and low levels of gene expression per cluster (Figure 3.39). Mainly, cluster 1 contained genes with lower expression than the other cells, whereas cluster 4 showed some genes with higher expression than ARR and CCRF-CEM (Figure 3.39).



**Figure 3.39 RNA-seq analysis of HSB2 cell line.**

Heat maps showing hierarchical clustering of differentially expressed genes. These genes are either the nearest 5' or 3' gene or contain the PHF6/LMO2/TAL1/LDB1 peaks within the gene body. Scale bars represent colour index for the log2 FPKM values. Self-organising tree analysis identified 5 clusters.

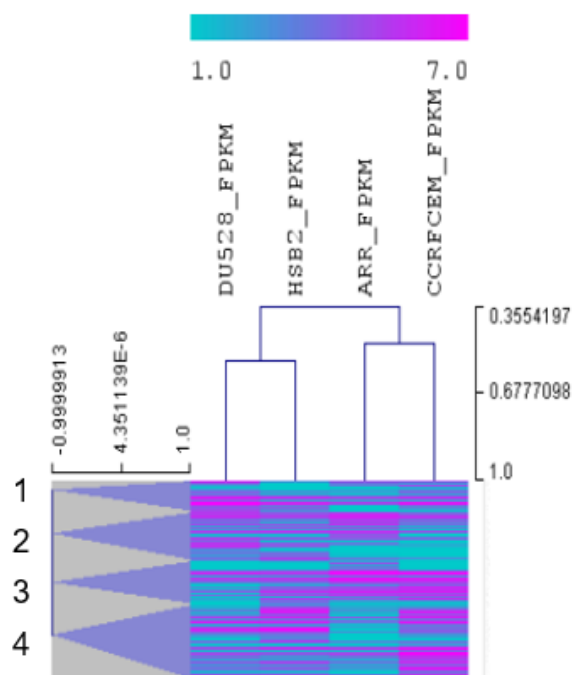
The GO terms of cluster 2 revealed several genes associated with signal transduction, B-cell processes, regulation of transcription, and haematopoiesis (Figure 3.40A). In cluster 4 there were genes involved in protein phosphorylation, signal transduction, regulation of ERK1 and ERK2 cascade, and stem cell differentiation (Figure 3.40B).

<b>A</b>	GO term BP	%	Pvalue	Genes
	positive regulation of signal transduction	4.445	0.004149	VAV3, SH2B3, SH3BP2, SLA
	intracellular signal transduction	8.889	0.00555	FYB, PTPN6, PRKCQ, GNB1L, SPSB1, SH2B3, RGS9, DAPK1
	positive regulation of phosphatidylinositol 3-kinase activity	3.334	0.01169	CCR7, VAV3, FLT1
	Fc-epsilon receptor signalling pathway	5.556	0.01483	CARD11, PRKCQ, VAV3, NFKB1, NFATC3
	negative regulation of vitamin D biosynthetic process	2.223	0.021034	NFKB1, GFI1
	negative regulation of B cell receptor signalling pathway	2.223	0.021034	PTPN6, LPXN
	apoptotic process	8.889	0.030806	TNS4, DNASE1, PTPN6, BRMS1, DFFB, TNFAIP8, NFKB1, DAPK1
	regulation of B cell differentiation	2.223	0.031387	CARD11, PTPN6
	negative regulation of transcription from RNA polymerase II promoter	10	0.036965	BRMS1, ATN1, WFS1, NFKB1, PTCH1, GFI1, ETV6, WT1, HDAC7
	haematopoiesis	3.334	0.039105	RTKN2, SH2B3, GFI1
<b>B</b>	GO term BP	%	Pvalue	Genes
	protein phosphorylation	10.834	9.54E-05	TNIK, STK24, PIM1, STK17B, MMD, AKAP13, PRKCH, ST3GAL1, RPS6KA5, APP, MAP3K8, STK38L, MATK
	intracellular signal transduction	10	1.37E-04	SH3BP5, RPS6KA5, PLCL2, TNK1, PREX1, PREX2, STK17B, CHN2, PRKCH, AKAP13, STK38L, LCP2
	transmembrane receptor protein tyrosine kinase signalling pathway	5	6.01E-04	IGF1R, DOK2, PIK3CB, AHI1, LCP2, MATK
	positive regulation of GTPase activity	10.834	6.73E-04	SNX9, IL2RB, PREX1, PREX2, CYTH4, AKAP13, ITGB1, FARP1, ARHGAP25, RGS10, ARRB1, RASGRP1, CHN2
	trophoblast giant cell differentiation	2.5	0.003138	LIF, SOCS3, NR2F2
	protein autophosphorylation	5	0.007606	EPHA4, IGF1R, TNK1, STK24, PIM1, STK17B
	signal transduction	14.167	0.007824	IL2RB, STK24, PIK3CB, PRKCH, CXXC5, ARHGAP25, SH3BP5, IGF1R, DOK2, ARRB1, RASGRP1, IL4R, RALB, CHN2, STAM, VOPP1, NR2F2
	negative regulation of ERK1 and ERK2 cascade	3.334	0.008066	LIF, NLRP6, RANBP9, ARRB1
	stem cell differentiation	2.5	0.015572	LIF, PHF19, RUNX2
	cardiac muscle cell differentiation	2.5	0.015572	AKAP13, ITGB1, FOXP1

**Figure 3.40 GO analysis on differentially expressed genes associated with PHF6/LMO2/TAL1/LDB1 shared peaks in HSB2 cells show genes involved in B-cell processes, haematopoiesis, and stem cell differentiation.**

GO analyses for biological processes performed on (A) cluster 2, (B) cluster 4 as identified in Figure 3.39. Table of GO term biological processes, percentage of genes, P-value and genes.

Hierarchical clustering in CCRF-CEM cells showed four clusters, where genes with lower gene expression were grouped in clusters 1 and 2. Conversely, clusters 3 and 4 contained genes with higher gene expression compared to ARR and DU.528 (Figure 3.41). Because of the limited number of peaks that were called in CCRF-CEM cells, a relatively small number of differentially expressed genes were associated with them. These 114 genes in all 4 clusters were involved in cell surface receptor signalling pathway, regulation of GTPase activity, cell migration, vascular endothelial growth factor receptor signalling pathway, regulation of apoptotic process, and mitotic cell cycle (Figure 3.42).



**Figure 3.41 RNA-seq analysis of CCRF-CEM cell line.**

Heat maps showing hierarchical clustering of differentially expressed genes. These genes are either the nearest 5' or 3' gene or contain the PHF6/LMO2/TAL1/LDB1 peaks within the gene body. Scale bars represent colour index for the log2 FPKM values. Self-organising tree analysis identified 4 clusters.

GO term BP	%	Pvalue	Genes
cell surface receptor signalling pathway	7.018	0.00133	<i>DOK2, MAGI1, DTX1, TSPAN7, EVL, ADGRG1, ADAP1, ADGRG3</i>
positive regulation of GTPase activity	9.65	0.00212	<i>RCBTB2, VAV3, PREX2, ARHGAP19, DOCK9, CHN2, AKAP13, DENND2D, VAV2, IQGAP1, ADAP1</i>
cell migration	5.264	0.003819	<i>ERG, PIK3CB, FAT1, FOXC1, VAV2, ADGRG1</i>
leukocyte migration	4.386	0.006299	<i>CD84, CD47, CD244, DOK2, PIK3CB</i>
vascular endothelial growth factor receptor signalling pathway	3.509	0.009331	<i>VAV3, PIK3CB, FOXC1, VAV2</i>
positive regulation of apoptotic process	6.141	0.009512	<i>PNMA2, IGF2R, RASSF2, AKAP13, VAV2, CAMK1D</i>
signal transduction	13.158	0.00955	<i>CD244, ERG, PIK3CB, ARHGAP19, CXXC5, VAV2, IQGAP1, DOK2, CCND3, ANK2, DGKG, IGF2R, CHN2, STAM, RUNX1</i>
regulation of Rho protein signal transduction	3.509	0.012841	<i>VAV3, PREX2, AKAP13, VAV2</i>
transmembrane receptor protein tyrosine kinase signalling pathway	3.509	0.020164	<i>MTSS1, DOK2, PIK3CB, AHI1</i>
mitotic cell cycle	2.632	0.025225	<i>FER, MYB, PPP2R2A</i>

**Figure 3.42 GO analysis on differentially expressed genes associated with PHF6/LMO2/TAL1/LDB1 shared peaks in CCRF-CEM cells show genes involved in regulation of GTPase activity, vascular endothelial growth factor signalling pathway, and mitotic cell cycle.**

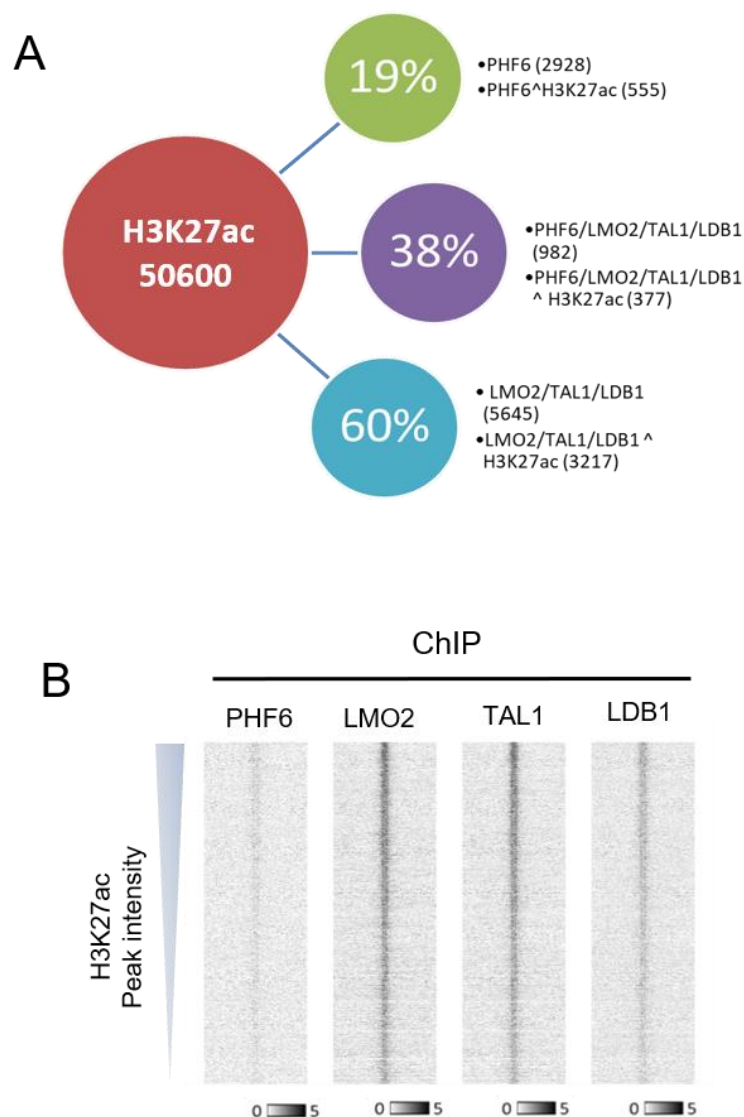
GO analyses for biological processes performed on all 4 clusters as identified in Figure 3.41. Table of GO term biological processes, percentage of genes, P-value and genes.

Altogether, these analyses showed differentially expressed genes that are possibly regulated by PHF6/LMO2/TAL1/LDB1 binding sites. Many of these genes encode for haematopoietic TFs such as TAL1, MYB, RUNX1, and GATA2 (Hoang et al., 2016, Baer, 1993, Imperato et al., 2015, Palis et al., 1999, Ciau-Uitz et al., 2013, Wilson et al., 2011). Moreover, several genes encoding for members of the ETS family of TFs were identified namely *ETS1*, *ERG*, *ETV5*, *ETV6* (Sharrocks, 2001). A reoccurring observation from GO analyses are genes associated with regulation of transcription, haematopoiesis, cell proliferation, and lymphocyte cell processes. Collectively this indicates a transcriptional regulatory function of PHF6 with the LMO2 complex.

### **3.2.16 PHF6 is associated with active enhancers in DU.528 cells**

The ChIP-seq analyses in T-ALL cell lines indicated a transcriptional regulatory function for PHF6 and the genomic regions bound PHF6/LMO2/TAL1/LDB1 common peaks were linked to genes involved in haematopoiesis and transcriptional regulation. Taken that many of PHF6/LMO2/TAL1/LDB1 common peaks were located far from TSSs, we speculate that these factors bind enhancers rather than promoters. Gene promoters mostly reside at TSSs where TFs can bind and regulate gene expression (Suzuki et al., 2001). The histone 3 lysine 27 acetylation (H3K27ac) marks both active promoters and distal enhancers (Ernst et al., 2011, Creighton et al., 2010). Therefore, we sought to compare the ChIP-seq data of DU.528 cell line with published ChIP-Seq datasets for H3K27ac enriched DNA (Abraham et al., 2017).

The histone mark H3K27ac sequence reads were generated from regulatory sites in DU.528 cells, providing a good indicator of active enhancer state. These adapted ChIP-Seq datasets showed 50600 H3K27ac enriched regions (Abraham et al., 2017). The percentage of PHF6 which occupied these regions was 19% of the total PHF6 peaks (Figure 3.43A). Comparably, 38% of PHF6/LMO2/TAL1/LDB1 common peaks were enriched for H3K27ac. A higher percentage *i.e* 60% of overlap was found between the regions bound by LMO2/TAL1/LDB1 and H3K27ac modified regions (Figure 3.43A). This was also reflected on the heat maps where PHF6 showed less binding to H3K27ac marked regions. LMO2 and TAL1 showed stronger binding to these regions while, LDB1 showed slightly less occupancy (Figure 3.43B). This indicated that the regions bound by PHF6 are not highly associated with enhancer rich H3K27ac histone marks.

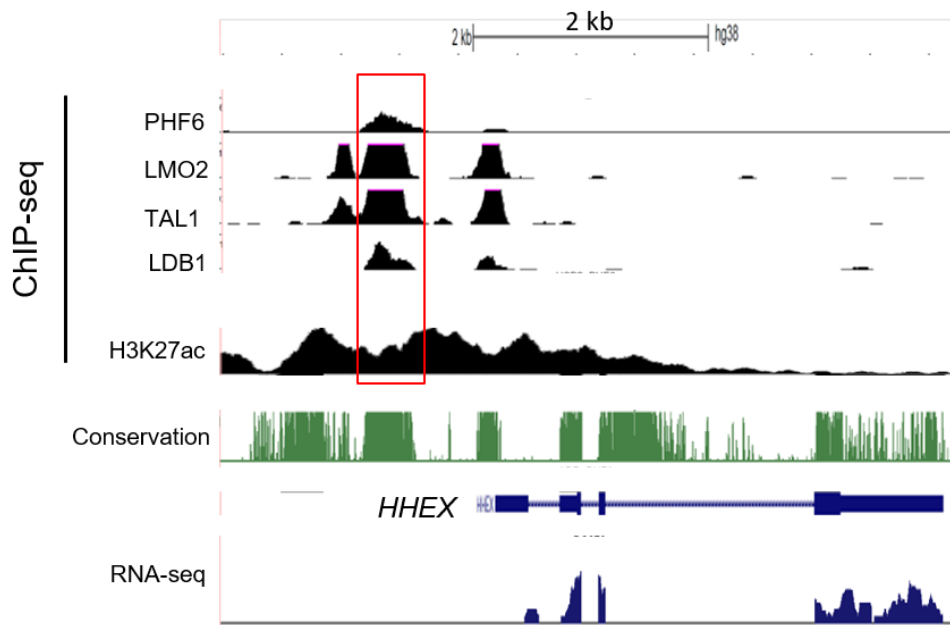


**Figure 3.43 PHF6 peaks bind H3K27ac enriched regions in DU5.528 cells.**

(A) A diagram showing the number of H3K27ac enriched regions (red circle), the percentage of PHF6 bound regions marked with H3K27ac (green circle) with the total number of PHF6 enriched regions and number of overlapping regions with H3K27ac indicated, the percentage of PHF6/LMO2/TAL1/LDB1 bound regions marked with H3K27ac (purple circle) with the total number of PHF6/LMO2/TAL1/LDB1 enriched regions and number of overlapping regions with H3K27ac indicated, the percentage of LMO2/TAL1/LDB1 bound regions marked with H3K27ac (blue circle) with the total number of LMO2/TAL1/LDB1 enriched regions and number of overlapping regions with H3K27ac indicated. (B) Heat maps of H3K27ac ChIP-seq data (region sets) ranked according to peak intensity score in descending order. The plots showing ChIP-seq data of PHF6, LMO2, TAL1, and LDB1 overlaid on H3K27ac ChIP-seq region-sets. The Y-axis represents individual positions of the regions, and the X-axis represents -2 kb to +2 kb window centred on the summits of H3K27ac peaks. At the bottom of each plot a bar showing the relationship between colouring and signal intensity.



There were PHF6/LMO2/TAL1/LDB1 peaks at H3K27ac enriched active enhancers in DU.528 cells. One example is the *HHEX* enhancer where PHF6 co-localised with LMO2 complex and H3K27ac was enriched (Figure 3.44). The RNA-seq track showed *HHEX* expression, which supports the notion that the *HHEX* enhancer is in an active state (Figure 3.44). Previous studies have reported *HHEX* as a direct transcriptional target of the LMO2 complex in human T-ALL (Cleveland et al., 2013, Oram et al., 2010, Smith et al., 2014). This supports the gene ontology analysis where GO terms identified genes that are involved in haematopoiesis and transcriptional regulation as potential targets of PHF6/LMO2/TAL1/LDB1 common peaks. Importantly, this gives a strong indication that PHF6 can bind with the LMO2 complex members at enhancer associated regions and activate gene expression of TFs that play important roles in transcriptional regulatory networks involved in haematopoietic development.

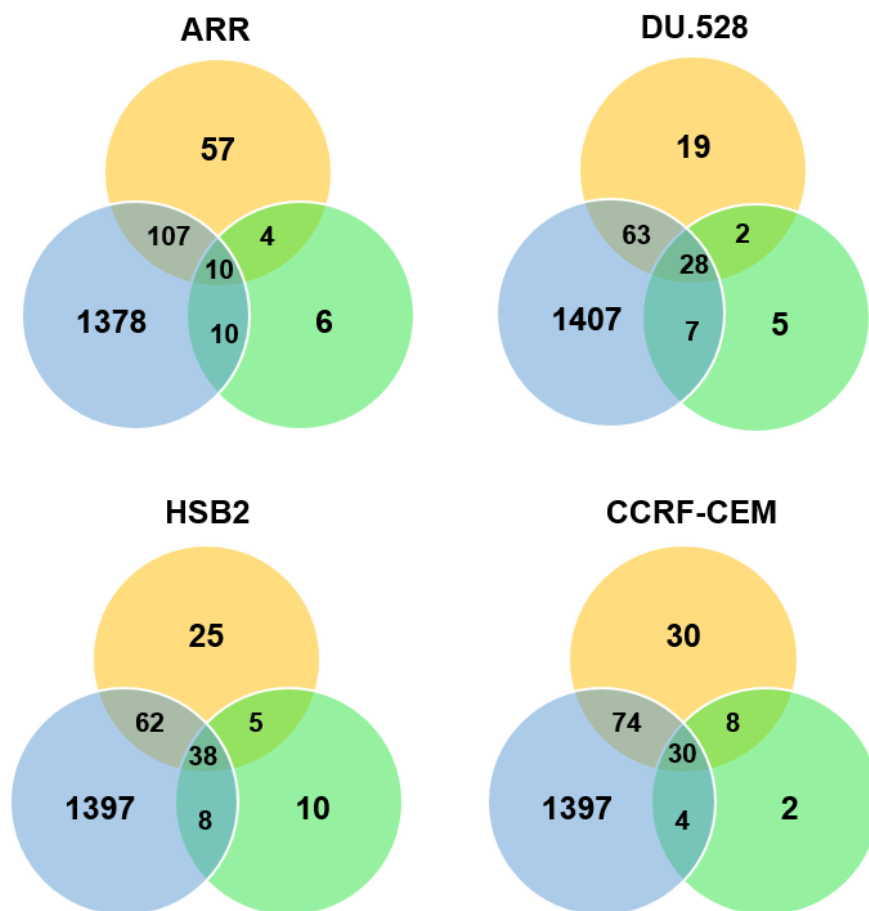


**Figure 3.44 Binding profiles of PHF6, LMO2, TAL1, LDB1, and H3K27ac in DU.528 cells.**

Screenshot from the UCSC browser showing ChIP-seq binding profiles at *HHEX* enhancer (red box), and the distribution of RNA-seq reads over the *HHEX* locus.

### 3.2.17 Identification of PHF6 interacting partners by mass spectrometry

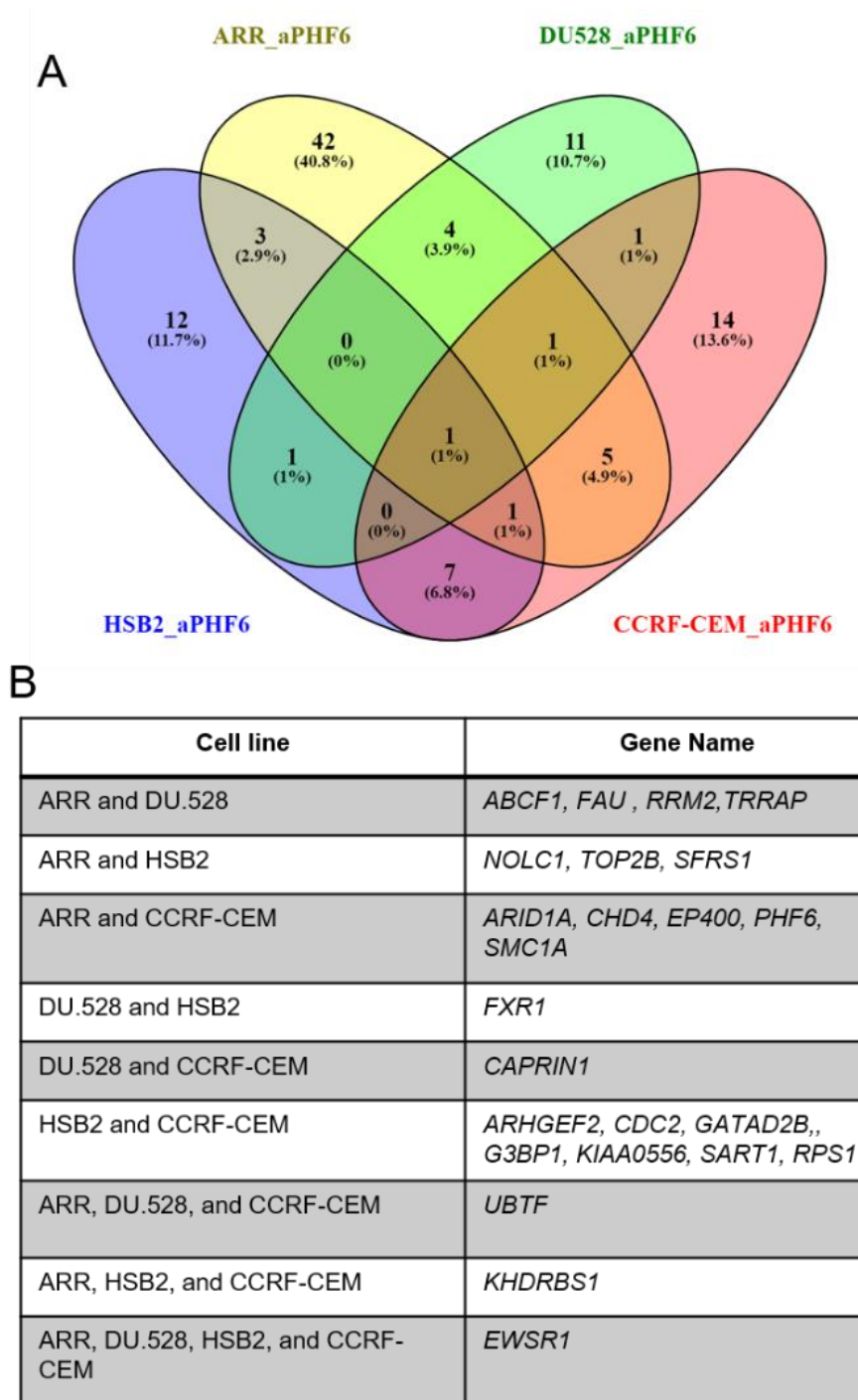
Having corroborated the interaction between PHF6 and LMO2, we next investigated the interacting partners of PHF6 through proteomic analysis. Antibodies against PHF6 and IgG (negative control) were used for the immunoprecipitations in all four T-ALL cell lines, which was followed by mass spectrometry. The data analysis was carried out as described in section 3.2.5. The numbers of peptides that coprecipitated exclusively with PHF6 were 57 in ARR, 19 in DU.528, 25 in HSB2, and 30 in CCRF-CEM (Figure 3.45).



**Figure 3.45 Mass Spectrometry assay, T-ALL cell lines immunoprecipitation with PHF6 and IgG antibodies.**

Venn diagrams showing overlap of proteins pulled down with PHF6 (yellow circle) versus IgG (green circle) and common contaminants (blue circle) in T-ALL cells.

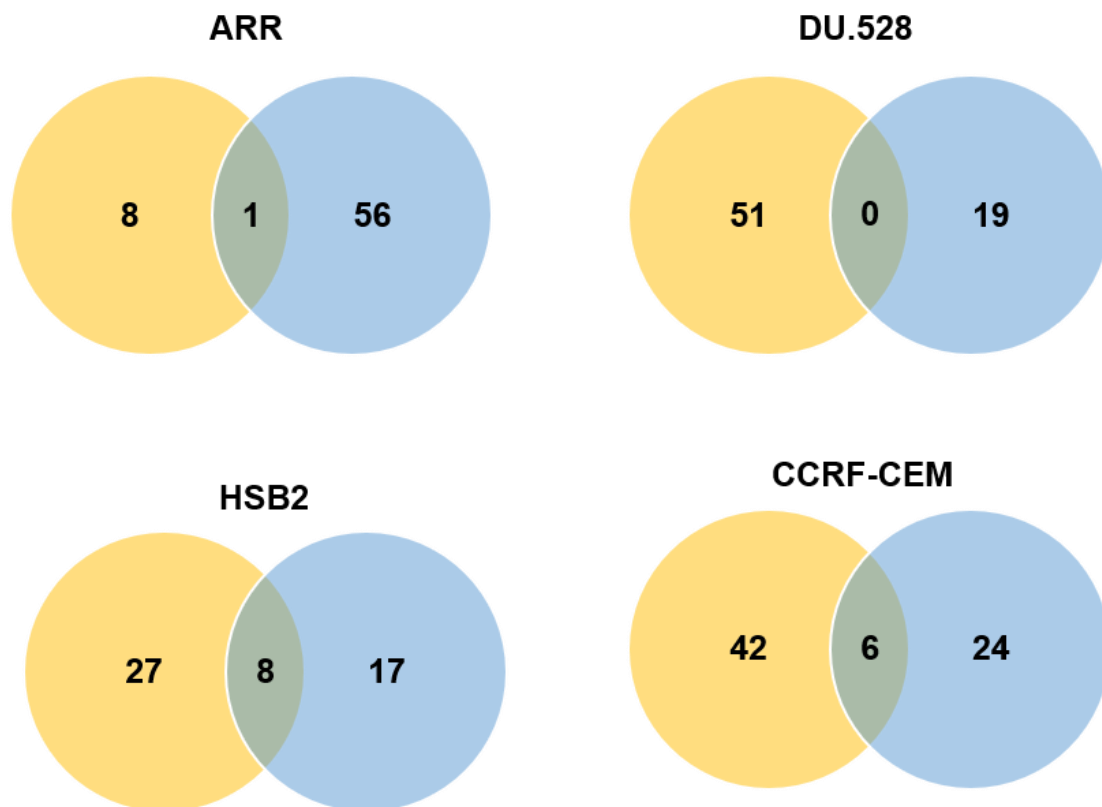
To investigate the proteins shared between cell lines, the gene names encoding for proteins that are specifically pulled down with PHF6 per cell line were intersected and the proteins that were common in two or more of the cell lines were selected (Figure 3.46A). This data showed PHF6 interaction with multiple components of the NuRD chromatin remodelling complex including CHD4, HDAC1, and RBBP4 (Todd and Picketts, 2012), in addition to the upstream binding factor UBF1 encoded by the *UBTF* gene (Wang et al., 2013). These proteins were not immunoprecipitated with PHF6 in all cell lines (Figure 3.46B). One protein that coprecipitated with PHF6 in all four cell lines was EWSR1 (Figure 3.46B). The EWS RNA Binding Protein 1 (EWSR1) protein is involved in RNA binding, transcription, RNA metabolism and rearrangements of *EWSR1* gene are associated with soft tissue tumours (Fisher, 2014).



**Figure 3.46 Mass Spectrometry analyses show coprecipitation of ESWR1 with PHF6 in all four cell lines.**

(A) Venn diagram showing Gene ID overlaps of proteins pulled down with PHF6 in T-ALL cells. (B) Table of gene names of proteins common in two cell lines or more.

To further explore the peptides in common between PHF6 and LMO2, the proteins specifically coprecipitated with LMO2 and PHF6 in each cell line were intersected. The Venn diagrams revealed 1 protein common in ARR which was PHF6. The intersection in HSB2 and CCRF-CEM showed 8 and 6 peptides in common, respectively (Figure 3.47). EWSR1 was identified as potential interactor of both PHF6 and LMO2 in HSB2 and CCRF-CEM cells. Finally, this suggests that EWSR1 interacts with PHF6 and LMO2 in these cells and further investigations may verify the interaction and study whether it plays role in T-ALL leukemogenesis.



**Figure 3.47 Venn diagrams showing overlap of proteins pulled down with LMO2 (yellow circle) versus PHF6 (blue circle) in T-ALL cells.**

### 3.3 Discussion

The existence of specific molecular subtypes of T-ALL has long been established (Ferrando et al., 2002, Homminga et al., 2011), but there has been no profound description of functional differences between them. Therefore, high-dose chemotherapy treatment regimens are applied uniformly across subtypes. Although cure rates of leukaemia have increased, it is usually coupled with high toxicity side effects (Pui et al., 2008, Liu et al., 2011). The TAL/LMO is the largest subgroup, accounting for 40–60% of all cases. Aberrant overexpression of LMO2, in concert with TAL1 leads to accumulation of immature thymocytes and a block in their differentiation (Zhang et al., 2012a, Homminga et al., 2011). Although several types of genetic alterations were observed across different subgroups of T-ALL, some abnormalities do not occur simultaneously (Belver and Ferrando, 2016, Van Vlierberghe et al., 2008). As an example, mutations of the *PHF6* gene were reported in T-ALL, but were largely mutually exclusive with the LMO2/TAL1 subgroup (Van Vlierberghe et al., 2010, Van Vlierberghe et al., 2011, Li et al., 2016, Wang et al., 2011, Vicente et al., 2015). In this study we describe functional commonalities and differences between the genes that are associated with the LMO2/TAL1 complex binding sites. Furthermore, we report a novel interaction between LMO2 and PHF6 which showed a potential transcriptional regulatory function.

#### 3.3.1 ChIP experiments confirmed LMO2/TAL1 binding to regulatory elements

The LMO2/TAL1 complex has been described in haematopoiesis and particularly erythropoiesis, however little is known about it in T-ALL and a single model has been demonstrated in T-cell cancer. The described model is comprised of LMO2, LDB1 and can bind DNA through two E-box motifs including TAL1 (Wadman et al., 1997, Grutz

et al., 1998). We confirmed the known interaction between LMO2 and TAL1 (Larson et al., 1996) in four human T-ALL cell lines ARR, DU.5228, HSB2, and CCRF-CEM. The latter three cell lines possess *SIL/TAL* deletion that leads to the abnormal expression of *TAL1* under the *SIL* gene regulatory element (Cave et al., 2004, Janssen et al., 1993). All the four cell lines have immature T-cell features according to the expression of CD markers (Sandberg et al., 2007).

We investigated binding of the LMO2/TAL1 complex throughout the genome by ChIP-seq experiments. In our attempt to find out whether the overexpression of TAL1 due to the *SIL/TAL1* mutation affects the binding of the complex to the DNA, we observed that LMO2/TAL1 complex localise at the same motifs in ARR and SIL/TAL1 cells. The enriched motifs were mainly RUNX1, GATA and ETS which reside at regulatory element regions (Nottingham et al., 2007, Anderson et al., 1998, Boros et al., 2009, Sanda et al., 2012). In a previous study the genomic regions occupied by LMO2 and TAL1 in CCRF-CEM cells have been investigated and *de novo* motif analyses of these bound regions identified E-box, GATA and RUNX motifs as well as known motifs for the ETS family of transcription factors (Sanda et al., 2012).

Furthermore, our ChIP confirmed binding of LMO2 and TAL1 to regulatory elements known to regulate transcription factors involved in haematopoiesis, such as *RUNX1* and *SPI1* (Wilson et al., 2010). Our data also showed co-localisation of LDB1 at the *RUNX1* +23 kb enhancer (Nottingham et al., 2007). This is in agreement with Sanda *et al.* which showed TAL1 and LMO2 binding at this *RUNX1* +23 kb enhancer in CCRF-CEM cells (Sanda et al., 2012). This indicated that the aberrant expression of LMO2 and TAL1 in T-ALL cells triggers the formation of a complex similar to the one described in haematopoietic progenitors that may regulate the expression of

transcription factors that play roles in haematopoietic lineage commitment and differentiation.

### **3.3.2 Commonalities and differences between gene expression in ARR and SIL/TAL1 cell lines**

LMO2 has a well-established function in transcriptional regulation via direct interaction with transcription factors including TAL1, LYL1, LDB1 and GATA proteins (Wadman et al., 1994, Lecuyer et al., 2007, El Omari et al., 2013). In this study, we sought to expand upon previous analyses of LMO2/TAL1 in T-ALL by identifying the target genes differentially expressed in T-ALLs, as well as their possible association with a distinct leukemic phenotype and/or outcome.

The genome wide data analysis showed specific and robust binding sites which strongly correlated with differences in gene expression shown by integrating ChIP-seq data with RNA-seq. Moreover, gene ontology analysis allowed the identification of the biological processes that are linked to genes that are differentially expressed between the cell lines and potential targets of the LMO2/TAL1 complex.

We observed four different groups of differentially expressed genes that were associated with genomic regions bound by LMO2 and TAL1 in ARR and SIL/TAL1 cells. In the first two groups were genes that associated with LMO2/TAL1 peaks that were either exclusively found in ARR cells, or in SIL/TAL1 cells. The GO analyses allowed for the identification of biological processes unique to cell lines. This showed that ARR cells might require regulation of genes specifically needed for RNA production and that they utilise metabolic pathways different from the SIL/TAL1 cells. However, three facts have to be taken into consideration (i) each cell line is blocked at



a different stage of differentiation, (ii) they are an *in vitro* model and the culture media conditions might not exactly mimic the *in vivo* environment leading to different patterns of gene expression and variable metabolic pathways, (iii) the LMO2/TAL1 complex does not have a function after the DN2 stage in normal T-cells. Hence, it is possible that the aberrant expression of LMO2/TAL1 complex is involved in the regulation of genes that are exclusive to cell lines and tailored according to the phenotypic requirements of each cell type. Because drug compounds may affect altered pathways between cell lines differently, the documentation of gene expression differences along signalling and metabolic pathways is important in drug discovery processes for T-ALL.

The third group included genes that associated with LMO2/TAL1 bound genomic sites in common between ARR and SIL/TAL1 cells. This group of genes were involved in regulation of translation, regulation of gene expression, T and B cell signalling pathways, and apoptosis. The fourth group had differentially expressed genes in common between the cell lines, but these genes were linked to different sites bound by LMO2/TAL1. These genes were essential for basic cellular functions such as leukocyte development and differentiation in addition to transcriptional regulation. The GO analyses in the latter two groups compared to the first groups, showed basic molecular and cellular processes that are essential for living cells. This suggest that regulation of these genes, irrespective of the binding sites of the LMO2/TAL1 complex, may be crucial for the survival of these leukemic cells. Additionally, many of these genes were associated with different peaks indicating that a single gene can possibly be regulated through multiple elements bound by this complex.

Importantly, these analyses revealed many differentially expressed genes encoding transcriptional regulators such as, *NOTCH1*, *ETV6*, *NKX3-1*, and *ARID5B*. Several

groups have identified *NKX3-1*, and *ARID5B* as TAL1 target genes in T-ALLs (Kusy et al., 2010, Leong et al., 2017, Sanda et al., 2012). It has been demonstrated that TAL1 regulates haematopoietic TFs including RUNX1, GATA3, MYB, and ETS family of genes (Palii et al., 2011, Sanda et al., 2012). There is also evidence that the ETS family transcription factor *ERG* gene is regulated through LMO2/TAL1 binding at the *ERG* enhancer region in primary human T-ALL cells (Thoms et al., 2011). Additionally, other ETS family members including, ETS1, ETV5, and ETV6 have been reported as TAL1 targets (Sanda et al., 2012).

On the other hand, Notch signalling is an important factor in T-cell development (Ciofani and Zuniga-Pflucker, 2005). Activating mutations of *NOTCH1* occur in over 50% of human T-ALL cases and many of TAL/LMO patients develop *NOTCH1* mutations (Breit et al., 2006, Mansour et al., 2007). In mouse models, activation of *Notch* cooperates with *Tal1/Lmo1* in inducing T-ALL (Tremblay et al., 2010). *NOTCH1* can activate the MYC oncogene in T-ALL cells, thereby driving cell proliferation, however the significance of MYC expression in LMO2/TAL1 T-ALLs remains unclear (Weng et al., 2006). Different studies argued that TAL1 and NOTCH1 regulate the same targets, suggesting that they compensate and potentiate each other to synergistically promote T-cell leukemogenesis (Joshi et al., 2009, Wouters et al., 2007, Sanda et al., 2012). Constitutive Notch signalling allows T-cell development to occur independent of the thymic cell niche (Pui et al., 1999). Therefore, Curtis and McCormack argued that because Notch signalling is not increased in immature thymocytes in *Lmo2* transgenic mice, *Notch* mutations may allow self-renewing thymocytes to escape the cell niche requirement of DN3 thymocytes prior to overt T-ALL (Curtis and McCormack, 2010). It is not certain that all these genes are

downstream targets of LMO2/TAL1 complex in the T-ALL cells used in this study, albeit finding them differentially expressed between the cell lines and associated with the complex binding sites. However, many of our results are in line with data from other studies, implicating a possible mechanism for the LMO2/TAL1 complex in the pathogenesis of T-ALL, through which it regulates the expression of key transcription factors and genes that play a role in maintaining an immature state of thymocytes with self renewal capacity.

### **3.3.3 PHF6 as a novel interactor of LMO2**

In this study, we sought to expand upon previous investigations of the LMO2 complex and its interacting partners in T-ALL by identifying novel members of the complex through mass spectrometry analysis. We discovered PHF6 and HDAC1 as new protein-protein interactors of LMO2. A previous study by Todd and Picketts presented coimmunoprecipitation of HDAC1 with PHF6 and proved the interaction of these two proteins with the NuRD complex which is a multifunctional epigenetic regulator (Todd and Picketts, 2012). However, we were interested in the interplay between PHF6 and LMO2. Vlierberghe *et al.* previously demonstrated high levels of *PHF6* transcript expression in the human thymus and PHF6 was also detected by immunostainings in mouse thymus. Consistently, variable levels of *PHF6* expression were detected in human T-cell progenitors at different stages of development (Van Vlierberghe et al., 2010). This supported the gene expression profiles of *PHF6* in our T-ALL cells, where RNA-seq data showed the expression of intact *PHF6* and western blots presented detectable levels of PHF6 protein in these cells. We also showed the interaction between LMO2 and PHF6 through two additional experimental techniques besides

mass spectrometry, including protein pull down assays followed by western blot, and proximity ligation assays.

Moreover, the immunostainings presented PHF6 and LMO2 localisation in the nuclei of T-ALLs. While previous studies showed the localisation of PHF6 in the nucleus and nucleolus of HeLa cells and HEK293T-cells (Todd et al., 2016, Todd and Picketts, 2012), we were able to visualise the localisation of PHF6 with LMO2 in the nuclei of T-cells, but not in the nucleoli. This indicated that the interaction is occurring preferentially within the nucleus. Similarly, the PHF6/NuRD interaction was only found within the nucleoplasm because the NuRD components localised to the nucleus with PHF6 and had very little nucleolar localisation (Todd and Picketts, 2012).

#### **3.3.4 PHF6, LMO2, TAL1, and LDB1 co-localise at regulatory elements and TF motifs**

Previous genome wide data of the LMO2 complex reported that LMO2, TAL1, and LDB1 form a complex in T-ALL that is similar to the one described in early haematopoietic progenitors and erythrocytes (Love et al., 2014, Nam and Rabbitts, 2006, Rabbitts et al., 1999, Curtis and McCormack, 2010). Indeed, our ChIP experiments supported these findings. Moreover, ChIP-seq analyses of PHF6 showed a remarkable number of overlapping peaks with the LMO2 complex members, particularly in ARR cells. Many of these bound regions were found at known gene regulatory elements, such as the *RUNX1* +23 kb enhancer. We observed that strong PHF6 peaks co-localised with LMO2, TAL1, and LDB1. Moreover, LDB1 co-occupied sites with the strongest PHF6 binding where LMO2 and TAL1 were not bound. This observation has yet to be further explored. The common peaks between all four factors also showed binding to accessible DNA regions. DHSs are known to possess *cis*-

regulatory elements, including enhancers and promoters (Lyu et al., 2018, Cockerill, 2011). This means that the complex access to these open chromatin regions may allow it to bind regulatory elements to regulate gene expression. Although TAL1 is more frequently observed at gene promoters near transcriptional start sites, LMO2 is rarely found close to TSSs by ChIP-seq in multipotent progenitors (Wilson et al., 2010). On the other hand, PHF6 was shown to occupy gene bodies and proximal promoter/enhancer regions (Meacham et al., 2015).

Our analyses showed that many PHF6/LMO2/TAL1/LDB1 common peaks were not at close proximity to TSSs. The majority of peaks were located at a distance farther than 50 kb from TSSs. This was not unexpected because enhancers can bypass neighbouring genes to regulate genes located more distantly along a chromosome and individual enhancers have been found to regulate multiple genes (Visel et al., 2009b, Ron et al., 2017). In addition, LMO2 is known to form a multi-protein complex where LDB1 multimerize, bringing distal enhancers along the DNA to close proximity to the target gene promoter through DNA looping, thereby facilitating activation of transcription (Cross et al., 2010, Matthews and Visvader, 2003). *De novo* motif analyses of all the common peaks and the peaks that were less than 50 kb away from TSSs revealed enrichment of TFs binding motifs. The three most significantly enriched haematopoietic TFs motifs were RUNX1, ETS, and GATA in addition to an E-box motif that was observed in ARR and CCRF-CEM cells. The over representation of these motifs, is highly similar to the LMO2 complex motifs identified in haematopoietic cells, erythrocytes, and T-ALL (Bach, 2000, Wilkinson-White et al., 2011, Sanda et al., 2012, Grutz et al., 1998).

Critically, the motif analyses of all PHF6 bound regions showed unknown long DNA sequences in ARR, DU5.528, and HSB2 cells. Similar findings were reported in B-ALL cells where PHF6 was unable to recognise specific DNA binding sequences, and the enriched motifs resembled the ones bound by known TFs but no significant *de novo* motifs were identified for PHF6 itself (Soto-Feliciano et al., 2017). The unknown motifs were not observed in CCRF-CEM cells. However, PHF6 ChIP-seq in these cells identified a low number of reads and only the strongest ones were recognised as peaks and almost all PHF6 binding sites showed co-localisation of LMO2, TAL1 and LDB1. The absence of defined DNA binding motifs for PHF6 indicates that PHF6 is not acting as a transcription factor in these T-ALL cells and is rather recruited to DNA by the LMO2 complex at open chromatin regions that harbour *cis*-regulatory elements. Therefore, it could potentially co-regulate gene expression. Supporting this notion, PHF6 was unable to recognise specific DNA binding sequence but showed interaction with the transcription factors TCF12 and NFκB, implying that PHF6 augments their capacity to activate transcription (Soto-Feliciano et al., 2017, Miyagi et al., 2019).

### **3.3.5 PHF6, LMO2, TAL1, and LDB1 common peaks are associated with haematopoietic TF genes**

The gene targets of the LMO2 complex have been investigated in many studies (Gerby et al., 2014, McCormack et al., 2010, Sanda et al., 2012, Lahlil et al., 2004, Kusy et al., 2010, Love et al., 2014, Gilmour et al., 2018). However, PHF6 has recently been implicated as a transcriptional regulator and became a protein of interest for genomic studies in haematopoiesis (Meacham et al., 2015, Wang et al., 2013, Miyagi et al., 2019, Soto-Feliciano et al., 2017). In concordance, the ChIP-seq and DNaseI-seq data

in this study suggested a transcriptional regulatory function for PHF6 cooperatively with the LMO2 complex.

Hence, we analysed the genome wide data through two different approaches. First, we looked at the GO terms of the genes that were allocated by GREAT analyses to PHF6/LMO2/TAL1/LDB1 common peaks that were close to TSSs; particularly less than 50 kb distant. Second, ChIP-seq data of the common peaks in each cell line were integrated with RNA-seq data and pairwise differential expression analyses between the four T-ALL cell lines were performed.

A recurrent GO term was regulation of transcription and was in common between ARR, DU.528, and HSB2 cells. This term highlighted many TF genes involved in haematopoiesis. For example, genes encoding STAT5A, MYB, TAL1, LYL1, RUNX1, HHEX, and NOTCH1 were found as potential targets of common peaks that were close to TSSs in these cell lines.

Genes including *HHEX*, *STAT5A*, and *LYL1* have known roles in HSC function (Paz et al., 2010, Schuringa et al., 2004, Souroullas et al., 2009). Upregulation of *Hhex*, *Lyl1*, *Stat5a*, and *Nfe2* were reported in transgenic mice with *Lmo2* induced DN thymocytes. Examining the relative expression of *LYL1* and *NFE2* in a data set of human T-ALL samples, including distinct subclasses expressing LYL1 or TAL1 revealed a significant association between the mouse *Lmo2* induced expression profile and LYL1 associated T-ALL cases (McCormack et al., 2010). Although in the latter study these genes were negatively correlated with TAL1 positive T-ALL patients, we found them as potential targets of the PHF6/LMO2/TAL1/LDB1 complex in our LMO2/TAL1 expressing T-ALL cells. In addition, genes such as *RUNX1* and *NOTCH* have been associated with LMO2

and TAL1 overexpression in T-ALL (Sanda et al., 2012, Curtis and McCormack, 2010), as discussed in section 3.3.2. These two genes were also reported as PHF6 targets in B-ALL cells. Suppression of *Phf6* led to a significant decrease in *Runx1* and *Notch* levels. Moreover, ChIP-seq data generated from human T-ALL cells presented PHF6 binding sites at the promoters of *RUNX1* and *NOTCH1* (Meacham et al., 2015). We tried to reproduce these ChIP experiments in T-ALL cells, however, PHF6 was not enriched across these sites by qPCR. The ChIP data in our study showed binding of PHF6 and LMO2 complex members at *RUNX1* enhancer but not at the reported promotor (Meacham et al., 2015). A possible reason for having different outcomes could be due to the different systems used. Jurkat cells express LMO1 instead of LMO2 (Palomero et al., 2006), but we were not able to regenerate their data even in this cell line. Looking at their published data, we were able to align the reads and see a small enrichment on active promoters, but the program with normal setting was not able to call peaks. Critically, active promoters are located on active chromatin regions which are easier to sonicate therefore ChIP-seq analysis could falsely call peaks at these sites.

Another two independent studies revealed enrichment of PHF6 across the entire rDNA coding sequence and the rDNA promoter by ChIP qPCR (Todd et al., 2016, Wang et al., 2013). We used some of the primer pairs reported in these studies in addition to our own designed primers against the rDNA region to validate ChIP experiments. The limitations to these primer sets were that rDNA sequence is highly repetitive and present at multiple loci, therefore the primers aligned to multiple complementary regions throughout the genome. Thus, the melting curves showed peaks at multiple melting temperatures. The melting curves with multiple products would make the



calculation invalid. Therefore, we did not find PHF6 binding at these regions because we could not accurately calculate it due to the very low cT values of the standards.

The GO terms in CCRF-CEM cells showed T-cell processes and mitotic cell cycle. These terms revealed *MYB* as potential target for PHF6/LMO2/TAL1/LDB1 common peaks. It is noteworthy that LMO2, TAL1, MYB, and RUNX1 regulate each other in HSCs, thus forming a connected autoregulatory loop structure, which is likely to be essential for the maintenance of gene expression programming (Wilkinson and Gottgens, 2013). The autoregulatory loop of these TFs has also been demonstrated in T-ALL cells including CCRF-CEM (Sanda et al., 2012). Moreover, the genes regulated after knockdown of TAL1 in these cells closely overlapped with those after knockdown of *MYB*, and *RUNX1* (Sanda et al., 2012). Therefore, these complexes of transcription factors may be arranged in an autoregulatory loop to regulate each other and coordinately regulate downstream gene expression in T-ALL cells.

Generally, our data implicate that the LMO2 complex recruits PHF6 and possibly drives a transcriptional program which regulates the transcription of several HSC-associated genes and other haematopoietic TF genes that are involved in blood lineage commitment and development. There were similarities and differences between the cell lines regarding the potential target genes of the complex. This limited the insight into the exact mechanism of how LMO2 complex is facilitating the leukemogenesis in these T-ALL cells. Some of the candidate target genes were HSC associated and therefore their upregulation may be responsible for the self-renewal capacity observed in T-ALL. Other potential targets were genes encoding TFs involved in haematopoietic transcriptional regulation, including genes with established roles in leukaemia, and T-cell processes such as development and activation. Fine tuning of these genes

assures normal development of T-cells which is disrupted in T-ALLs. Because LMO2 and TAL1 are normally silenced during T-cell maturation, the complex may not have targets in normal T-cells past lineage commitment. This suggests one possible mechanism that contribute to the pathogenesis of T-ALL, where the aberrant expression of LMO2 in T-ALL cells triggers the formation of the LMO2 complex and causes abnormal activation of target genes.

Observations in DU.528 cells showed that many of the PHF6/LMO2/TAL1/LDB1 common peaks did not correlate to enhancer sites marked with H3K27ac therefore they may be located at distal H3K4me1 regions. Genes can be associated with multiple enhancers that are either H3K4me1 positive, H3K27ac positive, or a combination of both (Creyghton et al., 2010). Most of H3K4me1 marked enhancer regions tested in reporter assays displayed activity, but also a significant percentage was not active (Heintzman et al., 2009, Visel et al., 2009a). Thus, it has been suggested that H3K4me1 can mark active enhancers as well as those in a poised or predetermined state. The lineage-specific enhancers that become activated upon differentiation are protected from DNA methylation in the parental cells, where the enhancers are not yet active (Xu et al., 2009). In fact, several studies implied that enhancers contain information about the current and future developmental potential of a cell (Xu et al., 2007, Xu et al., 2009).

This suggests another possible mechanism through which LMO2 complex induces leukemogenesis in T-cells, where LMO2 complex members with PHF6 bind active enhancers of genes that are expressed in progenitor cells in addition to poised enhancers serving as “placeholders” for T-cell specific TFs. When the T-cell specific TFs are upregulated during T-cell differentiation they bind to the enhancers of T-cell

genes which were reserved by LMO2 complex. Because LMO2 is aberrantly expressed in T-cells it may occupy enhancers of genes essential for T-cell development and keeping them in an inactive state while its binding needs to be replaced by other factors to activate these genes. At the same time, it is binding active enhancers of genes that are expressed in T-cell progenitors such as *HHEX* and maintaining their expression. Therefore, the thymocytes fail to upregulate genes that are T-cell specific and downregulate progenitor genes. Overall, the observations throughout this study suggest that the LMO2 complex and its members including the novel interactor PHF6, form a transcriptional regulatory network that drives self-renewal, proliferation, and survival of leukemic thymocytes.

A limitation to this study was the inability to knockdown *PHF6* in the utilised cell lines. Several shPHF6 knockdown methods were used including, lentiviral mediated transduction and electroporation of integrating vectors based on transposase activity. These attempts were unsuccessful, even though the virus efficiency was validated in other cell lines and showed successful results. Additionally, the cells were not able to survive beyond the integration of shPHF6 by electroporation in comparison to cells integrated with GFP which raise the possibility that PHF6 is crucial for the survival of these cells. For a future study, dissection of the generated data for each cell line with a comparison to LMO2 and PHF6 knockdown phenotypes would shed more light into specific mechanisms that contribute to the pathogenesis of the oncogenic overexpression of LMO2 in T-ALL. It will also help in the identification of specific targets of LMO2 and PHF6 individually, as well as LMO2/PHF6 complex.

### 3.3.6 PHF6 has multiple interacting partners

As mentioned above, it has previously been reported that PHF6 can interact with components of the NuRD complex (Todd and Picketts, 2012). Likewise, we found the interactions between PHF6 and members of the NuRD complex RBBP4, HDAC1 and CHD4. Given that components of the NuRD complex, *i.e.* RBBP4 and CHD4 immunoprecipitated with PHF6 only but not LMO2 indicated that PHF6 may be interacting with NuRD independently. However, further studies should consider investigating the participation of LMO2 in the PHF6/NuRD complex because NuRD complex has multiple functions, including transcriptional regulation, chromatin remodelling, and DNA repair. Hence, exploring whether LMO2 participate in such activities other than transcriptional regulation can give more clues about its oncogenic function in T-ALL.

Furthermore, we found an interaction between UBF1, an rDNA transcriptional activation factor that associates with the RNA Pol I pre-initiation complex (Learned et al., 1986). Wang *et al.* reported that UBF1 recruits PHF6 to the rDNA promoter and represses rRNA transcription. They showed that loss of PHF6 results in increased genomic instability at rDNA genes, accumulation of  $\gamma$ H2AX DNA damage marker and a cell cycle delay at G2/M (Wang et al., 2013). In our study, EWSR1 immunoprecipitated with PHF6 in all T-ALL cell lines. In addition, EWSR1 co-precipitated with LMO2 in HSB2 and CCRF-CEM. EWSR1 belongs to a family of RNA binding proteins, the encoding gene is known to translocate with several partner genes encoding TFs which are associated with sarcoma and leukaemia such as the ETS transcription family members (Erkizan et al., 2010, Delattre et al., 1992). EWSR1 is a well-known player in cancer biology for the specific translocations occurring in solid

tumours including sarcomas (Paronetto, 2013). The exact function of EWSR1 in cancer is not clear but a role in DNA repair and cell-cycle progression has been suggested for this protein (Paronetto, 2013, Li et al., 2007). It is possible that PHF6 and LMO2 are team players in different complexes according to the cell type therefore exerting multiple roles including transcriptional regulation and genome stability. Understanding these functions and how they contribute to the progression of cancer will help to provide better approaches for identifying new therapeutic strategies.

## 4 Chapter 4: The role of PHF6 in myeloid cells

### 4.1 Introduction

A major aim of this study was to understand the role of LMO2 and PHF6 expression and their interaction in haematopoietic cells and whether the PHF6 and LMO2 interaction exerts a regulatory function in these cells. Moreover, we wanted to understand whether they target genes that are involved in important biological processes for differentiation, proliferation, or cell survival. Because we experienced technical difficulties to knockdown PHF6 and LMO2 in T-ALL cells, as discussed in the previous chapter, we expanded our investigations by using an easy to manipulate myeloid mouse system. In this chapter the interaction between PHF6 and LMO2 along with known LMO2 interacting partners, such as TAL1 and LDB1, in mouse myeloid progenitors is explored. The differences between wild type and *Phf6* knockdown myeloid cells in terms of differentiation, gene expression, and LMO2 complex interactions are also investigated. Additionally, chromosomal stability is compared between wild type, *Phf6* knockdown, *Lmo2* knockdown, and *Lmo2* overexpressing myeloid cells.

To achieve this, we utilised a cell line model (PUER) of PU.1-driven macrophage differentiation. The ETS family transcription factor PU.1 is essential for the development of myeloid lineages, such as macrophages, mast cells and neutrophils (Singh et al., 1999). PUER cells are murine haematopoietic progenitors, derived from the foetal liver of PU.1 knockout mice. During myeloid differentiation, these cells are blocked at the common myeloid progenitor (CMP) stage. After isolation they were retrovirally transduced to express PU.1 protein fused to the ligand-binding domain of the estrogen receptor (ER). Under normal cell culture conditions, this fusion protein is

excluded from the nucleus, however upon treatment of the cells with the synthetic oestrogen hydroxytamoxifen (OHT), PU.1 translocates to the nucleus where its function re-activates differentiation to macrophages. PUER cells have been used to study myeloid progenitors before, during and after differentiation into macrophages (Walsh et al., 2002).

## **4.2 Results**

### **4.2.1 Identification of new LMO2 Protein–Protein interactions in myeloid progenitors**

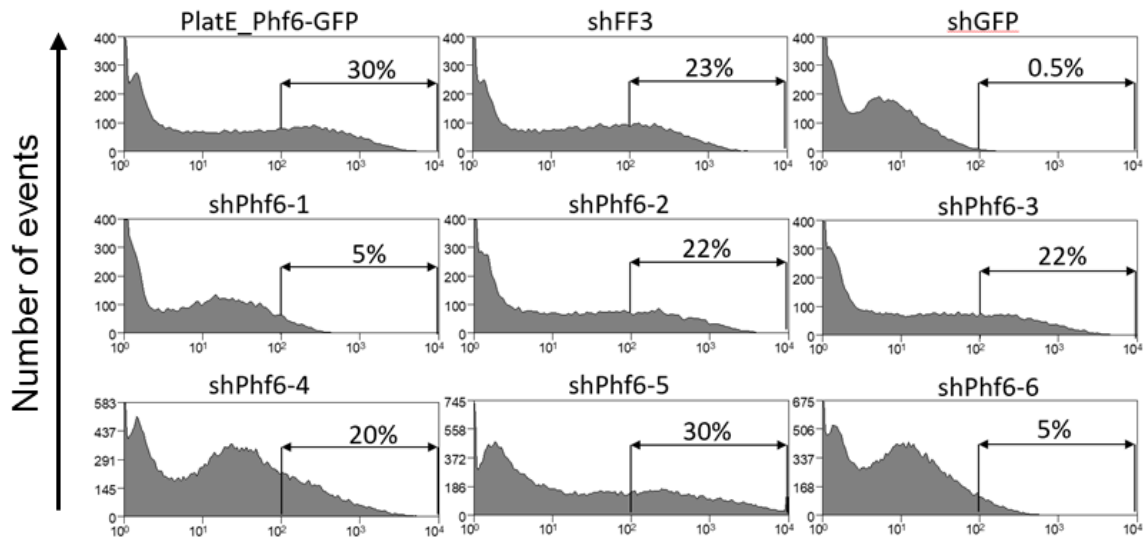
Taken that T-ALL cells data demonstrated an interplay between LMO2 and PHF6, the LMO2 interactions in myeloid cells were explored by performing co-immunoprecipitation followed by mass spectrometry experiments. In addition to known interacting partners, such as TAL1 and LDB1, PHF6 was identified as a novel interactor of LMO2 in PUER myeloid cells. The interaction between LMO2 and PHF6 was further confirmed by reciprocal immunoprecipitations using antibodies against the endogenous LMO2 and PHF6 followed by western blotting (data not shown). These experiments were performed by an undergraduate student (Ian Dorrington).

### **4.2.2 Retroviral mediated *Phf6* knockdown in PUER cells**

To gain insight into the role of the LMO2 complex in myeloid cells and to study the cellular functions of PHF6, stable *Phf6* knockdown (KD) cells and control cells using *Phf6*-targeting shRNAs (shPhf6) were generated in addition to a negative control vector targeting firefly luciferase (shCtrl) in PUER myeloid cells. Six different shRNAs were used to target *Phf6* as described in chapter 2 materials and methods section 2.3.

First, the viral packaging cell line PlatE was co-transfected with *Phf6* MigR1 and pMSCVhyg shPhf6 plasmids to test the efficiency of shRNA. The *Phf6* MigR1 allows the expression of the mouse *Phf6* cDNA followed by an IRES GFP. Thus, *Phf6* and GFP share the same mRNA and knockdown of either *Phf6* or GFP results in loss of the GFP signal. The pMSCVhyg containing shPhf6 allows the expression of shRNA that targets the *Phf6* mRNA expressed from the *Phf6* MigR1 plasmid and thereby depletes the GFP. Additionally, pMSCVhyg containing shFF3 and shGFP were used as the negative and positive control, respectively. The co-transfection allowed the assessment of shRNA efficiency and selection of the shPhf6 sequence which successfully knocked down *Phf6* through detection of GFP depletion. The cells were visualised by fluorescent microscopy for GFP expression and compared to the positive and negative controls. These results were further validated by flow cytometry analysis of the PlatE cells (Figure 4.1). Two of the shPhf6 (shphf6-1 and shPhf6-6) hairpins successfully depleted the GFP signal (Figure 4.1) and were subcloned into pMSCV-IRES-GFP and used to transfect PlatE cells. These cells were generated by the Hoogenkamp lab.

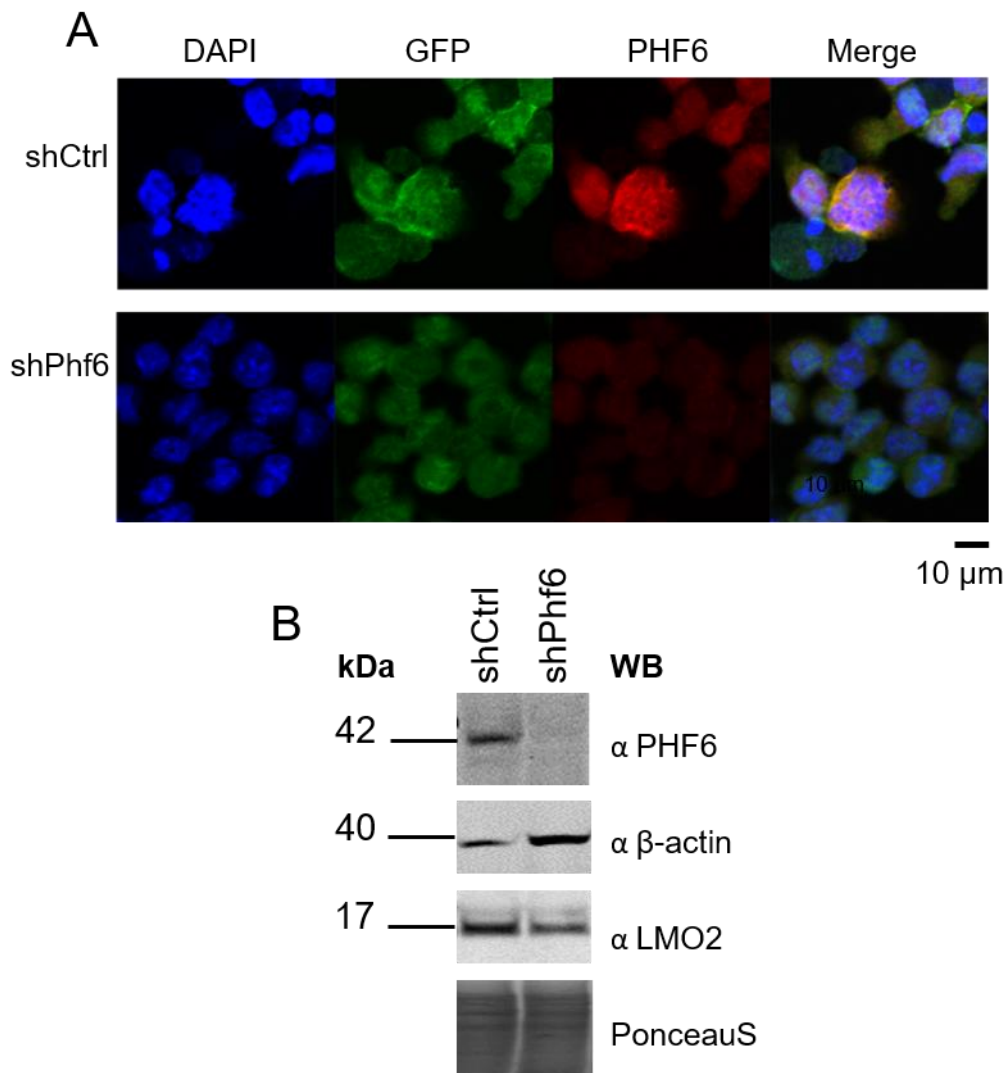




**Figure 4.1 Flow cytometry analysis of transfected PlatE cells.**

The top row graphs show PlatE cells that were transfected with MigR1 *Phf6* and either pMSCV shFF3 (negative control) or shGFP (positive control). The middle and bottom rows graphs show PlatE cells that were transfected with MigR1 *Phf6* and different shPhf6 sequences. The efficiency of shPhf6 and the percentages represent all events that have GFP Log > 10<sup>2</sup>. Cells were analysed with CyAn<sup>TM</sup> ADP flow cytometer (DakoCytomation-Beckman Coulter).

Based on this, PUER cells were transduced with pMSCV-IRES-GFP shPhf6-1 and shPhf6-6. To enrich for successfully transduced cells, they were FACS sorted on the basis of GFP expression. Subsequently, stable cell lines were generated through the outgrowth from single cell clones. To verify the efficiency of *Phf6* knockdown immunostainings were performed and showed GFP expression in both shCtrl and shPhf6, while PHF6 was not expressed in shPhf6 cells compared to shCtrl cells (Figure 4.2A). Additionally, the protein expression levels of PHF6 were compared by western blotting which showed loss of PHF6 protein expression in shPhf6 cells (Figure 4.2B).



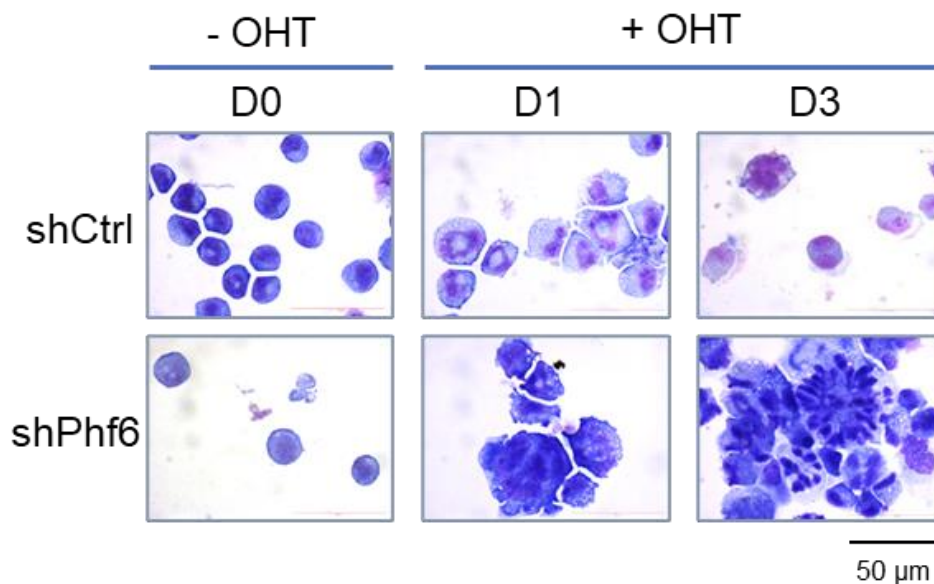
**Figure 4.2 *Phf6* knockdown in PUER cells.**

(A) Confocal imaging of transduced PUER cells expressing the green fluorescent protein (GFP). PUER shCtrl cells were transduced with shFF3 and shPhf6 cells were transduced with shRNA targeting *Phf6*. shCtrl and shPhf6 were stained with PHF6 (red) and counterstained with DAPI (blue) to visualise the nucleus. Scale bar 10  $\mu$ m. Images were taken using Carl Zeiss LSM510 Meta Confocal microscope (Carl Zeiss, Germany). (B) PHF6, LMO2 and  $\beta$ -actin protein levels detected by western blotting, using nuclear extracts from shCtrl and shPhf6 cells. PonceauS was used to show equal loading.

#### 4.2.3 Knockdown of *Phf6* slows myeloid differentiation

Taken that PUER cells can differentiate into macrophages when cultured in the presence of tamoxifen (OHT) (Walsh et al., 2002), we sought to study the effect of *Phf6*

deprivation on myeloid differentiation. We harvested shCtrl and shPhf6 cells at day 0 (D0), day 1 (D1), and day 3 (D3) of differentiation after treating the cells with OHT. Macrophage differentiation was verified in days 0, 1 and 3 by staining the cells with Kwik-Diff stain and examining them under the microscope (Figure 4.3). Noticeably, we observed morphological differences between the shCtrl and shPhf6 cells throughout differentiation days D1 and D3. The images of shPhf6 cells showed very large multi-nucleated cells with dense cytoplasm and scanty vacuoles, whereas shCtrl cells showed the common morphological features of macrophages.

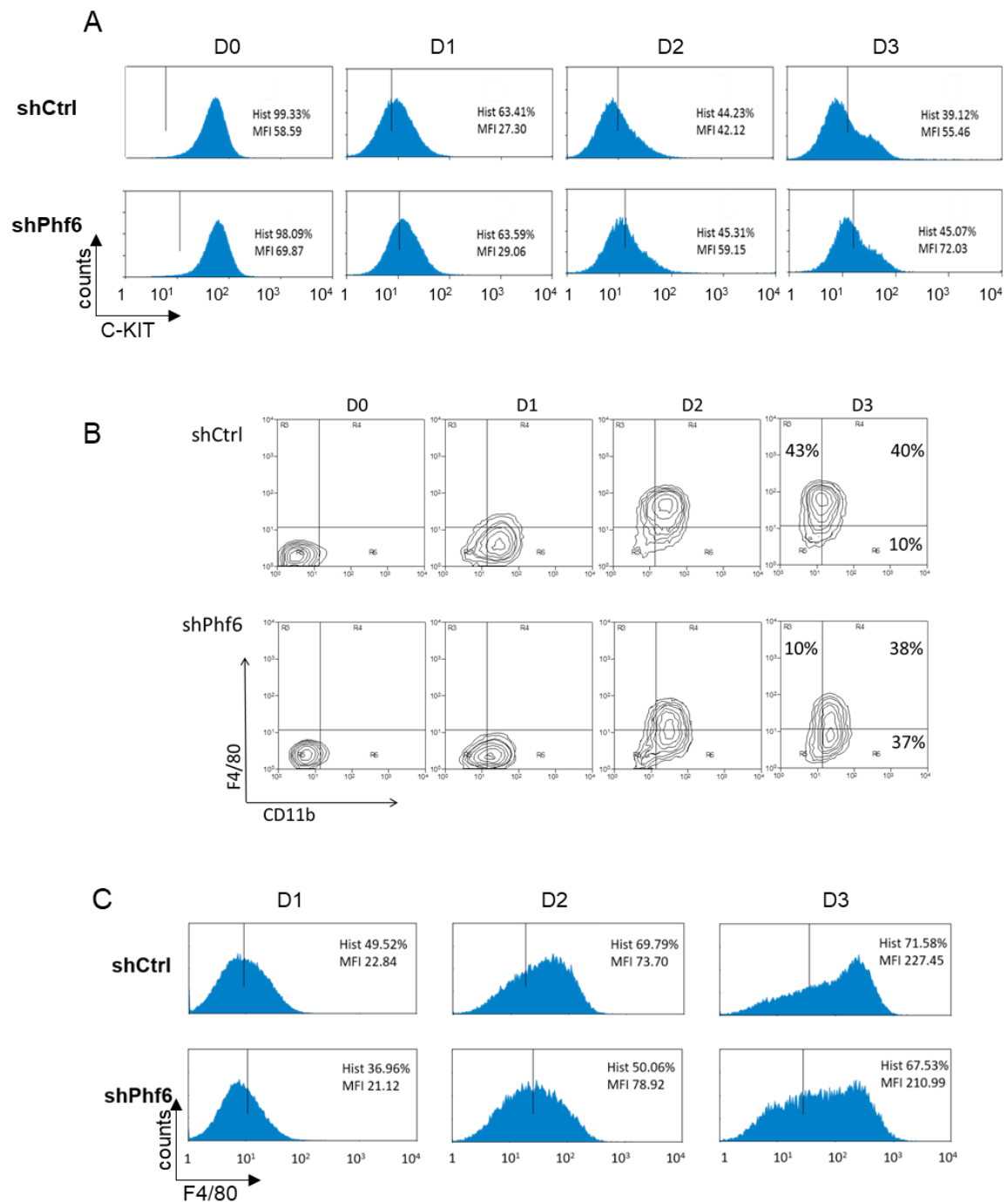


**Figure 4.3 Kwik-Diff staining of shCtrl and shPhf6 cells.**

The shCtrl and shPhf6 cells at D0 (undifferentiated), D1 and D3 of treatment with tamoxifen (OHT) to induce differentiation. Scale bar 50 µm. Images were taken using light microscope (Leica DM600).

Furthermore, cellular differentiation was assessed by flow cytometry analyses of myeloid cell surface antigens. The analyses showed expression of C-KIT, CD11b and F4/80 in both shCtrl and shPhf6 cells (Figure 4.4). As expected, C-KIT which is an HSC surface marker, was expressed in both cell lines in D0 at similar levels and

decreased gradually as the cells started to differentiate (Figure 4.4A). Starting from D1 the majority of shCtrl and shPhf6 cells indicated positive for CD11b marker, a macrophage differentiation marker, which increased throughout days of differentiation (Figure 4.4B). The expression of another macrophage surface marker (F4/80) was also tested. The flowcytometry data of shCtrl revealed that 69.8% of the cells at D2 and 71.58% of cells at D3 expressed F4/80. However, a lower percentage of shPhf6 cells, *i.e.* 50% at D2 and 67.3% at D3, expressed F4/80 (Figure 4.4B-C). This indicates that *Phf6* deprived cells experienced a slower rate of differentiation into macrophages compared to the wild type cells.

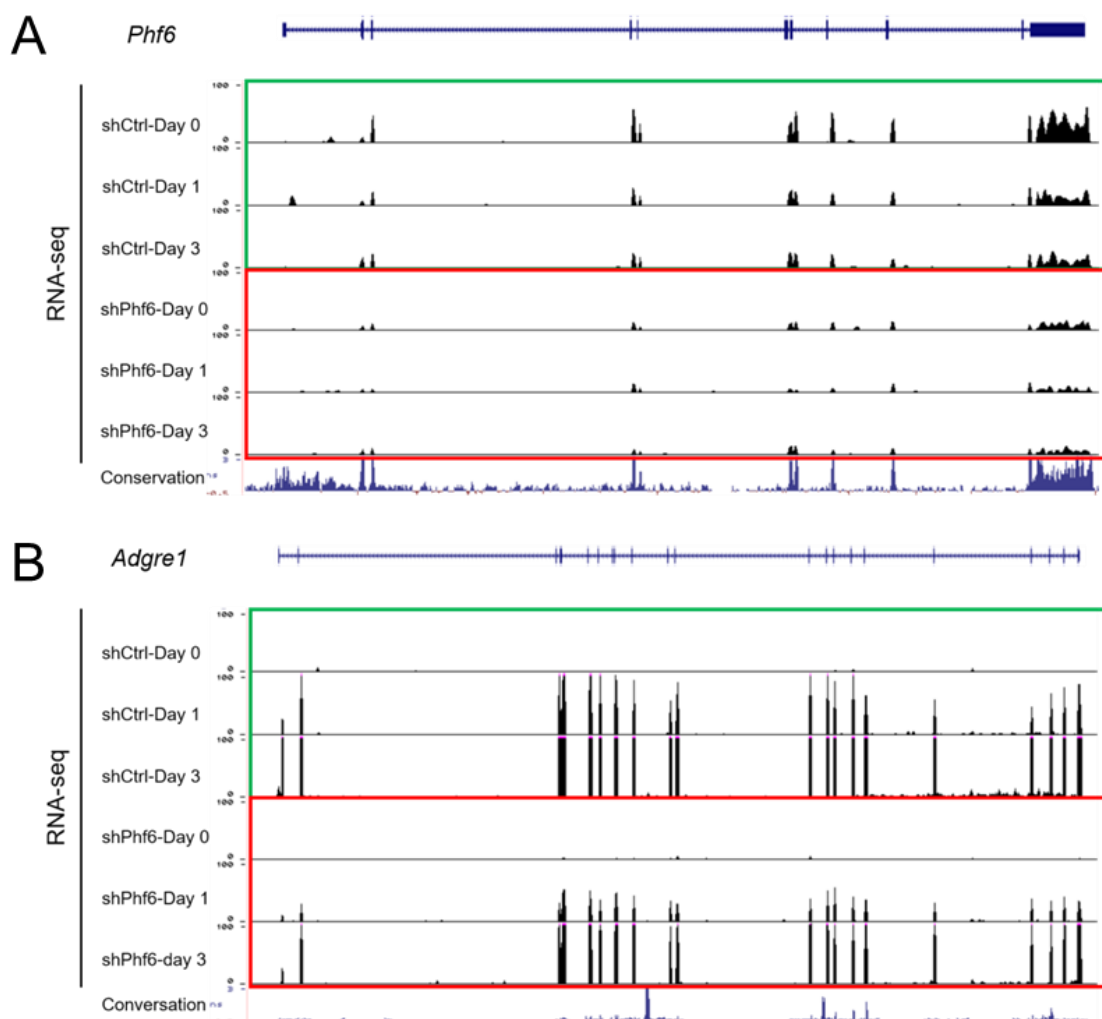


**Figure 4.4 Flow cytometry analyses of shCtrl and shPhf6 cell differentiation.**

The shCtrl and shPhf6 cells were harvested at D0, D1, D2, and D3 of differentiation. Flow cytometry analyses represented in a histogram (A) to evaluate C-KIT surface expression marker. Hist: histogram, indicate the percentage of positive population; MFI: Mean fluorescent intensity. (B) to evaluate the relative expression of CD11b and F4/80. The percentage of cells reflect the number of gated positive cells. (C) to evaluate F4/80 macrophage expression marker. Cells were analysed with CyAn<sup>TM</sup> ADP flow cytometer (DakoCytomation-Beckman Coulter)

#### 4.2.4 *Phf6* deprivation affects gene expression

To elucidate the molecular mechanisms underlying the delayed differentiation in *Phf6* knockdown cells, RNA-seq was performed using shPhf6 and shCtrl cells. We investigated the differences in gene expression between the knockdown cells and wild type cells. The comparisons between these cells uncovered a number of differences in their gene expression profiles. We initially looked at *Phf6* expression. The screen shot from the UCSC browser showed the suppression of *Phf6* in shPhf6 cells, while it was still being expressed in shCtrl cells (Figure 4.5A). This further confirmed the knockdown of *Phf6* in these cells. We also checked *Adgre1*, the gene encoding the F4/80 antigen, and found that it started being expressed in D1 and D3 of differentiation in the shCtrl cells, which was expected because it encodes a surface marker for macrophages but not progenitors, and is in line with the flow cytometry data presented in (Figure 4.4). However, *Adgre1* expression levels were much lower in shPhf6 KD cells, especially in D1 of differentiation (Figure 4.5B). Collectively, flow cytometry analyses showed delayed expression of the macrophage surface antigen F4/80 in shPhf6 cells, which was confirmed to be at the mRNA expression level. This indicates that shPhf6 cells experienced a delay in their differentiation kinetics.



**Figure 4.5 RNA-seq analyses of shCtrl and shPhf6 at D0, D1 and D3 of differentiation.**

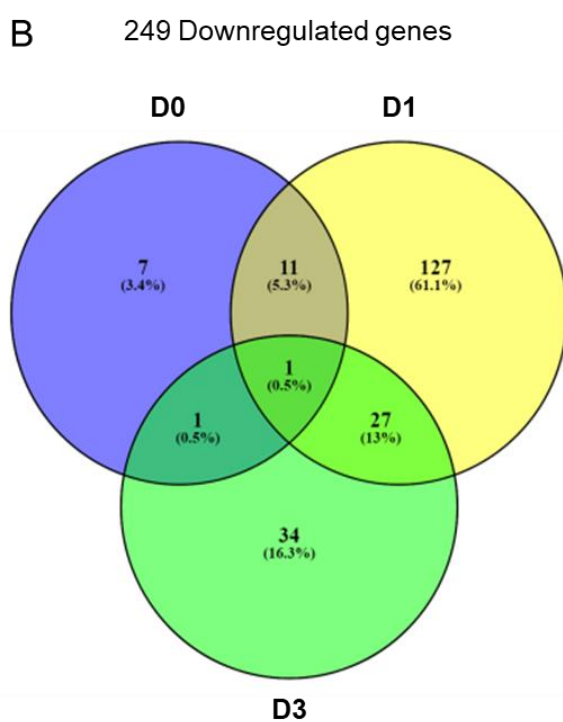
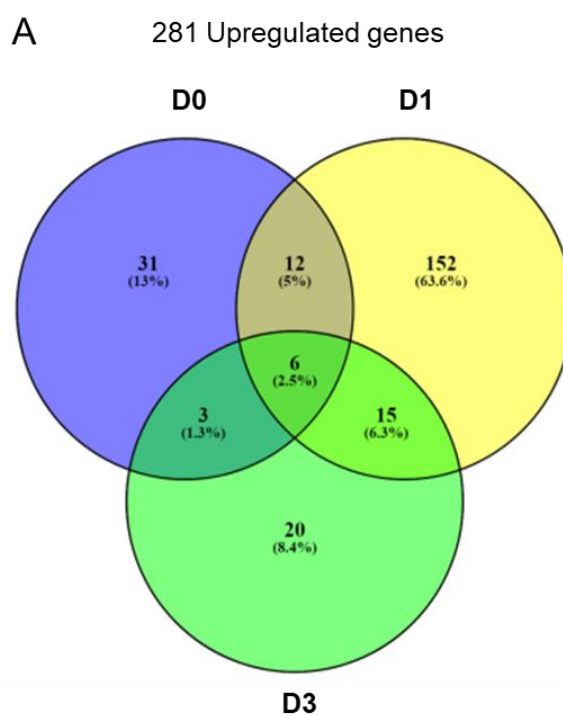
Screenshot from the UCSC browser showing (A) *Phf6* gene expression in shCtrl cells (green box), and suppression in shPhf6 cells (red box) at D0, D1 and D3. (B) *Adgre1* gene expression in shCtrl cells (green box), and lower expression in shPhf6 (red box) at D1 and D3.

To further understand how *Phf6* deficiency impacts global gene expression during D0, D1 and D3 of myeloid differentiation towards macrophages, we performed pairwise comparisons. This comparison identified 530 differentially expressed genes in shPhf6 in the three days of differentiation combined. We then classified them into upregulated

and downregulated genes per day. The total number of upregulated genes was 281 in shPhf6 compared to shCtrl with 52, 185 and 44 genes in D0, D1 and D3, respectively. On the other hand, 249 genes were repressed with only 20 of them in D0, 166 in D1 and 63 in D3. Further analysis of the upregulated genes overlaps showed that 6 of the expressed genes were shared between the three days, 12 common between D0 and D1, 15 common between D1 and D3, only 3 common between D0 and D3 (Figure 4.6A). Likewise, from the 249 repressed genes, 11 were common between D0 and D1, 27 between D1 and D3, 1 between D0 and D3, with only 1 shared between the 3 days of differentiation (Figure 4.6B).

Notably, the number of these differentially expressed genes was restricted to 72 and 107 between D0 and D3 respectively, whereas the majority of them were observed in the first day of differentiation with 185 genes upregulated and 166 downregulated. These changes in gene expression at D1 are in line with the previous observation of F4/80 suppression at D1, which indicates that *Phf6* deprivation mostly affected gene expression at the onset of myeloid differentiation. Altogether, differentiation from myeloid progenitor to macrophage was characterised by changes in gene expression between shCtrl and shPhf6 cells mostly in D1. At the progenitor stage D0 and macrophage stage D3, a lower number of genes were differentially expressed after loss of *Phf6*.

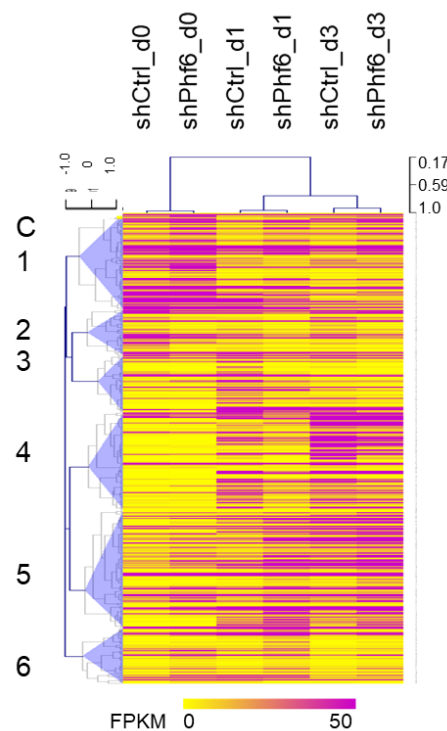




**Figure 4.6 Intersection of differentially expressed genes in shCtrl and shPhf6 PUER cells.**

Venn diagrams showing overlap between (A) Upregulated genes (B) Downregulated genes in shPhf6 cells compared to shCtrl cells at D0 (blue circle), D1 (yellow circle) and D3 (green circle).

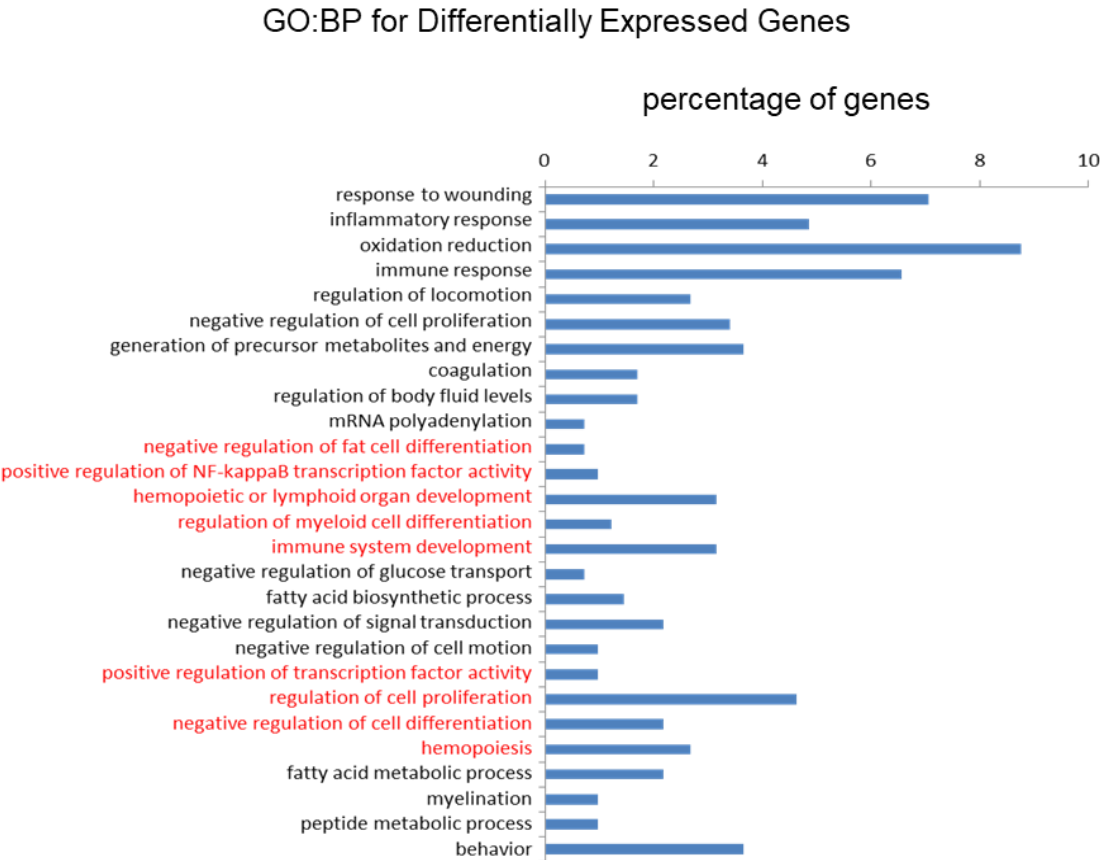
Hierarchical clustering for the differentially expressed genes revealed differences between shCtrl and shPhf6 cells. Self-organising tree clustering of all the differentially expressed genes identified six clusters (Figure 4.7). It showed that at D1 of differentiation shPhf6 cells failed to upregulate certain genes, demonstrated by clusters 3 and 4 and to suppress genes, as indicated by cluster 5, compared to shCtrl cells. By contrast, in cluster 2, shPhf6 cells were not able to upregulate genes in D0 despite maintaining a similar level of expression as shCtrl in D1 and D3. The differences in gene expression between the control and KD cells are mainly represented at D1. Hence, loss of *Phf6* resulted in changes in gene expression with marked alteration in D1 of differentiation (Figure 4.7).



**Figure 4.7 *Phf6* deprivation causes changes in gene expression levels.**

Heat map showing hierarchical clustering of differentially expressed genes comparing shCtrl and shPhf6 cell lines at D0, D1 and D3 of differentiation. Scale bar represents colour index for the log<sub>2</sub> FPKM values. Self-organising tree analysis identified 6 clusters. Analysis was performed using MultiExperimentViewer v4.9.0.

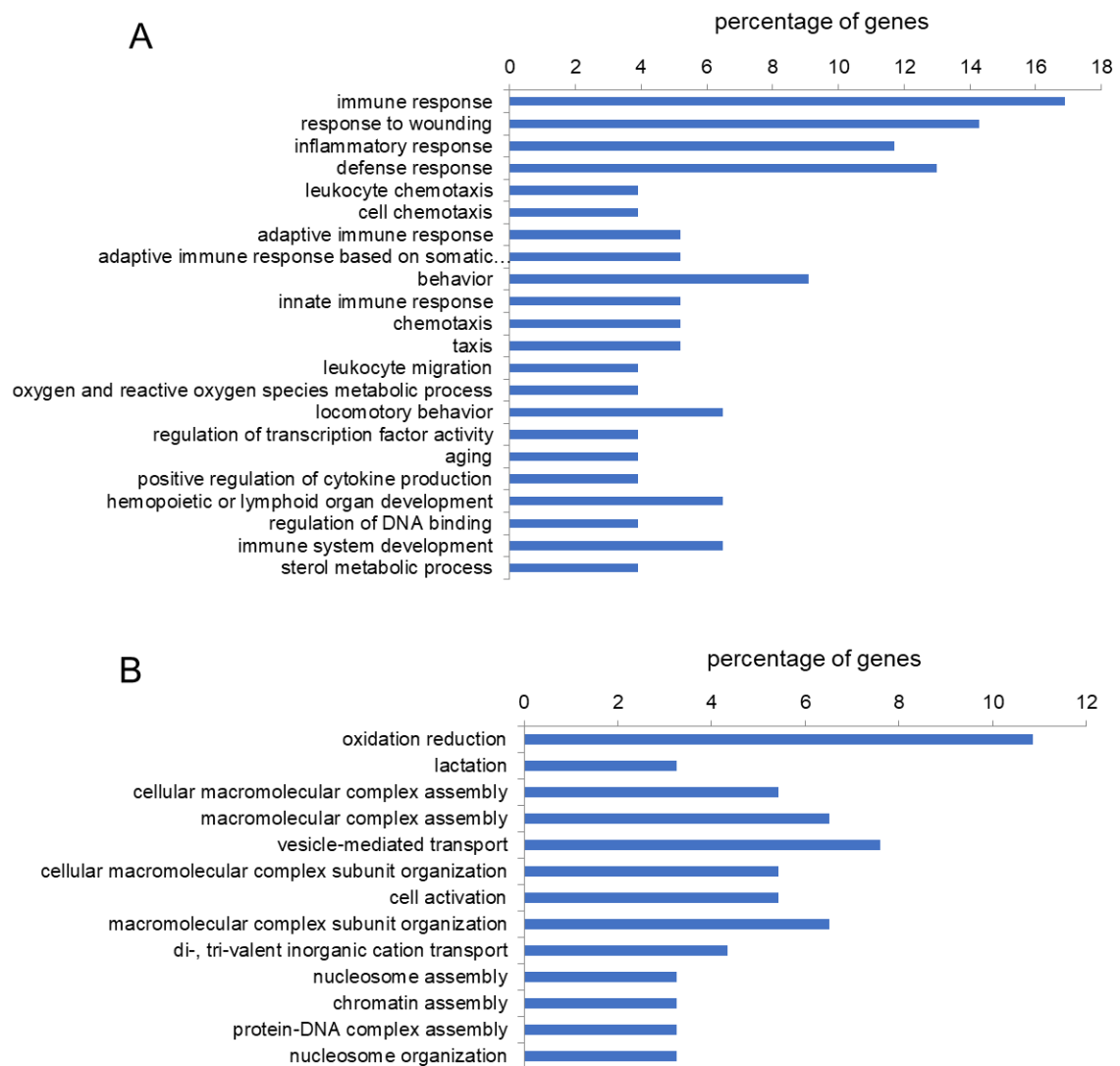
We next wanted to assess how the regulation of gene expression seen in shPhf6 impact on processes within myeloid cells. To achieve this all the differentially expressed genes underwent Gene Ontology (GO) analyses. The resulting GO terms showed processes involved in immune and inflammatory response, mRNA polyadenylation, NFκB transcription factor activity, haematopoietic or lymphoid organ development, regulation of myeloid cell differentiation, immune system development, regulation of transcription factor activity, regulation of cell proliferation, and haematopoiesis (Figure 4.8).



**Figure 4.8 Gene ontology analysis performed on all differentially expressed genes between shCtrl and shPhf6 cells.**

Terms were ordered according to their Modified Fisher Exact P-value and only terms with  $P < 0.05$  were considered significant. Analysis was performed using DAVID 6.8.

Individual clusters were also subjected to GO analyses. Cluster 3 was dominated by terms involving immune response, immune system development, regulation of transcription factor activity and regulation of DNA binding (Figure 4.9A), whereas cluster 4 revealed macromolecular complex assembly, cell activation, nucleosome assembly, chromatin assembly, protein-DNA complex assembly, nucleosome organisation (Figure 4.9B). Interestingly, within the list of transcriptional regulators differentially expressed in shPhf6 cells we found several proteins with known functions in haematopoietic development. Transcription factor genes such as *Lmo2*, *Id2* and *Egr1* were upregulated upon loss of PHF6. This suggests that PHF6 might exert a regulatory role in myeloid cells and target genes that are essential for blood cells development including myeloid cells.



**Figure 4.9 Gene ontology enrichment analyses for biological process performed on (A) cluster 3, (B) cluster 4 as identified in Figure 4.7.**

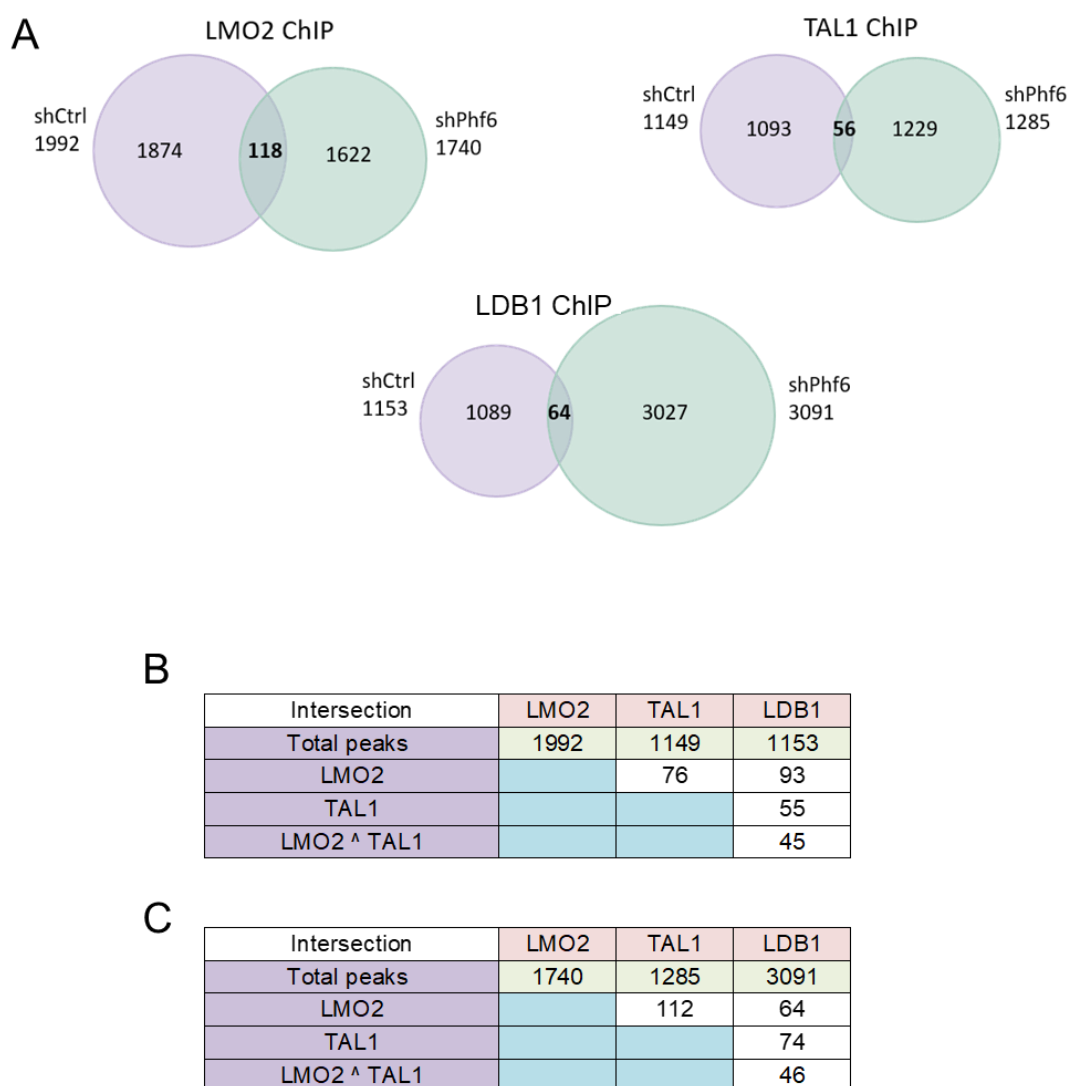
Terms were ordered according to their Modified Fisher Exact P-value and only terms with  $P < 0.05$  were considered significant. Analysis was performed using DAVID 6.8.

#### 4.2.5 LMO2, TAL1 and LDB1 binding is redistributed in the absence of PHF6

Having substantiated the changes in gene expression upon *Phf6* knockdown, we aimed to explore PHF6 binding across the genome. Moreover, we wanted to examine whether PHF6 co-localised with the LMO2 complex members on the genome and study the consequences of *Phf6* loss on the LMO2, TAL1, and LDB1 binding to DNA.

To accomplish this, ChIP-seq experiments were performed with antibodies against LMO2, TAL1 and LDB1 in shCtrl and shPhf6 cells in addition to PHF6 in shCtrl cells. PHF6 ChIP-seq resulted in a low number of peaks *i.e* 868 which were mostly located at highly repetitive, telomeric and centromeric regions. The intersection of PHF6 ChIP-seq with LMO2 ChIP-seq data showed a low number of overlapping peaks (data not shown). PHF6 ChIP-seq data was not very reliable, and we were not able show a functional interaction between PHF6 and LMO2.

The ChIP-seq experiments in shCtrl cells detected 1992 LMO2, 1149 TAL1, and 1153 LDB1 peaks. After loss of *Phf6* the number of bound sites by LMO2 decreased to 1740 and the number of TAL1 peaks increased to 1285. Surprisingly, the number of LDB1 peaks increased to 3091, with approximately 1938 extra bound regions in shPhf6 cells. To compare the region sets bound by the three factors LMO2, TAL1 and LDB1 in shCtrl and shPhf6, ChIP-seq data were intersected. The Venn diagrams demonstrated only 118 peaks of LMO2 shared between shCtrl and shPhf6 cells, 56 TAL1 shared peaks and 64 shared LDB1 peaks (Figure 4.10A). This indicates that loss of *Phf6* redistributed the binding sites of LMO2, TAL1, and LDB1 drastically. The interplay between LMO2, TAL1, and LDB1 was then assessed by intersecting LMO2 with each factor individually and both factors together. LMO2 had 76 peaks in common with TAL1 and 93 with LDB1 in shCtrl cells. The LMO2, TAL1, and LDB1 associated together to 45 sites in shCtrl cells (Figure 4.10B). On the other hand, the number of bound regions by LMO2/TAL1 increased to 112 and by LMO2/LDB1 decreased to 64 after depletion of *Phf6* and all three factors bound to 46 sites in shPhf6 cells (Figure 4.10C).

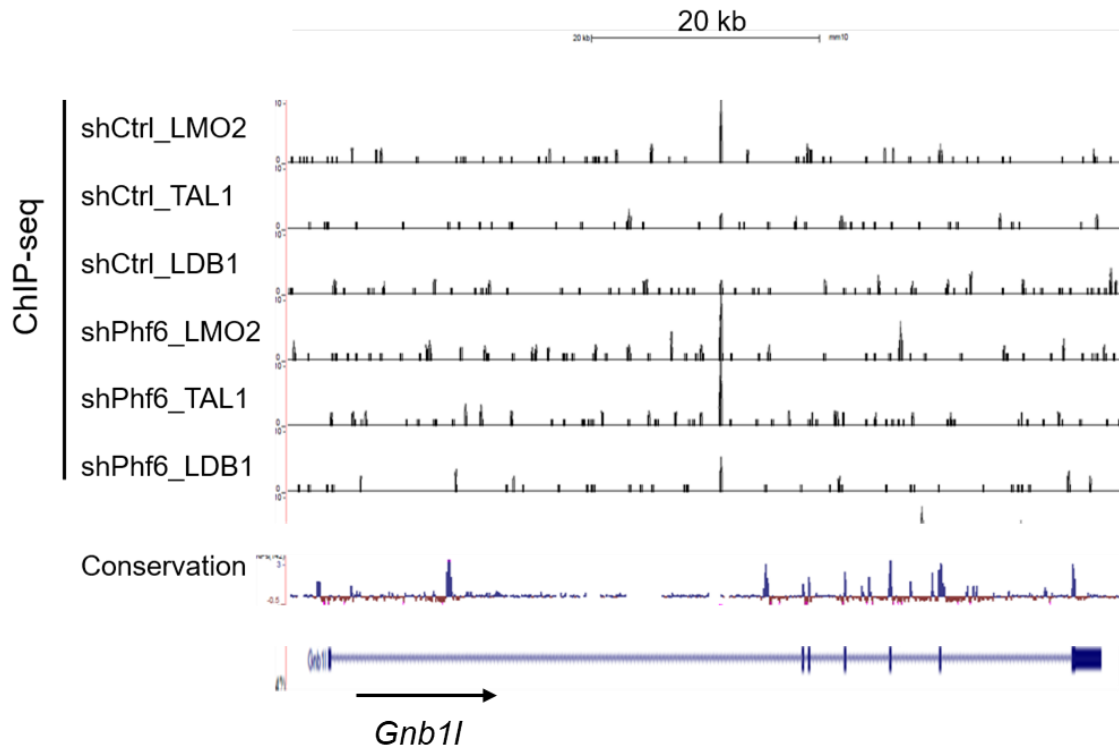


**Figure 4.10 Intersections of LMO2, TAL1 and LDB1 ChIP-seq data in shCtrl and shPhf6 cells.**

(A) Venn diagrams showing the number of LMO2, TAL1 and LDB1 common peaks between shCtrl and shPhf6 cell lines. (B, C) Tables showing the common peaks between LMO2, TAL1 and LDB1 in (B) shCtrl cells, and (C) shPhf6 cells.

These ChIP-seq analyses showed that there was no significant change in the number of regions bound by LMO2/TAL1/LDB1 complex after *Phf6* knockdown. However, TAL1 and LDB1 had more peaks in shPhf6 cells, with greater overlap between LMO2 and TAL1. This was exemplified by the peaks shown in UCSC genome browser

screenshots (Figure 4.11). These data indicate that LMO2 binds a relatively low number of genomic regions with TAL1 and LDB1 in myeloid cells and knockdown of *Phf6* led to an increased number of LMO2/TAL1 peaks.



**Figure 4.11 Screenshot from the UCSC browser showing LMO2, TAL1 and LDB1 binding profiles in shCtrl and shPhf6 cells at *Gnb1l* locus.**

#### 4.2.6 LMO2 binds to novel sites after *Phf6* depletion

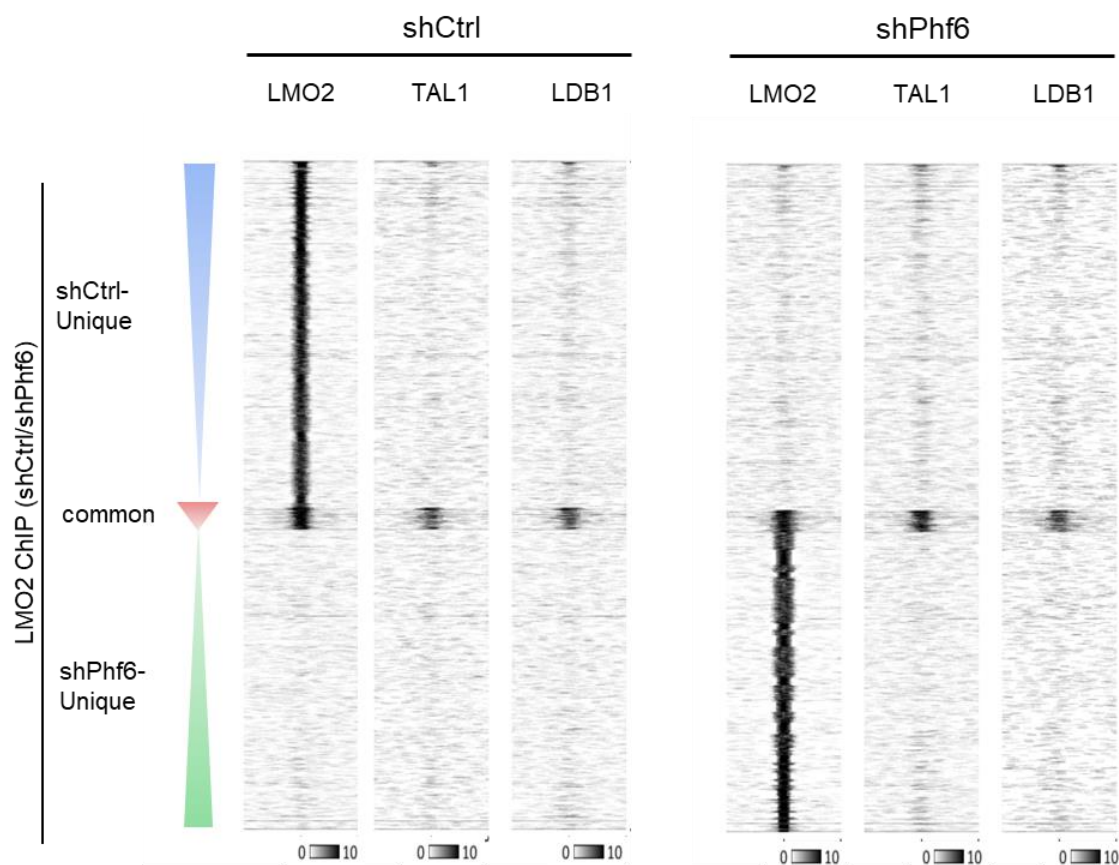
The ChIP-seq analyses showed that LMO2 binds to different genomic regions upon loss of *Phf6* (Figure 4.10A), therefore we compared the ChIP-seq signals of TAL1 and LDB1 with the LMO2 signal. We used the LMO2 ChIP-seq intersection data and classified them into (i) shCtrl unique peaks, (ii) common peaks between shCtrl and shPhf6, and (iii) shPhf6 unique peaks. The three categories were sorted according to peak intensity ratios of shCtrl/shPhf6



Figure 4.12). This classification revealed 1874 peaks unique to shCtrl, 1622 peaks unique to shPhf6 and 118 common in both.

Next, we compared LMO2 ChIP signals to the corresponding TAL1 and LDB1 peaks in shCtrl and shPhf6. This comparison demonstrated that the strongest shCtrl unique LMO2 peaks were mirrored by weak TAL1 peaks and stronger LDB1 peaks in shCtrl cells (Figure 4.12). When looking at these sites in shPhf6, LMO2 binding was not completely lost, but corresponded to weaker LMO2 peaks. These regions that remained bound by LMO2 showed more TAL1 co-localisation and LDB1 overlap upon loss of *Phf6* (Figure 4.12). This was in line with the data shown in Figure 4.10C and Figure 4.11, where stronger binding of TAL1 with LMO2 was observed in shPhf6 cells.

On the other hand, the new regions specifically bound by LMO2 in shPhf6 cells did not show strong occupancy by TAL1 and LDB1. The LMO2 bound sites with the highest score at the bottom of the heat map were bound by TAL1 with minimal binding of LDB1 (Figure 4.12). These regions also showed peaks of LMO2, TAL1 and LDB1 with very low signal intensity in shCtrl cells. The LMO2 peaks that were common in both cell lines and did not relocate after *Phf6* knockdown were mirrored by the strongest TAL1, and LDB1 peaks. This indicates that these common peaks may be the strongest binding sites for the LMO2 complex. Therefore, these are the regions that showed LMO2, TAL1, and LDB1 occupancy in wild type and *Phf6* knockdown cells.



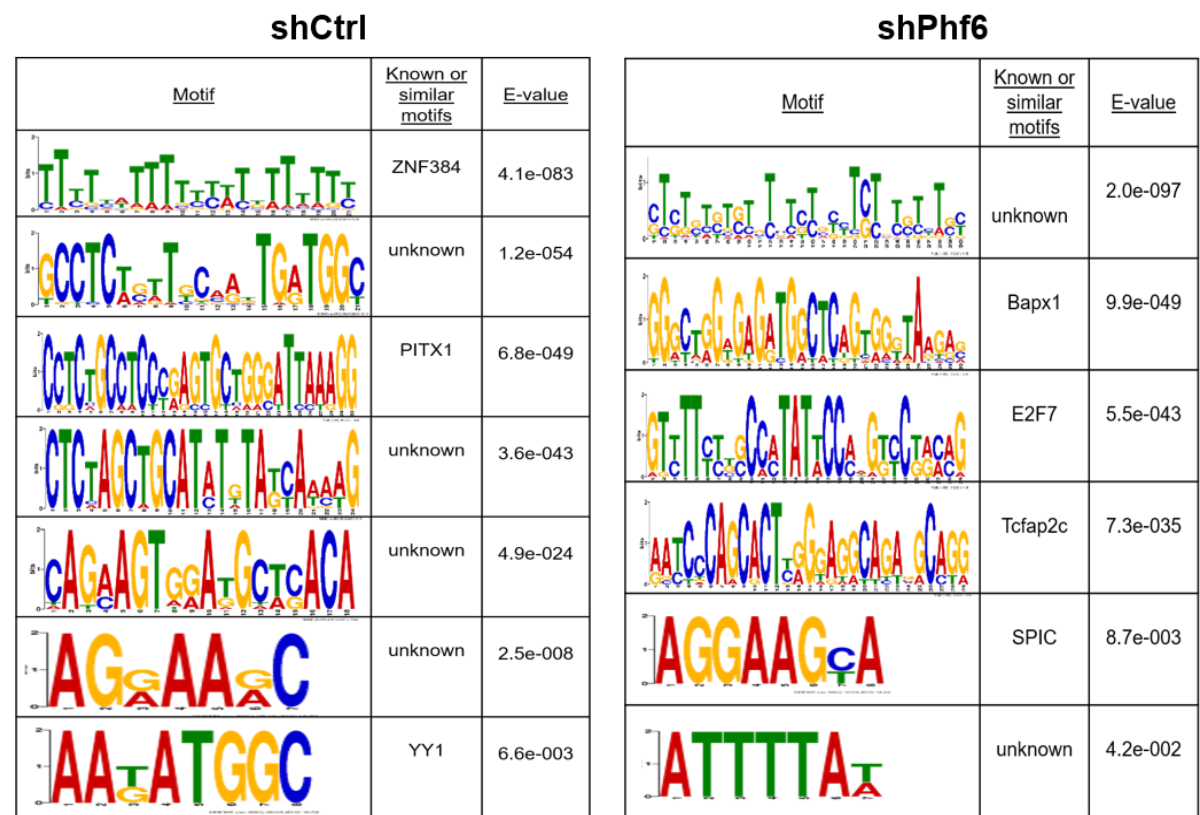
**Figure 4.12 Heat map plots of ChIP-seq data for LMO2, PHF6, TAL1 and LDB1 in shCtrl and shPhf6 cells.**

The LMO2 ChIP-seq peaks (region-sets) ranked according to the peak intensity ratio score shCtrl/shPhf6. ShCtrl and common peaks in descending order (blue and red), shPhf6 unique peaks in ascending order (green). The heat map plots showing ChIP-seq data of LMO2, TAL1 and LDB1, overlaid on LMO2 ChIP-seq region-sets. The Y-axis represents individual positions of the regions, and the X-axis represents a window from  $-1$  kb to  $+1$  kb centred on the summits of LMO2 peaks. A bar showing the relationship between colouring and signal intensity at the bottom of each plot. Heatmaps were generated using EaSeq analysis software.

#### 4.2.7 shPhf6 cells feature distinctive binding motifs

Considering that *Phf6* deficiency redistributed LMO2, TAL1 and LDB1 binding, the differences in binding motifs between shCtrl and shPhf6 cell lines were examined. Long DNA sequences that were identified by the MEME program as either unknown

or similar to zinc finger motifs were among LMO2, TAL1, and LDB1 enriched motifs in both shCtrl and shPhf6 cells. Comparing the binding motifs of each factor before and after loss of *Phf6* showed differences in their binding patterns. For example, in shCtrl cells LMO2 featured enrichment of motifs similar to ZNF384, PITX1 and the known YY1 motif. Different motifs were identified within long DNA sequences in shPhf6 cells including Bapx1, E2F7, and Tcfap2c in addition to a motif similar to SPIC (Figure 4.13).



**Figure 4.13** *De novo* motif enrichment analyses within LMO2 peaks in shCtrl and shPhf6 cells.

The results are presented as sequence of the motif, name of the motif, and the E-value. Analyses were performed using MEME-ChIP version 4.12.0.

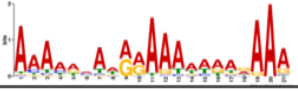
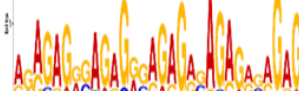
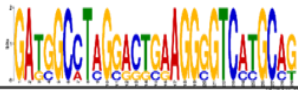
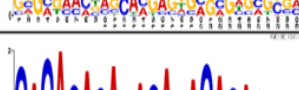
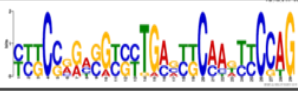
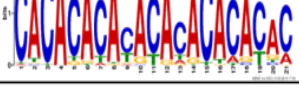
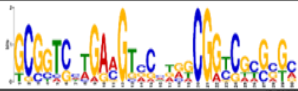



On the other hand, TAL1 showed binding at motifs identified as ZNF384, PITX1 and Nkx3-1 in both cell lines. The STAT motif was enriched in shCtrl cells in contrast to SPIB motif enrichment in shPhf6 cells (Figure 4.14).

[illegible]

**Figure 4.14** *De novo* motif enrichment analyses within TAL1 peaks in shCtrl and shPhf6 cells.

The results are presented as sequence of the motif, name of the motif, and the E-value. Analyses were performed using MEME-ChIP version 4.12.0.

Similar to LMO2, the analysis of LDB1 ChIP did not show consistent binding motifs in shCtrl and shPhf6 cells. Motifs similar to TBX20, RXRG and KLF were enriched in shCtrl compared to the overrepresented ZNF motifs in shPhf6 cells (Figure 4.15).

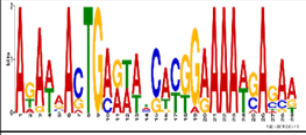
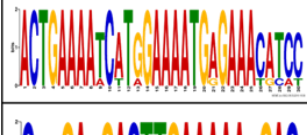
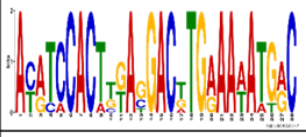
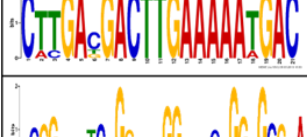
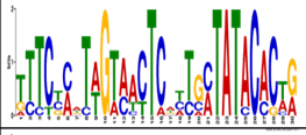
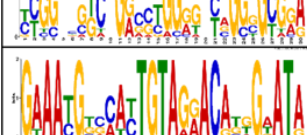
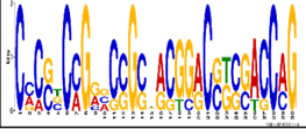
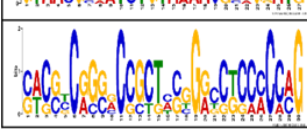
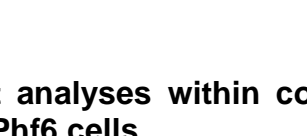
shCtrl			shPhf6		
Motif	Known or similar motifs	E-value	Motif	Known or similar motifs	E-value
	ZNF384	7.9e-071		ZNF263	5.4e-209
	TBX20	2.3e-048		unknown	3.1e-191
	unknown	1.6e-023		ZNF384	1.1e-065
	RXRG	3.5e-003		unknown	9.0e-048
	KLF14 SP8	8.3e-003			
	unknown	4.2e-002			

**Figure 4.15** *De novo* motif enrichment analyses within LDB1 peaks in shCtrl and shPhf6 cells.

The results are presented as sequence of the motif, name of the motif, and the E-value. Analyses were performed using MEME-ChIP version 4.12.0.

Next, the motifs bound by the LMO2, TAL1, and LDB1 common peaks were compared between shCtrl and shPhf6 cells. In shCtrl cells, the motif analysis showed binding to unknown DNA sites composed of long stretches of DNA sequence. In shPhf6 cells, the complex also showed binding to long sequence motifs where one of these unknown motifs was enriched in shCtrl as well (Figure 4.16). The shPhf6 cells demonstrated enrichment of a long DNA sequence with an NFAT binding motif (5'-GGAAA-3') located within it. An Esrra secondary motif also resided in another stretched sequence that was enriched in these cells. The rest of the motifs that passed the significance threshold had no resemblance to motifs bound by known TFs (Figure 4.16). Overall,

De novo Motif analyses did not show enrichment for motifs of myeloid associated TFs such as RUNX or ETS family proteins and revealed long sequences of DNA.

shCtrl			shPhf6		
Motif	Known or similar motifs	E-value	Motif	Known or similar motifs	E-value
	unknown	1.1e-043		NFATC2 NFAT5 SOX8	3.1e-042
	unknown	1.7e-022		unknown	7.6e-018
	unknown	3.4e-020		Esrra secondary	1.1e-021
	unknown	7.7e-009		unknown	5.3e-011
				unknown	6.2e-003

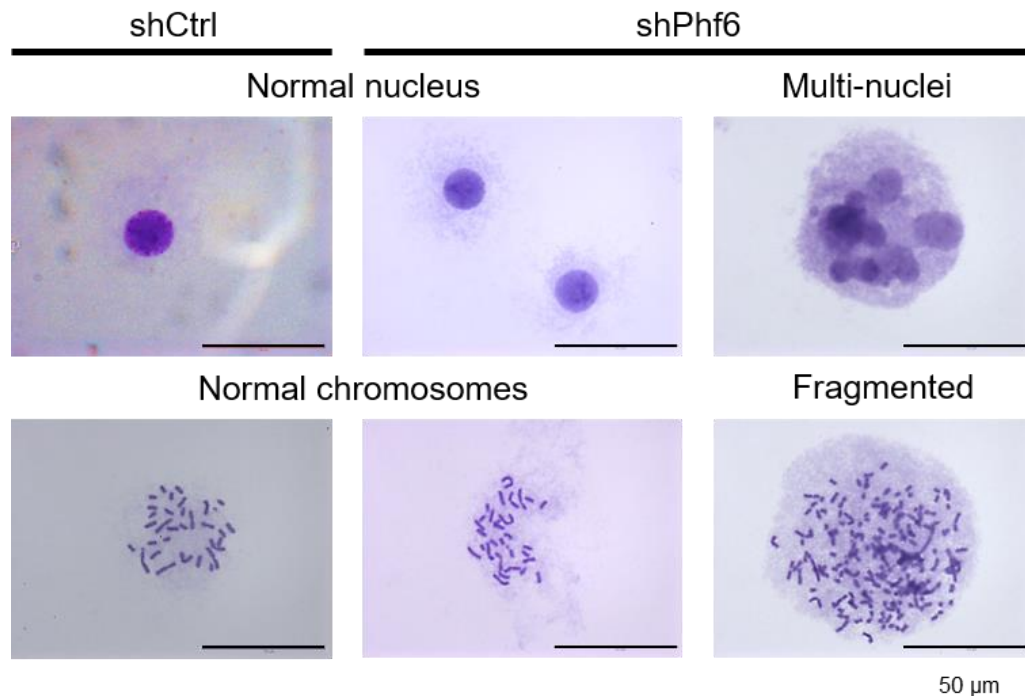
**Figure 4.16** *De novo* motif enrichment analyses within common peaks for LMO2, TAL1 and LDB1 in shCtrl and shPhf6 cells.

The results are presented as sequence of the motif, name of the motif, and the E-value. Analyses were performed using MEME-ChIP version 4.12.0.

#### 4.2.8 *Phf6* knockdown affects genome and chromosome stability

The RNA-seq analyses showed that *Phf6* deprivation affects gene expression. In addition, Kiwik-Diff staining showed distinct morphological features in shPhf6 cells where large multi-nucleated cells were observed (Figure 4.3), which could indicate a replication or genome stability problem. Hence, we speculated that PHF6 might be required for genome stability. To test this, chromosomal spread analyses were performed. Indeed, *Phf6* deprived cells exhibited polyploidy chromosomes and genome instability compared to shCtrl. Consistent with this, we observed significantly

elevated levels of multi-nuclei and possibly fragmented or completely pulverized metaphase chromosomes (Figure 4.17).



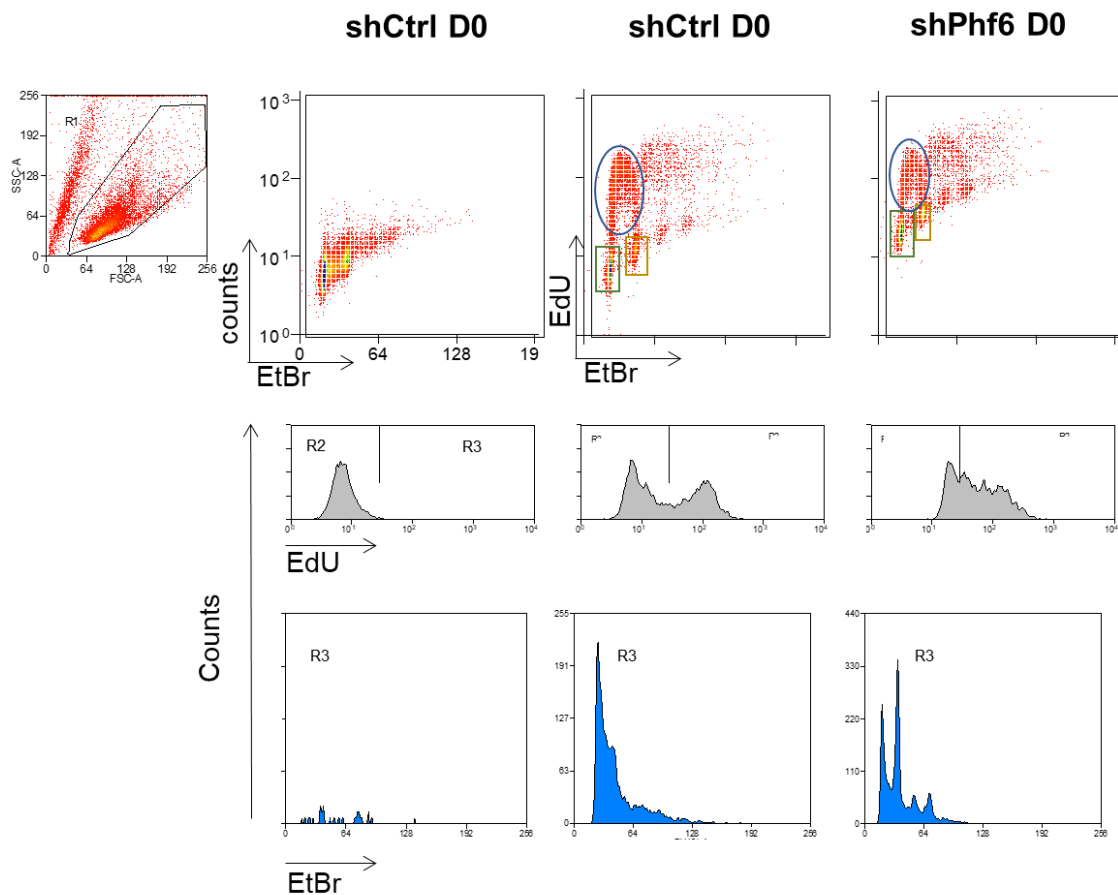
**Figure 4.17 Chromosomal spread analyses in shCtrl and shPhf6 cells.**

Representative image from chromosome spread experiments of normal chromosomes and fragmented chromosomes (n=3 Independent experiments;  $\geq 20$  metaphases per sample per experiment). Scale bar, 50  $\mu\text{m}$ . Images were taken using light microscope (Leica DM600).

To compare the DNA content and the cells in S-phase between shCtrl and shPhf6, ethidium bromide staining coupled with ethynyldeoxyuridine (EdU) labelling was performed by the Hoogenkamp lab. Flow cytometry analysis showed that shCtrl cells had discrete amounts of DNA and groups of dividing cells that go through regular cell cycle phases G1, S, and G2/M (Figure 4.18). On the other hand, the dual positive cells in S-phase cells were more in shPhf6. This indicates an increased DNA content and higher amount of replicating DNA (Figure 4.18). Moreover, the histograms in shCtrl showed a clear separation of proliferating cells which have incorporated EdU and



nonproliferating cells which have not. The histogram of shPhf6 cells showed more EdU positive continuously proliferating cells which seemed to fail to complete timely replication (Figure 4.18). These observations are in line with the chromosomal spread analyses, which demonstrated shPhf6 cells with abnormal numbers of chromosomes and multi-nuclei.

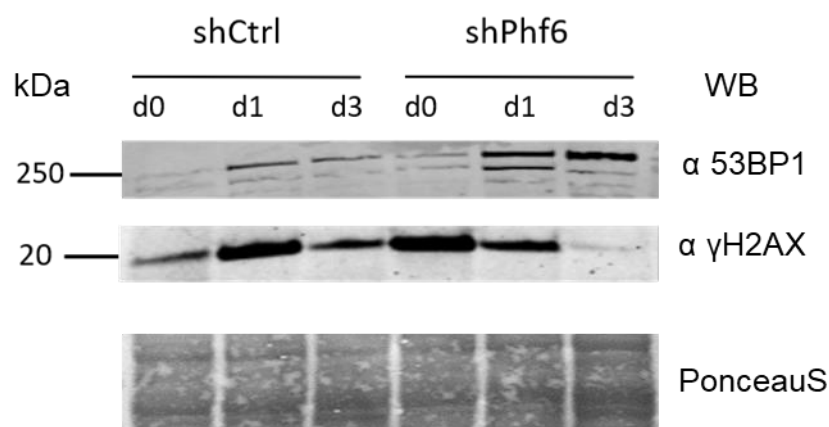


**Figure 4.18 Flow cytometry analysis of ethidium bromide (EtBr) stained cells labelled with ethynyldeoxyuridine (EdU).**

Cells were cultured in the presence of 10  $\mu$ M EdU for two hours before fixation, staining, and detection by flow cytometry. Plot (left) showing shCtrl cells stained with EtBr indicating DNA content. Dual-parameter plots (upper middle and right) showing shCtrl and shPhf6 cells, EdU indicating newly synthesised DNA, cells in the green box are in G1 phase, in the blue oval are in S-phase (dual positive), and in the yellow box are in G2 phase of the cell cycle. Histograms (grey) showing EdU labelling (x-axis) and cell count (y-axis). Histograms (blue) showing EtBr staining (x-axis) and cell count (y-axis).



To further investigate whether *Phf6* deprivation would lead to increased DNA damage, DNA damage markers were tested. This was carried out by preparing nuclear extracts from shCtrl, shPhf6 cells followed by western blotting with 53BP1 and  $\gamma$ H2AX (markers of double-strand breaks and DNA damage, respectively) (Figure 4.19). These analyses showed that shPhf6 accumulated 53BP1 in days 1 and 3 of differentiation compared to the shCtrl. Conversely, phosphorylated H2AX ( $\gamma$ H2AX) was observed in D0 and showed lower levels in D1 and was undetectable in D3 in shPhf6 cells. The shCtrl cells showed similar levels of  $\gamma$ H2AX in D0 and D3 with higher levels at D1 (Figure 4.19). Normally, DNA double-strand breaks attract the 53BP1 protein to the surrounding chromatin, where the 53BP1 signals chromatin/DNA damage in a  $\gamma$ -H2AX-dependent manner. The outcomes of the western blot experiments were not conclusive because the co-localisation of 53BP1 and  $\gamma$ H2AX must be detected which is indicative of DNA repair. Thus, further investigations by immunostaining must be carried out to visualise 53BP1 and  $\gamma$ H2AX foci along with cell cycle distribution analysis to explore the possibility of DNA damage accumulation or replication problems in these cells.



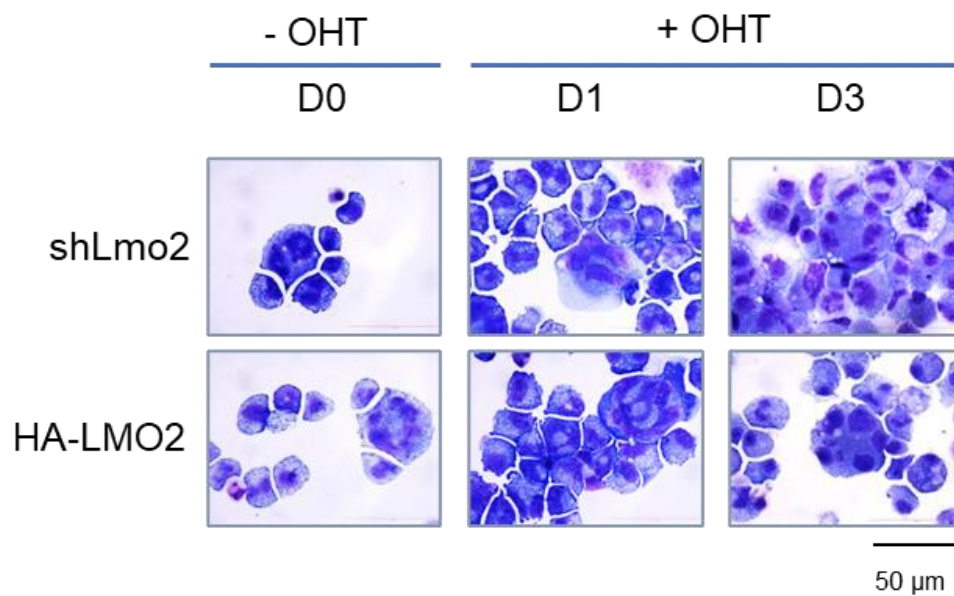
**Figure 4.19 *Phf6* depletion shows signs of DNA damage.**

The shPhf6 cells or shCtrl cells at day 0, day 1 and day 3 of differentiation were lysed and subjected to Western blotting (WB) using the indicated antibodies.

#### **4.2.9 *Lmo2* knockdown and overexpression shows a similar phenotype to shPhf6**

To gain better insight into the cellular functions of *Lmo2* and compare the effect of *Lmo2* deprivation to *Phf6* depletion in myeloid cells, we utilised stable *Lmo2* knockdown cells, that were generated using *Lmo2*-targeting shRNAs (shLmo2) in PUER myeloid cells, following the same protocols and validation methods that were used in shPhf6 cells as described in 4.2.2. Moreover, to study the overexpression of LMO2, we used a stable PUER cell line, which was generated previously by the Hoogenkamp lab. This PUER cell line exhibits overexpression of *Lmo2* with an HA-tag (HA-LMO2).

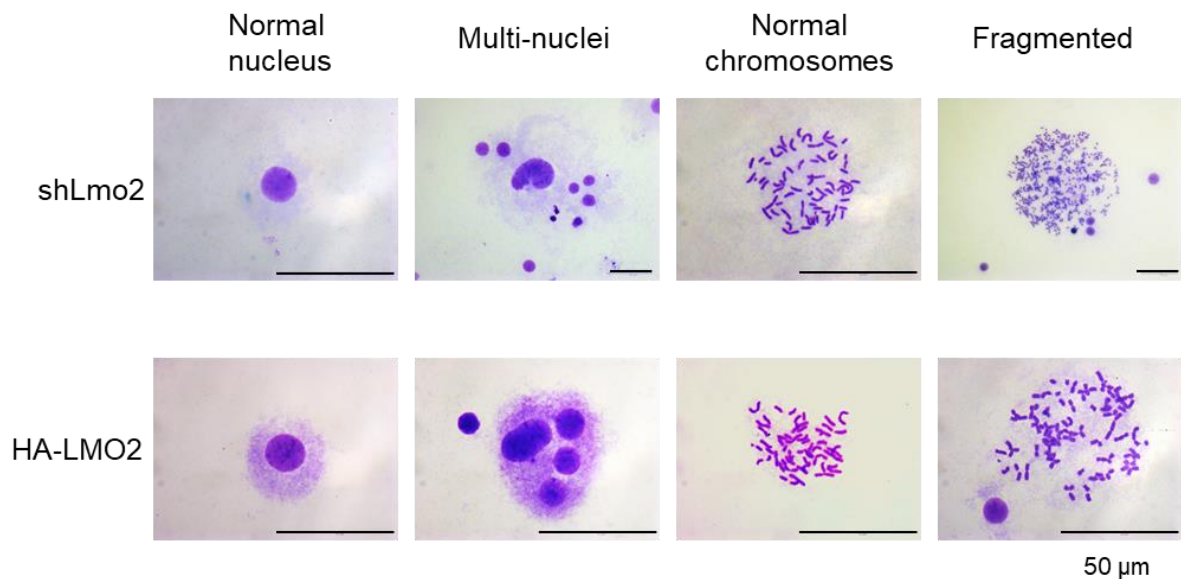
Macrophage differentiation was obtained by treatment of the cells with OHT. The cells in D0, D1 and D3 were stained with Kwik-diff stains and visualised under the microscope. Interestingly, both shLmo2 and HA-LMO2 cells showed very large multi-nucleated cells with coarse cytoplasm and few vacuoles (Figure 4.20). These morphological features were similar to the ones observed in shPhf6 cells (Figure 4.3).



**Figure 4.20 Kwik-Diff staining of shLmo2 and HA-LMO2 cells.**

The shLmo2 and HA-LMO2 cells at D0 (undifferentiated), D1 and D3 of treatment with tamoxifen (OHT) to induce differentiation. Scale bar 50  $\mu$ m. Images were taken using light microscope (Leica DM600).

To explore the effect of *Lmo2* depletion or overexpression on chromosome stability, chromosomal spread analyses were performed. The visualised metaphases revealed polyploidy and possible chromosomal breakage in both cell lines. Additionally, the cells were multi-nucleated, and the chromosomes looked fragmented or pulverized (Figure 4.21).

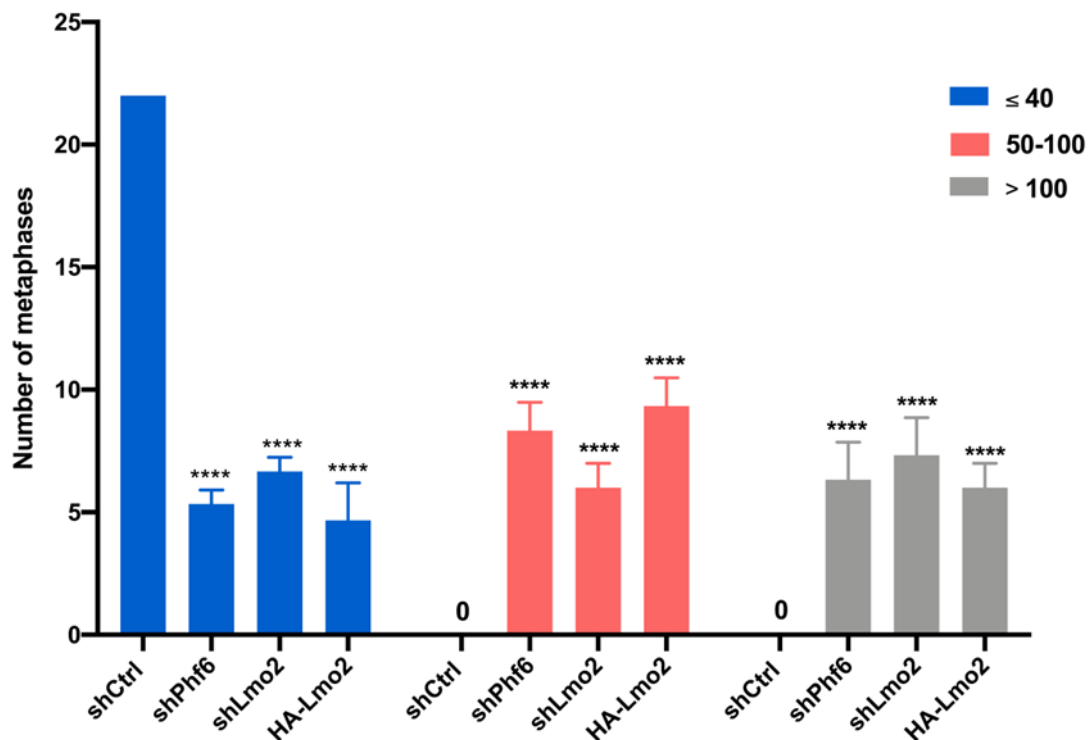


**Figure 4.21 Chromosome spreads of shLmo2 and HA-LMO2 cells.**

Representative image from chromosome spread experiments of normal chromosomes and fragmented chromosomes (n=3 Independent experiments;  $\geq 20$  metaphases per sample per experiment). Scale bar, 50  $\mu\text{m}$ . Images were taken using light microscope (Leica DM600).

Finally, the number of chromosomes was counted in at least 20 metaphases per experiment in each cell line to investigate the cells that had abnormal metaphases and compare them to shCtrl cells. The metaphases were categorised into three groups according to the number of chromosomes per metaphase (i) 40 chromosomes, which is the normal number of  $2n$  chromosomes in mouse cells, or less than 40, (ii) 50-100 chromosomes, (iii) more than 100. Statistical analyses by two-way ANOVA test were performed using prism 8, version 8.4.1 (460) from GraphPad company (Figure 4.22). The chromosomal counts revealed that all of the control cells had  $\leq 40$  chromosomes, while shPhf6, shLmo2, and HA-LMO2 cells showed significantly less metaphases with normal or lower than normal chromosome count. The shCtrl cells did not have any metaphases with high number of chromosomes, whereas the other cell lines had a significantly high number of metaphases that showed 50-100 chromosomes and more

than 100 chromosomes (Figure 4.22). Remarkably, *Lmo2* and *Phf6* deprivation and *Lmo2* overexpression caused comparable changes in chromosomal counts, which indicates that LMO2 and PHF6 may play a role in genome and chromosome stability or replication.



**Figure 4.22 Chromosome counts of shCtrl, shPhf6, shLmo2, and HA-LMO2 cells.**

Bar graph showing the number of metaphases counted in each cell line. Error bars represent SEM, \*\*\*\* indicates  $p < 0.0001$ . Statistical analysis using two-way ANOVA test were performed using prism 8, version 8.4.1 (460) from GraphPad company ( $n=3$  Independent experiments;  $\geq 20$  metaphases per sample per experiment).

## **4.3 Discussion**

### **4.3.1 LMO2 physically interacts with PHF6 in common myeloid progenitors**

The interaction between LMO2 and TAL1 has been investigated previously in the myeloid cells, however the LMO2 complex has not been fully described within the normal myeloid lineage and its members remain unknown. Our LMO2 Mass spectrometry analysis in murine myeloid progenitor cells (PUER) revealed the interacting partner TAL1 in addition to LDB1. Consistent with what we discovered in T-ALL cells, PHF6 was revealed as a novel interactor with LMO2 in these cells. Previous studies have also implied that PHF6 physically interacts with other proteins and complexes, such as NFκB (Soto-Feliciano et al., 2017), the NuRD complex (Todd and Picketts, 2012), the PAF1 transcriptional elongation complex (Zhang et al., 2013) and UBF1 (Wang et al., 2013). These studies have shown that the interplay of PHF6 in these complexes augmented their capacity to regulate transcription. Despite finding a physical interaction between LMO2 and PHF6, it was challenging to explore their functional interaction. Obtaining robust PHF6 ChIP-seq data results was an ongoing challenge throughout this study and the set of data we generated showed enrichment of highly repetitive regions. These regions were mostly marked by possessing long sequence motifs. Because of the shortcomings of this dataset we were not able to conclude whether PHF6 co-localises with LMO2 at genomic regions that possess regulatory elements and consequently regulate gene expression. Comparably, one group reported the physical interaction between PHF6 and NFκB (Soto-Feliciano et al., 2017), but another group was not able to reproduce these results (Miyagi et al., 2019). Instead ChIP-seq experiments by the latter group revealed a functional interaction between PHF6 and NFκB where the peaks significantly overlapped and were found in

intergenic regions, introns, and promoters (Miyagi et al., 2019). This might be due to technical limitations or because different cell models were used in these studies. It is also common that some antibodies produce successful immunoprecipitation results but do not work for ChIP and vice versa.

#### **4.3.2 Knockdown of *Phf6* shows signs of morphological abnormalities and delayed differentiation**

The first thing we noticed in *Phf6* knockdown cells were the morphological differences compared to the control cells. Allowing the cells to differentiate by treatment with tamoxifen showed incapability of shPhf6 cells to acquire morphological features of macrophages. They had a distinctive appearance of very large cells that were multi-nucleated, had dense cytoplasm, and minimal vacuoles.

PU.1 instructs progenitors to upregulate myeloid specific genes, including cell surface markers, and to downregulate other cell specific antigens and transcription factors (Nerlov and Graf, 1998). The vast majority of gene expression changes to allow myeloid cells to differentiate into macrophages occur in the first few days of PU.1 expression. Through examining macrophage surface markers, we observed latency in the expression of F4/80 differentiation antigen among shPhf6 cells. The expression of PU.1 via the addition of tamoxifen was not enough to allow for the macrophage maturation in shPhf6 cells at the same rate as shCtrl cells. Therefore, *Phf6* depletion slowed down the differentiation of myeloid progenitors into macrophages. This suggests that PHF6 might be required by myeloid cells to ensure a normal differentiation process. However, future studies should investigate whether PHF6 is more important earlier in myeloid differentiation than later, or for the initiation of differentiation. It is also worth exploring whether the delay in differentiation caused by

*Phf6* deprivation affects the macrophages phagocytic function. This can be tested through methods based on engulfment of fluorescently-labelled latex mini-beads by cultured macrophages. The phagocytic activity can be assessed either by counting individual cells using a fluorescence microscope or measuring fluorescence intensity using a flow cytometer (Sharma et al., 2014).

#### **4.3.3 *Phf6* depletion affects genes involved in haematopoiesis and chromatin assembly**

Although lack of *Phf6* expression resulted in slowly differentiating macrophages with irregular morphology, the RNA-seq data revealed that the total number of differentially expressed genes was restricted to 530 throughout all days of differentiation. The genes that were affected by *Phf6* loss were involved in regulation of myeloid cell differentiation, regulation of transcription factor activity, regulation of cell proliferation and regulation of DNA binding. Because PHF6 might exert a transcriptional regulatory function it may have targeted genes that are crucial to ensure proper development of myeloid cells and, as a consequence, *Phf6* deprived cells were not able to develop at the same rate as wild type cells.

Likewise, several studies have found a link between PHF6 and transcriptional regulation of haematopoietic genes. Despite shortcomings of ChIP-seq experiments on Jurkat T-ALL cells as discussed in chapter 3 section 3.3.5, these data revealed PHF6 binding to a set of target genes associated with haematopoiesis including *RUNX1*, *DMNT3A*, *NOTCH1*, and *JAG1*. The suppression of *Phf6* also resulted in a significant decrease in *Runx1* and *Notch* levels in B-ALL cells (Meacham et al., 2015). It has also been suggested that PHF6 directs the NuRD complex to a subset of gene targets that promote the development of hematopoietic cells through lineage



commitment or maturation (Todd and Picketts, 2012). Moreover, integrated genomics revealed PHF6 binding to nucleosomes flanking the TSSs of B-cell-specific genes, which allows B-cell-specific transcription factors to bind and activate genes in B-ALL cells. Conversely, PHF6 binding to nucleosomes surrounding the TSSs of T-cell-specific genes promoted chromatin compaction, thereby restricting the binding of T-cell-specific transcription factors. Consequently, the deletion of *PHF6* led to the downregulation of B-cell-specific genes and ectopic activation of the T-cell program in B-ALL cell lines (Soto-Feliciano et al., 2017). In a recent study, *Phf6* deletion led to the suppression of a set of genes associated with TNF $\alpha$  signalling and permitted resistance against the TNF $\alpha$ -mediated growth inhibition on HSCs by directly and indirectly activating the downstream effectors of TNF $\alpha$  signalling, such as *Nr4a1*, *Egr1*, and *Junb*. TNF $\alpha$  is a pro-inflammatory cytokine that drives the immune response and is induced by haematopoietic stress (Miyagi et al., 2019). Indeed, our data showed that among the upregulated genes in shPhf6 cells were TFs involved in haematopoiesis including *Egr1*. This gene encodes the transcription factor EGR1 which is crucial for normal myeloid cell differentiation through transcriptional regulation (Tian et al., 2016). The mechanisms for finely controlling TFs activity to ensure proper development of the cells are not confined to transcriptional control of TF genes. The presence of posttranslational modifications, such as acetylation, methylation, phosphorylation also plays roles in regulating TF activities. Therefore, it is not certain that *Egr1* is a direct target of PHF6 and the possibility of PHF6 contribution with the NuRD complex in posttranslational modifications in myeloid cells should be taken into consideration.

Remarkably, we also identified differential expression of genes responsible for chromatin assembly, protein-DNA complex assembly, nucleosome assembly and

organisation. These genes are essential for the chromatin structure which plays an important role in all nuclear processes, such as DNA repair, replication, transcription, and recombination. Altogether, this highlights an indication that PHF6 might play a transcriptional regulatory function and targets genes that are required for hematopoietic cell development. Additionally, this suggests that PHF6 may target genes that ensure a proper maintenance of chromatin structure and nucleosome organisation and assembly in myeloid cells.

#### **4.3.4 *Phf6* depletion repositions LMO2, TAL1, and LDB1 binding sites**

Insight into the distribution of LMO2, TAL1 and LDB1 bound sites across the genome of myeloid cells was achieved through ChIP-seq experiments. The intersection of ChIP-seq data revealed that LMO2 associated with TAL1 and LDB1 in 45 regions within the genome. These regions maintained the occupancy of LMO2/TAL1/LDB1 after *Phf6* knockdown. By contrast, LMO2 repositioned into a novel set of genomic regions after knockdown of *Phf6*. In addition, TAL1 and LDB1 showed a greater number of peaks in shPhf6 than in shCtrl cells.

Motif enrichment analyses were performed to investigate whether a specific motif was enriched or whether binding of other TFs might contribute to the establishment of the DNA binding patterns of LMO2, TAL1 and LDB1. The motifs that were bound by LMO2 and LDB1 were replaced by different bound sequences after *Phf6* depletion. TAL1 bound regions were consistent in shCtrl and shPhf6 cells, with some different motifs following PHF6 loss. However, we found the motif of Yin Yang 1 (YY1) transcription factor exclusively enriched in LMO2 ChIP-seq in shCtrl cells. The YY1 binding motif was not significantly enriched in cells lacking PHF6. The YY1 transcription factor is known to occupy interacting enhancers and promoters (Gordon et al., 2006). It can

form homodimers, thereby facilitating DNA loop formation, bringing distant regulatory regions in close proximity. Depletion of YY1 can disrupt enhancer-promoter contacts and normal gene expression (Weintraub et al., 2017). Thus, the participation of LMO2 in a complex that involves YY1 through which it binds to DNA and regulates gene expression should not be neglected.

The motif analyses mainly identified motifs with complex sequences, and we did not observe enrichment of hematopoietic transcription factors motifs like ETS, GATA or RUNX. Although LMO2/TAL1/LDB1 common peaks occupied the same regions before and after *Phf6* knockdown the motif sequences were different in shCtrl than in shPhf6. One possible explanation is that the number of overlapping peaks is very low which might not be enough to be very robust to identify specifically enriched motifs and making a bias towards longer motifs as they are more easily identified as significant. It was also puzzling why LMO2, TAL1, and LDB1 almost completely relocated their binding sites after *Phf6* knockdown. It is possible that the regions where these factors can bind were multiplied, deleted, or translocated due to the chromosomal abnormalities caused by PHF6 deprivation which will be discussed in the next section. Thus, peaks were called in shPhf6 cells at distinct binding sites from the regions found in the wild type cells.

#### **4.3.5 Genome and chromosome stability are affected by *Phf6* and *Lmo2* expression levels**

Having explored the potential regulatory functions of PHF6 in addition to observing the effect of PHF6 depletion on cell morphology and differentiation, we further investigated chromosomal integrity prior to and after *Phf6* knockdown through cytogenetic tests. Furthermore, cytogenetic experiments were also performed using shLmo2 and HA-

LMO2 cell lines to compare whether *Lmo2* depletion or overexpression have similar effect to *Phf6* deprivation in myeloid cells.

Our imaging analyses revealed very high numbers of chromosomes in *Phf6* and *Lmo2* depleted cells, as well as in LMO2 overexpressing cells compared to the control cells. It was difficult to distinguish whether the chromosomes were polyploidy or fragmented and some metaphases showed likelihood of having pulverized chromosomes. Premature chromosome compaction is the mechanism of pulverization by which compaction of partially replicated chromosomes is induced by mitotic cyclin-dependent kinase (CDK) activity (Obe et al., 1975). Chromosome pulverization has been demonstrated in cell-fusion experiments where chromosomes from an S-phase cell are pulverized due to exposure to signals from mitotic cytoplasm (Johnson and Rao, 1970). Pulverization has also been reported when aberrant late-replicating chromosome translocations are produced (Smith et al., 2001). This indicates that PHF6 and a balanced expression of LMO2 may be required to maintain genome and chromosome stability in myeloid cells.

These assumptions are corroborated by other studies on the individual roles of LMO2 and PHF6 in replication and DNA damage. Sincennes *et al.*, reported the influence of LMO2 on cell cycle progression and DNA replication in hematopoietic cells through its association with three replication proteins, mini-chromosome 6 (MCM6), DNA primase (PRIM1), and DNA polymerase delta (POLD1) (Sincennes et al., 2016). Moreover, previous studies showed that *PHF6* depletion causes increased genomic instability, accumulation of the DNA damage marker  $\gamma$ H2AX and cell cycle delay at G2/M (Wang et al., 2013). Also, knockdown of *PHF6* in HEK 293T cells resulted in elevated levels of  $\gamma$ H2AX (Van Vlierberghe et al., 2010). Therefore, we postulate that DNA damage is

correlated to the chromosomal abnormalities we observed upon *Phf6* and *Lmo2* suppression or LMO2 overexpression.

It has been reported that loss of *Phf6* promoted significant cell growth in AML, while it had a neutral effect in T-cell lymphomas, suggesting differential requirements for *Phf6* gene function in T-cell, and myeloid cancers (Meacham et al., 2015). *PHF6* mutations have also been reported in patients with AML (Van Vlierberghe et al., 2011, Cancer Genome Atlas Research et al., 2013). Cancers acquiring *PHF6* mutations have not been correlated with any specific myeloid progenitor cell type, however, higher *PHF6* expression levels were found in HSCs and megakaryocyte/erythroid progenitors than in common myeloid progenitors or granulocyte/macrophage progenitors (Van Vlierberghe et al., 2011).

Additionally, the role of LMO2 in myeloid cells is not well defined although LMO2 was found overexpressed in 51.1% of AML patients, compared to normal bone marrow donors (Meng et al., 2009). Immunostainings performed on AML patient bone marrow biopsies showed that LMO2 is expressed in 77% of the cases, but it was not a predictor of overall survival (Cobanoglu et al., 2010). Nonetheless, LMO2 upregulation has been correlated to overall survival and haematological remission in a group of chronic myeloid leukaemia (CML) patients undergoing therapy (Sonmez et al., 2009). Furthermore, *Lmo2* overexpression has been reported in the evolution of leukaemia stem cell activity and disease generation in a murine model of AML (Kvinlaug et al., 2011).

Cancer cells adapt to external stress by acquiring mechanisms such as DNA mutation, chromosomal translocation, gene amplification and chromosomal instability (Lengauer et al., 1998). Chromosomal instability is a feature of cancer genomes and can directly

lead to DNA damage (Janssen et al., 2011). Although the DNA damage experiments, we performed were not conclusive, our findings underline the value of these mouse cell models to better understand the cellular and molecular functions that *Phf6* and *Lmo2* play in myeloid cells. They also serve as a relevant tool to gain insights into human cancer. A useful approach to further understand the counter effects of *Phf6* and *Lmo2* deprivation is gene rescue experiments. These experiments enable the regain of the knockdown effects by adding shRNA rescue construct. I have already designed shPhf6 resistant sequences with multiple nucleotide mismatches which will allow the rescue of the shPhf6 phenotype. Due to time limitations these experiments were not further conducted but are within the future plans. After the reconstitution of PHF6 the cells can be assessed for gene expression, chromosome stability, and DNA damage which will clarify the specific effects of *Phf6* depletion in myeloid cells. These future studies in combination with the data we generated in addition to *Lmo2* rescue experiments will provide more evidence of *Phf6* and *Lmo2* important functions in myeloid cells and their potential mechanisms in cancer.

## **5 Chapter 5: Overall discussion**

### **5.1 Role of the LMO2/TAL1 complex in T-ALL**

Previous studies have identified an LMO2 complex which involves TAL1 and LDB1 in T-ALL cells that resembles the complex described in erythroid cells. The bipartite DNA motif recognised by the LMO2 complex in erythrocytes is an E-box followed downstream by a GATA motif (Wadman et al., 1997). Alternatively, the LMO2 complex described in T-ALL binds to DNA through a dual E-box motif (Grutz et al., 1998). However, this complex has not been extensively studied in T-ALL cells and its oncogenic function remains widely unknown.

Indeed, analyses in this study performed on T-ALL cells reinforced the notion that LMO2 functionally co-operates with TAL1 protein to regulate genes that are essential for basic cellular functions, transcriptional regulation, RNA processing, cellular biosynthetic processes, and metabolic processes. Many genes were identified as potential targets of the LMO2/TAL1 complex in the different T-ALL cell lines through the complex binding to distinct binding sites. These observations indicated that the LMO2/TAL1 complex may regulate the expression of a specific gene through binding to different regulatory elements. Different potential targets of the complex were also confined to particular cell lines. Hence, we suggest that a transcriptional regulatory function of the LMO2/TAL1 complex may allow tailored regulation of genes according to variable cell phenotypes and their dissimilar requirements leading to survival and proliferation. The genome wide data obtained from T-ALL cell lines that are blocked at different stages of differentiation represent a good model to investigate the role of LMO2 complex. It provided insight into differences between gene expression and regulation in these cells, despite being categorised in the TAL/LMO group of T-ALL.

This documentation of differences in regulation of gene expression is important in drug discovery processes for T-ALL besides studies of signalling and metabolic pathways.

## **5.2 Identifying PHF6 as a novel interactor of LMO2**

A recurring observation throughout this study was the interaction of PHF6 with LMO2. These findings were documented both in the context of human T-ALL cell lines (Chapter 3), as well as in the mouse myeloid cell line model PUER (Chapter 4). Furthermore, analyses of LMO2 and PHF6 binding using ChIP-seq datasets from human T-ALL cells revealed functional interactions between PHF6 and the LMO2/TAL1/LBD1 complex, which bind DNA through known haematopoietic TF motifs.

LMO2 is considered a transcriptional co-factor, given that it does not bind directly to DNA, but instead acts by forming a scaffold for DNA-binding proteins and other interacting proteins, creating a bipartite multi-protein DNA-binding complex (Lecuyer et al., 2007). The binding properties of PHF6 are still not well defined. It has been reported in a study that PHF6 is able to directly bind dsDNA *in vitro*. By conducting electrophoretic mobility shift assays (EMSA) to test whether the PHF6-ePHD2 domain directly binds to dsDNA, it was shown that PHF6 has a binding affinity to dsDNA via its C-terminal tail. However, PHF6 did not exhibit sequence preferences for AT- or GC-rich sequences (Liu et al., 2014).

In this study, the direct binding of PHF6 to DNA was not investigated in depth. Nonetheless, the *de novo* motif analyses of ChIP-seq datasets represented in human T-ALL cells did not identify recurrent significant motifs for PHF6 itself. It showed that PHF6 does not seem to recognise specific DNA-binding sequences, and the enriched



motifs were long DNA sequences. Similar findings were reported in B-ALL cells (Soto-Feliciano et al., 2017) and HSCs (Miyagi et al., 2019), where no specific DNA motifs were associated with PHF6 binding.

These observations raise the possibility that PHF6 promotes the binding of a TF to target genomic sites without directly interacting with the DNA itself. Indeed, this study and other studies have explored this prospect. PHF6 ChIP-seq datasets in this study showed enriched binding to TFs motifs involved in haematopoiesis including RUNX1, GATA, RUNX, TCF12 and SPI1, in addition to the unknown DNA sequences.

Soto *et al.* investigated whether PHF6 is part of a transcription factor complex in B-ALL cells. They analysed the  $\pm 1$  kb DNA regions surrounding the TSSs of differentially expressed genes in these cells and assessed the enrichment of known motifs. The enriched motifs were for TFs that are important during haematopoiesis and lymphopoiesis, including PU.1, NF $\kappa$ B, EGR1, EBF1, TCF3 (E2A), and TCF12 (HEB). They further confirmed that PHF6 interacts with TCF12 and NF $\kappa$ B in B-ALL cells through endogenous coimmunoprecipitation experiments. The study concluded that PHF6 does not act as a canonical transcription factor in B-ALL cells, because of its inability to recognise specific DNA-binding motifs (Soto-Feliciano et al., 2017). Experiments by Miyagi *et al.* could not confirm this physical interaction between PHF6 and NF $\kappa$ B in CML cells. However, they detected overlapping peaks of PHF6 and NF $\kappa$ B mainly in the intergenic regions, introns, and promoter TSSs through ChIP-seq analyses. They suggested a functional interaction between PHF6 and NF $\kappa$ B in CML cells, which exerts a regulatory function in these cells (Miyagi et al., 2019). Findings from this thesis provide novel evidence for physical and functional interaction between PHF6 and LMO2 complex in the context of human T-ALL, where PHF6 is recruited to

the DNA through TFs that are members of the LMO2 complex. Additionally, the physical interaction of PHF6 with LMO2 was identified in myeloid cells and further investigations into the functional interactions will be carried out through ChIP-seq experiments.

### **5.3 The transcriptional role of PHF6**

We and other groups have suggested that PHF6 possibly plays a transcriptional regulatory function and targets genes involved in haematopoiesis. In this study, the integrated ChIP-seq to RNA-seq data in T-ALL cells revealed many TF genes that are involved in haematopoiesis as potential targets of PHF6/LMO2/TAL1/LDB1 peaks. Moreover, among the differentially expressed genes in PUER myeloid cells after *Phf6* depletion were haematopoietic TFs.

Experiments on B-ALL cells showed that loss of *Phf6* caused suppression of *Runx1* and *Notch* levels (Meacham et al., 2015). Other experiments performed on *Phf6* knockout B-ALL cells, demonstrated enrichment of genes associated with T-cell signal transduction and function including *Tcf7*, *Bcl11b*, and *Gata3* (Soto-Feliciano et al., 2017). Recently, RNA-seq data identified a set of downregulated genes in *Phf6* deficient HSCs relative to wild type HSCs, indicating that PHF6 is primarily implicated in transcriptional activation in HSCs (Miyagi et al., 2019). Overall, this indicates that PHF6 might regulate transcription of genes necessary for haematopoietic cell differentiation and proliferation. Supporting this notion, we showed *HHEX* gene expression through PHF6 binding with the LMO2 complex members at active enhancer sites supported by adapted H3K27ac ChIP-seq dataset in DU.528 cells (Abraham et al., 2017). In a study using *CD2-Lmo2* mice, two different mechanisms through which the LMO2 complex can cause T-cell differentiation block has been described. One

involves the activation of *Hhex*, *Lyl1*, and *Mycn* resulting in an ETP-ALL like pathway and the other showing overexpression of *Notch1* (Smith et al., 2014). The findings described in this thesis build upon this concept and propose a mechanism of PHF6 collaboration with the LMO2 complex in activating genes that are expressed in immature thymocytes, thus enhancing self-renewal capacity of the cells and obstructing the upregulation of T-cell specific genes, leading to an accumulation of undifferentiated leukemic T-cells.

Additionally, the genome wide data revealed potential PHF6 targets, other than genes directly involved in haematopoiesis. PUER cells showed that the expression of genes involved in chromatin assembly, as well as nucleosome organisation and assembly, were altered upon *Phf6* depletion. Chromatin undergoes transient disruption and restoration to replicate parental chromatin during the S-phase (Krude and Keller, 2001). Chromatin restoration on newly synthesised DNA includes nucleosome assembly and remodelling, restoration of DNA methylation and histone marks, deposition of histone variants and establishment of higher order chromosomal structures including sister-chromatid cohesion. Chromatin replication facilitates an opportunity to change the chromatin structure, thereby altering the transcription factor profile on chromatin (Alabert and Groth, 2012, Margueron and Reinberg, 2010). A proteomic analysis reported PHF6 as a component of nascent chromatin, which may indicate a role for PHF6 in the loading of transcription factors on chromatin or the functional regulation of transcription factors in the S-phase (Alabert et al., 2014). These assumptions were supported by our observations of extreme effects on chromosome integrity upon knockdown of *Phf6* and *Lmo2* in myeloid cells.

Overall, the findings in this study could have two explanations, which are not mutually exclusive: (i) complexes composed of PHF6 bind DNA through TFs, thus regulating transcription of genes essential for haematopoietic development and/or chromatin structure; (ii) PHF6 may serve as an essential protein that participates in facilitating chromatin and chromosome packaging. These findings highlight the value of the genome wide data generated by this study and the advantages of utilising these mouse cell models to understand the biological and molecular roles of *Phf6* and *Lmo2* in normal cells and cancer cells. Thereafter, providing a useful tool to gain more insight into human leukaemia to find better treatments for this type of cancer.

#### **5.4 Limitations**

A first limitation, was the challenge to generate *PHF6* knockdown T-ALL cells despite the continuous efforts in applying different methods to silence *PHF6*, as discussed in chapter 3. T-ALL cell lines were difficult to transduce with lentiviral plasmid and did not proliferate after the integration of shPHF6 through a transposase-mediated method. Alternatively, the CRISPR-Cas9 system is an efficient way to manipulate gene expression and function by genome editing to study loss-of-function phenotypes (Sander and Joung, 2014). The system requires introducing two components: the nuclease Cas9 and a single-guide RNA (sgRNA) which targets the nuclease to the genomic locus of interest (Ran et al., 2013). Vehicles used to deliver the CRISPR-Cas9 system include: (i) physical delivery methods, such as microinjection and electroporation, while methods such as hydrodynamic delivery are still under investigation, (ii) viral delivery vectors, including specifically engineered adeno-associated virus (AAV), and full-sized adenovirus and lentivirus vehicles, (iii) non-viral vector delivery vehicles, such as lipid nanoparticles, cell-penetrating peptides, DNA

‘nanoclews’, and gold nanoparticles (Lino et al., 2018). Because there are variable delivery methods for the CRISPR-Cas9 system it is an appealing technique to overcome the drawbacks we encountered by using the shRNA method. It can also be the system of choice for cells that are difficult to transduce and proliferate like the T-ALL cell lines utilised in this study.

A second limitation, there was no corresponding normal sample for comparison in T-ALL because LMO2 is not expressed in normal T-cells beyond the DN2 stage. Therefore, studying the LMO2 complex and the interaction with PHF6 was only in the context of cancer T-cells.

A third limitation, PHF6 ChIP-seq was an ongoing challenge throughout this study because the quality of ChIP-seq data depended on the quality of the antibody used for immunoprecipitation. Validation of the PHF6 antibody for ChIP-seq experiments was laborious and time consuming. Thus, distinct sets of PHF6 ChIP-seq data generated different results. This can be expected when using different antibodies because antibody epitopes may be masked or partially masked at some gene loci depending on other DNA-bound proteins and the extent of masking can vary across different gene loci. Successful sets of PHF6 ChIP-seq data in T-ALL cells were produced eventually as reported in chapter 3. However, the PHF6 ChIP-seq experiments in the PUER myeloid cells were only performed using an antibody that did not seem to be suitable for ChIP experiments (data not shown). Taken that ChIP experiments yield better results using the antibody used in T-ALL cells, we speculate that PHF6 ChIP with this antibody will identify more genomic regions associated with LMO2. This will allow the investigation of PHF6 and LMO2 functional interaction in myeloid cells, which will shed the light on the role of PHF6 in myeloid cells development.

A fourth limitation was the DNaseI-seq experiments in myeloid cells (data not shown) which generated a very limited number of DHSs which were not considered informative. A fair amount of pre-testing to identify the right incubation conditions was performed, but the sequencing results showed under-digestion which influenced the detection of open chromatin sites. One solution is to further optimise the conditions for the myeloid cells and repeat the DNaseI-seq experiment. Another solution is to apply the assay for transposase-accessible chromatin (ATAC-seq) technique which is also used to detect open chromatin regions. ATAC-seq has some experimental advantages over DNaseI-seq. The ATAC-seq assay and library preparation are carried out in a single enzymatic step and requires a very low number of cells which enables rapid library generation (Buenrostro et al., 2013). However, DNaseI-seq is the method of choice when sufficient cells are available because it generates better data.

Lastly, *Phf6* and *Lmo2* rescue experiments should be conducted in myeloid cells as discussed in chapter 4. These experiments will provide useful information regarding the chromosome instability, and changes in gene expression that occurred in the knockdown phenotypes. This can give more insight into *Phf6* and *Lmo2* roles in normal myeloid development and highlight their possible oncogenic roles.

## **5.5 Final conclusions**

To the best of our knowledge, this study has shown the first integrative genome-wide analysis of chromatin elements and gene expression as related to expression of PHF6, in both human T-ALL cells and in mouse myeloid cells. A novel interaction between PHF6 and LMO2 in these lymphoid and myeloid cells has been identified.

Relating to the ChIP-seq results in T-ALL cells, we assume that PHF6/LMO2 are team players in a complex that includes TAL1 and LDB1, which plays a transcriptional regulatory function that regulates genes involved in haematopoietic lineage determination and differentiation.

Although the analyses are not yet conclusive, we have postulated from several independent lines of investigation that PHF6 may represent a factor that plays a major transcriptional regulatory function. This is of potential relevance to cancers where PHF6 have been implicated as a tumour suppressor. Mutations of *PHF6* have been described in T-ALL, AML, CML and hepatocellular carcinoma patient tumours (Van Vlierberghe et al., 2010, Van Vlierberghe et al., 2011, Yoo et al., 2012, Li et al., 2013b).

In this study, PHF6 seems to promote expression of genes involved in the development of haematopoietic cells, chromatin and nucleosome structure maintenance and DNA repair. However, the genome-wide events accompanying these changes have yet to be explored. Moreover, *Phf6* suppression caused changes in gene expression, chromosomal abnormality, and delayed differentiation. Further characterisation of PHF6 and PHF6/LMO2 interactions will need to focus on their participation in different complexes and what their functions are in normal haematopoiesis, in order to fully understand how they contribute to oncogenesis.

## **5.6 Future perspectives**

To answer the remaining questions throughout this study, it would be useful to follow up on the findings by applying the following:

1. A different approach such as *CRISPR-CAS9* mediated PHF6 knockout in order to deplete *PHF6* in T-ALL cells. This will enable to further investigate the role of

PHF6 and the PHF6/LMO2 complex in oncogenesis. Which can be followed up by RNA-seq, DNA damage, and chromosome spread experiments.

2. Extend the findings of chromosome breakage and genome instability in *Phf6* and *Lmo2* deprived myeloid cells by investigating their role in cell cycle progression and DNA damage. This can be done by performing immunostainings coupled with western blotting to test DNA damage markers such as 53BP1 and  $\gamma$ H2AX in these cells compared to wild type cells. The immunostainings of these DNA damage markers can be combined with EdU labelling followed by flowcytometry analysis to count the cells in S-phase and determine the accumulation of DNA damage.
3. The rescue of *Phf6* in shPhf6 cells to study the consequences of PHF6 reconstitution on gene expression, cell cycle progression, and DNA repair.
4. RNA-seq experiments for shLmo2 and HA-Lmo2 cells have already been performed and the sequences ready for analysis to be compared to shCtrl and shPhf6 RNA-seq. ChIP experiments with PHF6 antibody in shCtrl have also been performed and being analysed, after which the data will be integrated with the RNA-seq data to further explore the functions of *Phf6* in these myeloid cells.



## References

- ABRAHAM, B. J., HNISZ, D., WEINTRAUB, A. S., KWIATKOWSKI, N., LI, C. H., LI, Z., WEICHERT-LEAHEY, N., RAHMAN, S., LIU, Y., ETCHIN, J., LI, B., SHEN, S., LEE, T. I., ZHANG, J., LOOK, A. T., MANSOUR, M. R. & YOUNG, R. A. 2017. Small genomic insertions form enhancers that misregulate oncogenes. *Nat Commun*, 8, 14385.
- ADOLFSSON, J., BORGE, O. J., BRYDER, D., THEILGAARD-MONCH, K., ASTRAND-GRUNDSTROM, I., SITNICKA, E., SASAKI, Y. & JACOBSEN, S. E. 2001. Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity*, 15, 659-69.
- ADOLFSSON, J., MANSSON, R., BUZA-VIDAS, N., HULTQUIST, A., LIUBA, K., JENSEN, C. T., BRYDER, D., YANG, L., BORGE, O. J., THOREN, L. A., ANDERSON, K., SITNICKA, E., SASAKI, Y., SIGVARDSSON, M. & JACOBSEN, S. E. 2005. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell*, 121, 295-306.
- AIFANTIS, I., RAETZ, E. & BUONAMICI, S. 2008. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nat Rev Immunol*, 8, 380-90.
- AKASHI, K., TRAVER, D., MIYAMOTO, T. & WEISSMAN, I. L. 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, 404, 193-7.
- ALABERT, C., BUKOWSKI-WILLS, J. C., LEE, S. B., KUSTATSCHER, G., NAKAMURA, K., DE LIMA ALVES, F., MENARD, P., MEJLVANG, J., RAPPSILBER, J. & GROTH, A. 2014. Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nat Cell Biol*, 16, 281-93.
- ALABERT, C. & GROTH, A. 2012. Chromatin replication and epigenome maintenance. *Nat Rev Mol Cell Biol*, 13, 153-67.
- ANDERSON, K. P., CRABLE, S. C. & LINGREL, J. B. 1998. Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. *J Biol Chem*, 273, 14347-54.
- ANDERSON, M. K. 2006. At the crossroads: diverse roles of early thymocyte transcriptional regulators. *Immunol Rev*, 209, 191-211.
- AOYAMA, M., OZAKI, T., INUZUKA, H., TOMOTSUNE, D., HIRATO, J., OKAMOTO, Y., TOKITA, H., OHIRA, M. & NAKAGAWARA, A. 2005. LMO3 interacts with neuronal transcription factor, HEN2, and acts as an oncogene in neuroblastoma. *Cancer Res*, 65, 4587-97.
- APLAN, P. D., JONES, C. A., CHERVINSKY, D. S., ZHAO, X., ELLSWORTH, M., WU, C., MCGUIRE, E. A. & GROSS, K. W. 1997. An scl gene product lacking the transactivation domain induces bony abnormalities and cooperates with LMO1 to generate T-cell malignancies in transgenic mice. *EMBO J*, 16, 2408-19.
- APPERT, A., NAM, C. H., LOBATO, N., PRIEGO, E., MIGUEL, R. N., BLUNDELL, T., DRYNAN, L., SEWELL, H., TANAKA, T. & RABBITS, T. 2009. Targeting LMO2 with a peptide aptamer establishes a necessary function in overt T-cell neoplasia. *Cancer Res*, 69, 4784-90.

- ATCHLEY, W. R. & FITCH, W. M. 1997. A natural classification of the basic helix-loop-helix class of transcription factors. *Proc Natl Acad Sci U S A*, 94, 5172-6.
- BACH, I. 2000. The LIM domain: regulation by association. *Mech Dev*, 91, 5-17.
- BAER, R. 1993. TAL1, TAL2 and LYL1: a family of basic helix-loop-helix proteins implicated in T cell acute leukaemia. *Semin Cancer Biol*, 4, 341-7.
- BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W. & NOBLE, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*, 37, W202-8.
- BASSING, C. H., SWAT, W. & ALT, F. W. 2002. The mechanism and regulation of chromosomal V(D)J recombination. *Cell*, 109 Suppl, S45-55.
- BASTA, J. & RAUCHMAN, M. 2015. The nucleosome remodeling and deacetylase complex in development and disease. *Transl Res*, 165, 36-47.
- BAUMSTARK, A., LOWER, K. M., SINKUS, A., ANDRIUSKEVICIUTE, I., JURKENIENE, L., GECZ, J. & JUST, W. 2003. Novel PHF6 mutation p.D333del causes Borjeson-Forssman-Lehmann syndrome. *J Med Genet*, 40, e50.
- BECKER, A. J., MC, C. E. & TILL, J. E. 1963. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature*, 197, 452-4.
- BELVER, L. & FERRANDO, A. 2016. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat Rev Cancer*, 16, 494-507.
- BLANKENBERG, D., VON KUSTER, G., CORAOR, N., ANANDA, G., LAZARUS, R., MANGAN, M., NEKRUTENKO, A. & TAYLOR, J. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol*, Chapter 19, Unit 19 10 1-21.
- BLOM, B. & SPITS, H. 2006. Development of human lymphoid cells. *Annu Rev Immunol*, 24, 287-320.
- BLOM, B., VERSCHUREN, M. C., HEEMSKERK, M. H., BAKKER, A. Q., VAN GASTEL-MOL, E. J., WOLVERS-TETTERO, I. L., VAN DONGEN, J. J. & SPITS, H. 1999. TCR gene rearrangements and expression of the pre-T cell receptor complex during human T-cell differentiation. *Blood*, 93, 3033-43.
- BOEHM, T., BAER, R., LAVENIR, I., FORSTER, A., WATERS, J. J., NACHEVA, E. & RABBITS, T. H. 1988. The mechanism of chromosomal translocation t(11;14) involving the T-cell receptor C delta locus on human chromosome 14q11 and a transcribed region of chromosome 11p15. *EMBO J*, 7, 385-94.
- BOEHM, T., FORONI, L., KANEKO, Y., PERUTZ, M. F. & RABBITS, T. H. 1991. The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. *Proc Natl Acad Sci U S A*, 88, 4367-71.
- BOEHM, T., FORONI, L., KENNEDY, M. & RABBITS, T. H. 1990. The rhombotin gene belongs to a class of transcriptional regulators with a potential novel protein dimerisation motif. *Oncogene*, 5, 1103-5.
- BOROS, J., O'DONNELL, A., DONALDSON, I. J., KASZA, A., ZEEF, L. & SHARROCKS, A. D. 2009. Overlapping promoter targeting by Elk-1 and other divergent ETS-domain transcription factor family members. *Nucleic Acids Res*, 37, 7368-80.
- BOWMAN, R. L., BUSQUE, L. & LEVINE, R. L. 2018. Clonal Hematopoiesis and Evolution to Hematopoietic Malignancies. *Cell Stem Cell*, 22, 157-170.

- BRAUNSTEIN, M. & ANDERSON, M. K. 2012. HEB in the spotlight: Transcriptional regulation of T-cell specification, commitment, and developmental plasticity. *Clin Dev Immunol*, 2012, 678705.
- BREIT, S., STANULLA, M., FLOHR, T., SCHRAPPE, M., LUDWIG, W. D., TOLLE, G., HAPPICH, M., MUCKENTHALER, M. U. & KULOZIK, A. E. 2006. Activating NOTCH1 mutations predict favorable early treatment response and long-term outcome in childhood precursor T-cell lymphoblastic leukemia. *Blood*, 108, 1151-7.
- BRIDWELL, J. A., PRICE, J. R., PARKER, G. E., MCCUTCHAN SCHILLER, A., SLOOP, K. W. & RHODES, S. J. 2001. Role of the LIM domains in DNA recognition by the Lhx3 neuroendocrine transcription factor. *Gene*, 277, 239-50.
- BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*, 10, 1213-8.
- CANCER GENOME ATLAS RESEARCH, N., LEY, T. J., MILLER, C., DING, L., RAPHAEL, B. J., MUNGALL, A. J., ROBERTSON, A., HOADLEY, K., TRICHE, T. J., JR., LAIRD, P. W., BATY, J. D., FULTON, L. L., FULTON, R., HEATH, S. E., KALICKI-VEIZER, J., KANDOTH, C., KLCO, J. M., KOBOLDT, D. C., KANCHI, K. L., KULKARNI, S., LAMPRECHT, T. L., LARSON, D. E., LIN, L., LU, C., MCLELLAN, M. D., MCMICHAEL, J. F., PAYTON, J., SCHMIDT, H., SPENCER, D. H., TOMASSON, M. H., WALLIS, J. W., WARTMAN, L. D., WATSON, M. A., WELCH, J., WENDL, M. C., ALLY, A., BALASUNDARAM, M., BIROL, I., BUTTERFIELD, Y., CHIU, R., CHU, A., CHUAH, E., CHUN, H. J., CORBETT, R., DHALLA, N., GUIN, R., HE, A., HIRST, C., HIRST, M., HOLT, R. A., JONES, S., KARSAN, A., LEE, D., LI, H. I., MARRA, M. A., MAYO, M., MOORE, R. A., MUNGALL, K., PARKER, J., PLEASANCE, E., PLETTNER, P., SCHEIN, J., STOLL, D., SWANSON, L., TAM, A., THIESSEN, N., VARHOL, R., WYE, N., ZHAO, Y., GABRIEL, S., GETZ, G., SOUGNEZ, C., ZOU, L., LEISERSON, M. D., VANDIN, F., WU, H. T., APPLEBAUM, F., BAYLIN, S. B., AKBANI, R., BROOM, B. M., CHEN, K., MOTTER, T. C., NGUYEN, K., WEINSTEIN, J. N., ZHANG, N., FERGUSON, M. L., ADAMS, C., BLACK, A., BOWEN, J., GASTIER-FOSTER, J., GROSSMAN, T., LICHTENBERG, T., WISE, L., DAVIDSEN, T., DEMCHOK, J. A., SHAW, K. R., SHETH, M., SOFIA, H. J., YANG, L., DOWNING, J. R., et al. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*, 368, 2059-74.
- CASTELLO, A., FISCHER, B., EICHELBAUM, K., HOROS, R., BECKMANN, B. M., STREIN, C., DAVEY, N. E., HUMPHREYS, D. T., PREISS, T., STEINMETZ, L. M., KRIJGSVELD, J. & HENTZE, M. W. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 149, 1393-406.
- CAVE, H., SUCIU, S., PREUDHOMME, C., POPPE, B., ROBERT, A., UYTTEBROECK, A., MALET, M., BOUTARD, P., BENOIT, Y., MAUVIEUX, L., LUTZ, P., MECHINAUD, F., GRARDEL, N., MAZINGUE, F., DUPONT, M., MARGUERITTE, G., PAGES, M. P., BERTRAND, Y., PLOUVIER, E., BRUNIE, G., BASTARD, C., PLANTAZ, D., VANDE VELDE, I., HAGEMEIJER, A., SPELEMAN, F., LESSARD, M., OTTEN, J., VILMER, E.,

- DASTUGUE, N. & EORTC, C. L. G. 2004. Clinical significance of HOX11L2 expression linked to t(5;14)(q35;q32), of HOX11 expression, and of SIL-TAL fusion in childhood T-cell malignancies: results of EORTC studies 58881 and 58951. *Blood*, 103, 442-50.
- CHAMBERS, J. & RABBITS, T. H. 2015. LMO2 at 25 years: a paradigm of chromosomal translocation proteins. *Open Biol*, 5, 150062.
- CHEN, F., ROWEN, L., HOOD, L. & ROTHENBERG, E. V. 2001. Differential transcriptional regulation of individual TCR V beta segments before gene rearrangement. *J Immunol*, 166, 1771-80.
- CHEN, M. J., LI, Y., DE OBALDIA, M. E., YANG, Q., YZAGUIRRE, A. D., YAMADA-INAGAWA, T., VINK, C. S., BHANDoola, A., DZIERZAK, E. & SPECK, N. A. 2011. Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell*, 9, 541-52.
- CHERVINSKY, D. S., ZHAO, X. F., LAM, D. H., ELLSWORTH, M., GROSS, K. W. & APLAN, P. D. 1999. Disordered T-cell development and T-cell malignancies in SCL LMO1 double-transgenic mice: parallels with E2A-deficient mice. *Mol Cell Biol*, 19, 5025-35.
- CHIARETTI, S. & FOA, R. 2009. T-cell acute lymphoblastic leukemia. *Haematologica*, 94, 160-2.
- CHRISTENSEN, J. L. & WEISSMAN, I. L. 2001. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc Natl Acad Sci U S A*, 98, 14541-6.
- CIAU-UITZ, A., WANG, L., PATIENT, R. & LIU, F. 2013. ETS transcription factors in hematopoietic stem cell development. *Blood Cells Mol Dis*, 51, 248-55.
- CIOFANI, M. & ZUNIGA-PFLUCKER, J. C. 2005. Notch promotes survival of pre-T cells at the beta-selection checkpoint by regulating cellular metabolism. *Nat Immunol*, 6, 881-8.
- CIOFANI, M. & ZUNIGA-PFLUCKER, J. C. 2007. The thymus as an inductive site for T lymphopoiesis. *Annu Rev Cell Dev Biol*, 23, 463-93.
- CLEVELAND, S. M., GOODINGS, C., TRIPATHI, R. M., ELLIOTT, N., THOMPSON, M. A., GUO, Y., SHYR, Y. & DAVE, U. P. 2014. LMO2 induces T-cell leukemia with epigenetic deregulation of CD4. *Exp Hematol*, 42, 581-93 e5.
- CLEVELAND, S. M., SMITH, S., TRIPATHI, R., MATHIAS, E. M., GOODINGS, C., ELLIOTT, N., PENG, D., EL-RIFAI, W., YI, D., CHEN, X., LI, L., MULLIGHAN, C., DOWNING, J. R., LOVE, P. & DAVE, U. P. 2013. Lmo2 induces hematopoietic stem cell-like features in T-cell progenitor cells prior to leukemia. *Stem Cells*, 31, 882-94.
- COBANOGU, U., SONMEZ, M., OZBAS, H. M., ERKUT, N. & CAN, G. 2010. The expression of LMO2 protein in acute B-cell and myeloid leukemia. *Hematology*, 15, 132-4.
- COCKERILL, P. N. 2011. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J*, 278, 2182-210.
- CONDORELLI, G. L., FACCHIANO, F., VALTIERI, M., PROIETTI, E., VITELLI, L., LULLI, V., HUEBNER, K., PESCHLE, C. & CROCE, C. M. 1996. T-cell-directed TAL-1 expression induces T-cell malignancies in transgenic mice. *Cancer Res*, 56, 5113-9.

- CONROY, L. A. & ALEXANDER, D. R. 1996. The role of intracellular signalling pathways regulating thymocyte and leukemic T cell apoptosis. *Leukemia*, 10, 1422-35.
- CRAWFORD, J., LOWER, K. M., HENNEKAM, R. C., VAN ESCH, H., MEGARBANE, A., LYNCH, S. A., TURNER, G. & GECZ, J. 2006. Mutation screening in Borjeson-Forssman-Lehmann syndrome: identification of a novel de novo PHF6 mutation in a female patient. *J Med Genet*, 43, 238-43.
- CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, 107, 21931-6.
- CROSS, A. J., JEFFRIES, C. M., TREWHELLA, J. & MATTHEWS, J. M. 2010. LIM domain binding proteins 1 and 2 have different oligomeric states. *J Mol Biol*, 399, 133-44.
- CURTIS, D. J. & MCCORMACK, M. P. 2010. The molecular basis of Lmo2-induced T-cell acute lymphoblastic leukemia. *Clin Cancer Res*, 16, 5618-23.
- DAVID-FUNG, E. S., BUTLER, R., BUZI, G., YUI, M. A., DIAMOND, R. A., ANDERSON, M. K., ROWEN, L. & ROTHENBERG, E. V. 2009. Transcription factor expression dynamics of early T-lymphocyte specification and commitment. *Dev Biol*, 325, 444-67.
- DAWID, I. B., BREEN, J. J. & TOYAMA, R. 1998. LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *Trends Genet*, 14, 156-62.
- DEL REAL, M. M. & ROTHENBERG, E. V. 2013. Architecture of a lymphomyeloid developmental switch controlled by PU.1, Notch and Gata3. *Development*, 140, 1207-19.
- DELATTRE, O., ZUCMAN, J., PLOUGASTEL, B., DESMAZE, C., MELOT, T., PETER, M., KOVAR, H., JOUBERT, I., DE JONG, P., ROULEAU, G. & ET AL. 1992. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature*, 359, 162-5.
- DI DONATO, N., ISIDOR, B., LOPEZ CAZAUX, S., LE CAIGNEC, C., KLINK, B., KRAUS, C., SCHROCK, E. & HACKMANN, K. 2014. Distinct phenotype of PHF6 deletions in females. *Eur J Med Genet*, 57, 85-9.
- DIK, W. A., PIKE-OVERZET, K., WEERKAMP, F., DE RIDDER, D., DE HAAS, E. F., BAERT, M. R., VAN DER SPEK, P., KOSTER, E. E., REINDERS, M. J., VAN DONGEN, J. J., LANGERAK, A. W. & STAAL, F. J. 2005. New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J Exp Med*, 201, 1715-23.
- DOHNER, H., ESTEY, E. H., AMADORI, S., APPELBAUM, F. R., BUCHNER, T., BURNETT, A. K., DOMBRET, H., FENAUX, P., GRIMWADE, D., LARSON, R. A., LO-COCO, F., NAOE, T., NIEDERWIESER, D., OSSENKOPPELE, G. J., SANZ, M. A., SIERRA, J., TALLMAN, M. S., LOWENBERG, B., BLOOMFIELD, C. D. & EUROPEAN, L. 2010. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*, 115, 453-74.
- DOW, L. E., PREMSRIRUT, P. K., ZUBER, J., FELLMANN, C., MCJUNKIN, K., MIETHING, C., PARK, Y., DICKINS, R. A., HANNON, G. J. & LOWE, S. W.

2012. A pipeline for the generation of shRNA transgenic mice. *Nat Protoc*, 7, 374-93.
- DRAHEIM, K. M., HERMANCE, N., YANG, Y., AROUS, E., CALVO, J. & KELLIHER, M. A. 2011. A DNA-binding mutant of TAL1 cooperates with LMO2 to cause T cell leukemia in mice. *Oncogene*, 30, 1252-60.
- DZIERZAK, E. & BIGAS, A. 2018. Blood Development: Hematopoietic Stem Cell Dependence and Independence. *Cell Stem Cell*, 22, 639-651.
- EGAWA, T., TILLMAN, R. E., NAOE, Y., TANIUCHI, I. & LITTMAN, D. R. 2007. The role of the Runx transcription factors in thymocyte differentiation and in homeostasis of naive T cells. *J Exp Med*, 204, 1945-57.
- EL OMARI, K., HOOSDALLY, S. J., TULADHAR, K., KARIA, D., HALL-PONSELE, E., PLATONOVA, O., VYAS, P., PATIENT, R., PORCHER, C. & MANCINI, E. J. 2013. Structural basis for LMO2-driven recruitment of the SCL:E47bHLH heterodimer to hematopoietic-specific transcriptional targets. *Cell Rep*, 4, 135-47.
- ELLENBERGER, T., FASS, D., ARNAUD, M. & HARRISON, S. C. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev*, 8, 970-80.
- ENGEL, I., JOHNS, C., BAIN, G., RIVERA, R. R. & MURRE, C. 2001. Early thymocyte development is regulated by modulation of E2A protein activity. *J Exp Med*, 194, 733-45.
- ENGEL, I. & MURRE, C. 2004. E2A proteins enforce a proliferation checkpoint in developing thymocytes. *EMBO J*, 23, 202-11.
- ERKIZAN, H. V., UVERSKY, V. N. & TORETSKY, J. A. 2010. Oncogenic partnerships: EWS-FLI1 protein interactions initiate key pathways of Ewing's sarcoma. *Clin Cancer Res*, 16, 4077-83.
- ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R., COYNE, M., KU, M., DURHAM, T., KELLIS, M. & BERNSTEIN, B. E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473, 43-9.
- FERRANDO, A. A., NEUBERG, D. S., STAUNTON, J., LOH, M. L., HUARD, C., RAIMONDI, S. C., BEHM, F. G., PUI, C. H., DOWNING, J. R., GILLILAND, D. G., LANDER, E. S., GOLUB, T. R. & LOOK, A. T. 2002. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell*, 1, 75-87.
- FISHER, C. 2014. The diversity of soft tissue tumours with EWSR1 gene rearrangements: a review. *Histopathology*, 64, 134-50.
- FORONI, L., BOEHM, T., WHITE, L., FORSTER, A., SHERRINGTON, P., LIAO, X. B., BRANNAN, C. I., JENKINS, N. A., COPELAND, N. G. & RABBITS, T. H. 1992. The rhombotin gene family encode related LIM-domain proteins whose differing expression suggests multiple roles in mouse development. *J Mol Biol*, 226, 747-61.
- FRANZONI, E., BOOKER, S. A., PARTHASARATHY, S., REHFELD, F., GROSSER, S., SRIVATSA, S., FUCHS, H. R., TARABYKIN, V., VIDA, I. & WULCZYN, F. G. 2015. miR-128 regulates neuronal migration, outgrowth and intrinsic excitability via the intellectual disability gene Phf6. *Elife*, 4.
- GARCIA-OJEDA, M. E., KLEIN WOLTERINK, R. G., LEMAITRE, F., RICHARD-LE GOFF, O., HASAN, M., HENDRIKS, R. W., CUMANO, A. & DI SANTO, J. P.

2013. GATA-3 promotes T-cell specification by repressing B-cell potential in pro-T cells in mice. *Blood*, 121, 1749-59.
- GARCIA-RAMIREZ, I., BHATIA, S., RODRIGUEZ-HERNANDEZ, G., GONZALEZ-HERRERO, I., WALTER, C., GONZALEZ DE TENA-DAVILA, S., PARVIN, S., HAAS, O., WOESSMANN, W., STANULLA, M., SCHRAPPE, M., DUGAS, M., NATKUNAM, Y., ORFAO, A., DOMINGUEZ, V., PINTADO, B., BLANCO, O., ALONSO-LOPEZ, D., DE LAS RIVAS, J., MARTIN-LORENZO, A., JIMENEZ, R., GARCIA CRIADO, F. J., GARCIA CENADOR, M. B., LOSSOS, I. S., VICENTE-DUENAS, C., BORKHARDT, A., HAUER, J. & SANCHEZ-GARCIA, I. 2018. Lmo2 expression defines tumor cell identity during T-cell leukemogenesis. *EMBO J*, 37.
- GARCIA, I. S., KANEKO, Y., GONZALEZ-SARMIENTO, R., CAMPBELL, K., WHITE, L., BOEHM, T. & RABBITTS, T. H. 1991. A study of chromosome 11p13 translocations involving TCR beta and TCR delta in human T cell leukaemia. *Oncogene*, 6, 577-82.
- GAULARD, P., BOURQUELOT, P., KANAVAROS, P., HAIOUN, C., LE COUEDIC, J. P., DIVINE, M., GOOSSENS, M., ZAFRANI, E. S., FARCET, J. P. & REYES, F. 1990. Expression of the alpha/beta and gamma/delta T-cell receptors in 57 cases of peripheral T-cell lymphomas. Identification of a subset of gamma/delta T-cell lymphomas. *Am J Pathol*, 137, 617-28.
- GAVVA, N. R., GAVVA, R., ERMEKOVA, K., SUDOL, M. & SHEN, C. J. 1997. Interaction of WW domains with hematopoietic transcription factor p45/NF-E2 and RNA polymerase II. *J Biol Chem*, 272, 24105-8.
- GEIMER LE LAY, A. S., ORAVECZ, A., MASTIO, J., JUNG, C., MARCHAL, P., EBEL, C., DEMBELE, D., JOST, B., LE GRAS, S., THIBAUT, C., BORGGREFE, T., KASTNER, P. & CHAN, S. 2014. The tumor suppressor Ikaros shapes the repertoire of notch target genes in T cells. *Sci Signal*, 7, ra28.
- GERBY, B., TREMBLAY, C. S., TREMBLAY, M., ROJAS-SUTTERLIN, S., HERBLOT, S., HEBERT, J., SAUVAGEAU, G., LEMIEUX, S., LECUYER, E., VEIGA, D. F. & HOANG, T. 2014. SCL, LMO1 and Notch1 reprogram thymocytes into self-renewing cells. *PLoS Genet*, 10, e1004768.
- GERMAIN, R. N. 2002. T-cell development and the CD4-CD8 lineage decision. *Nature Reviews Immunology*, 2, 309-322.
- GERMAR, K., DOSE, M., KONSTANTINOU, T., ZHANG, J., WANG, H., LOBRY, C., ARNETT, K. L., BLACKLOW, S. C., AIFANTIS, I., ASTER, J. C. & GOUNARI, F. 2011. T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. *Proc Natl Acad Sci U S A*, 108, 20060-5.
- GIARDINE, B., RIEMER, C., HARDISON, R. C., BURHANS, R., ELNITSKI, L., SHAH, P., ZHANG, Y., BLANKENBERG, D., ALBERT, I., TAYLOR, J., MILLER, W., KENT, W. J. & NEKRUTENKO, A. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res*, 15, 1451-5.
- GILMOUR, J., ASSI, S. A., NOAILLES, L., LICHTINGER, M., OBIER, N. & BONIFER, C. 2018. The Co-operation of RUNX1 with LDB1, CDK9 and BRD4 Drives Transcription Factor Complex Relocation During Haematopoietic Specification. *Sci Rep*, 8, 10410.
- GIMFERRER, I., HU, T., SIMMONS, A., WANG, C., SOUABNI, A., BUSSLINGER, M., BENDER, T. P., HERNANDEZ-HOYOS, G. & ALBEROLA-ILA, J. 2011.

- Regulation of GATA-3 expression during CD4 lineage differentiation. *J Immunol*, 186, 3892-8.
- GOECKS, J., NEKRUTENKO, A., TAYLOR, J. & GALAXY, T. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*, 11, R86.
- GOLUB, R. & CUMANO, A. 2013. Embryonic hematopoiesis. *Blood Cells Mol Dis*, 51, 226-31.
- GORDON, S., AKOPYAN, G., GARBAN, H. & BONAVIDA, B. 2006. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene*, 25, 1125-42.
- GOTTGENS, B. 2015. Regulatory network control of blood stem cells. *Blood*, 125, 2614-20.
- GOUNARI, F., AIFANTIS, I., MARTIN, C., FEHLING, H. J., HOEFLINGER, S., LEDER, P., VON BOEHMER, H. & REIZIS, B. 2002. Tracing lymphopoiesis with the aid of a pTalpha-controlled reporter gene. *Nat Immunol*, 3, 489-96.
- GREENBAUM, S. & ZHUANG, Y. 2002. Identification of E2A target genes in B lymphocyte development by using a gene tagging-based chromatin immunoprecipitation system. *Proc Natl Acad Sci U S A*, 99, 15030-5.
- GROSSMANN, V., HAFERLACH, C., WEISSMANN, S., ROLLER, A., SCHINDELA, S., POETZINGER, F., STADLER, K., BELLOS, F., KERN, W., HAFERLACH, T., SCHNITTGER, S. & KOHLMANN, A. 2013. The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL. *Genes Chromosomes Cancer*, 52, 410-22.
- GROWNEY, J. D., SHIGEMATSU, H., LI, Z., LEE, B. H., ADELSPERGER, J., ROWAN, R., CURLEY, D. P., KUTOK, J. L., AKASHI, K., WILLIAMS, I. R., SPECK, N. A. & GILLILAND, D. G. 2005. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*, 106, 494-504.
- GRUTZ, G. G., BUCHER, K., LAVENIR, I., LARSON, T., LARSON, R. & RABBITTS, T. H. 1998. The oncogenic T cell LIM-protein Lmo2 forms part of a DNA-binding complex specifically in immature T cells. *EMBO J*, 17, 4594-605.
- GWIN, K. A., SHAPIRO, M. B., DOLENCE, J. J., HUANG, Z. L. & MEDINA, K. L. 2013. Hoxa9 and Flt3 signaling synergistically regulate an early checkpoint in lymphopoiesis. *J Immunol*, 191, 745-54.
- HAAS, S., TRUMPP, A. & MILSOM, M. D. 2018. Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell*, 22, 627-638.
- HACEIN-BEY-ABINA, S., GARRIGUE, A., WANG, G. P., SOULIER, J., LIM, A., MORILLON, E., CLAPPIER, E., CACCAVELLI, L., DELABESSE, E., BELDJORD, K., ASNAFI, V., MACINTYRE, E., DAL CORTIVO, L., RADFORD, I., BROUSSE, N., SIGAUX, F., MOSHOUS, D., HAUER, J., BORKHARDT, A., BELOHRADSKY, B. H., WINTERGERST, U., VELEZ, M. C., LEIVA, L., SORENSEN, R., WULFFRAAT, N., BLANCHE, S., BUSHMAN, F. D., FISCHER, A. & CAVAZZANA-CALVO, M. 2008. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest*, 118, 3132-42.



- HACEIN-BEY-ABINA, S., VON KALLE, C., SCHMIDT, M., MCCORMACK, M. P., WULFFRAAT, N., LEBOULCH, P., LIM, A., OSBORNE, C. S., PAWLIUK, R., MORILLON, E., SORENSEN, R., FORSTER, A., FRASER, P., COHEN, J. I., DE SAINT BASILE, G., ALEXANDER, I., WINTERGERST, U., FREBOURG, T., AURIAS, A., STOPPA-LYONNET, D., ROMANA, S., RADFORD-WEISS, I., GROSS, F., VALENSI, F., DELABESSE, E., MACINTYRE, E., SIGAUX, F., SOULIER, J., LEIVA, L. E., WISSLER, M., PRINZ, C., RABBITTS, T. H., LE DEIST, F., FISCHER, A. & CAVAZZANA-CALVO, M. 2003. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*, 302, 415-9.
- HAYDAY, A. C. & PENNINGTON, D. J. 2007. Key factors in the organized chaos of early T cell development. *Nat Immunol*, 8, 137-44.
- HEINTZMAN, N. D., HON, G. C., HAWKINS, R. D., KHERADPOUR, P., STARK, A., HARP, L. F., YE, Z., LEE, L. K., STUART, R. K., CHING, C. W., CHING, K. A., ANTOSIEWICZ-BOURGET, J. E., LIU, H., ZHANG, X., GREEN, R. D., LOBANENKOV, V. V., STEWART, R., THOMSON, J. A., CRAWFORD, G. E., KELLIS, M. & REN, B. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459, 108-12.
- HENG, T. S., PAINTER, M. W. & IMMUNOLOGICAL GENOME PROJECT, C. 2008. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol*, 9, 1091-4.
- HERBLOT, S., STEFF, A. M., HUGO, P., APLAN, P. D. & HOANG, T. 2000. SCL and LMO1 alter thymocyte differentiation: inhibition of E2A-HEB function and pre-T alpha chain expression. *Nat Immunol*, 1, 138-44.
- HERNANDEZ-HOYOS, G., ANDERSON, M. K., WANG, C., ROTHENBERG, E. V. & ALBEROLA-ILA, J. 2003. GATA-3 expression is controlled by TCR signals and regulates CD4/CD8 differentiation. *Immunity*, 19, 83-94.
- HERZENBERG, L. A., SWEET, R. G. & HERZENBERG, L. A. 1976. Fluorescence-activated cell sorting. *Sci Am*, 234, 108-17.
- HOANG, T., LAMBERT, J. A. & MARTIN, R. 2016. SCL/TAL1 in Hematopoiesis and Cellular Reprogramming. *Curr Top Dev Biol*, 118, 163-204.
- HOFFMAN, E. S., PASSONI, L., CROMPTON, T., LEU, T. M., SCHATZ, D. G., KOFF, A., OWEN, M. J. & HAYDAY, A. C. 1996. Productive T-cell receptor beta-chain gene rearrangement: coincident regulation of cell cycle and clonality during development in vivo. *Genes Dev*, 10, 948-62.
- HOMMINGA, I., PIETERS, R., LANGERAK, A. W., DE ROOI, J. J., STUBBS, A., VERSTEGEN, M., VUERHARD, M., BUIJS-GLADDINES, J., KOOL, C., KLOUS, P., VAN VLIERBERGHE, P., FERRANDO, A. A., CAYUELA, J. M., VERHAAF, B., BEVERLOO, H. B., HORSTMANN, M., DE HAAS, V., WIEKMEIJER, A. S., PIKE-OVERZET, K., STAAL, F. J., DE LAAT, W., SOULIER, J., SIGAUX, F. & MEIJERINK, J. P. 2011. Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell*, 19, 484-97.
- HOSOYA, T., KUROHA, T., MORIGUCHI, T., CUMMINGS, D., MAILLARD, I., LIM, K. C. & ENGEL, J. D. 2009. GATA-3 is required for early T lineage progenitor development. *J Exp Med*, 206, 2987-3000.
- HOWE, S. J., MANSOUR, M. R., SCHWARZWAELDER, K., BARTHOLOMAE, C., HUBANK, M., KEMPSKI, H., BRUGMAN, M. H., PIKE-OVERZET, K.,

- CHATTERS, S. J., DE RIDDER, D., GILMOUR, K. C., ADAMS, S., THORNHILL, S. I., PARSLEY, K. L., STAAL, F. J., GALE, R. E., LINCH, D. C., BAYFORD, J., BROWN, L., QUAYE, M., KINNON, C., ANCLIFF, P., WEBB, D. K., SCHMIDT, M., VON KALLE, C., GASPAS, H. B. & THRASHER, A. J. 2008. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest*, 118, 3143-50.
- HSU, H. L., WADMAN, I., TSAN, J. T. & BAER, R. 1994. Positive and negative transcriptional control by the TAL1 helix-loop-helix protein. *Proc Natl Acad Sci U S A*, 91, 5947-51.
- HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4, 44-57.
- HUANG, Y., SITWALA, K., BRONSTEIN, J., SANDERS, D., DANDEKAR, M., COLLINS, C., ROBERTSON, G., MACDONALD, J., CEZARD, T., BILENKY, M., THIESSEN, N., ZHAO, Y., ZENG, T., HIRST, M., HERO, A., JONES, S. & HESS, J. L. 2012. Identification and characterization of Hoxa9 binding sites in hematopoietic cells. *Blood*, 119, 388-98.
- HUH, H. J., LEE, S. H., YOO, K. H., SUNG, K. W., KOO, H. H., JANG, J. H., KIM, K., KIM, S. J., KIM, W. S., JUNG, C. W., LEE, K. O., KIM, S. H. & KIM, H. J. 2013. Gene mutation profiles and prognostic implications in Korean patients with T-lymphoblastic leukemia. *Ann Hematol*, 92, 635-44.
- HUNGER, S. P. & MULLIGHAN, C. G. 2015. Acute Lymphoblastic Leukemia in Children. *N Engl J Med*, 373, 1541-52.
- IGARASHI, H., GREGORY, S. C., YOKOTA, T., SAKAGUCHI, N. & KINCADE, P. W. 2002. Transcription from the RAG1 locus marks the earliest lymphocyte progenitors in bone marrow. *Immunity*, 17, 117-30.
- IKAWA, T., HIROSE, S., MASUDA, K., KAKUGAWA, K., SATOH, R., SHIBANO-SATOH, A., KOMINAMI, R., KATSURA, Y. & KAWAMOTO, H. 2010. An essential developmental checkpoint for production of the T cell lineage. *Science*, 329, 93-6.
- IKAWA, T., KAWAMOTO, H., GOLDRATH, A. W. & MURRE, C. 2006. E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. *J Exp Med*, 203, 1329-42.
- IMPERATO, M. R., CAUCHY, P., OBIER, N. & BONIFER, C. 2015. The RUNX1-PU.1 axis in the control of hematopoiesis. *Int J Hematol*, 101, 319-29.
- INABA, H., GREAVES, M. & MULLIGHAN, C. G. 2013. Acute lymphoblastic leukaemia. *Lancet*, 381, 1943-55.
- ISOGAI, E., OHIRA, M., OZAKI, T., OBA, S., NAKAMURA, Y. & NAKAGAWARA, A. 2011. Oncogenic LMO3 collaborates with HEN2 to enhance neuroblastoma cell growth through transactivation of Mash1. *PLoS One*, 6, e19297.
- JAGANNATHAN-BOGDAN, M. & ZON, L. I. 2013. Hematopoiesis. *Development*, 140, 2463-7.
- JANG, W., PARK, J., KWON, A., CHOI, H., KIM, J., LEE, G. D., HAN, E., JEKARL, D. W., CHAE, H., HAN, K., YOON, J. H., LEE, S., CHUNG, N. G., CHO, B., KIM, M. & KIM, Y. 2019. CDKN2B downregulation and other genetic characteristics in T-acute lymphoblastic leukemia. *Exp Mol Med*, 51, 4.

- JANSSEN, A., VAN DER BURG, M., SZUHAI, K., KOPS, G. J. & MEDEMA, R. H. 2011. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science*, 333, 1895-8.
- JANSSEN, J. W., LUDWIG, W. D., STERRY, W. & BARTRAM, C. R. 1993. SIL-TAL1 deletion in T-cell acute lymphoblastic leukemia. *Leukemia*, 7, 1204-10.
- JOHNSON, R. T. & RAO, P. N. 1970. Mammalian cell fusion: induction of premature chromosome condensation in interphase nuclei. *Nature*, 226, 717-22.
- JOJIC, V., SHAY, T., SYLVIA, K., ZUK, O., SUN, X., KANG, J., REGEV, A., KOLLER, D., IMMUNOLOGICAL GENOME PROJECT, C., BEST, A. J., KNELL, J., GOLDRATH, A., JOIC, V., KOLLER, D., SHAY, T., REGEV, A., COHEN, N., BRENNAN, P., BRENNER, M., KIM, F., RAO, T. N., WAGERS, A., HENG, T., ERICSON, J., ROTHAMEL, K., ORTIZ-LOPEZ, A., MATHIS, D., BENOIST, C., BEZMAN, N. A., SUN, J. C., MIN-OO, G., KIM, C. C., LANIER, L. L., MILLER, J., BROWN, B., MERAD, M., GAUTIER, E. L., JAKUBZICK, C., RANDOLPH, G. J., MONACH, P., BLAIR, D. A., DUSTIN, M. L., SHINTON, S. A., HARDY, R. R., LAIDLAW, D., COLLINS, J., GAZIT, R., ROSSI, D. J., MALHOTRA, N., SYLVIA, K., KANG, J., KRESLAVSKY, T., FLETCHER, A., ELPEK, K., BELLEMARTE-PELLETIER, A., MALHOTRA, D. & TURLEY, S. 2013. Identification of transcriptional regulators in the mouse immune system. *Nat Immunol*, 14, 633-43.
- JONES, D. L. & WAGERS, A. J. 2008. No place like home: anatomy and function of the stem cell niche. *Nat Rev Mol Cell Biol*, 9, 11-21.
- JONES, S. 2004. An overview of the basic helix-loop-helix proteins. *Genome Biol*, 5, 226.
- JOSHI, I., MINTER, L. M., TELFER, J., DEMAREST, R. M., CAPOBIANCO, A. J., ASTER, J. C., SICINSKI, P., FAUQ, A., GOLDE, T. E. & OSBORNE, B. A. 2009. Notch signaling mediates G1/S cell-cycle progression in T cells via cyclin D3 and its dependent kinases. *Blood*, 113, 1689-98.
- KADRMAS, J. L. & BECKERLE, M. C. 2004. The LIM domain: from the cytoskeleton to the nucleus. *Nat Rev Mol Cell Biol*, 5, 920-31.
- KANG, J. & DER, S. D. 2004. Cytokine functions in the formative stages of a lymphocyte's life. *Curr Opin Immunol*, 16, 180-90.
- KENNY, D. A., JURATA, L. W., SAGA, Y. & GILL, G. N. 1998. Identification and characterization of LMO4, an LMO gene with a novel pattern of expression during embryogenesis. *Proc Natl Acad Sci U S A*, 95, 11257-62.
- KIEL, M. J., YILMAZ, O. H., IWASHITA, T., YILMAZ, O. H., TERHORST, C. & MORRISON, S. J. 2005. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell*, 121, 1109-21.
- KOCH, U., FIORINI, E., BENEDITO, R., BESSEYRIAS, V., SCHUSTER-GOSSLER, K., PIERRES, M., MANLEY, N. R., DUARTE, A., MACDONALD, H. R. & RADTKE, F. 2008. Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. *J Exp Med*, 205, 2515-23.
- KONDO, M., WEISSMAN, I. L. & AKASHI, K. 1997. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*, 91, 661-72.
- KRANGEL, M. S. 2009. Mechanics of T cell receptor gene rearrangement. *Curr Opin Immunol*, 21, 133-9.

- KRESLAVSKY, T., GLEIMER, M., MIYAZAKI, M., CHOI, Y., GAGNON, E., MURRE, C., SICINSKI, P. & VON BOEHMER, H. 2012. beta-Selection-induced proliferation is required for alphabeta T cell differentiation. *Immunity*, 37, 840-53.
- KRUDE, T. & KELLER, C. 2001. Chromatin assembly during S phase: contributions from histone deposition, DNA replication and the cell division cycle. *Cell Mol Life Sci*, 58, 665-72.
- KUSY, S., GERBY, B., GOARDON, N., GAULT, N., FERRI, F., GERARD, D., ARMSTRONG, F., BALLERINI, P., CAYUELA, J. M., BARUCHEL, A., PFLUMIO, F. & ROMEO, P. H. 2010. NKX3.1 is a direct TAL1 target gene that mediates proliferation of TAL1-expressing human T cell acute lymphoblastic leukemia. *J Exp Med*, 207, 2141-56.
- KVINLAUG, B. T., CHAN, W. I., BULLINGER, L., RAMASWAMI, M., SEARS, C., FOSTER, D., LAZIC, S. E., OKABE, R., BENNER, A., LEE, B. H., DE SILVA, I., VALK, P. J., DELWEL, R., ARMSTRONG, S. A., DOHNER, H., GILLILAND, D. G. & HUNTLY, B. J. 2011. Common and overlapping oncogenic pathways contribute to the evolution of acute myeloid leukemias. *Cancer Res*, 71, 4117-29.
- LAHLIL, R., LECUYER, E., HERBLOT, S. & HOANG, T. 2004. SCL assembles a multifactorial complex that determines glycophorin A expression. *Mol Cell Biol*, 24, 1439-52.
- LAI, A. Y. & WADE, P. A. 2011. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer*, 11, 588-96.
- LARSON, R. C., LAVENIR, I., LARSON, T. A., BAER, R., WARREN, A. J., WADMAN, I., NOTTAGE, K. & RABBITS, T. H. 1996. Protein dimerization between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. *EMBO J*, 15, 1021-7.
- LARSON, R. C., OSADA, H., LARSON, T. A., LAVENIR, I. & RABBITS, T. H. 1995. The oncogenic LIM protein Rbtn2 causes thymic developmental aberrations that precede malignancy in transgenic mice. *Oncogene*, 11, 853-62.
- LEARNED, R. M., LEARNED, T. K., HALTINER, M. M. & TJIAN, R. T. 1986. Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. *Cell*, 45, 847-57.
- LECUYER, E. & HOANG, T. 2004. SCL: from the origin of hematopoiesis to stem cells and leukemia. *Exp Hematol*, 32, 11-24.
- LECUYER, E., LARIVIERE, S., SINCENNES, M. C., HAMAN, A., LAHLIL, R., TODOROVA, M., TREMBLAY, M., WILKES, B. C. & HOANG, T. 2007. Protein stability and transcription factor complex assembly determined by the SCL-LMO2 interaction. *J Biol Chem*, 282, 33649-58.
- LENGAUER, C., KINZLER, K. W. & VOGELSTEIN, B. 1998. Genetic instabilities in human cancers. *Nature*, 396, 643-9.
- LEONG, W. Z., TAN, S. H., NGOC, P. C. T., AMANDA, S., YAM, A. W. Y., LIAU, W. S., GONG, Z., LAWTON, L. N., TENEN, D. G. & SANDA, T. 2017. ARID5B as a critical downstream target of the TAL1 complex that activates the oncogenic transcriptional program and promotes T-cell leukemogenesis. *Genes Dev*, 31, 2343-2360.

- LERDRUP, M., JOHANSEN, J. V., AGRAWAL-SINGH, S. & HANSEN, K. 2016. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat Struct Mol Biol*, 23, 349-57.
- LI, H., WATFORD, W., LI, C., PARMELEE, A., BRYANT, M. A., DENG, C., O'SHEA, J. & LEE, S. B. 2007. Ewing sarcoma gene EWS is essential for meiosis and B lymphocyte development. *J Clin Invest*, 117, 1314-23.
- LI, L., LEE, J. Y., GROSS, J., SONG, S. H., DEAN, A. & LOVE, P. E. 2010a. A requirement for Lim domain binding protein 1 in erythropoiesis. *J Exp Med*, 207, 2543-50.
- LI, L., LEID, M. & ROTHENBERG, E. V. 2010b. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science*, 329, 89-93.
- LI, L., ZHANG, J. A., DOSE, M., KUEH, H. Y., MOSADEGHI, R., GOUNARI, F. & ROTHENBERG, E. V. 2013a. A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. *Blood*, 122, 902-11.
- LI, M., XIAO, L., XU, J., ZHANG, R., GUO, J., OLSON, J., WU, Y., LI, J., SONG, C. & GE, Z. 2016. Co-existence of PHF6 and NOTCH1 mutations in adult T-cell acute lymphoblastic leukemia. *Oncol Lett*, 12, 16-22.
- LI, X., YAO, H., CHEN, Z., WANG, Q., ZHAO, Y. & CHEN, S. 2013b. Somatic mutations of PHF6 in patients with chronic myeloid leukemia in blast crisis. *Leuk Lymphoma*, 54, 671-2.
- LINO, C. A., HARPER, J. C., CARNEY, J. P. & TIMLIN, J. A. 2018. Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv*, 25, 1234-1257.
- LITZOW, M. R. & FERRANDO, A. A. 2015. How I treat T-cell acute lymphoblastic leukemia in adults. *Blood*, 126, 833-41.
- LIU, H., CHIANG, M. Y. & PEAR, W. S. 2011. Critical roles of NOTCH1 in acute T-cell lymphoblastic leukemia. *Int J Hematol*, 94, 118-125.
- LIU, L., QIN, S., ZHANG, J., JI, P., SHI, Y. & WU, J. 2012. Solution structure of an atypical PHD finger in BRPF2 and its interaction with DNA. *J Struct Biol*, 180, 165-73.
- LIU, Z., LI, F., RUAN, K., ZHANG, J., MEI, Y., WU, J. & SHI, Y. 2014. Structural and functional insights into the human Borjeson-Forssman-Lehmann syndrome-associated protein PHF6. *J Biol Chem*, 289, 10069-83.
- LIU, Z., LI, F., ZHANG, B., LI, S., WU, J. & SHI, Y. 2015. Structural basis of plant homeodomain finger 6 (PHF6) recognition by the retinoblastoma binding protein 4 (RBBP4) component of the nucleosome remodeling and deacetylase (NuRD) complex. *J Biol Chem*, 290, 6630-8.
- LOVE, P. E., WARZECHA, C. & LI, L. 2014. Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends Genet*, 30, 1-9.
- LOWER, K. M., TURNER, G., KERR, B. A., MATHEWS, K. D., SHAW, M. A., GEDEON, A. K., SCHELLEY, S., HOYME, H. E., WHITE, S. M., DELATYCKI, M. B., LAMPE, A. K., CLAYTON-SMITH, J., STEWART, H., VAN RAVENSWAAY, C. M., DE VRIES, B. B., COX, B., GROMPE, M., ROSS, S., THOMAS, P., MULLEY, J. C. & GECZ, J. 2002. Mutations in PHF6 are associated with Borjeson-Forssman-Lehmann syndrome. *Nat Genet*, 32, 661-5.
- LYU, C., WANG, L. & ZHANG, J. 2018. Deep learning for DNase I hypersensitive sites identification. *BMC Genomics*, 19, 905.

- MA, P. C., ROULD, M. A., WEINTRAUB, H. & PABO, C. O. 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell*, 77, 451-9.
- MANSOUR, M. R., DUKE, V., FORONI, L., PATEL, B., ALLEN, C. G., ANCLIFF, P. J., GALE, R. E. & LINCH, D. C. 2007. Notch-1 mutations are secondary events in some patients with T-cell acute lymphoblastic leukemia. *Clin Cancer Res*, 13, 6964-9.
- MARGUERON, R. & REINBERG, D. 2010. Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet*, 11, 285-96.
- MARTON, H. A. & DESIDERIO, S. 2008. The Paf1 complex promotes displacement of histones upon rapid induction of transcription by RNA polymerase II. *BMC Mol Biol*, 9, 4.
- MASSA, S., BALCIUNAITE, G., CEREDIG, R. & ROLINK, A. G. 2006. Critical role for c-kit (CD117) in T cell lineage commitment and early thymocyte development in vitro. *Eur J Immunol*, 36, 526-32.
- MASUDA, K., KAKUGAWA, K., NAKAYAMA, T., MINATO, N., KATSURA, Y. & KAWAMOTO, H. 2007. T cell lineage determination precedes the initiation of TCR beta gene rearrangement. *J Immunol*, 179, 3699-706.
- MATTHEWS, J. M., LESTER, K., JOSEPH, S. & CURTIS, D. J. 2013. LIM-domain-only proteins in cancer. *Nat Rev Cancer*, 13, 111-22.
- MATTHEWS, J. M. & VISVADER, J. E. 2003. LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO Rep*, 4, 1132-7.
- MAURICE, D., HOOPER, J., LANG, G. & WESTON, K. 2007. c-Myb regulates lineage choice in developing thymocytes via its target gene Gata3. *EMBO J*, 26, 3629-40.
- MCCORMACK, M. P., FORSTER, A., DRYNAN, L., PANNELL, R. & RABBITS, T. H. 2003. The LMO2 T-cell oncogene is activated via chromosomal translocations or retroviral insertion during gene therapy but has no mandatory role in normal T-cell development. *Mol Cell Biol*, 23, 9003-13.
- MCCORMACK, M. P., SHIELDS, B. J., JACKSON, J. T., NASA, C., SHI, W., SLATER, N. J., TREMBLAY, C. S., RABBITS, T. H. & CURTIS, D. J. 2013. Requirement for Lyl1 in a model of Lmo2-driven early T-cell precursor ALL. *Blood*, 122, 2093-103.
- MCCORMACK, M. P., YOUNG, L. F., VASUDEVAN, S., DE GRAAF, C. A., CODRINGTON, R., RABBITS, T. H., JANE, S. M. & CURTIS, D. J. 2010. The Lmo2 oncogene initiates leukemia in mice by inducing thymocyte self-renewal. *Science*, 327, 879-83.
- MCGUIRE, E. A., HOCKETT, R. D., POLLOCK, K. M., BARTHOLDI, M. F., O'BRIEN, S. J. & KORSMEYER, S. J. 1989. The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including Ttg-1, a gene encoding a potential zinc finger protein. *Mol Cell Biol*, 9, 2124-32.
- MCLEAN, C. Y., BRISTOR, D., HILLER, M., CLARKE, S. L., SCHAAR, B. T., LOWE, C. B., WENGER, A. M. & BEJERANO, G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*, 28, 495-501.
- MCRAE, H. M., GARNHAM, A. L., HU, Y., WITKOWSKI, M. T., CORBETT, M. A., DIXON, M. P., MAY, R. E., SHEIKH, B. N., CHIANG, W., KUEH, A. J., NGUYEN, T. A., MAN, K., GLOURY, R., AUBREY, B. J., POLICHENI, A., DI

- RAGO, L., ALEXANDER, W. S., GRAY, D. H. D., STRASSER, A., HAWKINS, E. D., WILCOX, S., GECZ, J., KALLIES, A., MCCORMACK, M. P., SMYTH, G. K., VOSS, A. K. & THOMAS, T. 2019. PHF6 regulates hematopoietic stem and progenitor cells and its loss synergizes with expression of TLX3 to cause leukemia. *Blood*, 133, 1729-1741.
- MEACHAM, C. E., LAWTON, L. N., SOTO-FELICIANO, Y. M., PRITCHARD, J. R., JOUGHIN, B. A., EHRENBERGER, T., FENOUILLE, N., ZUBER, J., WILLIAMS, R. T., YOUNG, R. A. & HEMANN, M. T. 2015. A genome-scale in vivo loss-of-function screen identifies Phf6 as a lineage-specific regulator of leukemia cell growth. *Genes Dev*, 29, 483-8.
- MEDVINSKY, A. & DZIERZAK, E. 1996. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*, 86, 897-906.
- MELLACHERUVU, D., WRIGHT, Z., COUZENS, A. L., LAMBERT, J. P., ST-DENIS, N. A., LI, T., MITEVA, Y. V., HAURI, S., SARDIU, M. E., LOW, T. Y., HALIM, V. A., BAGSHAW, R. D., HUBNER, N. C., AL-HAKIM, A., BOUCHARD, A., FAUBERT, D., FERMIN, D., DUNHAM, W. H., GOUDREAU, M., LIN, Z. Y., BADILLO, B. G., PAWSON, T., DUROCHER, D., COULOMBE, B., AEBERSOLD, R., SUPERTI-FURGA, G., COLINGE, J., HECK, A. J., CHOI, H., GSTAIGER, M., MOHAMMED, S., CRISTEA, I. M., BENNETT, K. L., WASHBURN, M. P., RAUGHT, B., EWING, R. M., GINGRAS, A. C. & NESVIZHSKII, A. I. 2013. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods*, 10, 730-6.
- MENG, Y. S., WEI, R., AI, G. W., MENG, X. Q. & ZHANG, Y. X. 2009. [Abnormal expression of transcription factors LYL1 and LMO2 and interaction between them in myeloid leukemia]. *Zhonghua Yi Xue Za Zhi*, 89, 890-3.
- MIKKOLA, H. K., KLINTMAN, J., YANG, H., HOCK, H., SCHLAEGER, T. M., FUJIWARA, Y. & ORKIN, S. H. 2003. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature*, 421, 547-51.
- MINGUENEAU, M., KRESLAVSKY, T., GRAY, D., HENG, T., CRUSE, R., ERICSON, J., BENDALL, S., SPITZER, M. H., NOLAN, G. P., KOBAYASHI, K., VON BOEHMER, H., MATHIS, D., BENOIST, C., IMMUNOLOGICAL GENOME, C., BEST, A. J., KNELL, J., GOLDRATH, A., JOIC, V., KOLLER, D., SHAY, T., REGEV, A., COHEN, N., BRENNAN, P., BRENNER, M., KIM, F., NAGESWARA RAO, T., WAGERS, A., HENG, T., ERICSON, J., ROTHAMEL, K., ORTIZ-LOPEZ, A., MATHIS, D., BENOIST, C., BEZMAN, N. A., SUN, J. C., MIN-OO, G., KIM, C. C., LANIER, L. L., MILLER, J., BROWN, B., MERAD, M., GAUTIER, E. L., JAKUBZICK, C., RANDOLPH, G. J., MONACH, P., BLAIR, D. A., DUSTIN, M. L., SHINTON, S. A., HARDY, R. R., LAIDLAW, D., COLLINS, J., GAZIT, R., ROSSI, D. J., MALHOTRA, N., SYLVIA, K., KANG, J., KRESLAVSKY, T., FLETCHER, A., ELPEK, K., BELLEMARE-PELLETIER, A., MALHOTRA, D. & TURLEY, S. 2013. The transcriptional landscape of alphabeta T cell differentiation. *Nat Immunol*, 14, 619-32.
- MIYAGI, S., SROCZYNSKA, P., KATO, Y., NAKAJIMA-TAKAGI, Y., OSHIMA, M., RIZQ, O., TAKAYAMA, N., SARAYA, A., MIZUNO, S., SUGIYAMA, F., TAKAHASHI, S., MATSUZAKI, Y., CHRISTENSEN, J., HELIN, K. & IWAMA,

- A. 2019. The chromatin-binding protein Phf6 restricts the self-renewal of hematopoietic stem cells. *Blood*, 133, 2495-2506.
- MONTANEZ-WISCOVICH, M. E., SHELTON, M. D., SEACHRIST, D. D., LOZADA, K. L., JOHNSON, E., MIEDLER, J. D., ABDUL-KARIM, F. W., VISVADER, J. E. & KERI, R. A. 2010. Aberrant expression of LMO4 induces centrosome amplification and mitotic spindle abnormalities in breast cancer cells. *J Pathol*, 222, 271-81.
- MORI, T., NAGATA, Y., MAKISHIMA, H., SANADA, M., SHIOZAWA, Y., KON, A., YOSHIZATO, T., SATO-OTSUBO, A., KATAOKA, K., SHIRAISHI, Y., CHIBA, K., TANAKA, H., ISHIYAMA, K., MIYAWAKI, S., MORI, H., NAKAMAKI, T., KIHARA, R., KIYOI, H., KOEFFLER, H. P., SHIH, L. Y., MIYANO, S., NAOE, T., HAFERLACH, C., KERN, W., HAFERLACH, T., OGAWA, S. & YOSHIDA, K. 2016. Somatic PHF6 mutations in 1760 cases with various myeloid neoplasms. *Leukemia*, 30, 2270-2273.
- MORITA, S., KOJIMA, T. & KITAMURA, T. 2000. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther*, 7, 1063-6.
- MORRISON, S. J., UCHIDA, N. & WEISSMAN, I. L. 1995. The biology of hematopoietic stem cells. *Annu Rev Cell Dev Biol*, 11, 35-71.
- MORRISON, S. J., WANDYCH, A. M., HEMMATI, H. D., WRIGHT, D. E. & WEISSMAN, I. L. 1997. Identification of a lineage of multipotent hematopoietic progenitors. *Development*, 124, 1929-39.
- MUKHOPADHYAY, M., TEUFEL, A., YAMASHITA, T., AGULNICK, A. D., CHEN, L., DOWNS, K. M., SCHINDLER, A., GRINBERG, A., HUANG, S. P., DORWARD, D. & WESTPHAL, H. 2003. Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development*, 130, 495-505.
- MULLER-SIEBURG, C. E., WHITLOCK, C. A. & WEISSMAN, I. L. 1986. Isolation of two early B lymphocyte progenitors from mouse marrow: a committed pre-pre-B cell and a clonogenic Thy-1-lo hematopoietic stem cell. *Cell*, 44, 653-62.
- MULLER, A. M., MEDVINSKY, A., STROUBOULIS, J., GROSVELD, F. & DZIERZAK, E. 1994. Development of hematopoietic stem cell activity in the mouse embryo. *Immunity*, 1, 291-301.
- MURRE, C., BAIN, G., VAN DIJK, M. A., ENGEL, I., FURNARI, B. A., MASSARI, M. E., MATTHEWS, J. R., QUONG, M. W., RIVERA, R. R. & STUIVER, M. H. 1994. Structure and function of helix-loop-helix proteins. *Biochim Biophys Acta*, 1218, 129-35.
- MURRE, C., MCCAWE, P. S. & BALTIMORE, D. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell*, 56, 777-83.
- NAIK, S. H., PERIE, L., SWART, E., GERLACH, C., VAN ROOIJ, N., DE BOER, R. J. & SCHUMACHER, T. N. 2013. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature*, 496, 229-32.
- NAITO, T., TANAKA, H., NAOE, Y. & TANIUCHI, I. 2011. Transcriptional control of T-cell development. *Int Immunol*, 23, 661-8.
- NAM, C. H. & RABBITTS, T. H. 2006. The role of LMO2 in development and in T cell leukemia after chromosomal translocation or retroviral insertion. *Mol Ther*, 13, 15-25.



- NERLOV, C. & GRAF, T. 1998. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev*, 12, 2403-12.
- NOTTINGHAM, W. T., JARRATT, A., BURGESS, M., SPECK, C. L., CHENG, J. F., PRABHAKAR, S., RUBIN, E. M., LI, P. S., SLOANE-STANLEY, J., KONG, A. S. J. & DE BRUIJN, M. F. 2007. Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood*, 110, 4188-97.
- O'NEIL, J., BILLA, M., OIKEMUS, S. & KELLIHER, M. 2001. The DNA binding activity of TAL-1 is not required to induce leukemia/lymphoma in mice. *Oncogene*, 20, 3897-905.
- O'NEIL, J., SHANK, J., CUSSON, N., MURRE, C. & KELLIHER, M. 2004. TAL1/SCL induces leukemia by inhibiting the transcriptional activity of E47/HEB. *Cancer Cell*, 5, 587-96.
- OBE, G., BEEK, B. & VAIDYA, V. G. 1975. The human leukocyte test system. III. Premature chromosome condensation from chemically and x-ray induced micronuclei. *Mutat Res*, 27, 89-101.
- OOSTERWEGEL, M., VAN DE WETERING, M., DOOIJES, D., KLOMP, L., WINOTO, A., GEORGOPOULOS, K., MEIJLINK, F. & CLEVERS, H. 1991. Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers. *J Exp Med*, 173, 1133-42.
- ORAM, S. H., THOMS, J. A., PRIDANS, C., JANES, M. E., KINSTON, S. J., ANAND, S., LANDRY, J. R., LOCK, R. B., JAYARAMAN, P. S., HUNTLY, B. J., PIMANDA, J. E. & GOTTGENS, B. 2010. A previously unrecognized promoter of LMO2 forms part of a transcriptional regulatory circuit mediating LMO2 expression in a subset of T-acute lymphoblastic leukaemia patients. *Oncogene*, 29, 5796-808.
- ORKIN, S. H. 2000. Diversification of haematopoietic stem cells to specific lineages. *Nat Rev Genet*, 1, 57-64.
- PADDISON, P. J., CLEARY, M., SILVA, J. M., CHANG, K., SHETH, N., SACHIDANANDAM, R. & HANNON, G. J. 2004. Cloning of short hairpin RNAs for gene knockdown in mammalian cells. *Nat Methods*, 1, 163-7.
- PAI, S. Y., TRUITT, M. L., TING, C. N., LEIDEN, J. M., GLIMCHER, L. H. & HO, I. C. 2003. Critical roles for transcription factor GATA-3 in thymocyte development. *Immunity*, 19, 863-75.
- PALII, C. G., PEREZ-IRATXETA, C., YAO, Z., CAO, Y., DAI, F., DAVISON, J., ATKINS, H., ALLAN, D., DILWORTH, F. J., GENTLEMAN, R., TAPSCOTT, S. J. & BRAND, M. 2011. Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *EMBO J*, 30, 494-509.
- PALIS, J., ROBERTSON, S., KENNEDY, M., WALL, C. & KELLER, G. 1999. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development*, 126, 5073-84.
- PALOMERO, T., ODOM, D. T., O'NEIL, J., FERRANDO, A. A., MARGOLIN, A., NEUBERG, D. S., WINTER, S. S., LARSON, R. S., LI, W., LIU, X. S., YOUNG, R. A. & LOOK, A. T. 2006. Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. *Blood*, 108, 986-92.

- PARONETTO, M. P. 2013. Ewing sarcoma protein: a key player in human cancer. *Int J Cell Biol*, 2013, 642853.
- PATEL, J. P., GONEN, M., FIGUEROA, M. E., FERNANDEZ, H., SUN, Z., RACEVSKIS, J., VAN VLIERBERGHE, P., DOLGALEV, I., THOMAS, S., AMINOVA, O., HUBERMAN, K., CHENG, J., VIALE, A., SOCCI, N. D., HEGUY, A., CHERRY, A., VANCE, G., HIGGINS, R. R., KETTERLING, R. P., GALLAGHER, R. E., LITZOW, M., VAN DEN BRINK, M. R., LAZARUS, H. M., ROWE, J. M., LUGER, S., FERRANDO, A., PAIETTA, E., TALLMAN, M. S., MELNICK, A., ABDEL-WAHAB, O. & LEVINE, R. L. 2012. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med*, 366, 1079-89.
- PAZ, H., LYNCH, M. R., BOGUE, C. W. & GASSON, J. C. 2010. The homeobox gene Hhex regulates the earliest stages of definitive hematopoiesis. *Blood*, 116, 1254-62.
- PERKINS, A. C., SHARPE, A. H. & ORKIN, S. H. 1995. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, 375, 318-22.
- PERRY, J. 2006. The Epc-N domain: a predicted protein-protein interaction domain found in select chromatin associated proteins. *BMC Genomics*, 7, 6.
- PEVNY, L., SIMON, M. C., ROBERTSON, E., KLEIN, W. H., TSAI, S. F., D'AGATI, V., ORKIN, S. H. & COSTANTINI, F. 1991. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, 349, 257-60.
- PIETERS, R. & CARROLL, W. L. 2008. Biology and treatment of acute lymphoblastic leukemia. *Pediatr Clin North Am*, 55, 1-20, ix.
- PORCHER, C., SWAT, W., ROCKWELL, K., FUJIWARA, Y., ALT, F. W. & ORKIN, S. H. 1996. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell*, 86, 47-57.
- PUI, C. H. & EVANS, W. E. 2006. Treatment of acute lymphoblastic leukemia. *N Engl J Med*, 354, 166-78.
- PUI, C. H., MULLIGHAN, C. G., EVANS, W. E. & RELLING, M. V. 2012. Pediatric acute lymphoblastic leukemia: where are we going and how do we get there? *Blood*, 120, 1165-74.
- PUI, C. H., PEI, D., CAMPANA, D., CHENG, C., SANDLUND, J. T., BOWMAN, W. P., HUDSON, M. M., RIBEIRO, R. C., RAIMONDI, S. C., JEHA, S., HOWARD, S. C., BHOJWANI, D., INABA, H., RUBNITZ, J. E., METZGER, M. L., GRUBER, T. A., COUSTAN-SMITH, E., DOWNING, J. R., LEUNG, W. H., RELLING, M. V. & EVANS, W. E. 2014. A revised definition for cure of childhood acute lymphoblastic leukemia. *Leukemia*, 28, 2336-43.
- PUI, C. H., ROBISON, L. L. & LOOK, A. T. 2008. Acute lymphoblastic leukaemia. *Lancet*, 371, 1030-43.
- PUI, J. C., ALLMAN, D., XU, L., DEROCOCCO, S., KARNELL, F. G., BAKKOUR, S., LEE, J. Y., KADESCH, T., HARDY, R. R., ASTER, J. C. & PEAR, W. S. 1999. Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity*, 11, 299-308.
- PUNT, J. A., OSBORNE, B. A., TAKAHAMA, Y., SHARROW, S. O. & SINGER, A. 1994. Negative selection of CD4+CD8+ thymocytes by T cell receptor-induced

- apoptosis requires a costimulatory signal that can be provided by CD28. *J Exp Med*, 179, 709-13.
- RABBITS, T. H. 1998. LMO T-cell translocation oncogenes typify genes activated by chromosomal translocations that alter transcription and developmental processes. *Genes Dev*, 12, 2651-7.
- RABBITS, T. H., BUCHER, K., CHUNG, G., GRUTZ, G., WARREN, A. & YAMADA, Y. 1999. The effect of chromosomal translocations in acute leukemias: the LMO2 paradigm in transcription and development. *Cancer Res*, 59, 1794s-1798s.
- RADTKE, F., MACDONALD, H. R. & TACCHINI-COTTIER, F. 2013. Regulation of innate and adaptive immunity by Notch. *Nat Rev Immunol*, 13, 427-37.
- RAHMAN, S., MAGNUSSEN, M., LEON, T. E., FARAH, N., LI, Z., ABRAHAM, B. J., ALAPI, K. Z., MITCHELL, R. J., NAUGHTON, T., FIELDING, A. K., PIZZEY, A., BUSTRAAN, S., ALLEN, C., POPA, T., PIKE-OVERZET, K., GARCIA-PEREZ, L., GALE, R. E., LINCH, D. C., STAAL, F. J. T., YOUNG, R. A., LOOK, A. T. & MANSOUR, M. R. 2017. Activation of the LMO2 oncogene through a somatically acquired neomorphic promoter in T-cell acute lymphoblastic leukemia. *Blood*, 129, 3221-3226.
- RAMIRO, A. R., TRIGUEROS, C., MARQUEZ, C., SAN MILLAN, J. L. & TORIBIO, M. L. 1996. Regulation of pre-T cell receptor (pT alpha-TCR beta) gene expression during human thymic development. *J Exp Med*, 184, 519-30.
- RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, 8, 2281-2308.
- REMKE, M., PFISTER, S., KOX, C., TOEDT, G., BECKER, N., BENNER, A., WERFT, W., BREIT, S., LIU, S., ENGEL, F., WITTMANN, A., ZIMMERMANN, M., STANULLA, M., SCHRAPPE, M., LUDWIG, W. D., BARTRAM, C. R., RADLWIMMER, B., MUCKENTHALER, M. U., LICHTER, P. & KULOZIK, A. E. 2009. High-resolution genomic profiling of childhood T-ALL reveals frequent copy-number alterations affecting the TGF-beta and PI3K-AKT pathways and deletions at 6q15-16.1 as a genomic marker for unfavorable early treatment response. *Blood*, 114, 1053-62.
- REYNOLDS, N., SALMON-DIVON, M., DVINGE, H., HYNES-ALLEN, A., BALASOORIYA, G., LEAFORD, D., BEHRENS, A., BERTONE, P. & HENDRICH, B. 2012. NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *EMBO J*, 31, 593-605.
- RIEGER, M. A. & SCHROEDER, T. 2009. Analyzing cell fate control by cytokines through continuous single cell biochemistry. *J Cell Biochem*, 108, 343-52.
- RON, G., GLOBERSON, Y., MORAN, D. & KAPLAN, T. 2017. Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nat Commun*, 8, 2237.
- ROTHENBERG, E. V., MOORE, J. E. & YUI, M. A. 2008. Launching the T-cell-lineage developmental programme. *Nat Rev Immunol*, 8, 9-21.
- ROTHENBERG, E. V., UNGERBACK, J. & CHAMPHEKAR, A. 2016. Forging T-Lymphocyte Identity: Intersecting Networks of Transcriptional Control. *Adv Immunol*, 129, 109-74.

- ROTHENBERG, E. V., ZHANG, J. & LI, L. 2010. Multilayered specification of the T-cell lineage fate. *Immunol Rev*, 238, 150-68.
- ROYER-POKORA, B., LOOS, U. & LUDWIG, W. D. 1991. TTG-2, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukaemia with the t(11;14)(p13;q11). *Oncogene*, 6, 1887-93.
- SANCHEZ-GARCIA, I. & RABBITS, T. H. 1993. LIM domain proteins in leukaemia and development. *Semin Cancer Biol*, 4, 349-58.
- SANCHEZ, R. & ZHOU, M. M. 2011. The PHD finger: a versatile epigenome reader. *Trends Biochem Sci*, 36, 364-72.
- SANDA, T., LAWTON, L. N., BARRASA, M. I., FAN, Z. P., KOHLHAMMER, H., GUTIERREZ, A., MA, W., TATAREK, J., AHN, Y., KELLIHER, M. A., JAMIESON, C. H., STAUDT, L. M., YOUNG, R. A. & LOOK, A. T. 2012. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell*, 22, 209-21.
- SANDBERG, Y., VERHAAF, B., VAN GASTEL-MOL, E. J., WOLVERS-TETTERO, I. L., DE VOS, J., MACLEOD, R. A., NOORDZIJ, J. G., DIK, W. A., VAN DONGEN, J. J. & LANGERAK, A. W. 2007. Human T-cell lines with well-defined T-cell receptor gene rearrangements as controls for the BIOMED-2 multiplex polymerase chain reaction tubes. *Leukemia*, 21, 230-7.
- SANDER, J. D. & JOUNG, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, 32, 347-55.
- SANG, M., MA, L., SANG, M., ZHOU, X., GAO, W. & GENG, C. 2014. LIM-domain-only proteins: multifunctional nuclear transcription coregulators that interacts with diverse proteins. *Mol Biol Rep*, 41, 1067-73.
- SARAN, N., LYSZKIEWICZ, M., POMMERENCKE, J., WITZLAU, K., VAKILZADEH, R., BALLMAIER, M., VON BOEHMER, H. & KRUEGER, A. 2010. Multiple extrathymic precursors contribute to T-cell development with different kinetics. *Blood*, 115, 1137-44.
- SAWADA, S. & LITTMAN, D. R. 1993. A heterodimer of HEB and an E12-related protein interacts with the CD4 enhancer and regulates its activity in T-cell lines. *Mol Cell Biol*, 13, 5620-8.
- SCHMEICHEL, K. L. & BECKERLE, M. C. 1994. The LIM domain is a modular protein-binding interface. *Cell*, 79, 211-9.
- SCHULZ, C., GOMEZ PERDIGUERO, E., CHORRO, L., SZABO-ROGERS, H., CAGNARD, N., KIERDORF, K., PRINZ, M., WU, B., JACOBSEN, S. E., POLLARD, J. W., FRAMPTON, J., LIU, K. J. & GEISSMANN, F. 2012. A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science*, 336, 86-90.
- SCHURINGA, J. J., WU, K., MORRONE, G. & MOORE, M. A. 2004. Enforced activation of STAT5A facilitates the generation of embryonic stem-derived hematopoietic stem cells that contribute to hematopoiesis in vivo. *Stem Cells*, 22, 1191-204.
- SEITA, J., SAHOO, D., ROSSI, D. J., BHATTACHARYA, D., SERWOLD, T., INLAY, M. A., EHRLICH, L. I., FATHMAN, J. W., DILL, D. L. & WEISSMAN, I. L. 2012. Gene Expression Commons: an open platform for absolute gene expression profiling. *PLoS One*, 7, e40321.

- SERWOLD, T., EHRLICH, L. I. & WEISSMAN, I. L. 2009. Reductive isolation from bone marrow and blood implicates common lymphoid progenitors as the major source of thymopoiesis. *Blood*, 113, 807-15.
- SHARMA, L., WU, W., DHOLAKIYA, S. L., GORASIYA, S., WU, J., SITAPARA, R., PATEL, V., WANG, M., ZUR, M., REDDY, S., SIEGELAUB, N., BAMBA, K., BARILE, F. A. & MANTELL, L. L. 2014. Assessment of phagocytic activity of cultured macrophages using fluorescence microscopy and flow cytometry. *Methods Mol Biol*, 1172, 137-45.
- SHARROCKS, A. D. 2001. The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol*, 2, 827-37.
- SHIVDASANI, R. A., MAYER, E. L. & ORKIN, S. H. 1995. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature*, 373, 432-4.
- SIMONIS, M., KLOUS, P., HOMMINGA, I., GALJAARD, R. J., RIJKERS, E. J., GROSVELD, F., MEIJERINK, J. P. & DE LAAT, W. 2009. High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat Methods*, 6, 837-42.
- SINCENNES, M. C., HUMBERT, M., GRONDIN, B., LISI, V., VEIGA, D. F., HAMAN, A., CAZAUX, C., MASHTALIR, N., AFFAR EL, B., VERREAULT, A. & HOANG, T. 2016. The LMO2 oncogene regulates DNA replication in hematopoietic cells. *Proc Natl Acad Sci U S A*, 113, 1393-8.
- SINGH, H., DEKOTER, R. P. & WALSH, J. C. 1999. PU.1, a shared transcriptional regulator of lymphoid and myeloid cell fates. *Cold Spring Harb Symp Quant Biol*, 64, 13-20.
- SIVE, J. I. & GOTTGENS, B. 2014. Transcriptional network control of normal and leukaemic haematopoiesis. *Exp Cell Res*, 329, 255-64.
- SMITH, L., PLUG, A. & THAYER, M. 2001. Delayed replication timing leads to delayed mitotic chromosome condensation and chromosomal instability of chromosome translocations. *Proc Natl Acad Sci U S A*, 98, 13300-5.
- SMITH, S., TRIPATHI, R., GOODINGS, C., CLEVELAND, S., MATHIAS, E., HARDAWAY, J. A., ELLIOTT, N., YI, Y., CHEN, X., DOWNING, J., MULLIGHAN, C., SWING, D. A., TESSAROLLO, L., LI, L., LOVE, P., JENKINS, N. A., COPELAND, N. G., THOMPSON, M. A., DU, Y. & DAVE, U. P. 2014. LIM domain only-2 (LMO2) induces T-cell leukemia by two distinct pathways. *PLoS One*, 9, e85883.
- SONMEZ, M., AKAGUN, T., COBANOGU, U., TOPBAS, M., ERKUT, N., YILMAZ, M., OVALI, E. & OMAI, S. B. 2009. Effect of LMO2 protein expression on survival in chronic myeloid leukemia patients treated with imatinib mesylate. *Hematology*, 14, 220-3.
- SOTO-FELICIANO, Y. M., BARTLEBAUGH, J. M. E., LIU, Y., SANCHEZ-RIVERA, F. J., BHUTKAR, A., WEINTRAUB, A. S., BUENROSTRO, J. D., CHENG, C. S., REGEV, A., JACKS, T. E., YOUNG, R. A. & HEMANN, M. T. 2017. PHF6 regulates phenotypic plasticity through chromatin organization within lineage-specific genes. *Genes Dev*, 31, 973-989.
- SOUPLIER, J., CLAPPIER, E., CAYUELA, J. M., REGNAULT, A., GARCIA-PEYDRO, M., DOMBRET, H., BARUCHEL, A., TORIBIO, M. L. & SIGAUX, F. 2005. HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood*, 106, 274-86.

- SOUROULLAS, G. P., SALMON, J. M., SABLITZKY, F., CURTIS, D. J. & GOODELL, M. A. 2009. Adult hematopoietic stem and progenitor cells require either Lyl1 or Scl for survival. *Cell Stem Cell*, 4, 180-6.
- SPANGRUDE, G. J., HEIMFELD, S. & WEISSMAN, I. L. 1988. Purification and characterization of mouse hematopoietic stem cells. *Science*, 241, 58-62.
- SPITS, H. 2002. Development of alphabeta T cells in the human thymus. *Nat Rev Immunol*, 2, 760-72.
- STAAL, F. J., WEERKAMP, F., LANGERAK, A. W., HENDRIKS, R. W. & CLEVERS, H. C. 2001. Transcriptional control of t lymphocyte differentiation. *Stem Cells*, 19, 165-79.
- STAAL, F. J., WIEKMEIJER, A. S., BRUGMAN, M. H. & PIKE-OVERZET, K. 2016. The functional relationship between hematopoietic stem cells and developing T lymphocytes. *Ann N Y Acad Sci*, 1370, 36-44.
- STANULOVIC, V. S., CAUCHY, P., ASSI, S. A. & HOOGENKAMP, M. 2017. LMO2 is required for TAL1 DNA binding activity and initiation of definitive haematopoiesis at the haemangioblast stage. *Nucleic Acids Res*, 45, 9874-9888.
- SUZUKI, Y., TSUNODA, T., SESE, J., TAIRA, H., MIZUSHIMA-SUGANO, J., HATA, H., OTA, T., ISOGAI, T., TANAKA, T., NAKAMURA, Y., SUYAMA, A., SAKAKI, Y., MORISHITA, S., OKUBO, K. & SUGANO, S. 2001. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res*, 11, 677-84.
- SWAIN, S. L. 1983. T cell subsets and the recognition of MHC class. *Immunol Rev*, 74, 129-42.
- TABRIZIFARD, S., OLARU, A., PLOTKIN, J., FALLAHI-SICHANI, M., LIVAK, F. & PETRIE, H. T. 2004. Analysis of transcription factor expression during discrete stages of postnatal thymocyte differentiation. *J Immunol*, 173, 1094-102.
- TAGHON, T., VAN DE WALLE, I., DE SMET, G., DE SMEDT, M., LECLERCQ, G., VANDEKERCKHOVE, B. & PLUM, J. 2009. Notch signaling is required for proliferation but not for differentiation at a well-defined beta-selection checkpoint during human T-cell development. *Blood*, 113, 3254-63.
- TAGHON, T., YUI, M. A., PANT, R., DIAMOND, R. A. & ROTHENBERG, E. V. 2006. Developmental and molecular characterization of emerging beta- and gammadelta-selected pre-T cells in the adult mouse thymus. *Immunity*, 24, 53-64.
- TAKEUCHI, A., YAMASAKI, S., TAKASE, K., NAKATSU, F., ARASE, H., ONODERA, M. & SAITO, T. 2001. E2A and HEB activate the pre-TCR alpha promoter during immature T cell development. *J Immunol*, 167, 2157-63.
- TAN, T. K., ZHANG, C. & SANDA, T. 2019. Oncogenic transcriptional program driven by TAL1 in T-cell acute lymphoblastic leukemia. *Int J Hematol*, 109, 5-17.
- TEAGUE, T. K., TAN, C., MARINO, J. H., DAVIS, B. K., TAYLOR, A. A., HUEY, R. W. & VAN DE WIELE, C. J. 2010. CD28 expression redefines thymocyte development during the pre-T to DP transition. *Int Immunol*, 22, 387-97.
- TEH, H. S., KISIELOW, P., SCOTT, B., KISHI, H., UEMATSU, Y., BLUTHMANN, H. & VON BOEHMER, H. 1988. Thymic major histocompatibility complex antigens and the alpha beta T-cell receptor determine the CD4/CD8 phenotype of T cells. *Nature*, 335, 229-33.

- THOMS, J. A., BIRGER, Y., FOSTER, S., KNEZEVIC, K., KIRSCHENBAUM, Y., CHANDRAKANTHAN, V., JONQUIERES, G., SPENSBERGER, D., WONG, J. W., ORAM, S. H., KINSTON, S. J., GRONER, Y., LOCK, R., MACKENZIE, K. L., GOTTGENS, B., IZRAELI, S. & PIMANDA, J. E. 2011. ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer. *Blood*, 117, 7079-89.
- TIAN, J., LI, Z., HAN, Y., JIANG, T., SONG, X. & JIANG, G. 2016. The progress of early growth response factor 1 and leukemia. *Intractable Rare Dis Res*, 5, 76-82.
- TILL, J. E. & MC, C. E. 1961. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res*, 14, 213-22.
- TODD, M. A., HUH, M. S. & PICKETTS, D. J. 2016. The sub-nucleolar localization of PHF6 defines its role in rDNA transcription and early processing events. *Eur J Hum Genet*, 24, 1453-9.
- TODD, M. A. & PICKETTS, D. J. 2012. PHF6 interacts with the nucleosome remodeling and deacetylation (NuRD) complex. *J Proteome Res*, 11, 4326-37.
- TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7, 562-78.
- TREMBLAY, M., TREMBLAY, C. S., HERBLOT, S., APLAN, P. D., HEBERT, J., PERREAULT, C. & HOANG, T. 2010. Modeling T-cell acute lymphoblastic leukemia induced by the SCL and LMO1 oncogenes. *Genes Dev*, 24, 1093-105.
- TSAI, F. Y., KELLER, G., KUO, F. C., WEISS, M., CHEN, J., ROSENBLATT, M., ALT, F. W. & ORKIN, S. H. 1994. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, 371, 221-6.
- TYDELL, C. C., DAVID-FUNG, E. S., MOORE, J. E., ROWEN, L., TAGHON, T. & ROTHENBERG, E. V. 2007. Molecular dissection of prethymic progenitor entry into the T lymphocyte developmental pathway. *J Immunol*, 179, 421-38.
- UCHIDA, N. & WEISSMAN, I. L. 1992. Searching for hematopoietic stem cells: evidence that Thy-1.1<sup>lo</sup> Lin<sup>-</sup> Sca-1<sup>+</sup> cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. *J Exp Med*, 175, 175-84.
- VAN DER WEERD, K., DIK, W. A., SCHRIJVER, B., BOGERS, A. J., MAAT, A. P., VAN NEDERVEEN, F. H., VAN HAGEN, P. M., VAN DONGEN, J. J., LANGERAK, A. W. & STAAL, F. J. 2013. Combined TCRG and TCRA TREC analysis reveals increased peripheral T-lymphocyte but constant intra-thymic proliferative history upon ageing. *Mol Immunol*, 53, 302-12.
- VAN GROTEL, M., MEIJERINK, J. P., BEVERLOO, H. B., LANGERAK, A. W., BUYS-GLADDINES, J. G., SCHNEIDER, P., POULSEN, T. S., DEN BOER, M. L., HORSTMANN, M., KAMPS, W. A., VEERMAN, A. J., VAN WERING, E. R., VAN NOESEL, M. M. & PIETERS, R. 2006. The outcome of molecular-cytogenetic subgroups in pediatric T-cell acute lymphoblastic leukemia: a retrospective study of patients treated according to DCOG or COALL protocols. *Haematologica*, 91, 1212-21.
- VAN VLIERBERGHE, P. & FERRANDO, A. 2012. The molecular basis of T cell acute lymphoblastic leukemia. *J Clin Invest*, 122, 3398-406.

- VAN VLIERBERGHE, P., PALOMERO, T., KHIABANIAN, H., VAN DER MEULEN, J., CASTILLO, M., VAN ROY, N., DE MOERLOOSE, B., PHILIPPE, J., GONZALEZ-GARCIA, S., TORIBIO, M. L., TAGHON, T., ZUURBIER, L., CAUWELIER, B., HARRISON, C. J., SCHWAB, C., PISECKER, M., STREHL, S., LANGERAK, A. W., GECZ, J., SONNEVELD, E., PIETERS, R., PAIETTA, E., ROWE, J. M., WIERNIK, P. H., BENOIT, Y., SOULIER, J., POPPE, B., YAO, X., CORDON-CARDO, C., MEIJERINK, J., RABADAN, R., SPELEMAN, F. & FERRANDO, A. 2010. PHF6 mutations in T-cell acute lymphoblastic leukemia. *Nat Genet*, 42, 338-42.
- VAN VLIERBERGHE, P., PATEL, J., ABDEL-WAHAB, O., LOBRY, C., HEDVAT, C. V., BALBIN, M., NICOLAS, C., PAYER, A. R., FERNANDEZ, H. F., TALLMAN, M. S., PAIETTA, E., MELNICK, A., VANDENBERGHE, P., SPELEMAN, F., AIFANTIS, I., COOLS, J., LEVINE, R. & FERRANDO, A. 2011. PHF6 mutations in adult acute myeloid leukemia. *Leukemia*, 25, 130-4.
- VAN VLIERBERGHE, P., PIETERS, R., BEVERLOO, H. B. & MEIJERINK, J. P. 2008. Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. *Br J Haematol*, 143, 153-68.
- VAN VLIERBERGHE, P., VAN GROTEL, M., BEVERLOO, H. B., LEE, C., HELGASON, T., BUIJS-GLADDINES, J., PASSIER, M., VAN WERING, E. R., VEERMAN, A. J., KAMPS, W. A., MEIJERINK, J. P. & PIETERS, R. 2006. The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. *Blood*, 108, 3520-9.
- VELTEN, L., HAAS, S. F., RAFFEL, S., BLASZKIEWICZ, S., ISLAM, S., HENNIG, B. P., HIRCHE, C., LUTZ, C., BUSS, E. C., NOWAK, D., BOCH, T., HOFMANN, W. K., HO, A. D., HUBER, W., TRUMPP, A., ESSERS, M. A. & STEINMETZ, L. M. 2017. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*, 19, 271-281.
- VICENTE, C., SCHWAB, C., BROUX, M., GEERDENS, E., DEGRYSE, S., DEMEYER, S., LAHORTIGA, I., ELLIOTT, A., CHILTON, L., LA STARZA, R., MECUCCI, C., VANDENBERGHE, P., GOULDEN, N., VORA, A., MOORMAN, A. V., SOULIER, J., HARRISON, C. J., CLAPPIER, E. & COOLS, J. 2015. Targeted sequencing identifies associations between IL7R-JAK mutations and epigenetic modulators in T-cell acute lymphoblastic leukemia. *Haematologica*, 100, 1301-10.
- VICENTE, R., SWAINSON, L., MARTY-GRES, S., DE BARROS, S. C., KINET, S., ZIMMERMANN, V. S. & TAYLOR, N. 2010. Molecular and cellular basis of T cell lineage commitment. *Semin Immunol*, 22, 270-5.
- VIGANO, M. A., IVANEK, R., BALWIERZ, P., BERNINGER, P., VAN NIMWEGEN, E., KARJALAINEN, K. & ROLINK, A. 2014. An epigenetic profile of early T-cell development from multipotent progenitors to committed T-cell descendants. *Eur J Immunol*, 44, 1181-93.
- VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJSER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009a. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457, 854-8.
- VISEL, A., RUBIN, E. M. & PENNACCHIO, L. A. 2009b. Genomic views of distant-acting enhancers. *Nature*, 461, 199-205.



- VITELLI, L., CONDORELLI, G., LULLI, V., HOANG, T., LUCHETTI, L., CROCE, C. M. & PESCHLE, C. 2000. A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts. *Mol Cell Biol*, 20, 5330-42.
- VON LINTIG, F. C., HUVAR, I., LAW, P., DICCIANNI, M. B., YU, A. L. & BOSS, G. R. 2000. Ras activation in normal white blood cells and childhood acute lymphoblastic leukemia. *Clin Cancer Res*, 6, 1804-10.
- VOSS, A. K., GAMBLE, R., COLLIN, C., SHOUBRIDGE, C., CORBETT, M., GECZ, J. & THOMAS, T. 2007. Protein and gene expression analysis of Phf6, the gene mutated in the Borjeson-Forssman-Lehmann Syndrome of intellectual disability and obesity. *Gene Expr Patterns*, 7, 858-71.
- WADMAN, I., LI, J., BASH, R. O., FORSTER, A., OSADA, H., RABBITTS, T. H. & BAER, R. 1994. Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia. *EMBO J*, 13, 4831-9.
- WADMAN, I. A., OSADA, H., GRUTZ, G. G., AGULNICK, A. D., WESTPHAL, H., FORSTER, A. & RABBITTS, T. H. 1997. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J*, 16, 3145-57.
- WAKABAYASHI, Y., WATANABE, H., INOUE, J., TAKEDA, N., SAKATA, J., MISHIMA, Y., HITOMI, J., YAMAMOTO, T., UTSUYAMA, M., NIWA, O., AIZAWA, S. & KOMINAMI, R. 2003. Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. *Nat Immunol*, 4, 533-9.
- WALSH, J. C., DEKOTER, R. P., LEE, H. J., SMITH, E. D., LANCKI, D. W., GURISH, M. F., FRIEND, D. S., STEVENS, R. L., ANASTASI, J. & SINGH, H. 2002. Cooperative and antagonistic interplay between PU.1 and GATA-2 in the specification of myeloid cell fates. *Immunity*, 17, 665-76.
- WANG, D., CLAUS, C. L., VACCARELLI, G., BRAUNSTEIN, M., SCHMITT, T. M., ZUNIGA-PFLUCKER, J. C., ROTHENBERG, E. V. & ANDERSON, M. K. 2006. The basic helix-loop-helix transcription factor HEBAt is expressed in pro-T cells and enhances the generation of T cell precursors. *J Immunol*, 177, 109-19.
- WANG, J., LEUNG, J. W., GONG, Z., FENG, L., SHI, X. & CHEN, J. 2013. PHF6 regulates cell cycle progression by suppressing ribosomal RNA synthesis. *J Biol Chem*, 288, 3174-83.
- WANG, Q., QIU, H., JIANG, H., WU, L., DONG, S., PAN, J., WANG, W., PING, N., XIA, J., SUN, A., WU, D., XUE, Y., DREXLER, H. G., MACLEOD, R. A. & CHEN, S. 2011. Mutations of PHF6 are associated with mutations of NOTCH1, JAK1 and rearrangement of SET-NUP214 in T-cell acute lymphoblastic leukemia. *Haematologica*, 96, 1808-14.
- WANG, Y., ZHANG, H., CHEN, Y., SUN, Y., YANG, F., YU, W., LIANG, J., SUN, L., YANG, X., SHI, L., LI, R., LI, Y., ZHANG, Y., LI, Q., YI, X. & SHANG, Y. 2009. LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell*, 138, 660-72.
- WARREN, A. J., COLLEDGE, W. H., CARLTON, M. B., EVANS, M. J., SMITH, A. J. & RABBITTS, T. H. 1994. The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. *Cell*, 78, 45-57.

- WAY, J. C. & CHALFIE, M. 1988. *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell*, 54, 5-16.
- WEBER, B. N., CHI, A. W., CHAVEZ, A., YASHIRO-OHTANI, Y., YANG, Q., SHESTOVA, O. & BHANDoola, A. 2011. A critical role for TCF-1 in T-lineage specification and differentiation. *Nature*, 476, 63-8.
- WEERKAMP, F., LUIS, T. C., NABER, B. A., KOSTER, E. E., JEANNOTTE, L., VAN DONGEN, J. J. & STAAL, F. J. 2006. Identification of Notch target genes in uncommitted T-cell progenitors: No direct induction of a T-cell specific gene program. *Leukemia*, 20, 1967-77.
- WEINTRAUB, A. S., LI, C. H., ZAMUDIO, A. V., SIGOVA, A. A., HANNETT, N. M., DAY, D. S., ABRAHAM, B. J., COHEN, M. A., NABET, B., BUCKLEY, D. L., GUO, Y. E., HNISZ, D., JAENISCH, R., BRADNER, J. E., GRAY, N. S. & YOUNG, R. A. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, 171, 1573-1588 e28.
- WENDORFF, A. A., QUINN, S. A., RASHKOVAN, M., MADUBATA, C. J., AMBESI-IMPIOMBATO, A., LITZOW, M. R., TALLMAN, M. S., PAIETTA, E., PAGANIN, M., BASSO, G., GASTIER-FOSTER, J. M., LOH, M. L., RABADAN, R., VAN Vlierberghe, P. & FERRANDO, A. A. 2019. Phf6 Loss Enhances HSC Self-Renewal Driving Tumor Initiation and Leukemia Stem Cell Activity in T-ALL. *Cancer Discov*, 9, 436-451.
- WENG, A. P., FERRANDO, A. A., LEE, W., MORRIS, J. P. T., SILVERMAN, L. B., SANCHEZ-IRIZARRY, C., BLACKLOW, S. C., LOOK, A. T. & ASTER, J. C. 2004. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*, 306, 269-71.
- WENG, A. P., MILLHOLLAND, J. M., YASHIRO-OHTANI, Y., ARCANGELI, M. L., LAU, A., WAI, C., DEL BIANCO, C., RODRIGUEZ, C. G., SAI, H., TOBIAS, J., LI, Y., WOLFE, M. S., SHACHAF, C., FELSHER, D., BLACKLOW, S. C., PEAR, W. S. & ASTER, J. C. 2006. c-Myc is an important direct target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. *Genes Dev*, 20, 2096-109.
- WIEKMEIJER, A. S., PIKE-OVERZET, K., BRUGMAN, M. H., VAN EGGERMOND, M., CORDES, M., DE HAAS, E. F. E., LI, Y., OOLE, E., VAN, I. W. F. J., EGELER, R. M., MEIJERINK, J. P. & STAAL, F. J. T. 2016. Overexpression of LMO2 causes aberrant human T-Cell development in vivo by three potentially distinct cellular mechanisms. *Exp Hematol*, 44, 838-849 e9.
- WILKINSON-WHITE, L., GAMSJAEGER, R., DASTMALCHI, S., WIENERT, B., STOKES, P. H., CROSSLEY, M., MACKAY, J. P. & MATTHEWS, J. M. 2011. Structural basis of simultaneous recruitment of the transcriptional regulators LMO2 and FOG1/ZFPM1 by the transcription factor GATA1. *Proc Natl Acad Sci U S A*, 108, 14443-8.
- WILKINSON, A. C. & GOTTGENS, B. 2013. Transcriptional regulation of haematopoietic stem cells. *Adv Exp Med Biol*, 786, 187-212.
- WILLIAMS, C. J., NAITO, T., ARCO, P. G., SEAVITT, J. R., CASHMAN, S. M., DE SOUZA, B., QI, X., KEABLES, P., VON ANDRIAN, U. H. & GEORGOPOULOS, K. 2004. The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. *Immunity*, 20, 719-33.

- WILSON, N. K., CALERO-NIETO, F. J., FERREIRA, R. & GOTTGENS, B. 2011. Transcriptional regulation of haematopoietic transcription factors. *Stem Cell Res Ther*, 2, 6.
- WILSON, N. K., FOSTER, S. D., WANG, X., KNEZEVIC, K., SCHUTTE, J., KAIMAKIS, P., CHILARSKA, P. M., KINSTON, S., OUWEHAND, W. H., DZIERZAK, E., PIMANDA, J. E., DE BRUIJN, M. F. & GOTTGENS, B. 2010. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, 7, 532-44.
- WOJCIECHOWSKI, J., LAI, A., KONDO, M. & ZHUANG, Y. 2007. E2A and HEB are required to block thymocyte proliferation prior to pre-TCR expression. *J Immunol*, 178, 5717-26.
- WOUTERS, B. J., JORDA, M. A., KEESHAN, K., LOUWERS, I., ERPELINCK-VERSCHUEREN, C. A., TIELEMANS, D., LANGERAK, A. W., HE, Y., YASHIRO-OHTANI, Y., ZHANG, P., HETHERINGTON, C. J., VERHAAK, R. G., VALK, P. J., LOWENBERG, B., TENEN, D. G., PEAR, W. S. & DELWEL, R. 2007. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood*, 110, 3706-14.
- WU, L., XU, Y., WANG, Q., RUAN, C., DREXLER, H. G., WU, D., MACLEOD, R. A. & CHEN, S. 2015. High frequency of cryptic chromosomal rearrangements involving the LMO2 gene in T-cell acute lymphoblastic leukemia. *Haematologica*, 100, e233-6.
- WYSOCKA, J., SWIGUT, T., XIAO, H., MILNE, T. A., KWON, S. Y., LANDRY, J., KAUER, M., TACKETT, A. J., CHAIT, B. T., BADENHORST, P., WU, C. & ALLIS, C. D. 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, 442, 86-90.
- XU, J., POPE, S. D., JAZIREHI, A. R., ATTEMA, J. L., PAPATHANASIOU, P., WATTS, J. A., ZARET, K. S., WEISSMAN, I. L. & SMALE, S. T. 2007. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc Natl Acad Sci U S A*, 104, 12377-82.
- XU, J., WATTS, J. A., POPE, S. D., GADUE, P., KAMPS, M., PLATH, K., ZARET, K. S. & SMALE, S. T. 2009. Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev*, 23, 2824-38.
- XU, W., CARR, T., RAMIREZ, K., MCGREGOR, S., SIGVARDSSON, M. & KEE, B. L. 2013. E2A transcription factors limit expression of Gata3 to facilitate T lymphocyte lineage commitment. *Blood*, 121, 1534-42.
- YADEN, B. C., SAVAGE, J. J., HUNTER, C. S. & RHODES, S. J. 2005. DNA recognition properties of the LHX3b LIM homeodomain transcription factor. *Mol Biol Rep*, 32, 1-6.
- YAMADA, T., YANG, Y., HEMBERG, M., YOSHIDA, T., CHO, H. Y., MURPHY, J. P., FIORAVANTE, D., REGEHR, W. G., GYGI, S. P., GEORGOPOULOS, K. & BONNI, A. 2014. Promoter decommissioning by the NuRD chromatin remodeling complex triggers synaptic connectivity in the mammalian brain. *Neuron*, 83, 122-34.

- YAMADA, Y., PANNELL, R., FORSTER, A. & RABBITS, T. H. 2000. The oncogenic LIM-only transcription factor Lmo2 regulates angiogenesis but not vasculogenesis in mice. *Proc Natl Acad Sci U S A*, 97, 320-4.
- YAMADA, Y., WARREN, A. J., DOBSON, C., FORSTER, A., PANNELL, R. & RABBITS, T. H. 1998. The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 3890-3895.
- YAMAMOTO, R., MORITA, Y., OOEHARA, J., HAMANAKA, S., ONODERA, M., RUDOLPH, K. L., EMA, H. & NAKAUCHI, H. 2013. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell*, 154, 1112-1126.
- YANG, L., BRYDER, D., ADOLFSSON, J., NYGREN, J., MANSSON, R., SIGVARDSSON, M. & JACOBSEN, S. E. 2005. Identification of Lin(-)Sca1(+)kit(+)CD34(+)Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood*, 105, 2717-23.
- YASHIRO-OHTANI, Y., HE, Y., OHTANI, T., JONES, M. E., SHESTOVA, O., XU, L., FANG, T. C., CHIANG, M. Y., INTLEKOFER, A. M., BLACKLOW, S. C., ZHUANG, Y. & PEAR, W. S. 2009. Pre-TCR signaling inactivates Notch1 transcription by antagonizing E2A. *Genes Dev*, 23, 1665-76.
- YOO, N. J., KIM, Y. R. & LEE, S. H. 2012. Somatic mutation of PHF6 gene in T-cell acute lymphoblastic leukemia, acute myelogenous leukemia and hepatocellular carcinoma. *Acta Oncol*, 51, 107-11.
- YOSHIDA, T., HAZAN, I., ZHANG, J., NG, S. Y., NAITO, T., SNIPPERT, H. J., HELLER, E. J., QI, X., LAWTON, L. N., WILLIAMS, C. J. & GEORGOPOULOS, K. 2008. The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation. *Genes Dev*, 22, 1174-89.
- YOSHIMOTO, M., PORAYETTE, P., GLOSSON, N. L., CONWAY, S. J., CARLESSO, N., CARDOSO, A. A., KAPLAN, M. H. & YODER, M. C. 2012. Autonomous murine T-cell progenitor production in the extra-embryonic yolk sac before HSC emergence. *Blood*, 119, 5706-14.
- YUI, M. A. & ROTHENBERG, E. V. 2014. Developmental gene networks: a triathlon on the course to T cell identity. *Nat Rev Immunol*, 14, 529-45.
- ZHANG, C., MEJIA, L. A., HUANG, J., VALNEGRI, P., BENNETT, E. J., ANCKAR, J., JAHANI-ASL, A., GALLARDO, G., IKEUCHI, Y., YAMADA, T., RUDNICKI, M., HARPER, J. W. & BONNI, A. 2013. The X-linked intellectual disability protein PHF6 associates with the PAF1 complex and regulates neuronal migration in the mammalian brain. *Neuron*, 78, 986-93.
- ZHANG, J., DING, L., HOLMFELDT, L., WU, G., HEATLEY, S. L., PAYNE-TURNER, D., EASTON, J., CHEN, X., WANG, J., RUSCH, M., LU, C., CHEN, S. C., WEI, L., COLLINS-UNDERWOOD, J. R., MA, J., ROBERTS, K. G., POUNDS, S. B., ULYANOV, A., BECKSFORT, J., GUPTA, P., HUETHER, R., KRIWACKI, R. W., PARKER, M., MCGOLDRICK, D. J., ZHAO, D., ALFORD, D., ESPY, S., BOBBA, K. C., SONG, G., PEI, D., CHENG, C., ROBERTS, S., BARBATO, M. I., CAMPANA, D., COUSTAN-SMITH, E., SHURTLEFF, S. A., RAIMONDI, S. C., KLEPPE, M., COOLS, J., SHIMANO, K. A., HERMISTON, M. L., DOULATOV, S., EPPERT, K., LAURENTI, E., NOTTA, F., DICK, J. E.,

- BASSO, G., HUNGER, S. P., LOH, M. L., DEVIDAS, M., WOOD, B., WINTER, S., DUNSMORE, K. P., FULTON, R. S., FULTON, L. L., HONG, X., HARRIS, C. C., DOOLING, D. J., OCHOA, K., JOHNSON, K. J., OBENAUER, J. C., EVANS, W. E., PUI, C. H., NAEVE, C. W., LEY, T. J., MARDIS, E. R., WILSON, R. K., DOWNING, J. R. & MULLIGHAN, C. G. 2012a. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature*, 481, 157-63.
- ZHANG, J. A., MORTAZAVI, A., WILLIAMS, B. A., WOLD, B. J. & ROTHENBERG, E. V. 2012b. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell*, 149, 467-82.
- ZHANG, Y., GARCIA-IBANEZ, L. & TOELLNER, K. M. 2016. Regulation of germinal center B-cell differentiation. *Immunol Rev*, 270, 8-19.
- ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, 9, R137.
- ZHENG, Q. & ZHAO, Y. 2007. The diverse biofunctions of LIM domain proteins: determined by subcellular localization and protein-protein interaction. *Biol Cell*, 99, 489-502.