INVESTIGATING THE ROLE OF HEB AND ITS INTERACTION WITH GFI1 IN T-CELL ACUTE LYMPHOBLASTIC LEUKAEMIA (T-ALL)

BY

SHIMAA ABDULRAHMAN ALAZMI

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

DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomics Sciences
College of Medical and Dental Sciences
University of Birmingham
November 2023

UNIVERSITYOF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

ABSTRACT

An intricate network of transcription factors determines gene expression patterns that define cellular identity. Development of T-cells involves dynamic changes, including the activation and silencing of transcription factors. As part of the transcriptional landscape, TAL1 forms a heterodimer with E-proteins such as E2A, E2-2, and HEB, which bind DNA and activate transcription as part of a complex, including LIM-only protein 2 (LMO2), Lim domain-binding protein 1 (LDB1), and GATA proteins. The overexpression of LMO proteins, as well as an abnormal expression of TAL1, has been identified as hallmarks of T-cell Acute Lymphoblastic Leukaemia (T-ALL). Considering the high relapse rates, drug resistance, and poor prognosis of T-ALL, it is imperative to investigate the molecular mechanisms underlying T-ALL in order to develop effective therapeutic strategies.

E-proteins have important roles in the progression of haematopoiesis, orchestrating thymocyte maturation and developmental progression. HEB, a member of the E-protein protein family, plays a pivotal role in guiding thymocytes toward the double positive stage and beyond. HEB exists in two isoforms: HEB canonical and HEB alternative. A previous study of HEB immunoprecipitation followed by mass spectrometry identified the transcription factor GFI1 as a potential new member of the complex. Recent research has shown that GFI1 has dual functions, acting as a transcriptional activator in conjunction with IKAROS and as a transcriptional repressor by recruiting KDM1A. However, the role of GFI1 in T-ALL remains unexplored.

This study identified the interaction between HEB and GFI1, particularly with the HEB alternative isoform, in four human T-ALL cell lines (ARR, DU528, HSB2, and CCRFCEM), using proteomic analysis and genome-wide studies. The genomic distribution of GFI1 was examined using chromatin immunoprecipitation (ChIP) followed by next generation

sequencing (ChIP-seq). Comparison of these data with previous ChIP-seq data revealed that distinct binding complexes exists in ARR and in the other three cell lines.

The role of HEB in T-ALL was further dissected using CRISPR-Cas9 gene editing. Targeting *TCF12*, the gene encoding HEB, led to the loss of the expression of the canonical isoform of HEB. This reduced the expression of its alternative isoform, providing evidence for a regulatory interaction between the two isoforms, as well as reduced the expression of GFI1. There was no significant morphological change or chromosomal instability in the ARR cells following HEB deletion, indicating that *TCF12* deletion does not cause overt cellular abnormalities. In response to the loss of HEB, LMO2 is observed to possess increased association with the DNA, increasing chromatin accessibility, and potentially compensating for the loss of HEB through enhanced DNA-cofactor interactions. Overall, these findings provide valuable insight into the molecular basis of T-ALL, which could lead to the development of targeted therapeutic interventions.

ACKNOWLEDGMENTS

First and foremost, I would like to praise Allah the Almighty, the Most Gracious and Merciful, for blessing me with the strength, wisdom and perseverance to complete this journey. With His guidance, all things are possible. My PhD started just before the Covid-19 pandemic, a period filled with uncertainties and challenges. Words are inadequate to express my gratitude to everyone who supported me in completing my doctorate, despite many unforeseen obstacles.

I would like to express my deepest gratitude to my supervisors, Dr Maarten Hoogenkamp and Dr Vesna Stanulovic, for their invaluable guidance, great patience and continuous support during this academic journey. I could not have done this work without them. I also extend my deepest appreciation to my colleagues for the rich experiences we have shared.

I owe enormous gratitude to my beloved parents, my father Dr Abdulrahman and my great mother Ghaliah, who inspired me and instilled in me the values of hard work and dedication. Your love, prayers and unwavering support have been the pillars of my motivation and strength.

Neither words nor actions could express my special acknowledgment to my husband Majed and my children, Ali, Manal and new baby Ghaliah for accompanying me during these four years. Their patience, understanding and sacrifices have been the cornerstone of my ability to pursue this dream. The unwavering support and the love and joy they bring to my life have been my sanctuary and my sources of comfort.

I would also like to thank my sisters, particularly, Tahreer, as well as Amna, Fatma and Ghaliah and my brothers, Homoud, Mohammed, Ahmed and Abdulaziz, for their kind words, encouragement and support whenever I needed them.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

°C	Degree Celsius
AGM	Aorta-gonado-mesonephros
AML	Acute myeloid leukaemia
B-ALL	B-cell acute lymphoblastic leukaemia
BCL11B	B-cell Leukaemia11 B
bHLH	basic Helix loop helix
CD	Cluster of differentiation
CDK	Cyclin-dependent kinase
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CLP	Common lymphoid progenitors
CMP	Common myeloid progenitors
Co-IP	Co-immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
DCs	Dendritic cells
DHS	DNaseI hypersensitivity site
DNA	Double negative
DNA	Deoxyribonucleic acid
DNAseI-seq	Deoxyribonuclease I sequencing
DP	Double positive
EHT	Endothelial-to-haematopoietic transition
EMP	erythro-myeloid progenitors
ES cells	Embryonic stem cells
ETP	Early T-cell progenitor
ETP-ALL	Early T-cell precursors acute lymphoblastic leukaemia
FACS	Fluorescence-activated cell sorting
FBS	Foetal bovine serum
FPKM	Fragments per kilobase of transcript per million mapped reads
GATA	GATA binding protein
GFI1	Growth Factor Independent 1
GFI1B	Growth Factor Independent 1B
GFP	Green fluorescent protein

GMP	Granulocyte/monocyte progenitors
GO	Gene ontology
GREAT	Genomic Regions Enrichment of Annotations Tool
HDAC	Histone deacetylase
HDR	Homology-directed repair
HE	Hemogenic Endothelium
HEB	Helix-loop-helix E-box
HEBalt	Alternative HEB
HEBcan	Canonical HEB
hESCs	human embryonic stem cells
HOXA	home box genes
HSC	Haematopoietic stem cell
HSCs	Haematopoietic stem cells
IC	Input control samples
ID	Inhibitors of DNA-binding
IP	Immunoprecipitation
ISP	Immature single positive
kDa	Kilodalton
KLF6	Kruppel-like factor 6
Kme1	Monomethylated lysine
KO	Knockout
LB	Lysogeny Broth
LDB1	LIM domain binding1
LMO	LIM-domain only
LMP	lympho-myeloid progenitors
LMPP	Lymphoid-primed multipotent progenitor
LSD1	Lysine specific demethylase
LSK	Lin/Sca-1/c-Kit
LYL1	Lymphoblastic leukaemia associated haematopoiesis regulator 1
MACS2	Model-based Analysis of ChIP-Seq 2
MeDIP	Methylated DNA immunoprecipitation
MEP	Megakaryocyte/Erythroid progenitor
MHC	Major histocompatibility complex

MLP	Multilineage progenitor
MPP	Multipotent progenitor
MS	Mass spectrometry
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NK	Natural killer
NOTCH1	NOTCH Receptor 1
PCI	Phenol/chloroform/isoamyl alcohol
PCR	Polymerase chain reaction
PHF6	Plant homeodomain zinc finger protein 6 PLA
PU.1	Transcription factor PU.1
qPCR	Quantitative polymerase chain reaction
RAG	Recombination activating gene
RAG1	Recombination activating protein 1
RFP	red fluorescent protein
RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
rRNA	ribosomal RNA
RT	Room temperature
RUNX1	RUNX Family Transcription Factor 1
sgRNA	single-guide RNA
SLAM	signalling lymphocyte activation molecule
SIL	SCL interrupting locus
SP	Single positive
SPI1	Spi-1 Proto-Oncogene
SREBF2	Sterol Regulatory Element Binding Transcription Factor 2
ST-HSCs	Short-term haematopoietic stem cells
T-ALL	T-cell Acute lymphoblastic leukaemia
TAL1	T-cell acute lymphoblastic leukaemia
TBP	TATA -box-binding protein
TCF-1	Transcription factor 1
TCF12	Transcription factor 12
TCR	T-cell receptor

TF	Transcription factor
TFAP2C	transcription factor AP-2 gamma
TSS	Transcription start site
μl	Micro litre
WB	Western Blot
WT	Wildtype

Chapter 1 General introduction

1.1 Haematopoiesis

Haematopoiesis refers to the process whereby the cellular components of blood are produced, during embryonic development and throughout adulthood, to establish and refresh the circulatory system. Different blood cell types (lineages) in the blood system have diverse functions. Leukocytes, white blood cells, are cell types required for the immune system (Bain, 2017). Erythrocytes, red blood cells, are responsible for delivering oxygen to tissues (Hamasaki and Yamamoto, 2000). Megakaryocytes form and release platelets, which have a vital role in controlling blood clotting, inflammation, wound healing, and tumour metastasis (Jurk and Kehrel, 2005). All blood lineages arise from haematopoietic stem cells (HSCs). Scientists have studied the development of the blood system in detail using a variety of models to enhance the understanding of this process, as well as blood disorders and cancers. The sites of haematopoiesis change anatomically during development. Before birth, it begins with primitive haematopoiesis in the yolk sac, whereas definitive haematopoiesis originates from the aorta-gonad-mesonephros (AGM), expands within the developing liver in mammals, and finally establishes itself in the bone marrow and thymus around birth (Jagannathan-Bogdan and Zon, 2013).

A mammal embryo undergoes several waves of haematopoietic development, each of which produces cohorts of cells with increasing complexity of blood lineage potential. There are three sequential waves or stages of haematopoiesis: the primitive wave, the pro-definitive wave, and the definitive wave (Figure 1.1). The first wave, the primitive wave is responsible for the formation of unipotent blood cells, it starts in the extraembryonic yolk sac during the gestation period which is at 7 days in mouse embryos and 18–20 days in human embryos, and their formation occurs primarily in the aortic endothelium of the AGM (Palis and Yoder, 2001). The main purpose of the primitive wave is to generate nucleated erythroid progenitors (primitive erythrocytes) that are essential for oxygen distribution required for the rapidly growing embryo,

while the first macrophages and megakaryocytes play an important role in tissue remodelling (Tober et al., 2007, Orkin and Zon, 2008).

The second wave called the "pro-definitive wave" arises in the yolk sac, it generates the HSCs that originates from haemogenic endothelial cells and give rise to multipotential haematopoietic progenitors. During this wave cells undergo a transition process called endothelial-to-haematopoietic transition (EHT) then migrate to the foetal liver and differentiate into erythro-myeloid progenitors (EMPs) and lympho-myeloid progenitors (LMPs) (Jagannathan-Bogdan and Zon, 2013). These cells lack the self-renewal capability which makes this a limited and transitory wave. Therefore, it is replaced by the third wave of haematopoiesis, that is permanent, called the definitive wave. The production of definitive HSCs and differentiated blood cells including erythrocytes, megakaryocytes, monocytes, B and T lymphocytes occur during this intraembryonic wave (Dzierzak and Bigas, 2018). Around birth, these cells further migrate to the spleen, where myeloid lineage differentiation occurs, and then finally reside in bone marrow, where they acquire the stem cell characteristics and adult HSC surface markers (Canu and Ruhrberg, 2021).
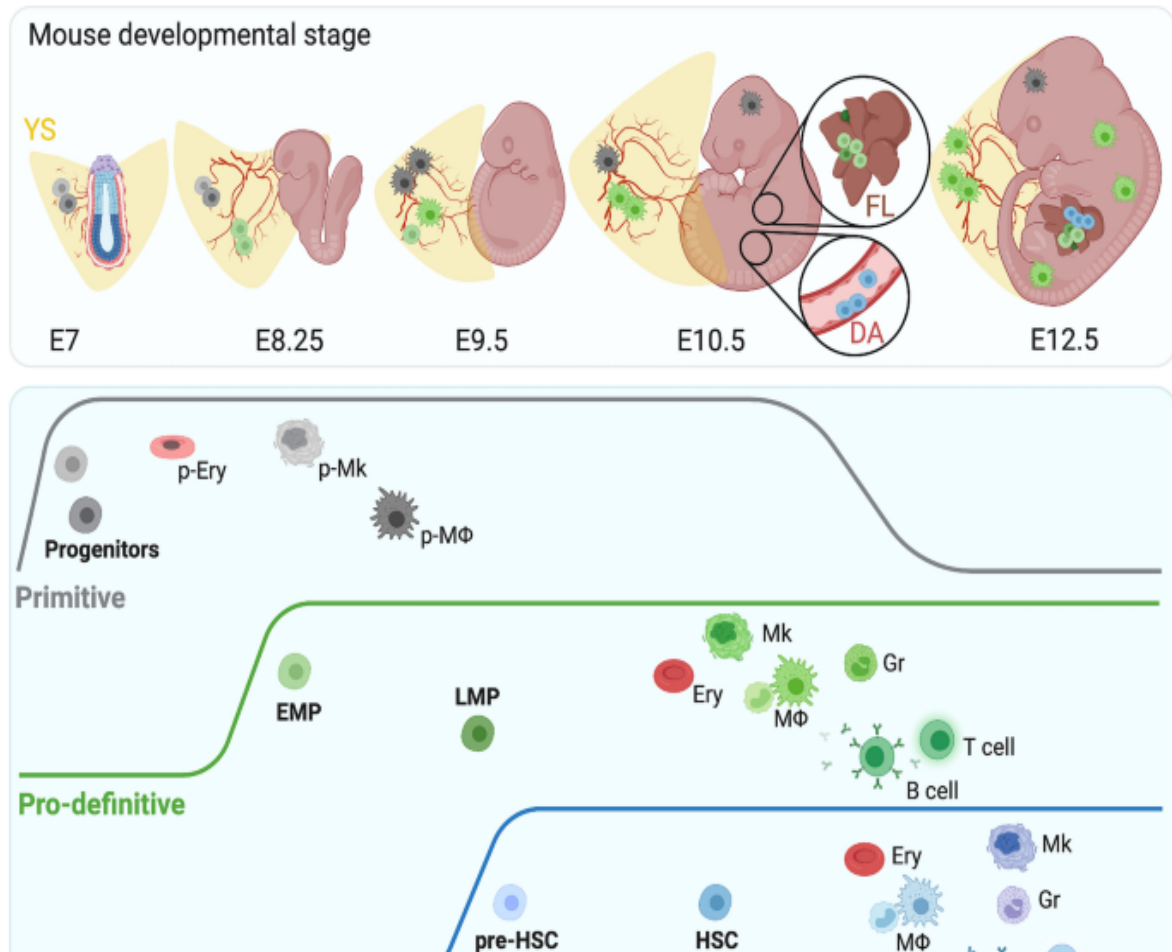
Figure 1.1: Embryonic haematopoietic development.

Schematic illustration of the haematopoietic development three sequential waves; primitive, pro-definitive and definitive waves. A mouse developmental stage timeline is shown at the top from embryonic day 7 (E7) to embryonic day 12.5 (E12.5). Highlights the foetal liver (FL) and the dorsal aorta (DA), which are important sites for blood cell development. The "Progenitors" section below details the types of primitive cells found at these stages, including progenitors for erythrocytes (p-Ery), megakaryocytes (p-Mk), and macrophages (p-Mo). In the middle section, the green line indicates blood cell maturation process. EMP (Erythro-Myeloid Progenitor) and LMP (Lymphoid-Myeloid Progenitor) line up, resulting in HSC at the "Definitive" stage. Each progenitor type is indicated by the following icons: erythrocytes (Ery), megakaryocytes (Mk), macrophages (Mo), granulocytes (Gr), T cells, and B cells. Taken from (Canu and Ruhrberg, 2021).

4

In adult vertebrates, the haematopoiesis system is a hierarchical system, with the adult HSCs located at the apex of this hierarchy. Within the haematopoietic system, HSCs are the only cells that can self-renew, giving rise to identical daughter cells without differentiation, in addition to generating multipotent progenitor cells that further differentiate into all blood cell types (Seita and Weissman, 2010). The differentiation starts from HSCs, which give rise to multipotent progenitors (MPPs) (Figure 1.2). At this stage, the self-renewal ability is lost, and cells retain the capacity to differentiate into various types of mature blood cell lineages. Subsequently, two cell lineages emerge from the MPP stage: common myeloid progenitors (CMPs) and Lymphoid-primed multipotent progenitors (LMPP). CMPs segregate into megakaryocyte/erythroid progenitors (MEPs) and granulocyte/monocyte progenitors (GMPs). MEPs further differentiate into mature megakaryocytes and erythrocytes, whereas GMPs maintain the production of mature granulocytes and monocytes. LMPP, on the other hand, have the potential to develop into granulocyte/monocyte progenitors or it can develop into common lymphoid progenitor that give rise to natural killer (NK) cells, and B-cells or T-cells.

Although the above description of haematological development helps to understand the generation of the different progenitor cell types, a newer continuum model of haematopoiesis shows that HSCs lead to the development of a mixture of biased cells rather than truly multipotent cells (Ceredig, R et al., 2009). This model indicates that HSCs can be affiliated with a single pathway leading to the production of a specific cell lineage. In this pathway, HSCs may not end in each type of cell but may prefer to produce cells based on certain shared characteristics, such as near-neighbour relationships. In a continuum model of haematopoiesis, cytokines have also been found to have an important role in lineage fate (Ceredig R et al., 2006). These haematopoietic cytokines not only guide cell lineage but also enhance the chances of survival and production of lineage-related cells (Brown, Sanchez and Sanchez-Garcia, 2020).

In a continuum model, HSCs may select a lineage from a range of possibilities, leading to a mixture of cells with specific lineage traits. HSC sub-populations develop low levels of mRNA for different receptors, such as macrophage colony-stimulating factor (M-CSFR), erythropoietin (EpoR), and granulocyte colony-stimulating factor (G-CSFR). These cytokines can generate chemical gradients to move cells to particular areas and can direct the development of the erythroid, monocyte, and neutrophil lineages, respectively (Eizenberg-Magar I et al., 2017). This shows that cytokine signals received by HSCs can affect their differentiation towards specific blood and immune cell types, showing the complex relationship between intrinsic factors, extrinsic signals, and cell lineage determination in haematopoiesis (Brown, 2021).

The functional characteristics of the multipotent stem cell concept was first discovered by Till and McCulloch in 1961. They identified the first quantitative, clonal method, when they studied the effect of radiation on the bone marrow of mice and noticed cells in mice spleens with the capability to colonise and self-renew (Till and Mc, 1961), (Becker et al., 1963). Subsequently, several groups of scientists identified cell surface markers to purify and understand the differentiation process of HSCs. HSCs were then subdivided and characterised based on their immunophenotype using FACS technology (fluorescence-activated cell sorting) (Spangrude et al., 1988, Akashi et al., 2000).

Haematopoietic stem cells and multipotent haematopoietic progenitors (MMP) were identified and purified based on a combination of surface markers to identify different progenitor populations, the signalling lymphocyte activation molecule (SLAM) family markers and Lineage, Stem cell antigen-1 (Sca-1), Kit (CD117) LSK markers (Morrison and Weissman, 1994). The expression of SLAM family receptors is differential among haematopoietic progenitors, which correlates with the primitiveness of the progenitor. It has been identified that receptors with combinatorial expression can accurately differentiate stem cells from

progenitor cells. Cell surface receptors of SLAM family include CD150 (also called Slamf1), CD48 (Slamf2), CD229 and CD244. The expression of CD150 was observed in HSCs but not in MPPs or restricted haematopoietic progenitors. The CD244 marker was found to be expressed by MPPs and some restricted progenitors, but not by HSCs. The CD48 marker was expressed by restricted B lineage and myeloerythroid lineage progenitors, but not by HSCs and MPPs. The expression of CD229 was found to distinguish lymphoid-biased HSCs from myeloid-biased HSCs (Kiel et al., 2005, Oguro et al., 2013).
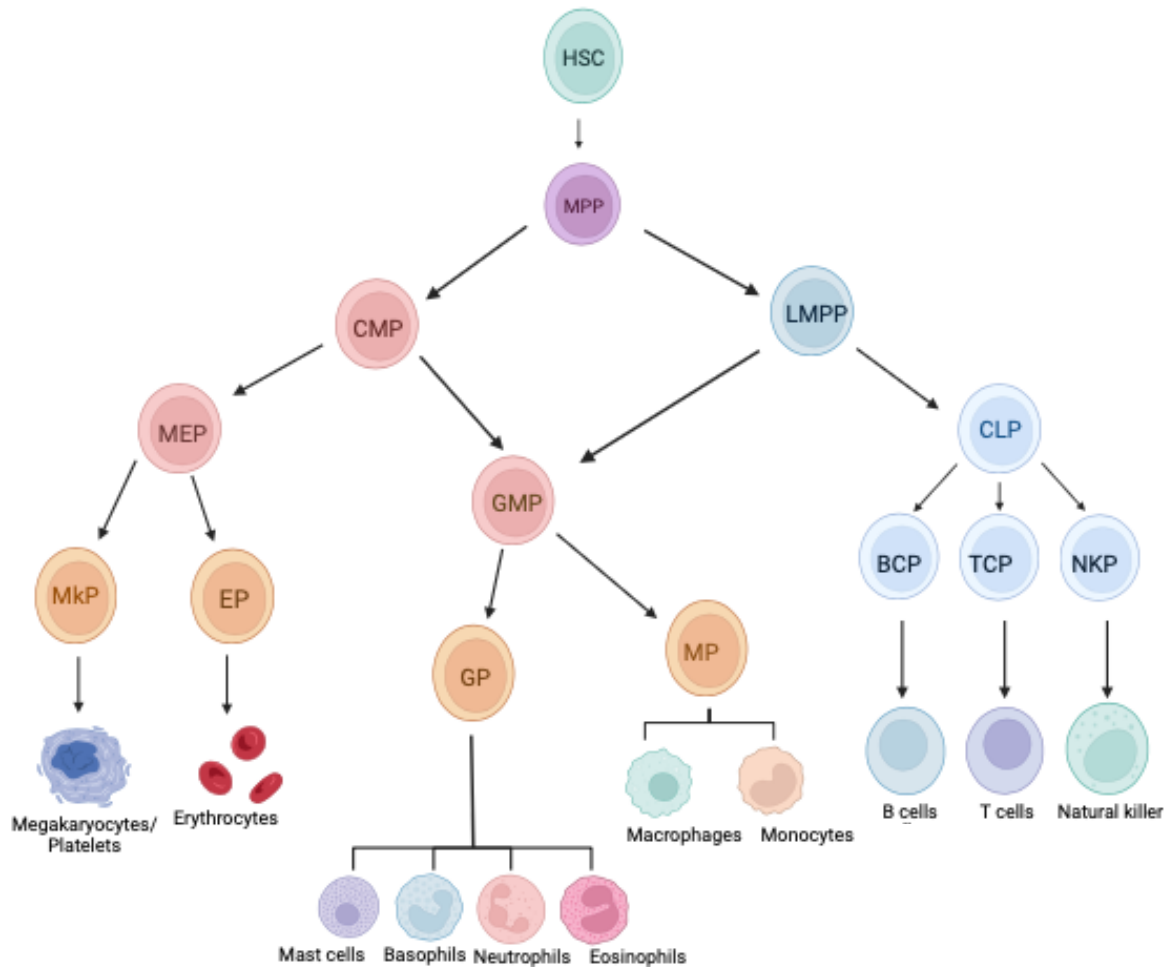
Figure 1.2: A schematic representation showing haematopoietic cells differentiation.

The diagram depicts the stepwise process of haematopoiesis, starting with haematopoietic stem cell (HSC) that differentiate into multipotent progenitor cells (MPP). MPPs branch into both myeloid or lymphoid lineages - the common myeloid progenitor (CMP) or Lymphoid-primed multipotent progenitor cell (LMPP). CMP develop into myeloid-erythroid progenitors (MEP), which then differentiate into megakaryocyte progenitors, platelets, or erythrocytes. Granulocyte-monocyte progenitors (GMP) may originate from CMP or LMPP. GMPs then further differentiate into mast cells, basophils, neutrophils, eosinophils, monocytes, and macrophages. LMPP can also produce lymphoid progenitors (CLP) which differentiate into; B cell progenitors (BCP) that become B cells, T cell progenitors (TCP) that become T cells, and natural killer progenitors (NKP) that become natural killer cells. Created in Biorender.com

1.1.1 Haematopoiesis regulatory factors

Haematopoiesis is a complex process regulated by multiple intrinsic and extrinsic factors. A combination of factors is needed to maintain an accurate balance between mature and immature cells (Robertson et al., 2000). For example, the fluctuation of gene expression and interactions between transcription factors are essential intrinsic factors that control the differentiation of cells toward a particular cell lineage (Tsai et al., 1994). Several studies have identified that Runt-related transcription factor 1 (RUNX1), LIM-only 2 (LMO-2) and GATA2, amongst others, are necessary for the development of HSCs (Lie et al., 2018, Yamada et al., 1998).

Additionally, blood lineage specification requires other transcription factors for differentiation and lineage commitment. For instance, GATA1 is known to mediate the differentiation of common myeloid progenitors towards megakaryocyte/erythroid progenitors and granulocyte/monocyte progenitors (Heyworth et al., 2002). The overexpression of GATA1, P21 (cyclin-dependent kinase inhibitor 1) and Krüppel-like factor (KLF-1) directs the cell differentiation toward erythroid. Additionally, PU.1 (gene name *Spi1*) and C/EBPα play an important role in lineage specification, leading to the commitment of cells towards macrophages, granulocytes, and B lymphocytes, whereas the PU.1 interaction with GATA3 stimulates the differentiation of T-cells (Xie et al., 2004, Laiosa et al., 2006).

The extrinsic factors are also necessary for the regulation of HSC expansion, cell fate determination and self-renewal. The bone marrow niche, the bone marrow microenvironment that consists of different types of haematopoietic and non-haematopoietic cells such as osteocytes, adipocytes, stromal cells, and extracellular matrix, is essential for the HSC senescence and regulation of stem and progenitor cell differentiation (Lee et al., 2017). A component of this microenvironment is the extracellular matrix, which is composed of collagen, proteoglycans, and glycoproteins. It plays an important role in the haematopoietic environment and controls haematopoiesis; it is also involved in a variety of biological

processes, including cell adhesion, growth factor interactions, various cytokines binding and presentation and cell signalling pathways (Verma et al., 2020).

1.2 T-cell development

The thymus is the site where T-lineage commitment and T-cell differentiation take place. Haematopoietic progenitors that have the potential to become T-cells need to migrate from the bone marrow to the thymus through circulation. The thymus provides an environment that comprises non-lymphoid cells, chemokines, cytokines, extracellular matrix elements, and other soluble proteins that support the commitment of progenitors to the T-cell lineage, their proliferation, T-cell receptor gene (TCR) rearrangement and the development of thymocytes into mature T-cells. TCR rearrangement is an important process that occurs during the development of T-cells in the thymus, which allows T-cells to recognize specific antigens. In the thymus, the early lymphoid progenitors lack T-cell receptors (TCRs) and undergo phenotypically distinct stages with the expression of specific surface markers and TCR variations (Germain, 2002).

For T-cell development and cell fate determination, somatic assembly of TCR genes (*Tcra, Tcrb, Tcrg, Tcrd*) is critical. An assembly of these four TCR genes is controlled by the lymphoid-specific recombinase complex RAG, consisting of RAG1 and RAG2. TCR coding sequences are assembled by the RAG1/RAG2 complexes that catalyse the recombination of variable (V), diversity (D), and joining (J) segments. As a result of RAG enzyme activity, double-strand breaks are created between recombination signal sequences flanking the V(D)J gene segments, and nonhomologous end joining is performed to promote assembly of αβ or γδ TCRs. The recombination of the TCR gene instructs the thymocyte to be committed to either the αβ or γδ lineages, which have different characteristics and functions. (Bassing et al., 2002, Hayday and Pennington, 2007). The majority of T-cells are of the αβ lineage, they express TCRs with α and β chains. They can detect antigen peptides presented by major histocompatibility complex (MHC) molecules on other cells' surfaces. While the γδ T-cell lineage express TCRs consisting of γ and δ chains, they are capable of recognizing a wider

range of antigens such as non-peptide molecules, on the surface of cells in a non-MHC-dependent manner (Attaf et al., 2015).

Tcrd, Tcrg, and Tcrb are recombined at the stage when most T-cell identity genes are active and RAG enzymes are activated (Mingueneau et al., 2013). The recombination of *Tcrd* and *Tcrg* successfully leads to the assembly of a TCRγδ, while recombination of *Tcrb* result in assembly of TCRβ (Krangel, 2009). Thus, functional and effective TCR facilitates differentiation and positive selection during T-cell development, whereas the unsuccessful expression of either TCRγδ or TCRβ chain, will result in cellular apoptosis (Taghon et al., 2006).

T-cells develop by passing through a series of developmental stages. Differentiation stages are classified by the expression of several important membrane molecules. In mice, there are four double-negative stages (DN1–DN4), meaning that the cells lack the expression of the two glycoprotein markers that determine T-cell type (CD4 for T helper cells and CD8 for cytotoxic T-cells). This is followed by the CD8 intermediate single-positive (ISP), then CD4/CD8 double-positive stage, and finally functionally mature CD4 or CD8 single-positive (SP) cells (Koch and Radtke, 2011). The DN stages are categorised through the expression of different surface markers (CD44 and CD25). The pre-thymic progenitors migrate from the bone marrow to the thymus and at the corticomedullary junction and become early thymic progenitor cells (ETPs) (Figure 1.3). In the first double-negative stage (ETP/DN1), in which the cells are immature T-cell progenitors in the thymus, the transition from DN1 to DN2 is characterised by the expression of CD44, DN2a cells with high KIT expression transform into DN2b cells expressing low KIT, indicating the initiation of T-cell commitment (Yui et al., 2010). This is followed by the expression of CD25 in stages DN2 and DN3. The DN3 stage is divided into DN3a and DN3b according to β-selection. β-selection is a checkpoint that allow cells to verify the successful rearrangement of their TCR-β chain locus, a successful TCRβ expression

is required for proliferation, while an unsuccessful TCRβ expression results in apoptosis (Famili et al., 2017, Carpenter and Bosselut, 2010). Cells containing functional TCRβ express the pre-TCR, which consists of TCRβ and pre-TCRα, and proceed to the double-positive stage (CD4$^+$CD8$^+$)(Jones and Zhuang, 2007).

In the double-positive stage, thymocytes go through TCRα rearrangement, resulting in the production of mature αβ TCR. Double-positive cells that express effective TCR recognise antigens in the context of major histocompatibility complex and receive positive selection signals. These cells then differentiate into single-positive cells, expressing either CD4$^+$ or CD8$^+$. Cells that fail to express effective TCR will die through a process called death-by-neglect, which is a default death pathway for non-functional cells with the inability to bind to MHC (Szondy et al., 2012).

T-cell differentiation stages in humans are similar to those in mice, but they are classiefied according to different surface markers such as CD34, CD7, CD5 and CD1a (Figure 1.4) (Famili et al., 2017). The pre-thymic progenitors express the marker of development CD34, which can further differentiate into multiple lineages. Starting with the DN stages, the early T-cell progenitors (ETPs) are CD34$^+$ CD38^{low}, which is equivalent to the DN1 in mice. The next stage, called the Pro-T1, is characterised by the expression of CD7$^+$ CD5$^-$ CD1a$^-$. This is followed by the Pro-T2 stage that expresses CD7$^+$ CD5$^+$ CD1a$^-$, and the last DN stage is the pre-T stage, which is the expression of the surface marker CD1a. This is the T-cell commitment marker, where the ability of cells to differentiate into other lineages is lost (Capone et al., 1998). Following is a CD4$^+$ ISP stage (mouse ISPs are CD8$^+$) that will further differentiate into CD4$^+$ CD8$^+$ double-positive cells and then to functional CD4$^+$ or CD8$^+$ cells (Awong et al., 2009).
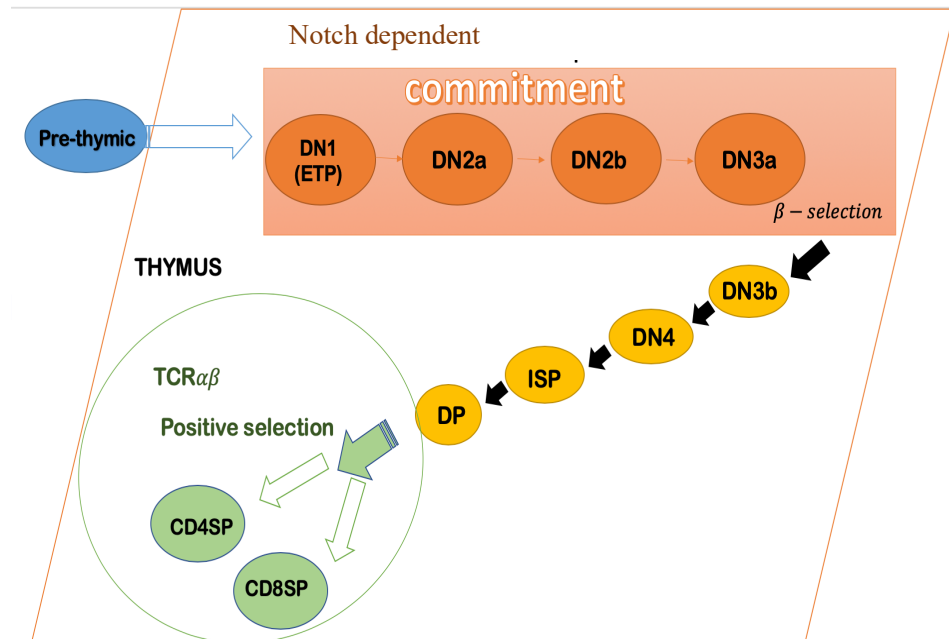
Figure 1.3: Illustration of mice T-cell developmental stages in the Thymus.

Early progenitor cells travel from the bone marrow (pre-thymic) and enter the thymus as DN1 cells. DN2a stages show Notch dependence and T-cell commitment starts at the DN2b stage followed by DN3a stage, where cells undergo β-selection. Cells that arranged their TCR-β successfully differentiate into DN3b, followed by the DN4, ISP and DP stages. The double positive stage undergoes positive selection through the TCR-αβ, giving rise to CD4 or CD8 single positive cells. Created using Biorender.com.
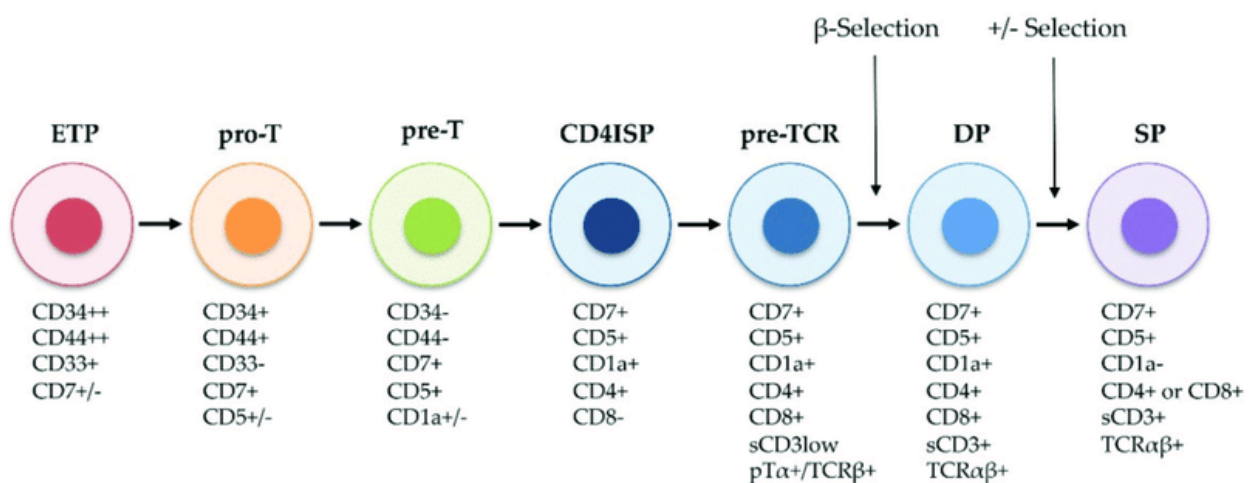
Figure 1.4: Illustration of T-cell differentiation stages in human.

Upon entering the thymus, early thymic precursors (ETP) undergo sequential developmental stages. Using CD7, CD5 and CD1a as surface markers, CD4 and CD8 double negative stages are further classified. The ETP differentiates into CD7$^+$, CD5$^-$, and CD1a$^-$ (pro-T). There are then CD7$^+$, CD5$^+$, and CD1a$^-$ cells (Pre-T). Intermediate single positive (ISP) cells in humans express CD4, which differentiate into CD4/CD8 double positive cells and eventually become CD4 or CD8 single positive cells. Taken from (Bayón-Calderón et al., 2020).

1.2.1 Transcriptional regulation of T-cell development

The cellular identity of cells is defined by their gene expression patterns and therefore by the underlying networks of transcription factor activity. The development of T-cells, through the above-mentioned transition stages that generate functional T-cells from lymphoid primed progenitors, is characterised by rapid coordinated changes in the activation and silencing of transcription factor genes in response to the changing environment they are in. Lineage commitment starts at the DN1 stage, which has a specific molecular signature in which it expresses *Gata3* and *Spi1*. Overexpression of these transcription factors leads to a block in T-cell development (Anderson et al., 2002a) and an increase in myeloid (Anderson et al., 2002b) and/or mast cell development (Taghon et al., 2007). At the DN1 stage Notch signalling is activated, which is required for T-cell lineage development and inhibition of the B-cell lineage. Notch signaling is also involved in the upregulation of *Bcl11b*, reducing NK potential (Li et al., 2010). It was shown that *Bcl11b* deficiency and high expression of Inhibitor of DNA binding proteins (ID) factors enhance NK development (Heemskerk et al., 1997).

The transcriptional regulation in T-cell development has been classified into three overlapping stages, (Figure 1.5) (Yui and Rothenberg, 2014). Stage 1 involves the transcription factors that have a critical role in the proliferation and survival of HSCs. These are expressed in the pre-thymic stage and then downregulated. *Meis1*, *Gata2*, and *Hoxa9* are downregulated before the transition to DN2a (Gwin et al., 2013). *Lmo2* and *Mef2c* remain expressed until the DN2a phase and are then suppressed. Transcription factor genes *Spi1, Hhex, Gfi1b, Erg, Tal1, Lyl1,* and *N-Myc* continue to be expressed until reaching the DN2b stage and are then inhibited. Upon reaching the DN3 stage, cells lose their potential to become natural killer cells, dendritic cells, B-cells, or myeloid cells. To support this lineage commitment, stage 2 of transcriptional regulation sees increased expression of key transcription factors that identify T-cell lineage, such as *Gata3*, *Bcl11b*, and *Tcf7*. Their expression helps the cells to progress to the DN3 stage,

where the cells undergo TCR rearrangement (Zhang et al., 2012b). The activation of NOTCH signalling at this stage and the interaction of its ligands with NOTCH receptors on the surface of lymphoid precursor ensures the sequential activation of T-cell transcription factors and promote T-cell development and differentiation (Koch et al., 2008). GATA3, BCL11B and TCF1 in stage 2 play an essential role in inhibiting the expression of many stem and progenitor-cell genes from stage 1. This is required for normal T-cell development, because their abnormal expression could lead to malignant transformation of T-cells. Also, the inhibition of the class II helix–loop–helix (HLH) transcription factors, *Tal1*, and *Lyl1* in stage 1 is crucial for the activation of T-cell-specific genes. The overexpression of *Tal1* induces the sequestration of E-proteins which form the TAL1/E-protein heterodimer. This heterodimer indirectly inhibits the transcription of genes regulated by E-protein dimers. The silencing of *Tal1* and/or *Lyl1* is therefore critical for the formation of E-protein dimers (Tan et al., 2019). E-protein transcription factors are known to regulate T-cell transition processes and important for TCRβ thymocytes. E-protein dimers regulate *Rag1* and *Rag2*, facilitate the integrity of β-selection checkpoint and regulate the thymocytes that passed the β-checkpoint (Zhong et al., 2007, Yui et al., 2010, Rothenberg, 2014)

The third stage involves gene expression following the β-selection. T-cells undergo rapid proliferation during the DN3b to DP transition phases in response to TCR signals and ultimately lose their dependence upon the Notch receptor. In the absence of Notch signalling, thymocytes rely on pre-TCR signals that inhibit NOTCH-activated genes by the IKAROS transcription factor (Geimer Le Lay et al., 2014, Yui and Rothenberg, 2014). In response to pre-TCR signalling, *Id3* expression is transiently turned on, temporarily inhibiting E-protein activity, but the E-proteins then resume their function (Engel et al., 2001). DN3 cells that have not completed V(D)J recombination while being selected may be inhibited by high levels of constitutively expressed Gfi1, considering an important role of Gfi1 in this process (Schmidt

et al., 1998). Furthermore, RUNX1 has been demonstrated to play a role in regulating the transition from the DN to the DP stage of developing thymocytes, as well as the transition from DP to SP (Egawa et al., 2007). The gene expression of *Ets1/2*, *Tcf7*, *Lef1*, *Gfi1* and *Tcf12* continues from stage 2 and increases in stage 3 for efficient T-cell thymocyte production. A precise control of regulatory networks and the inhibition of Notch target genes during and post β-selection is essential for maintaining healthy T-cells, as well as protecting against potentially malignant transformations (Yui and Rothenberg, 2014).
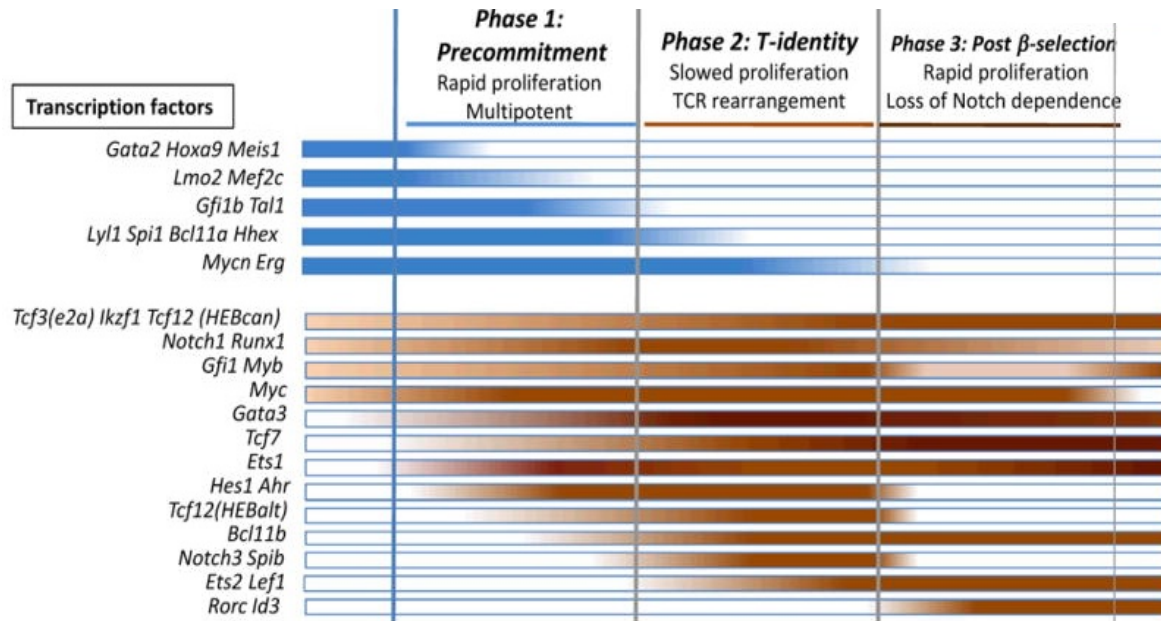
Figure 1.5: Illustration of the pattern of important transcription factors expression levels during T- cell development.

The figure shows the three phases of transcriptional regulation during T-cell development. The blue bar represents the genes of proteins involved in the transcriptional regulation of stem cells and progenitors. A brown bar represents the expression of the T-cell specification and commitment factors. Intensity of colour indicates dynamic changes in gene expression (dark = high, light = low) (Yui and Rothenberg, 2014).

1.3 Helix-loop-helix E-box binding factor - HEB

The E-protein transcription factors belong to the basic helix-loop helix (bHLH) family, a large family of proteins characterised by two distinct conserved domains: a basic region and the HLH structure. The basic region located at the N-terminus, consists of mostly basic residues and is essential for binding of the transcription factor to DNA at the specific hexanucleotide sequence 'CANNTG', referred as an E-box motif. The HLH domain is at the C-terminal end and comprises mainly hydrophobic amino acid residues and contains two α-helical regions connected by a loop. The HLH region is important for the interaction between proteins and the formation of homo- and heterodimers (Jones, 2004, Murre et al., 1994). E-proteins are involved in a variety of developmental processes such as myogenesis, neurogenesis, heart, pancreatic development and lymphopoiesis (Massari and Murre, 2000).

The HLH family is classified into different groups based on dimerization capabilities and DNA-binding specificity. Class I includes members of the E-protein family; E2A (*TCF3*), E2-2 (*TCF4*), and HEB (gene name: *TCF12*) (Belle and Zhuang, 2014). These regulate transcription by forming homodimers or heterodimers with other E-proteins or with class II bHLH proteins. Class II includes proteins such as myodenin, MyoD, TAL1, and LYL1 and require heterodimerization with class I E-proteins to function as transcriptional activators or repressors (Murre c, et al1994). E-proteins can also form as heterodimer with Inhibitor of DNA binding proteins (ID) to prevent DNA binding by the E-protein dimers. ID proteins are HLH proteins lacking the DNA binding activity (Benezra et al., 1990). The activities of E-proteins and their antagonist ID proteins play a critical role in haematopoiesis as they orchestrate developmental progression and thymocyte maturation (Lazorchak et al., 2005).

HEB was first described in 1992 (Hu et al., 1992). *TCF12* contains 21 exons, with the total gene spanning more than 200 kb (Wang et al., 2006a). It is essential in many developmental processes. It can promote myogenesis by forming a heterodimer with MyoD, which is a

myogenic regulatory factor, and binds in the promoter regions of muscle-specific genes at E-box sequences. HEB plays a fundamental role in controlling thymocyte transition to the DP stage and beyond and controls thymocytes survival. HEB is important in the activation of the recombinase genes Rag1 and Rag2, which mediate TCR-β, TCR-γ and TCR-δ gene segment rearrangements (Langerak et al., 2001). It is also involved in the regulation of pTα expression, which is essential for efficient selection and proliferation of immature thymocytes (Tremblay et al., 2003).

HEB is expressed in the form of two major isoforms, due to differential transcription start site usage; HEB canonical (HEBcan), which starts at exon 1 and HEB alternative (HEBalt), which starts between exons 8 and 9 (Figure 1.6A). There are different domain compositions in these isoforms (Figure 1.6B). Both consists of the HLH domain that is required for DNA binding and dimerization with other bHLH proteins. The activation domains 1 (AD1) and AD2 mediate coactivators recruitment to the transcriptional complex is found in canonical isoform. However, AD1 activation domain is absent from the alternative. HEB isoforms are expressed differently during T-cell development stages. HEBalt is highly expressed in early T-cell progenitors and then gradually inhibited towards the double-positive transition stage. In contrast, HEBcan is expressed throughout the stages and is upregulated in the DP stage (Figure 1.6C) (Wang et al., 2006a). Unique roles have been played by the two isoforms: HEBcan was shown to inhibit natural killer cell potential, while HEBalt did not show this effect (Braunstein and Anderson, 2011). HEBalt restricts myeloid fate during haematopoiesis (Wang et al., 2006b).
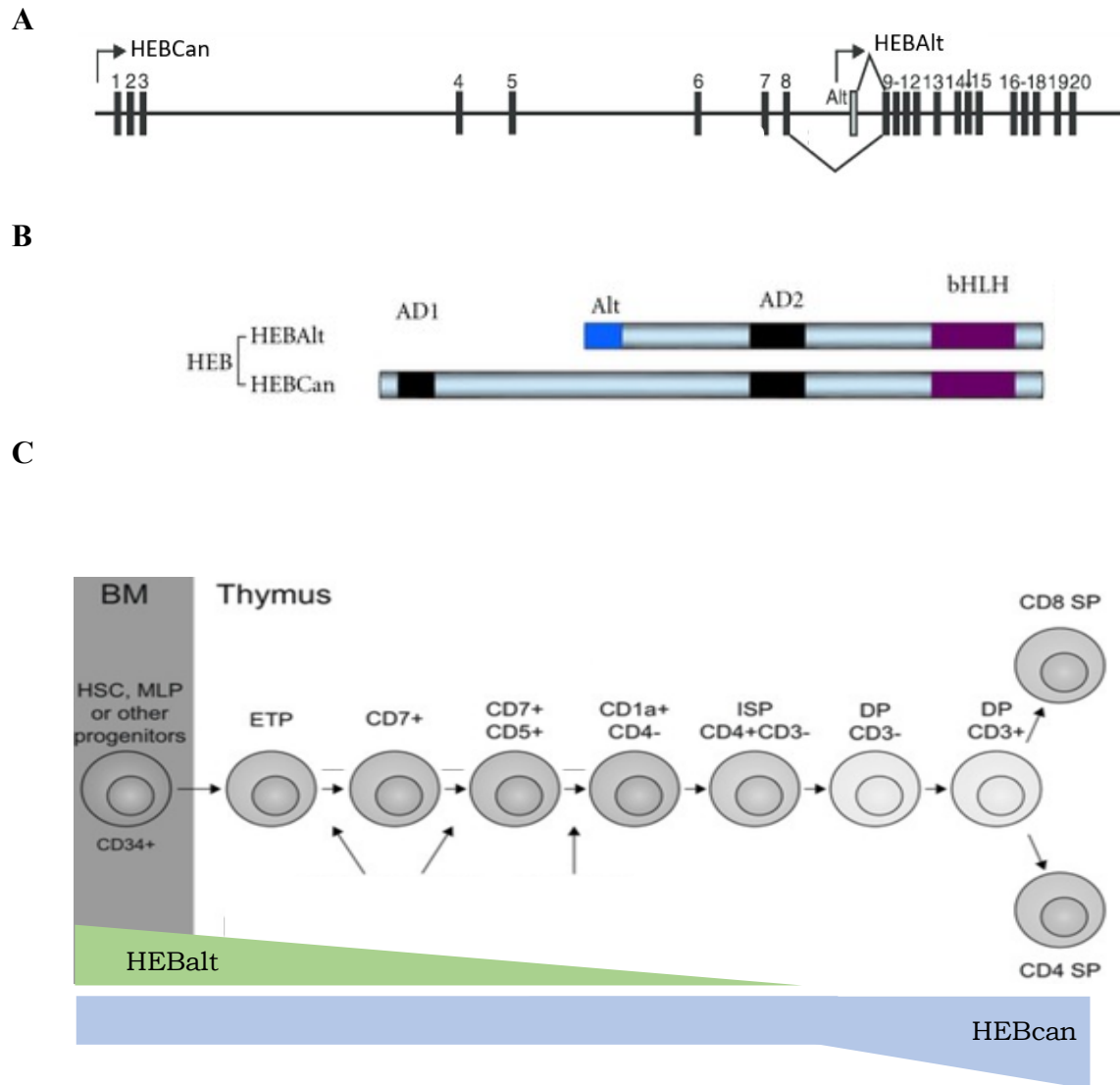
Figure 1.6: Schematic representation of HEB genomic structure and expression level.

(a) Genomic organization of *TCF12*, arrows indicate the two alternative transcription start sites, resulting in HEBcan and HEBalt expression, vertical bars represent exons. (b) The structural differences between the two HEB isoforms. The figure was adapted from (Braunstein and Anderson, 2012) and modified using Biorender.com. (c) Dynamic expression of HEB isoforms during T-cell development stages. Created using Biorender.com.

1.3.1 HEB in Haematopoiesis

There is considerable evidence supporting an important role of HEB in haematopoiesis, both in lineage commitment and differentiation. Tcf12 expression level is found to be more abundant in HSCs as compared to LMPP and CLP. As such, Tcf12 is believed to have a role in regulating haematopoiesis. The result of targeted deletion of Tcf12 in HSCs showed that their reconstitution capacity is significantly reduced in comparison to wild-type (WT) HSCs. A significant decrease in white blood cell and lymphocyte numbers was observed in the peripheral blood of $Tcf12^{-/-}$ mice compared with WT mice. In the meantime, neutrophil, red blood cell, and platelet counts remained unchanged. Also, the percentage of B cells and CD4$^+$ cells declined significantly, but the percentage of myeloid and CD8$^+$ cells did not change in the peripheral blood of $Tcf12^{-/-}$ mice. In the thymus of $Tcf12^{-/-}$ mice, CD4$^+$ and double-positive T-cell frequencies and absolute numbers decreased significantly while CD8$^+$ and double-negative T-cells increased. The results suggest that there are two effects: a block in differentiation, and a shift between CD4-positive and CD8-positive cells (Liao and Wang, 2021).

Moreover, the lineage cells in the spleen of $Tcf12^{-/-}$ mice showed a decrease in B cells and CD4$^+$ cells, suggesting that *Tcf12* protein is needed for the development of these cells. At the same time, the number of myeloid, CD8$^+$ cells, and DP cells remained relatively unchanged. The effect originates primarily from a defect in B and T cell differentiation, rather than myeloid differentiation. Also, ablation of *Tcf12* led to a significant reduction in the expression of differentiation genes such as *Runx1, Klf5, Cebpα, and Cebpe* in HSCs, indicating a role of *Tcf12* in regulating HSCs differentiation. However, there was no difference in the HSC self-renewal gene expression, including *Hoxa5, Hoxa7, Hoxa9, Hoxa10 and Mn1* (Liao and Wang, 2021). These last two observations may also explain the expansion of early T-cell progenitors in the absence of HEB; persistent self-renewal with reduced differentiation.

Recent research has revealed that HEB has an indispensable effect on human mesodermal specification and haematopoiesis during embryogenesis. Using CRISPR/Cas9 gene-editing technology, human embryonic stem cells (hESCs) were transfected with a plasmid that disrupted both HEBalt and HEBcan expression, by targeting exon 9 of the gene locus. The results of this study revealed a profound defect in mesodermal development, failure of T cell development as well as a decrease in the expression of *RUNX1* and *NOTCH1*, two key regulators of the generation of hemogenic endothelium. Ectopic expression of HEBcan largely corrected these defects. Additionally, analysis of RNA-seq data showed a disruption of the expression of genes required for the development of heart, neural crest and endothelial development, which confirmed that HEB is important in the differentiation of all these lineages (Li et al., 2017).

Moving on to the interaction between the E-proteins during haematopoiesis, E2A has been found to be essential in B-cell lineage development and differentiation. The E2A gene encodes alternative splice variants, E12 and E47, which vary in their basic DNA binding region. HEB, in co-operation with E2-2, can help to modulate E2A protein activity in B-cell lineage development via two potential mechanisms: 1) HEB and E2-2 dimerize with the inhibitory ID proteins that inhibit both HEB and E2-2, resulting in increasing E2A protein availability, 2) HEB and E2-2 heterodimerize with E2A proteins and contribute directly in B-cell-specific gene transcription (Zhuang et al., 1996, Hu et al., 1992). Overall, these studies provide an important insight into the role of E-proteins in the regulation of a wide range of genes involved in various aspects of cellular development and haematopoiesis.

1.4 TAL1/LMO2 multiprotein complex

T-cell Acute Lymphoblastic Leukaemia protein 1 (TAL1, previously SCL) belongs to the class II basic helix–loop–helix (bHLH) transcription factors. It has an essential function in the generation of haematopoietic stem cells but is not required for their maintenance (O'Neil et al.,

2004). TAL1 is highly expressed in HSCs and early multipotent progenitors and is then suppressed during haematopoietic differentiation in all but the erythroid lineage (Kallianpur et al., 1994). TAL1 forms a heterodimer with E-proteins (E2A, E2-2 or HEB) to bind to DNA and activate transcription as part of a transcription complex that consists of LIM-only protein 2 (LMO2), Lim-domain-binding protein 1 (LDB1) and GATA family proteins (Hsu et al., 1994 Murre, 2000). This multiprotein complex functions in the regulation of gene expression of T-cell development and differentiation. LMO2 expression is largely overlapping with that of TAL1 and is supressed at the DN2 stage, the beginning of T-cell commitment.

LMO2 is a transcriptional regulator important for normal haematopoietic development and is required for HSCs differentiation to committed to erythrocyte lineage (Sincennes et al., 2016). LMO2 belongs to the LIM-only class of the LIM zinc finger protein family, it is composed of two LIM domains. For many biological processes, protein-protein interactions are critical for the formation of complexes with other proteins. LIM domains do not bind DNA, but rather serve as scaffolds for the assembly of protein complexes. It has consistently been demonstrated to play a significant role in regulating gene expression via transcriptional regulation (Zheng and Zhao, 2007). The LMO2 complex, through multimerization of LDB1, was found to facilitate long-range interactions between enhancers and promoters, for example in the β-globin locus (Meier et al., 2006).

As a result of a study using LMO2 knockout mice, it was found that LMO2 is necessary for erythropoiesis to occur in the yolk sac, and mice lacking LMO2 died at E9-10 (Warren, et al. 1995). In 1997 Wadman *et al.* demonstrated that LMO2 regulates transcription by forming a complex with GATA-1, TAL1 (Scl) and LDB1. Since then, other studies have confirmed that this transcription factor complex is essential to the development of erythropoiesis, which explains why knockout mice did not survive past early embryonic development when LMO2 was deleted (Kerenyi and Orkin, 2010).

Overexpression of LMO proteins (LMO1 and LMO2) was found with aberrant TAL1 expression in T-cell Acute Lymphoblastic Leukaemia (T-ALL) (Ferrando et al., 2002a). More than 70% of T-ALL cases that are due to a chromosomal translocation involve a master transcriptional regulator (Sincennes et al., 2016). TAL1 and LMO2 were found aberrantly expressed in T-ALL as a result of chromosomal translocation involving their loci at 1p32 and 11p13 respectively (Wadman et al., 1997). TAL1 is commonly overexpressed due to the SIL-TAL deletion, which is an approximately 90 kb deletion from the 5'UTR of the SCL-interrupting locus (SIL) gene to the 5'UTR of the TAL1 gene, forming the *SIL/TAL1* fusion gene. The cell cycle regulator, SIL gene is expressed throughout the development of T cells, consequently, SIL/TAL1 fusion activates TAL1 expression even when TAL1 is supposed to be downregulated leading to constitutive expression of TAL1 (Delabesse et al., 1998).

1.5 T-cell Acute Lymphoblastic Leukaemia

Cancer is a class of diseases that can be defined as the abnormal and uncontrolled proliferation of cells which exceed their normal boundaries, invading surrounding tissues and spreading to distant tissues and organs (Roy and Saikia, 2016). Leukaemia is a type of cancer that forms in the tissues responsible for the production of blood cells, which includes bone marrow and the lymphatic system. One type of leukaemia, Acute Lymphoblastic Leukaemia (ALL), involves the expansion and malignant transformation of lymphoid progenitor cells in bone marrow and peripheral blood (Szychot et al., 2014). ALL is a frequently observed type of leukaemia; it is the second most common cancer among children under the age of one, although it can also occur in adults, and has a high mortality rate (Ibagy et al., 2013). It is classified into two broad types, known as B cell-lineage and T cell-lineage ALL. B-ALL shows heterogeneity with 7 recurrent genetic abnormalities, based on the 2008 World Health Organization classification. T-ALL on the other hand, is often the result of chromosomal rearrangements and aberrant expression of transcription factors (Montaño et al., 2018b).

T-ALL represents approximately 10-15% of paediatric cases and 25% of adult cases of ALL. With a diagnosis of T-ALL, patients show widespread bone marrow infiltration by immature T-cell lymphocytes, increased white blood cells counts, mediastinal masses of thymus, pleural effusions and central nervous system infiltration (Van Vlierberghe and Ferrando, 2012). T-ALL diagnosis has gradually improved as a result of intensified treatment protocol and chemotherapy, leading to a 5-year cure rate without relapse in over 70% of children and over 50% of adults. However, relapse cases, which are resistant to chemotherapeutics and tend to have a poor prognosis, occurs in up to 20% of children and 40% of adult (Girardi et al., 2017). Despite the high amount of research and clinical studies, T-ALL is considered to be a clinical challenge due to its high relapse rates, drug resistance and poor initial prognosis. Therefore,

further research on T-ALL is necessary to better understand the underlying molecular causes and to evolve therapeutic approaches.

Cytogenetic alteration of healthy T-cell progenitors can result in T-ALL. T-ALL development is a multi-steps process; the malignant transformation occurs via the accumulation of several genetic and epigenetic abnormalities, causing defects in cellular proliferation, survival, cell cycle and differentiation of T cells. Several studies indicated that more than 70% of T-ALL cases are associated with chromosomal translocation and aberrant expression of transcription factors that are the key regulators responsible for normal control of T-cell proliferation, differentiation, and survival (Belver and Ferrando, 2016). The first genetic changes observed in T-ALL patients were chromosome aberrations, discovered by Kowalcyzk *et al.* in 1983. T-ALL is frequently caused by a spectrum of chromosomal aberrations such as translocations, deletions, or mutations (Kowalczyk and Sandberg, 1983). Van Vlierberghe *et al.* classified genetic abnormalities into two types. Type A abnormalities include abnormal expression of oncogenes encoding transcription factors that are crucial in T-cell development and differentiation. Examples of these aberrantly expressed genes are *TAL1, LMO1/2, HOXA, TLX1, TLX2 and MEF2C.* Type A abnormalities may be mutually exclusive and correlate to specific molecular-cytogenetic T-ALL subgroups. Type B abnormalities consist of genes that are involved in cellular processes such as signalling pathways, self-renewal, or aberrant activation of tyrosine kinases, such as *NOTCH1, CDKN2A/B and ETV6* (Van Vlierberghe et al., 2011).

1.5.1 Classification of T-ALL

Analysis of gene expression and immunophenotype profiles classified T-ALL into subgroups. Early T-lineage progenitor (ETP) T-ALL is the first group, which is characterised by a block at the early stage of T-cell differentiation and exhibit immunophenotypes of CD4$^-$ and CD8$^-$. ETP represent approximately 10% of T-ALL paediatric cases and 50% of adult cases and is

associated with poor prognosis (Coustan-Smith et al., 2009, Van Vlierberghe et al., 2011). In contrast, the second group show more favourable prognosis which is the early cortical subgroup with phenotype of CD1a$^+$ CD4$^+$ CD8$^+$ (Niehues et al., 1999). The third group with a relatively more mature immunophenotype of CD4$^+$ CD8$^+$ CD3$^+$, is the late cortical subgroup (Ferrando et al., 2002b).

Homminga *et al.*, classified T-ALL cases according to gene expression. As shown in Figure 1. 7, these data were hierarchically clustered based on microarray gene expression data of paediatric T-ALL samples. The major clusters were associated with elevated expression of TAL and LMO proteins due to genetic abnormalities, while TLX/HOXA transcription factors were the predominant transcription factors in the second largest cluster. TLX1-translocated cases with high CD1 expression were found in the third cluster, referred to as the "proliferative cluster." Among the fourth cluster was the so-called "immature cluster", consisting of immunophenotypically immature CD4/CD8 double negative cases expressing early progenitor genes (Homminga et al., 2011).
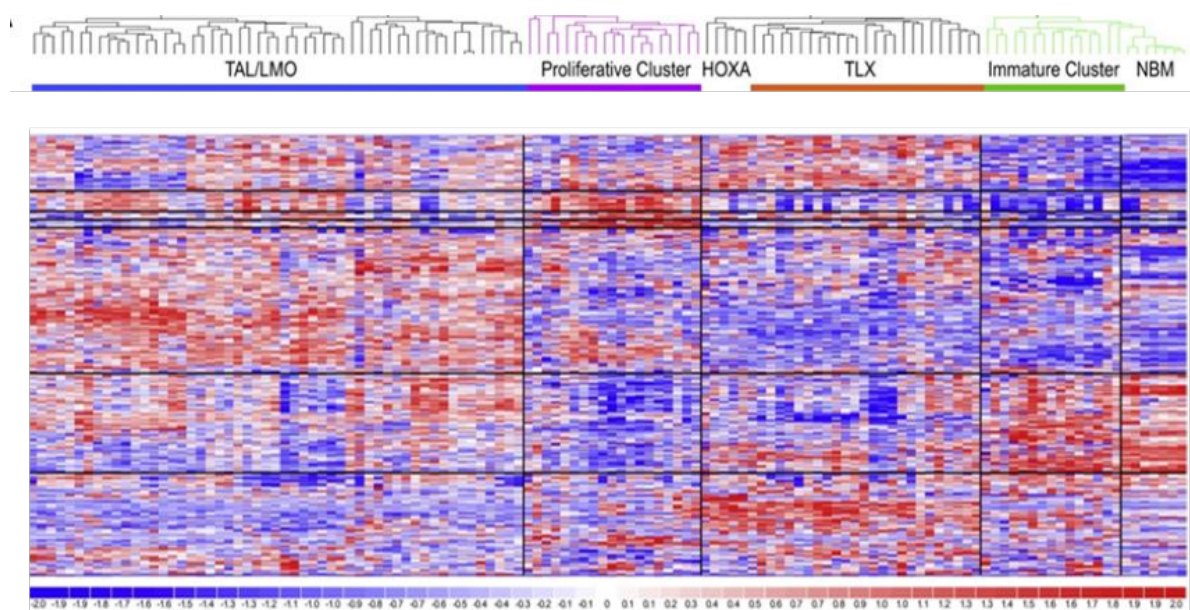
Figure 1.7: The genetic subgroups of T-ALL identified based on cluster analyses of patient samples.

Four groups were identified based on a hierarchical cluster analysis of 117 diagnostic T-ALL samples and seven controls of normal bone marrow (NBM). Taken from Homminga et al., 2011.

1.5.2   HEB in T-cell Acute lymphoblastic leukaemia

HEB knockout/knockdown studies were found to display greater defects in T-cell development compared to knockout/knockdown studies of other E-protein genes (Barndt et al., 2000). HEB deficiency in mice resulted in an accumulation of Intermediate single-positive (ISP) cells along with approximate 5–10-fold total thymocyte reduction (Barndt et al., 1999). In contrast, knockout of *TCF3* (E2A) and *TCF4* (E2-2) showed a partial block at the DN1 stage with normal thymocyte development (Barndt et al., 2000). HEB and E2A play an important role in thymocyte development as they control the transition from DP to single-positive (SP) stage. E2A is downregulated upon transition from the DP to the SP stage and this downregulation promotes thymocytes maturation. During positive selection, TCR signalling induces the expression of ID3 to decrease E2A activity and progress to the SP stage, whereas HEB is required to be expressed during the transition stage. E2A-deficient mice showed excess maturation from DP to SP, whereas ID3-deficient mice showed a reduction in the transition from DP to SP (Bain et al., 1999). Deletion of HEB and E2A simultaneously in the T-cell lineage clearly affected the balance between CD4 and CD8 populations. The results showed a reduction of CD4 SP cells while CD8 SP cells increased, while the overall thymic cellularity remained the same. As a consequence of the deletion of HEB or E2A, numerous $CD8^+$ cells lacked surface TCR expression in the periphery, showing that HEB and E2A prevent the cells from further development until there is an effective TCR-positive selection signal to transit to SP (Jones and Zhuang, 2007).

In early T-cell development, HEB functions to control levels of GATA3. By limiting GATA3 and IL-7Rα expression, HEB restricts thymic NK development (Vosshenrich et al., 2006). HEB deficiency leads to the inhibition of T-cell development at the β-selection checkpoint (DN3) (Braunstein and Anderson, 2010). Deficiency of HEB resulted in the generation of cells that had phenotypic characteristics similar to those found in DN1, in the DN3 stage. The

conversion of DN3 to a DN1-like precursor resulted in compromised NOTCH1 function and thereby loss of T-cell potential, which was replaced by intra-thymic B-cell development (Pui et al., 1999). It was shown that the DN1-like thymocytes differed from the WT DN1 thymocytes, as the DN1-like precursors did not have the potential to develop into different lineages. By comparing the mRNA levels of HEB knockout precursors with that in wild type DN1 thymocytes, B-cells, myeloid cells, and NK, it was found that the levels of *Pax5, Spi1,* and *SpiB,* which enable the cells to differentiate into B-cells, decreased in DN1-like precursors. *E2A, Id2,* and *Gata3* levels were similar in DN1-like to DN1 thymocyte. In contrast, *Notch1* and *Bcl11b* expression in DN1-like were reduced in contrast to DN1 thymocytes, which promotes the development of NK cell (Braunstein and Anderson, 2011).

TAL1 acts as an activator and/or repressor of target genes by binding to gene regulatory regions (Palomero et al., 2006). One study examined the overexpression of *Tal1* and deficiency of *E2A* and *HEB* in a transgenic mouse model and reported a 50% reduction of thymocyte cellularity, a decrease in the number of DP thymocytes that led to reduced production of CD4 SP thymocytes, and an increase in immature CD8 thymocyte production. Activation of *Tal1* perturbs the development of thymocytes, induces and accelerates leukaemia by interfering with E47 and HEB*,* and represses the *E47/HEB* target genes. Ectopic expressing of *Tal1* in mouse cells deficient in E2A and/or HEB expression (*Tal1/E2A$^{+/-}$ and Tal1/HEB$^{+/-}$)* revealed a reduction of the expression of genes essential for thymocyte development, such as *CD5*, *Rag1* and *Rag2*, which are regulated by the *E47/HEB* heterodimer, comparing to transgenic mouse cells expressing *Tal1* on a wild type background. CD4 expression was decreased in *Tal1* cells and demonstrated a further decrease in cells expressing *Tal1* and deficient in *HEB* and/or *E2A*. These results indicate that *Tal1* contributes to the leukaemia by direct repression of *E47/HEB* target gene expression (O'Neil et al., 2004).

1.6 The zinc finger transcription factors GFI1 and GFI1B

Growth factor independence 1 and 1B (GFI1 and GFI1B) are zinc-finger containing transcriptional repressors. They share a similar structure consisting of three identifiable domains; the N terminus domain contains a small conserved SNAG (Snail/GFI1) domain, that is important for its repressor activity via recruitment of histone modifiers (Figure 1.8). The C-terminus domain contains of six C_2H_2-type zinc fingers essential for DNA-binding – notably zinc fingers 3–5. Finally, the intermediate domain, located between the SNAG and zinc finger domains, functions by binding proteins that are involved in transcriptional regulation, mRNA splicing and protein modifications (Möröy, 2005, van der Meer et al., 2010). The intermediate domain distinguishes the GFI1 from the GFI1B protein based on the difference in size between the two proteins: GFI1 consists of 422 amino acids, while GFI1B comprises 330 amino acids (Anguita et al., 2017), generating a mass of 55 kDa and 37 kDa for GFI1 and GFI1B respectively.
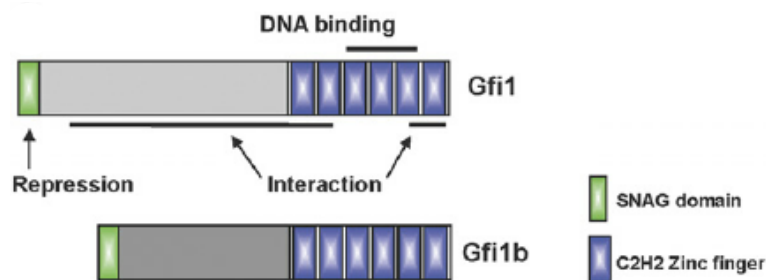
Figure 1.8: Structure of Gfi1 and Gfi1b.

The figure illustrates the domain structure of two proteins, Gfi1 and Gfi1b. As indicated by the blue boxes (the zinc finger) both proteins contain regions designated for DNA binding, the C-terminus domain. The green box at one end is called the "SNAG domain," the N-terminus domain, which has a specific role in repression, as indicated by the adjacent arrow. An intermediate domain is indicated by the grey box. This is the location where the protein interacts with other molecules, taken from (Möröy et al., 2015).

GFI1 and GFI1b have essential roles in haematopoietic stem cell self-renewal, proliferation and differentiation. *Gfi1b* is highly expressed in early HSCs and then downregulated at the multipotent progenitor stage, and gradually further downregulated during B and T cell differentiation. In contrast, *Gfi1* expression level is upregulated at these stages. It is required for early T-cell progenitors transition stages, where it is(Duan et al., 2005) expressed at the DN1 and DN2 stage and then increases to high expression at DN3, where it controls the beta-selection and encourages DP cell formation (van der Meer et al., 2010) (Möröy et al., 2015).

GFI1 modifies chromatin and controls target gene expression by recruiting histone modifying enzymes, such as histone lysine methyltransferase G9a, histone demethylase LSD1 and histone deacetylase 1 (HDAC1) (Duan et al., 2005). The oncogenic potential of GFI1, in part arises from its ability to recruit LSD1 (Lysine specific demethylase) to the p53 tumour suppressor protein (Khandanpour *et al.*, 2013). Subsequently, LSD1 de-methylates p53 at C-terminal lysines which in turn prevents p53 from activating pro-apoptotic proteins, such as Bax, Puma and Noxa, in response to tumorigenic mutations. Ultimately, GFI1 recruitment of LSD1 prevents p53 induced cell apoptosis in response to DNA damage, an important defence mechanism against tumour formation.

Several studies have shown a role for GFI1 in regulating oncogenesis. *Gfi1* mutation can give rise to human neutropenia (Person et al., 2003). Aberrant expression of *Gfi1* can result in lung tumours (Kazanjian et al., 2004), and *GFI1* may cooperate with oncoproteins such as C-MYC and PIM-1 in T-cell lymphomagenesis (Zörnig et al., 1996). Aberrantly expressed *Gfi1* also affected differentiation of healthy cells. *Gfi1* deficiency in myeloid progenitor cells prevented cell differentiation into mature granulocytes (Duan and Horwitz, 2003). T-cell development was partially arrested in mice lacking *Gfi1*, and there was a reduction of mature B-cells in these mice (Karsunky et al., 2002b). The overexpression of *Gfi1* in transgenic mice accelerated T-

cell entry into the S phase after antigenic simulation, whereas *Gfi1* deficiency led to a reduced T cell response to TCR stimulation (Karsunky et al., 2002a). In murine disease models, *Gfi1* deletion slowed down T-ALL progression, prevented malignant transformation in transplanted hosts, and moreover, conditional *Gfi1* knockout in established lymphomas was observed to attenuate the tumour and promote survival (Khandanpour and Möröy, 2013).

Since the 1990s, emerging evidence has implicated aberrant *Gfi1* expression in haematopoietic transformation, leading to the development of T-cell lymphoid leukaemia. The involvement of GFI1 in T-ALL progression was first identified in studies using Moloney murine leukaemia virus (MoMLV) induced models of leukaemia (Gilks et al., 1993). In this model viral insertional mutagenesis of tumour suppressor and proto-oncogenes drives oncogenesis. Accumulation of MoMLV proviral insertions in these models confers interleukin-2 independent growth to T cells (Lazo et al., 1990), a critical process in the oncogenic transformation of lymphoid progenitors. In these studies, *Gfi1* was identified as a prevalent proviral integration site.

However, it has become clear that aberrant *Gfi1* expression is not the major aetiological mechanism driving lymphoid progenitor transformation, as transgenic overexpression of GFI1 alone is insufficient for leukemogenesis (Zörnig *et al.*, 1996). Rather, the role of GFI1 in transformation appears to arise from its potential as an 'oncorequisite' factor that assists the progression and development of T-ALL through cooperation with other oncogenes. Indeed, MoMLV-upregulated *Gfi1* operates as a partner of constitutively expressed *Pim-1* or *Myc* oncogenes in the tumorigenesis of murine T-cell lymphoma modules (Zörnig *et al.*, 1996), highlighting the oncogenic potential of GFI1 and its ability to operate in concert with other oncogenes.

Critically, the role of *Gfi1* as an oncogenic progression factor has been demonstrated outside of viral integration-induced leukaemia. Indeed, elevated expression of *Gfi1* has been reported

in subsets of human NOTCH-driven T-ALL oncogenesis (Weng *et al.*, 2004). *Gfi1* is intricately involved in mutant Notch-dependent T-ALL, where it appears to act in concert with Notch signalling to initiate and maintain lymphoid leukaemia. *Gfi1* knockdown significantly inhibited Notch activity and disease progression (Phelan *et al.*, 2013). Furthermore, GFI1 appears to have a role in tumorigenesis and T-ALL progression, as maintained and upregulated *Gfi1* expression is required during the preleukemic phase, but the gene later becomes downregulated during progression to leukaemia (Phelan *et al.*, 2013).

A recent study identified the GFI1-interacting protein IKAROS, encoded by the IKZF1 gene, which is a zinc finger transcription factor (Sun W et al., 2022). IKAROS plays an important role in B and T cells development, and in T-ALL function as a tumour suppressor. Gene expression profiling revealed a previously unrecognized function of GFI1 in transcriptional activation of a subgroup of genes involved in T cell development. Ectopic expression of GFI1 and/or knockout of IKAROS resulted in positively transactivated genes such as Notch3, CD3 and RAG1 (Figure 1.9). They performed GFI1 and IKAROS ChIP-seq to investigate the potential common target genes, which revealed approximately 80% of GFI1-bound peaks overlap with IKAROS peaks. This result suggested that GFI1 and IKAROS regulate common proteins and that genes co-occupied by GFI1 and IKAROS have fundamental roles in T-cell development, such as GFI1 itself, NOTCH3, CD3, TCF3, MYC, RUNX3, MYB and HES (Sun W et al., 2022).
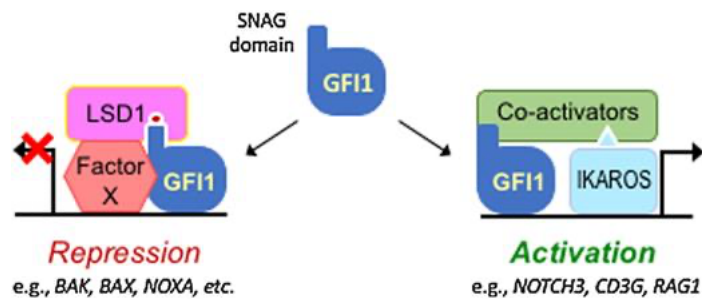
Figure 1.9: Schematic representation of GFI1 role.

GFI1 acts as transcriptional activator when cooperates with IKAROS and activates genes such as *NOTCH3, CD3* and *RAG1*, or acts as a transcriptional repressor by recruiting LSD1 via its suppressor domain SNAG (Sun W et al., 2022).

It has been shown that that GFI1 is involved in the regulation of alternative splicing of mRNA. This is a regulated post-transcriptional process in which exons are linked in different combinations from the same mRNA, leading to the generation of more than one functional protein isoform from the same gene (Black, 2003). In T-cells, GFI1 was shown to interact and bind with U2AF26, a small nuclear RNA auxiliary splice factor. As a consequence, this cooperation directly regulates alternative splicing of the transmembrane tyrosine phosphatase CD45. The activity of CD45, which is one of the important regulators of T-cell receptor (TCR)-mediated signalling and T-cell activation, is controlled by alternative splicing of its pre-mRNA. In the absence of GFI1, U2AF26 induces exon skipping in CD45 pre-mRNA, generating the less active CD45RO isoform, which negatively affects T cell activation. In contrast, U2AF26 with GFI1 facilitates the formation of the more active CD45RB isoform, associated with T-cell activation. In *Gfi1*-deficient mice, the uncontrolled activity of U2AF26 leads to defective splicing of CD45, especially in CD4$^+$ T-cells, thereby causing a decreased level of antigen-mediated activation of T-cells. This indicates that GFI1 plays an essential role in the mediation of alternative splicing of CD45 in T- cell and controls T-cell receptor signalling (Heyd et al., 2006).

Additionally, *Gfi1* deficiency was found to influence the expression of genes. It acts as an upstream negative regulator of the Id1 and Id2 proteins. As a result, ID proteins negatively

regulate the expression of HEB, which function as a critical effector in early T cell development and proliferation (Murre, 2005). Aberrant regulation of HEB expression has been implicated in T-ALL (O'Neil *et al.*, 2004). The oncogenic function of Gf*i1* in this regard is mediated by its loss-of-function, which inhibits its repressive function on its downstream targets, *Id1* and *Id2*, leading to elevated expression of these genes (Yücel et al., 2003). As mentioned above, deficits in expression of E-protein genes have been reported to drive accelerated T-cell selection and proliferation (Bain *et al.*, 1999; Yücel *et al.*, 2003). Therefore, it can be hypothesised that GFI1 exerts an additional oncogenic role through its ability to regulate HEB E-protein dosage via regulation of *Id1* and *Id2*.

To conclude, GFI1 displays dynamic and intricate roles as "oncorequisite" factor of T-ALL progression. Despite not providing the initiating force underlying lymphoid transformation, it clearly has cooperative functions by inhibiting apoptosis, and for sustaining oncogenic Notch signalling. Furthermore, GFI1 expression is not simply required in an upregulated state, but rather is modulated during disease progression from high to low expression levels, where modulation of downstream E-protein dosage is involved as well. Future work will be required to better define the regulatory mechanisms behind GFI1 expression, and to better characterise its role in lymphoid malignancy. Finally, GFI1 presents as a promising target for novel T-ALL treatments, which is currently being explored (Geimer Le Lay et al., 2014).

1.7 Genome editing and CRISPR-Cas9

The development of genome editing technologies has offered new possibilities for tackling diseases such as lymphoblastic leukaemia and to improved therapeutic potential of diseases. There are several genome editing tools that have been developed in the last decade. The most commonly used technologies are Zinc-Finger Nucleases (ZFNs), Transcription activator-like effector nucleases (TALENs) and Clustered regularly interspaced short palindromic repeats (CRISPR)(Gaj et al., 2016). Following is a brief explanation of how these three gene editing techniques differ in term of the nucleases used and the type of repair required.

The ZFNs editing tool involves the fusion of the zinc finger domain and the cleavage domain of the restriction enzyme Fokl from the bacterium *Flavobacterium okeanokoites* (Fokl). These ZF DNA-binding domains are made up of multiple C2H2 zinc fingers that determine the specificity of DNA binding, thereby directing ZFNs to their target sites. The ZF domains recognize three to four base pairs, and the tandem domain can bind to longer nucleotide sequences. For specific site cleavage in the genome, zinc finger domains are designed in pairs to bind two sequences flanking upstream and downstream the cut site. The FokI domains as a result bind and cut DNA at the site and create a DNA DSB that has a 5'overhang. DNA DSB is repaired via two repair mechanisms; the Homology-directed repair (HDR) is a precise mechanism, which occurs primarily during the S or G2 phase and involves the use of homologous donor DNA to repair DNA damage, or the Non-homologous end joining (NHEJ), is an error-prone mechanism, involves connecting broken ends of DNA together, leading to the formation of a heterogeneous pool of insertions and deletions and occurs throughout any cell cycle phases (Cathomen and Joung, 2008, Urnov et al., 2010).

TALENs is another tool with a structure similar to ZFNs. Both involve using the FokI nuclease domain to create a DSB at a specific target site in the genome. However, the DNA binding

domains in TALENs are different. TALENs use transcription activator-like effectors domains that are composed of a series of 33-35 amino acids repeats, each recognising single nucleotides. This increases the potential for specificity and target capabilities in comparison to ZFNs. Genome editing has been successfully accomplished using these techniques, however there are several disadvantages associated with these approaches. In the case of ZFNs, the requirement for three base pairs made the construction more challenging. Sites rich in guanine appeared to produce more efficient editing than sites lacking guanine. Furthermore, because ZFNs interact with DNA modularly (*i.e.* independently), editing efficiency was compromised as well. TALENs as a gene-editing tool was time-consuming and expensive. Although TALENs have a 5' thymine and 3' adenine structure, they limit the customization of targets. In addition, regions with high levels of DNA methylation were less efficiently edited with TALENs (Gupta and Musunuru, 2014, Li et al., 2020).

In recent years, studies have presented the use of CRISPR-Cas9 technique as a simpler, faster, cheaper method, with the ability to multiplex genome editing (Mali et al., 2013). In this way, multiple guide RNAs can be used simultaneously in a single cell to target multiple sites. CRISPR are repeated DNA sequences composed of 23 to 47 bp separated by non-repetitive sequences or spacers found in the genome of prokaryotic, CRISPR was initially discovered in *Escherichia coli*, by Dr Nakata's team in 1987 (Ishino et al., 1987). However, its role as a protection against bacteriophages was not clarified until 2007. Jansen *et al*. discovered in the early 2000s that these DNA repeat elements were found in both bacteria and archaea. The repeat elements consist of direct repeats separated by spacer regions of DNA between 21 and 37 bp in length. In addition, cas genes were identified, though the exact function of the system was unclear at the time (Jansen et al., 2002).

In 2005 Francisco Mojica and Christine Pourcel reported independently that the spacer regions between the repeat sequences are homologous to sequences found in bacteriophages, prophages, and plasmids (Mojica et al., 2005, Pourcel et al., 2005). In addition, Bolotin and colleagues confirmed these observations in a third influential paper published in the same year (Bolotin et al., 2005). The significance of these seminal publications was, however, grossly underappreciated.

Researchers have spent almost a decade improving the system, since the discovery in 2005, for it to be used as a tool that is safe and efficient in a variety of circumstances. Since the beginning of 2013, more than a thousand papers have been published containing the acronym CRISPR in the title or abstract. CRISPR-Cas9 has developed into an incredibly fast-paced field where laboratories around the world have used this technology to edit the genomes of a variety of different cell types and organisms. CRISPR-Cas9 system (Type II) was first introduced into mammalian organisms in 2013 and was considered the newest genome editing tool for inducing site-specific DSBs and subsequent mutagenesis in eukaryotic cells, plant, mouse, and human (Cong et al., 2013). The timeline in Figure 1.10 summarizes the major advances that have been made in this field.
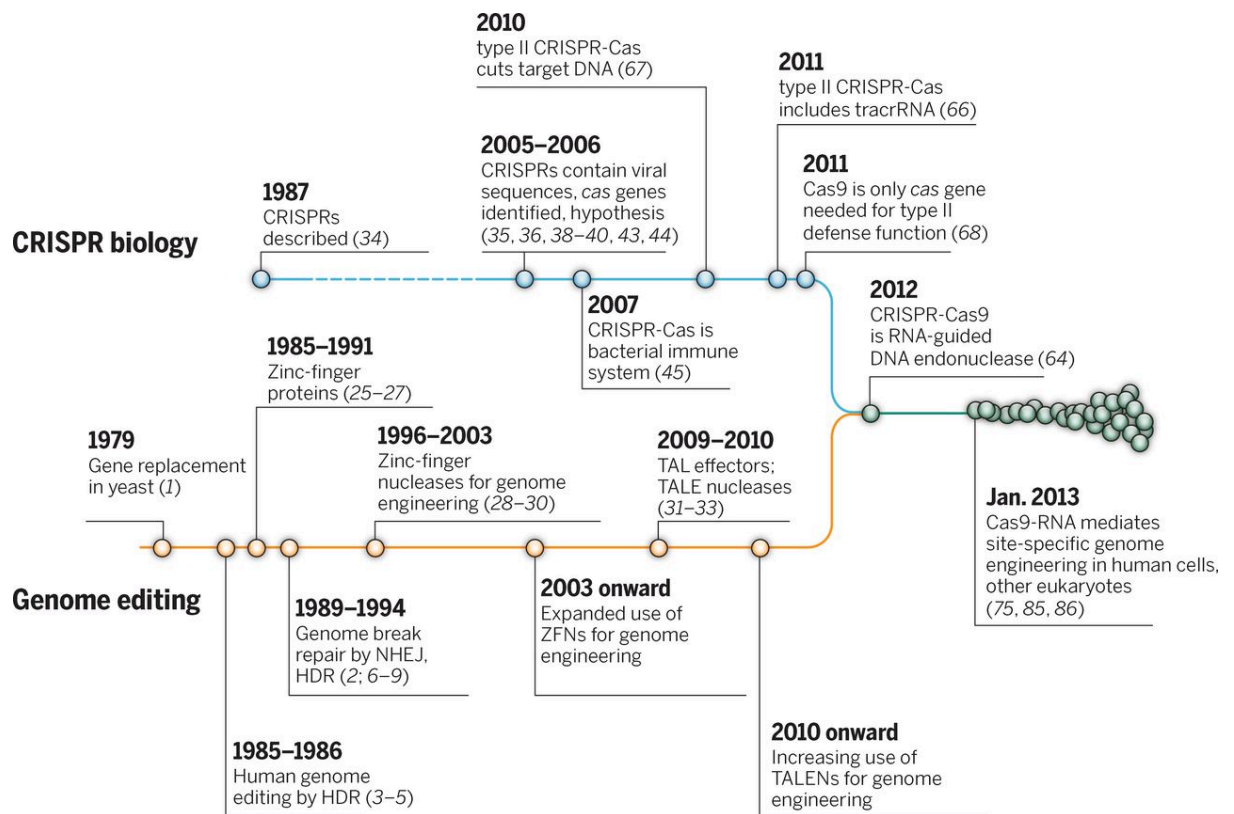
Figure 1.10: Timeline of genome editing and CRISPR-Cas9 system development.

An overview of key milestones in the development of genome editing technologies is outlined in this figure, with a particular emphasis on the emergence and advancement of CRISPR-Cas9. From the beginning of genetic manipulation to the most recent advances in precise gene editing is included in the timeline. Taken from (Doudna and Charpentier, 2014)

1.7.1 CRISPR-Cas9 principle

CRISPR-Cas9 is a novel targeted genomic editing method that enable genomic modifications in different mammalian cells. When using CRISPR-Cas9 as an editing tool, two parts are essential: (i) the Cas9 protein is an endonuclease that can recognize and cut the DNA at specific locations and initiates the DSBs in the target DNA, (ii) the gRNA, which is an RNA sequence of 20 nucleotides long that is complementary to the targeted DNA sequence to be edited. This gRNA assists in directing Cas9 to the DNA target region. Additionally, the protospacer adjacent motif (PAM) sequence, a short guanine-enriched sequence (5'NGG sequence) plays an important role in this system, since only those genomic DNA regions that contain PAMs can be recognised and bound by the Cas9, indicating their significance (Barrangou, 2013, Jinek et al., 2013).

Briefly, the CRISPR method involves using a gRNA to guide the Cas9 nuclease to the target region in the genome, where it can mediate DNA cleavage and create double strand breaks. DNA DSBs generated by Cas9 can result in endogenous cellular DNA repair processes, which can be exploited to engineer the genome. An illustration of the CRISPR/Cas9 genome editing system can be seen in Figure 1.11. The repair of DSBs can generally be accomplished in one of two ways: either by HDR or by NHEJ. During HDR repair, a donor DNA template with homology in the ends is used, this DNA can be inserted into the target site of the gene, causing frameshifts or insertion of DNA, is vital for single gene editing. In NHEJ no template is used, broken ends can be rapidly ligated. However, this process generates small insertions or deletions at specific sites, causing target genes' function to be altered or disrupted, which can lead to a knock-out phenotype resulting from frameshift mutations (Hryhorowicz et al., 2017). As a result of CRISPR-Cas9's simplicity and high efficiency it has become one of the most

widely used tool for studying gene expression alterations and has a great potential for modelling the treatment of cancer and other genetic disorders.

An advantage of CRISPR technology is that it allows the reproduction of tumour-associated chromosomal translocations, which result from illegitimate nonhomologous chromosome joining during carcinogenesis. With CRISPR-Cas9's ability to introduce DSBs at defined positions in cells, it has been possible to generate human cell lines and primary cells containing chromosomal translations that resemble those described in cancers, enabling researchers to diagnose and treat the disease such as lung cancer (Choi and Meyerson, 2014), acute myeloid leukaemia, and Ewing's sarcoma (Torres et al., 2014, Chen et al., 2014).
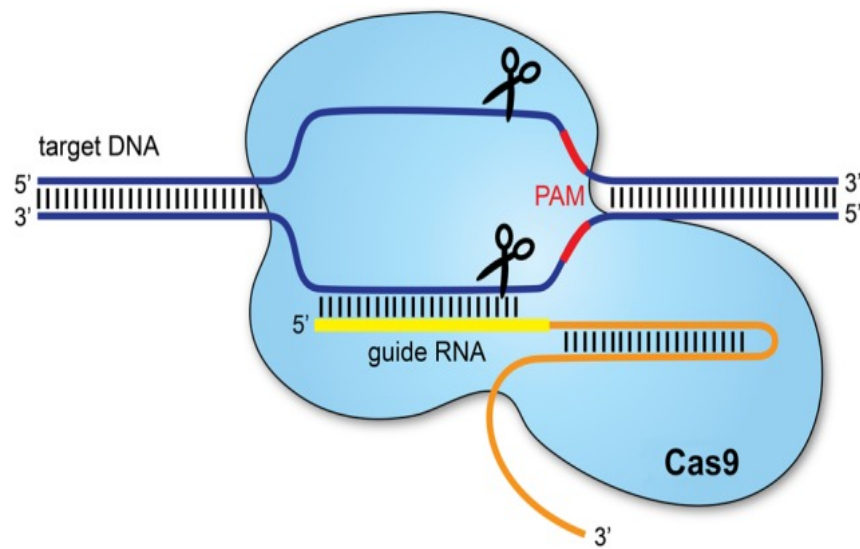
Figure 1.11: Schematic representing the CRISPR/Cas9 system.

The target DNA molecule represents the genomic region to be edited or modified. Cas9 is needed to recognize and bind to a short, specific DNA sequence adjacent to the target site, protospacer adjacent motif (PAM). gRNA consists of a sequence complementary to the target DNA. An endonuclease protein, Cas9 is illustrated as a molecular scissor, function to break double strands of DNA at the targeted sites. The 3' and 5' ends of the target DNA indicate that the DNA strands will be cleaved and repaired following Cas9-induced breaks. Taken from (Redman et al., 2016).

1.7.2 CRISPR-Cas9 applications

As previously mentioned, the development of new approaches, including CRISPR-Cas9, is crucial for functional studies, it has been possible to develop powerful animal genetic models that mimic the cooperative oncogenic lesions affecting genes that have an established role in the proliferation and establishment of leukemic clones (Nemudryi et al., 2014). CRISPR-Cas9 has already been used to study many diseases, including haematological diseases. In ALL most genetic modifications have been conducted using this technique, most of these experiments involved the mutation, insertion, or deletion of genes. Transcription factor deregulation is a common mechanism in the pathogenesis of human cancer, especially in leukaemia cells. Through chromosomal translocations, transcription factor genes are often amplified, deleted, or rearranged, it may also be subjected to point mutations that result in functional gains or losses. It may therefore be possible to treat ALL effectively by targeting transcription factors (Montaño et al., 2018a).

In nearly 80% of B-ALL patients, transcription factors such as PAX5 and IKZF1 were altered. In response to these alterations, glucose metabolism and energy supply have been adversely affected, where transcription factors function as metabolic repressors by limiting the production of ATP. The CRISPR-Cas9-based screening of PAX5 and IKZF1 transcriptional targets identified several target genes including NR3C1, TXNIP and CB2 as central effectors of B-lymphoid glucose and energy restriction, and therefore as potential targets for treating B-ALL (Chan et al., 2017).

In the case of T-ALL, TAL1, which is one of the most frequently deregulated oncogenes and TRIB1 (TRB1) genes, were disrupted with CRISPR-Cas9 technique to investigate their biological function. In this study, two known changes to TAL1 protein (insertion and deletion) were reproduced in a cell line and that resulted in alterations in their expression. Additionally,

there was also a change in the methylation and acetylation of H3K27, which suggests a causal relationship between mutagenesis, epigenetic modulation, and TAL1 expression (Navarro et al., 2015).

Overexpression of LMO2 has been related to the development of T-ALL. However, the cause of this overexpression remains unclear since few mutations have been reported. A study conducted by Rahman's group in 2017 using diagnostic samples from paediatric and adult T-ALL patients as well as PF-382 and DU528 T-ALL cell lines. They identified a recurrent mutation hotspot resulting in an aberrant promoter within intron 1 of LMO2. Disruption of the mutant allele by CRISPR/Cas9 in PF-382 T-ALL cell lines, significantly decreased LMO2 expression, demonstrating a clear causal relationship between the mutation and dysregulation of oncogene expression. Moreover, the spectrum of CRISPR/Cas9-derived mutations reveals important insights into how functional transcription factor binding is interconnected. These mutations create high-confidence binding sites that may be utilized by MYB, ETS1, or RUNX1, all of which belong to the highly oncogenic TAL1-LMO2 complex (Rahman et al., 2017).

1.7.3 CRISPR-Cas9 limitations

Despite its wide usage as a genome editing tool, CRISPR-Cas9 still has a number of limitations that drive researchers to propose solutions to develop it. Its main difficulty, as with all gene editing, is that Cas9 nuclease has a diverse tolerance to mismatch between gRNA and targeted genomic DNA, which can lead to off-target effects. Another cause of off-target effects may be PAM mismatch (Fu et al., 2013). It is interesting to note that whole-genome sequencing has demonstrated that off-target sites typically occur in mouse ESCs but occur very rarely in human stem cells (Koo et al., 2015, Schwank et al., 2013, Smith et al., 2014). However, it is important to improve the targeting specificity. This could be by creating a well-designed gRNA, to ensure

high efficiency of genome cleavage and low rates of off-target effects (Kim et al., 2016, Naito et al., 2015, Montague et al., 2014, Cradick et al., 2014). Additionally, by reengineering the Cas9 nuclease structure, CRISPR-Cas9 could be rendered more specific in its targeting (Ran et al., 2013). The use of Cas9 mutants, for example, can introduce a single nick into the DNA strand bound to gRNA without destroying the entire double strand. In this case, single stranded breaks (SSBs) are caused by mutants that combine two single-stranded RNA pairs to cause double-stranded breaks (DSBs) at the exact same location (Ran et al., 2013). There are several mutant Cas9 proteins that are highly efficient, such as Cas9 nickase (Cas9n), Streptococcus pyogenes Cas9 (SpCas9), SpCas9-high fidelity-1, eSpCas9 1.1, and hyper-accurate Cas9 (Slaymaker et al., 2016, Kleinstiver et al., 2016, Chen et al., 2017). Furthermore, it has been shown by several studies that altering the target specificity of gRNA by modifying its length either decreasing or increasing, can increase the efficiency of on-target cleavage without sacrificing the targeting specificity (Fu et al., 2014). In addition, a hairpin structure was demonstrated to significantly reduce off-target rates by 55-fold by designing it at the 5′ end of the gRNA spacer (Kocak et al., 2019).

1.8 Aim and objectives

This study aims to understand how transcription factors determine gene expression patterns defining cellular identity, particularly in the context of T-cell Acute Lymphoblastic Leukaemia (T-ALL). A number of transcription factors, such as TAL1, as well as E-proteins, such as HEB, play critical roles in the development of T-cells. T-ALL is characterized by abnormal expression patterns, including overexpression of LMO proteins and TAL1. Therefore, in order to develop effective therapeutic strategies, it is necessary to explore these processes in greater depth. In this study, we investigate the role of HEB, a key factor in guiding thymocytes to maturation stages, and its interaction with GFI1, whose functions are still unknown in T-ALL. Using proteomic analysis, genome-wide studies, and CRISPR-Cas9 gene editing experiments, this study reveals distinct binding complexes, regulatory interactions between HEB isoforms, and the effect on GFI1 expression. The specific objectives of this study were to:

- Investigating the role of the transcription factor HEB in T-cell Acute Lymphoblastic Leukaemia.

- Investigating the functional interaction between HEB and GFI1 in T-ALL.

- Exploring the relation between DNA methylation and HEB binding in T-ALL.

- Performing CRISPR/Cas9 experiment to delete HEB gene and assess the impact on T-ALL cell proliferation, apoptosis, and transcription factors.

- Understanding how GFI1 can influence the expression of genes associated with T-ALL, which can help to develop novel treatments for the disease.

Chapter 2 Materials and Methods

2.1 Cell culture

T-ALL cell lines (ARR, DU528, HSB2, and CCRFCEM) were cultured in RPMI medium (Sigma), supplemented with 10 % Fetal Bovine Serum (FBS), 2 mM GlutaMAX (Gibco), 100 units/mL penicillin, 100 μg/mL streptomycin (Gibco), 0.075 mM monothioglycerol (MTG; Sigma). Cells were maintained at a density range of 0.4 - $2x10^6$ cells/ml by centrifugation for 5 min at 300 g and resuspension in fresh medium every other day. Cells were incubated at 37 °C and 5 % CO_2 in a humidified incubator.

Cell lines

Our study was conducted using proteomic analysis and genome-wide approaches in four human T-ALL cell lines. In all four cell lines (ARR, DU528, HSB2 and CCRFCEM), LMO2 is expressed (Sandberg et al., 2007a). The cell lines ARR and DU528 were obtained from Dr A.W. Langerak, (Erasmus Medical Centre, Rotterdam, NL). HSB2 and CCRFCEM cell lines were provided by Prof. P. N. Cockerill (University of Birmingham).

ARR is an early T-cell precursor cell line that carries some myeloid features. ARR cells were derived from a $CD7^+$ $CD33^+$ $CD3^-$ T-ALL patient sample. DU528 is a leukaemia-derived immature T-cell line that originated from a 16-year-old male patient by Kurtzberg et al. in 1985. HSB2, an immature T-cell line, was derived in 1966 from an 11-year-old male patient with acute lymphoblastic leukaemia and subsequently passaged in new-born Syrian hamsters. CCRFCEM comprises lymphoblastic leukaemia cells obtained from a 4-year-old Caucasian female patient with T-ALL secondary to lymphosarcoma, as reported by Kaplan et al. in 1974. The cell lines DU528, HSB2 and CCRFCEM are all associated with small interstitial deletion (1p32), that fuses the regulatory region of the SCL-interrupting locus (SIL) with the TAL1 gene, resulting in continuous expression of TAL1 (Cavé et al., 2004), (Janssen et al., 1993). On the basis of their surface marker expression, ARR is the least differentiated T-ALL cell line as it is only positive for CD7. Therefore, its development has been blocked at the DN1 stage.

This is followed by DU528 which is arrested at the DN2 stage as it expresses CD5 in addition to CD7. HSB2 has cytoplasmic CD3 (CyCD3) and therefore belongs to the DN3 group, while CCRFCEM is a mature T-cell, arrested at the single positive stage, as it is CD3 and CD4 positive with rearranged TCRαβ. Table 2.1 illustrates the specific immunophenotypic markers expressed in each cell line. Several reasons led to the decision to use all four cell lines for our subsequent studies. The most important is that these cell lines, blocked at the different stages of T-cell development, typify the whole range of T-ALLs, allowing us to maximise the possibility of identifying oncogenic mechanisms that are common in all T-ALL.

Table 2.1: Immunophenotype of human T-ALL cell lines at different stages of development.

This table lists cell line names, origins, and surface marker expressions.

Cell line	T-cell malignancy of origin	CD7	CD5	CD1a	CyCD3	CD3	TCR	CD4	CD8	TdT	CD2
ARR	CD3- T-ALL	+	−	−	ND	−	−	−	−	ND	−
DU528	CD3- T-ALL	+	+	−	ND	−	−	−	−	−	−
HSB2	CD3- T-ALL	+	+	−	+	−	−	−	−	−	−
CCRFCEM	TCRαβ+ T-ALL	+	+	+	+	+	αβ	+	−	+	−

Abbreviations: CD, cluster of differentiation; TCR, T-cell receptor; TdT, terminal deoxynucleotidyl transferase; ND, not done. Adapted from Sandberg et al. (2007).

2.2 Total protein extraction

Cells were pelleted by centrifugation at 300 g for 5 min, followed by a wash with phosphate-buffered saline (PBS). IP cell lysis buffer (150 mM NaCl, 20 mM Tris pH8, 2 mM EDTA, 0.5 % NP-40, 1:1000 protease/phosphatase inhibitor cocktail (PPI; Roche)) was used to resuspend the cell pellet at 100×10^7 cells/ml and incubated on ice for 10 min. Samples were centrifuged for 10 min at 4 °C, after which the supernatant was collected and stored at -20 °C. Protein concentration was measured using a BCA Protein Assay Kit (Thermo scientific).

2.3 Nuclear protein extraction

For enrichment of nuclear proteins, cells were pelleted by centrifugation at 300 g for 5 min, followed by two washes with phosphate-buffered saline (PBS) and hypotonic buffer (10 mM HEPES pH7.6, 10 mM KCl, 1.5 mM MgCl2, 1:1000 protease/phosphatase inhibitor) respectively. Cells were then resuspended in hypotonic buffer at 2×10^7 cells/ml and incubated on ice for 1 h. The lysate was centrifuged at 500 g for 5 min at 4 °C to pellet the nuclei, after which the pellet was resuspended at 3×10^8 cells/ml hypertonic buffer (20 mM HEPES pH 7.6, 420 mM NaCl, 1.5 mM MgCl2, 0.2 mM EDTA, 0.5 % NP40, 20 % Glycerol, 1:1000 PPI) and incubated on ice for 20 min to extract nuclear proteins. Samples were centrifuged at 15,000 g for 2 min at 4 °C to pellet the insoluble proteins. Supernatant was transferred to new tubes and diluted with 1.8 volume of No Salt Buffer (20 mM HEPES pH 7.6, 1.5 mM MgCl2, 0.2 mM EDTA, 0.5 % NP40 and 20 % Glycerol, 1:100 PPI) to adjust the salt concentration of the sample. Nuclear extracts were frozen and stored at -20 °C until use. Protein concentration was measured using a BCA Protein Assay Kit (Thermo scientific).

2.4 RNA isolation and cDNA synthesis

RNA was isolated using a NucleoSpin RNA isolation Kit (Macherey-Nagel), according to the manufacturer's protocol. In brief, up to 5×10^6 cells were collected, centrifuged, and resuspended in a lysis buffer containing β-mercaptoethanol. Then, lysate was filtered and

bound to an extraction column for DNA digestion, after which the column was washed and dried. The RNA was eluted using RNase-free H_2O and RNA concentrations were measured using a Nanodrop 2000 C spectrophotometer (Thermo Scientific). Following the addition of RNase inhibitor (Invitrogen) to RNA samples, they were stored at -80 °C.

For cDNA synthesis, 2 μg of total RNA was incubated with 0.5 μg/ml Oligo d(T) primer (Life technologies) at 70 °C for 5 min, followed by the addition of reverse transcriptase (RT) buffer, 20 mM dNTP, 0.2 mM Moloney Murine Leukaemia Virus (M-MLV) reverse transcriptase, 0.4 mM Recombinant RNase inhibitor and incubation at 42 °C for 45 min. The reaction was finished with 95 °C for 5 min to inactivate the reverse transcriptase. cDNA samples were stored at -20 °C.

2.5 Quantitative Polymerase chain Reaction (qPCR)

Gene expression levels were measured by real time qPCR. Reactions were performed in 96-wells plate, each well containing 5 μl of Luna qPCR master mix (NEB), 0.25 μM of each forward and reverse primer and 5-fold diluted (in H_2O) cDNA sample. Samples were run in duplicates for each amplicon. For relative quantitation, a standard curve was generated, using 5-fold serial dilutions of cDNA. Plates were run using an Applied Biosystems 7500 Step-One Plus Real-Time PCR system. Intron-spanning qPCR primers were designed using Primer3 software and purchased from Sigma. Sequences are shown in Table 2.2 and primer pairs were tested to generate a single product. TBP expression was used as a control.

Table 2.2: Primer sequences used in qPCR

Human	Forward	Reverse
TBP	CAGGAGCCAAGAGTGAAGAACA	AGCTGGAAAACCCAACTTCTGT
HEB	TCACACACTCCTCCCATCAA	CAAAGCCTTTCCAAGTGCAT
GFI1	GCTCCTACAAGTGCATCAAGTG	AGTGGATAAGCAGGTGTGTGG
GFI1B	GAGCATCTGCGGCAAAAC	TGGACGAGCGCTTGAAGGC
TAL1	GTTCTTTGGGGAGCCGGATG	TGAAGATACGCCFCACAACT
LMO2	GAAGAGCCTGGACCCTTCAG	TCGATGGCCTTCAGGAAGTA

2.6 Immunoprecipitation

For each immunoprecipitation, 60 μl of Dynabeads protein G (Invitrogen) was washed with PBS and incubated with 6 μg IgG or indicated antibody (Table 2) for 30 min on a rotating wheel at 4 °C. Then, 600 μg of protein extract was added and incubated for 2 h on a rotating wheel at 4 °C. After incubation, samples underwent three washes with 1.8:1 hypotonic: hypertonic buffer. Samples were eluted from the beads in loading buffer containing sample reducing agent. An input control of 100 μg of total protein extract was prepared and all samples were incubated at 95 °C for 10 min. Samples were loaded and analysed by western blotting.

2.7 Western blotting

Protein extracts were prepared in Loading Buffer (B0007 Novex-Life Technologies) with the addition of sample reducing agent (B0009 Novex -Life Technologies) and incubated at 95 °C for 10 min. Samples were loaded alongside a protein molecular weight ladder (Thermo Scientific- PageRuler) on 4-12 % gradient Bis-Tris-Plus-Gels (Invitrogen) and run at 200 V for 35 min in MES running buffer. Proteins were transferred to an iBlot nitrocellulose membrane (Invitrogen) using an iBlot Gel transfer machine (Invitrogen). Membranes were stained with Ponceau S to establish equal loading of proteins. After washing off the Ponceau S stain in PBS, membranes were blocked using 5 % milk in PBS for 1 h. Membranes were then washed with PBS and incubated with 1 μg/ml primary antibody in PBS (Table 2.3) on a roller at 4 °C overnight. Membranes subsequently underwent three washes with PBS for 10 mins each. Secondary antibodies were used at a concentration of 1:2000 in PBS (Table 2.4) and incubated for 30 min on a roller, followed by three washes with PBS. Blots were scanned using a ChemiDoc MP Imaging System.

Table 2.3: List of primary antibodies

Primary antibodies	Applications	Species	Source
αHEB	WB IP Immunostaining ChIP	Mouse monoclonal	R&D SYSTEM, MAB6928
αGFI1	WB IP ChIP MS	Rabbit polyclonal	Abcam, Ab21061
αGFI1	WB	Mouse monoclonal	Novus, NBP2-45833
αGFI1	WB	Goat polyclonal	R&D, AF3540
αLMO2	WB IP Immunostaining ChIP	Goat polyclonal	R&D SYSTEM, AF27126
αGFI-1B	WB	Mouse polyclonal	Abcam, Ab167304
aTAL1	WB ChIP	Goat polyclonal	SANTA CRUZ, sc-28356x
αGATA2	WB ChIP	Goat polyclonal	R&D system AF2046
αLDB1	WB ChIP	Rabbit polyclonal	Abcam ab96799
αPHF6	WB ChIP	Mouse monoclonal	Santa-Cruz sc-385237
αPRMT5	WB ChIP	Mouse monoclonal	SANTA CRUZ, Sc-376937
αLYL1	WB ChIP	Mouse monoclonal	SANTA CRUZ, sc-374164
αRUNX1	WB	Rabbit polyclonal	Abcam AB23980
αMono-Methyl Lysine (mme-K)	WB MS	Rabbit monoclonal	Cell signalling, 14679
5-Methylcytosine (5-mC)	MeDIP	Rabbit	Cell signalling, 28692
αTubulin	WB	Mouse monoclonal	Sigma, T6199
αIgG2B	WB	Mouse monoclonal	R&D SYSTEM MAB004
IgG kappa	WB	Mouse monoclonal	R&D SYSTEM MAB10050
Normal rabbit IgG	WB	Rabbit	Santa Cruz sc-2027
Normal mouse IgG	WB	Mouse	Santa Cruz sc-2025

Table 2.4: List of secondary antibodies

Secondary antibodies	Source
Alexa Fluor 790 -Monoclonal mouse Anti-Rabbit IgG	Jackson ImmunoResearch, 211-652-171
Alexa Fluor 790 - Monoclonal Mouse Anti-Goat IgG	Jackson ImmunoResearch, 205-652-176
Alexa Fluor 488 dye – Donkey Anti-Goat IgG	Invitrogen, A11055
Alexa Fluor 647 dye – Chicken Anti-Mouse IgG	Invitrogen, A21463
IRDye 800CW – Donkey Anti-Mouse IgG	Li-Cor, 926-32212
IRDye 800CW – Donkey Anti--Goat IgG	Li-Cor, 926-32214

2.8 Mass spectrometry

Immunoprecipitation was performed by incubating 1 mg of protein extract with 100 μg of Dynabeads Protein G (Invitrogen) and 10 μg of IgG or indicated antibody. Input control and samples were loaded and separated on a 4-12 % blot-Tris-Mini gel (Invitrogen) followed by Coomassie blue staining. After de-staining with 10 % acetic acid overnight, lanes were sliced into 12 pieces each. Prior to in-gel digestion, the gel slices underwent three washing steps at room temperature: 500 μl of 50 % acetonitrile/ 50 mM Ammonium Bicarbonate for 45 min with agitation, 100 mM Iodoacetamide in 10 % Acetonitrile/ 50 mM Ammonium Bicarbonate for 30 min in the dark, 500 μl of 10 % Acetonitrile in 50 mM Ammonium bicarbonate for 30 min with agitation. Gel slices were then dried using a vacuum micro-centrifuge overnight. Trypsin digestion was performed overnight by rehydrating the gel slices with 500 μl of 10 % Acetonitrile/ 50 mM Ammonium Bicarbonate containing 0.5 μg of Trypsin (Promega). To extract the peptides, gel slices were incubated in 100 μl of 1 % formic acid in 10 % Acetonitrile for one hour, followed by 100 μl of 2 % formic acid in 60 % acetonitrile. Supernatants were pooled and the peptide mixtures were dried in a vacuum micro-centrifugate and reconstituted in 20 μl of 1 % formic acid. Mass spectrometry was performed by Dr D. Ward, University of Birmingham, using a Bruker Impact ESI-TOF machine with running software Otof. MS/MS data was acquired in data dependent mode. Peptides were put on an exclusion list for 60 seconds after being selected for MS/MS twice. LC-MS/MS data were then searched against the Swiss-Prot human protein sequence database, using the MASCOT software package (www.matrixscience.com). For peptide sequence analysis and MS/MS ion search, peptides were reported if $p<0.05$ and proteins with a MOWSE score (-10 log (P), where P (α value / number of database sequences)). Mass spectrometry data were analyzed using online software (Venny v2.1.0) http://bioinfogp.cnb.csic.es/tools/venny/, Protein-protein networks and biological functions were demonstrated using STRING v11 https://string-db.org (Szklarczyk

et al., 2018). Figure 2.1 shows a schematic illustration of pulldown and mass spectrometry procedure.
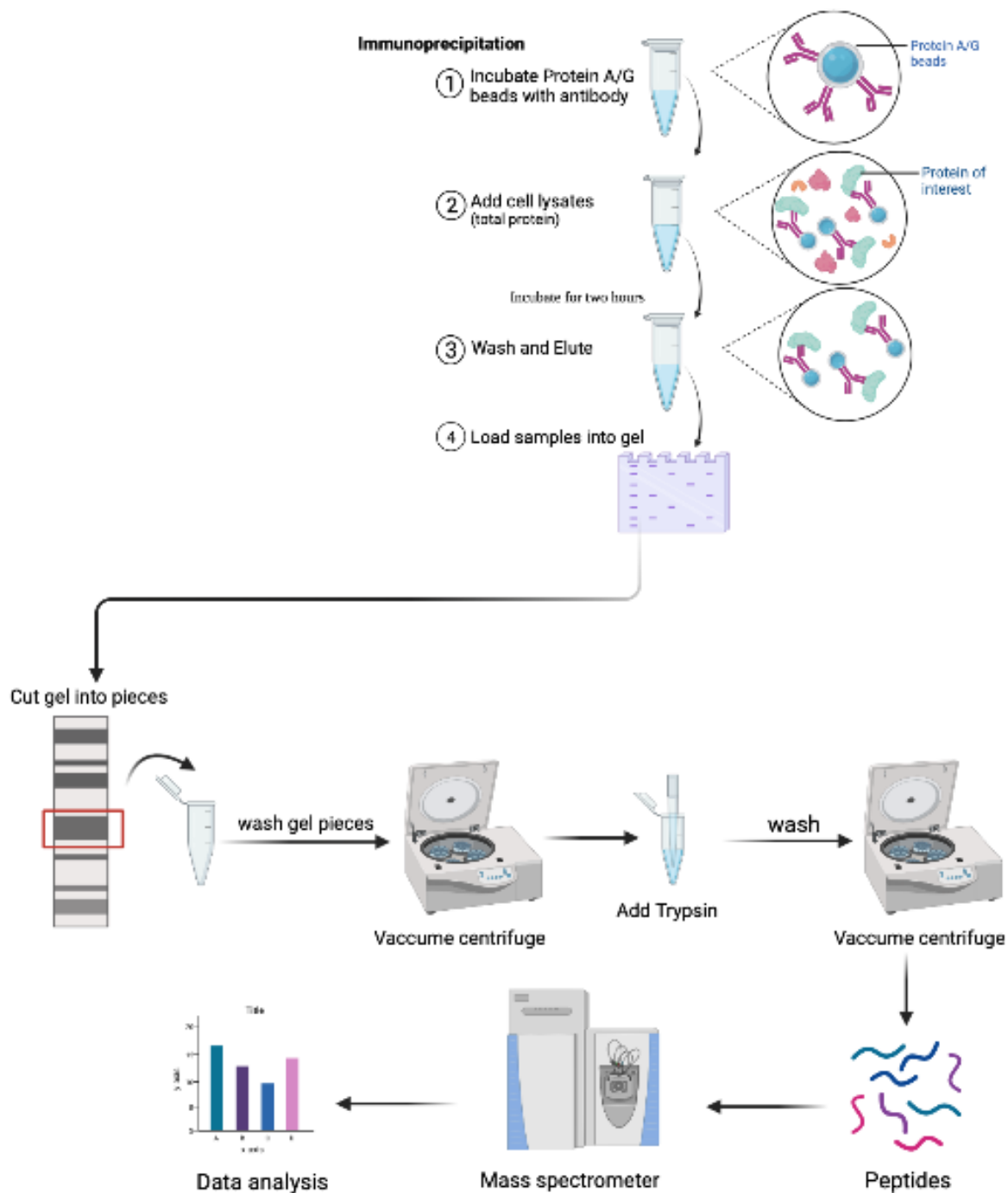
Figure 2.1: Illustration of Immunoprecipitation and Mass spectrometry protocol.

(A) A schematic showing the pull down and mass spectrometry experiment strategy. Incubation of magnetic beads with antibodies of interest was followed by incubation with nuclear extracts of T-ALL cells., samples were eluted and analysed by western blotting. (B) For mass spectrometry, procedure in part (A) followed by gel stained with Coomassie blue staining. Gel lanes were sliced into pieces and underwent washing steps and subjected to trypsin digestion then analysed by mass spectrometry analysis. Figure was created using Biorender.com.

2.9 HEB CRISPR-Cas9 plasmid preparation

The HEB CRISPR-Cas9 KO plasmid mix (Santa Cruz; sc-402508) consisted of 3 plasmids. To determine the genomic target site of each plasmid by sequencing, plasmids were expanded through bacterial transformation. Transformation of 10 ng of DNA plasmid was added to 50 µl of Alpha-Select Gold Efficiency Chemically Competent cells (Bioline, BIO-85027) in a tube, placed into a 42 °C water bath for 30 sec, then quickly transferred to ice and incubated for 2 min. Lysogeny Broth (LB) was added, and samples placed in a 37 °C shaker for 20 min. Bacteria were spread over LB plates containing 50 mg/ml ampicillin and incubated at 37 °C for 16 h . Subsequently, colonies were picked and placed in LB containing 50 µg/ml ampicillin and incubated on a shaker at 37 °C overnight. To isolate DNA from bacteria, the medium was spun for 2 min at 15,000 g, after which the supernatant was discarded. TENS buffer (10 mM Tris pH 8.0, 1 mM EDTA, 0.1 M NaOH, 10 % SDS) was added to the pellet and mixed by inverting the tube. Then, 3 M sodium acetate pH 5.2 was added and centrifuged for 2 min at 15,000 g. Supernatant was transferred to a new tube, two volumes of 100 % ethanol was added and centrifuged for 5 min at 15,000 g. The pellet was then washed with 70 % Ethanol and airdried. The dry DNA pellet was dissolved in 0.1 x TE (10 mM Tris pH 8.0, 1 mM EDTA) buffer containing 20 µg/ml of RNase A at final concentration. Samples were sent for sequencing (Source Bioscience) and the sequence was used to determine the relative position. After identification of the positive clones, the transformed samples were added to a flask containing LB and Ampicillin at final concentration 50 mg/ml and incubated on 37 °C shaker for 16 h in order to produce a large quantity of bacteria with the HEB Cas9 plasmids. DNA purification was performed using a NucleoBond Xtra Maxi kit (Machery-Nagel). DNA concentrations were measured using a Nanodrop 2000 C (Thermo scientific). Figure 2.2 shows a schematic illustration of CRISPR/cas9 experiment.

2.10 Transfection of ARR cells

The ARR cell line was transfected with N-terminal HEB CRISPR-Cas9 plasmid and Homology-directed repair (HDR) plasmid or Cas9 control plasmid and Homology-directed repair (HDR) plasmid. Cells were centrifuged at 300 g for 5 min and resuspended at 10^7 cells/ml in Opti-MEM medium. Electroporation was performed in 4 mm cuvettes, each containing 400 μl cell suspension with 25 μg/ml plasmid. Using a Gene Pulser Xcell (Bio-Rad), the transfection was performed using programme conditions capacitance 1000 F at 320 V. Transfected cells were transferred to 6-well plates containing 3 ml culture medium. After 48 h, transfection efficiency was examined by fluorescence microscopy. Transfected cells were selected by the addition of 0.1 μg/ml puromycin. Fluorescence-activated cell sorting was used to select the transfected cells more stringently and single cell dilutions were prepared, plating 0.25 cells/well in 96-well plates in order to achieve clonal populations using medium with 20 % FBS. Cells were monitored and expanded for total protein extract followed by western blot analysis.
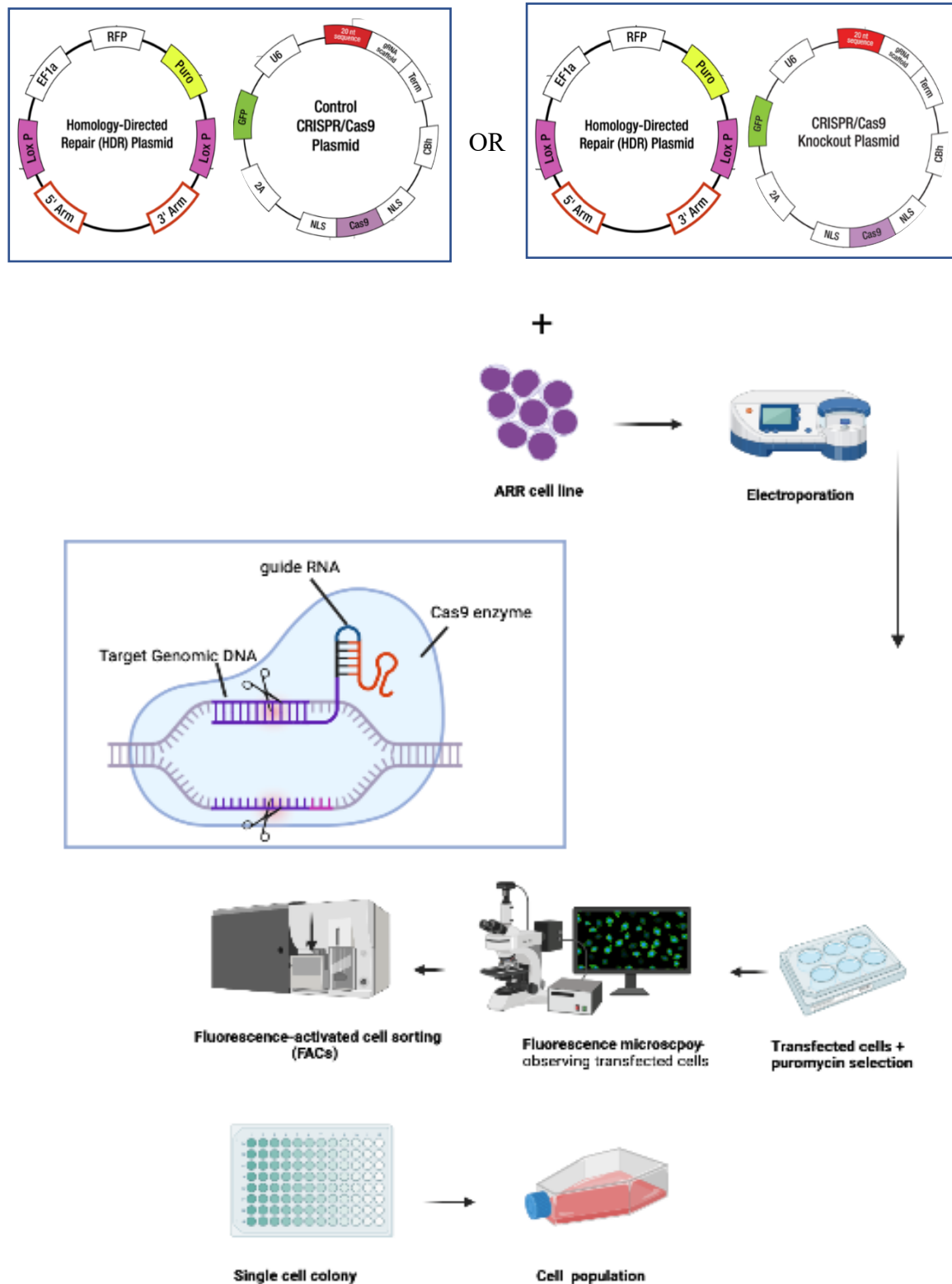
Figure 2.2: Illustration of CRISPR Cas9 HEB KO experiment.

ARR cell line transfected with HDR and Cas9 control and/or HDR and Cas9 HEB KO plasmids. Transfected cells underwent puromycin selection, then sorted using FACs. Cells were platted 0.25 cells/well in 96-well plates, monitored and expanded for protein extract. Figure was created using Biorender.com.

2.11 Immunofluorescence staining

Cells were pelleted by centrifugation at 300 g for 5 min and fixed by resuspended in 4 % formaldehyde and incubated at room temperature for 10 min. Cells were washed with PBS and incubated with 1 ml of permeabilization buffer (0.5 % Triton X-100 and 1 % FBS) for 1 h. Cells were then incubated overnight at 4 °C with 1 µg/ml primary antibody. After this, the cells were washed twice using PBS and incubated with the appropriate secondary antibody (1:1000) for 1 h. Cells were washed with PBS and loaded on slides using a Cytospin III centrifuge at 500 g for 2 min. DAPI (Life Technologies) antifade reagent was added then coverslips were placed, mounted and imaged using Zeiss LSM 880 Meta laser confocal microscope. Images were visualised using Zen 2012 Zeiss microscopy software.

2.12 Cytogenetic chromosomal spreading

Cells were treated with 20 ng/ml of KryoMax Colcemid solution (Gibco) at 37 °C for 2 h to arrest cell cycle at metaphase. Cells were then pelleted at 300 g for 8 min and resuspended and incubated with 75 mM potassium chloride (KCl) lysis buffer or 30 min at 37 °C. Cells were fixed with cold Carnoy's solution (3:1 methanol/glacial acetic acid) at room temperature for 30 min. Cells were washed and resuspended in cold Carnoy's solution then dropped onto humidified, chilled glass slides. The quality of metaphase spread, mitotic index, presence of cytoplasm and overlap were evaluated under a phase contrast microscope. Then slides were air-dried in a humidified bed and stained with Giemsa-modified solution (Gibco). Chromosome were imaged and counted using light microscope.

2.13 Kwik-Diff Morphological staining

Cellular morphology was visualized using Kwik-Diff (9990700-Thermo Fisher) staining. Per slide, 10^6 cells were washed with PBS, spun, and loaded onto microscopy slides using a Cytospin III centrifuge at 500 g for 5 min. Slides were air-dried at room temperature and then dipped in respectively fixative solution, eosinophilic stain (Eosin), and basophilic stain

(Methyle Blue), for 30 s each. The slides were washed in water, air-dried, and mounted. Slides were imaged using a light microscope.

2.14 Chromatin preparation

T-ALL cells were crosslinked with 1% formaldehyde for 12 min at room temperature and quenched with 0.4 M Glycine. Cells were centrifuged at 300 g for 5 min at 4 °C and washed with PBS. Nuclei were isolated by incubation in buffer A (10 mM HEPES pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25 % Triton X-100, 1:1000 dilution of protease inhibitor cocktail (Roche), 0.1 mM PMSF) for 10 min at 4 °C on a rotating wheel, followed by centrifugation at 500 g for 5 min at 4 °C. This was followed by incubation with buffer B (10 mM HEPES pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01 % Triton X-100, PIC 1:1000 dilution) for 10 min at 4 °C on a rotating wheel. The nuclear pellets were resuspended in IP Buffer I (25 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA, 1 % Triton X-100, 0.25 % SDS, PIC 1:1000 dilution) at 10^7 cells/ 300 µl for 20 cycles of 30 s on, 30 s off on ice using a Bioruptor sonicator (Diagenode). Following the sonication, samples were centrifuged at 15,000 g for 10 min at 4 °C, supernatant was collected, and diluted with 2X IP buffer II (25 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA, 1 % Triton X-100, 7.5 % Glycerol, PIC 1:1000 dilution). Chromatin samples were stored at -80 °C until used.

2.15 Chromatin Immunoprecipitation (ChIP)

ChIP assays were performed using protein G Dynabeads. The beads were placed on a tube on a magnetic rack, washed twice with PBS and then resuspended in PBS and incubated with antibodies of interest (4 µg antibody, 40 µg beads) for 30 min on a rotating wheel at 4 °C. As input control, 10 % of the chromatin sample was incubated with 20 g of RNaseA at 37 °C for 1 h, after which 1 mg/ml proteinase K was added along with 0.15 % SDS, and then the samples were incubated at 65 °C overnight to allow proteins to digest and crosslink to reverse. Chromatin samples were incubated with beads bound to antibody for 2 h at 4 °C on a rotating

65

wheel, then washed twice with ChIP buffer I (20 mM Tris 1 M, pH 8.0, 150 mM NaCl, 2 mM EDTA, pH 8.0, 1 % TritonX100, 0.1 % SDS), then twice with ChIP buffer II (20 mM Tris, pH 8.0, 500 mM NaCl, 2 mM EDTA, pH 8.0, 1 % TritonX100, 0.1 % SDS), followed by a wash with TE/NaCl buffer (TE, pH 8.0, and 50 mM NaCl). Chromatin samples were eluted using elution buffer (100 mM NaHCO3, 1 % SDS, 0.2 M NaCl, 250 µg proteinase K) and crosslinks were reversed overnight at 65 °C. Following this, ChIP supernatants were separated from beads using a magnetic rack. DNA was purified by AMPure XP beads (Bechman Coulter), according to the manufacturer's protocol.

2.16 ChIP sample validation.

The effectiveness of the ChIP experiments was validated by qPCR using specific primers against both positive and negative control regions (Table 2.5). Samples were validated by qPCR in 10 µl reactions using 5 µl of Luna qPCR master mix, 0.25 µM of each forward and reverse primer and 2.5 µl of 10-fold diluted ChIP sample. A 5-fold serial dilution curve of genomic DNA was used as a standard curve for quantification. The fold-enrichment was calculated based on the quantities relative to the negative control regions.

Table 2.5: ChIP qPCR primers sequences.

Human	Forward	Reverse
Chr.18	ACTCCCCTTTCATGCTTCTGATATCCATT	AGGTCCCAGGACATATCCATT
PU.1	TGTGCGGTGCCTGTGGTAAT	AACAGGAAGCGCCAGTCA
SDE2	GGGTTAGGTTTCGAGCTTCC	CAGCTTCTCTCCCCACACTC

2.17 ChIP-seq library preparation.

ChIP-seq libraries were prepared using the Illumina sample preparation protocol. For each sample, DNA fragments were first converted to blunt-ended DNA by incubation at 20 °C for 30 min in ligase buffer with ATP (NEB B0202S), additionally containing 0.12U/µl Klenow DNA polymerase (Invitrogen 18012-021), 0.5 U/µl T4 polynucleotide kinase (PNK) (NEB M0201L), 0.1 U/µl T4 DNA polymerase (NEB M0203L) and 0.4 mM dNTPs (NEB B02202S),

placed in PCR machine. Following each step, DNA was purified using a MinElute column purification kit (QIAGEN). To enable adapter ligation, a single A nucleotide was added to the 3` ends of the blunt-ended DNA that contain 0.3 U/μl Klenow exo⁻ (NEB M0212L), 0.2 mM dATP in NEB buffer 2 (NEB B7002S) placed in PCR machine at 37 °C for 30 min. Samples were then purified and eluted. Illumina adapters with unique tag sequences were ligated to DNA sample using Quick T4 DNA ligase (NEB E6056S) and 1:20 diluted adapters (Illumina) in Ligase buffer (NEB E6056S) and placed in PCR machine at 20 °C for 30 min. Purified DNA samples were amplified using Phusion hot-start polymerase (NEB M0535), 0.3 mM dNTPs, and primers in HF buffer (NEB F540S), followed by 18 cycles in a PCR machine. DNA samples were purified and were run on a 3 % agarose gel. DNA size was selected between 200-300 bp (Figure 2.3) and isolated using QIAGEN MinElute column. Libraries were quantified using a KAPA library Quantification Kit (Illumina) and sequenced with an Illumina Nextseq 500 sequencer.
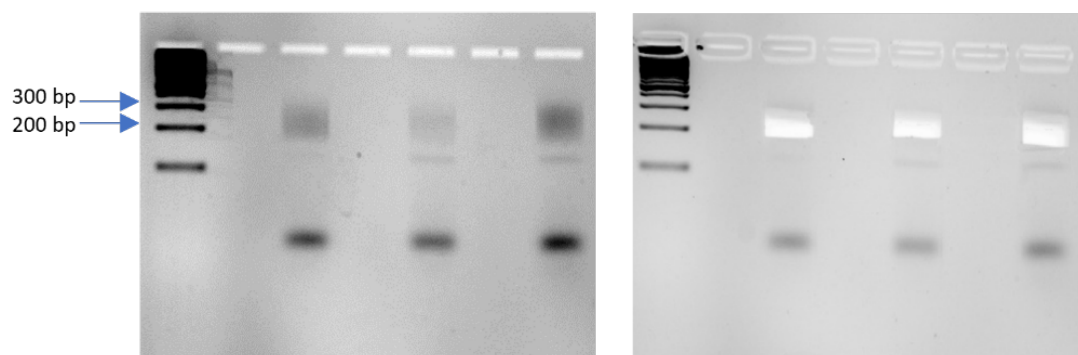
Figure 2.3: Representation of DNA samples of ChIP-seq libraries run on 3 % agarose gel.

The following figure illustrates two images of DNA samples from ChIP-seq libraries that have been run on a 3% agarose gel. As shown in the above image, the lane indicated by the blue arrow indicates the size of DNA fragments selected. On the right, DNA fragments are fragmented to a certain range of sizes in order to enable efficient sequencing.

2.18 Methylated DNA immunoprecipitation (MeDIP)

For genomic DNA extraction, cells were pelleted, washed with PBS, and then resuspended in lysis buffer (20 mM Tris pH8, 4 mM EDTA, 20 mM NaCl, 1 % SDS) and incubated overnight with 200 μg/ml Proteinase K at 55 °C. One volume of phenol was added to the samples and tubes were inverted 5-7 times then centrifuged at high speed for 20 mins. Phenol was removed and Phenol:Chloroform:Isoamyl alcohol (ratio 25:24:1; PCI) solution was added and mixed by inversion followed by centrifugation at 300 g for 20 mins. PCI solution was removed, and chloroform was added, mixed and centrifuged. The aqueous phase was collected and transferred into a new tube with the addition of 1/10 volume of 5 M NaCl and 1 volume isopropanol. Precipitated DNA was transferred to a new tube and resuspended in 1X TE buffer. Following this, samples were sonicated to fragment the DNA. This was monitored by running sonicated samples at intervals on an agarose gel to generate the ideal band size for precipitating the DNA (~300bp) (Figure 2.4). On a 1% agarose gel, 10μl of each sample was placed, along with 2μl of Orange G dye, and a DNA ladder, the gel was run at 25V for an hour before being imaged.

The selected sample was divided and placed in five Biorupter microtubes. Each microtube contained 20 μl of DNA, and were sonicated for a period of 5 minutes, 10 minutes, 15 minutes, 20 minutes, and 25 minutes, respectively. After this, a gel electrophoresis was performed in of further cycles required (Figure 2.5a). Another DNA sample of 40μl were sonicated for 20 minutes and 25 minutes (Figure 2.5b). Finally, after loading 8μg of DNA at a concentration of 500ng/l and sonicated for 30 minutes, the correct band size was achieved (Figure 2.5c).
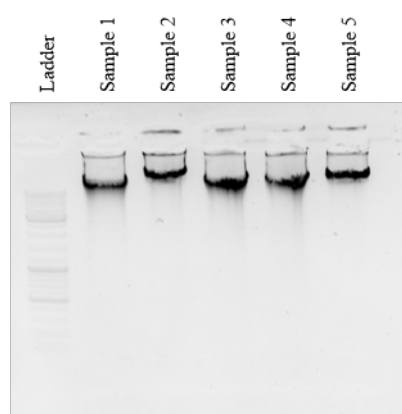
Figure 2.4: DNA samples, diluted in TE buffer and loaded on 1% agarose gel.

MeDIP was performed using the immunoprecipitation method described in section 2.6. Briefly, Dynabeads coated with Protein G were incubated with antibody recognising 5-Methylcytosine, followed by incubation with DNA samples and washing with IP buffer to remove unbound DNA. Samples were resuspended in digestion buffer (50 mM Tris pH8, 10 mM EDTA, 0.5 % SDS). Pulled-down DNA was isolated using AMPure XP beads (Beckman Coulter). The MeDIP libraries were prepared according to the Illumina protocol and sequenced with an Illumina Nextseq 500 sequencer.
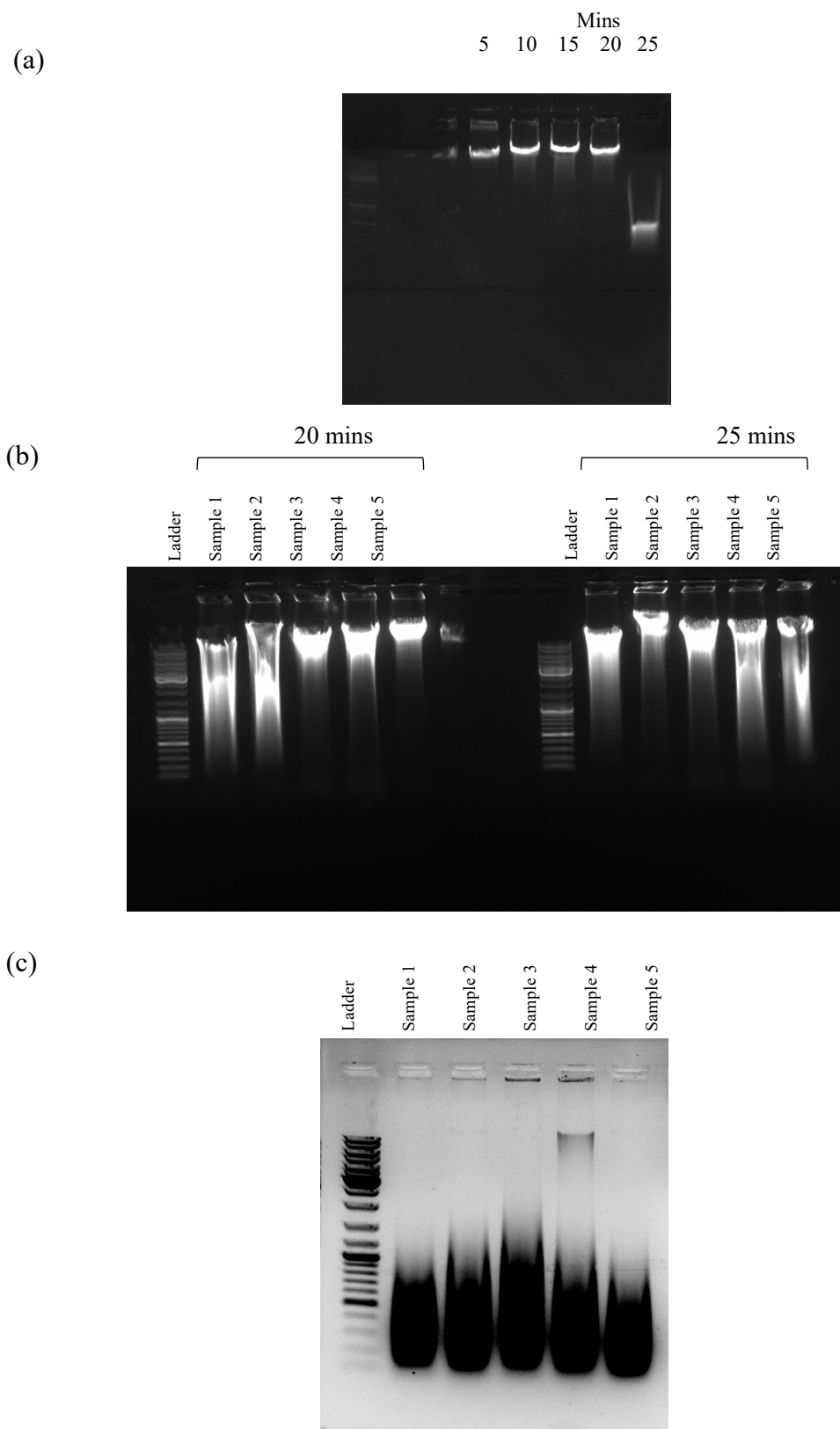
Figure 2.5: Representation of gel electrophoresis images for detecting the correct band size of DNA fragments.

DNA samples were run on 1 % agarose gel and underwent different sonication periods. (a) DNA sample divided into five tubes (20 μl each) and sonicated for different time period. (b) all 5 samples (40 μl) loaded and sonicated for 20 min and 25 min. (c) the correct size was achieved when loaded 8 μg of DNA at a concentration of 500 ng/l and sonicated for 30 minutes.

2.19 DNaseI hypersensitive site (DHS)

For the identification of DNaseI Hypersensitive Sites (DHS), cells were washed using ψ buffer (11 mM KPO4 pH 7.4, 108 mM KCl, 22 mM NaCl, 5 mM MgCl2, 1 mM CaCl2, 1 mM DTT), then resuspended in ψ buffer containing 1 mM ATP. Afterward, DNaseI mix was prepared using 80 μl of ψ buffer containing 1 mM ATP, 4 μl 10 % NP40 and 20 μl H_2O, containing series of DNaseI concentration (including 0 DNase, 1, 2, 3, and 4), 104 μl of the mix were added to every 100 μl of the cell suspension and incubated at RT for exactly 6 min. Reactions were terminated by the addition of 200 μl lysis buffer (100 mM Tris pH 8.0, 40 mM EDTA, 2 % SDS, 200 μg/ml proteinase K). The samples were incubated at 55 °C overnight to facilitate the digestion of proteins by proteinase K. The next day, 50 μg/ml of RNase A was added to samples, followed by incubation at 37 ° C for 1 h. After that, the sample was extracted twice using phenol/chloroform/isoamyl alcohol (PCI; 25:24:1) and once using chloroform. The DNA was precipitated by the addition of 0.5 M NaCl and 1 volume of isopropanol and washed with 70 % ethanol. The DNA was dissolved in 0.1x TE. Samples were loaded onto a 0.8 % agarose gel to isolate DNA fragments (size) using a QIAGEN mini-elute gel extraction kit according to manufacturer's protocol. Library preparation was carried out using the Illumina library preparation protocol as described in ChIP-seq library preparation section 2.17.

2.20 Genome-wide Bioinformatics Analysis

ChIP-seq, DHS and MeDIP data were analysed using the online European Galaxy resources https://usegalaxy.eu and https://usegalaxy.org. (Giardine et al., 2005, Blankenberg et al., 2010, Goecks et al., 2010). Briefly, reads were mapped to the human reference genome: GRch38/hg38 with HISAT2 2.0.9 (Trapnell et al., 2012). BAM tools were used to merge and sort the HISAT results, and MACS2 was used to identify the peaks (Zhang et al., 2008). The generated bed graph files were uploaded to the UCSC genome browser at https://genome.ucsc.edu/ and screenshots were taken (Kent et al., 2002). MEME-ChIP (https://meme-suite.org/meme/) was used to identify enriched de novo motifs. Heat maps were then generated with Easeq 1.5.0.0 software by plotting the BAM file data on a reference interval file. The Venn diagrams were produced using Venny version 2.1 application (http://bioinfogp.cnb.csic.es/tools/venny/). The Gene Ontology (GO) analysis was carried out using DAVID 6.8 (Huang et al., 2009). The first 10 categories were presented after filtering out redundant terms based on GO terms for biological processes with a p-value of 0.05. The Genomic Regions Enrichment of Annotations Tool (GREAT) was used to conduct gene association analysis (McLean et al., 2010b).

Chapter 3 GFI1 as HEB interacting partner in human T-cell Acute Lymphoblastic Leukaemia

3.1 Introduction

T-cell acute lymphoblastic leukaemia (T-ALL) is a type of haematological cancer caused by the malignant transformation of thymocytes. It is an aggressive cancer that is characterised by changes in the expression of several tumour suppressors and oncogenes, resulting in disruption of the normal development of T-cell progenitors (Ferrando et al., 2002b, Pui et al., 2008, Pui et al., 2014). Molecular changes as well as cytogenetic anomalies contribute to T-ALL pathogenesis, resulting in the disruption of specific cellular processes, for example cell cycle, signalling, proliferation, chromatin alteration, differentiation of T cells, and self-renewal.

In addition, aberrant transcriptional regulation was noted in T-ALL cases, involving genes such as *LMO2, TAL1, GATA3,* and *RUNX1*, culminating in an alteration of gene expression profiles in the cell that further contributes to the disease's development and progression (Van Vlierberghe et al., 2008, Zhang et al., 2012a).

The LMO2 complex is known to regulate haematopoietic cell differentiation. LMO2 and TAL1 are expressed in DN1 and DN2 immature thymocytes. Normal thymocyte maturation through the DN3, DN4 and DP stages requires suppression of *TAL1* and *LMO2*. The aberrant expression of LMO2 or TAL1 results in the blockage of T-cell differentiation and the development of T-ALL (Tremblay et al., 2010, Begley et al., 1989). The LMO2 multiprotein transcriptional complex consists of a heterodimer TAL1 or LYL1 with an E-protein such as E2A or HEB, which binds to LMO2. LMO2 also binds Zn-finger transcription factor GATA and transcriptional modulator LDB1. To study the oncogenic potential of the complex, it is necessary to specify which exact proteins are present in the complex and identify if there are several different complexes that have different functions in regulating normal and oncogenic gene expression. A mass spectrometry analysis of co-immunoprecipitation of TAL1 in four

74

human T-ALL cell lines that overexpress TAL1 confirmed the TAL1/HEB interaction (Omair, 2019). Further HEB IP/mass spectrometry experiments identified a potentially new member of the complex, the transcription factor GFI1.

E-proteins HEB and E2A are essential for maintaining the DP fate and preventing premature differentiation into the single positive stage (Zhang et al., 2012). This is at least partly due to the regulation of the transcriptional inhibitor *Gfi1*. In murine models where *Tcf3* (E2A) and *Tcf12* (HEB) were disrupted, E-proteins were demonstrated to activate the expression of *Gfi1* at the DP stage, which is then significantly downregulated at the single positive stage (Schwartz et al., 2006b). Furthermore, GFI1 suppresses the expression of the genes that are expressed in the single positive, on its own or in conjunction with E-proteins (Wang et al., 2007, Yücel et al., 2003).

Aim

It has recently been revealed that GFI1 can act as a transcriptional activator when it binds together with transcription factor IKAROS, or it can act as a transcriptional repressor by recruiting KDM1A (lysine-specific demethylase 1A, also known as LSD1) which is responsible for H3K4 demethylation. However, very little is known about the role of GFI1 in T-ALL, and this chapter aims to investigate this relationship. Understanding how GFI1 can influence the expression of genes associated with T-ALL can help to develop novel treatments for the disease.

3.2 Results

As our interest is in the TAL1/LMO2 complex, we first performed qPCR experiments to determine the mRNA expression levels and western blot experiments to determine protein expression levels in these cell lines. The relative expression in qPCR experiments was calculated by normalising to the TATA box binding protein (TBP), a stable housekeeping gene, as a control. For *TAL1*, qPCR data showed that it was expressed at high levels in DU528 and, to a lesser extent, in ARR, HSB2 and CRRFCEM (Figure 3.1 A). The western blot of TAL1 showed a difference in expression compared to qPCR data (Figure 3.1 B). Different bands in the western blot indicated the expression of different isoforms of *TAL1*, which has four major isoforms: 42, 39, 22 and 20 kDa (Jin et al., 2016, Calkhoven et al., 2003). The DU528 cell line expressed a high mRNA level of *TAL1* that can be explained by the SIL-TAL deletion DU528 carries. In contrast, full-length TAL1 protein levels in DU528 cells are low, whereas it expresses the short 22 kDa form of TAL1. This has been detected previously and could be due to different reasons, such as post-transcriptional mechanisms or protein degradation. The other three cell lines show dominant expressions of the two longest isoforms.

For *LMO2*, DU528 showed the highest mRNA expression level between the T-ALL cell lines and the strong band in the western blot analysis, followed by CCRFCEM. The level of *LMO2* mRNA expression in T-ALL cell lines was correlated with the level of LMO2 protein expression.
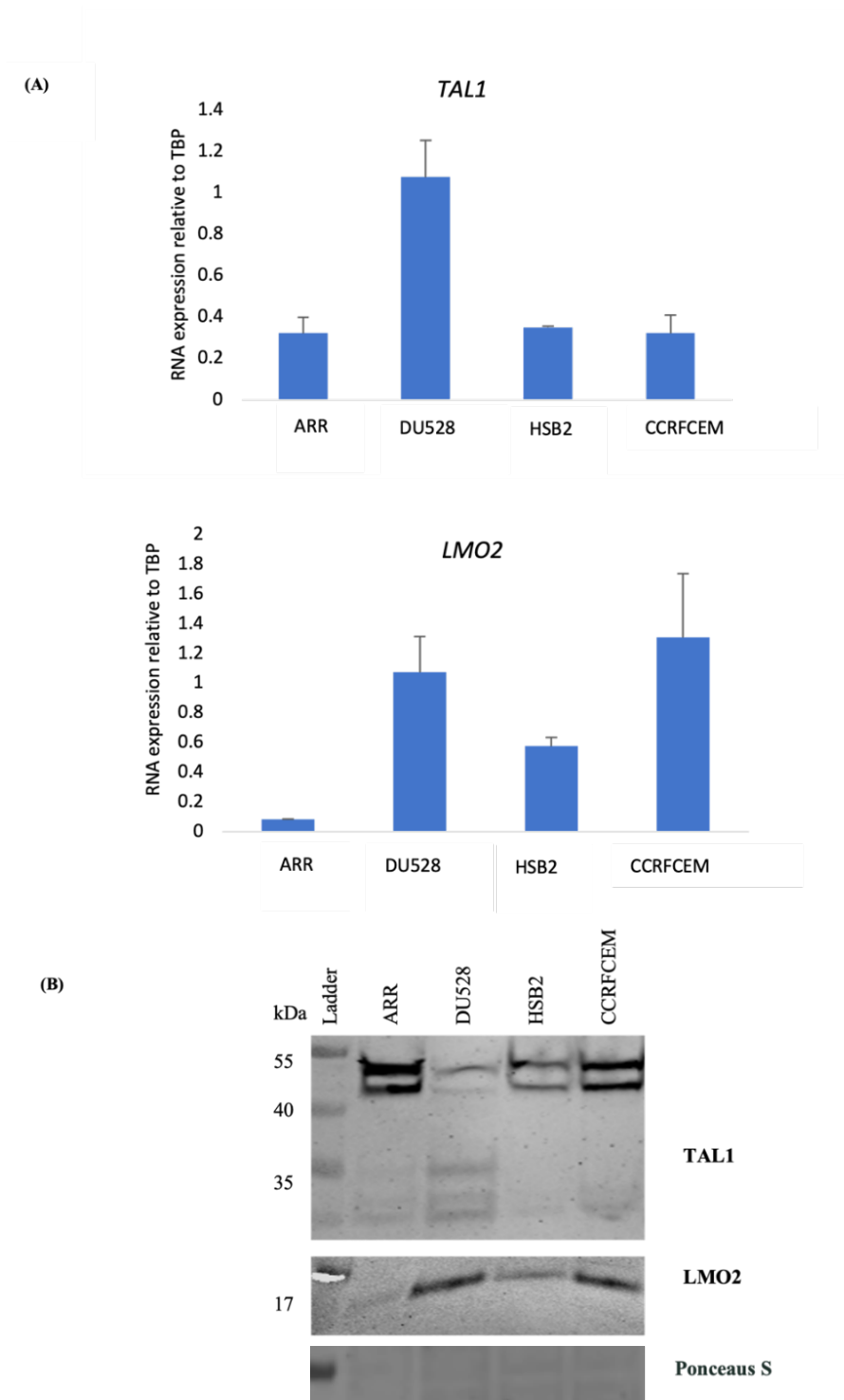
Figure 3.1: Expression levels of TAL1 and LMO2 mRNA and protein in T-ALL cell lines.

The figure shows (**A**) *TAL1* and *LMO2* relative mRNA expression levels measured by qPCR using TBP as a control (n=3). (**B**) TAL1 and LMO2 (17 kDa) protein level in T-ALL cell lines determined by western blotting (representative of three biological and technical repeats); Ponceau S was used to confirm equal loading.

Previous work done in the Hoogenkamp laboratory, performing HEB co-immunoprecipitation assays using T-ALL nuclear extracts followed by mass spectrometry, identified the presence of GFI1 in ARR and DU528 cell lines as a potential interacting partner of HEB. Additionally, the comparison of HEB ChIPseq data and methylated DNA immunoprecipitation in the T-ALL cell line ARR showed binding of HEB at regions containing CpG DNA methylation. Additionally, *de novo* motif analysis identified that HEB-bound regions are enriched for the GFI1 DNA binding motif. This indicated that GFI1 and HEB are binding to the same DNA elements and possibly directly interacting.

To study the role of HEB in T-ALL and its possible interaction with GFI1, we first decided to confirm the presence of mRNA and protein in different developmental stages of human T-cell development and T-ALL cell lines.

TCF12 is expressed in T-ALL cell lines.

To investigate the expression of *TCF12* in T-ALL cell lines, we performed quantitative PCR and western blot analyses. *TCF12* mRNA expression was quantified and normalised to TATA box binding protein. The primers used were designed to detect both isoforms, HEBcan and HEBalt. The qPCR results for *TCF12* revealed that it was expressed in all four cell lines (Figure 3.2 A). The highest *TCF12* expression was observed in the CCRFCEM, followed by ARR, HSB2 and DU528.

Western blot analyses on nuclear extracts from T-ALL cells showed that all four T-ALL cell lines express HEB canonical (HEBcan) (100 kDa). The results were in line with the qPCR results, so CCRFCEM had the highest HEB protein level, ARR and DU528 have both isoforms, HEBcan and HEB alternative (HEBalt), while the HEBalt (55 kDa) was not present in HSB2 or CCRFCEM (Figure 3.2 B). This could be because neither HSB2 nor CCRFCEM are most differentiated and HEBalt is expressed at a high level at early stages and then downregulated as T-cells develop (Wang et al., 2006a).

RNAseq data, previously obtained by the Hoogenkamp lab, from CD34$^+$ cord blood cells *in vitro* differentiated to the specific T-cell developmental stages from CD34$^+$, CD7$^+$/CD5$^-$, CD5$^+$/CD1a$^-$, CD1a$^+$/CD3$^-$ and CD3$^+$ and all four T-ALL cell lines, showed that *TCF12* is expressed not only in T-ALL but also at all stages of normal T-cell development (Figure 3.2 C).

Additionally, we investigated the cellular localisation of HEB in T-ALL cell lines by immunofluorescence staining. T-ALL cell lines were fixed, permeabilised and stained with antibodies against HEB and LMO2 and confocal microscopy was used to visualise the cells. DAPI was used as a fluorescent stain for nuclear staining. The results showed that HEB co-localised with LMO2 in the nucleus for all four T-ALL cell lines. Furthermore, we observed cytoplasmic staining in ARR, and this pattern was consistent for both LMO2 and HEB, providing strong confirmation of their co-localisation within the cell (Figure 3.3).
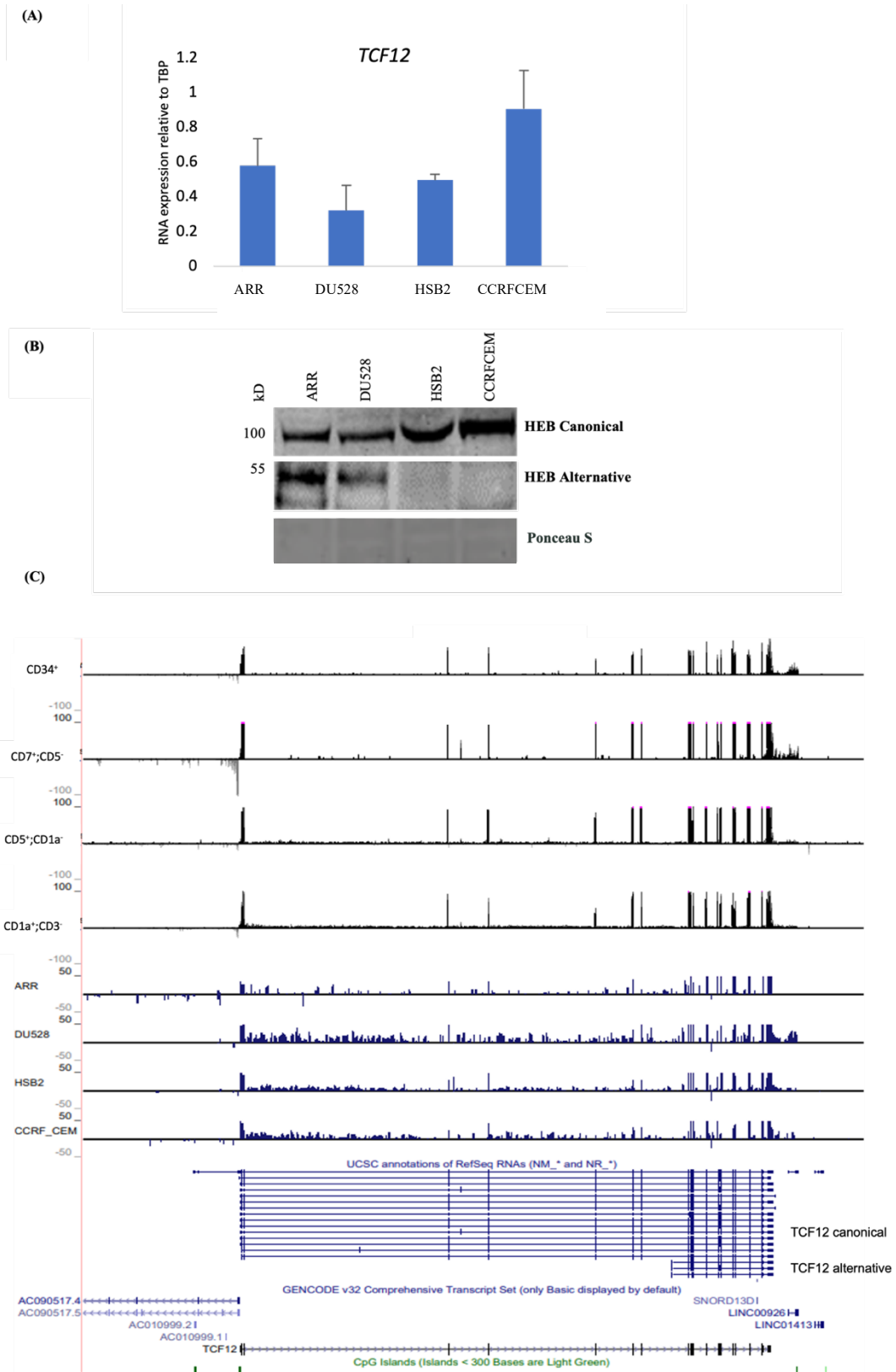
Figure 3.2: Expression profiles of *TCF12* mRNA and HEB protein in T-ALL cell lines.

(A) Bar graph showing the relative expression level of *TCF12* mRNA determined by qPCR using TBP as a control (n=3). **(B)** HEB protein expression levels in T-ALL cell lines determined by western blotting; Ponceau S was used to confirm equal loading (representative of three biological and technical repeats). **(C)** RNA-seq data of *TCF12* locus in normal T-cell development stages (CD34$^+$, CD7$^+$; CD5$^-$, CD5$^+$; CD1a$^-$, CD1a$^+$; CD3$^-$) and T-ALL cell lines using UCSC genome browser.

80
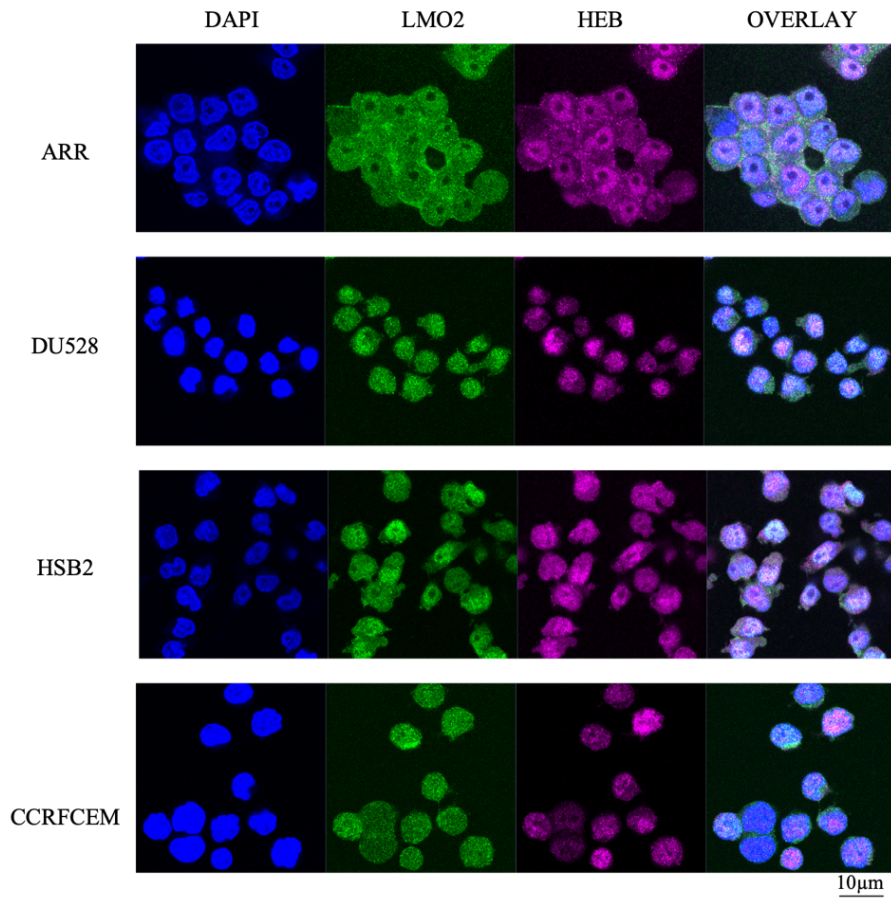
Figure 3.3: HEB and LMO2 co-localisation in T-ALL.

T-ALL cell lines were fixed and stained with anti-HEB (magenta) and anti-LMO2 (green) antibodies. Slides were imaged using immunofluorescence confocal microscopy, DAPI (blue) staining was used to visualise nuclei. Scale bar 10 μm (n=3).

In T-cell development, *GFI1* mRNA was upregulated gradually as the cells differentiated. In mice, it has been shown to be expressed in the first two stages, DN1 and DN2, with an upregulation at the DN3 stage when GFI1 is required for beta-selection (Möröy, 2005) (Möröy et al., 2015).

GFI1 qPCR showed the lowest mRNA expression level in ARR and the highest expression in CCRFCEM, which is consistent with the RNAseq data. GFI1 western blot results correlated with the mRNA expression levels. The RNAseq from the T-cell differentiation stages showed that *GFI1* mRNA was present throughout the development. It was the lowest in $CD34^+$ haematopoietic progenitors and the highest at the $CD7^+/CD5^-$ stage (Figure 3.4).

We also measured the expression of the closely related transcription factor *GFI1B*. Quantitative PCR showed that it was highly expressed in DU528, followed by ARR and HSB2, with low to undetected expression in CCRFCEM (Figure 3.5). This aligns with the literature, which shows that *GFI1B* is highly expressed in early haematopoietic stem cells and downregulated gradually upon differentiation. The GFI1B western blot results were correlated with the qPCR results, and the RNA-seq results confirmed the T-ALL cell line results.
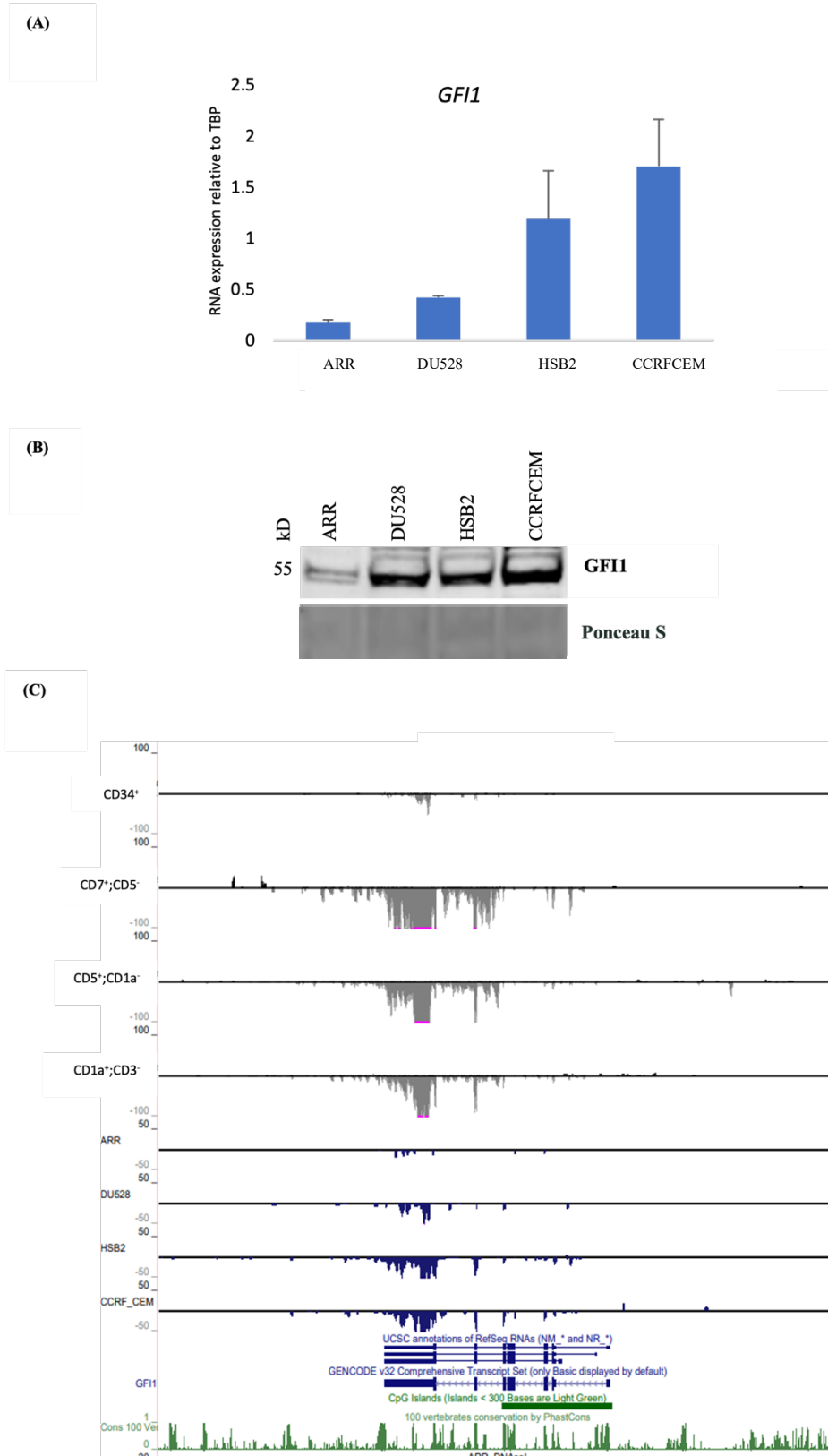
Figure 3.4: GFI1 mRNA and protein expression levels in T-ALL cell lines.

The figure shows **(A)** *GFI1* relative mRNA expression levels determined by qPCR using TBP as a control (n=3). **(B)** GFI1 protein expression levels in T-ALL cell lines determined by western blotting using GFI1 antibody (representative of three biological and technical repeats); Ponceau S was used to confirm equal loading. **(C)** RNA-seq data of GFI1 locus in normal T-cell development stages (CD34[+], CD7[+];CD5[-], CD5[+];CD1a[-], CD1a[+];CD3[-]) and T-ALL cell lines using UCSC genome browser.

(A) *GFI-1b* relative mRNA expression levels determined by qPCR using TBP as a control (n=3). **(B)** GFI-1B protein expression levels in T-ALL cell lines determined by western blotting; Ponceau S was used to confirm equal loading (representative of three biological and technical repeats). **(C)** RNA-seq data of GFI1B locus in normal T-cell development stages (CD34$^+$, CD7$^+$;CD5$^-$, CD5$^+$;CD1a$^-$, CD1a$^+$;CD3$^-$) and T-ALL cell lines using UCSC genome browser.
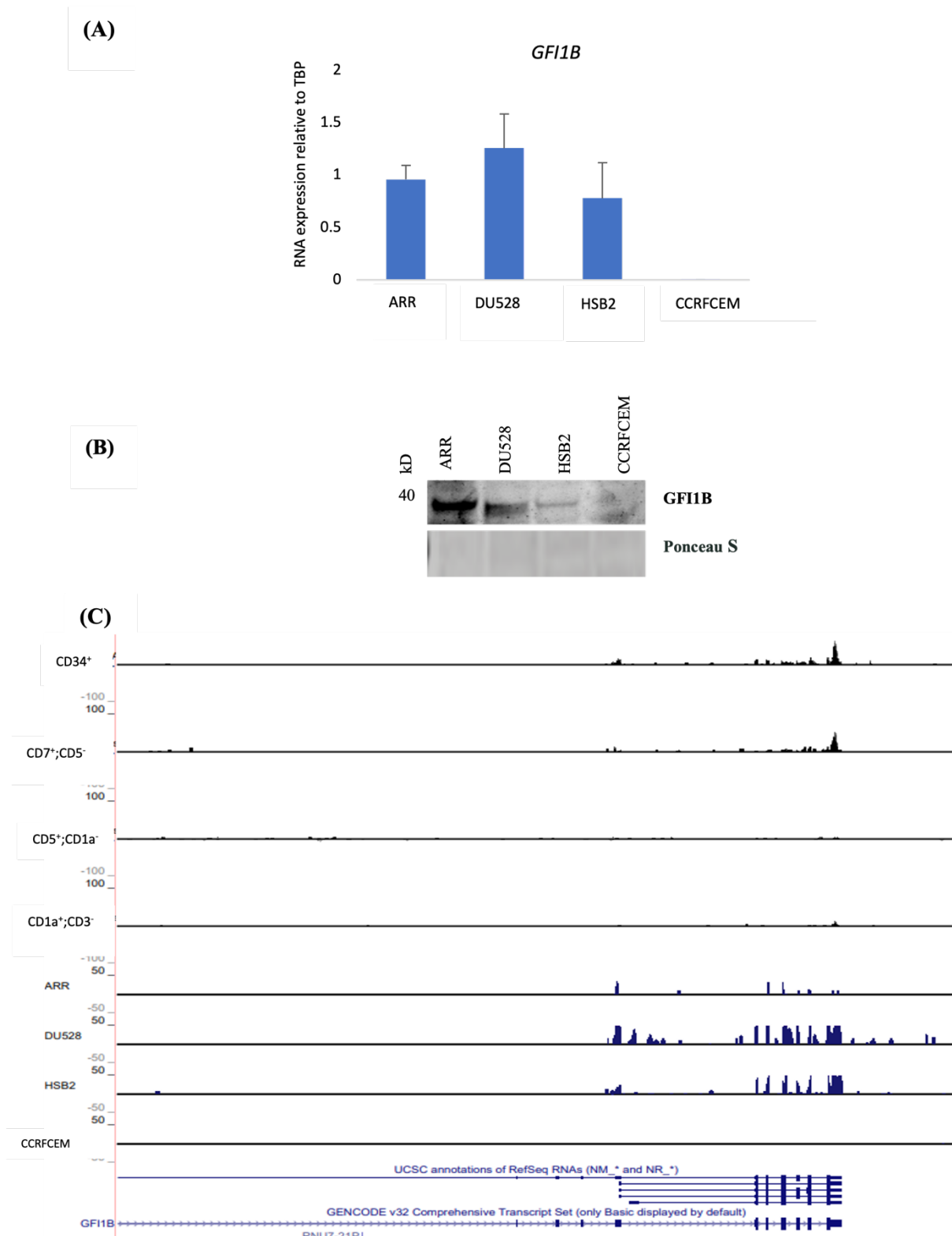
Figure 3.5: GFI-1b mRNA and protein expression levels in T-ALL cell lines.

As our data demonstrated that HEB and GFI1 are expressed in all four of our T-ALL cell line models, we proceeded to use them to investigate the possible HEB-GFI1 interaction by pulldown assays, followed by mass spectrometry. As immunofluorescent staining showed that HEB was present in the cytoplasm as well as in the nucleus, we decided to use total protein extracts. The protein extracts were incubated with magnetic beads coated with HEB or IgG antibodies. An acrylamide gel was used to separate the proteins from the pulldowns. Western blots probed with HEB, GFI1 or LMO2 antibodies were used to show the presence of the possible interactions (Figure 3.6).
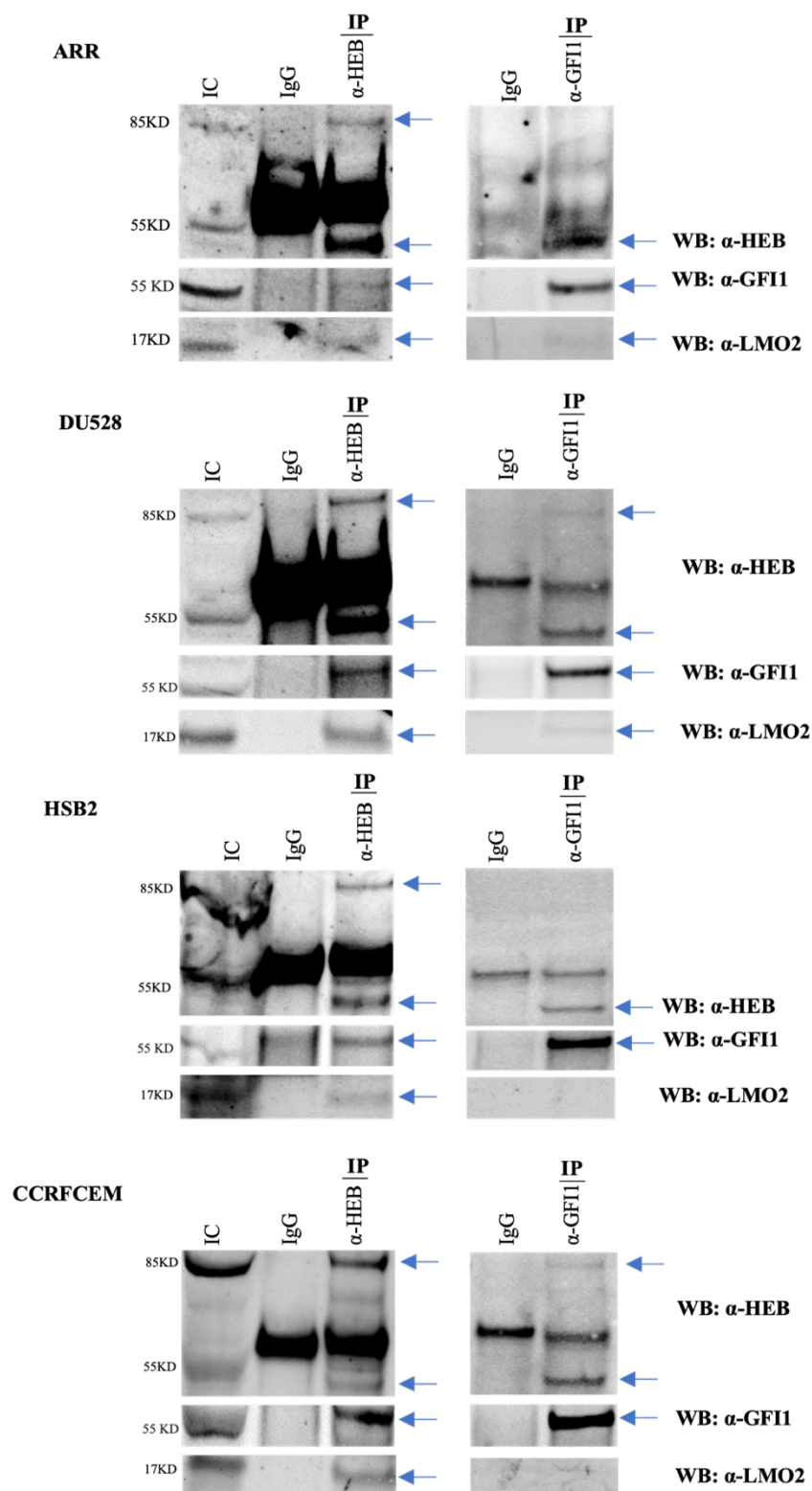
Figure 3.6: Co-immunoprecipitation of HEB and GFI1 using an extract from T-ALL cell lines.

The figure illustrate Co-immunoprecipitation followed by Western blotting to visualise the presence of HEB, GFI1 and LMO2. T-ALL protein lysates were immunoprecipitated with IgG protein-G incubated with beads coupled to IgG, αHEB or αGFI1. Protein extracts (100 ug) were used as a positive input control, mouse IgG antibody was used as a negative control for HEB, and rabbit IgG antibody was used as a negative control for GFI1 (representative of three biological and technical repeats).

86

In all four T-ALL cell lines, immunoprecipitation (IP) of HEB, followed by HEB Western blotting, confirmed the validity of the pulldown experiments by confirming that an aHEB antibody precipitates a HEB protein. The presence of two HEB isoforms, HEBcan (100 kDa) and HEBalt (55 kDa), was observed. ARR exhibited a more prominent band of HEBalt in comparison to HEBcan, while the CCRFCEM cell line showed a stronger band of the HEBcan isoform. An additional band at approximately 70 kDa was observed in the IgG control, as well as in the aHEB samples. The band of this size corresponds to the molecular weight of the IgGs, showing that our secondary antibody recognises both the HEB and aIgG antibodies used in the pulldown procedure.

The western blot analysis using the GFI1 antibody showed that the GFI1 protein (55 kDa) was detected in HEB but not in IgG IP samples in all four T-ALL cell lines These results verify the specificity of GFI1 enrichment and prove the interaction between HEB and GFI1.

In addition to this, we performed an LMO2 pulldown experiment to test if the LMO2 protein interacts with HEB and GFI1. The results showed that LMO2 (17 kDa) was present in all the cell lines tested and was detected (IC lane) in HEB IP, which confirms that LMO2 interacts with HEB in T-ALL cell lines. Taking into consideration the findings presented here, it appears that HEB interacts with GFI1, as well as with LMO2.

To establish if HEB, LMO2 and GFI1 are part of the same complex, we performed GFI1 IPs and used HEB, GFI1 or LMO2 antibodies to probe the western blots. Our results revealed that GFI1 successfully precipitates itself and HEBalt in all T-ALL cell lines, while much lower levels of HEBalt were detected in DU528 and CCRFCEM. Additionally, LMO2 was only observed in the pulldowns where ARR and DU528 cell extracts were used. At the same time, a negative control IgG IP showed only one band at 70kDa, confirming that the observed HEB-GFI1 interactions were specific.

Failure to observe GFI1/LMO2 interaction in the HSB2 and CCRFCEM cells even though LMO2 was present in the input control may indicate that LMO2's interaction with a GFI1-containing complex is much weaker and therefore more difficult to detect, or that LMO2 is not a part of the HEB-GFI1 protein complex in these cells.

Altogether, our results show that (1) HEB interacts with GFI1 and LMO2 in all four T-ALL cell lines, (2) GFI1 interacts mostly with HEBalt and (3) GFI1 interacts with LMO2 only in ARR and DU528 cells. These results suggest that HEBalt, GFI1 and LMO2 interact with each other only in ARR and DU528, the T-ALL cells that are blocked at the earlier stage of T-cell differentiation.

According to the previous HEB mass spectrometry analysis performed in our lab that detected the presence of GFI1, it was found that GFI1 was monomethylated at lysine 256. As GFI1 is known to interact with lysine methyltransferases such as G9a, we wanted to investigate this further (Duan et al., 2005). First, we tested for the presence of proteins carrying lysine aminomethylation in total cell extracts isolated from T-ALL cell lines 24 h and 48 h after the media change (Figure 3.7 A). We found a strong signal at 15kDa in all the samples, which corresponds to the size of histone proteins. The second strongest was a band at 55 kDa, which happens to be the size of the GFI1 protein. To test if this 55 kDa protein was GFI1, we employed an IP experiment using an antibody specifically recognising monomethylated lysine (Kme1) (Figure 3.7 B). The input controls confirmed that GFI1and LMO2 were present in all cell extracts. The Kme1 pulldown revealed the presence of GFI1-Kme1 in all cell lines, confirming the HEB IP-mass spectrometry results indicating that GFI1 has monomethylated lysine. Additionally, we detected LMO2 in ARR and DU528 cells. Controlling the GFI1 pulldown confirmed the GFI1 interaction with LMO2 in ARR and DU528.

Furthermore, we tested protein arginine methyltransferase 5 (PRMT5) and it was detected in both the GFI1 and Kme1 pulldowns of HSB2 and CCRFCEM cell. Protein arginine methylation is an important epigenetic regulation that is catalysed by protein arginine methyltransferases (PRMTs). Protein arginine methyltransferase 5 (PRMT5) is a type II PRMT that can symmetrically dimethylate histone and non-histone proteins, affecting chromatin regulation and gene transcription. It plays a critical role in T-cell development and has been shown to be expressed in different cancer types, such as lymphoma, glioblastoma and acute myeloid leukaemia. PRMT5 may repress transcription by the symmetrical dimethylation of the histone proteins H3 and H4, leading to chromatin remodelling (Hwang et al., 2021).
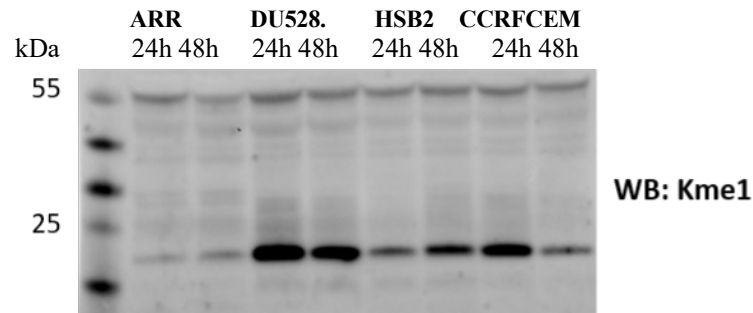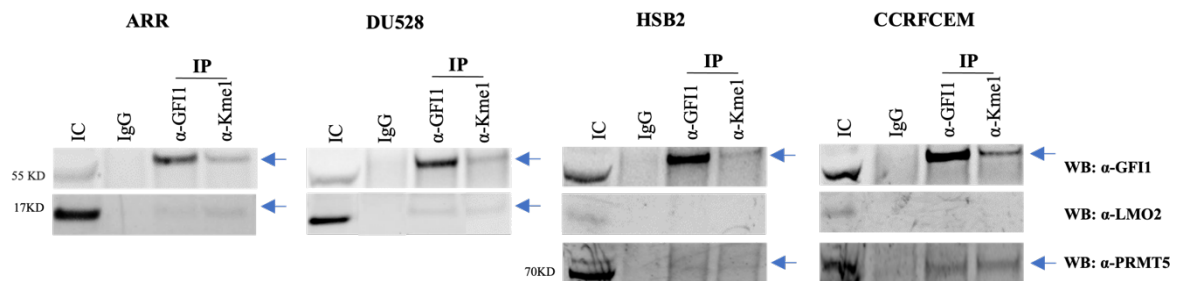
Figure 3.7: Co-Immunoprecipitation of GFI1 and aminomethylated lysine (Kme1) using extracts from T-ALL cell lines.

In the figure (A) represent western blot analysis using aminomethylated lysine (Kme1) antibody in T-ALL total cell extract 24h and 48h after the media change. The strong signal at 15 kDa correspond to the size of histone protein, the second strongest signal (55kDa) which is the size of GFI1 protein. (B) T-ALL protein lysates were immunoprecipitated with protein G incubated with beads coupled with IgG, αGFI1 or αKme1. Nuclear extracts were used as a positive input control and rabbit IgG antibody was used as negative. Co-immunoprecipitation was followed by Western blotting to visualise the presence of GFI1, LMO2, and PRMT5.

Therefore, in addition to observing HEBalt/GFI1/LMO2 in ARR and DU528 cells, we can further specify that GFI1 carries lysine monomethylation (GFI-Kme1). Complementary to this, PRMT5 protein was detected in pulldowns of GFI1 and Kme1 in HSB and CCRFCEM. So far, we have established that GFI-containing complex in ARR and DU528 cells consists of GFI1-Kme1/HEBalt/LMO2, while in HSB2 and CCRFCEM we could only detect the GFI1/HEBalt interaction. To identify if the GFI1 complex has other binding partners we performed protein pull-down, followed by mass spectrometry analysis. IPs were carried out for each of the four different T-ALL cell lines using antibodies directed against GFI1, Kme1 and IgG as a negative control. Following the pull-down, proteins were loaded on acrylamide gels and stained with Coomassie blue stain as shown in Figure 3.8. Sections were made and processed for mass spectrometry analysis. Our initial mass spectrometry experiments were conducted using both total and nuclear protein extracts (TE and NE) of the ARR cell line (Figure 3.9). The results were compiled after subtracting proteins detected in the IgG controls from the analysis.
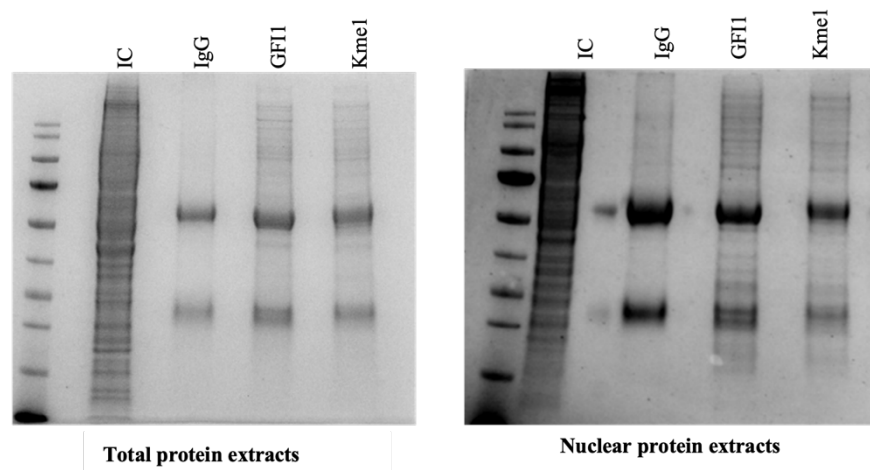
Figure 3.8: Representative of Coomassie blue stained gel used for mass spectrometry.

Coomassie blue gels showing pulldown of IgG, GFI1 and Monomethylated lysine (Kme1) using total protein extracts and nuclear protein extracts prepared from ARR.

In the experiment using total protein extracts, we identified 394 proteins from the pulldown for GFI1 and 336 proteins from the Kme1 co-IP, of which 217 were common to both experiments. In contrast to the results obtained using total protein lysates, IP involving nuclear protein extracts yielded different results. We detected 530 proteins interacting with GFI1 and 350 proteins interacting with Kme1. These two IPs shared 256 proteins. The data are represented graphically using a Venn diagram to illustrate the commonalities between the two experiments. As shown in Figure 3.9, each circle represents a set, and the overlapping region refers to the elements that belong to more than one set. We then compared the two different pulldowns of GFI1 from total protein extracts and nuclear protein extracts and found that 85 proteins were detected for both total and nuclear protein extracts. The second step consisted of conducting a STRING analysis of the mass spectrometry results to analyse the known interactions and relationship between the proteins based on published data. Among the proteins that are shared between the IPs of GFI1 and Kme1 using ARR total protein extracts, GFI1 exhibits only an interaction with protein Tyrosine Phosphate Receptor Type C. On the other hand, GFI1 was found to interact with RUNX1, RUNX2, RUNX3, HDAC1 and HAX1 among the common nuclear protein extracts between Kme1 and GFI1. As we were able to identify more proteins from the experiments employing nuclear extracts, we decided to proceed with IP/mass spectrometric experiments using only nuclear extracts.
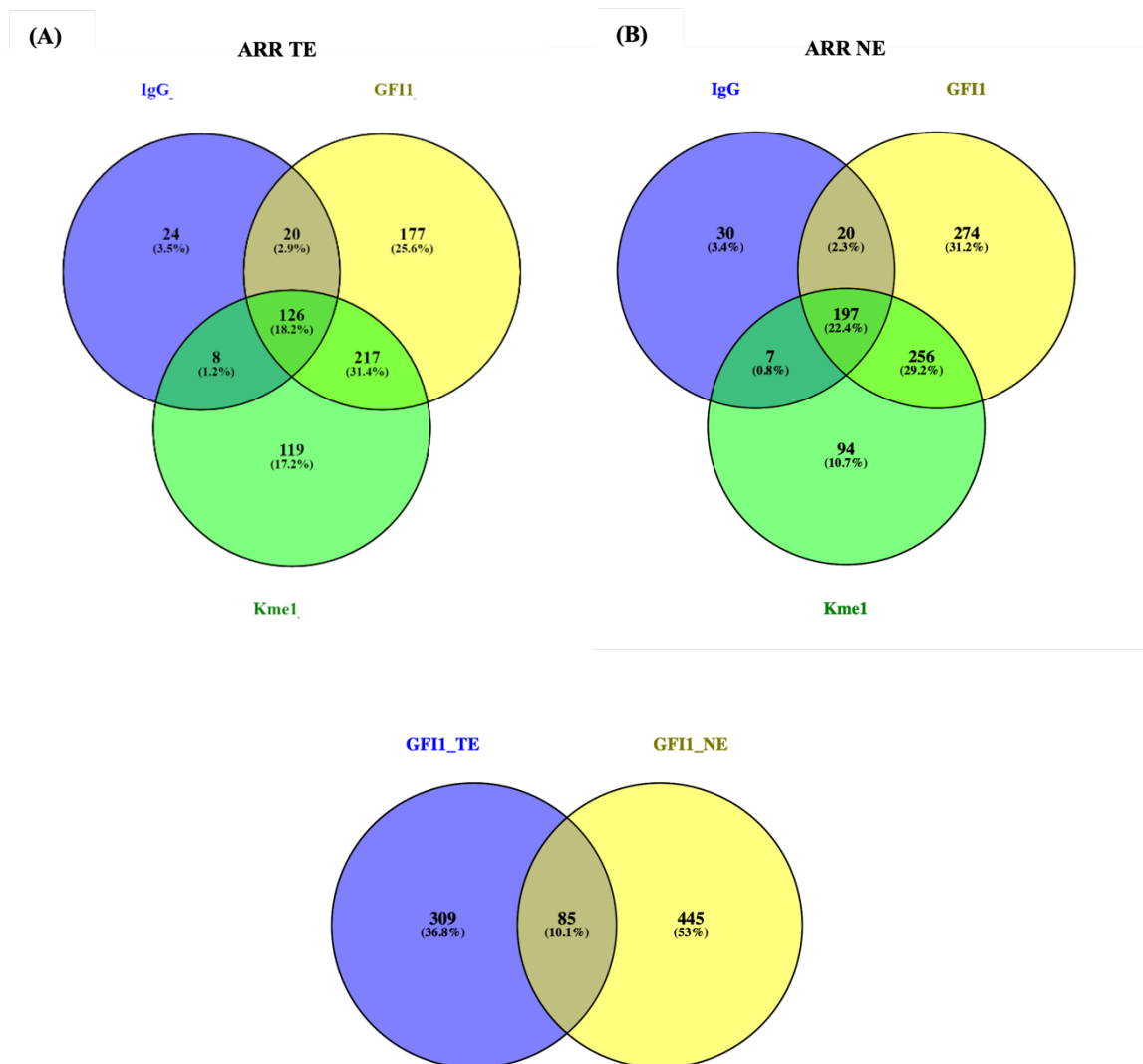
Figure 3.9: Venn diagrams of GFI1 and Mono-methylated Lysine (Kme1) mass spectrometry in ARR.

(A) the diagram demonstrates the comparison between GFI1 pulldown and Kme1 pulldown using total protein extracts (TE) (share 217). (B) Comparison between GFI1 pulldown and Kme1 pulldown using nuclear protein extracts (NE) (share 256). (C) Comparison between GFI1pulldown excluding IgG in total and nuclear protein extract.

IP of the remaining three T-ALL cell lines, DU528, HSB2 and CCRFCEM, using antibodies against GFI1, Kme1 and IgG was followed by resolving the precipitated protein on the Coomassie blue stained gels (Figure 3.10). Protein peptides identified by mass spectrometry analysis of GFI1, Kme1 and IgG were sorted in descending order by highest Mascot score. The peptides that were common between GFI1 and IgG, as well as Kme1 and IgG were eliminated, and only proteins that were exclusively found in GFI1 mass spectrometry or Kme1 mass spectrometry were selected, as shown by the Venn diagrams in Figure 3.11. Around 48% of all identified proteins were common between the two pulldowns in ARR, 48.9% in DU528, 40% in HSB2 and 32% in CCRFCEM. These findings suggest that a significant number of GFI1-interacting proteins also interact with proteins containing monomethylated lysine, which can be explained by the fact that a known GFI1 binding partner is lysine methyl transferase EHMT2 (Duan et al., 2005). Although we did not find EHMT2 in GFI1 IP-mass spectrometry, the Kme1 IP-mass spectrometry experiment with the CCRFCEM cell extract found EHMT2, and in ARR and HSB2 the closely related protein EHMT1 that has the same function in lysine methylation. Additionally, the lysine demethylase KDM1 (LSD1) was found in GFI1 and Kme1 IP with the CCRFCEM extract, confirming GFI1 involvement with lysine methylation.
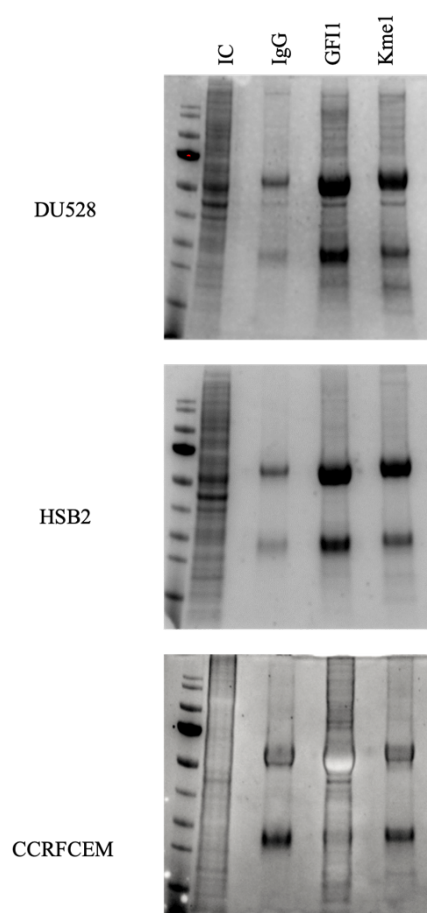
Figure 3.10: Coomassie blue gels of Mass spectrometry of GFI1 and Kme1 on DU528, HSB2 and CCRFCEM cell lines.
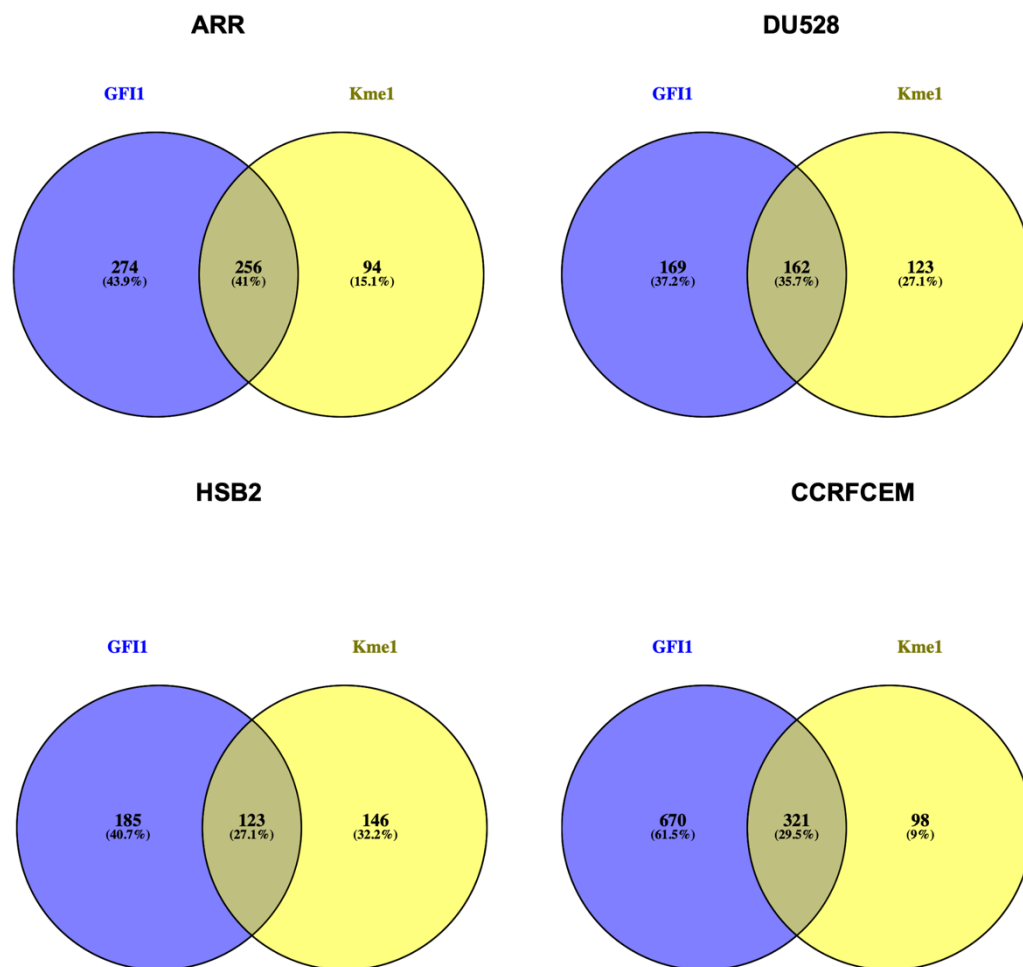
Figure 3.11: Overlap of common proteins detected in GFI1 and Kme1 MS analysis.

Venn diagrams illustrating the common shared proteins identified between GFI1 and Kme1 mass spectrometry (MS) analyses of four T-ALL cell lines.

Aiming to elucidate protein association networks among proteins identified through mass spectrometry analysis, the STRING database was used to assess known protein interactions within the GFI1 mass spectrometry dataset. First, we established a network of proteins that are known to interact in ARR, DU528, HSB2 and CCRFCEM, as shown in Figure 3.12 A, B, C and D, respectively. Proteins are represented by nodes, while known interactions are represented by connecting lines. The GFI1 protein is indicated by a red arrow in the mass spectrometry network.

In all four cell lines, a known interactor of GFI1, histone deacetylase 1 (HDAC1), was found. Additionally, in ARR, DU528 and CCRFCEM cells, GFI1 and HDAC1 interacting partners RCOR1 and NCOR1 were detected. Together, these proteins mediate histone deacetylation, chromatin compaction and epigenetic silencing of gene expression. These findings confirm that in T-ALL, GFI1 mediates suppression of gene expression through the known interactions with HDAC1 and N/RCOR complex. Further to this, GFI1 interactions with DNA-damage repair machinery through MRE11 and DNA phosphor kinase C PRKDS/DNPKc (all four T-ALL cell lines), p53 binding protein TP53B, RAD50 (ARR, HSB2 and CCRCEM), arginine methyl transferase PRMT1/ANM1 (HSB2, CCRFCEM), PARP1/2 (DU528, HSB2, CCRFCEM), Ku70,NBN (HSB2, CCRFCEM and Ku86 (ARR, DU528) were identified. Out of these MRE11, PRMT1/ANM1, PARP1/2, Ku70/86 and PRKDS were also found in the Kme1 IP. This multitude of interaction is in line with the previously reported finding that GFI1 facilitates efficient DNA repair through its interactions with PRMT1/ANM1, MRE11 and TP53B but also that the same mechanism is in action during the T-cell receptor VDJ rearrangements (Vadnais et al., 2018).

Additionally, an interaction with RUNX transcription factors (RUNX1, RUNX2, RUNX3) was found in ARR cells, while RUNX-binding partner PEBP2B/CBFβ was detected in three cell

lines. In CCRFCEM, we observed that GFI1 interacts with transcription factors TCF7, IKZF1, RUNX1, and LMO2 binding partner LDB1. Furthermore, the CCRFCEM mass spectrometry showed the presence of arginine methyl transferase PRMT5, which was detected in our previous IP-mass spectrometry data employing LMO2 antibodies and has since been experimentally confirmed.

A

C
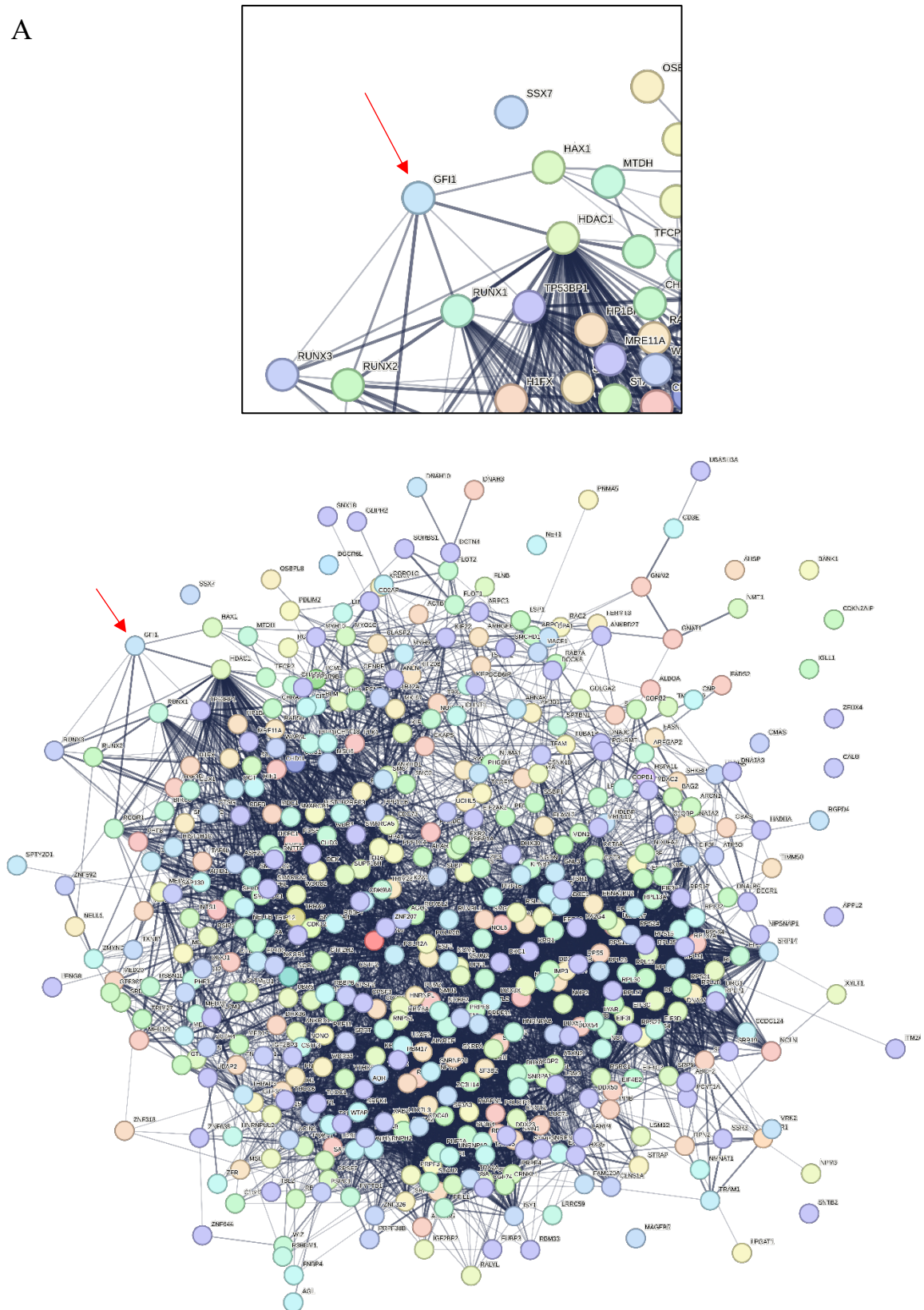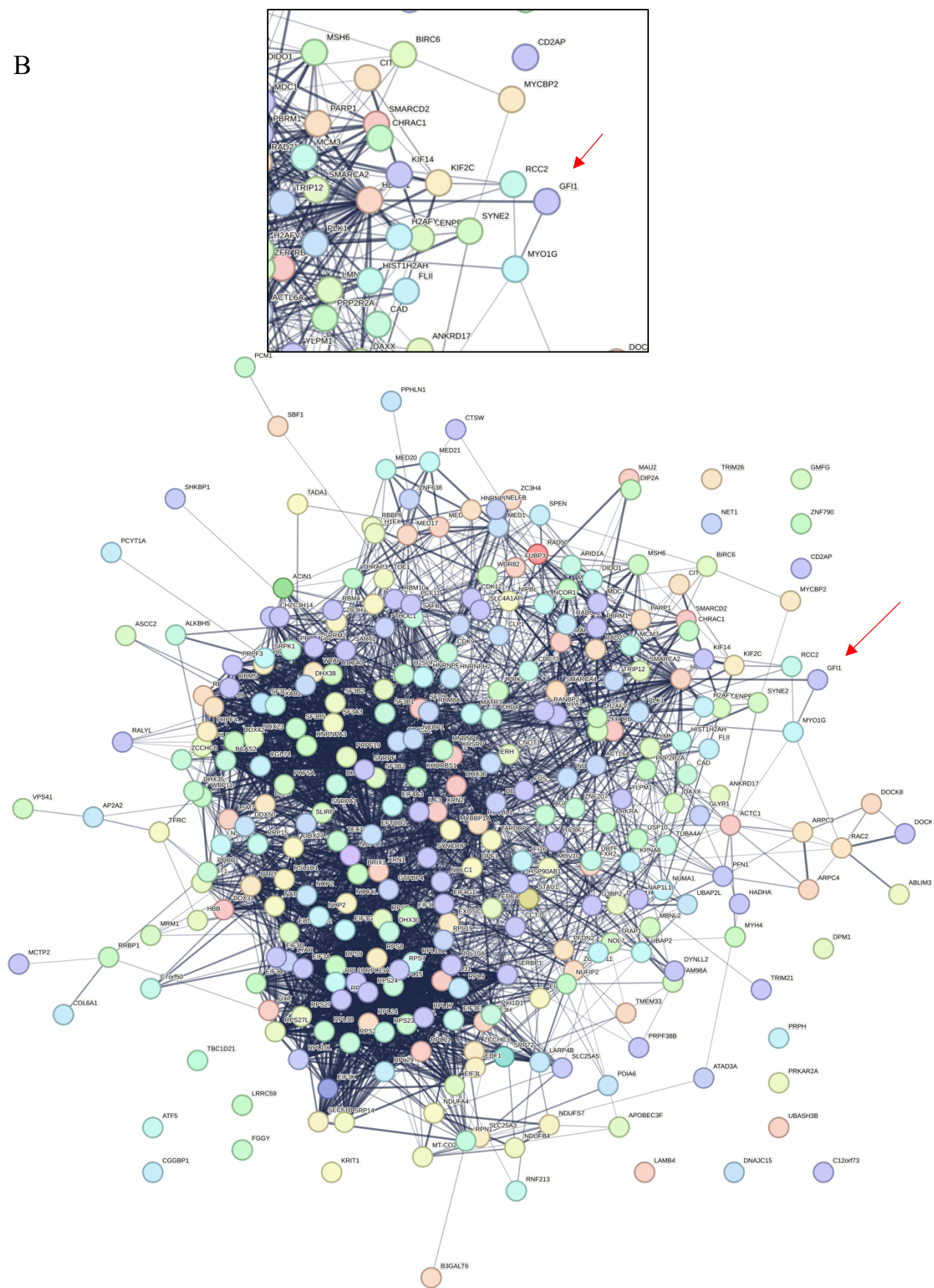
D
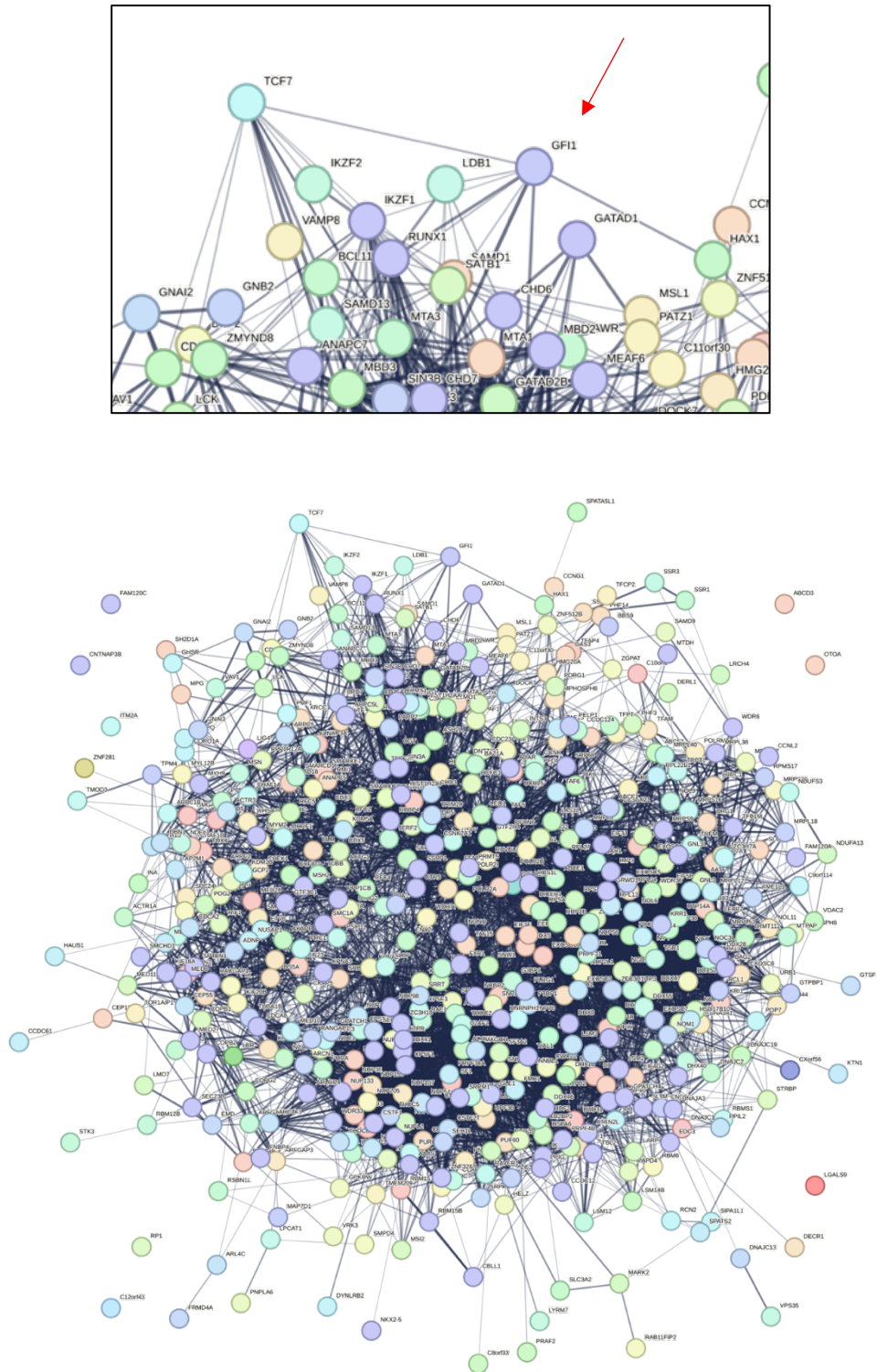
Figure 3.12: Networks of protein-protein known interactions in GFI1 MS.

Proteins detected by mass spectrometry analysis of GFI1 pulldowns were analysed for known interactions using STRING database. The top panel of each figure provides a detailed zoom-in view of the larger network depicted in the bottom panel. This close-up shows the interactions around GFI1 in (A) ARR, (B) DU528, (C) HSB2, and (D) CCRFCEM. Nodes demonstrate proteins, lines represent the known interactions, red arrow indicates GIF1 protein.

102

Afterward, we examined 218 proteins shared between at least three cell lines, (Figure 3.13). The analysis of protein interactions between the three cell lines identified the known interactions between GFI1 and HDAC1, as well as TP53BP1. Furthermore, we detected the presence of PEBP2B/CBFβ (the core-binding factor subunit beta).

Following this, we performed a four-way comparison of GFI1 mass spectrometry conducted on the T-ALL cell lines. The results revealed 42 proteins shared by all four lines (Figure 3.14). These 42 proteins are listed in Table 3.1. Out of the common proteins between the four cell lines, HDAC1 is known to directly interact with GFI1. We completed a functional annotation analysis of biological processes, which revealed that there was statistically significant enrichment for proteins involved in RNA splicing, p53 regulation of signal transduction, termination of translation, and rRNA processing. The known interactions between the 42 common proteins identified through the four-way comparison of T-ALL cell lines and biological processes are shown in Figures 3.15 and 3.16.
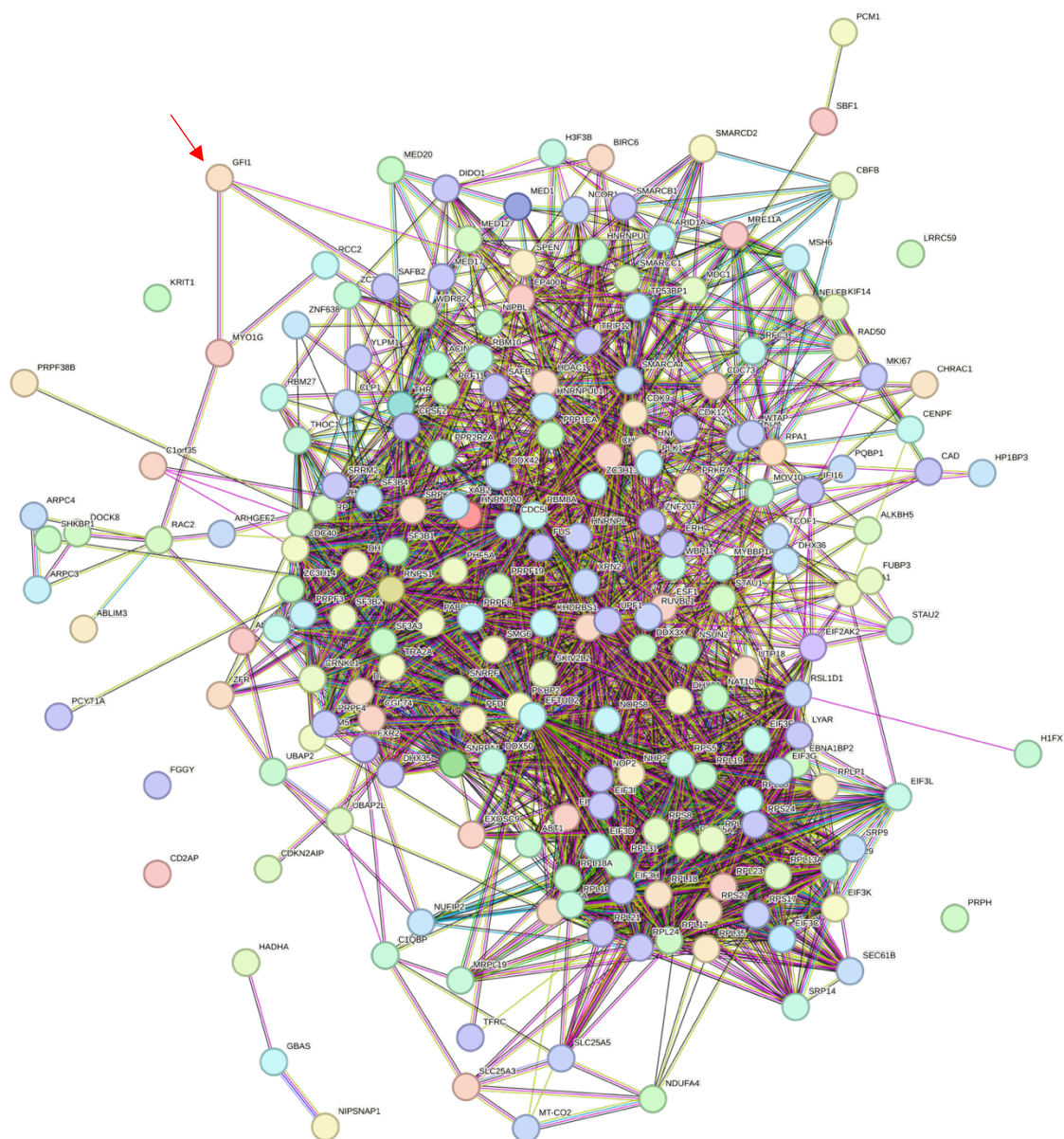
Figure 3.13: STRING analysis of common proteins between three T-ALL cell lines.
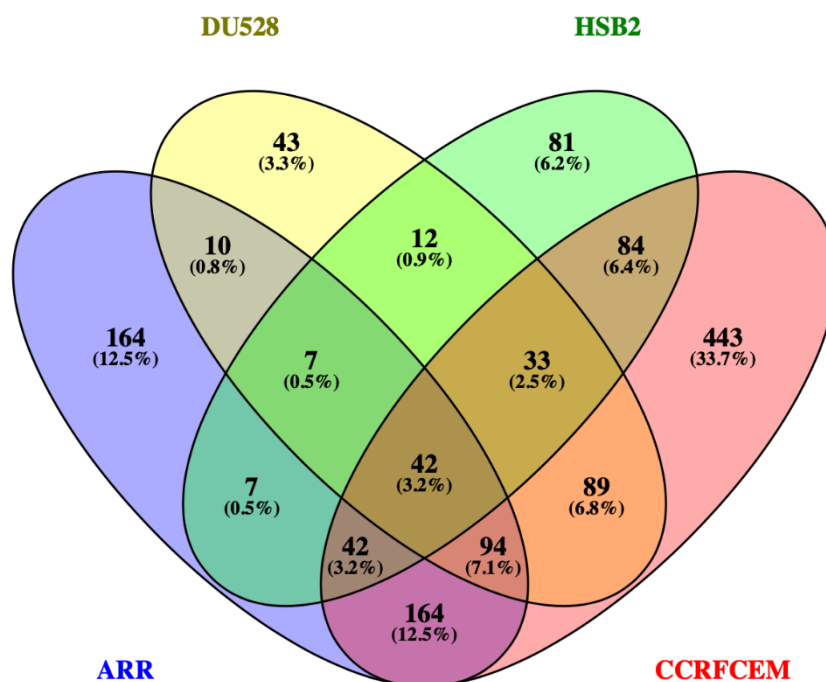
Figure 3.14: Four-way comparison of GFI1 MS on T-ALL cell line.

The diagram illustrates the quantitative comparison of common proteins between DU528, HSB2, ARR, and CCRCEM cell lines identified in GFI1 mass spectrometry analysis. Numbers indicates the count of shared proteins. intersecting areas indicate the count of shared proteins, with corresponding percentages.

Table 3.1: The list of 42 proteins common between all four cell lines in GFI MS analysis.

Protein	Mapped gene symbol
U5 small nuclear ribonucleoprotein component	U5S1
Death-inducer obliterator 1	DIDO1
Serine/arginine repetitive matrix protein 2	SRRM2
U2-associated protein SR140	SR140
Pre-mRNA cleavage complex 2 protein Pcf11	PCF11
Pre-mRNA-splicing factor ATP-dependent RNA helicase PRP16	PRP1
Putative ATP-dependent RNA helicase DHX30	DHX30
40S ribosomal protein S24	RS24
DNA-dependent protein kinase catalytic subunit	PRKDC
Zinc finger CCCH domain-containing protein 4	ZC3H4
Scaffold attachment factor B1	SAFB1
60S ribosomal protein L10a	RL10A
Zinc finger protein 638	ZN638
Double-strand break repair protein MRE11A	MRE11
Eukaryotic translation initiation factor 3 subunit D	EIF3D
Histone deacetylase 1	HDAC1
Cell division protein kinase 12	CDK12
60S ribosomal protein L13a	RL13A
Pericentriolar material 1 protein	PCM1
Eukaryotic translation initiation factor 3 subunit B	EIF3B
Polyglutamine-binding protein 1	PQBP1
WW domain-binding protein 11	WBP11
Polyribonucleotide 5'-hydroxyl-kinase Clp1	CLP1
60S ribosomal protein L17	RL17
Nuclear mitotic apparatus protein 1	NUMA1
Eukaryotic translation initiation factor 3 subunit G	EIF3G
Polyadenylate-binding protein 2	PABP2
Activator of basal transcription 1	ABT1
CDKN2A-interacting protein	CARF
KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDR1
Uncharacterized protein C17orf85	CQ085
ATP-dependent DNA helicase 2 subunit 2	KU86
40S ribosomal protein S5	RS5
Protein unc-84 homolog B	UN84B
Dedicator of cytokinesis protein 8	DOCK8
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin	SNF5
U3 small nucleolar RNA-associated protein 18 homologs	UTP18
Serine/threonine-protein kinase SRPK1	SRPK1
Putative ribosomal RNA methyltransferase NOP2	NOP2
Uncharacterized protein C21orf70	CU070
Histone H1x	H1X
60S ribosomal protein L15	RL15
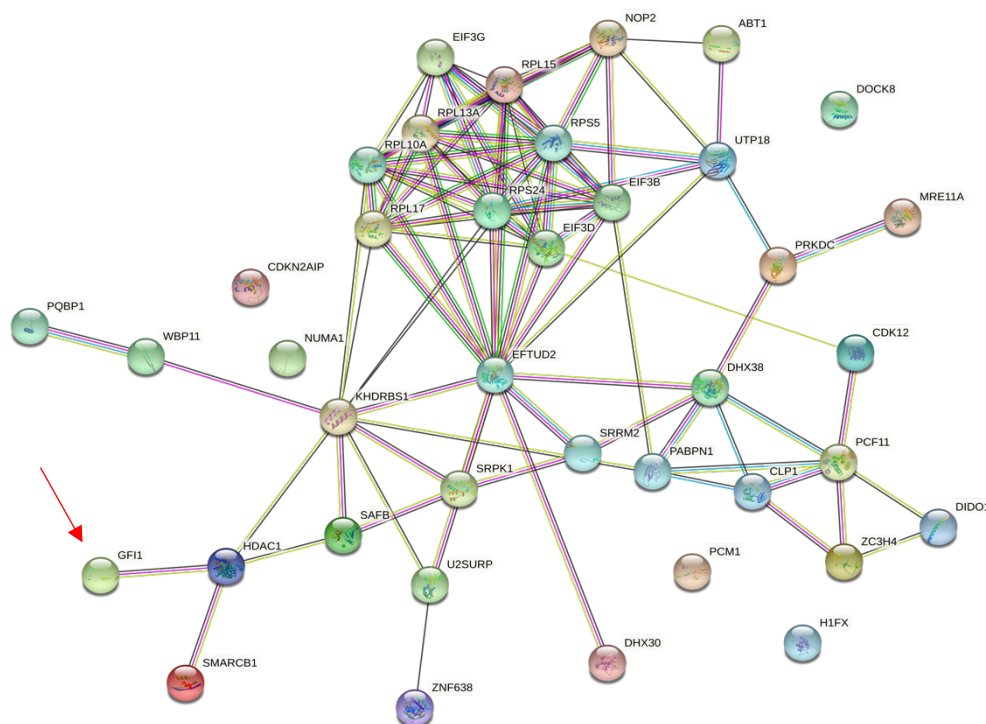
Figure 3.15: STRING analysis of 42 common proteins in GFI1 MS of T-ALL.

The figure demonstrates a visualization network generated by the STRING database of the 42 common proteins identified in the GFI1 mass spectrometry analysis of T-ALL cell lines. Nodes represent proteins, lines represent known protein-protein interactions.
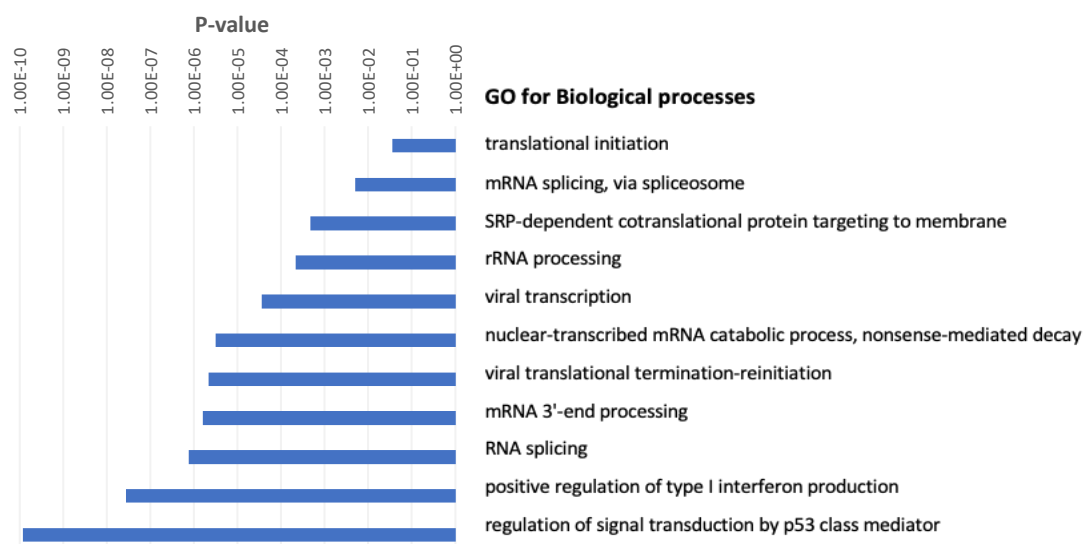
Figure 3.16: Enriched biological processes in GFI1 MS of T-ALL cell.

Bar chart illustrating the top biological processes significantly enriched in GFI1 MS of T-ALL cell lines.

3.3 Discussion

Proteins and their interactions control most biological processes in a cell, and they facilitate cells' fates by establishing cell-type-specific gene expression programmes. Haematopoietic differentiation is characterised by the activation of developmental stage-specific transcription factors at certain stages to regulate the transcription of genes specific to each stage. In addition to the genes they regulate, transcription factors form dynamic networks, with different transcription factors activating sequentially to drive differentiation as the lineage progresses (Goode et al., 2016). Thus, it is critical to understand the relationship between proteins and how they interact to result in progressive treatment responses. The development of targeted therapies will be greatly aided by a deeper understanding of molecular mechanisms and protein complex interactions. In this study, we describe a novel interaction between HEB and GFI1 that has a potential transcriptional regulatory function.

Our approach for confirming and understanding protein interactions started with measuring the expression level of the mRNA and protein level in T-ALL cell lines. A stable housekeeping gene, *TBP* (*TATA* box binding protein), was used as a control gene in qPCR experiments. Housekeeping genes like -actin, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) or -tubulin are the most widely used normalisation controls. There is, however, a variation in the RNA and protein expression level of these housekeeping genes across T-ALL cell types as well as changes during developmental stages. Considering that we used cell lines that were at different stages of differentiation, we decided to normalise our samples based on the quantification relative to BSA, which is then confirmed by a total protein stain with Ponceau S. Ponceau S is considered a reliable method for assessing total protein loading on western blots (Sander et al., 2019).

Our investigation began by determining the mRNA and protein expression levels of TAL1 and LMO2. The results were consistent with the RNA-seq data previously obtained by the Hoogenkamp group. There was high expression of TAL1 and LMO2 in the DU528 cell line, which is the cell line carrying the SIL-TAL rearrangement as the initiating oncogenic transformation. The results indicate that while LMO2 is expressed as a 17 kDa protein in the four T-ALL cell lines, TAL1 in DU528 cell is only present as short 22 and 19kDa isoform. Tal1 isoforms originate from different translation start sites; therefore, the two short isoforms do not contain the transactivation domain, but only the DNA-binding and heterodimerization domains (Sharma et al., 2023). Therefore, TAL1 in DU528 cells can interact with HEB and LMO2 and bind to E-box motif but it lacks the ability to interact with a full set for transcriptional regulators as 44 and 39kDa TAL1.

Examination of the HEB levels revealed that *TCF12* is expressed in T-ALL cell lines, and the protein HEB can isoform was found in all four cell lines. Meanwhile, HEBalt was only observed in ARR and DU528. However, our HEB IPs, using protein extract from all four T-ALL cell lines showed that both HEBcan and HEBalt are present in all four T-ALL cell lines. The fact that HEBalt was not detected initially is most likely be due to lower HEBalt levels in HSB2 and CCRFCEM cells that could not be detected in western blot experiments where 100 μg of nuclear extract used, while in HEB-IP experiments utilising 600 μg made the detection of HEBalt possible.

When we looked at the GFI1 levels we saw from our results that the lowest mRNA and protein levels were found in the least differentiated ARR, cells and in the most differentiated CCRFCEM cell line. Therefore, we see a negative correlation between the HEBalt and GFI1 levels so that in less differentiated cells (ARR) HEBalt and low level of GFI1 proteins are

present while higher GFI1 level and less HEBalt were noted in the more differentiate cells (CCRFCEM).

There was a correlation between the protein level and the mRNA level, with the highest level being expressed by the CCRFCEM cell line, which is the most differentiated cell line used (Sandberg et al., 2007b), where *TCF12* is essential at this stage and is involved in the development of $CD4^+$ T-cells (Wojciechowski et al., 2007). Additionally, the results were also in agreement with those obtained by RNA-seq.

To investigate the localisation of HEB within T-ALL cell lines, we conducted an immunocytochemistry assay. Our results proved in accordance with previous studies that HEB and LMO2 have nuclear localisation (Labreche et al., 2015a). Different studies have identified HEB as being expressed in the nucleus of mesenchymal stem cells in mice and pre-osteoblasts, as well as in human brain tumours, where it was found to have nuclear expression (Labreche et al., 2015b, Yi et al., 2017). As a result of the comprehensive analysis conducted by Natkunam et al. (2007), strong evidence was found for the nuclear localisation of LMO2 protein in normal B-cells, based on an analysis of more than 1,200 normal and neoplastic tissues and cell lines. Our own findings support this observation, which demonstrates that LMO2 is nuclear localised in acute lymphoblastic leukaemia cell lines (Natkunam et al., 2007). In addition to these, we have performed experiments to examine the GFI1 localisation and despite using several different antibodies we could not produce reliable results most likely due to the antibodies not being able to recognise GFI1 after the fixation with formaldehyde.

A number of studies have previously explored the crucial role that GFI1 plays in T-cell development as well as its interactions with a variety of molecular partners. Research has demonstrated that GFI1 is indispensable for the maintenance of haematopoietic stem cells (Duan and Horwitz, 2005). It is also indispensable for promoting the transition from early T-

cell precursors to committed T-cell progenitors. Additionally, GFI1 promotes the balance between self-renewal and differentiation in T-cell development, allowing functional T-cells to be produced.

Several studies have proven the precision and sensitivity of co-IP and mass spectrometry methods in characterising protein complexes and their reactivity to various regulatory mechanisms (ten Have et al., 2011). As previous HEB IP-mass spectrometry experiments from Hoogenkamp lab had detected GFI1, this project was to explore if GFI1 was indeed a HEB interacting partner. Our HEB and GFI1 IPs, using protein extract from all four T-ALL cell lines confirmed the interaction between HEB and GFI1 and additionally revealed that HEBalt is the isoform responsible for this interaction. Additionally, both HEB and GFI1 showed an interaction with LMO2 in ARR and DU528 cells while in more differentiated cell lines HSB2 and CCRFCEM, only HEB interacted with LMO2. This could indicate that GFI1 and the HEBalt isoform are involved in a transcriptional complex distinct from the one involving LMO2 and HEBcan. Additionally, it is worth noticing that the HEBalt/GFI1/LMO2-containing complex was detected in less differentiated, ARR and DU528 cells, which have higher HEBalt levels while in more differentiated HSB2 and CCRCEM cells, where HEBalt was hard to detect, HEBcan/LMO2 complex is more prevalent. An obvious question is if these two complexes could be involved in the transcriptional regulation of different genes and if their presence would also be at the different stages of the normal T-cell development. Further investigation is needed to understand the precise role of this complex and its implications in T-ALL and T-cell development.

HEBalt and HEBcan are two HEB isoforms that have previously been identified as having distinct functions. Retroviral overexpression of HEBcan or HEBalt has provided insight into the role of each HEB factor in the specification of lineages and the determination of

developmental fate. For instance, within the thymus, HEBalt induces T-lineage gene expression and suppresses myelopoiesis while overexpression of HEBalt ectopically in LSK cells led to enhanced differentiation into T-cells and reduced production of myeloid cells (Wang et al., 2010, Wang et al., 2006b). Moreover, HEBalt overexpression suppresses B-cell potential during B-cell development (Braunstein et al., 2010). On the other hand, HEBcan plays a critical role in the initiation of CD4 gene expression (Sawada and Littman, 1993). It also appears to play a more prominent role in repressing natural killer cell differentiation. Therefore, considering that each isoform has a unique function, each isoform may interact with a different set of binding partners, resulting in binding to a different set of DNA elements and hence regulating different group of responsive genes. Even binding to the same DNA-response elements, two distinct complexes might be recruiting a different set of activators/ inhibitors which would further lead to activation or suppression of the target gene expression.

In view of the previous HEB mass spectrometry in which GFI1 was found to be monomethylated, the results of our co-IP experiment show that GFI1 was detected after Kme1 pulldown followed by GFI1 western blot. We have therefore determined that the complex containing GFI1 in ARR and DU528 cells consists of GFI1-Kme1, HEBalt, and LMO2. However, in HSB2 and CCRF-CEM cells, we only observed an interaction between GFI1 and HEBalt. A possible explanation is that GFI1, once methylated, acts as a repressor, and interacts with HEBalt. Further studies are needed to verify this hypothesis using more sensitive assays. This result also requires further investigation since LMO2 in ARR and DU528 showed an association with lysine mono-methylation pulldown.

As part of this study, we sought to expand upon previous investigations and identify members of the GFI1 complex through mass spectrometry analysis that may interact with GFI1 in T-ALL. It is unfortunate that we were unable to detect HEB in any of the GFI1-IP mass

spectrometry experiments. This might be due to the strong GFI1 interaction with the GFI1 antibody and the possibility that the treatment with the reducing agent did not break the antigen-antibody bonds, therefore reducing the level of detectable GFI1. Similar problem was detected in previous experiments in Hoogenkamp group, and the curated inspection of the mass-spec data revealed that the peptides of the immunoprecipitated proteins can be found. Moreover, post-translational modification such as methylation and phosphorylation alter the mass of the peptides, resulting in peptides not being identified and a reduction in the overall likelihood of the identification of modified protein (Parker et al., 2010). Additionally, since we have identified HEB presence in GFI co-IP experiments, which are more sensitive and a better approach to address the question of a specific protein-protein interaction we believe that the HEB-GFI1 interaction is genuine, and we used the results of the IP-mass spectrometry experiments as a guide to uncovering more distant interaction.

In general, the mass spectrometry results of GFI1 and Kme1 indicate that approximately 50% of the proteins detected in the GFI1 pulldown interact with proteins that have monomethylated lysine. In Kme1 IP/mass spectrometry using CCRFCEM cell extract we identified the presence of EHMT2 (G9a). In ARR and HSB2 cells, we identified EHMT1 (GLP), which is also lysine methylase protein, 80% identical to EHMT2 in its catalytic domain. Previous studies have demonstrated that GFI1 and G9a are two proteins that regulate gene expression by adding methyl groups to histone H3, lysine 9 (H3K9) leads to chromatin compaction and suppression of gene expression (Duan et al., 2005). Due to the GFI1/G9a interaction, GFI1 acts as a transcriptional repressor. Furthermore, our results revealed the interaction with RCOR, HDAC1 and LSD1. RCOR is known transcriptional repressor that recruits HDAC and LSD which promote histone deacetylation and H3K4 demethylation, modifications that together with H3K9 methylation required for epigenetic silencing of gene expression. It has been demonstrated that in T-cells, GFI1 can interact with and recruit HDAC1, contributing to

epigenetic regulation of gene expression, which is essential for the development and function of T-cells (Duan et al., 2005). . However, LSD1 was only detected in CCRFCEM experiments, it will be important to further investigate if GFI1 recruits LSD1 in all the T-ALL and to address the function of GFI1 in suppression of gene expression by ChIP-seq and other genome-wide approaches. Also, as we have previously identified that GFI1 carries lysine methylation, it would be interesting to investigate whether this modification affects GFI1's interaction with other proteins and its function.

In our analyses of the mass spectrometry data, we explored GFI1 direct interactions for each cell line separately, especially focusing on any transcription factors. One of the identified interactions was found in ARR cells, the Runt-domain transcription factors, RUNX1, RUNX2 and RUNX3 were detected, and the STRING analysis recognised that RUNX proteins as directly interacting with GFI1. Moreover, the RUNX obligatory binding factor CBFβ (core binding factor b), was found in three of the T-ALL samples. Both, RUNX and CBFβ are known to be important in haematopoiesis and are known to facilitate the selection of T-cell fates during various stages of the lineage specification process (Taniuchi et al., 2002) (Egawa et al., 2007, Botezatu et al., 2016). Literature describes that RUNX/CBFβ regulate the expression of GFI1 but the direct interaction between these factors has not been demonstrated yet and this would be one of the findings that can be further explored in Co-IP experiments as well as through the genome-wide approaches, especially as we already have the RUNX1 ChIPseq data from the T-ALL cell lines (Botezatu et al., 2016). Direct RUNX/GFI1 interaction might suggest that RUNX not only activates the GFI gene expression but by direct interaction with GFI, it recruits the suppressor of the gene expression to, at least, a subset of its own target genes, therefore through the interaction with GFI1, RUNX would also be able not only to activate but also to suppress the gene expression.

GFI1 was also found to have direct interaction with the majority of proteins that are part of the DNA-damage repair machinery such as MRE11 and DNA phosphor kinase C PRKDS/DNPKc, p53 binding protein TP53B, RAD50, arginine methyl transferase PRMT1/ANM1, PARP1/2, Ku70, Ku86 and NBN. This finding is in consistent with Vadnais et al. (2018), where GFI1 was found to acts as a cofactor for PRMT1, facilitating the methylation of DNA repair proteins, such as TP53BP1 in T-cells thus establishing a critical link between GFI1 and TP53BP1 in the DNA damage response pathway as well as by its interaction with DNA-damage repair machinery through MRE11 and DNA phosphor kinase C PRKDS/DNPKc (Vadnais et al., 2018). It is of particular interest since previous research conducted by our group showed that LMO2 through its interaction with PHF6 is responsible for the recruitment of the DNA repair machinery, through TP53BP to γH2AX, which marks sites where DNA integrity has been compromised (Stanulović et al., 2020). As our Co-IP experiments confirm the GFI1/LMO2 interaction, a question arises if the LMO2/HEB/GFI1 interaction is not only important for the regulation of gene expression but also for the surveillance of DNA integrity.

Among the identified possible interacting partners of GFI, in CCRFCEM experiment, IKAROS and PRMT5 have been identified, both known GFI1 interacting partners, A study revealed that GFI1 forms a transcriptional partnership with IKAROS in T-ALL cells. Together, they occupy regulatory regions of genes essential for T-cell development. This partnership activates gene expression independent of GFI1 LSD1 binding but requires GFI1 SNAG domains (Sun et al., 2022). It has been previously established that GFI1 forms a complex with a protein called AJUBA, facilitating PRMT5 recruitment to the transcriptional repressor Snail, and suggesting that the GFI1 protein may be involved in the process of arginine methylation (Möröy et al., 2015).

Finally, the analysis identified 42 common proteins between the four GFI1-IP mass spec data sets. These proteins involve biological processes, including RNA or mRNA splicing, viral translation termination and viral transcription. The STRING analyses showed that GFI1/HDAC1 interaction is linked to PRKDC and MRE11 through a nuclear matrix-associated protein SAFB protein. Therefore, SAFB is identified as a link between the GFI1 function as a repressor of gene expression to the GFI function in DNA repair machinery. This is of particular interest as raises a possibility that the two functions of GFI1 are connected by localisation in the nuclear periphery (Altmeyer et al., 2013) (Norman et al., 2016)

Chapter 4 The dynamics of GFI1 genomic occupancy

4.1 Introduction

After the interaction between HEB and GFI1 in T-ALL cell lines was established in the previous chapter, we examined the distribution of GFI1 across the genome using chromatin immunoprecipitation (ChIP). ChIP is a technique used to investigate protein association with the DNA. Briefly, cells are crosslinked to fix protein-DNA interactions, and then chromatin is fragmented. GFI1-specific antibodies are subsequently used to selectively pull-down fragments of chromatin bound to GFI1. These fragments are then purified. Quantitative polymerase chain reaction (qPCR) and sequencing techniques are used to analyse the ChIP experiment and identify the genomic regions to which GFI1 is bound. The analysis of GFI1 ChIP can provide insight into the regulatory mechanisms underlying GFI1 and its role in a wide range of cellular processes. By identifying the genomic regions where GFI1 binds, we will be able to gain a deeper understanding of the genes and pathways that are directly regulated by GFI1.

4.2 Results

An antibody that recognizes GFI1 was used to analyse the genomic distribution of GFI1 in chromatin samples that were isolated from T-ALL cell lines. We initially analysed the ChIP samples by qPCR. For this, we measured the pulled down DNA at a region near the SDE2 gene, which has previously shown to be highly enriched for several haematopoietic transcription factors, and at a region of chromosome 18, which has shown to be heterochromatic in haematopoietic cells and without any genes nearby. In all four T-ALL cell lines, ChIP experiments followed by quantitative PCR (qPCR), demonstrated an enrichment of GFI1 at the SDE2 region compared to the chr. 18 negative control amplicon we used. The qPCR measurements of GFI1 ChIP were correct for the amount measured in input samples of chromatin. Bar charts shown in Figure 4.1 were used to illustrate the relative enrichment of GFI1 ChIP-seq samples over the chr. 18 negative control.

Following the results of the qPCR, we generated genome-wide libraries for these ChIP samples. The samples were loaded onto agarose gels, as shown in Figure 4.2. The gel electrophoresis shows a visual representation of the size distribution of DNA fragments obtained. The DNA fragments were selected for size between 200 and 300 base pairs then were subsequently sequenced. The resulting ChIP-seq library reads were aligned and mapped to the reference genome human hg38 genome. Peaks were identified by using MACS software for identifying genomic regions with significantly more reads in the correct distribution. The GFI1 ChIP-seq analysis for ARR, DU528, HSB2, and CCRFCEM, were 7531, 14908, 7845, and 3376 peaks, respectively. Figure 4.3 illustrates a screenshot of the ChIP-seq data of GFI1 showing distribution of reads in T-ALL cell line.

Correlation analysis of ChIP-seq results of the transcription factor GFI1 between different T-ALL cell lines can reveal the degree of similarities of the binding pattern observed in the different cell lines. Using the plot correlation software within the Galaxy portal, we analysed

the patterns within the GFI1 ChIP-seq data of the different T-ALL cell lines. In Figure 4.4 A, the results show the different comparisons with the strength of the correlation indicated by a colour scale, with the scale bar below. The diagonal line represents the perfect correlation, it is comparing samples with themselves (correlation coefficient = 1). The correlation analysis demonstrated that HSB2 and CCRFCEM cell lines have a strong correlation, followed by DU528, whereas ARR showed a weak correlation. This is in line with HSB2, CCRFCEM and DU528 harbouring the SIL-TAL deletion, while ARR does not.

Next, we performed an intersection analysis to determine the overlap of GFI1 ChIP peaks between the four T-ALL cell lines. Using the intersection tool within Galaxy, intersections between the four data sets were determined and plotted as a table (Figure 4.4 B). The table clearly illustrates the peak distribution among the four cell lines, showing that DU528 had the highest peak count, followed by ARR, HSB2, then CCRFCEM. We observed a significant overlap in the binding patterns between the HSB2 and CCRFCEM cell lines of 1543 peaks, suggesting a strong similarity between these two cell lines. Additionally, 326 peaks have been identified as common between cell lines containing the SIL/TAL deletion (DU528, HSB2, and CCRFCEM) and 11 peaks have been found to be shared between the four cell lines. Taking these results into account, we decided for further analysis to separate ARR from the three SIL/TAL cell lines.
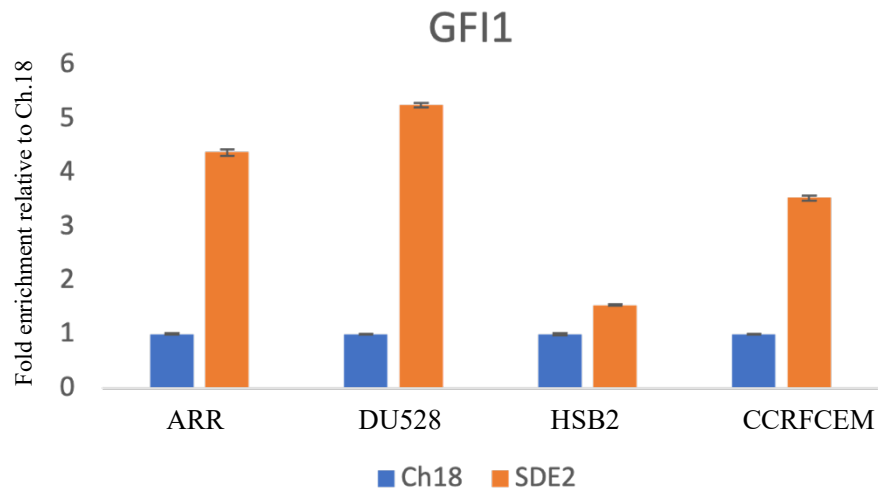
Figure 4.1: qPCR analyses of GFI1 ChIP in T-ALL cells.

The figure demonstrates a qPCR experiment using the indicated primer sets. The ChIP qPCR results showed a relative enrichment of GFI1 at SDE2 region in T-ALL, relative to Ch18 negative control. (n=4).
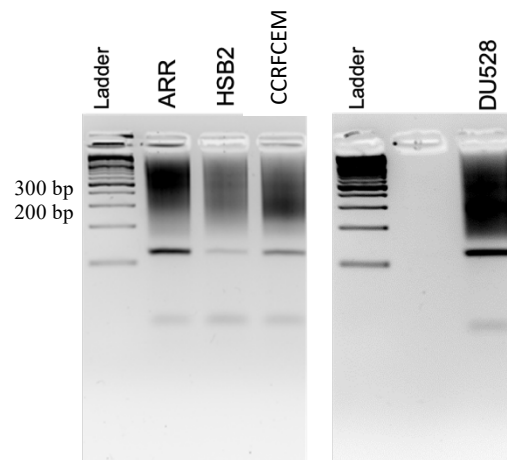
Figure 4.2: ChIP samples run on agarose gel.

An illustration of ChIP-seq libraries loaded onto 3% agarose gels. Fragments of 200-300 bp was selected, pooled and sent for sequencing.
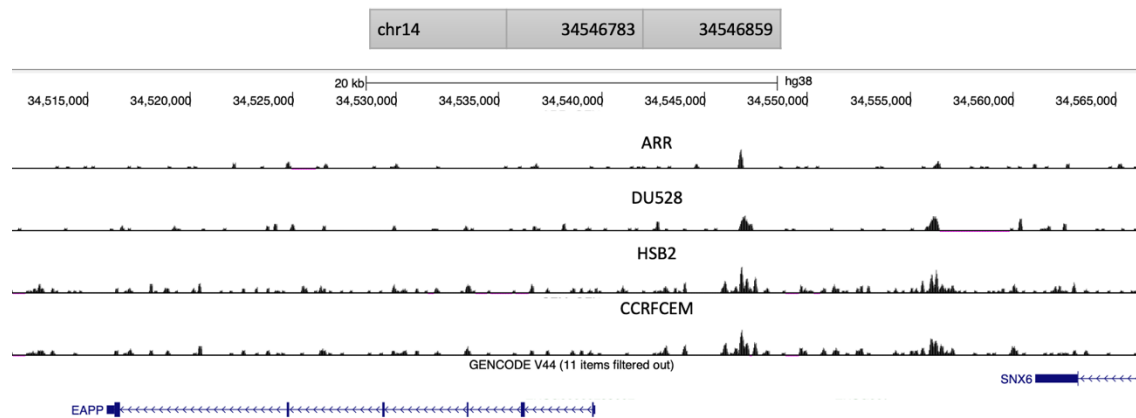
Figure 4.3: GFI1 binding in T-ALL cell lines.

Screenshot from UCSC browser illustrating binding profile of GFI1 in T-ALL cell lines.

(A)
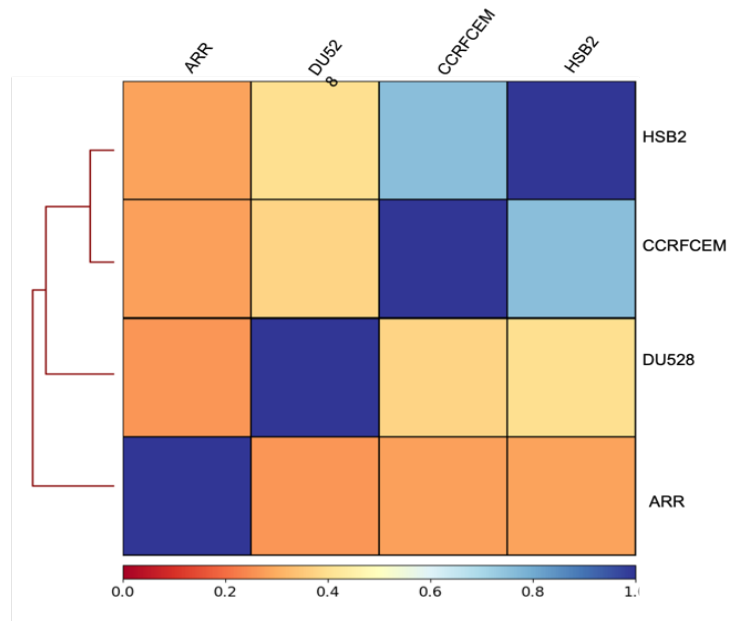
(B)

ARR (7531)				X	X	X	X		X	X	X
DU528 (14908)	X	X			X	X	X	X			X
HSB2 (7845)	X	X	X	X	X		X		X		
CCRFCEM (3376)	X	X			X		X	X		X	X
	1543	326	171	45	16	16	11	10	4	3	1

Figure 4.4: Correlation pattern and number of overlap peaks of GFI1 ChIP-seq using T-ALL cell lines.

(A) Heat map showing correlation between T-ALL cell lines, scale bar (from 0-1) below indicates colours relationship. (B) Table showing the overlap between GFI1 ChIP-seq peaks in T-ALL cell lines. The first column lists the name of the cell line and the total number of GFI1 peaks. The other columns represent the overlaps and the frequency of their occurrence.

To gain a better understanding of transcription factor binding at GFI1 binding sites, we performed *de novo* motif analysis to identify DNA sequences statistically over-represented within GFI1 binding regions. For this, we analysed the top 500 highest peaks of GFI1 ChIP-seq for the ARR cell line. Using the online tool MEME suite, the ChIP-seq FASTA format sequences of GFI1 peaks were analysed for enrichment of sequence motifs. We additionally analysed the methylocytosine DNA methylation database, which contains sequences of methylated DNA regions, to identify motifs commonly associated with methylated regions. The results are shown together in Table 4.1, with the methylation-specific data shown in red. There was a strong enrichment for a sequence that could bind MEF2A, SOX10/11and ZN384, as well as sequences recognised by ETV5 and ZNF257, MYBL2 and SP1/2/4 factors (Table 4.1).

Surprisingly, our search did not yield any GFI1 motifs. There is a possibility that GFI1 may be recruited by other proteins in the 500 highest-scoring peaks, rather than binding directly to DNA itself. To examine this further, we focused on a different set of sequences - the lowest-scoring peaks, where GFI1 might directly bind. Indeed, we found a GFI1 sequence in the 500 lowest-scoring peaks. Additionally, TFAP2C, KLF6 and SREBF2 binding sites were identified in this data set (Table 4.2).

Using the Genomic Regions Enrichment of Annotations Tool (GREAT), analyses were conducted to determine the locations and distances of these peaks in comparison to the nearest transcription start sites (TSS). The tool associates input genomic regions, bound to a protein of interest, with putative target genes nearest to those genomic regions (McLean et al., 2010a). The ChIP-seq BED files were analysed using the online tool. The GREAT analysis of the highest scoring 500 peaks in ARR revealed that approximately 5% of the peaks are located within 5 kb of a gene transcription start sites. About 20-30% of the bound regions were located

within 5-50 kb from the TSS, while 50-60% were within a distance of 50-500 kb. The percentage of sites located more than 500 kb from the TSS was approximately 15% (Figure 4.5 A). In the analysis of the 500 peaks with the lowest scores in ARR, less than 5% are located within 5 kb of the transcription start sites. Nearly 20% of bound regions were within a distance of 5-50 kb from the TSS, and 50-60% were located between 50 and 500 kb. Approximately 20% of the sites were located more than 500 kb from the TSS (Figure 4.5 B).

Table 4.1: Enriched motifs in the top 500 highest-scoring peaks from the GFI1 ChIP-seq data in ARR. Motifs associated with methylated regions are shown in red.

Presented here is a table showing DNA sequences that are highly enriched in the highest scoring 500 peaks. Each row represents a motif enriched in the GFI1 ChIP-seq data. Proteins predicted to interact with these motifs are listed under (known or similar motifs). Motifs associated with methylated regions are highlighted in red. E-value: Indicates the statistical significance of the motif enrichment. Sites: Indicates how many times the motif occurs among the top 500 highest scoring peaks.
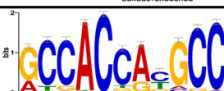
Motifs	*E*-value	Sites	Known or similar motifs
	8.87e+00	106	Prdm14,
	1.54e-04	426	MEF2A, SOX10, SOX11, ZNF384
	1.71e+00	87	TFEC, USF2, SREBF1, MLX, PAX5
	9.13e+00	113	MYBL2_DBD_2, TBX20
	5.10e-01	150	ZNF257, ETV5
	7.88e-01	127	ZNF740, GLIS1, ZIC4, E2F2
	9.41e-01	35	Hoxb3
	3.45e+00	26	SP1, SP2, SP4,

Table 4.2: Enriched motifs in the selected 500 lowest-scoring peaks from the GFI1 ChIP-seq data in ARR.

Motifs	*E*-value	Sites	Known or similar motifs
	2.06e+00	14	GFI1
	3.8e-068	37	ZNF460, TFAP2C, ZNF708
	3.5e-066	31	KLF6
	7.8e-006	12	SREBF2

A
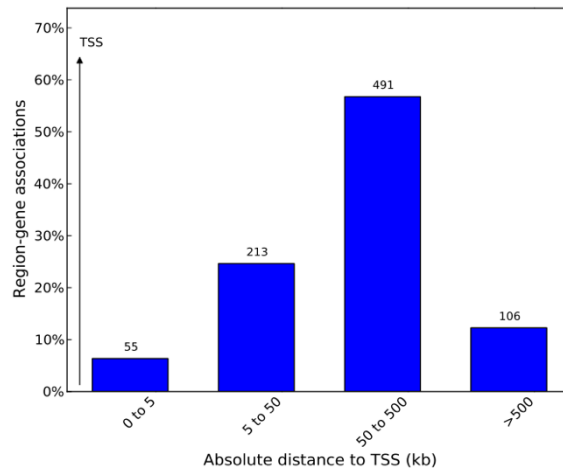
B
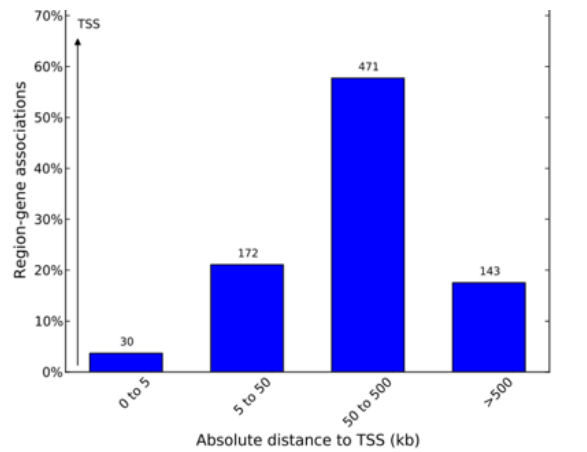
Figure 4.5: GREAT Region-Gene Association graphs of genomic regions bound by GFI1 peaks in ARR.

Graphs showing the distance and orientation of GFI1 bound regions in ARR to transcription start sites-TSS (A) Highest scoring 500 peaks (B) Lowest scoring 500 peak. On the y-axis, the percentage of genes being counted is given, and above each bar in the graph is the absolute number of genes being counted.

The interaction results in Chapter 3 led us to believe that GFI1 and HEB alternative may be involved in a complex that is distinct from the TAL1/LMO2 complex. To address this, we integrated GFI1 ChIP-seq data from ARR cell line with previously obtained ChIP-seq data for LMO2, TAL1, LYL1, HEB, GATA2, LDB1, RUNX1, LMO4, PHF6 and MethylCpG in ARR (Figure 4.6). As a result of the correlation analysis of the data set, we discovered a high correlation between RUNX1, LDB1, GATA2, HEB and GFI1. Both LMO2 and TAL1 showed a significant correlation in ARR not associated with the previously mentioned complexes.
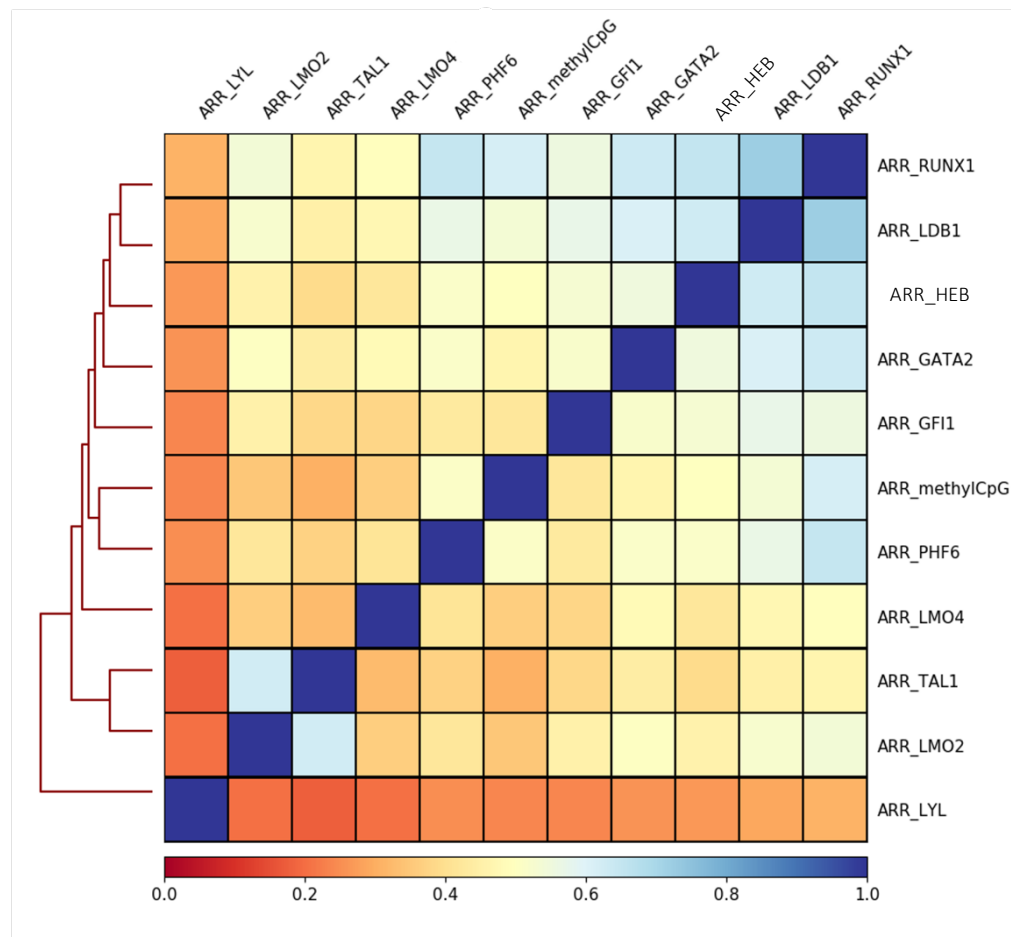
Figure 4.6: The correlation pattern of GFI1 ChIP-seq in ARR cell line.

Heat map showing hierarchical clustering representing the correlation between GFI1, HEB, LMO2, TAL1, GATA2, PHF6, RUNX1, LMO4, LDB1, LYL and MethylCpG ChIP-seq data. Scale bar (from 0-1) below indicates colours relationship.

Similar to the above analysis for ARR, we conducted the identification of motifs and correlations among the common peaks overlap between the SIL/TAL cell lines, DU528, HSB2 and CCRFCEM. In the *de novo* motif analysis of SIL/TAL cell lines, sequences representing binding sites for TCF3, TCF7L2, SOX18, and SOX10 were enriched. Furthermore, sequences for IKZF1, SPI1, TEAD and ZFP14 were identified as well (Table 4.3)

With the GREAT analysis, we identified the location and distance of common SIL/TAL peaks from transcription start sites. The results showed that approximately 10% of peaks were located within 0-5 kb and 10% were between 5-50 kb, approximately 33% were located within 50-500 kb, and around 47% were within more than 500 kb from TSS (Figure 4.7). There is a difference between the distribution of ARR peaks and the distribution of SIL/TAL peaks in relation to the TSS. Peaks in the SIL/TAL are located farther from the TSS, with a greater percentage found more than 500 kb away, whereas ARR peaks are more frequently observed closer to the TSS, especially within 50-500 kb.

Following this, we integrated our GFI1 ChIP-seq results from the three cell lines with previously performed ChIP-seq data from the Hoogenkamp group. Using hierarchical clustering, we observed that our GFI1 results grouped together between them with a relatively weak correlation to the results of other ChIP-seq experiments (Figure 4.8). Moreover, our results suggest that GFI1 binding sites are distinct to the other ChIP-seq data, implying that the overall binding profile of GFI1 is distinct to that of the LMO2 complex and associated factors. There is a strong correlation between RUNX1 in HSB2 and CCRFCEM and PHF6 in DU528 and HSB2, however RUNX1 in DU528 and PHF6 in CCRFCEM are not among these two groups. HEB, GATA2, LDB1 and LMO4 were in the same group and exhibited a strong correlation pattern. The binding patterns of TAL1 and LMO2 is a common feature across SIL/TAL cell lines and this binding was consistent across the different cell lines. Additionally,

the strong correlation pattern between HEB, GATA2, LDB1 and LMO4 indicates that they may be involved in a common complex, while the distinct functions of RUNX1, LDB1, and PHF6 in different cell lines suggests that they may also be involved in a different complex. The presence of a transcription factor in a subset of one TF complex as well as an additional TF complex would reduce the direct correlation between the two factors. This does however not mean that they do not interact with each other in one of these subsets of peaks.

Table 4.3: Enriched motifs analysis from common peaks in GFI1 ChIP-seq data in SIL/TAL cell lines.

Table showing DNA sequences that are highly enriched found in common peaks of GFI1 in DU528, HSB2, CCRFCEM cell lines. Each row represents a motif enriched in the GFI1 ChIP-seq data. Proteins predicted to interact with these motifs are listed under (known or similar motifs). E-value: Indicates the statistical significance of the motif enrichment. Sites: Indicates how many times the motif occurs.

Motifs	*E*-value	Sites	Known or similar motifs
	8.19e-01	36	SOX10
	3.12e-01	254	TCF7L2 Lef1 Tcf3
	5.89e+00	145	TEAD1, TEAD3, TEAD4, ZFP14
	8.28e+00	50	IKZF1
	4.26e+00	64	SPI1, SPIB
	1.72e+00	237	SOX18
	9.76e+00	80	Hbp1
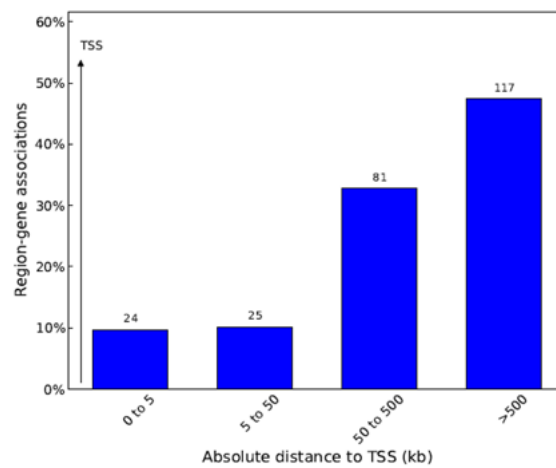
Figure 4.7: GREAT Region-Gene Association graphs of genomic regions bound by GFI1 peaks common between DU528, HSB2, and CCRFCEM.

Graphs illustrating the distance and orientation of GFI1 bound regions of peaks overlapped between SIL/TAL cell lines to transcription start sites- TSS. On the y-axis, the percentage of genes is given, and above each bar in the graph is the absolute number of genes being counted.
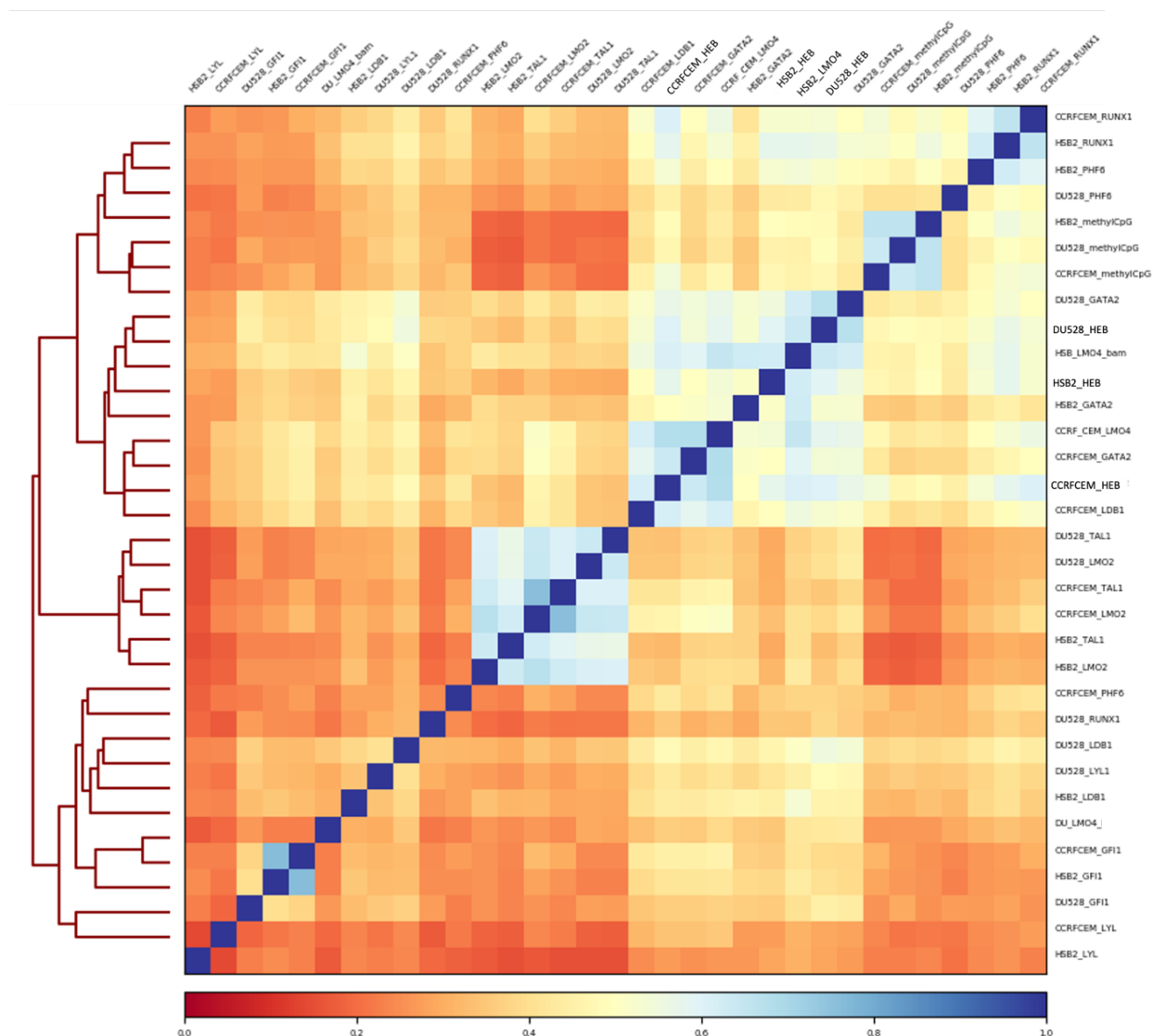
Figure 4.8: Correlation pattern of ChIP-seq data of SIL/TAL cell lines.

Heat map showing hierarchical clustering of the correlation between GFI1, HEB, LMO2, TAL1, GATA2, PHF6, RUNX1, LMO4, LDB1, LYL and MethylCpG ChIP-seq data in DU528, HSB2, and CCRFCEM cell lines. Scale bar indicates correlation coefficient (0-1).

The analysis of the GFI1 and transcriptional complex containing HEB was extended using the same panel of genome-wide datasets previously generated by the Hoogenkamp lab. Utilizing the GFI1 peak intensity scores for each cell line, we generated a matrix for each cell line and sorted them in descending order of peak intensity to visualize the GFI1 co-localisation with the other members of the complex. EaSeq analysis software was used to generate heat maps by overlaying these matrices with Bam files of the other ChIPseq. The Bam files contained aligned sequences from ChIP-seq data for HEB, LDB1, PHF6, GATA2, RUNX1, LYL1, LMO4, TAL1, LMO2, and MethylCpG.

In the heatmaps of ARR, the strongest peaks of GFI1 are mirrored by the strongest peaks of HEB, LDB1, PHF6, GATA2, RUNX1, LYL and LMO4, showing co-localisation of these proteins (Figure 4.9). This pattern was also observed in MethylCpG but was not found with TAL1 or LMO2. For the DU528 cell line, the co-localisation was not as prominent as observed for ARR. Only few of the strongest peaks showed co-localisation, whereas less intense GFI1 peaks showed only very weak co-localisation with the other proteins. For the HSB2 and CCRFCEM cell lines we observed a co-localisation pattern that was between ARR and DU528. Based on the heatmap results, we conclude that GFI1 does co-localise with HEB, in line with a direct interaction. This complex would also involve the other transcription factors mentioned above, except for LMO2 and TAL1. Although HEB correlates with GFI1, the TAL1/LMO2 complex does not show much overlap with GFI1. As a result, it is likely that HEB is a member of at least two distinct complexes: one involving GFI1 and the other involving the TAL1/LMO2 complex. Interestingly, the co-localisation between GFI1 and HEB seems for a large part to correlate with the presence of MethylCpG signal as well.

Following this, we observed the peak overlays on basis of the HEB signal intensity (Figure 4.10). In contrast to the GFI1 matrix, in ARR almost all of the HEB associated regions were

bound by GFI1, LDB1, PHF6, GATA2, RUNX1, LMO4 and to a lesser extent, LYL1. With TAL1 and LMO2, there is a different pattern, but there is an overlap with HEB. Based on MethylCpG analysis, the main differences between ARR and SIL/TAL cells can be seen since there is a correlation with HEB in ARR but not with SIL/TAL cell lines. The three cell lines followed a similar pattern, exhibiting a lower correlation between GFI1 and HEB compared to the ARR results. All other transcription factors were strongly correlated and overlayed with HEB.
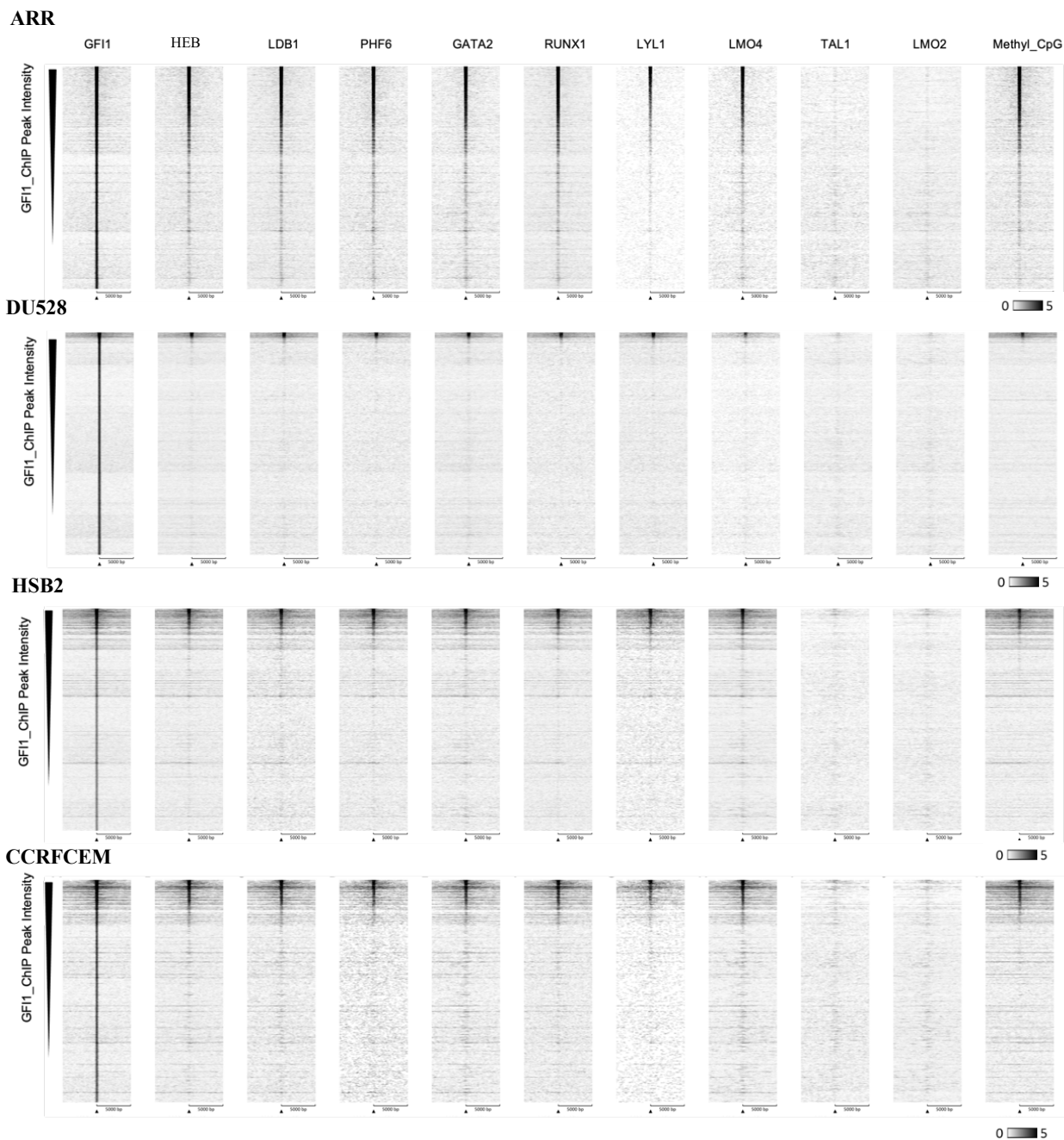
Figure 4.9: Heat maps plots showing GFI1 ChIP-seq results ranked according to intensity of binding.

The heat maps show GFI1, HEB, LMO2, TAL1, GATA2, PHF6, RUNX1, LMO4, LDB1, LYL and methylCpG ChIP-seq results ranked according to intensity of GFI1 binding, overlayed on GFI1 ChIP-seq matrix in T-ALL cell lines. Easeq software was used to generate heatmap plots, depicting windows from ± 5 kb around the centre of the GFI1 peaks. Greyscale intensity is indicated below the heat maps.

136
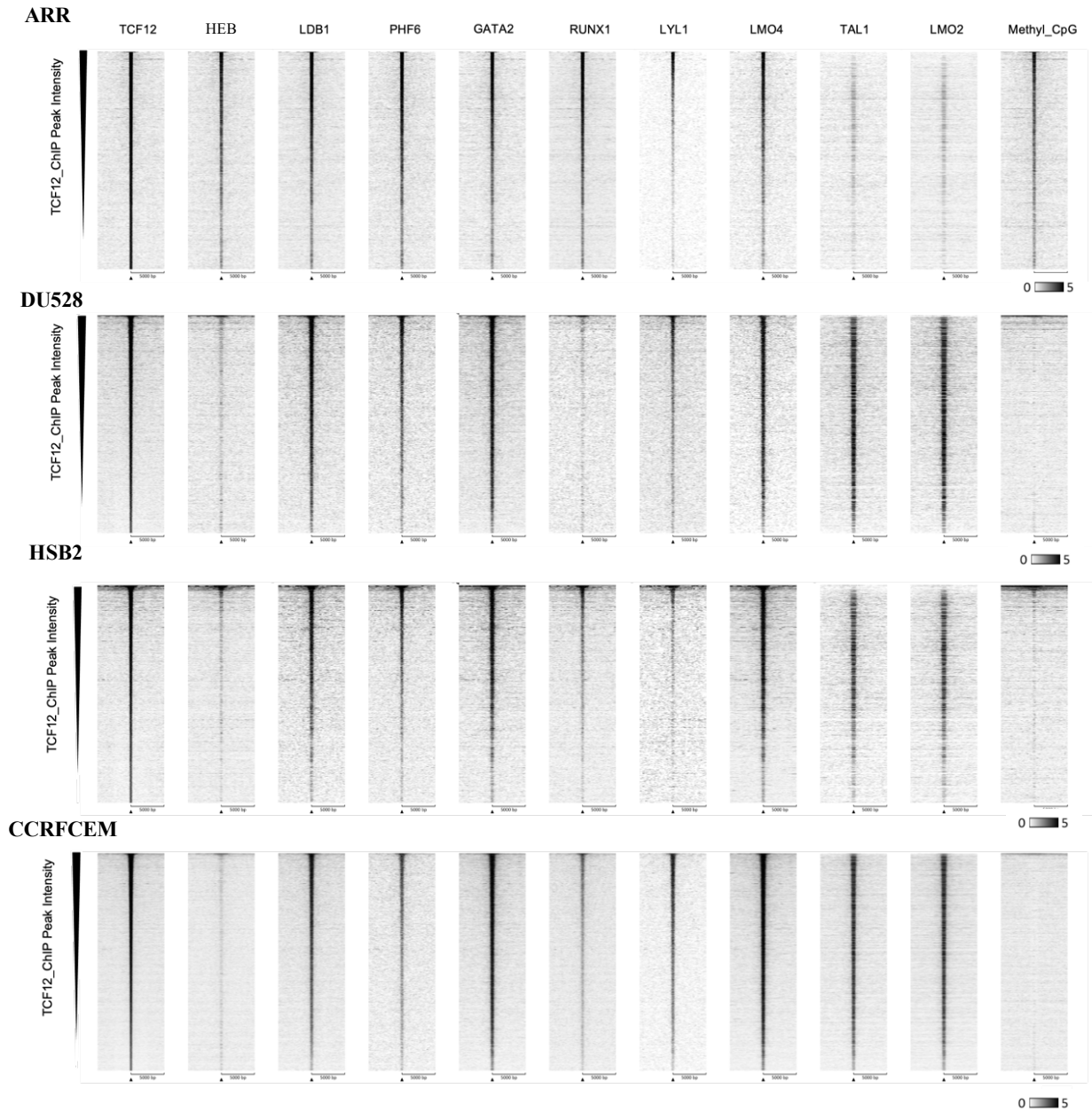
Figure 4.10: Heat map plots showing correlation between ChIP-seq results in T-ALL cell lines.

The heat maps show GFI1, HEB, LMO2, TAL1, GATA2, PHF6, RUNX1, LMO4, LDB1, LYL and methylCpG ChIP-seq results ranked according to intensity of GFI1 binding, overlayed on HEB ChIP-seq matrix in T-ALL cell lines. Easeq software was used to generate heatmap plots, depicting windows from ±5 kb around the centre. Greyscale intensity is indicated below the heat maps.

4.3 Discussion

Growth factor independence-1 (GFI1) is a zinc finger transcription factor that plays an important role in myeloid and lymphoid haematopoiesis in normal and malignant haematopoietic cells (Hock et al., 2004, Zeng et al., 2004). GFI1 is known to repress genes by recruiting LSD1-containing protein complexes, whereas in cooperation with IKAROS, GFI1 activates gene expression by acting as a transcriptional activator (Khandanpour and Möröy, 2013), (Sun et al., 2022). We were interested to examine the role of GFI1 in T-ALL cells, and particularly how it interacts with HEB and members of the LMO2/TAL1 complex. We investigated the binding of GFI1 throughout the genome by ChIP-seq experiments. We found that the ChIPseq data from the three SIL/TAL cell lines, *i.e.* DU528, HSB2 and CCRFCEM, demonstrated clear correlations between each other, but that ARR was clearly distinct. Upon intersecting GFI1 peaks of all cell lines, we discovered 326 shared peaks of the three SIL/TAL cell lines, with only 11 shared peaks with ARR. We therefore focused our analysis on ARR separately and combined the three SIL-TAL cell lines together.

We then selected 500 highest peaks in ARR and 500 lowest peaks in ARR. *De novo* motif analysis identified in the top 500 peaks of ARR include critical regulators of many developmental processes, as well as key transcription factors involved in differentiation, such as myocyte enhancer factor-2 (MEF2) (Subramanian and Nadal-Ginard, 1996), MYBL2, SRY-box transcription factors (SOX10 and SOX11). This analysis also identified ETV5, which is a member of the ETS family that is involved in the regulation of γδ T-cell development (Jojic et al., 2013), and PAX5, which is involved in B cell development (Souabni et al., 2002). This is in line with ARR representing an ETP leukaemia, which is characterised by the expression of genes normally associated with other lineages, rather than only factors specific to T cells.

In the lowest scoring 500 GFI1 peaks in ARR, our analysis identified the GFI1 motif. GFI1 motif identified within lower-scoring peaks in GFI1 ChIP-seq data suggest different

explanations such as it is possibly recruited by the top 500 proteins, weaker or more transient binding affinity of GFI1 for these DNA regions. The nuanced binding pattern of GFI1 makes it interact with a different set of genes than GFI1 with a higher affinity. Additionally, GFI1's binding dynamics are context-dependent, influenced by various factors such as the presence of cofactors, epigenetic marks, and the current state of the cell. Other enriched motifs included potential binding sites for transcription factor AP-2 gamma (TFAP2C), and Sterol Regulatory Element Binding Transcription Factor 2 (SREBF2). These were found to be target genes regulated by IKZF1(Vastrad et al., 2020). Kruppel-like factor 6 (KLF6) was also detected, which is a zinc finger functions as a transcription activator and is a tumour suppressor (Laitman et al., 2016).

In the motif analysis of SIL/TAL cell lines we detected IKZF1 (IKAROS) and TCF3. IKAROS plays a critical role in B and T cells development, and in T-ALL function as a tumour suppressor. GFI1 was found to acts as a transcriptional activator to activate gene expression when it cooperates with IKAROS. This is in agreement with a recent study in which BioID proximity labelling was used in human CCRFCEM T-ALL cells to detect GFI1-interacting proteins. The researchers identified 500 direct and indirect interacting proteins with GFI1 itself. These included previously identified GFI1-interacters, such as HMG20B, STAG1, LSD1, CoREST, and GSE1. Using GFI1 and IKAROS ChIP-seq to investigate potential targets, approximately 80% of GFI1-bound peaks overlapped with IKAROS peaks. As a result of this study, it was suggested that GFI1 and IKAROS regulate similar genes and that genes co-occupied by GFI1 and IKAROS play an important role in T-cell development, including GFI1 itself, NOTCH3, CD3, TCF3, MYC, RUNX3, MYB and HES (Sun et al., 2022).

Additionally, we identified enrichment of binding sites of proteins associated with CD8[+] T cell progression such as TCF7L2 and LEF (Greenough et al., 2015). The potential binding site of the protein TEAD, which were identified as a stage-specific regulator of murine blood

differentiation *in vitro* and *in vivo* by a study, which identified TEAD4 binding sites share a significant number with LMO2 and TAL1 (Goode et al., 2016). SPI1 (PU.1) and SPIB binding sites were enriched in these cells, which genes are activated during early stages of T-cell development. Silencing of these factors takes place during the β-selection. If these factors are not silenced it could result in the development of malignant T cells (Yui and Rothenberg, 2014). According to a previous study in myeloid cells, GFI1 antagonizes PU.1 activity via protein-protein interactions and plays a critical role in repressing macrophage gene expression during granulocyte development (Dahl et al., 2007).

The distance of genomic regions from transcription start sites is indicative of their potential functional roles within the genome. Proximal regions, typically within a few hundred to thousands of base pairs of the TSS, are commonly associated with promoter elements responsible for initiating transcription. Enhancer elements, on the other hand, tend to be located further away from the TSS, ranging from 50 to 500 kb, and can exert regulatory influence on gene expression over long distances. Distal regions, located more than 500 kb from the TSS, may serve structural functions or participate in long-range chromosomal interactions.

In the context of the study, the distribution of peaks from ChIP-seq data at varying distances from the TSS provides insights into the regulatory landscape of the examined genomic regions. The GREAT analysis revealed that a significant proportion of GFI1 peaks in both the ARR and SIL/TAL datasets were located within 50-500 kb from the nearest TSS. This suggests a preference to binding at enhancer elements, which could modulate the expression of nearby or even distal target genes.

Interestingly, the distribution patterns differed between the ARR and SIL/TAL datasets. Peaks within the ARR results were most frequently observed within the 50-500 kb range, indicating a potential role in binding to enhancers. Peaks associated with SIL/TAL were found to be located farther from the TSS, with a higher percentage located more than 500 kb away. This

may suggest a greater involvement of GFI1 with distal regulatory elements in the SIL/TAL dataset, possibly influencing gene expression through long-range chromosomal interactions.

To determine how GFI1 correlated with previously generated ChIP-seq data of other proteins in ARR, we created a heat map that showed hierarchical clustering of GFI1 ChIP-seq data with ChIP-seq data from HEB, LMO2, LMO4, LYL1, TAL1, LDB1, PHF6, RUNX1, GATA2 and methylCpG. We then used EaSeq analysis software to overlay the matrices for each cell line with Bam files of these factors sorted in descending order to create heat maps to visualize the genome-wide association of GFI1 localization. In the hierarchical clustering, we found strong correlation between GFI1, HEB, RUNX1, GATA2, LDB1, PHF6 and MethylCpG, while TAL1 and LMO2 showed very weak correlations. Approximately half of the highest scoring peaks of GFI1 on the heatmap plots of ARR appeared to co-localise with and overlap with these factors, except not with TAL1 and LMO2. Previous research has demonstrated that GFI1 is regulated by a number of key players that drive HSC development. These include the transcription factors RUNX1, TAL1, and GATA2 (Wilson et al., 2010). Considering the fact that ARR is an ETP cell line, it is not surprising that RUNX1 correlates strongly, since RUNX1 is primarily responsible to generate myeloid and lymphoid cells from definitive HSCs (Link et al., 2010), (Kurokawa, 2006). Researchers have demonstrated that LDB1 can self-associate as trimers and is able to assemble large multi-transcriptional factor complexes (Cross et al., 2010). LDB1 binds to proteins containing the LIM domain, such as the LMO proteins (Ryan et al., 2006). Through these interactions, LDB1 can bring together several enhancers and promoters, which was shown during erythroid β-globin regulation and included HEB and LMO4 (Song et al., 2010). Considering our results, it may be possible to conclude that the complex in ARR that contains LDB1 binds to LMO4 instead of LMO2. This complex includes HEB, RUNX1 and GFI1 differs from the complex containing LMO2/TAL1.

Chapter 5 Exploring the impact of HEB deletion on T-cell acute lymphoblastic leukaemia: a CRISPR/Cas9 study

5.1 Introduction

As a powerful gene editing technique, CRISPR-Cas9 can be used to disrupt, delete, or alter the normal activity of specific genes. It is a precise and efficient method of gene editing that can be used to study gene function, develop disease models, or even create novel genetic modifications. Gene knockout models using CRISPR-Cas9 provide researchers with a better understanding of gene function, the molecular pathways it regulates, and the role it plays in normal development and disease. As a result of these studies, we are able to gain a better understanding of the mechanisms underlying gene regulation as well as identify potential targets for therapeutic intervention in relevant diseases.

The transcription factor *TCF12*, also known as HEB, is a protein that regulates various cellular processes, such as embryonic development, neurogenesis, and mesoderm development. Through the use of gene editing techniques such as CRISPR/Cas9, human embryonic stem cells (hESCs) were transfected with plasmids that disrupted HEB expression. The results of this study revealed a profound defect in mesodermal development, failure of T cell development, and a decrease in the expression of two key regulators of hemogenic endothelium development (Li et al., 2017). In addition, the ablation of *Tcf12* in HSCs led to a significant reduction in the expression of differentiation genes, suggesting that the expression of differentiation genes may be influenced by *Tcf12* in its role in regulating the differentiation process (Liao and Wang, 2021).

HEB has been demonstrated to play a role in T-cell differentiation and development in several studies. Dysregulation or aberrant expression of *TCF12* may contribute to T-ALL progression and development (O'Neil et al., 2004). *Tcf12* deletion affected the balance between the CD4 and CD8 T-cell populations with significant numbers of peripheral CD8$^+$ cells lacking surface TCR. These results suggest that *TCF12* prevents cells from developing into single positive

144

cells until there is an efficient selection signal for TCR-positive cells (Jones and Zhuang, 2007).

The purpose of this study was to investigate the role of HEB, in T-cell acute lymphoblastic leukaemia. To accomplish this, we performed CRISPR/Cas9 experiment to delete HEB and analyse its effect on the proliferation, apoptosis, and differentiation stages of T-ALL cells. These knockout experiments are expected to provide a better understanding of the molecular basis of T-cell acute lymphoblastic leukaemia, as well as provide insight into potential treatments for this type of cancer.

5.2 Results

To perform *TCF12* knock out experiments, we decided to purchase the required plasmids. We chose the Santa Cruz; sc-402508, HEB CRISPR-Cas9 KO plasmid mix. This contains two plasmid mixes; one containing up to three plasmids for guide RNA and one containing plasmids harbouring genomic regions surrounding the guide RNAs. We determined the genomic target site of each plasmid, expanded them through bacterial transformation, and then sequenced the samples to identify their relative genome positions (Figure 5.1A). In the Homology-directed repair (HDR) plasmid, the HDR guide plasmid was only provided for one of the 3 different guide positions in the gene, the N-terminal. The ARR cell line was transfected with N-terminal HEB CRISPR-Cas9 plasmid and HDR plasmid or Cas9 control plasmid and HDR plasmid. HDR plasmids contain red fluorescent protein (RFP) and cas9 plasmids contain green fluorescent protein (GFP) and a U6 promoter for regulating the expression of the gRNA (Figure 5.1B). In the twenty-four hours following transfection, puromycin was used to select cells, and then FACS was used to sort the GFP or RFP-positive cells (Figure 5.1C). Single cell colonies were generated by plating cells in 96-well plates, monitoring and expanding them to achieve clonal populations for protein extraction.

Our objective was to delete HEB from other T-ALL cell lines as well, therefore we electroporated DU528, HSB2 and CCRFCEM cell lines. Our attempts to transfect HSB2 were unsuccessful because this cell line was very difficult to transfect. We tried various voltages, but at higher voltages the cells died immediately after electroporation and at lower voltages there were no transfected cells. The transfection of DU528 and CCRFCEM was successful up to the single cell colony stage. However, these cell lines were very difficult to grow in isolation. Consequently, we decided to focus on the ARR cell lines and continue our investigation into HEB deletion.
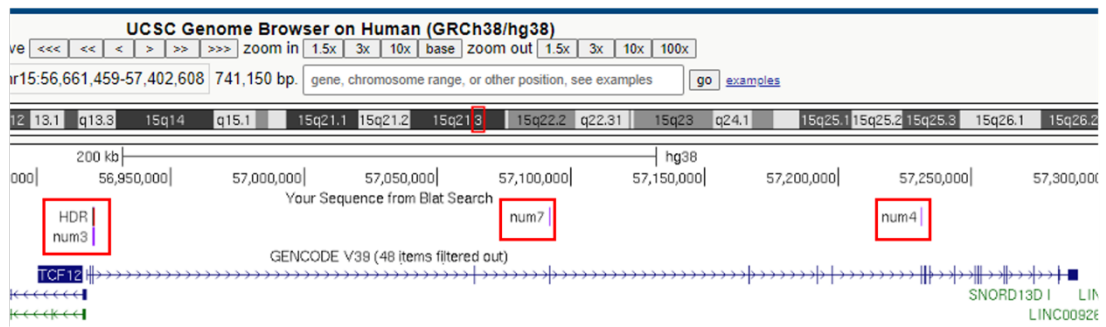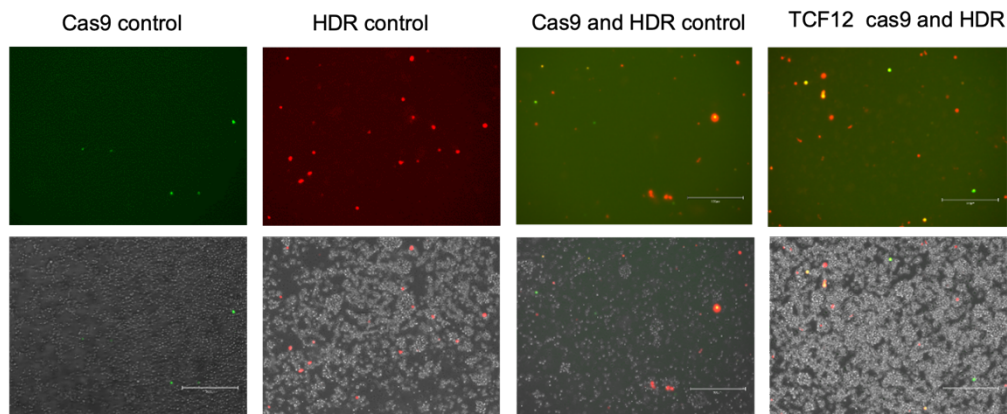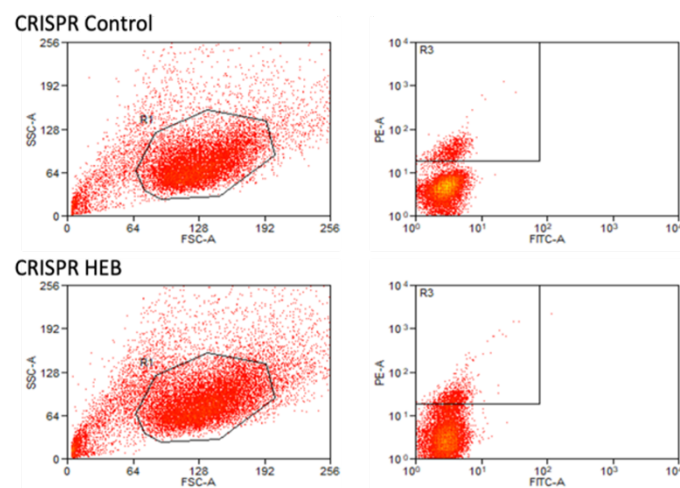
Figure 5.1: CRISPR Cas9 experiment to knockout HEB.

ARR cell line transfected with HEB CRISPR/Cas9 KO plasmid, Cas9 control plasmid and Homology-directed repair (HDR) plasmid. (A) The genomic target site of each plasmid, highlighted in red square and labelled as; N-terminal (num3), Middle (num7) and C-terminal (num4). (B) Transfected cells were examined by fluorescence microscopy to confirm transfection efficiency. (C) Cells sorted using FACs.

A total of six samples were generated from the control transfected cells, and eleven samples were generated from the HEB KO cells. All 18 samples including wild type ARR were extracted, and western blots were performed using an antibody against HEB (Figure 5.2A). In comparison with ARR WT, sample number 5 showed a clear loss of HEB at a size of 100 kDa. The expression of HEB was reduced in several other samples in comparison with the normal cell line; these samples were likely to be heterozygotes. Following this, two controls (control 1 and control 5), two likely heterozygote samples ($TCF12^{+/-}2$ and $TCF12^{+/-}3$), and the knockout sample ($TCF12^{-/-}$) were selected for determining the level of expression of multiple proteins of interest (Figure 5.2B).

We observed that the knockout was for the canonical isoform of HEB, whereas the alternative isoform showed a decrease in expression compared to the control. There was a reduction in GFI1 expression in the $TCF12^{-/-}$ sample detected by all three antibodies from different companies. The expression of TAL1, LMO2, PHF6 and LYL1 was reduced when HEBcan was deleted, RUNX1, LDB1 and PRMT5 displayed no difference in protein level in all samples, while G9a exhibited an increase in expression level when HEBcan was deleted compared to the control sample. Ponceau S was used to confirm equal protein loading.
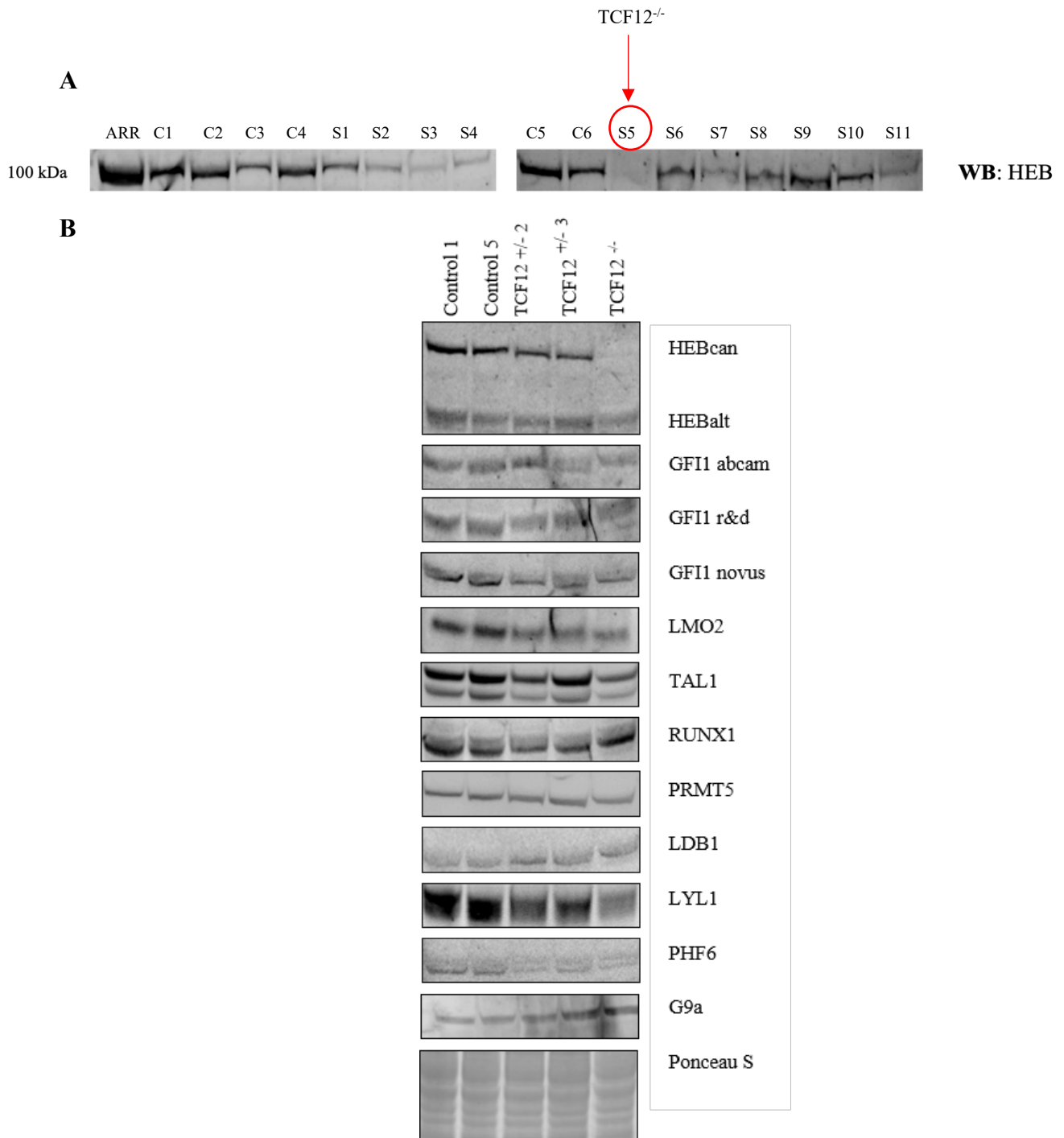
Figure 5.2: Protein expression level measured using Western blot.

(A) HEB Western blot to measure proteins expression level for 6 control samples and 11 KO HEB samples. (B) control 1, control 5, two likely heterozygote samples (*TCF12*$^{+/-}$2 and *TCF12*$^{+/-}$3), and the knockout sample 5 (*TCF12*$^{-/-}$) selected to measure the expression level of different proteins. Ponceau S staining was used to establish equal loading (representative of three technical repeats).

Following this, immunocytochemistry staining was performed on ARR, Cas9 control sample, *TCF12*$^{+/-}$ and *TCF12*$^{-/-}$ samples to determine whether HEB knockout affects protein localization. The cells were fixed with formaldehyde then permeabilized and blocked with FBS. Incubation of the cells with anti-HEB antibody overnight at 4°C was followed by incubation with a secondary antibody to mouse IgG for 1 hour at room temperature. The cells were washed and loaded onto slides, then imaged using a laser confocal microscope. DAPI was used to stain nuclear DNA (in blue) (Figure 5.3A). We observed nuclear localization of HEB, in line with what we described in chapter three. Based on the results of ARR, it is clear that HEB (magenta colour) exhibits the same pattern as shown by DAPI localization in the nucleus. The *TCF12*$^{+/-}$ sample exhibited the same pattern as the ARR sample and the cas9 sample. Compared to wild type ARR and cas9 control, HEB knockout by CRISPR cas9 resulted in weak to absent nuclear staining pattern and intensity of colour, this also confirms the specificity of the antibody used. Immunostaining in Figure 5.3B was conducted without the use of HEB primary antibody in order to confirm that the results observed were not due to nonspecific binding of fluorescent secondary antibodies to the samples.

A

	DAPI	HEB	Overlay
ARR			
Cas9 Control			
TCF12$^{+/-}$			
TCF12$^{-/-}$			
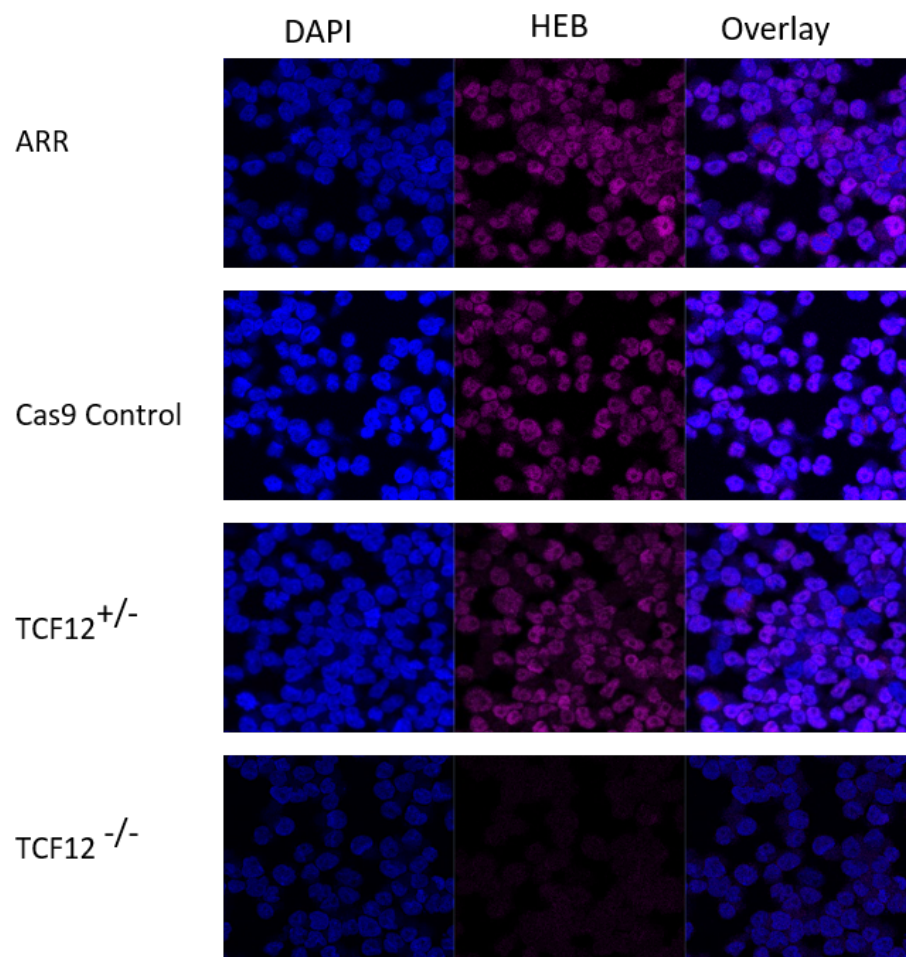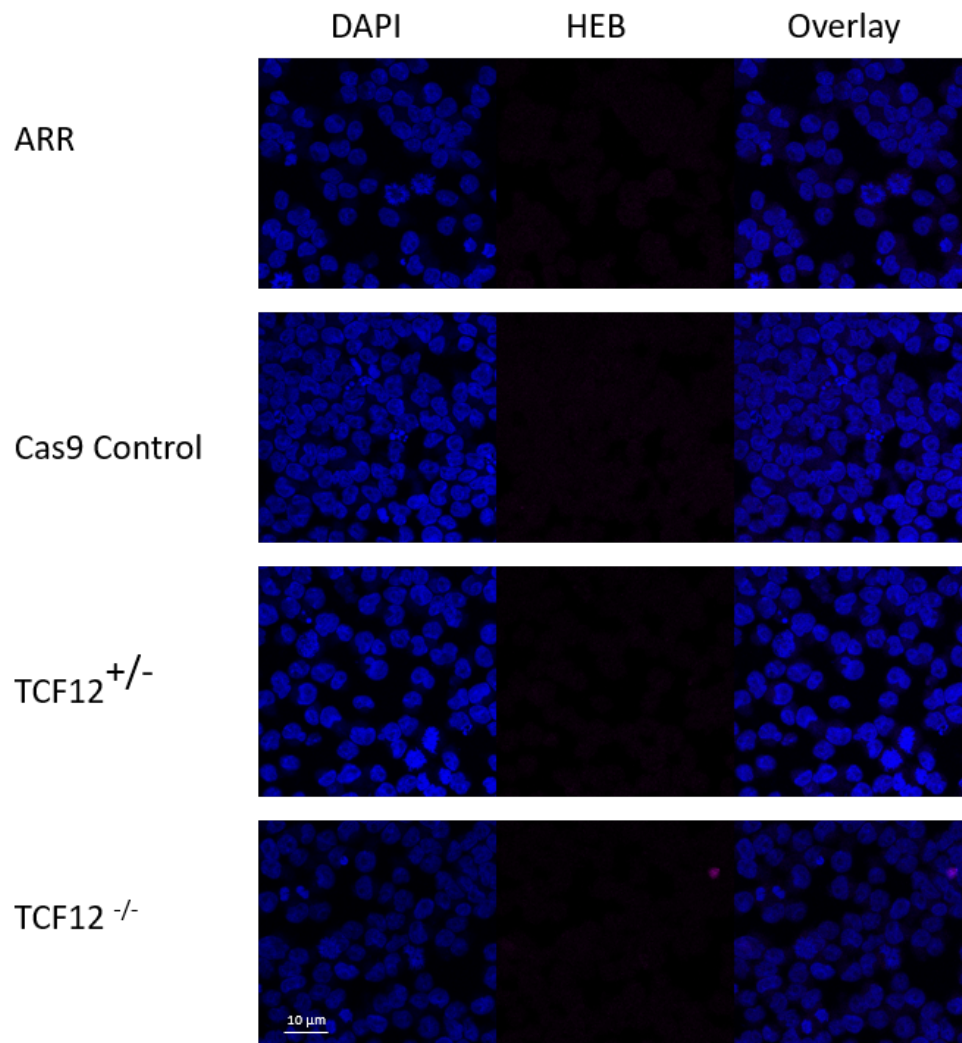
Figure 5.3: HEB Knockout affects cellular localisation in ARR.

(A) ARR cell line, Cas9 control 1, control 5, two likely heterozygote samples (*TCF12*^{+/-}2 and *TCF12*^{+/-}3), and the knockout sample (*TCF12*^{-/-}) were fixed and stained with anti-HEB (magenta) antibody. Slides were imaged using immunofluorescence confocal microscopy, DAPI (blue) staining was used to visualise nuclei. Scale bar 10 µm. (B) immunostaining experiment without primary antibody to test for nonspecific binding of fluorescent secondary antibodies to the samples (n=3).

To exclude the possibility that the single cell clone had an aberrant karyotype and to investigate potential defects in chromosome stability, we performed chromosomal spread analyses to detect any abnormalities. Mitotic chromosome spreads were performed on ARR wild-type, cas9 control sample and *TCF12*^{-/-} sample to evaluate chromosomal integrity and karyotype. A minimum of 20 metaphases were counted in each sample to determine the number of chromosomes. To arrest mitotic cells at metaphase, Colcemid was applied 24 hours following passaging. After pelleting the cells, they were swollen by hypotonic treatment with KCl. Following this, a cold Carnoys fixative solution was added. Cells were then fixed, and then twice washed in cold Carnoys fixative. They were then dropped onto a humidified, chilled glass slide. Phase contrast microscopy was used to assess mitotic index, quality of metaphase spread, and the presence of cytoplasm. The slides were placed overnight at 60° C on a heating block. Giemsa banding (G-banding) technique was used to stain chromosomes during metaphase spreading. The chromosomes were visually analysed and counted under a light microscope. The results of our analysis indicate that all samples including *TCF12*^{-/-} exhibited a normal karyotype, suggesting that the selected clones had a normal karyotype, and that chromosome stability was not majorly affected by *TCF12* knockout (Figure 5.4).

Following this we examined the cells' morphology using Kwik-Diff staining. The cells were loaded into cytospin slide chambers and centrifuged. Following the drying of the slides, the slides were dipped in Fixative solution, and immediately transferred to the Eosin solution to stain the cytoplasm and then to the Methylene blue solution to stain the nucleus. Slides were rinsed in water to remove excess staining, air-dried, then slides were mounted, observed, and scanned under bright field microscopy.

Our results showed that there were no significant morphological changes in *TCF12*^{-/-} ARR cell line in comparison to wild type ARR and control (Figure 5.5). We assessed 10 slides for

To exclude the possibility that the single cell clone had an aberrant karyotype and to investigate potential defects in chromosome stability, we performed chromosomal spread analyses to detect any abnormalities. Mitotic chromosome spreads were performed on ARR wild-type, cas9 control sample and *TCF12*[-/-] sample to evaluate chromosomal integrity and karyotype. A minimum of 20 metaphases were counted in each sample to determine the number of chromosomes. To arrest mitotic cells at metaphase, Colcemid was applied 24 hours following passaging. After pelleting the cells, they were swollen by hypotonic treatment with KCl. Following this, a cold Carnoys fixative solution was added. Cells were then fixed, and then twice washed in cold Carnoys fixative. They were then dropped onto a humidified, chilled glass slide. Phase contrast microscopy was used to assess mitotic index, quality of metaphase spread, and the presence of cytoplasm. The slides were placed overnight at 60° C on a heating block. Giemsa banding (G-banding) technique was used to stain chromosomes during metaphase spreading. The chromosomes were visually analysed and counted under a light microscope. The results of our analysis indicate that all samples including *TCF12*[-/-] exhibited a normal karyotype, suggesting that the selected clones had a normal karyotype, and that chromosome stability was not majorly affected by *TCF12* knockout (Figure 5.4).

Following this we examined the cells' morphology using Kwik-Diff staining. The cells were loaded into cytospin slide chambers and centrifuged. Following the drying of the slides, the slides were dipped in Fixative solution, and immediately transferred to the Eosin solution to stain the cytoplasm and then to the Methylene blue solution to stain the nucleus. Slides were rinsed in water to remove excess staining, air-dried, then slides were mounted, observed, and scanned under bright field microscopy.

Our results showed that there were no significant morphological changes in *TCF12*[-/-] ARR cell line in comparison to wild type ARR and control (Figure 5.5). We assessed 10 slides for

each sample at 24 hours and 48 hours after passaging and found that there were no morphological changes between the samples.
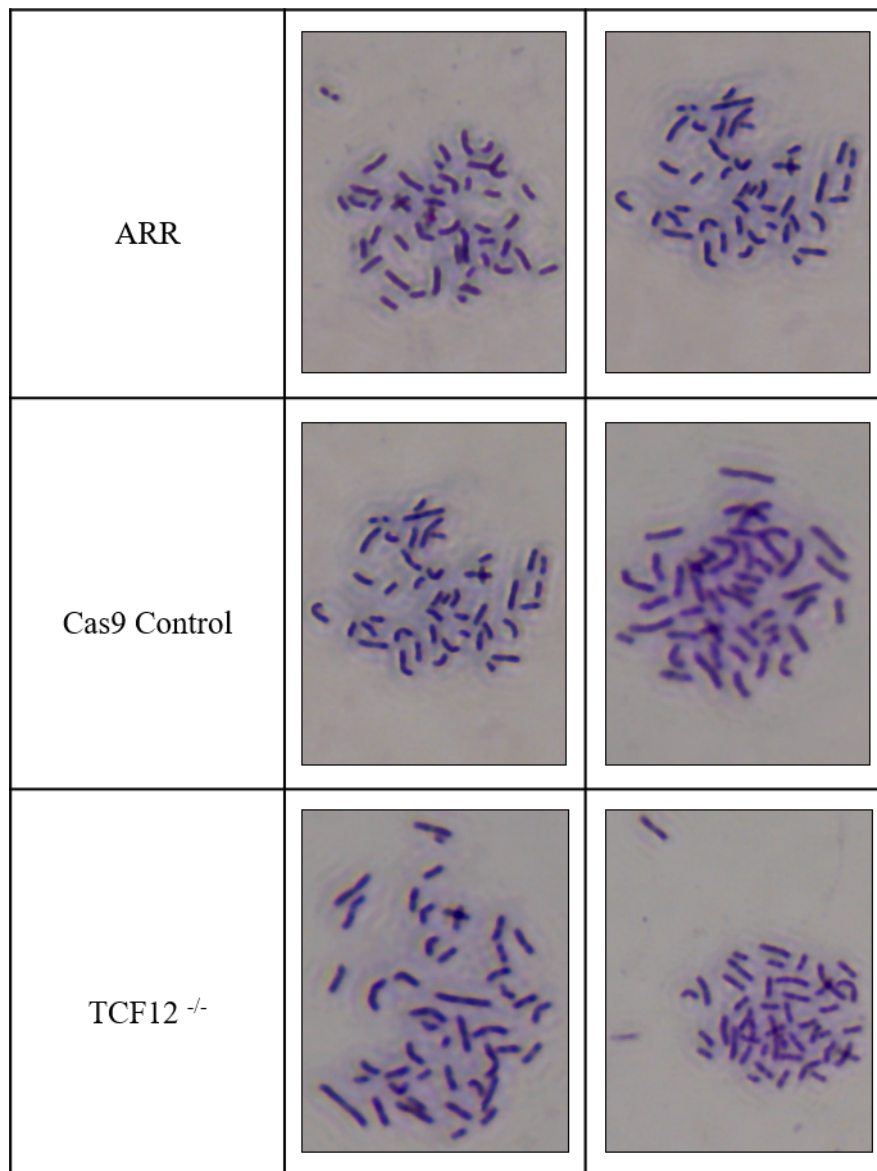
Figure 5.4: Karyotype analysis of ARR wild-type, Cas9 Control, and *TCF12*[-/-] samples.

To assess the chromosomal integrity and karyotype of ARR wild-type, cas9 control, and *TCF12*[-/-] samples, mitotic chromosome spreads were prepared. Each sample was examined for a minimum of 20 metaphases in order to determine its chromosome count. Light microscopes were used to examine and enumerate chromosomes.
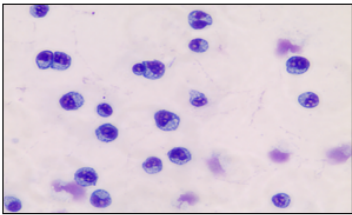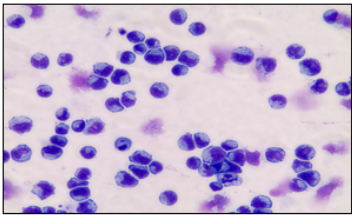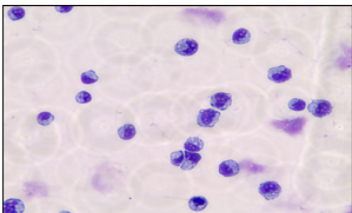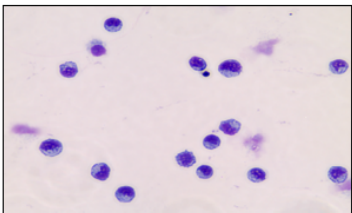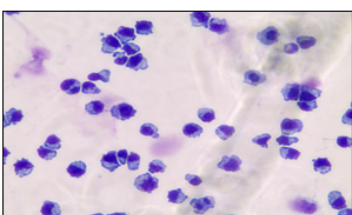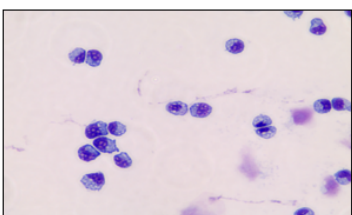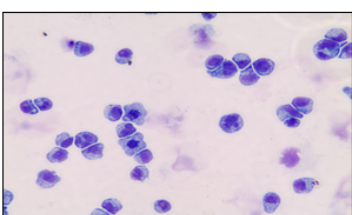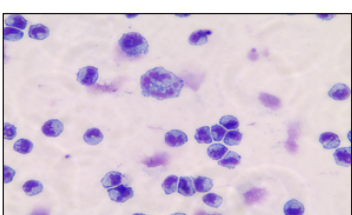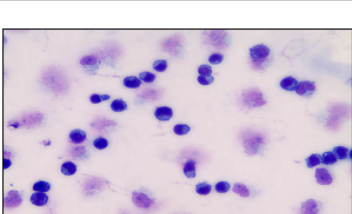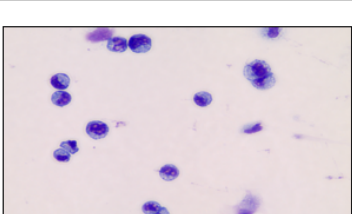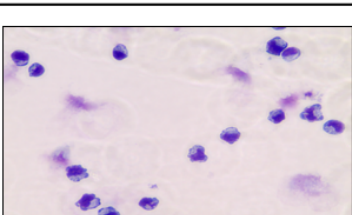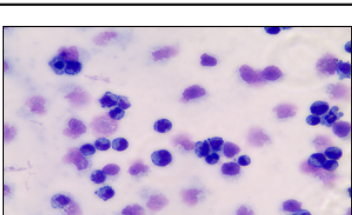
Figure 5.5: Morphological assessment and cell proliferation in ARR following *TCF12* knockout.

For each sample, morphological assessment was conducted on ten slides 24 hours and 48 hours post-total extraction, revealing a consistent pattern among all samples. The *TCF12⁻/⁻* ARR cell line showed no noticeable morphological differences from wild-type ARR and control samples.

DNase I hypersensitive sites (DHSs) are regions of chromatin that are more accessible than the rest of the chromatin and therefore are preferentially cleaved by DNase I when added into the cells. This remodelled state is necessary for the binding of proteins, for example TF, which makes these accessible chromatin sites functionally associated with transcriptional regulation (Boyle et al., 2008). We wanted to know whether the absence of HEB impacted on the presence or characteristics of DHSs. The DNaseI hypersensitive site mapping in our project was carried out to investigate the accessibility of chromatin of two samples of cas9 control, two samples of likely heterozygote and $TCF12^{-/-}$ in ARR.

In brief, DNaseI hypersensitive sites digestions were performed using different concentrations of DNase I to determine the optimal conditions for the assay. The cells were centrifuged, washed, then resuspended in DNase I resuspension buffer. Each sample was then incubated with a DNaseI mix containing different DNaseI concentrations. Cell lysis buffer was added to stop the digestion reaction. Proteinase K was used to digest proteins and RNase A for RNA degradation. DNA samples were isolated by organic extraction followed by precipitation using isopropanol. Agarose gel electrophoresis was performed. DNA fragments from 100-400 base pairs were isolated using a QIAGEN mini-elute gel extraction kit. Finally, the Illumina protocol was used to prepare the library followed by sequencing. Reads from the DHS were mapped to the human genome hg38 and analysed (Figure 5.6).

Figure 5.6: DNase I hypersensitivity profiling of ARR Cas9 control and *TCF12*^{-/-}.

This figure illustrates genomic analysis of DNase I hypersensitive sites (DHS) of Cas9 control and *TCF12*^{-/-}. UCSC Genome Browser tracks are shown, with vertical bars indicating regions of DNase I hypersensitivity. It can be observed that DHS sites are distributed differently in the control and *TCF12*^{-/-} deficient samples, suggesting that chromatin accessibility has been altered.

DNaseI-seq experiments identified 18,896 peaks for cas9 control, 9,459 peaks for *TCF12*^{-/-}, with 5,800 in common between the two. We were interested in determining whether deletion of *TCF12* from ARR cells had any noticeable differences in motifs enrichment comparing to ARR WT cells. Thus, we used MEME to analyse the enriched DNA motifs. The unique peaks of Cas9 control and *TCF12*^{-/-} were analysed. We identified potential binding sites of RUNX, STAT, NFYB, IKZF3, and Spi1 enriched motifs in the control DHS and *TCF12*^{-/-} DHS (Table 5.1). In addition to that, the control DHS showed enriched motifs for KLF1, ETS2, JDP2, NFE2, while the *TCF12*^{-/-} DHS was enriched for Atf3, FOSL1 (Table 5.2).

The next step was to analyse the association between DHS control and *TCF12*^{-/-} accessible genomic regions, as well as their putative target genes. As a result of the analysis conducted using the GREAT online tool, more than 15% of the peaks were found at gene transcription start sites (TSSs) in both samples. Approximately 30% of the bound regions were located between 5 and 50 kb from the TSS, while 40 to 50% were located between 50 and 500 kb away from the TSSs. Less than 10% of the sites were more than 500 kb distant. There was no

Figure 5.6: DNase I hypersensitivity profiling of ARR Cas9 control and *TCF12*$^{-/-}$.

This figure illustrates genomic analysis of DNase I hypersensitive sites (DHS) of Cas9 control and *TCF12*$^{-/-}$. UCSC Genome Browser tracks are shown, with vertical bars indicating regions of DNase I hypersensitivity. It can be observed that DHS sites are distributed differently in the control and *TCF12*$^{-/-}$ deficient samples, suggesting that chromatin accessibility has been altered.

DNaseI-seq experiments identified 18,896 peaks for cas9 control, 9,459 peaks for *TCF12*$^{-/-}$, with 5,800 in common between the two. We were interested in determining whether deletion of *TCF12* from ARR cells had any noticeable differences in motifs enrichment comparing to ARR WT cells. Thus, we used MEME to analyse the enriched DNA motifs. The unique peaks of Cas9 control and *TCF12*$^{-/-}$ were analysed. We identified potential binding sites of RUNX, STAT, NFYB, IKZF3, and Spi1 enriched motifs in the control DHS and *TCF12*$^{-/-}$ DHS (Table 5.1). In addition to that, the control DHS showed enriched motifs for KLF1, ETS2, JDP2, NFE2, while the *TCF12*$^{-/-}$ DHS was enriched for Atf3, FOSL1 (Table 5.2).

The next step was to analyse the association between DHS control and *TCF12*$^{-/-}$ accessible genomic regions, as well as their putative target genes. As a result of the analysis conducted using the GREAT online tool, more than 15% of the peaks were found at gene transcription start sites (TSSs) in both samples. Approximately 30% of the bound regions were located between 5 and 50 kb from the TSS, while 40 to 50% were located between 50 and 500 kb away from the TSSs. Less than 10% of the sites were more than 500 kb distant. There was no

difference between the two DHS samples, control and *TCF12*$^{-/-}$, in these percentages. Interestingly, the majority of genomic sites were between 50 and 500 kb away from the TSSs (Figure 5.7)

.

Table 5.1: Enriched motif analysis of cas9 control DHSs.

The table below illustrates the results of DNA motif analysis carried out on Cas9 control and *TCF12*^{-/-} samples from the DNAse I Hypersensitivity Sites (DHS) region. Using the MEME suite, each sample was analysed independently, excluding common peaks. (Table 5.1) demonstrate enriched DNA motifs identified in control sample, (Table 5.2) demonstrate enriched DNA motifs identified in HEB knockout sample. Each row represents a motif enriched in the DHS ChIP-seq data. Proteins predicted to interact with these motifs are listed under (known or similar motifs). E-value: Indicates the statistical significance of the motif enrichment. Sites: Indicates how many times the motif occurs.

Motifs	*E*-value	Sites	Known or similar motifs
	8.06e-03 7.24e-01	52	RUNX1, RUNX2, RUNX3
	1.81e-02	83	KLF1, KLF5, SP1, E2F6, ETS2, ELK1
	1.27e-03	185	IKZF1, Ikzf3, Spi1, EHF, GABPA
	2.22e-02	32	STAT1,STAT3, STAT5
	2.25e-04	94	NFYA, NFYB, NFYC, Hoxa13
	1.32e-04	25	JDP2, NFE2

Table 5.2: Enriched motif analysis of *TCF12*-/- DHSs.

Motifs	*E*-value	Sites	Known or similar motifs
	5.21e-03	40	Bcl11B, RUNX1, RUNX2, RUNX3
	2.66e-04	52	Atf3, FOSL1,
	2.75e-03	40	IKZF3, Spi1, ELF1, ELK4
	7.32e-02	24	STAT1,STAT3, STAT5
	3.31e-01	37	NFYB
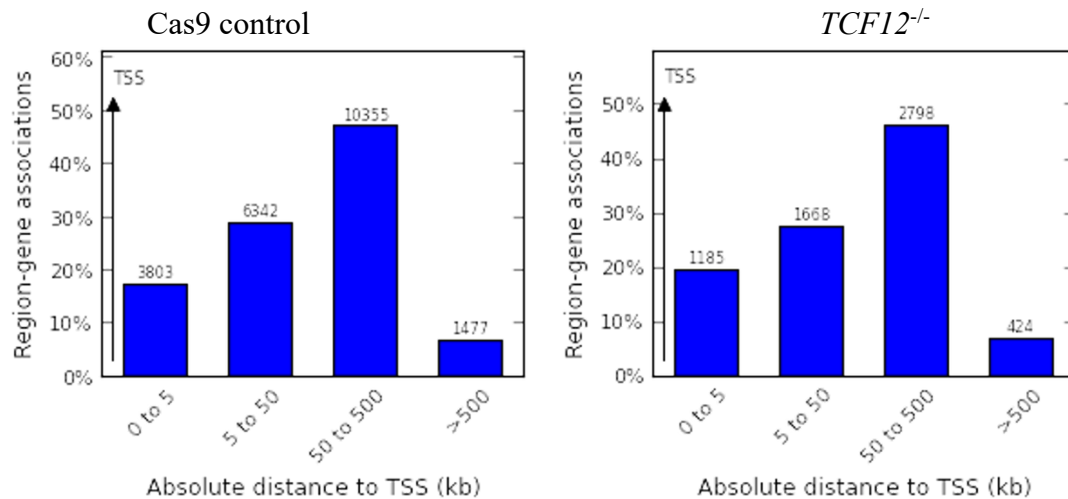
Figure 5.7: Genomic distribution of DHS peaks.

Bar chart illustrates the genomic distribution, distance and orientation of DHS peaks to transcription start sites- TSS, using GREAT online tool, in cas9 control (left) and *TCF12* knockout (right). On the y-axis, the percentage of genes is given, and above each bar in the graph is the absolute number of genes being counted.

To determine the roles of these genes in biological processes, we further analysed the gene ontology of the peaks specific to the Cas9 control or *TCF12*[-/-], using DAVID 6.8 (Huang da et al., 2009). For DHS cas9 control regions, gene ontology analysis revealed an enrichment of genes associated with proton transport regulation, transcription of mRNA catabolic process, apoptotic processes modulation by viral transcription by the host, and phagocytosis negative regulation (Figure 5.8A). On the other hand, the gene ontology for *TCF12*[-/-] DHSs revealed the enrichment of genes involved in myeloid leukocyte differentiation, regulatory T cell differentiation, and cytokine biosynthesis processes, lipid kinase activity, and histone acetylation (Figure 5.8B).

(A)

GO Biological Process

-log10(Binomial p value)

	0	1	2	3	4	5	6	7	8	9

positive regulation of mRNA catabolic process — 9.38
regulation of proton transport — 8.71
positive regulation of transcription from RNA polymerase II promoter in response to stress — 8.31
positive regulation of nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay — 8.26
epithelial cell apoptotic process — 7.82
modulation by host of viral transcription — 7.62
negative regulation of phagocytosis — 6.78
response to interleukin-15 — 3.93

(B)

GO Biological Process

-log10(Binomial p value)

	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5

regulation of cytokine biosynthetic process — 6.65
positive regulation of myeloid leukocyte differentiation — 5.23
serine transport — 5.03
histone H4 acetylation — 4.27
L-serine transport — 4.26
response to fluid shear stress — 4.15
negative regulation of receptor binding — 3.94
interleukin-12-mediated signaling pathway — 3.88
regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway — 3.62
regulation of regulatory T cell differentiation — 3.51
cellular response to fluid shear stress — 3.27
positive regulation of lipid kinase activity — 3.11

Figure 5.8: Gene ontology analysis of DHS peaks.

Gene ontology was performed on DHS peaks in both (A) Cas9 control and (B) *TCF12*$^{-/-}$ samples, analysing unique peaks, excluding common regions. The chart represents the negative logarithm base 10 of the binomial p-values, reflecting the statistical significance of biological processes enrichment among DHS peaks. Each bar corresponds to distinct biological process category.

According to previous results from our lab, many HEB ChIPseq peaks in T-ALL occurred and overlapped with regions of methylated DNA. Consequently, it was interesting investigate the methylation status of the genomic DNA in ARR after HEB knockout and how that influenced transcription factor binding. Methylated DNA Immunoprecipitation (MeDIP), using antibodies specific for 5-Methylcytosine (5mC), was used to determine genome-wide DNA methylation patterns in ARR. An initial step in the MeDIP experiment involved isolating genomic DNA from the control ARR cas9 and the mutant ARR *TCF12*-/-. Sonication was then used to fragment the DNA into smaller pieces. The fragments of DNA were then incubated with anti-5-methylcytosine antibody, allowing it to bind to the methylated regions of DNA. The DNA-antibody complexes were immunoprecipitated with protein G magnetic beads. Finally, methylated DNA was eluted from the beads and purified. The purified DNA was then used to prepare a MeDIP library, which was carried out in accordance with the Illumina protocol.

Bioinformatic analysis of MeDIP-seq data identified 50,065 peaks in control cas9 ARR and 28,407 peaks in *TCF12*-/-. A total of 11566 peaks were found to be common between the two MeDIP samples. To identify DNA sequences statistically overrepresented within MeDIP-seq data binding regions, we performed *de novo* motif analysis. In order to identify motifs specific to Cas9 control or *TCF12*-/- ARR, the common peaks were excluded from the analysis. Interestingly, in the cas9 control sample, there was a significant enrichment for sequences that could bind GFI1B, Ikzf3, and SOX9 (Table 5.3). In contrast, GFI1 was enriched in *TCF12*-/- samples as well as TCF3 (E2A), prdm5, and Zbtb3 (Table 5.4).

We then examined the association between MeDIP control and *TCF12*-/- mutually bound genomic regions. A GREAT online analysis revealed that less than 10% of the peaks were located at gene TSSs in both samples. Between 20% and 30% of the control sample bound

regions were located between 5 and 50 kb away, while over 30% of the *TCF12*^{-/-} sample bound regions were located between 5 and 50 kb away from TSSs. In the control sample, approximately 60% were located between 50 and 500 kb from the TSSs, while approximately 55% were found in the *TCF12*^{-/-} sample. Finally, there were about 10% of sites more than 500 kb apart in both samples (Figure 5.9).

Table 5.3: Enriched motifs in Cas9 control sample of MeDIP ChIP-seq data.

The tables below show DNA motif analysis carried out on Cas9 control and *TCF12*-/- samples from the MeDIP region. Using the MEME suite, each sample was analysed independently, excluding common peaks. (Table 5.3) demonstrate enriched DNA motifs identified in control sample, (Table 5.4) demonstrate enriched DNA motifs identified in HEB knockout sample. DNA motifs identified in HEB knockout sample. Each row represents a motif enriched in the DHS ChIP-seq data. Proteins predicted to interact with these motifs are listed under (known or similar motifs). E-value: Indicates the statistical significance of the motif enrichment. Sites: Indicates how many times the motif occurs
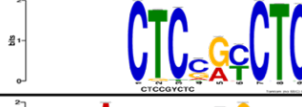
Motifs	*E*-value	Sites	Known or similar motifs
	8.37e-01	218	Gfi1B
	1.96e-02	170	SOX8, SOX9, SOX10
	5.90e+00	182	Ikzf3

Table 5.4: Enriched motifs in *TCF12*-/- sample of MeDIP ChIP-seq data.

Motifs	*E*-value	Sites	Known or similar motifs
	1.04e-02 9.16e-01	394	SOX10, SOX8, SOX9 GFI1
	1.84e-01	363	Zbtb3
	9.64e-01	384	Prdm5
	9.64e-01 8.29e+00	442	MA1710 (ZNF257) TCF3 (E2A)

165
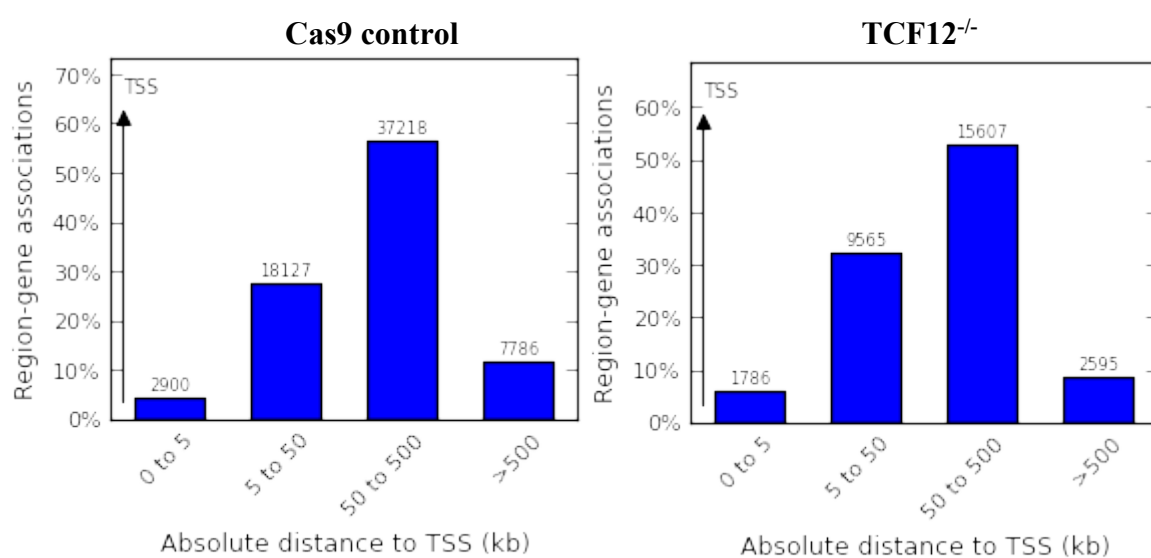
Figure 5.9: Genomic distribution of MeDIP peaks.

The bar chart illustrates the genomic distribution of MeDIP peaks to transcritption start sites (TSS), using GREAT online tool, in cas9 control (left) and *TCF12* knockout (right).

Following this, we conducted a ChIP experiment to determine the effect of deleting HEB on other transcription factors. ChIP was performed for LMO2, TAL1, PRMT5, PHF6, and LDB1. Quantification of the enrichment was conducted with qPCR, and the results were normalized based on the input control (Figure 5.10). The enrichment was determined using the PU.1 enhancer and the SDE2 element. The PU.1 enhancer is well-characterised and the PU.1 transcription factor plays a crucial role in the development of the haematopoietic system. The SDE2 amplicon was selected based on previous experiments that demonstrated a strong binding of several transcription factors before (Stanulović et al., 2020). As a negative control, a heterochromatic region of chromosome 18 was used that is devoid of genes and does not support binding of transcription factors, as shown in previous ChIP and ChIP-seq experiments.

qPCR is used in this case to assess the suitability of the sample prior to library preparation for sequencing. By doing so, it can help ensure that the starting material is of sufficient quality to produce reliable sequencing data, rather than to perform direct comparisons of the data between the cell lines. Higher qPCR values in ChIP samples indicate higher enrichment or expression levels for these genes. The enrichment at SDE2 is higher than at PU.1 enh in ChIP of LMO2, TAL1, and PHF6. These ChIP experiments suggest that LMO2, TAL1, and PHF6 genes appear to bind to or be more strongly associated with SDE2 genomic regions than PU.1. PRMT5 and LDB1 preferentially bind to or are more strongly associated with the PU.1 genomic region than SDE2. Altogether, we clearly detected protein binding for all the performed ChIP experiments at one or more of the tested amplicons, indicating that our ChIP experiment was successful.
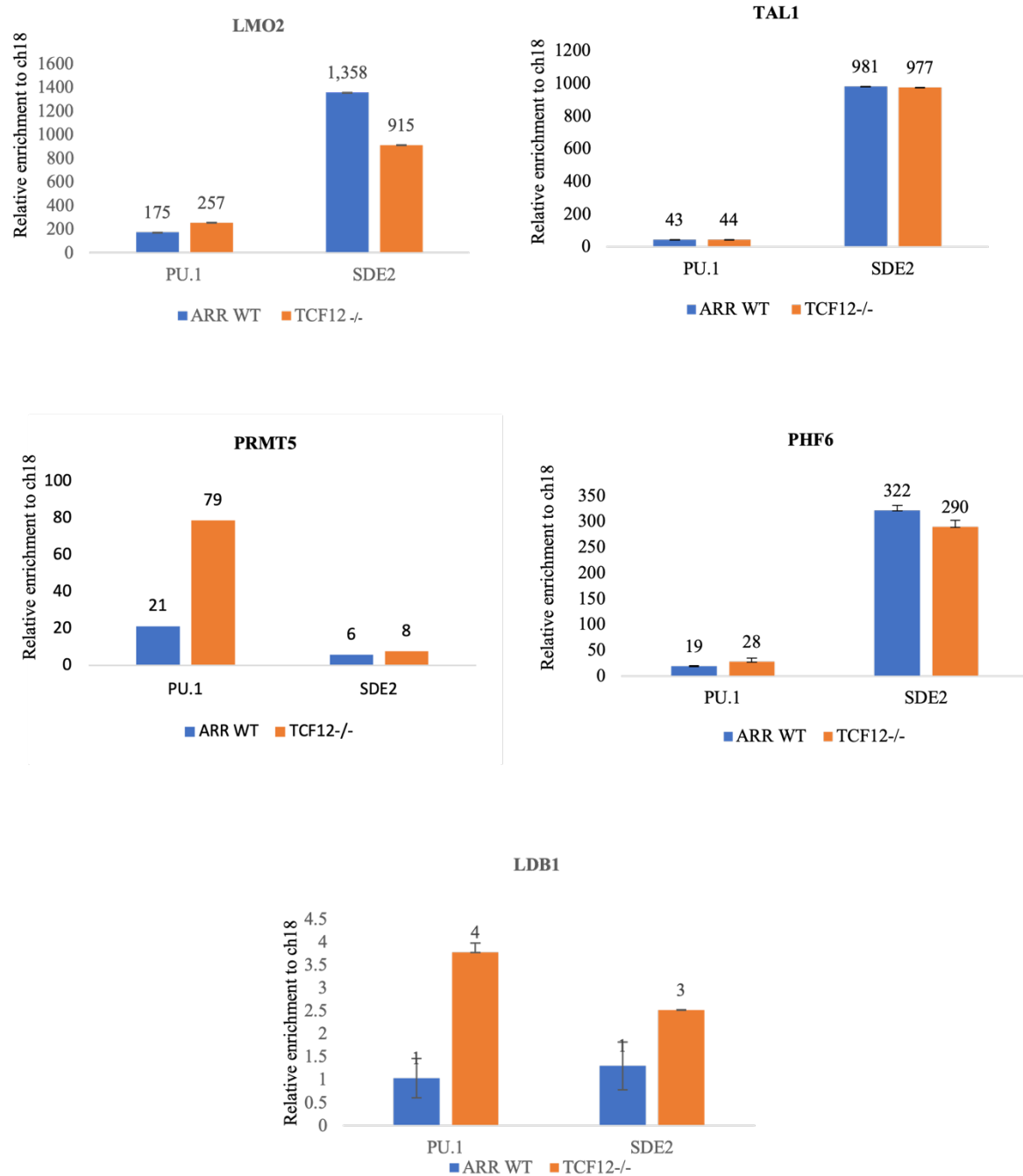
Figure 5.10: Enrichment of control and *TCF12*⁻/⁻ samples at the PU.1 and SDE2 in ARR.

ChIP was perfumed on ARR cas9 control and *TCF12*⁻/⁻ samples using LMO2, TAL1, PRMT5, PHF6 and LDB1 antibodies, then quantified using qPCR using the indicated primers. The fold enrichment relative to ch18.

After the ChIP experiments were validated by qPCR, ChIP libraries were prepared then samples were loaded onto agarose gels. The gel electrophoresis in Figure 5.11 shows the size distribution of DNA fragments obtained after immunoprecipitating chromatin with a specific protein of interest and subsequent library preparation. The DNA fragments selected to be isolated were between 200 and 300 base pairs in size. DNA fragments in this range are more likely to represent specific binding sites of the protein of interest than larger fragments that may contain multiple binding sites. In addition, they reduce the likelihood of nonspecific background signals in ChIP-seq data. The samples that did not show any results in panel A underwent a second attempt and are presented in panel B. GATA2 and LYL1 ChIP did not work therefore have been excluded from further sequencing analysis.
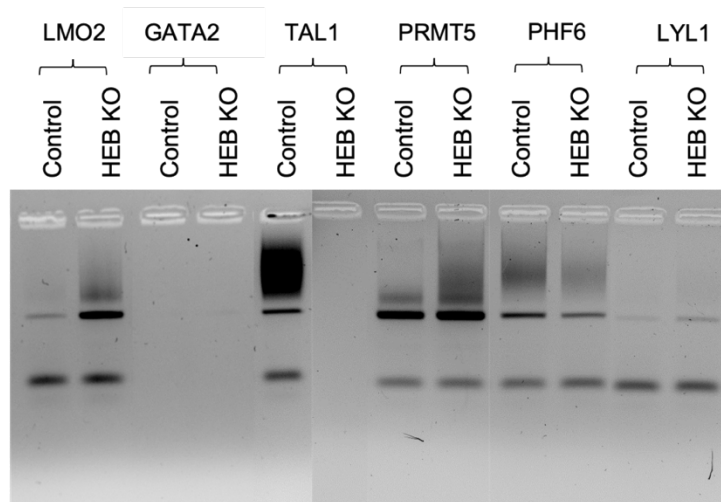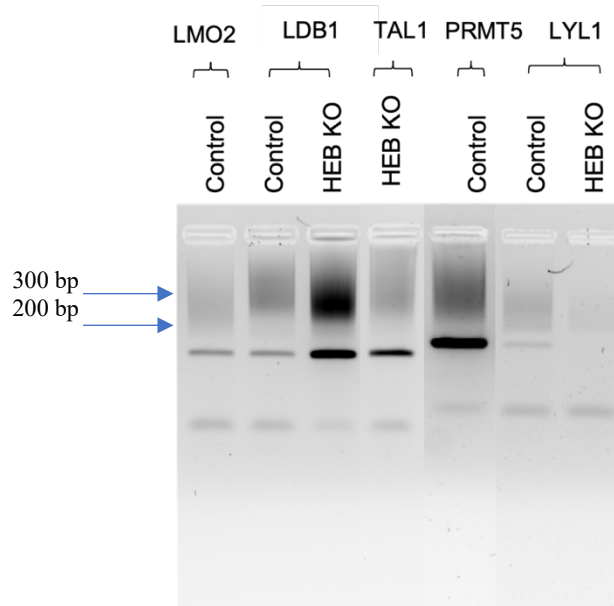
Figure 5.11: Agarose gel analysis of ChIP samples.

Electrophoresis analysis of ChIP samples loaded on 3% agarose gel, the highlighted DNA fragments falling within the 200-300 base pair (bp) range were cut and sent for sequencing.

As a result of the ChIP-seq analysis performed on the ARR cell line following HEB knockout ($TCF12^{-/-}$), transcription factor genomic binding patterns have been found. Among the factors analysed, LMO2 showed the highest number of peaks, with 76,551 identified binding sites, indicating that this factor has a significant role in this cell line. this was followed by TAL1, with a peak count of 36,393. In this cell line, PRMT5, PHF6, and LDB1 contribute to the complex regulatory landscape with respective peaks of 27,015, 23,110, and 17,495 peaks. These findings indicate that these proteins are likely to exert their influence at specific genomic locations, providing a basis for understanding their roles in gene regulation as well as potential implications for cellular function when HEB is absent.

To understand relationships between the different data sets, we used the data visualization method Upset. A Venn Diagram, which is the most used visualization approach for overlaps between data sets, is not effective beyond four sets. The UpSet data analysis method, however, is well suited for the quantitative analysis of data that consists of multiple sets. Data from the Upset plot was plotted as a table (table 5.5). In the following table, the intersections of a set are plotted as a matrix, each row representing a set, with possible intersection; indicated by the letter X. The table illustrates the results of overlap and intersection analyses performed on ChIP-seq data obtained from ARR HEBKO cells involving transcriptional regulatory proteins LMO2, TAL1, PRMT5, PHF6, and LDB1. We have also incorporated ChIP-seq data previously generated in our laboratory for HEB using the ARR cell line with a peak count of 8,615, as a reference. In this study, we examine the genomic binding regions between HEB and the aforementioned factors, with the objective of identifying potential co-binding events between HEB and these factors.

Table 5.5: illustrates the intersections of ChIP-seq data obtained in ARR following HEB knockout.

Table showing the overlap between ChIP-seq peaks of different transcription factors in ARR following HEB knockout (*TCF12*[-/-]). The first column lists the name of the cell line and the total number of peaks. The other columns represent the overlaps and the frequency of their occurrence.

TCF12 (8,615)	X	X				X	X	X	X			X
LDB1 *TCF12*-/- (17,495)	X	X	X		X	X	X	X				
PHF6 *TCF12*-/- (23,110)		X					X			X		
PRMT5 *TCF12*-/- (27,015)		X					X	X			X	
TAL1 *TCF12*-/- (36,393)		X		X	X	X	X	X				X
LMO2 *TCF12*-/- (76,551)		X	X	X					X	X	X	
	958	709	588	481	437	435	300	297	275	269	260	254

Our next step was to create a plotCorrelation analysis using Galaxy software, which can be used to assess and identify patterns and relationships between ChIP-seq samples. A plotCorrelation image shows the correlation coefficients between samples. Dendrograms demonstrate graphically how data points are grouped or clustered based on their similarity or dissimilarity.

The correlations between ChIP-seq data from different transcriptional regulator proteins and experimental conditions can be shown in Figure 5.12. Starting with the ChIP-seq results for HEB under wild-type ARR (*TCF12*) conditions and LDB1 under *TCF12* knockout conditions (*TCF12*[-/-]_LDB1), there appears to be a moderate to strong positive correlation ranging between 0.6 to 0.8. This suggests that HEB and LDB1 may share common genomic binding regions or have a functional relationship within the context of the *TCF12* knockout in the ARR cell line. ChIP-seq data for LDB1 and *TCF12* in the control condition (Ctrl_LDB1) have lower

correlation, around 0.4. This suggests that these two factors may not have a strong association or overlap in binding regions under normal (control) conditions.

As shown in the ChIP-seq data for PRMT5 under the control condition, PRMT5 correlated moderately (0.5-0.6) with both *TCF12* under the control condition (HEB) and LDB1 under the *TCF12* knockout condition (*TCF12*^{-/-}_LDB1). A correlation of approximately 0.3 exists between ChIP-seq data for *TCF12* in the WT ARR condition and *TCF12*^{-/-}_PRMT5 (PRMT5 in the knockout ARR condition). Thus, the knockout of *TCF12* in the ARR cell line has resulted in a reduced level of similarity between HEB binding patterns and PRMT5 binding patterns.

Observing that results of ChIP-seq data for LMO2 in *TCF12* knockout shows weaker correlations with HEB and all other transcriptional factors in both conditions (control and knockout) indicates a significant relationship between HEB and other transcriptional factors, including LMO2, in the regulation of gene expression. Accordingly, LMO2 regulatory activity depends heavily on the presence and functional activity of *TCF12*.

In the absence of HEB, TAL1 exhibits a moderate correlation with HEB and LDB1, but the correlation is weaker in comparison to the control condition of TAL1. Based on these findings, it seems that HEB could enhance the binding pattern or functional interaction between TAL1 and LDB1. In contrast, PHF6 and LMO2 show a similar correlation pattern to PRMT5 when correlated with HEB. The correlation is moderate under the control condition but weakens under the knockout condition. It is therefore concluded that the absence of HEB negatively impacts the alignment or similarity of the binding patterns between these proteins.
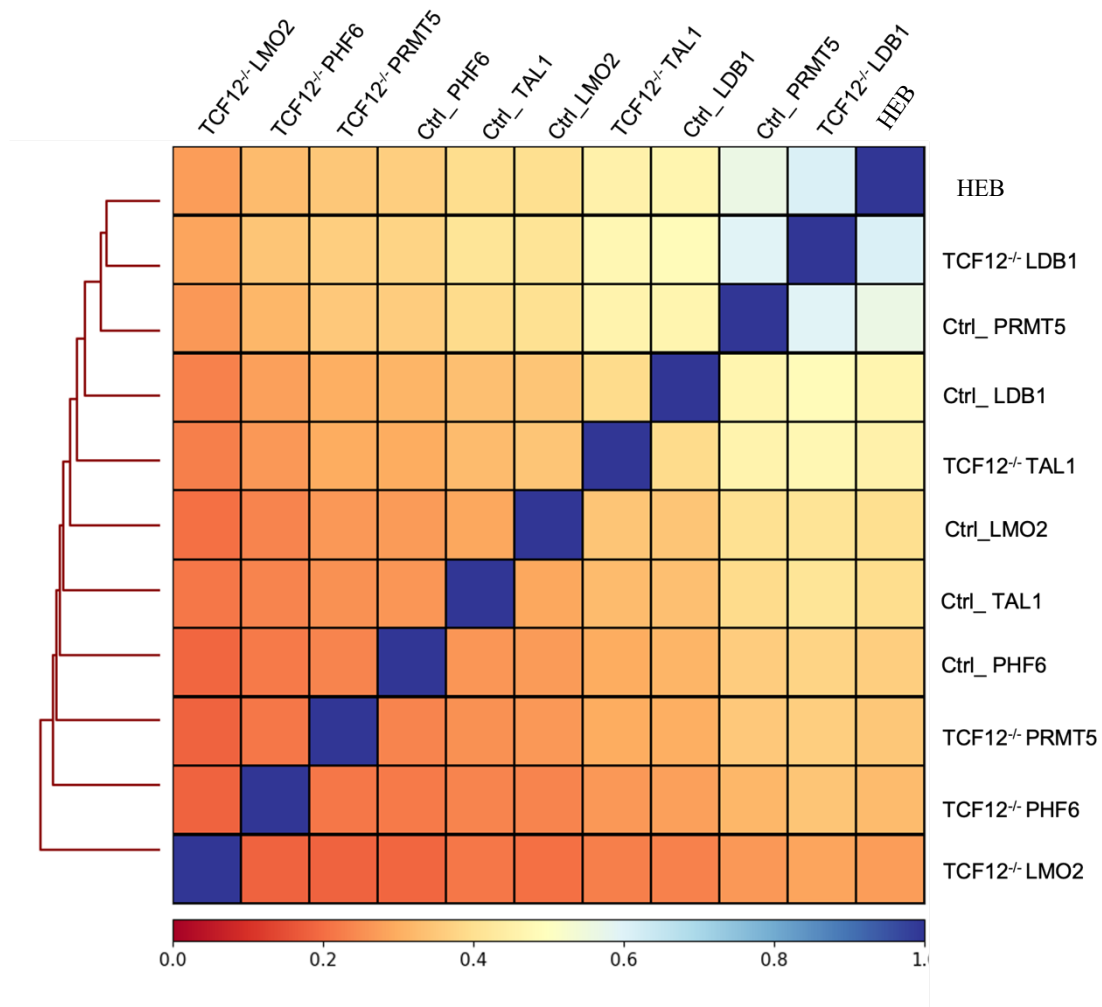
Figure 5.12: Correlation analysis of ChIP-seq data.

The figure represents a plot-correlation analysis generated using Galaxy software, revealing relationships between various transcriptional proteins and experimental conditions.

As a final analysis, we generated heatmaps using the EaSeq software. This analysis uses peak size as an indicator of occupancy, which adds a valuable element to ChIP-seq data comparison. The genomic coordinates of the ChIP-seq peaks are integrated with the read counts within a window of +/- 5kb on either side of each peak, creating a ranking matrix based on read counts. As a result, peaks with higher read counts are given greater prominence in the analysis, providing an opportunity to identify regions of heightened regulatory activity and stronger binding affinity. In designing the figure illustrating DHS peak intensity as a matrix, we divided the peaks into three sections (Figure 5.13). Only reads identified from the control results were included in the top section, resulting in a robust dataset of 13,096 reads. A total of 5,800 reads was shared by control and knockout samples in the middle section. Additionally, the bottom section includes only data from the *TCF12* knockout sample, consisting of 3,695 reads. On the left side of the matrix, we overlay the DHS results, which were used as a foundational control, these results were sorted in descending order. In addition, ChIP-seq data for transcriptional regulators such as LDB1, PHF6, PRMT5, TAL1, and LMO2, as well as data obtained from MeDIP ChIP-seq experiments were included.

The results show that PRMT5 displayed a lower intensity in $TCF12^{-/-}$ compared to the control condition. Other heatmaps representing transcriptional regulatory factors such as PHF6, TAL1, and LDB1 exhibit similar density pattern in the DHS control and $TCF12^{-/-}$. In contrast, LMO2 exhibits a contrasting pattern. Comparing LMO2 in the control and $TCF12^{-/-}$ conditions, the intensity is greater on the $TCF12^{-/-}$ side. Hence, LMO2's may become stronger in the absence of TCF12, resulting in greater DNA accessibility and binding. Upon overlaying MeDIP results on top of the DHS matrix heatmap, the data appear almost empty, indicating that signal or DNA methylation events are not evident in the regions of interest.

According to the heat map depicting the common data found between the DHS control and DHS TCF12 knockout samples, several conclusions can be drawn. Both the common data

between the two samples exhibit the same pattern as the control only sample. The similarity indicates that a significant proportion of the regulatory elements and regions identified in the DHS control sample remain unchanged in the HEB knockout condition, which indicates that there is a substantial overlap between DNA accessibility and regulatory activity under these conditions. There is a strong, dark band at the top of each heat map that indicates a cluster of genomic regions or features with a high level of DNA accessibility. As a result of the dark intensity of this band, it may indicate a group of elements that are particularly active or accessible in both the DHS control and DHS of *TCF12* knockout samples. This band may indicate that key regulatory elements, such as enhancers and promoters, are active in both conditions.

In the analysis of the DHS results of *TCF12* knockout reads (3659 reads) overlaid on the DHS matrix of *TCF12* knockout only, a distinct dark band is observed at the top of the map. This prominent band indicates that there is a group of genomic regions or features that show highly intense DNA accessibility or regulatory activity in the *TCF12* knockout condition. Data from ChIP-seq for PRMT5 overlaid on DHS of *TCF12*$^{-/-}$ demonstrate differences between the control and *TCF12* knockout samples. There were more reads in the control sample than in the *TCF12* knockout sample. The opposite pattern is shown by LMO2, where it is more significant in the knockout sample of *TCF12*. Notably, there were no significant differences between the control and knockout conditions for LDB1, PHF6, and TAL1, which suggests that the binding of these factors was not strongly influenced by the presence of HEB. It can be seen from the MeDIP heatmaps again that both the control and knockout conditions are nearly empty.
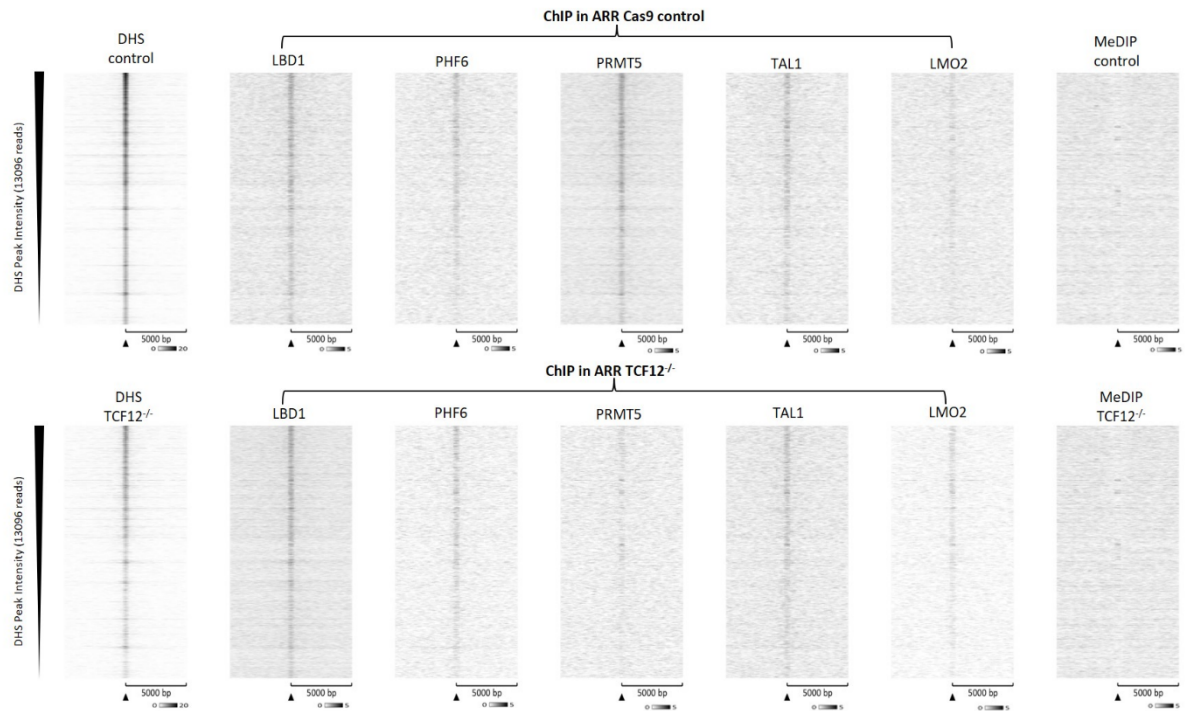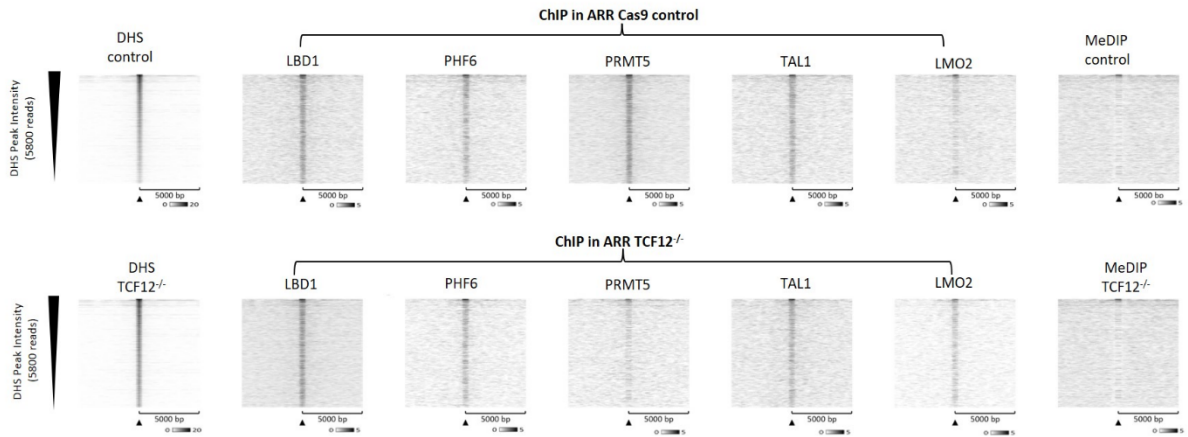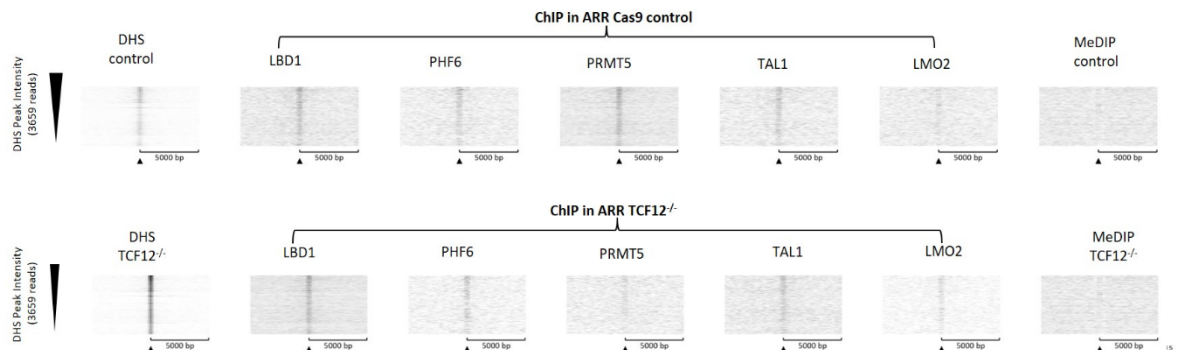
Figure 5.13: Heatmaps analysis of ChIP-seq data using DHS as a matrix.
The figure illustrates heatmaps generated using Easeq software to assess DNA binding strength. ChIP-seq results are ranked according to intensity. Easeq software was used to generate heatmap plots,

depicting windows from ±5 kb around the centre. Greyscale intensity is indicated below the heat maps. The figure illustrates DHS peak intensity as a matrix divided into three sections. DHS ChIP-seq results were categorized into three groups: peaks exclusive to the control sample (13,096 peaks), peaks shared between control and knockout samples (5,800 peaks), and peaks unique to the TCF12 knockout sample (3,695 peaks). DHS results, sorted in descending order, were overlaid on the left side of the matrix. ChIP-seq data for transcriptional regulators (LDB1, PHF6, PRMT5, TAL1, and LMO2) and MeDIP ChIP-seq experiments were included.

To investigate this further, we generated heatmaps using MeDIP data as the matrix. By overlaying MeDIP results for control and knockout samples, as well as integrating ChIP-seq data for LDB1, PHF6, PRMT5, TAL1, LMO2, and DHS ChIP-seq data, we can provide comprehensive insights into the epigenomic landscape, allowing us to explore how DNA methylation interacts with these factors. MeDIP ChIP-seq results were divided into three distinct groups. There were 38,517 peaks that corresponded exclusively to peaks identified in the control sample. There were also 11,566 peaks that were shared between the control and knockout samples. The *TCF12* knockout sample contained 16,841 peaks associated with its unique peak set (Figure 5.14).

Results were sorted in descending order based on peak scores, revealing intriguing patterns. In the first part, the heatmap representing MeDIP data for both the control and $TCF12^{-/-}$ conditions demonstrated a distribution of peaks throughout the heatmap. The distribution patterns of ChIP-seq data for transcriptional factors LDB1, PHF6, TAL1, and PRMT5 were observed to be similar in control and knockout. In contrast, LMO2 displayed a slightly weaker and less intense pattern in the knockout sample. A notable observation was a small line of data in the DHS data that was overlaid on top of the MeDIP-seq data in both control and knockout samples. This indicates a specific region or feature where DNA methylation and chromatin accessibility coincide, shown in red in Figure 5.14.

In the analysis of common peaks, the overlapping data from ChIP experiments revealed an interesting colocalization pattern: the peaks with the highest scores corresponded exactly to the MeDIP regions in both control and knockout samples. ChIP results for transcription factors

LDB1, PHF6, LMO2, PRMT5, and TAL1 consistently showed higher peak intensities within the MeDIP-associated regions. ChIP experiments produced an intriguing observation based on overlapping data: the peaks with the highest scores were precisely colocalized with MeDIP regions both in control and knockout samples. this suggest that DNA methylation has a strong and consistent association with transcriptional factor binding.

Interpretation of the peaks found exclusively in the knockout sample provides several important findings. Overlaid on the knockout matrix, the MeDIP of $TCF12^{-/-}$ showed a darker signal and higher peak intensity than the control sample. It appears that more genomic regions exhibit DNA methylation in the absence of $TCF12$. It is apparent from the darker signal in the $TCF12^{-/-}$ knockout sample that the amount of methylation was more comparing to the wild type cells.

MeDIP Cas9 control only matrix
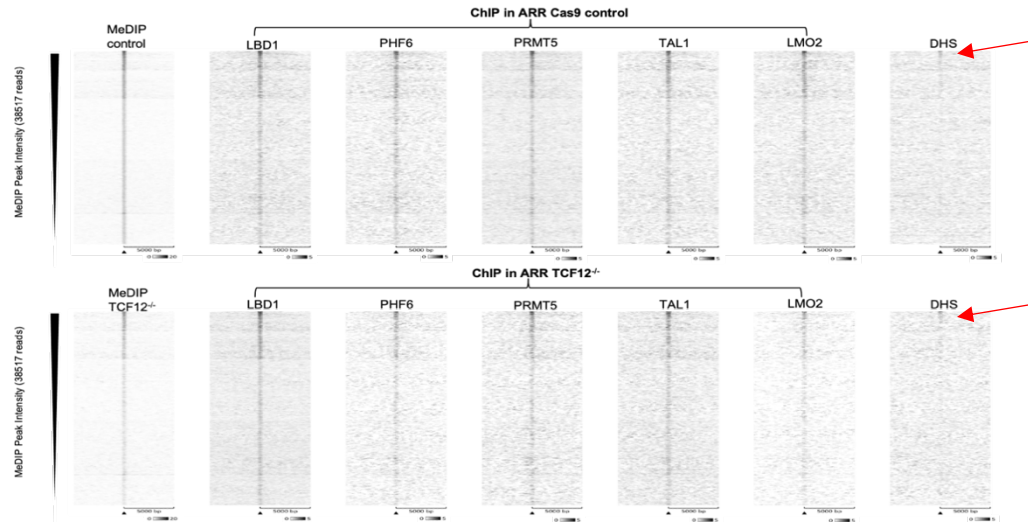
MeDIP common in control and TCF12⁻/⁻ matrix
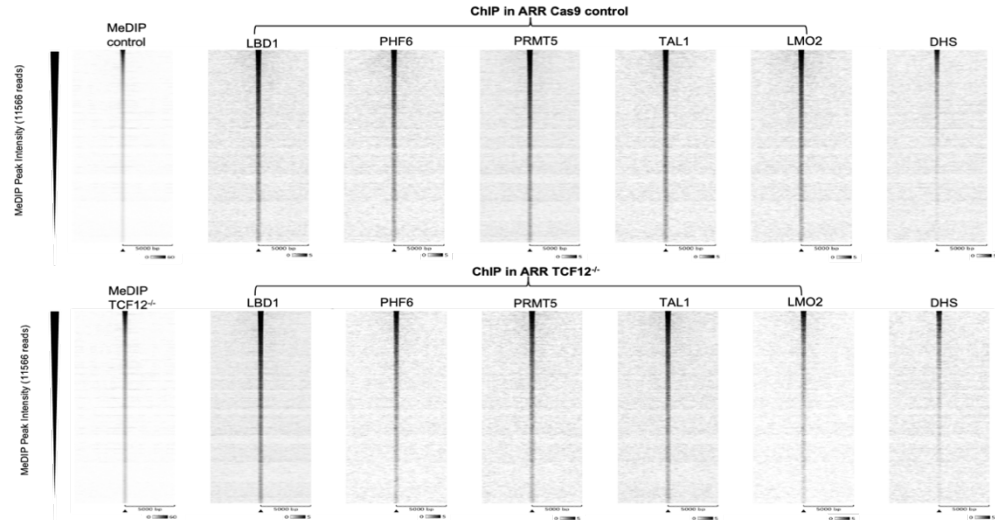
MeDIP TCF12⁻/⁻ only matrix
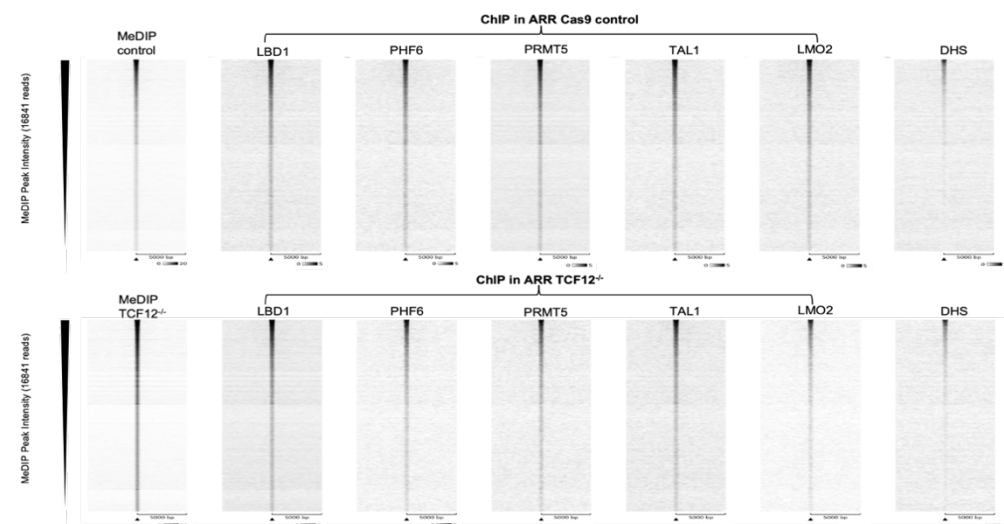
Figure 5.14: Heatmaps analysis of ChIP-seq data using MeDIP as a matrix.

The figure illustrates heatmaps generated using Easeq software to assess DNA binding strength using MeDIP as a matrix. ChIP-seq results ranked according to intensity. Easeq software was used to generate heatmap plots, depicting windows from ±5 kb around the centre. Greyscale intensity is indicated below

the heat maps. MeDIP ChIP-seq results were categorized into three groups: peaks exclusive to the control sample (38,517 peaks), peaks shared between control and knockout samples (11,566 peaks), and peaks unique to the TCF12 knockout sample (16,841 peaks).

5.3 Discussion

Many studies have explored the consequences of the deletion or disruption of HEB (*TCF12*) in different biological contexts. For example, studies have revealed its importance in the development of the craniofacial system, in plasmacytoid dendritic cell development and immune responses (Cisse et al., 2008 , Wedel et al., 2020) , and in neuronal specification during the development of the spinal cord (Davis et al., 2020). There have also been studies conducted on the impact of HEB (*TCF12*) deletion and disruption on T-cell development and T-cell acute lymphoblastic leukaemia (T-ALL). Researchers have found that HEB plays an important role in promoting T-cell development, orchestrating the differentiation of various T-cell subsets, and maintaining haematopoietic stem cells (Braunstein and Anderson, 2012, Jones and Zhuang, 2007). There has been extensive research on the molecular mechanisms through which HEB influences T-cell development, as well as how its disruption can alter these processes, which offers opportunities for developing therapeutic targets for T-ALL and other related disorders (Miyazaki et al., 2017). Collectively, these studies provide valuable insights into the disease mechanisms and potential interventions of T-ALL, emphasizing the critical role of HEB in T-cell biology. To date, no extensive research has been conducted on HEB's interactions with other transcription factors and how the loss of HEB in T-cell acute lymphoblastic leukaemia (T-ALL) may affect these interactions.

Our initial objective was to delete HEB in T-ALL cell lines in order to examine the effects of HEB deletion in various cell contexts. We transfected the four T-ALL cell lines with Cas9 and gRNA expression constructs and successfully generated a targeted ARR cell line. Subsequent

experiments showed that this resulted in the absence of the HEB canonical isoform, while expression of the alternative isoform remained. For the other cell lines, we encountered significant challenges. Despite experimenting with the transfection settings, particularly using different voltages, the HSB2 cells did not demonstrate sufficient transfection efficiency and had a high post-transfection death rate, making further experimentation with this cell line impractical. On the other hand, we successfully transfected and isolated single-cell colonies of DU528 and CCRFCEM cells. The cell lines, however, presented challenges as they were not able to grow as single cell colonies after being FACS sorted. A cell line's susceptibility to transfection varies depending on its characteristics (Rose, 2003). Transfection efficiency may vary among cell lines due to membrane properties, cellular machinery, or other factors. Among these factors are factors such as genetic heterogeneity (Sakuma et al., 2018), or cas9 toxicity. Cas9 toxicity is where expression of cas9 protein in cells may be toxic, resulting in cell death (Ihry et al., 2018). It is necessary to conduct optimization experiments to identify the most effective transfection and delivery protocols for a specific cell line in order to overcome these challenges. The use of alternative genome editing delivery techniques, such as viral transduction and ribonucleoprotein delivery, may also prove helpful for achieving successful gene editing in difficult cell lines.

After successfully transfecting the ARR cell line with CRISPR cas9, we conducted a western blot experiment to confirm that the knockout was successful. In this study, the knockout was specifically targeted at the canonical isoform of HEB. It was interesting to observe that the alternative isoform showed a decrease in expression in the targeted cells when compared to the control. This suggests that there may be a regulatory relationship between these isoforms, or that the generated genomic alteration impacts on HEBalt mRNA expression. It is most likely the first, as our findings are consistent with previous research suggesting that the HEBalt isoform is controlled by a mechanism that involves both the canonical isoform and the Delta-

Notch pathway (Wang et al., 2006b). The researchers created retroviral constructs which expressed HEBalt or HEBcan and transduced adh.2C2 pro-T cells. After isolation, they performed qPCR analysis. The results of this analysis indicated that the levels of HEBalt mRNA increased in cells expressing the HEBalt construct, while the levels of HEBcan remained unchanged. As a result, cells that expressed HEBcan mRNA displayed increased levels of HEBalt and HEBcan mRNA, suggesting that HEBcan promotes HEBalt transcription.

Our Western blot analysis revealed that the knockout of the HEBcan affected the expression of several other proteins as well. The HEB knockout sample was found to have a reduction in GFI1 expression. This reduction was validated using antibodies from three different companies. These findings align with previous research where disruption of E2A and HEB indicated a role for these E-proteins in regulating *Gfi1* expression in murine models. Moreover, a microarray study revealed that *Gfi1* is directly regulated by E2A. Taking this evidence into account, it is clear that HEB, E2A, and GFI1 form an intricate regulatory network that contributes to thymocyte differentiation and development (Yucel et al., 2003, Schwartz et al., 2006a). HEB knockout reduced the expression of TAL1, LMO2, PHF6, and LYL1. These reduced expressions suggest that HEB or its downstream effectors act as positive regulators of these genes. For TAL1, LMO2 and LYL1, which all can contribute to the same transcription factor complex as HEB, autoregulation of their genes by this complex is known. To which extend HEBcan and HEBalt each can contribute to this process has not been investigated before.

Following HEB knockout in ARR cells, there were no significant changes in morphology and chromosome stability was preserved, suggesting that HEB deletion does not induce overt cellular abnormalities or chromosome instability. Our results were in line with a study of pluripotent embryonic stem cells (ESCs) in which researchers developed double E-

protein knockouts of E2A and HEB and found that colony morphology, as seen by immunostaining, remained the same between knockout and parental cells (Rao et al., 2020).

The analysis of DNaseI-seq experiments conducted in this study yielded valuable insights into the genomic characteristics and functional implications of DHS regions in ARR cells, both under control and *TCF12*[-/-] conditions. Using unique peaks in DHS control or *TCF12*[-/-], DNA motif analysis identified similar enriched motifs corresponding to transcription factor binding sites, such as RUNX, STAT, NFYB, and others, suggesting that these transcription factors play a role in the regulation of DHS regions and not affected with the loss of HEB. Furthermore, the analysis of genomic locations revealed that a significant proportion of DHS peaks are located near transcription start sites (TSSs) in both conditions (control and knockout), including promotors. Interestingly, there is a substantial decrease in the absolute number of genes with DHS peaks near their TSS between Cas9 control compared to the knockout. Specifically, from 10,355 genes in the control down to 2,798 in the *TCF12*[-/-]. This finding suggests that the HEB plays a key role in maintaining open and accessible chromatin at many genomic locations, which is essential for transcription. The functional enrichment analysis revealed distinct biological processes associated with DHS control and *TCF12*[-/-] samples, indicating that a loss of HEB affects the regulatory landscape, and enrichment of genes particularly involved in myeloid differentiation, immune response and mRNA processing.

We performed ChIP-seq analysis following knock out of *TCF12* in the ARR cell line for several gene regulatory proteins, which provided valuable genome-wide insight into transcription factor binding patterns. We overlapped the LMO2 peaks with the peaks from other data sets. The LMO2 data displays the largest set size, closely followed by TAL1, PRMT5, PHF6, LDB1, and *TCF12*. The largest overlap was between all five proteins, indicating that these factors together function as a complex. A plot-correlation analysis was conducted to assess the correlation between the ChIP-seq data sets. A notable finding was that HEB and LDB1

exhibited moderate to strong positive correlations under wild-type ARR conditions, indicating that the two genes share genomic binding regions. In contrast, knocking out *TCF12* in the ARR cell line reduced the similarity between HEB and PRMT5 binding patterns, indicating that the knockout of *TCF12* has an adverse effect on their interaction. This therefore shows that alterations to individual binding partners within the complex can have a direct impact on the recruitment of other proteins.

Using the ChIP-seq and DHS data, we ranked the ChIP-seq peaks based on read counts within a window of +/- 5kb to identify regions with heightened regulatory activity. The PRMT5 intensity was lower in *TCF12*$^{-/-}$ compared to control, which is consistent with reduced regulatory impact of PRMT5 in the absence of *TCF12*. However, LMO2 showed a distinct pattern, with higher intensity in *TCF12*$^{-/-}$. We hypothesised that the loss of HEB may result in LMO2 exhibiting increased DNA binding affinity, resulting in greater chromatin accessibility. This could be due to LMO2 compensating for the loss of *TCF12* by enhancing its interactions with DNA and other cofactors. The absence of HEB might also influence the binding dynamics of TAL1. In the normal scenario, TAL1 would form a complex with bHLH E proteins, including HEB, to promote gene expression. It may, however, be necessary for TAL1 to partner with alternative E proteins for DNA binding without HEB.

Based on the integration of MeDIP-seq data with ChIP-seq results and epigenomic analysis performed in this study, it has been revealed that DNA methylation, transcription factors, and genomic regions interact dynamically within ARR cells. The study of the methylation state of ARR cells following HEB knockout, along with the examination of transcription factor-mediated gene expression, has provided us with valuable insight. We identified a significant number of MeDIP-seq peaks in the control and *TCF12*$^{-/-}$ samples, with some of them shared between the two conditions. *De novo* motif analysis revealed enriched DNA sequence motifs associated with different transcription factors in each sample, which indicates distinct

transcriptional regulatory mechanisms. Notably, the control sample showed motifs corresponding to GFI1B, Ikzf3, and SOX9, while the *TCF12*[-/-] sample featured GFI1, TCF3 (E2A), prdm5, and Zbtb3 motifs, suggesting a shift in transcription factor binding upon the deletion of HEB. A subsequent analysis of MeDIP-seq peaks revealed a diverse distribution, with many peaks located at different distances from gene transcription start sites. Using functional enrichment analysis, different biological processes were identified associated with MeDIP peaks in control and *TCF12*[-/-] samples, which shed light on DNA methylation's potential role in regulating gene expression in these contexts. Additional insights were obtained from the generation of heatmaps and the analysis of peak intensities, particularly in relation to the relationship between DNA methylation and transcription factor binding. There is an association between DNA methylation and transcriptional regulation, especially in regions with high ChIP-seq scores, in regions where MeDIP regions colocalize with transcription factor peaks. The heatmap results showed that the highest peak intensities binding sites of transcription factors LDB1, PHF6, LMO2, PRMT5, and TAL1 associated with DNA methylation. It is intriguing to note that we observed a higher intensity of DNA methylation events in the absence of *TCF12*, indicating that *TCF12* might have a role to play in DNA methylation. Further, LMO2 was found to be involved in DNA methylation in the control sample, whereas this colocalization was less pronounced in the *TCF12*[-/-] sample, suggesting that *TCF12* may affect LMO2's role in DNA methylation regulation. Or that other factors that recognize methylated DNA may recruit LMO2 to these regions. It is possible that certain proteins contain domains that bind specifically to methylated DNA, and as a result might recruit LMO2 as part of a larger protein complex.

Chapter 6 General discussion

When early haematopoietic progenitors reach the thymus to differentiate into functional T-cells, they undergo many changes. During the initial differentiation steps, genes encoding transcription factors required for stem and progenitor cell identity need to be silenced, whereas genes encoding transcription factors required for T-cell identity and function need to be turned on (Yui and Rothenberg, 2014). Two transcription factors that work together in haematopoietic stem and progenitor cells are LMO2 and TAL1. They are expressed in immature DN1 and DN2 thymocytes, but the suppression of their genes is essential for normal thymocyte maturation at the DN3, DN4, and DP stages (Tan et al., 2019). It has been demonstrated that an aberrant continuation of LMO2 and TAL1 expression blocks the differentiation of T cells, resulting in the development of T-cell acute lymphoblastic leukaemia (T-ALL; Tremblay et al., 2010; Begley et al., 1989). T-ALL is a type of blood cancer caused by the uncontrolled proliferation of thymocytes (Pui et al., 2014). Interestingly, ectopic expression of LMO2 leads to T-ALL with long latency, whereas continued expression of TAL1 does not lead to T-ALL. Expression of both together reduces latency significantly (Aifantis et al., 2008). Both TAL1 and LMO2 are expressed in the majority of T-ALL cases, and their expression typically overlaps (Ferrando et al., 2002a) (Sincennes et al., 2016). This shows that LMO2 and TAL1 together play crucial roles in the onset of T-ALL and highlights how transcription factors rely on each other to perform their functions. Likewise, the abnormal expression of TAL1 in T-ALL results in the formation of heterodimers with E-proteins, including E-2A, HEB, and E-2. It disrupts the transcriptional programme of the E-proteins, which normally regulates the differentiation of T-cells. As a result of this sequestration, E-proteins are prevented from binding to their target genes, resulting in a blockage of T-cell differentiation (O'Neil et al., 2004).

Previous research conducted in our laboratory examined the role of TAL1 in T-ALL and its protein–protein interactions. TAL1 co-immunoprecipitation experiments identified the presence of HEB in T-ALL cell lines (ARR, DU528, HSB2, and CCRFCEM). The genomic

occupancy of the TAL1 and HEB overlapped significantly with that of the LMO2, LDB1, and GATA2. In the SIL/TAL T-ALL cell lines, the HEB peaks overlapped with TAL1, whereas in the ARR cell line, HEB binding did not overlap greatly with TAL1 and LMO2. In these cells, HEB binding correlated with regions of increased DNA methylation. Our research group has previously identified potential interactions between TAL1/HEB and regulators of T-cell development using mass spectrometry and motif analysis. In particular, GFI1 was identified in ChIP-seq experiments using ARR and DU528, with its motif being notably enriched in the HEB-only regions of DU528. Furthermore, they detected the known T-cell regulator IKAROS in HEB and TAL1 MS of the CCRFCEM cell line.

The primary objective of this thesis was to increase our understanding of HEB's function in T-ALL. One particular goal was to investigate the potential interaction between GFI1 and HEB, which we observed in T-ALL cell lines and may lead to novel insight into the molecular mechanisms responsible for the pathogenesis of T-ALL. Furthermore, we aimed to investigate the link between HEB binding and DNA methylation patterns, thus providing information on the epigenetic regulation of genomic regions associated with HEB.

As described in the introduction, GFI1 is known to play an important role in regulating the activity of the E-protein E2A, through ID1, in haematopoietic progenitors, highlighting an interplay between GFI1 levels and E2A (Fraszczak et al., 2016). Furthermore, a study using T-cells demonstrated that Gfi1 regulates E2A (specifically the isoform E47) in a manner that enforces a crucial developmental and proliferative checkpoint in maturing T-cells (Schwartz et al., 2006). Conversely, in murine mouse models, E-proteins were demonstrated to be responsible for regulating *Gfi1* expression, as the disruption of *Tcf3* (E2A) resulted in significantly downregulated *Gfi1* expression (Jones and Zhuang, 2007).

Considering that GFI1 plays dual roles as a suppressor and activator, and given its association with the E-protein E2A, it was important to further investigate the interaction between GFI1 and HEB observed in the mass spectrometry experiment. To further put this in the context of other transcription factors associated with the LMO2/TAL1 complex, we can gain valuable insight into the mechanisms governing T-ALL and the broader network of transcription factors that contribute to the development and oncogenesis of T-cells. Using human T-ALL cell lines, we studied the expression of HEB and GFI1. A total of four human T-ALL cells were selected, each blocked at a different stage of differentiation, which belonged to the early T-cell progenitor (ETP) and TAL/LMO subgroups of T-ALL. In these cell lines, all proteins contained in the LMO2-containing complex are expressed. Additionally, using these cell lines provides the opportunity to cross-reference the findings with results from prior studies conducted in the Hoogenkamp lab.

Using proteomic analysis, we substantiated the expression of HEB and GFI1 protein in T-ALL cell lines in Chapter 3. Using this approach, we were able to confirm the presence of these proteins, which play a significant role in T-ALL. Additionally, immunostaining analysis was used to identify specific proteins' subcellular localisation. It is noteworthy that LMO2 and HEB were observed to colocalise within the nucleus, providing insight into their potential interaction in cellular processes. Co-immunoprecipitation (Co-IP) assays successfully identified an interaction between HEB and GFI1. This interaction was particularly observed for the HEBalt isoform rather than for HEBcan. In the context of T-ALL, this observation is interesting, as it suggests a novel interaction that warrants further investigation to gain a greater understanding of how it impacts the disease process.

Mass spectrometry analysis of GFI1 in the ARR cell line revealed significant insights into the interaction network of the protein. We identified several proteins already known to have direct

interactions with GFI1, including important proteins such as RUNX and tumour protein 53 binding protein 1 (TP53BP1). The latter is consistent with previous studies emphasising the important role of GFI1 in cellular processes. GFI1 is involved in the regulation of DNA damage signalling and repair proteins in T cells (Vadnais et al., 2018). This study provides evidence that GFI1 regulates post-translational modifications of proteins involved in DNA repair through the interaction between GFI1 and the arginine methyltransferase PRMT1. An intriguing finding was the discovery that GFI1 serves as a cofactor for PRMT1, enhancing its activity (Vadnais et al., 2018). It is interesting to note that our findings are consistent with previous mass spectrometry analyses of HEB, which also identified PRMT1 (OMAIR, 2019). In addition to reinforcing the importance of PRMT1 in T-cell biology, these parallel findings suggest that HEB, GFI1, and PRMT1 might be connected in a potential multiprotein network. The intersection of these pathways may provide a fertile ground for future research into targeted therapies and diagnostics for disorders relating to T-cells.

As a result of our GFI1 co-IP and mass spectrometry analyses of the CCRFCEM cell line, PRMT5 was identified. PRMT5 is another arginine methyltransferase, and its involvement suggests that GFI1 in a more complex manner. GFI1 has been shown to interact with the AJUBA protein functions, independent of its SNAG domain, in a manner important for recruiting PRMT5 to the snail SNAG domain (Möröy et al., 2015). Like LMO2, AJUBA is a protein that contains a LIM domain and can interact with LDB1 (Ayyanathan et al., 2007, Witzel et al., 2012). Together, this suggests that GFI1 could play an important role in methylation processes, influencing gene expression and cellular differentiation. As a result of this post-translational modification, the interaction of GFI1 with other proteins may change, altering in turn its role in various processes. Both PRMT1 and PRMT5 are known for their roles in the arginine methylation of histones and other proteins; PRMT5 catalyses the

symmetric demethylation of arginine, whereas PRMT1 catalyses the asymmetric demethylation of arginine. This can affect gene expression, RNA processing, DNA repair, and signalling. As a result of the interaction of GFI1 with these methyltransferases, GFI1 may play a role in recruiting these proteins to the DNA. Consequently, GFI1 may regulate the activity of PRMT1 and PRMT5, or it may be a substrate of these enzymes. As a result, the cell landscape is altered.

Chapter 4 of our study investigated the genomic distribution of GFI1 and its specific binding sites. We used chromatin immunoprecipitation (ChIP) to identify GFI1 binding sites. We observed distinct binding patterns for GFI1 in different cell lines, with a notable difference between the ARR and SIL/TAL cell lines (DU528, HSB2, CCRFCEM). In particular, many more common peaks were identified in the SIL/TAL cell lines than overlap across binding sites was observed in ARR. We therefore analysed the ARR cell line and the SIL/TAL cell lines separately. Our *de novo* motif analysis yielded several significant insights: the top 500 highest-ranked peaks in ARR contained binding sites for key regulators of development processes and transcription factors involved in T-cell and B-cell development but lacked the GFI1 motif. This could mean that these top peaks recruited GFI1 particularly through protein–protein interactions, as was previously observed for TAL1 binding adjacent to GATA binding sites (Stanulović et al., 2017) When we subsequently investigated lower-scoring peaks, we did identify GFI1 binding motifs. This suggests that GFI1 is context dependent in its binding dynamics. Consequently, we further found that binding sites of IKZF1 were enriched in the lowest 500 peaks, along with Kruppel-like factor 6 (KLF6), a zinc finger protein that plays a role in tumour suppression. This could explain the presence of GFI1 motifs in the lowest 500 peaks in which it interacts with IKZF1 in contrast to higher-affinity GFI1 binding sets. A significant enrichment of genes related to CD8+ T-cell progression and the early stages of T-

cell development was found in the SIL/TAL cell lines, with prominent binding sites observed for IKZF1, TCF3, TCF7L2, LEF, SPI1, and SPIB.

ARR and SIL/TAL GFI1 ChIPseq peaks were distributed differently in relation to the transcription start site (TSS) in our GREAT analysis, with more SIL/TAL GFI1 peaks located farther away from the TSS. This could mean that in the SIL/TAL cell lines, GFI1 is more often associated with distal regulatory elements, such as enhancers. Our correlation analysis revealed a strong association between GFI1 and several transcription factors, including RUNX1, LDB1, GATA2, and HEB. However, different patterns were observed for LMO2 and TAL1 complexes. Additionally, heatmap analysis further indicated that GFI1 co-localised with transcription factors including HEB, LDB1, PHF6, and GATA2, but not with TAL1 or LMO2, indicating that these proteins may be involved in separate complexes. Finally, our findings support the hypothesis that LDB1, known for its role in forming multi-transcriptional factor complexes and binding to LIM domain proteins, might form a distinct complex in ARR involving HEB, RUNX1, GFI1, and possibly LMO4 instead of LMO2. This new understanding of GFI1's genomic distribution illuminates its role in gene regulation. It could influence future research on cellular development and gene expression regulation.

Chapter 5 investigates the role of HEB in T-cell development and its implications for T-ALL. Our initial goal was to delete all HEB expression in four T-ALL cell lines by using CRISPR/Cas9. We were successful in knocking out the canonical isoform of HEB only in the ARR cell line. This was confirmed by western blot analysis, which also revealed changes in the expression of several proteins, including a reduction in GFI1. It is interesting to note that while the HEB knockout was not observed to significantly affect chromosome stability or cell morphology in the ARR cells, it did have a significant impact on the accessibility of chromatin and the regulatory landscape. Using DNaseI hypersensitive site mapping, we found differences

in biological processes between the control and *TCF12*[-/-] samples, suggesting that HEB plays a role in myeloid differentiation and immune response. Furthermore, we carried out a correlation analysis of ChIP-seq datasets, revealing that HEB and LDB1 exhibit a moderate-to-strong correlation under wild-type conditions, whereas the correlation was reduced when *TCF12* was knocked out. Therefore, the absence of HEB affected the interaction between these proteins. A significant decrease in PRMT5 intensity and a distinct pattern of LMO2 expression were also observed in the *TCF12*[-/-] sample, indicating that the loss of HEB contributed to increased DNA-binding affinity in LMO2. In this chapter, MeDIP-seq data and ChIP-seq results were integrated with epigenomic analysis to investigate the dynamic interaction between DNA methylation, transcription factors, and genomic regions in ARR cells. In the control and *TCF12*[-/-] samples, various DNA sequence motifs associated with different transcription factors were overrepresented, indicating changes in transcriptional regulation mechanisms following *TCF12* deletion. Notably, DNA methylation increased in the absence of HEB. This suggests not that the observed binding of HEB at methylCpG-enriched areas is likely to cause DNA methylation but rather that HEB is recruited to sites with elevated DNA methylation. The findings of this chapter provide important insights into the molecular mechanisms involved in the interactions between HEB and other transcription factors and how the loss of HEB may affect these interactions in T-ALL.

Limitations

There were several limitations to the work conducted in this thesis. We used mass spectrometry because it provided an unbiased tool with which to confirm or identify new partners interacting with the pulled-down protein. The mass spectrometer we used for this was inexpensive to operate but not very sensitive. This meant that we could get many interesting results but also that the absence of specific proteins in the results did not necessarily imply the absence of interaction. For example, based on prior research in the Hoogenkamp group, we know that

LMO2 is normally not detectable, even if western blot confirms that it is in the sample. Likewise, we did not detect GFI1.

A considerable difficulty was encountered in our attempt to target the deletion of the HEB using CRISPR-Cas9 and homology-directed repair (HDR). We designed multiple primers to be able to sequence the introduced genomic changes. Unfortunately, we did not succeed in this and consequently were not able to determine whether particular cellular clones were heterozygous for the CRISPR-Cas9-induced changes. A significant limitation of our study was that only a single HEB knockout (KO) clone was generated among all the T-ALL cell lines. This limitation makes it more difficult to draw robust conclusions from our data. To account for variations between clonally generated cell lines, multiple samples would have been more suitable. We cannot exclude the possibility that off-target effects or clonal anomalies might have impacted our results. Additionally, we were unable to perform GFI1 ChIP assays of the HEB knockout sample to determine whether the interaction would persist following the deletion of canonical HEB.

Additionally, in our *TCF12*$^{-/-}$ clone, only full-length HEBcan was deleted, whereas HEBalt continued to be expressed. Although this provided an interesting opportunity to determine the effect of HEBcan in isolation, we would likely have had more striking results if we had been able to knock out both forms. It is likely that the presence of HEBalt compensated partially for the loss of the canonical form, thereby reducing the impact of the deletion on the cellular processes. This is an important issue, particularly in the context of diseases such as T-ALL, in which the expression of different isoforms of a gene might have varying impacts on disease progression and cellular behaviour.

Our laboratory work was significantly impacted by the extreme challenges posed by the COVID-19 pandemic. Since the global health crisis started, we have had difficulty obtaining a

number of consumables for the lab. In particular, we were planning to perform proximity ligation assays (PLAs) to further confirm the interaction between HEB and GFI. The kit has been on continuous back order. Moreover, our access to the laboratory was further restricted by intermittent lockdowns. In particular, ongoing experiments in which we were attempting to generate further *TCF12* KO cell lines had to be prematurely terminated during a period of self-isolation following exposure to the virus. This resulted in the loss of valuable time and resources.

Future work

There are a number of experiments that should be performed to further develop this work. To further confirm the interaction between HEB and GFI1, we plan to use PLA techniques. With PLA, we will be able to obtain a more robust and visual indication of the physical proximity and interaction between these two proteins at the molecular level.

To gain a better understanding of the transcriptional impact of HEB KO and identify direct target genes, we would like to conduct RNA-seq experiments on HEB KO samples. In addition to providing insights into gene expression changes and potential downstream effects of HEB loss, these experiments will also enable us to integrate expression data with transcription factor binding and epigenetic implications, such as chromatin accessibility and DNA methylation status. Additionally, it would be interesting to determine whether the interaction between HEBcan and GFI1 persists or changes when the HEBalt is deleted by selectively targeting HEBalt, thereby shedding light on the isoform-specific functions of HEB in the context of protein–protein interactions. Furthermore, due to the difficulty in deleting HEB from SIL/TAL cell lines using CRISPR/cas9 technology, using different knockout approaches instead, such as viral transduction and ribonucleoprotein delivery, would allow us to gain a deeper understanding of the role of HEB across different cellular contexts at different stages of T-ALL.

Together, these approaches will contribute to a more comprehensive understanding of HEB–GFI1 interactions and their biological significance, which could ultimately lead to more targeted and effective therapeutic interventions in the future.

In conclusion, this study sheds light on the intricate interaction between HEB and GFI1 in human T-ALL cell lines and highlighting the significance of the HEB alternative isoform. Our proteomic and genome-wide studies revealed distinct binding complexes of HEB and GFI1 in different cell lines, suggesting context-specific regulatory mechanisms. CRISPR-Cas9 gene editing targeting TCF12 revealed a regulatory interaction between HEB isoforms, impacting GFI1 expression without causing significant cellular abnormalities. In addition, the increase in LMO2 association with DNA suggests a compensatory mechanism that enhances chromatin accessibility in response to HEB loss. These findings deepen our understanding of T-ALL at a molecular level, which may lead to the identification of future therapeutic targets.

REFERENCES

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.

ALTMEYER, M., TOLEDO, L., GUDJONSSON, T., GRØFTE, M., RASK, M. B., LUKAS, C., AKIMOV, V., BLAGOEV, B., BARTEK, J. & LUKAS, J. 2013. The chromatin scaffold protein SAFB1 renders chromatin permissive for DNA damage signaling. *Mol Cell,* 52**,** 206-20.

ANDERSON, M. K., HERNANDEZ-HOYOS, G., DIONNE, C. J., ARIAS, A. M., CHEN, D. & ROTHENBERG, E. V. 2002a. Definition of regulatory network elements for T cell development by perturbation analysis with PU.1 and GATA-3. *Dev Biol,* 246**,** 103-21.

ANDERSON, M. K., WEISS, A. H., HERNANDEZ-HOYOS, G., DIONNE, C. J. & ROTHENBERG, E. V. 2002b. Constitutive expression of PU.1 in fetal hematopoietic progenitors blocks T cell development at the pro-T cell stage. *Immunity,* 16**,** 285-96.

ANGUITA, E., CANDEL, F. J., CHAPARRO, A. & ROLDÁN-ETCHEVERRY, J. J. 2017. Transcription Factor GFI1B in Health and Disease. *Front Oncol,* 7**,** 54.

ATTAF, M., LEGUT, M., COLE, D. K. & SEWELL, A. K. 2015. The T cell antigen receptor: the Swiss army knife of the immune system. *Clin Exp Immunol,* 181**,** 1-18.

AWONG, G., HERER, E., SURH, C. D., DICK, J. E., LA MOTTE-MOHS, R. N. & ZÚÑIGA-PFLÜCKER, J. C. 2009. Characterization in vitro and engraftment potential in vivo of human progenitor T cells generated from hematopoietic stem cells. *Blood,* 114**,** 972-982.

AYYANATHAN, K., PENG, H., HOU, Z., FREDERICKS, W. J., GOYAL, R. K., LANGER, E. M., LONGMORE, G. D. & RAUSCHER, F. J., III 2007. The Ajuba LIM Domain Protein Is a Corepressor for SNAG Domain–Mediated Repression and Participates in Nucleocytoplasmic Shuttling. *Cancer Research,* 67**,** 9097-9106.

BAIN, B. J. 2017. Structure and function of red and white blood cells. *Medicine,* 45**,** 187-193.

BAIN, G., QUONG, M. W., SOLOFF, R. S., HEDRICK, S. M. & MURRE, C. 1999. Thymocyte maturation is regulated by the activity of the helix-loop-helix protein, E47. *J Exp Med,* 190**,** 1605-16.

BARNDT, R., DAI, M.-F. & ZHUANG, Y. 1999. A Novel Role for HEB Downstream or Parallel to the Pre-TCR Signaling Pathway During αβ Thymopoiesis. *The Journal of Immunology,* 163**,** 3331-3343.

BARNDT, R. J., DAI, M. & ZHUANG, Y. 2000. Functions of E2A-HEB heterodimers in T-cell development revealed by a dominant negative mutation of HEB. *Mol Cell Biol,* 20**,** 6677-85.

BARRANGOU, R. 2013. CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip Rev RNA,* 4**,** 267-78.

BASSING, C. H., SWAT, W. & ALT, F. W. 2002. The mechanism and regulation of chromosomal V(D)J recombination. *Cell,* 109 Suppl**,** S45-55.

BAYÓN-CALDERÓN, F., TORIBIO, M. L. & GONZÁLEZ-GARCÍA, S. 2020. Facts and Challenges in Immunotherapy for T-Cell Acute Lymphoblastic Leukemia. *International Journal of Molecular Sciences,* 21**,** 7685.

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.

BEGLEY, C. G., APLAN, P. D., DAVEY, M. P., NAKAHARA, K., TCHORZ, K., KURTZBERG, J., HERSHFIELD, M. S., HAYNES, B. F., COHEN, D. I., WALDMANN, T. A. & ET AL. 1989. Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region

and results in a previously unreported fusion transcript. *Proc Natl Acad Sci U S A,* 86**,** 2031-5.

BELLE, I. & ZHUANG, Y. 2014. E proteins in lymphocyte development and lymphoid diseases. *Current topics in developmental biology,* 110**,** 153-187.

BELVER, L. & FERRANDO, A. 2016. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat Rev Cancer,* 16**,** 494-507.

BENEZRA, R., DAVIS, R. L., LOCKSHON, D., TURNER, D. L. & WEINTRAUB, H. 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell,* 61**,** 49-59.

BLACK, D. L. 2003. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry,* 72**,** 291-336.

BLANKENBERG, D., KUSTER, G. V., CORAOR, N., ANANDA, G., LAZARUS, R., MANGAN, M., NEKRUTENKO, A. & TAYLOR, J. 2010. Galaxy: A Web-Based Genome Analysis Tool for Experimentalists. *Current Protocols in Molecular Biology,* 89**,** 19.10.1-19.10.21.

BOLOTIN, A., QUINQUIS, B., SOROKIN, A. & EHRLICH, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading),* 151**,** 2551-2561.

BOTEZATU, L., MICHEL, L. C., HELNESS, A., VADNAIS, C., MAKISHIMA, H., HÖNES, J. M., ROBERT, F., VASSEN, L., THIVAKARAN, A., AL-MATARY, Y., LAMS, R. F., SCHÜTTE, J., GIEBEL, B., GÖRGENS, A., HEUSER, M., MEDYOUF, H., MACIEJEWSKI, J., DÜHRSEN, U., MÖRÖY, T. & KHANDANPOUR, C. 2016. Epigenetic therapy as a novel approach for GFI136N-associated murine/human AML. *Experimental Hematology,* 44**,** 713-726.e14.

BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell,* 132**,** 311-22.

BRAUNSTEIN, M. & ANDERSON, M. K. 2010. Developmental progression of fetal HEB(-/-) precursors to the pre-T-cell stage is restored by HEBAlt. *Eur J Immunol,* 40**,** 3173-82.

BRAUNSTEIN, M. & ANDERSON, M. K. 2011. HEB-deficient T-cell precursors lose T-cell potential and adopt an alternative pathway of differentiation. *Mol Cell Biol,* 31**,** 971-82.

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.

BRAUNSTEIN, M., RAJKUMAR, P., CLAUS, C. L., VACCARELLI, G., MOORE, A. J., WANG, D. & ANDERSON, M. K. 2010. HEBAlt enhances the T-cell potential of fetal myeloid-biased precursors. *Int Immunol,* 22**,** 963-72.

BROWN, G., SANCHEZ, L. AND SANCHEZ-GARCIA, I. 2020. Lineage Decision-Making within Normal Haematopoietic and Leukemic Stem Cells. *International Journal of Molecular Sciences* 21(6).

CALKHOVEN, C. F., MULLER, C., MARTIN, R., KROSL, G., PIETSCH, H., HOANG, T. & LEUTZ, A. 2003. Translational control of SCL-isoform expression in hematopoietic lineage choice. *Genes Dev,* 17**,** 959-64.

CANU, G. & RUHRBERG, C. 2021. First blood: the endothelial origins of hematopoietic progenitors. *Angiogenesis,* 24**,** 199-211.

CAPONE, M., HOCKETT, R. D., JR. & ZLOTNIK, A. 1998. Kinetics of T cell receptor beta, gamma, and delta rearrangements during adult thymic development: T cell receptor rearrangements are present in CD44(+)CD25(+) Pro-T thymocytes. *Proc Natl Acad Sci U S A,* 95**,** 12522-7.

CARPENTER, A. C. & BOSSELUT, R. 2010. Decision checkpoints in the thymus. *Nature immunology,* 11**,** 666-673.

CATHOMEN, T. & JOUNG, J. K. 2008. Zinc-finger nucleases: the next generation emerges. *Mol Ther,* 16**,** 1200-7.

CAVÉ, H., SUCIU, S., PREUDHOMME, C., POPPE, B., ROBERT, A., UYTTEBROECK, A., MALET, M., BOUTARD, P., BENOIT, Y., MAUVIEUX, L., LUTZ, P., MÉCHINAUD, 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. 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.

CHAN, L. N., CHEN, Z., BRAAS, D., LEE, J. W., XIAO, G., GENG, H., COSGUN, K. N., HURTZ, C., SHOJAEE, S., CAZZANIGA, V., SCHJERVEN, H., ERNST, T., HOCHHAUS, A., KORNBLAU, S. M., KONOPLEVA, M., PUFALL, M. A., CAZZANIGA, G., LIU, G. J., MILNE, T. A., KOEFFLER, H. P., ROSS, T. S., SÁNCHEZ-GARCÍA, I., BORKHARDT, A., YAMAMOTO, K. R., DICKINS, R. A., GRAEBER, T. G. & MÜSCHEN, M. 2017. Metabolic gatekeeper function of B-lymphoid transcription factors. *Nature,* 542**,** 479-483.

CHEN, C., LIU, Y., RAPPAPORT, A. R., KITZING, T., SCHULTZ, N., ZHAO, Z., SHROFF, A. S., DICKINS, R. A., VAKOC, C. R., BRADNER, J. E., STOCK, W., LEBEAU, M. M., SHANNON, K. M., KOGAN, S., ZUBER, J. & LOWE, S. W. 2014. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell,* 25**,** 652-65.

CHEN, F., DING, X., FENG, Y., SEEBECK, T., JIANG, Y. & DAVIS, G. D. 2017. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nature Communications,* 8**,** 14958.

CHOI, P. S. & MEYERSON, M. 2014. Targeted genomic rearrangements using CRISPR/Cas technology. *Nature Communications,* 5**,** 3728.

CISSE, B., CATON, M. L., LEHNER, M., MAEDA, T., SCHEU, S., LOCKSLEY, R., HOLMBERG, D., ZWEIER, C., DEN HOLLANDER, N. S., KANT, S. G., HOLTER, W., RAUCH, A., ZHUANG, Y. & REIZIS, B. 2008. Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. *Cell,* 135**,** 37-48.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

COUSTAN-SMITH, E., MULLIGHAN, C. G., ONCIU, M., BEHM, F. G., RAIMONDI, S. C., PEI, D., CHENG, C., SU, X., RUBNITZ, J. E., BASSO, G., BIONDI, A., PUI, C. H., DOWNING, J. R. & CAMPANA, D. 2009. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol,* 10**,** 147-56.

CRADICK, T. J., QIU, P., LEE, C. M., FINE, E. J. & BAO, G. 2014. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Mol Ther Nucleic Acids,* 3**,** e214.

CEREDIG R, Rauch M, Balciunaite G, Rolink AG. Increasing Flt3L availability alters composition of a novel bone marrow lymphoid progenitor compartment. Blood. 2006 Aug 15;108(4):1216-22. doi: 10.1182/blood-2005-10-006643. Epub 2006 May 4. PMID: 16675711.

CEREDIG, R., ROLINK, A. & BROWN, G. Models of haematopoiesis: seeing the wood for the trees. *Nat Rev Immunol* **9**, 293–300 (2009). https://doi.org/10.1038/nri2525

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.

DAHL, R., IYER, S. R., OWENS, K. S., CUYLEAR, D. D. & SIMON, M. C. 2007. The transcriptional repressor GFI-1 antagonizes PU.1 activity through protein-protein interaction. *J Biol Chem,* 282**,** 6473-83.

DAVIS, E. E., BALASUBRAMANIAN, R., KUPCHINSKY, Z. A., KEEFE, D. L., PLUMMER, L., KHAN, K., MECZEKALSKI, B., HEATH, K. E., LOPEZ-GONZALEZ, V., BALLESTA-MARTINEZ, M. J., MARGABANTHU, G., PRICE, S., GREENING, J., BRAUNER, R., VALENZUELA, I., CUSCO, I., FERNANDEZ-ALVAREZ, P., WIERMAN, M. E., LI, T., LAGE, K., BARROSO, P. S., CHAN, Y. M., CROWLEY, W. F. & KATSANIS, N. 2020. TCF12 haploinsufficiency causes autosomal dominant Kallmann syndrome and reveals network-level interactions between causal loci. *Hum Mol Genet,* 29**,** 2435-2450.

DELABESSE, E., BERNARD, M., MEYER, V., SMIT, L., PULFORD, K., CAYUELA, J. M., RITZ, J., BOURQUELOT, P., STROMINGER, J. L., VALENSI, F. & MACINTYRE, E. A. 1998. TAL1 expression does not occur in the majority of T-ALL blasts. *Br J Haematol,* 102**,** 449-57.

DOUDNA, J. A. & CHARPENTIER, E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science,* 346**,** 1258096.

DUAN, Z. & HORWITZ, M. 2003. Targets of the transcriptional repressor oncoprotein Gfi-1. *Proceedings of the National Academy of Sciences,* 100**,** 5932-5937.

DUAN, Z. & HORWITZ, M. 2005. Gfi-1 takes center stage in hematopoietic stem cells. *Trends in Molecular Medicine,* 11**,** 49-52.

DUAN, Z., ZAREBSKI, A., MONTOYA-DURANGO, D., GRIMES, H. L. & HORWITZ, M. 2005. Gfi1 coordinates epigenetic repression of p21Cip/WAF1 by recruitment of histone lysine methyltransferase G9a and histone deacetylase 1. *Mol Cell Biol,* 25**,** 10338-51.

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.

EIZENBERG-MAGAR I, RIMER J, ZARETSKY I, LARA-ASTIASO D, REICH-ZELIGER S, FRIEDMAN N. Diverse continuum of CD4[+] T-cell states is determined by hierarchical additive integration of cytokine signals. Proc Natl Acad Sci U S A. 2017 Aug 1;114(31):E6447-E6456. doi:10.1073/pnas.1615590114. Epub 2017 Jul 17. PMID: 28716917; PMCID: PMC5547583.

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.

FAMILI, F., WIEKMEIJER, A. S. & STAAL, F. J. 2017. The development of T cells from stem cells in mice and humans. *Future Sci OA,* 3**,** Fso186.

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. 2002a. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell,* 1**,** 75-87.

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. 2002b. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell,* 1**,** 75-87.

FRASZCZAK, J., HELNESS, A., CHEN, R., VADNAIS, C., ROBERT, F., KHANDANPOUR, C. & MÖRÖY, T. 2016. Threshold Levels of Gfi1 Maintain E2A Activity for B Cell Commitment via Repression of Id1. *PLoS One,* 11**,** e0160344.

FU, Q., LI, H., MOORJANI, P., JAY, F., SLEPCHENKO, S. M., BONDAREV, A. A., JOHNSON, P. L. F., AXIMU-PETRI, A., PRÜFER, K., DE FILIPPO, C., MEYER, M., ZWYNS, N., SALAZAR-GARCÍA, D. C., KUZMIN, Y. V., KEATES, S. G., KOSINTSEV, P. A., RAZHEV, D. I., RICHARDS, M. P., PERISTOV, N. V., LACHMANN, M., DOUKA, K., HIGHAM, T. F. G., SLATKIN, M., HUBLIN, J.-J., REICH, D., KELSO, J., VIOLA, T. B. & PÄÄBO, S. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature,* 514**,** 445-449.

FU, Y., FODEN, J. A., KHAYTER, C., MAEDER, M. L., REYON, D., JOUNG, J. K. & SANDER, J. D. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol,* 31**,** 822-6.

GAJ, T., SIRK, S. J., SHUI, S. L. & LIU, J. 2016. Genome-Editing Technologies: Principles and Applications. *Cold Spring Harb Perspect Biol,* 8.

GEIMER LE LAY, A.-S., ORAVECZ, A., MASTIO, J., JUNG, C., MARCHAL, P., EBEL, C., DEMBÉLÉ, D., JOST, B., LE GRAS, S., THIBAULT, C., BORGGREFE, T., KASTNER, P. & CHAN, S. 2014. The tumor suppressor Ikaros shapes the repertoire of notch target genes in T cells. *Science signaling* [Online], 7. Available: http://europepmc.org/abstract/MED/24643801

https://doi.org/10.1126/scisignal.2004545 [Accessed 2014/03//].

GERMAIN, R. N. 2002. T-cell development and the CD4-CD8 lineage decision. *Nat Rev Immunol,* 2**,** 309-22.

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.

GILKS, C. B., BEAR, S. E., GRIMES, H. L. & TSICHLIS, P. N. 1993. Progression of interleukin-2 (IL-2)-dependent rat T cell lymphoma lines to IL-2-independent growth following activation of a gene (Gfi-1) encoding a novel zinc finger protein. *Molecular and Cellular Biology,* 13**,** 1759-1768.

GIRARDI, T., VICENTE, C., COOLS, J. & DE KEERSMAECKER, K. 2017. The genetics and molecular biology of T-ALL. *Blood,* 129**,** 1113-1123.

GOECKS, J., NEKRUTENKO, A., TAYLOR, J. & THE GALAXY, T. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology,* 11**,** R86.

GOODE, D. K., OBIER, N., VIJAYABASKAR, M. S., LIE, A. L. M., LILLY, A. J., HANNAH, R., LICHTINGER, M., BATTA, K., FLORKOWSKA, M., PATEL, R., CHALLINOR, M., WALLACE, K., GILMOUR, J., ASSI, S. A., CAUCHY, P., HOOGENKAMP, M., WESTHEAD, D. R., LACAUD, G., KOUSKOFF, V., GÖTTGENS, B. & BONIFER, C. 2016. Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Dev Cell,* 36**,** 572-87.

GREENOUGH, T. C., STRAUBHAAR, J. R., KAMGA, L., WEISS, E. R., BRODY, R. M., MCMANUS, M. M., LAMBRECHT, L. K., SOMASUNDARAN, M. & LUZURIAGA,

K. F. 2015. A Gene Expression Signature That Correlates with CD8+ T Cell Expansion in Acute EBV Infection. *J Immunol,* 195**,** 4185-97.

GUPTA, R. M. & MUSUNURU, K. 2014. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest,* 124**,** 4154-61.

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.

HAMASAKI, N. & YAMAMOTO, M. 2000. Red blood cell function and blood storage. *Vox Sang,* 79**,** 191-7.

HAYDAY, A. C. & PENNINGTON, D. J. 2007. Key factors in the organized chaos of early T cell development. *Nature Immunology,* 8**,** 137-144.

HEEMSKERK, M. H., BLOM, B., NOLAN, G., STEGMANN, A. P., BAKKER, A. Q., WEIJER, K., RES, P. C. & SPITS, H. 1997. Inhibition of T cell and promotion of natural killer cell development by the dominant negative helix loop helix factor Id3. *J Exp Med,* 186**,** 1597-602.

HEYD, F., TEN DAM, G. & MÖRÖY, T. 2006. Auxiliary splice factor U2AF26 and transcription factor Gfi1 cooperate directly in regulating CD45 alternative splicing. *Nat Immunol,* 7**,** 859-67.

HEYWORTH, C., PEARSON, S., MAY, G. & ENVER, T. 2002. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *Embo j,* 21**,** 3770-81.

HOCK, H., HAMBLEN, M. J., ROOKE, H. M., SCHINDLER, J. W., SALEQUE, S., FUJIWARA, Y. & ORKIN, S. H. 2004. Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. *Nature,* 431**,** 1002-7.

HOMMINGA, I., PIETERS, R., LANGERAK, ANTON W., DE ROOI, JOHAN J., STUBBS, A., VERSTEGEN, M., VUERHARD, M., BUIJS-GLADDINES, J., KOOI, C., KLOUS, P., VAN VLIERBERGHE, P., FERRANDO, ADOLFO A., CAYUELA, JEAN M., VERHAAF, B., BEVERLOO, H. B., HORSTMANN, M., DE HAAS, V., WIEKMEIJER, A.-S., PIKE-OVERZET, K., STAAL, FRANK J. T., DE LAAT, W., SOULIER, J., SIGAUX, F. & MEIJERINK, JULES P. P. 2011. Integrated Transcript and Genome Analyses Reveal NKX2-1 and MEF2C as Potential Oncogenes in T Cell Acute Lymphoblastic Leukemia. *Cancer Cell,* 19**,** 484-497.

HRYHOROWICZ, M., LIPIŃSKI, D., ZEYLAND, J. & SŁOMSKI, R. 2017. CRISPR/Cas9 Immune System as a Tool for Genome Engineering. *Arch Immunol Ther Exp (Warsz),* 65**,** 233-240.

HSU, H. L., HUANG, L., TSAN, J. T., FUNK, W., WRIGHT, W. E., HU, J. S., KINGSTON, R. E. & BAER, R. 1994. Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. *Mol Cell Biol,* 14**,** 1256-65.

HU, J. S., OLSON, E. N. & KINGSTON, R. E. 1992. HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. *Mol Cell Biol,* 12**,** 1031-42.

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols,* 4**,** 44-57.

HWANG, J. W., CHO, Y., BAE, G.-U., KIM, S.-N. & KIM, Y. K. 2021. Protein arginine methyltransferases: promising targets for cancer therapy. *Experimental & Molecular Medicine,* 53**,** 788-808.

IBAGY, A., SILVA, D. B., SEIBEN, J., WINNESHOFFER, A. P. F. F., COSTA, T. E. J. B., DACOREGIO, J. S., COSTA, I. & FARACO, D. 2013. Acute Lymphoblastic Leukemia in Infants: 20 years of Experience. *Jornal de Pediatria,* 89**,** 64-69.

IHRY, R. J., WORRINGER, K. A., SALICK, M. R., FRIAS, E., HO, D., THERIAULT, K., KOMMINENI, S., CHEN, J., SONDEY, M., YE, C., RANDHAWA, R., KULKARNI, T., YANG, Z., MCALLISTER, G., RUSS, C., REECE-HOYES, J., FORRESTER, W., HOFFMAN, G. R., DOLMETSCH, R. & KAYKAS, A. 2018. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat Med,* 24**,** 939-946.

ISHINO, Y., SHINAGAWA, H., MAKINO, K., AMEMURA, M. & NAKATA, A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol,* 169**,** 5429-33.

JAGANNATHAN-BOGDAN, M. & ZON, L. I. 2013. Hematopoiesis. *Development,* 140**,** 2463-7.

JANSEN, R., EMBDEN, J. D., GAASTRA, W. & SCHOULS, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol,* 43**,** 1565-75.

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.

JIN, S., TRAN, N.-T., SU, H., HUANG, S., ZHAO, X. & LIU, Y. 2016. The TAL1 Short Isoform Generated By PRMT1-Mediated Alternative RNA Splicing Promotes Erythroid Differentiation. *Blood,* 128**,** 1249.

JINEK, M., EAST, A., CHENG, A., LIN, S., MA, E. & DOUDNA, J. 2013. RNA-programmed genome editing in human cells. *Elife,* 2**,** e00471.

JOJIC, V., SHAY, T., SYLVIA, K., ZUK, O., SUN, X., KANG, J., REGEV, A., KOLLER, D., 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, M. E. & ZHUANG, Y. 2007. Acquisition of a functional T cell receptor during T lymphocyte development is enforced by HEB and E2A transcription factors. *Immunity,* 27**,** 860-70.

JONES, S. 2004. An overview of the basic helix-loop-helix proteins. *Genome Biol,* 5**,** 226.

JURK, K. & KEHREL, B. E. 2005. Platelets: physiology and biochemistry. *Semin Thromb Hemost,* 31**,** 381-92.

KALLIANPUR, A. R., JORDAN, J. E. & BRANDT, S. J. 1994. The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis. *Blood,* 83**,** 1200-8.

KARSUNKY, H., MENDE, I., SCHMIDT, T. & MÖRÖY, T. 2002a. High levels of the onco-protein Gfi-1 accelerate T-cell proliferation and inhibit activation induced T-cell death in Jurkat T-cells. *Oncogene,* 21**,** 1571-1579.

KARSUNKY, H., ZENG, H., SCHMIDT, T., ZEVNIK, B., KLUGE, R., SCHMID, K. W., DÜHRSEN, U. & MÖRÖY, T. 2002b. Inflammatory reactions and severe neutropenia in mice lacking the transcriptional repressor Gfi1. *Nat Genet,* 30**,** 295-300.

KAZANJIAN, A., WALLIS, D., AU, N., NIGAM, R., VENKEN, K. J. T., CAGLE, P. T., DICKEY, B. F., BELLEN, H. J., GILKS, C. B. & GRIMES, H. L. 2004. Growth Factor

Independence-1 Is Expressed in Primary Human Neuroendocrine Lung Carcinomas and Mediates the Differentiation of Murine Pulmonary Neuroendocrine Cells. *Cancer Research,* 64**,** 6874-6882.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

KERENYI, M. A. & ORKIN, S. H. 2010. Networking erythropoiesis. *J Exp Med,* 207**,** 2537-41.

KHANDANPOUR, C. & MÖRÖY, T. 2013. Growth factor independence 1 (Gfi1) as a regulator of p53 activity and a new therapeutical target for ALL. *Oncotarget,* 4**,** 374-5.

KIEL, M. J., YILMAZ, Ö. 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**,** 1109-1121.

KIM, D., KIM, S., KIM, S., PARK, J. & KIM, J. S. 2016. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Res,* 26**,** 406-15.

KLEINSTIVER, B. P., PATTANAYAK, V., PREW, M. S., TSAI, S. Q., NGUYEN, N. T., ZHENG, Z. & JOUNG, J. K. 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature,* 529**,** 490-5.

KOCAK, D. D., JOSEPHS, E. A., BHANDARKAR, V., ADKAR, S. S., KWON, J. B. & GERSBACH, C. A. 2019. Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nature Biotechnology,* 37**,** 657-666.

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. *Journal of Experimental Medicine,* 205**,** 2515-2523.

KOCH, U. & RADTKE, F. 2011. Mechanisms of T cell development and transformation. *Annu Rev Cell Dev Biol,* 27**,** 539-62.

KOO, T., LEE, J. & KIM, J. S. 2015. Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9. *Mol Cells,* 38**,** 475-81.

KOWALCZYK, J. & SANDBERG, A. A. 1983. A possible subgroup of ALL with 9p. *Cancer Genet Cytogenet,* 9**,** 383-5.

KRANGEL, M. S. 2009. Mechanics of T cell receptor gene rearrangement. *Curr Opin Immunol,* 21**,** 133-9.

KUROKAWA, M. 2006. AML1/Runx1 as a versatile regulator of hematopoiesis: regulation of its function and a role in adult hematopoiesis. *Int J Hematol,* 84**,** 136-42.

LABRECHE, K., SIMEONOVA, I., KAMOUN, A., GLEIZE, V., CHUBB, D., LETOUZÉ, E., RIAZALHOSSEINI, Y., DOBBINS, S. E., ELAROUCI, N., DUCRAY, F., DE REYNIÈS, A., ZELENIKA, D., WARDELL, C. P., FRAMPTON, M., SAULNIER, O., PASTINEN, T., HALLOUT, S., FIGARELLA-BRANGER, D., DEHAIS, C., IDBAIH, A., MOKHTARI, K., DELATTRE, J.-Y., HUILLARD, E., MARK LATHROP, G., SANSON, M., HOULSTON, R. S., ADAM, C., ANDRAUD, M., AUBRIOT-LORTON, M.-H., BAUCHET, L., BEAUCHESNE, P., BLECHET, C., CAMPONE, M., CARPENTIER, A., CARPENTIER, C., CARPIUC, I., CHENARD, M.-P., CHIFOREANU, D., CHINOT, O., COHEN-MOYAL, E., COLIN, P., DAM-HIEU, P., DESENCLOS, C., DESSE, N., DHERMAIN, F., DIEBOLD, M.-D., EIMER, S., FAILLOT, T., FESNEAU, M., FONTAINE, D., GAILLARD, S., GAUCHOTTE, G., GAULTIER, C., GHIRINGHELLI, F., GODARD, J., MARCEL GUEYE, E., SEBASTIEN GUILLAMO, J., HAMDI-ELOUADHANI, S., HONNORAT, J., LOUIS

KEMENY, J., KHALLIL, T., JOUVET, A., LABROUSSE, F., LANGLOIS, O., LAQUERRIERE, A., LECHAPT-ZALCMAN, E., LE GUÉRINEL, C., LEVILLAIN, P.-M., LOISEAU, H., LOUSSOUARN, D., MAURAGE, C.-A., MENEI, P., JANETTE MOTSUO FOTSO, M., NOEL, G., PARKER, F., PEOC'H, M., POLIVKA, M., QUINTIN-ROUÉ, I., RAMIREZ, C., RICARD, D., RICHARD, P., RIGAU, V., ROUSSEAU, A., RUNAVOT, G., SEVESTRE, H., CHRISTINE TORTEL, M., URO-COSTE, E., BUREL-VANDENBOS, F., VAULEON, E., VIENNET, G., VILLA, C., WAGER, M. & NETWORK, P. 2015a. TCF12 is mutated in anaplastic oligodendroglioma. *Nature Communications,* 6**,** 7207.

LABRECHE, K., SIMEONOVA, I., KAMOUN, A., GLEIZE, V., CHUBB, D., LETOUZÉ, E., RIAZALHOSSEINI, Y., DOBBINS, S. E., ELAROUCI, N., DUCRAY, F., DE REYNIÈS, A., ZELENIKA, D., WARDELL, C. P., FRAMPTON, M., SAULNIER, O., PASTINEN, T., HALLOUT, S., FIGARELLA-BRANGER, D., DEHAIS, C., IDBAIH, A., MOKHTARI, K., DELATTRE, J. Y., HUILLARD, E., MARK LATHROP, G., SANSON, M. & HOULSTON, R. S. 2015b. TCF12 is mutated in anaplastic oligodendroglioma. *Nat Commun,* 6**,** 7207.

LAIOSA, C. V., STADTFELD, M., XIE, H., DE ANDRES-AGUAYO, L. & GRAF, T. 2006. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. *Immunity,* 25**,** 731-44.

LAITMAN, B. M., ASP, L., MARIANI, J. N., ZHANG, J., LIU, J., SAWAI, S., CHAPOULY, C., HORNG, S., KRAMER, E. G., MITIKU, N., LOO, H., BURLANT, N., PEDRE, X., HARA, Y., NUDELMAN, G., ZASLAVSKY, E., LEE, Y. M., BRAUN, D. A., LU, Q. R., NARLA, G., RAINE, C. S., FRIEDMAN, S. L., CASACCIA, P. & JOHN, G. R. 2016. The Transcriptional Activator Krüppel-like Factor-6 Is Required for CNS Myelination. *PLoS Biol,* 14**,** e1002467.

LANGERAK, A. W., WOLVERS-TETTERO, I. L. M., VAN GASTEL-MOL, E. J., OUD, M. E. C. M. & VAN DONGEN, J. J. M. 2001. Basic helix-loop-helix proteins E2A and HEB induce immature T-cell receptor rearrangements in nonlymphoid cells. *Blood,* 98**,** 2456-2465.

LAZO, P. A., KLEIN-SZANTO, A. J. & TSICHLIS, P. N. 1990. T-cell lymphoma lines derived from rat thymomas induced by Moloney murine leukemia virus: phenotypic diversity and its implications. *Journal of Virology,* 64**,** 3948-3959.

LAZORCHAK, A., JONES, M. E. & ZHUANG, Y. 2005. New insights into E-protein function in lymphocyte development. *Trends Immunol,* 26**,** 334-8.

LEE, Y., DECKER, M., LEE, H. & DING, L. 2017. Extrinsic regulation of hematopoietic stem cells in development, homeostasis and diseases. *WIREs Developmental Biology,* 6**,** e279.

LI, H., YANG, Y., HONG, W., HUANG, M., WU, M. & ZHAO, X. 2020. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduction and Targeted Therapy,* 5**,** 1.

LI, P., BURKE, S., WANG, J., CHEN, X., ORTIZ, M., LEE, S. C., LU, D., CAMPOS, L., GOULDING, D., NG, B. L., DOUGAN, G., HUNTLY, B., GOTTGENS, B., JENKINS, N. A., COPELAND, N. G., COLUCCI, F. & LIU, P. 2010. Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. *Science,* 329**,** 85-9.

LI, Y., BRAUER, P. M., SINGH, J., XHIKU, S., YOGANATHAN, K., ZÚÑIGA-PFLÜCKER, J. C. & ANDERSON, M. K. 2017. Targeted Disruption of TCF12 Reveals HEB as Essential in Human Mesodermal Specification and Hematopoiesis. *Stem Cell Reports,* 9**,** 779-795.

LIAO, M. & WANG, J. 2021. Tcf12 balances the reconstitution and differentiation capacity of hematopoietic stem cell. *Blood Sci,* 3**,** 14-19.

LIE, A. L. M., MARINOPOULOU, E., LILLY, A. J., CHALLINOR, M., PATEL, R., LANCRIN, C., KOUSKOFF, V. & LACAUD, G. 2018. Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. *Development,* 145.

LINK, K. A., CHOU, F. S. & MULLOY, J. C. 2010. Core binding factor at the crossroads: determining the fate of the HSC. *J Cell Physiol,* 222**,** 50-6.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-6.

MASSARI, M. E. & MURRE, C. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol,* 20**,** 429-40.

MCLEAN, C. Y., BRISTOR, D., HILLER, M., CLARKE, S. L., SCHAAR, B. T., LOWE, C. B., WENGER, A. M. & BEJERANO, G. 2010a. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol,* 28**,** 495-501.

MCLEAN, C. Y., BRISTOR, D., HILLER, M., CLARKE, S. L., SCHAAR, B. T., LOWE, C. B., WENGER, A. M. & BEJERANO, G. 2010b. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology,* 28**,** 495-501.

MEIER, N., KRPIC, S., RODRIGUEZ, P., STROUBOULIS, J., MONTI, M., KRIJGSVELD, J., GERING, M., PATIENT, R., HOSTERT, A. & GROSVELD, F. 2006. Novel binding partners of Ldb1 are required for haematopoietic development. *Development,* 133**,** 4913-23.

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., 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 αβ T cell differentiation. *Nat Immunol,* 14**,** 619-32.

MIYAZAKI, M., MIYAZAKI, K., CHEN, K., JIN, Y., TURNER, J., MOORE, A. J., SAITO, R., YOSHIDA, K., OGAWA, S. & RODEWALD, H.-R. 2017. The E-Id protein axis specifies adaptive lymphoid cell identity and suppresses thymic innate lymphoid cell development. *Immunity,* 46**,** 818-834. e4.

MOJICA, F. J., DÍEZ-VILLASEÑOR, C., GARCÍA-MARTÍNEZ, J. & SORIA, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol,* 60**,** 174-82.

MONTAGUE, T. G., CRUZ, J. M., GAGNON, J. A., CHURCH, G. M. & VALEN, E. 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res,* 42**,** W401-7.

MONTAÑO, A., FORERO-CASTRO, M., HERNÁNDEZ-RIVAS, J.-M., GARCÍA-TUÑÓN, I. & BENITO, R. 2018a. Targeted genome editing in acute lymphoblastic leukemia: a review. *BMC Biotechnology,* 18**,** 45.

MONTAÑO, A., FORERO-CASTRO, M., MARCHENA-MENDOZA, D., BENITO, R. & HERNÁNDEZ-RIVAS, J. M. 2018b. New Challenges in Targeting Signaling Pathways in Acute Lymphoblastic Leukemia by NGS Approaches: An Update. *Cancers (Basel),* 10.

MÖRÖY, T. 2005. The zinc finger transcription factor Growth factor independence 1 (Gfi1). *The International Journal of Biochemistry & Cell Biology,* 37**,** 541-546.

MÖRÖY, T., VASSEN, L., WILKES, B. & KHANDANPOUR, C. 2015. From cytopenia to leukemia: the role of Gfi1 and Gfi1b in blood formation. *Blood,* 126**,** 2561-2569.

MORRISON, S. J. & WEISSMAN, I. L. 1994. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity,* 1**,** 661-73.

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.

NAITO, Y., HINO, K., BONO, H. & UI-TEI, K. 2015. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics,* 31**,** 1120-3.

NATKUNAM, Y., ZHAO, S., MASON, D. Y., CHEN, J., TAIDI, B., JONES, M., HAMMER, A. S., HAMILTON DUTOIT, S., LOSSOS, I. S. & LEVY, R. 2007. The oncoprotein LMO2 is expressed in normal germinal-center B cells and in human B-cell lymphomas. *Blood,* 109**,** 1636-42.

NAVARRO, J.-M., TOUZART, A., PRADEL, L. C., LOOSVELD, M., KOUBI, M., FENOUIL, R., LE NOIR, S., MAQBOOL, M. A., MORGADO, E., GREGOIRE, C., JAEGER, S., MAMESSIER, E., PIGNON, C., HACEIN-BEY-ABINA, S., MALISSEN, B., GUT, M., GUT, I. G., DOMBRET, H., MACINTYRE, E. A., HOWE, S. J., GASPAR, H. B., THRASHER, A. J., IFRAH, N., PAYET-BORNET, D., DUPREZ, E., ANDRAU, J.-C., ASNAFI, V. & NADEL, B. 2015. Site- and allele-specific polycomb dysregulation in T-cell leukaemia. *Nature Communications,* 6**,** 6094.

NEMUDRYI, A. A., VALETDINOVA, K. R., MEDVEDEV, S. P. & ZAKIAN, S. M. 2014. TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta Naturae,* 6**,** 19-40.

NIEHUES, T., KAPAUN, P., HARMS, D. O., BURDACH, S., KRAMM, C., KÖRHOLZ, D., JANKA-SCHAUB, G. & GÖBEL, U. 1999. A classification based on T cell selection-related phenotypes identifies a subgroup of childhood T-ALL with favorable outcome in the COALL studies. *Leukemia,* 13**,** 614-7.

NORMAN, M., RIVERS, C., LEE, Y. B., IDRIS, J. & UNEY, J. 2016. The increasing diversity of functions attributed to the SAFB family of RNA-/DNA-binding proteins. *Biochem J,* 473**,** 4271-4288.

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.

OGURO, H., DING, L. & MORRISON, SEAN J. 2013. SLAM Family Markers Resolve Functionally Distinct Subpopulations of Hematopoietic Stem Cells and Multipotent Progenitors. *Cell Stem Cell,* 13**,** 102-116.

OMAIR, S. A. 2019. *The consequences of the aberrant expression of TAL1 and its isoforms in T cell acute lymphoblastic leukaemia.* PhD, University of Birmingham.

ORKIN, S. H. & ZON, L. I. 2008. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell,* 132**,** 631-44.

PALIS, J. & YODER, M. C. 2001. Yolk-sac hematopoiesis: the first blood cells of mouse and man. *Exp Hematol,* 29**,** 927-36.

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.

PARKER, C. E., MOCANU, V., MOCANU, M., DICHEVA, N. & WARREN, M. R. 2010. Frontiers in Neuroscience

Mass Spectrometry for Post-Translational Modifications. *In:* ALZATE, O. (ed.) *Neuroproteomics.* Boca Raton (FL): CRC Press/Taylor & Francis

Copyright © 2010 by Taylor and Francis Group, LLC.

PERSON, R. E., LI, F. Q., DUAN, Z., BENSON, K. F., WECHSLER, J., PAPADAKI, H. A., ELIOPOULOS, G., KAUFMAN, C., BERTOLONE, S. J., NAKAMOTO, B., PAPAYANNOPOULOU, T., GRIMES, H. L. & HORWITZ, M. 2003. Mutations in proto-oncogene GFI1 cause human neutropenia and target ELA2. *Nat Genet,* 34**,** 308-12.

POURCEL, C., SALVIGNOL, G. & VERGNAUD, G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading),* 151**,** 653-663.

PUI, C.-H., ROBISON, L. L. & LOOK, A. T. 2008. Acute lymphoblastic leukaemia. *The Lancet,* 371**,** 1030-1043.

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, J. C., ALLMAN, D., XU, L., DEROCCO, 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.

RAHMAN, S., MAGNUSSEN, M., LEÓN, 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.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols,* 8**,** 2281-2308.

RAO, C., MALAGUTI, M., MASON, J. O. & LOWELL, S. 2020. The transcription factor E2A drives neural differentiation in pluripotent cells. *Development,* 147.

REDMAN, M., KING, A., WATSON, C. & KING, D. 2016. What is CRISPR/Cas9? *Archives of disease in childhood - Education & practice edition,* 101**,** 213-215.

ROBERTSON, S. M., KENNEDY, M., SHANNON, J. M. & KELLER, G. 2000. A transitional stage in the commitment of mesoderm to hematopoiesis requiring the transcription factor SCL/tal-1. *Development,* 127**,** 2447-59.

ROSE, J. K. 2003. Optimization of transfection. *Curr Protoc Cell Biol,* Chapter 20**,** Unit 20.7.

ROTHENBERG, E. V. 2014. Transcriptional control of early T and B cell developmental choices. *Annual review of immunology,* 32**,** 283-321.

ROTHENBERG, E. V., UNGERBÄCK, J. & CHAMPHEKAR, A. 2016. Forging T-Lymphocyte Identity: Intersecting Networks of Transcriptional Control. *Adv Immunol,* 129**,** 109-74.

ROY, P. S. & SAIKIA, B. J. 2016. Cancer and cure: A critical analysis. *Indian J Cancer,* 53**,** 441-442.

RUDOLPH, B., HUEBER, A. O. & EVAN, G. I. 2001. Expression of Mad1 in T cells leads to reduced thymic cellularity and impaired mitogen-induced proliferation. *Oncogene,* 20**,** 1164-75.

RYAN, D. P., SUNDE, M., KWAN, A. H. Y., MARIANAYAGAM, N. J., NANCARROW, A. L., VANDEN HOVEN, R. N., THOMPSON, L. S., BACA, M., MACKAY, J. P., VISVADER, J. E. & MATTHEWS, J. M. 2006. Identification of the key LMO2-binding determinants on Ldb1. *Journal of molecular biology,* 359**,** 66-75.

SAKUMA, T., MOCHIDA, K., NAKADE, S., EZURE, T., MINAGAWA, S. & YAMAMOTO, T. 2018. Unexpected heterogeneity derived from Cas9 ribonucleoprotein-introduced clonal cells at the HPRT1 locus. *Genes Cells,* 23**,** 255-263.

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. 2007a. 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.

SANDBERG, Y., VERHAAF, B., VAN GASTEL-MOL, E. J., WOLVERS-TETTERO, I. L. M., DE VOS, J., MACLEOD, R. A. F., NOORDZIJ, J. G., DIK, W. A., VAN DONGEN, J. J. M. & LANGERAK, A. W. 2007b. 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**,** 230-237.

SANDER, H., WALLACE, S., PLOUSE, R., TIWARI, S. & GOMES, A. V. 2019. Ponceau S waste: Ponceau S staining for total protein normalization. *Anal Biochem,* 575**,** 44-53.

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.

SCHMIDT, T., KARSUNKY, H., RODEL, B., ZEVNIK, B., ELSASSER, H. P. & MOROY, T. 1998. Evidence implicating Gfi-1 and Pim-1 in pre-T-cell differentiation steps associated with beta-selection. *Embo j,* 17**,** 5349-59.

SCHWANK, G., KOO, B. K., SASSELLI, V., DEKKERS, J. F., HEO, I., DEMIRCAN, T., SASAKI, N., BOYMANS, S., CUPPEN, E., VAN DER ENT, C. K., NIEUWENHUIS, E. E., BEEKMAN, J. M. & CLEVERS, H. 2013. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell,* 13**,** 653-8.

SCHWARTZ, R., ENGEL, I., FALLAHI-SICHANI, M., PETRIE, H. T. & MURRE, C. 2006a. Gene expression patterns define novel roles for E47 in cell cycle progression, cytokine-mediated signaling, and T lineage development. *Proceedings of the National Academy of Sciences,* 103**,** 9976-9981.

SCHWARTZ, R., ENGEL, I., FALLAHI-SICHANI, M., PETRIE, H. T. & MURRE, C. 2006b. Gene expression patterns define novel roles for E47 in cell cycle progression, cytokine-mediated signaling, and T lineage development. *Proc Natl Acad Sci U S A,* 103**,** 9976-81.

SEITA, J. & WEISSMAN, I. L. 2010. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med,* 2**,** 640-53.

SHARMA, A., MISTRIEL-ZERBIB, S., NAJAR, R. A., ENGAL, E., BENTATA, M., TAQATQA, N., DAHAN, S., COHEN, K., JAFFE-HERMAN, S., GEMINDER, O., BAKER, M., NEVO, Y., PLASCHKES, I., KAY, G., DRIER, Y., BERGER, M. & SALTON, M. 2023. Isoforms of the TAL1 transcription factor have different roles in hematopoiesis and cell growth. *PLoS Biol,* 21**,** e3002175.

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.

SLAYMAKER, I. M., GAO, L., ZETSCHE, B., SCOTT, D. A., YAN, W. X. & ZHANG, F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science,* 351**,** 84-8.

SMITH, C., GORE, A., YAN, W., ABALDE-ATRISTAIN, L., LI, Z., HE, C., WANG, Y., BRODSKY, R. A., ZHANG, K., CHENG, L. & YE, Z. 2014. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell,* 15**,** 12-3.

SONG, S. H., KIM, A., RAGOCZY, T., BENDER, M. A., GROUDINE, M. & DEAN, A. 2010. Multiple functions of Ldb1 required for beta-globin activation during erythroid differentiation. *Blood,* 116**,** 2356-64.

SOUABNI, A., COBALEDA, C., SCHEBESTA, M. & BUSSLINGER, M. 2002. Pax5 promotes B lymphopoiesis and blocks T cell development by repressing Notch1. *Immunity,* 17**,** 781-93.

SPANGRUDE, G. J., HEIMFELD, S. & WEISSMAN, I. L. 1988. Purification and characterization of mouse hematopoietic stem cells. *Science,* 241**,** 58-62.

STANULOVIĆ, V. S., BINHASSAN, S., DORRINGTON, I., WARD, D. G. & HOOGENKAMP, M. 2020. PHF6 Interacts with LMO2 During Normal Haematopoiesis and in Leukaemia and Regulates Gene Expression and Genome Integrity. *bioRxiv***,** 2020.08.18.255471.

STANULOVIĆ, 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 Research,* 45**,** 9874-9888.

SUBRAMANIAN, S. V. & NADAL-GINARD, B. 1996. Early expression of the different isoforms of the myocyte enhancer factor-2 (MEF2) protein in myogenic as well as non-myogenic cell lineages during mouse embryogenesis. *Mechanisms of Development,* 57**,** 103-112.

SUN, W., GUO, J., MCCLELLAN, D., POESCHLA, A., BAREYAN, D., CASEY, M. J., CAIRNS, B. R., TANTIN, D. & ENGEL, M. E. 2022. GFI1 Cooperates with IKZF1/IKAROS to Activate Gene Expression in T-cell Acute Lymphoblastic Leukemia. *Mol Cancer Res,* 20**,** 501-514.

SZKLARCZYK, D., GABLE, A., LYON, D., JUNGE, A., WYDER, S., HUERTA-CEPAS, J., SIMONOVIC, M., DONCHEVA, N., MORRIS, J., BORK, P., JENSEN, L. & VON MERING, C. 2018. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research,* 47.

SZONDY, Z., GARABUCZI, É., TÓTH, K., KISS, B. & KÖRÖSKÉNYI, K. 2012. Thymocyte death by neglect: contribution of engulfing macrophages. *Eur J Immunol,* 42**,** 1662-7.

SZYCHOT, E., BRODKIEWICZ, A. & PEREGUD-POGORZELSKI, J. 2014. How have advances in our understanding of the molecular genetics of paediatric leukaemia led to improved targeted therapies for these diseases? *Adv Clin Exp Med,* 23**,** 469-74.

TAGHON, T., YUI, M. A., PANT, R., DIAMOND, R. A. & ROTHENBERG, E. V. 2006. Developmental and Molecular Characterization of Emerging β- and γδ-Selected Pre-T Cells in the Adult Mouse Thymus. *Immunity,* 24**,** 53-64.

TAGHON, T., YUI, M. A. & ROTHENBERG, E. V. 2007. Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nat Immunol,* 8**,** 845-55.

TAN, T. K., ZHANG, C. & SANDA, T. 2019. Oncogenic transcriptional program driven by TAL1 in T-cell acute lymphoblastic leukemia. *International Journal of Hematology,* 109**,** 5-17.

TANIUCHI, I., OSATO, M., EGAWA, T., SUNSHINE, M. J., BAE, S. C., KOMORI, T., ITO, Y. & LITTMAN, D. R. 2002. Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell,* 111**,** 621-33.

TEN HAVE, S., BOULON, S., AHMAD, Y. & LAMOND, A. I. 2011. Mass spectrometry-based immuno-precipitation proteomics - the user's guide. *Proteomics,* 11**,** 1153-9.

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.

TOBER, J., KONISKI, A., MCGRATH, K. E., VEMISHETTI, R., EMERSON, R., DE MESY-BENTLEY, K. K. L., WAUGH, R. & PALIS, J. 2007. The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood,* 109**,** 1433-1441.

TORRES, R., MARTIN, M. C., GARCIA, A., CIGUDOSA, J. C., RAMIREZ, J. C. & RODRIGUEZ-PERALES, S. 2014. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat Commun,* 5**,** 3964.

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., HERBLOT, S., LECUYER, E. & HOANG, T. 2003. Regulation of pT alpha gene expression by a dosage of E2A, HEB, and SCL. *J Biol Chem,* 278**,** 12680-7.

TREMBLAY, M., TREMBLAY, C. S., HERBLOT, S., APLAN, P. D., HÉBERT, 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.

URNOV, F. D., REBAR, E. J., HOLMES, M. C., ZHANG, H. S. & GREGORY, P. D. 2010. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet,* 11**,** 636-46.

VADNAIS, C., CHEN, R., FRASZCZAK, J., YU, Z., BOULAIS, J., PINDER, J., FRANK, D., KHANDANPOUR, C., HÉBERT, J., DELLAIRE, G., CÔTÉ, J. F., RICHARD, S., ORTHWEIN, A., DROBETSKY, E. & MÖRÖY, T. 2018. GFI1 facilitates efficient DNA repair by regulating PRMT1 dependent methylation of MRE11 and 53BP1. *Nat Commun,* 9**,** 1418.

VAN DER MEER, L. T., JANSEN, J. H. & VAN DER REIJDEN, B. A. 2010. Gfi1 and Gfi1b: key regulators of hematopoiesis. *Leukemia,* 24**,** 1834-1843.

VAN VLIERBERGHE, P. & FERRANDO, A. 2012. The molecular basis of T cell acute lymphoblastic leukemia. *J Clin Invest,* 122**,** 3398-406.

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.

VASTRAD, B., VASTRAD, C. & TENGLI, A. 2020. Identification of potential mRNA panels for severe acute respiratory syndrome coronavirus 2 (COVID-19) diagnosis and treatment using microarray dataset and bioinformatics methods. *3 Biotech,* 10**,** 422.

VERMA, D., ZANETTI, C., GODAVARTHY, P. S., KUMAR, R., MINCIACCHI, V. R., PFEIFFER, J., METZLER, M., LEFORT, S., MAGUER-SATTA, V., NICOLINI, F. E., BURRONI, B., FONTENAY, M. & KRAUSE, D. S. 2020. Bone marrow niche-derived extracellular matrix-degrading enzymes influence the progression of B-cell acute lymphoblastic leukemia. *Leukemia,* 34**,** 1540-1552.

VOSSHENRICH, C. A., GARCIA-OJEDA, M. E., SAMSON-VILLEGER, S. I., PASQUALETTO, V., ENAULT, L., RICHARD-LE GOFF, O., CORCUFF, E., GUY-GRAND, D., ROCHA, B., CUMANO, A., ROGGE, L., EZINE, S. & DI SANTO, J. P. 2006. A thymic pathway of mouse natural killer cell development characterized by expression of GATA-3 and CD127. *Nat Immunol,* 7**,** 1217-24.

WADMAN, I. A., OSADA, H., GRÜTZ, 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. *The EMBO Journal,* 16**,** 3145-3157.

WANG, D., CLAUS, C. L., RAJKUMAR, P., BRAUNSTEIN, M., MOORE, A. J., SIGVARDSSON, M. & ANDERSON, M. K. 2010. Context-dependent regulation of hematopoietic lineage choice by HEBAlt. *J Immunol,* 185**,** 4109-17.

WANG, D., CLAUS, C. L., VACCARELLI, G., BRAUNSTEIN, M., SCHMITT, T. M., ZUNIGA-PFLUCKER, J. C., ROTHENBERG, E. V. & ANDERSON, M. K. 2006a. The basic helix-loop-helix transcription factor HEBAlt is expressed in pro-T cells and enhances the generation of T cell precursors. *J Immunol,* 177**,** 109-19.

WANG, D., CLAUS, C. L., VACCARELLI, G., BRAUNSTEIN, M., SCHMITT, T. M., ZÚÑIGA-PFLÜCKER, J. C., ROTHENBERG, E. V. & ANDERSON, M. K. 2006b. The Basic Helix-Loop-Helix Transcription Factor HEBAlt Is Expressed in Pro-T Cells and Enhances the Generation of T Cell Precursors1. *The Journal of Immunology,* 177**,** 109-119.

WANG, Z. X., KUEH, J. L., TEH, C. H., ROSSBACH, M., LIM, L., LI, P., WONG, K. Y., LUFKIN, T., ROBSON, P. & STANTON, L. W. 2007. Zfp206 is a transcription factor that controls pluripotency of embryonic stem cells. *Stem Cells,* 25**,** 2173-82.

WEDEL, M., FRÖB, F., ELSESSER, O., WITTMANN, M.-T., LIE, D C., REIS, A. & WEGNER, M. 2020. Transcription factor Tcf4 is the preferred heterodimerization partner for Olig2 in oligodendrocytes and required for differentiation. *Nucleic Acids Research,* 48**,** 4839-4857.

WILSON, N. K., TIMMS, R. T., KINSTON, S. J., CHENG, Y. H., ORAM, S. H., LANDRY, J. R., MULLENDER, J., OTTERSBACH, K. & GOTTGENS, B. 2010. Gfi1 expression is controlled by five distinct regulatory regions spread over 100 kilobases, with Scl/Tal1, Gata2, PU.1, Erg, Meis1, and Runx1 acting as upstream regulators in early hematopoietic cells. *Mol Cell Biol,* 30**,** 3853-63.

WITZEL, H. R., JUNGBLUT, B., CHOE, C. P., CRUMP, J. G., BRAUN, T. & DOBREVA, G. 2012. The LIM protein Ajuba restricts the second heart field progenitor pool by regulating Isl1 activity. *Dev Cell,* 23**,** 58-70.

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.

XIE, H., YE, M., FENG, R. & GRAF, T. 2004. Stepwise reprogramming of B cells into macrophages. *Cell,* 117**,** 663-76.

YAMADA, Y., WARREN, A. J., DOBSON, C., FORSTER, A., PANNELL, R. & RABBITTS, T. H. 1998. The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proc Natl Acad Sci U S A,* 95**,** 3890-5.

YI, S., YU, M., YANG, S., MIRON, R. J. & ZHANG, Y. 2017. Tcf12, A Member of Basic Helix-Loop-Helix Transcription Factors, Mediates Bone Marrow Mesenchymal Stem Cell Osteogenic Differentiation In Vitro and In Vivo. *Stem Cells,* 35**,** 386-397.

YUCEL, R., KARSUNKY, H., KLEIN-HITPASS, L. & MOROY, T. 2003. The transcriptional repressor Gfi1 affects development of early, uncommitted c-Kit+ T cell progenitors and CD4/CD8 lineage decision in the thymus. *J Exp Med,* 197**,** 831-44.

YÜCEL, R., KARSUNKY, H., KLEIN-HITPASS, L. & MÖRÖY, T. 2003. The transcriptional repressor Gfi1 affects development of early, uncommitted c-Kit+ T cell progenitors and CD4/CD8 lineage decision in the thymus. *J Exp Med,* 197**,** 831-44.

YUI, M. A., FENG, N. & ROTHENBERG, E. V. 2010. Fine-scale staging of T cell lineage commitment in adult mouse thymus. *J Immunol,* 185**,** 284-93.

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.

ZENG, H., YÜCEL, R., KOSAN, C., KLEIN-HITPASS, L. & MÖRÖY, T. 2004. Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells. *Embo j,* 23**,** 4116-25.

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., 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 Biology,* 9**,** R137.

ZHENG, Q. & ZHAO, Y. 2007. The diverse biofunctions of LIM domain proteins: determined by subcellular localization and protein—protein interaction. *Biology of the Cell,* 99**,** 489-502.

ZHONG, Y., JIANG, L., HIAI, H., TOYOKUNI, S. & YAMADA, Y. 2007. Overexpression of a transcription factor LYL1 induces T- and B-cell lymphoma in mice. *Oncogene,* 26**,** 6937-47.

ZHUANG, Y., CHENG, P. & WEINTRAUB, H. 1996. B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. *Mol Cell Biol,* 16**,** 2898-905.

ZÖRNIG, M., SCHMIDT, T., KARSUNKY, H., GRZESCHICZEK, A. & MÖRÖY, T. 1996. Zinc finger protein GFI-1 cooperates with myc and pim-1 in T-cell lymphomagenesis by reducing the requirements for IL-2. *Oncogene,* 12**,** 1789-801.

SUPPLEMENTARY

ARR

GFI1 Mass spectrometry	Kme1 Mass spectrometry	Common between GFI1 and Kme1
AHNK_HUMAN	ACACA_HUMAN	AHNK_HUMAN
U5S1_HUMAN	AHNK_HUMAN	DIDO1_HUMAN
DIDO1_HUMAN	HSP71_HUMAN	TBB2C_HUMAN
TBB2C_HUMAN	HS71L_HUMAN	SRRM2_HUMAN
SRRM2_HUMAN	K2C6C_HUMAN	SR140_HUMAN
TBA1B_HUMAN	SMCA5_HUMAN	PCF11_HUMAN
SR140_HUMAN	RL10A_HUMAN	PRP8_HUMAN
PCF11_HUMAN	MYH9_HUMAN	RENT1_HUMAN
PRP8_HUMAN	LAP2B_HUMAN	HNRH1_HUMAN
PRP16_HUMAN	SRRM2_HUMAN	EIF3C_HUMAN
RENT1_HUMAN	K1C16_HUMAN	NONO_HUMAN
HNRH1_HUMAN	PRKDC_HUMAN	DHX30_HUMAN
EIF3C_HUMAN	CKAP5_HUMAN	RS24_HUMAN
DDX42_HUMAN	CNOT1_HUMAN	HCFC1_HUMAN
TBA4A_HUMAN	NUFP2_HUMAN	NUFP2_HUMAN
NONO_HUMAN	H12_HUMAN	HS71L_HUMAN
DHX30_HUMAN	NUMA1_HUMAN	PRKDC_HUMAN
RS24_HUMAN	FLNB_HUMAN	CKAP5_HUMAN
HCFC1_HUMAN	WIZ_HUMAN	H12_HUMAN
ATX2L_HUMAN	IF16_HUMAN	RS17_HUMAN
NUFP2_HUMAN	RBBP7_HUMAN	EIF3F_HUMAN
HS71L_HUMAN	RENT1_HUMAN	DDX23_HUMAN
PRKDC_HUMAN	TBB2C_HUMAN	RL32_HUMAN
CKAP5_HUMAN	ATX2L_HUMAN	LAP2B_HUMAN
H12_HUMAN	RS17_HUMAN	FXR2_HUMAN
RS17_HUMAN	UBP10_HUMAN	EIF3L_HUMAN
EIF3F_HUMAN	DHX30_HUMAN	RPB1_HUMAN
G3BP2_HUMAN	NONO_HUMAN	RS15A_HUMAN
DDX23_HUMAN	SFRS1_HUMAN	HSP71_HUMAN
RL32_HUMAN	FXR2_HUMAN	RBBP7_HUMAN
LAP2B_HUMAN	RBBP4_HUMAN	RBM10_HUMAN
CNOT1_HUMAN	SR140_HUMAN	IF2B3_HUMAN
FXR2_HUMAN	RL13A_HUMAN	ATX2_HUMAN
EIF3L_HUMAN	PRP8_HUMAN	ZFR_HUMAN
RPB1_HUMAN	PSPC1_HUMAN	FLNB_HUMAN
RS15A_HUMAN	RL10_HUMAN	SAFB1_HUMAN
HSP71_HUMAN	U5S1_HUMAN	ELAV2_HUMAN
RBBP7_HUMAN	HCFC1_HUMAN	HNRL1_HUMAN
RBM10_HUMAN	TRAM1_HUMAN	ACINU_HUMAN
IF2B3_HUMAN	RS10_HUMAN	NOP58_HUMAN
ATX2_HUMAN	KIF22_HUMAN	RUNX1_HUMAN
ACTBL_HUMAN	RUVB1_HUMAN	SMCA4_HUMAN
ZFR_HUMAN	H2B1J_HUMAN	DDX3X_HUMAN
IF2B2_HUMAN	KI20B_HUMAN	THOC2_HUMAN
ZC3H4_HUMAN	BAZ1A_HUMAN	PININ_HUMAN
BAT2_HUMAN	EF1G_HUMAN	RUVB1_HUMAN
FLNB_HUMAN	PCM1_HUMAN	RL23_HUMAN
SAFB1_HUMAN	CBX3_HUMAN	MYH9_HUMAN
UBP10_HUMAN	RUVB2_HUMAN	ZN638_HUMAN
DDX50_HUMAN	SPTA2_HUMAN	PRP19_HUMAN
RL30_HUMAN	DECR_HUMAN	MRE11_HUMAN
SNW1_HUMAN	TBA1A_HUMAN	UAP56_HUMAN
ELAV2_HUMAN	RL17_HUMAN	RBM5_HUMAN
HNRL1_HUMAN	DDX3X_HUMAN	ZCCHV_HUMAN
ACINU_HUMAN	EP400_HUMAN	MBB1A_HUMAN
NOP58_HUMAN	HORN_HUMAN	PDIP3_HUMAN
SF3B2_HUMAN	CHD3_HUMAN	DHX36_HUMAN
RUNX1_HUMAN	ARI1A_HUMAN	YLPM1_HUMAN
SFRS1_HUMAN	LYRIC_HUMAN	SAFB2_HUMAN
SMCA4_HUMAN	DESP_HUMAN	IF16_HUMAN
DDX3X_HUMAN	RL32_HUMAN	PERQ2_HUMAN
TP53B_HUMAN	NAT10_HUMAN	DECR_HUMAN
SCRIB_HUMAN	ZCCHV_HUMAN	RL10_HUMAN
THOC2_HUMAN	RS15A_HUMAN	LC7L2_HUMAN
PININ_HUMAN	RBM25_HUMAN	SK2L2_HUMAN
RUVB1_HUMAN	ATX2_HUMAN	HDAC1_HUMAN
RL23_HUMAN	EIF3C_HUMAN	RL13A_HUMAN
MYH9_HUMAN	KU86_HUMAN	NAT10_HUMAN
AIFM1_HUMAN	RUNX1_HUMAN	PCM1_HUMAN
RL10A_HUMAN	EF1B_HUMAN	MINT_HUMAN
ZN638_HUMAN	ZFR_HUMAN	TRAM1_HUMAN
PRP19_HUMAN	KIF4A_HUMAN	F120A_HUMAN
MRE11_HUMAN	NOP58_HUMAN	RFA1_HUMAN
UAP56_HUMAN	SMCA4_HUMAN	RUVB2_HUMAN

DU528

GFI1 Mass spectrometry	Kme1 Mass spectrometry	Common between GFI1 and Kme1
TBB2C_HUMAN	ACTG_HUMAN	TBB2C_HUMAN
TBA4A_HUMAN	TBB2C_HUMAN	PRKDC_HUMAN
PRKDC_HUMAN	PARP1_HUMAN	RL18_HUMAN
ACTC_HUMAN	HSP71_HUMAN	DHX15_HUMAN
RL18_HUMAN	SMCA5_HUMAN	FUS_HUMAN
DHX15_HUMAN	ACACA_HUMAN	RS8_HUMAN
FUS_HUMAN	PRP6_HUMAN	IF4G1_HUMAN
RS8_HUMAN	U5S1_HUMAN	SF3B3_HUMAN
IF4G1_HUMAN	ACTS_HUMAN	U5S1_HUMAN
SF3B3_HUMAN	DDX23_HUMAN	NUFP2_HUMAN
U5S1_HUMAN	RS8_HUMAN	RS9_HUMAN
NUFP2_HUMAN	MYO1G_HUMAN	PRP6_HUMAN
RS9_HUMAN	SF3B3_HUMAN	G3BP2_HUMAN
PRP6_HUMAN	SP16H_HUMAN	HNRPQ_HUMAN
G3BP2_HUMAN	RL18_HUMAN	DDX50_HUMAN
HNRPQ_HUMAN	NAT10_HUMAN	DDX23_HUMAN
DDX50_HUMAN	EF1D_HUMAN	EIF3D_HUMAN
DDX23_HUMAN	POTEF_HUMAN	RL10A_HUMAN
EIF3D_HUMAN	SF3B1_HUMAN	RL1D1_HUMAN
RL10A_HUMAN	RENT1_HUMAN	RS5_HUMAN
RL1D1_HUMAN	PRKDC_HUMAN	RL24_HUMAN
RS5_HUMAN	DHX30_HUMAN	EIF3A_HUMAN
RL24_HUMAN	HS90B_HUMAN	RL17_HUMAN
EIF3A_HUMAN	RS9_HUMAN	MRE11_HUMAN
RL17_HUMAN	RFC1_HUMAN	RL18A_HUMAN
MRE11_HUMAN	H2A1H_HUMAN	H2A1H_HUMAN
RL18A_HUMAN	DNLI3_HUMAN	MYO1G_HUMAN
H2A1H_HUMAN	HS90A_HUMAN	RL13A_HUMAN
MYO1G_HUMAN	RL18A_HUMAN	UBP10_HUMAN
RL13A_HUMAN	EF1G_HUMAN	RENT1_HUMAN
UBP10_HUMAN	KU86_HUMAN	SF3B1_HUMAN
HELC1_HUMAN	H14_HUMAN	LYAR_HUMAN
RENT1_HUMAN	RL10A_HUMAN	SR140_HUMAN
SF3B1_HUMAN	H2B1N_HUMAN	NAT10_HUMAN
LYAR_HUMAN	FUS_HUMAN	DDX1_HUMAN
SR140_HUMAN	SR140_HUMAN	RU2A_HUMAN
NAT10_HUMAN	IF4G1_HUMAN	PARP1_HUMAN
RL38_HUMAN	FLII_HUMAN	PRP19_HUMAN
DDX1_HUMAN	CDC5L_HUMAN	LAP2B_HUMAN
RU2A_HUMAN	RL17_HUMAN	PCM1_HUMAN
PARP1_HUMAN	SNUT1_HUMAN	NOP2_HUMAN
EIF3L_HUMAN	CHERP_HUMAN	RL15_HUMAN
PRP19_HUMAN	CNOT1_HUMAN	SMCA4_HUMAN
LAP2B_HUMAN	TR150_HUMAN	RL19_HUMAN
PCM1_HUMAN	RS5_HUMAN	DHX30_HUMAN
NOP2_HUMAN	CHD4_HUMAN	KU86_HUMAN
WBP11_HUMAN	MSH6_HUMAN	DOCK8_HUMAN
RL15_HUMAN	RL24_HUMAN	ATD3A_HUMAN
SMCA4_HUMAN	HNRPQ_HUMAN	RL21_HUMAN
DDX3Y_HUMAN	ILF3_HUMAN	RUVB1_HUMAN
SMCA2_HUMAN	UBP10_HUMAN	H2AV_HUMAN
RL19_HUMAN	RL13A_HUMAN	PAIRB_HUMAN
DHX30_HUMAN	SF3B2_HUMAN	MSH6_HUMAN
KU86_HUMAN	TCPG_HUMAN	FXR2_HUMAN
MACF4_HUMAN	BCLF1_HUMAN	UN84B_HUMAN
DOCK8_HUMAN	DDX1_HUMAN	ILF3_HUMAN
ATD3A_HUMAN	TADBP_HUMAN	NOP58_HUMAN
MYCB2_HUMAN	EF2_HUMAN	SF3B2_HUMAN
ROA3_HUMAN	RFA1_HUMAN	H1X_HUMAN
RL21_HUMAN	XRN2_HUMAN	NHP2_HUMAN
RUVB1_HUMAN	GCP3_HUMAN	TRAP1_HUMAN
H2AV_HUMAN	NOLC1_HUMAN	RN213_HUMAN
RS28_HUMAN	RU2A_HUMAN	SNUT1_HUMAN
PAIRB_HUMAN	H2AV_HUMAN	PFD2_HUMAN
MSH6_HUMAN	IKZF1_HUMAN	NCBP1_HUMAN
FXR2_HUMAN	RL9_HUMAN	HS90B_HUMAN
UN84B_HUMAN	LYAR_HUMAN	IF16_HUMAN
ILF3_HUMAN	EF1B_HUMAN	SRP72_HUMAN
NOP58_HUMAN	ACL6A_HUMAN	HNRPF_HUMAN
SF3B2_HUMAN	COPA_HUMAN	CHERP_HUMAN
H1X_HUMAN	HNRPF_HUMAN	RALYL_HUMAN
NHP2_HUMAN	RL19_HUMAN	ARI1A_HUMAN
TRAP1_HUMAN	RL15_HUMAN	EIF3C_HUMAN
RN213_HUMAN	NOP2_HUMAN	CHD4_HUMAN

RBM5_HUMAN	SK2L2_HUMAN	PA2G4_HUMAN	ADT2_HUMAN	TRAP1_HUMAN	PIHD1_HUMAN
ZCCHV_HUMAN	SAFB1_HUMAN	MED14_HUMAN	PQBP1_HUMAN	LAP2B_HUMAN	EIF3B_HUMAN
MBB1A_HUMAN	RS24_HUMAN	EIF3B_HUMAN	SNUT1_HUMAN	H2AY_HUMAN	AGO1_HUMAN
PDIP3_HUMAN	DDX42_HUMAN	ARI1A_HUMAN	PFD2_HUMAN	DDX50_HUMAN	ARPC4_HUMAN
DHX36_HUMAN	SRPK1_HUMAN	RNPS1_HUMAN	NCBP1_HUMAN	AP2A1_HUMAN	RL31_HUMAN
YLPM1_HUMAN	HNRH1_HUMAN	CV028_HUMAN	HS90B_HUMAN	NUFP2_HUMAN	RS24_HUMAN
RS10_HUMAN	RBM10_HUMAN	KIF23_HUMAN	IF16_HUMAN	UBP2L_HUMAN	RS7_HUMAN
HNRH2_HUMAN	ADNP_HUMAN	THOC1_HUMAN	SRP72_HUMAN	PCM1_HUMAN	PERI_HUMAN
EIF3E_HUMAN	BLM_HUMAN	ZC11A_HUMAN	ERH_HUMAN	RBM10_HUMAN	TADBP_HUMAN
SAFB2_HUMAN	PERQ2_HUMAN	RL35_HUMAN	HNRH2_HUMAN	DHX15_HUMAN	IF4A3_HUMAN
IF16_HUMAN	IF2B3_HUMAN	RL17_HUMAN	HNRPF_HUMAN	RBM5_HUMAN	RS23_HUMAN
PERQ2_HUMAN	ARPC3_HUMAN	SET1A_HUMAN	CHERP_HUMAN	FXR1_HUMAN	HDAC1_HUMAN
EIF3D_HUMAN	BA2L2_HUMAN	RBBP6_HUMAN	EBP2_HUMAN	ZCHC3_HUMAN	CDC5L_HUMAN
MOV10_HUMAN	SUZ12_HUMAN	EP400_HUMAN	CDK9_HUMAN	SMRC1_HUMAN	ACL6A_HUMAN
DECR_HUMAN	HNRL1_HUMAN	RCC2_HUMAN	RALYL_HUMAN	IF4A3_HUMAN	RBP2_HUMAN
HNRPL_HUMAN	MRE11_HUMAN	KIF22_HUMAN	DOCK2_HUMAN	RL21_HUMAN	PRP4_HUMAN
RL10_HUMAN	DDX23_HUMAN	KI20B_HUMAN	NCOR1_HUMAN	PRP19_HUMAN	LAR4B_HUMAN
LC7L2_HUMAN	MBB1A_HUMAN	KI67_HUMAN	ARI1A_HUMAN	PR40A_HUMAN	RFA1_HUMAN
SK2L2_HUMAN	MTA1_HUMAN	FNBP4_HUMAN	RS27_HUMAN	PDIA6_HUMAN	SAFB1_HUMAN
HDAC1_HUMAN	SFRS5_HUMAN	LYRIC_HUMAN	EIF3C_HUMAN	ATX2L_HUMAN	SNUT2_HUMAN
CDK12_HUMAN	ACINU_HUMAN	SMRC2_HUMAN	CHD4_HUMAN	RBM25_HUMAN	XRN2_HUMAN
RL13A_HUMAN	P66A_HUMAN	CTRO_HUMAN	PHF5A_HUMAN	ATD3A_HUMAN	CNOT1_HUMAN
ACL6A_HUMAN	PEBB_HUMAN	WDR33_HUMAN	PIHD1_HUMAN	GTF2I_HUMAN	RBM10_HUMAN
RAD50_HUMAN	CV028_HUMAN	RLA1_HUMAN	EIF3B_HUMAN	RPN1_HUMAN	ZFR_HUMAN
NAT10_HUMAN	ZN592_HUMAN	MTA1_HUMAN	AGO1_HUMAN	SMHD1_HUMAN	MED12_HUMAN
RU17_HUMAN	ACL6A_HUMAN	SFRS5_HUMAN	EIF3G_HUMAN	RS7_HUMAN	RL9_HUMAN
RBM25_HUMAN	GNAI2_HUMAN	ARPC3_HUMAN	ARPC4_HUMAN	ARPC4_HUMAN	MATR3_HUMAN
SNUT2_HUMAN	NO66_HUMAN	BLM_HUMAN	CPSF2_HUMAN	HDAC1_HUMAN	SHKB1_HUMAN
PCM1_HUMAN	RL15_HUMAN	PSPC1_HUMAN	RL31_HUMAN	MATR3_HUMAN	MBB1A_HUMAN
MINT_HUMAN	PCF11_HUMAN	PHF8_HUMAN	RS24_HUMAN	RALYL_HUMAN	MPCP_HUMAN
TRAM1_HUMAN	RL23_HUMAN	RUNX3_HUMAN	NOG1_HUMAN	RUVB1_HUMAN	UBP2L_HUMAN
IKZF1_HUMAN	KIF2C_HUMAN	TCOF_HUMAN	RS7_HUMAN	H1X_HUMAN	UBF1_HUMAN
F120A_HUMAN	MYH10_HUMAN	TF3C3_HUMAN	KIF2C_HUMAN	PERI_HUMAN	UTP18_HUMAN
RFA1_HUMAN	RL35_HUMAN	TRA2A_HUMAN	PERI_HUMAN	SK2L2_HUMAN	SF3B4_HUMAN
RUVB2_HUMAN	PLAK_HUMAN	RL31_HUMAN	PCF11_HUMAN	MTA1_HUMAN	H2AY_HUMAN
PA2G4_HUMAN	TF3C3_HUMAN	FUBP3_HUMAN	TADBP_HUMAN	SF3B4_HUMAN	TR150_HUMAN
MED14_HUMAN	EF1D_HUMAN	ANKH1_HUMAN	HNRPL_HUMAN	NOP58_HUMAN	SRP14_HUMAN
EIF3B_HUMAN	VDAC2_HUMAN	NUMA1_HUMAN	IF4A3_HUMAN	PEBB_HUMAN	RBM4_HUMAN
ARI1A_HUMAN	RLA1_HUMAN	SMCA2_HUMAN	RS23_HUMAN	UBF1_HUMAN	CG050_HUMAN
SF3A3_HUMAN	UBN2_HUMAN	TRIPC_HUMAN	HDAC1_HUMAN	SMCA4_HUMAN	SRRM2_HUMAN
RNPS1_HUMAN	MED14_HUMAN	LUC7L_HUMAN	ZC3H4_HUMAN	ARHG8_HUMAN	BCLF1_HUMAN
CV028_HUMAN	RS5_HUMAN	BA2L2_HUMAN	CDC5L_HUMAN	LSP1_HUMAN	NUMA1_HUMAN
PQBP1_HUMAN	1433Z_HUMAN	NEUA_HUMAN	ACL6A_HUMAN	MPCP_HUMAN	PR40A_HUMAN
WBP11_HUMAN	UBAP2_HUMAN	FLOT1_HUMAN	SF3A3_HUMAN	THOC2_HUMAN	PDIA6_HUMAN
PRP4_HUMAN	TPX2_HUMAN	CPSF3_HUMAN	RBP2_HUMAN	RS23_HUMAN	MED1_HUMAN
KIF23_HUMAN	IKZF1_HUMAN	SMCA5_HUMAN	CQ085_HUMAN	RRBP1_HUMAN	SFRS2_HUMAN
CLP1_HUMAN	SMCA2_HUMAN	VDAC2_HUMAN	RBBP6_HUMAN	AZI1_HUMAN	NADAP_HUMAN
THOC1_HUMAN	YLPM1_HUMAN	ANR17_HUMAN	PRP4_HUMAN	SERPH_HUMAN	RPN1_HUMAN
ZC11A_HUMAN	TCOF_HUMAN	RS15_HUMAN	THOC1_HUMAN	PRDX1_HUMAN	RRBP1_HUMAN
RL35_HUMAN	ENOA_HUMAN	NO66_HUMAN	LAR4B_HUMAN	IF16_HUMAN	PRKRA_HUMAN
RL17_HUMAN	SMRC2_HUMAN	XYLT1_HUMAN	RFA1_HUMAN	SHKB1_HUMAN	RCC2_HUMAN
SET1A_HUMAN	PRDX1_HUMAN	TIF1B_HUMAN	DYHC1_HUMAN	TPX2_HUMAN	CALM_HUMAN
RBBP6_HUMAN	P66B_HUMAN	P80C_HUMAN	SAFB1_HUMAN	RS24_HUMAN	CV028_HUMAN
EP400_HUMAN	SET1A_HUMAN	NCOR1_HUMAN	SNUT2_HUMAN	EIF3A_HUMAN	DPM1_HUMAN
RCC2_HUMAN	RFA1_HUMAN	CHD8_HUMAN	XRN2_HUMAN	EMD_HUMAN	FA98A_HUMAN
KIF22_HUMAN	THOC2_HUMAN	PABP2_HUMAN	KHDR1_HUMAN	DHX36_HUMAN	LC7L2_HUMAN
KI20B_HUMAN	G3BP2_HUMAN	CDK9_HUMAN	PRP16_HUMAN	RL10L_HUMAN	E2AK2_HUMAN
KI67_HUMAN	EHMT1_HUMAN	SMC3_HUMAN	CNOT1_HUMAN	PAIRB_HUMAN	PB1_HUMAN
FNBP4_HUMAN	LUC7L_HUMAN	DKC1_HUMAN	RBM10_HUMAN	PFD2_HUMAN	KIF14_HUMAN
LYRIC_HUMAN	CHD8_HUMAN	ZC3H1_HUMAN	ZFR_HUMAN	G3BP2_HUMAN	ABLM3_HUMAN
SMRC2_HUMAN	ZC3H1_HUMAN	HAX1_HUMAN	MED12_HUMAN	BAZ1A_HUMAN	ARHG8_HUMAN
CTRO_HUMAN	NEUA_HUMAN	TPX2_HUMAN	RL9_HUMAN	MYH4_HUMAN	RLA1_HUMAN
NIPS1_HUMAN	SLAI2_HUMAN	EIF3H_HUMAN	MATR3_HUMAN	UBAP2_HUMAN	FLII_HUMAN
WDR33_HUMAN	BC11B_HUMAN	EF1G_HUMAN	CENPF_HUMAN	PPIA_HUMAN	DHX36_HUMAN
RSMN_HUMAN	KIF14_HUMAN	CHD3_HUMAN	RS27L_HUMAN	RN213_HUMAN	CO6A1_HUMAN
RLA1_HUMAN	ANR17_HUMAN	RS27_HUMAN	CDC73_HUMAN	PA2G4_HUMAN	RNPS1_HUMAN
MTA1_HUMAN	H2A1H_HUMAN	EBP2_HUMAN	SHKB1_HUMAN	UN84B_HUMAN	SMRC1_HUMAN
SFRS5_HUMAN	DREB_HUMAN	ZN592_HUMAN	YLPM1_HUMAN	MOV10_HUMAN	NOLC1_HUMAN
ARPC3_HUMAN	PNMA5_HUMAN	OGT1_HUMAN	ZN638_HUMAN	ENOA_HUMAN	RBM5_HUMAN
BLM_HUMAN	EIF3B_HUMAN	C1QBP_HUMAN	MBB1A_HUMAN	RL31_HUMAN	TMM33_HUMAN
SPF45_HUMAN	RL31_HUMAN	GNAI2_HUMAN	KAP2_HUMAN	LR16C_HUMAN	FUBP2_HUMAN
PSPC1_HUMAN	PRDX4_HUMAN	STAU1_HUMAN	MPCP_HUMAN	NHP2_HUMAN	MOV10_HUMAN
SMC1A_HUMAN	CNOT3_HUMAN	CDC2_HUMAN	RL29_HUMAN	UTP18_HUMAN	TFR1_HUMAN
PHF8_HUMAN	C1QBP_HUMAN	PEBB_HUMAN	UBP2L_HUMAN	MACF1_HUMAN	MRM1_HUMAN
H2A1J_HUMAN	PA2G4_HUMAN	HP1B3_HUMAN	UBF1_HUMAN	EIF3B_HUMAN	GLYR1_HUMAN
RRP1B_HUMAN	SF3B4_HUMAN	SPTA2_HUMAN	R39L5_HUMAN	RPOM_HUMAN	ASCC2_HUMAN
RUNX3_HUMAN	PDIP3_HUMAN	MED12_HUMAN	UTP18_HUMAN	MED1_HUMAN	RAC2_HUMAN
TCOF_HUMAN	NSUN2_HUMAN	RS12_HUMAN	EIF3E_HUMAN	RNPS1_HUMAN	MYH4_HUMAN
SMRC1_HUMAN	CUX1_HUMAN	CD2AP_HUMAN	SF3B4_HUMAN	DOCK8_HUMAN	UBAP2_HUMAN
SFR14_HUMAN	MSH6_HUMAN	RL34_HUMAN	H2AY_HUMAN	MRM1_HUMAN	TCOF_HUMAN
TF3C3_HUMAN	POGZ_HUMAN	CUX1_HUMAN	KV101_HUMAN	TCOF_HUMAN	RL10L_HUMAN
TRA2A_HUMAN	ATPO_HUMAN		TR150_HUMAN	SNUT2_HUMAN	ZCHC3_HUMAN

RL31_HUMAN	NOP2_HUMAN	SMC6_HUMAN	PRP17_HUMAN	RLA1_HUMAN	KV305_HUMAN
FUBP3_HUMAN	TIM50_HUMAN	ZN207_HUMAN	SRP14_HUMAN	ARP5L_HUMAN	HNRL2_HUMAN
ANKH1_HUMAN	DPM1_HUMAN	CN37_HUMAN	RS4Y2_HUMAN	CBX3_HUMAN	LAMB4_HUMAN
NUMA1_HUMAN	KI67_HUMAN	MSH6_HUMAN	RBM4_HUMAN	RRP1B_HUMAN	B3GT6_HUMAN
SMCA2_HUMAN	DOCK8_HUMAN	KHDR1_HUMAN	CG050_HUMAN	GAR1_HUMAN	DDX42_HUMAN
SNRPA_HUMAN	ELAV2_HUMAN	SF3B4_HUMAN	SRRM2_HUMAN	TCPA_HUMAN	TCPG_HUMAN
TRIPC_HUMAN	RL35A_HUMAN	TS101_HUMAN	BCLF1_HUMAN	NUMA1_HUMAN	
LUC7L_HUMAN	NACA2_HUMAN	FLOT2_HUMAN	CATW_HUMAN	NADAP_HUMAN	
BA2L2_HUMAN	H33_HUMAN	SF13A_HUMAN	DYL2_HUMAN	COPB_HUMAN	
NEUA_HUMAN	SMRC1_HUMAN	LEO1_HUMAN	EIF3K_HUMAN	SAFB1_HUMAN	
CPSF2_HUMAN	EIF3L_HUMAN	KU86_HUMAN	XRN1_HUMAN	CALM_HUMAN	
IF2P_HUMAN	RL29_HUMAN	MYO1C_HUMAN	BIRC6_HUMAN	BYST_HUMAN	
YTHD2_HUMAN	PCID2_HUMAN	HNRL2_HUMAN	CD2AP_HUMAN	TXND5_HUMAN	
CAF1B_HUMAN	PHF8_HUMAN	CHD1L_HUMAN	NUMA1_HUMAN	SON_HUMAN	
NEB2_HUMAN	AIFM1_HUMAN	UBAP2_HUMAN	PR40A_HUMAN	SRBD1_HUMAN	
FLOT1_HUMAN	KIF2A_HUMAN	RS5_HUMAN	KV102_HUMAN	RBM4_HUMAN	
CPSF3_HUMAN	HNRL2_HUMAN	NCBP2_HUMAN	PDIA6_HUMAN	LMO7_HUMAN	
EIF3G_HUMAN	HMMR_HUMAN	H33_HUMAN	MINT_HUMAN	ASCC2_HUMAN	
SMCA5_HUMAN	SRRM1_HUMAN	NUSAP_HUMAN	IMA7_HUMAN	PIHD1_HUMAN	
NACA_HUMAN	COPB_HUMAN	CC124_HUMAN	SYF1_HUMAN	HLTF_HUMAN	
THOC4_HUMAN	RBBP6_HUMAN	GPTC4_HUMAN	NDUA4_HUMAN	DHX8_HUMAN	
VDAC2_HUMAN	SSRG_HUMAN	EST1A_HUMAN	FL2D_HUMAN	C1TC_HUMAN	
ANR17_HUMAN	PRP19_HUMAN	CPSF7_HUMAN	NDUS7_HUMAN	KIF2A_HUMAN	
RS15_HUMAN	DDX1_HUMAN	LGAT1_HUMAN	MTMR5_HUMAN	PB1_HUMAN	
EIF3K_HUMAN	ZN638_HUMAN	DEK_HUMAN	RUXF_HUMAN	BZW2_HUMAN	
NO66_HUMAN	ROA0_HUMAN	FAS_HUMAN	SC61B_HUMAN	OLA1_HUMAN	
TR150_HUMAN	FLOT1_HUMAN	SSRG_HUMAN	CDK12_HUMAN	ARI1A_HUMAN	
XYLT1_HUMAN	CC124_HUMAN	PDC6I_HUMAN	CLP1_HUMAN	ABLM3_HUMAN	
TIF1B_HUMAN	STAU1_HUMAN	TIM50_HUMAN	ABC3F_HUMAN	PUR6_HUMAN	
P80C_HUMAN	KHDR1_HUMAN	DOCK8_HUMAN	MED1_HUMAN	SAFB2_HUMAN	
NCOR1_HUMAN	DNJB6_HUMAN	RL29_HUMAN	DDX27_HUMAN	SUZ12_HUMAN	
RU2A_HUMAN	TRIPC_HUMAN	PFD2_HUMAN	SFRS2_HUMAN	KIF2B_HUMAN	
RRP5_HUMAN	EIF3F_HUMAN	NSUN2_HUMAN	NADAP_HUMAN	ADNP_HUMAN	
CHD8_HUMAN	TRRAP_HUMAN	SNX18_HUMAN	RPN1_HUMAN	NCBP1_HUMAN	
PABP2_HUMAN	NUSAP_HUMAN	TCPG_HUMAN	RRBP1_HUMAN	CDK5_HUMAN	
ABT1_HUMAN	CHD1L_HUMAN	SPTB2_HUMAN	CTRO_HUMAN	6PGD_HUMAN	
RS21_HUMAN	RS27_HUMAN	SIN3A_HUMAN	PRKRA_HUMAN	ARHG2_HUMAN	
CDK9_HUMAN	CABIN_HUMAN	CNOT2_HUMAN	RCC2_HUMAN	FUBP2_HUMAN	
CA077_HUMAN	GPTC4_HUMAN	SNF5_HUMAN	EDF1_HUMAN	IKZF2_HUMAN	
SMC3_HUMAN	TCPQ_HUMAN	C170L_HUMAN	RBM8A_HUMAN	IF6_HUMAN	
ESF1_HUMAN	MYO1C_HUMAN	KIF2A_HUMAN	TEX10_HUMAN	CTND1_HUMAN	
DKC1_HUMAN	TCPG_HUMAN	SRP14_HUMAN	RAD21_HUMAN	GRWD1_HUMAN	
PLK1_HUMAN	SNUT2_HUMAN	SC61B_HUMAN	MCM3_HUMAN	RUVB2_HUMAN	
ZC3H1_HUMAN	EST1A_HUMAN	ATPO_HUMAN	CALM_HUMAN	DDX42_HUMAN	
ZC3HE_HUMAN	SAFB2_HUMAN	NMNA1_HUMAN	CV028_HUMAN	PRP4_HUMAN	
HAX1_HUMAN	RCC2_HUMAN	DNJB6_HUMAN	PROF1_HUMAN	PLEC1_HUMAN	
RBM26_HUMAN	CAPZB_HUMAN	PR38A_HUMAN	SLIRP_HUMAN	HELZ_HUMAN	
PPIL1_HUMAN	AT1A3_HUMAN	MED20_HUMAN	STAU1_HUMAN	PININ_HUMAN	
RA1L3_HUMAN	RPB1_HUMAN	PDLI2_HUMAN	SPF27_HUMAN	GOGA2_HUMAN	
EIF3I_HUMAN	IF2P_HUMAN	LARP4_HUMAN	RBMX2_HUMAN	EFTU_HUMAN	
TDIF1_HUMAN	RSMB_HUMAN	UTP18_HUMAN	ARPC3_HUMAN	AP2B1_HUMAN	
TPX2_HUMAN	LARP4_HUMAN	NHP2_HUMAN	DPM1_HUMAN	DKC1_HUMAN	
EIF3H_HUMAN	NIPBL_HUMAN	BAG2_HUMAN	FA98A_HUMAN	DPM1_HUMAN	
EF1G_HUMAN	ALDOA_HUMAN	AZI1_HUMAN	BAT2_HUMAN	WIZ_HUMAN	
RBM27_HUMAN	BAG2_HUMAN	TCPQ_HUMAN	EIF3H_HUMAN	HIRA_HUMAN	
CHD3_HUMAN	CDC2_HUMAN	NPM3_HUMAN	PPHLN_HUMAN	PRDX4_HUMAN	
RS27_HUMAN	SP16H_HUMAN	SRP54_HUMAN	2ABA_HUMAN	MBB1A_HUMAN	
EBP2_HUMAN	PSIP1_HUMAN	AGO1_HUMAN	RS29_HUMAN	SF3A1_HUMAN	
ZN592_HUMAN	RNPS1_HUMAN	RM19_HUMAN	LC7L2_HUMAN	SRRM2_HUMAN	
GNAT1_HUMAN	SRP14_HUMAN	TB182_HUMAN	PLK1_HUMAN	SRP72_HUMAN	
OGT1_HUMAN	LC7L2_HUMAN	IGLL1_HUMAN	KRIT1_HUMAN	EIF3D_HUMAN	
C1QBP_HUMAN	CNOT2_HUMAN	PSIP1_HUMAN	NELFB_HUMAN	B3GT6_HUMAN	
RALYL_HUMAN	SON_HUMAN	PPIB_HUMAN	UXT_HUMAN	KIF14_HUMAN	
GNAI2_HUMAN	HAX1_HUMAN	SRPK1_HUMAN	CGBP1_HUMAN	TMM33_HUMAN	
CARF_HUMAN	RFC4_HUMAN	NIPBL_HUMAN	TRIPC_HUMAN	UBA1_HUMAN	
STAU1_HUMAN	SMC6_HUMAN	TRRAP_HUMAN	CU070_HUMAN	TFR1_HUMAN	
CDC2_HUMAN	SMC3_HUMAN	MD12L_HUMAN	MBNL2_HUMAN	SC16A_HUMAN	
PEBB_HUMAN	BOD1L_HUMAN	SMC2_HUMAN	SFPQ_HUMAN	TOP3A_HUMAN	
HP1B3_HUMAN	RL18A_HUMAN	NOP2_HUMAN	E2AK2_HUMAN	HVCN1_HUMAN	
RBM15_HUMAN	GUF1_HUMAN	RL1D1_HUMAN	PB1_HUMAN	RAC2_HUMAN	
DSRAD_HUMAN	AZI1_HUMAN	DYH10_HUMAN	KIF14_HUMAN	UH1BL_HUMAN	
SPTA2_HUMAN	RFC2_HUMAN	WIZ_HUMAN	ABLM3_HUMAN	HCFC1_HUMAN	
MED12_HUMAN	PDLI2_HUMAN	GNL3_HUMAN	ARHG8_HUMAN	SRP14_HUMAN	
RS12_HUMAN	PTPM1_HUMAN	RL18A_HUMAN	RLA1_HUMAN	RL1D1_HUMAN	
CD2AP_HUMAN	KIF23_HUMAN	COPB_HUMAN	LMNA_HUMAN	MED12_HUMAN	
COR1C_HUMAN	MED20_HUMAN	TRI27_HUMAN	FLII_HUMAN	RRMJ3_HUMAN	
RL34_HUMAN	C170L_HUMAN	INT1_HUMAN	ALKB5_HUMAN	SPF45_HUMAN	
CD2B2_HUMAN	PPIB_HUMAN	TCRG1_HUMAN	ABT1_HUMAN	FA98A_HUMAN	
CUX1_HUMAN	OGT1_HUMAN	SERA_HUMAN	LSM6_HUMAN	MCM4_HUMAN	
SMC6_HUMAN	CTRO_HUMAN	MYH10_HUMAN	DHX36_HUMAN	DEOC_HUMAN	
SRRT_HUMAN	SRPK2_HUMAN	MYPT1_HUMAN	EXOS9_HUMAN	RBP2_HUMAN	
SMC4_HUMAN	NMNA1_HUMAN	RBM28_HUMAN	CO6A1_HUMAN	NAL13_HUMAN	

ZN207_HUMAN LEO1_HUMAN ALDOA_HUMAN DDX31_HUMAN SFRS2_HUMAN
CN37_HUMAN MINT_HUMAN RL15_HUMAN HBB_HUMAN CV028_HUMAN
MSH6_HUMAN P80C_HUMAN RAB7A_HUMAN TBC21_HUMAN AURKB_HUMAN
KHDR1_HUMAN SF13A_HUMAN SP16H_HUMAN PCY1A_HUMAN CG050_HUMAN
SF3B4_HUMAN AGO1_HUMAN SSX7_HUMAN SRPK1_HUMAN CHD2_HUMAN
ZN326_HUMAN G3P_HUMAN SFRS8_HUMAN GMFG_HUMAN EIF3C_HUMAN
TS101_HUMAN DHX36_HUMAN PNMA5_HUMAN RNPS1_HUMAN AGO1_HUMAN
FLOT2_HUMAN FUBP3_HUMAN CALU_HUMAN SMRC1_HUMAN HERC2_HUMAN
CQ085_HUMAN DCD_HUMAN CSTF2_HUMAN NOLC1_HUMAN LAMB4_HUMAN
SLBP_HUMAN KATL2_HUMAN CATG_HUMAN RBM5_HUMAN PRKRA_HUMAN
CDC73_HUMAN RFC3_HUMAN TFAM_HUMAN TMM33_HUMAN CQ049_HUMAN
SF13A_HUMAN FLOT2_HUMAN DNJA1_HUMAN ZC3HE_HUMAN RBM34_HUMAN
THOC5_HUMAN LAR4B_HUMAN UCHL5_HUMAN PRPF3_HUMAN UBS3A_HUMAN
LEO1_HUMAN DOCK9_HUMAN DNLI3_HUMAN EDC4_HUMAN ZC11A_HUMAN
KU86_HUMAN TLN1_HUMAN RFC3_HUMAN TOE1_HUMAN GLYR1_HUMAN
MYO1C_HUMAN DYH10_HUMAN ZN644_HUMAN FUBP2_HUMAN E2AK2_HUMAN
HNRL2_HUMAN IGLL1_HUMAN ROA0_HUMAN MED21_HUMAN ANKS3_HUMAN
CHD1L_HUMAN RL1D1_HUMAN SRPK2_HUMAN ZCHC8_HUMAN LAR4B_HUMAN
UBAP2_HUMAN HDAC1_HUMAN CHRC1_HUMAN RING2_HUMAN
RS5_HUMAN UTP18_HUMAN MOV10_HUMAN KCTD5_HUMAN
NCBP2_HUMAN RBM28_HUMAN TFR1_HUMAN BLM_HUMAN
PM14_HUMAN FNBP4_HUMAN RRP15_HUMAN ATPA_HUMAN
H33_HUMAN SNF5_HUMAN TRI26_HUMAN HNRL2_HUMAN
DHX8_HUMAN TIF1B_HUMAN MRM1_HUMAN KV305_HUMAN
NUSAP_HUMAN TCRG1_HUMAN COX2_HUMAN SYTC_HUMAN
CC124_HUMAN ELYS_HUMAN KV110_HUMAN MRE11_HUMAN
PCNT_HUMAN TFAM_HUMAN DIDO1_HUMAN TSR1_HUMAN
GPTC4_HUMAN NDUC1_HUMAN BRX1_HUMAN DNJC9_HUMAN
PP1RA_HUMAN TOPB1_HUMAN GLYR1_HUMAN PDIA1_HUMAN
EST1A_HUMAN ANKH1_HUMAN ASCC2_HUMAN ZFR_HUMAN
CPSF7_HUMAN XYLT1_HUMAN RAD50_HUMAN TRY2_HUMAN
LGAT1_HUMAN THOC1_HUMAN FUBP3_HUMAN RCC2_HUMAN
DEK_HUMAN DEK_HUMAN RAC2_HUMAN ATS14_HUMAN
ARHG2_HUMAN ZN207_HUMAN MYH4_HUMAN RP1_HUMAN
ECHA_HUMAN GDIR1_HUMAN SF3B5_HUMAN MED10_HUMAN
SON_HUMAN PDC6I_HUMAN UBAP2_HUMAN A2ML1_HUMAN
LSP1_HUMAN SPTB2_HUMAN TCOF_HUMAN PHB2_HUMAN
ARFG2_HUMAN TRI27_HUMAN NUP50_HUMAN BCAM_HUMAN
UN84B_HUMAN DIDO1_HUMAN SFR11_HUMAN POGZ_HUMAN
FAS_HUMAN CATG_HUMAN RL10L_HUMAN PPCEL_HUMAN
SSRG_HUMAN RS29_HUMAN RO52_HUMAN YB035_HUMAN
PDC6I_HUMAN SRP54_HUMAN SCC4_HUMAN RAVR1_HUMAN
TIM50_HUMAN RAN_HUMAN TUT4_HUMAN TYDP1_HUMAN
PCY1A_HUMAN RL34_HUMAN UBS3B_HUMAN SMAL1_HUMAN
UBS3A_HUMAN RAB10_HUMAN ACINU_HUMAN LC7L2_HUMAN
DOCK8_HUMAN DDX31_HUMAN LRC59_HUMAN CO6A1_HUMAN
RL29_HUMAN TB182_HUMAN ZCHC3_HUMAN RL36A_HUMAN
REQU_HUMAN KV305_HUMAN SFRS9_HUMAN
RIF1_HUMAN SC61B_HUMAN KV305_HUMAN
SFR12_HUMAN RM19_HUMAN RS26L_HUMAN
PFD2_HUMAN INT1_HUMAN HNRL2_HUMAN
COPD_HUMAN IF6_HUMAN CL073_HUMAN
LSM12_HUMAN PININ_HUMAN MED20_HUMAN
NSUN2_HUMAN AKAP8_HUMAN VPS41_HUMAN
NIPS2_HUMAN RS15_HUMAN DIP2A_HUMAN
U2AF2_HUMAN SNW1_HUMAN SYNE2_HUMAN
SNX18_HUMAN ZC11A_HUMAN LAMB4_HUMAN
CSTF3_HUMAN CPSF7_HUMAN GSCR2_HUMAN
EPN4_HUMAN Z3H7A_HUMAN PR38B_HUMAN
RL35A_HUMAN NCOR1_HUMAN ECHA_HUMAN
IF4G3_HUMAN DNJA1_HUMAN DHX35_HUMAN
TCPG_HUMAN UAP56_HUMAN MDC1_HUMAN
IMMT_HUMAN MD12L_HUMAN NOL7_HUMAN
SPTB2_HUMAN RCC1_HUMAN NSUN2_HUMAN
SIN3A_HUMAN TS101_HUMAN NDUB4_HUMAN
RUNX2_HUMAN WDR33_HUMAN CN166_HUMAN
MED17_HUMAN CD046_HUMAN MCTP2_HUMAN
ZN318_HUMAN TRA2A_HUMAN B3GT6_HUMAN
VRK2_HUMAN CSTF2_HUMAN NIPBL_HUMAN
CNOT2_HUMAN SERA_HUMAN NOC4L_HUMAN
EXOS9_HUMAN UCHL5_HUMAN ZN790_HUMAN
ECHB_HUMAN NHP2_HUMAN MED17_HUMAN
RBM8A_HUMAN QCR2_HUMAN LSM7_HUMAN
MDC1_HUMAN XRCC1_HUMAN PYR1_HUMAN
SNF5_HUMAN M10L1_HUMAN DDX42_HUMAN
RU2B_HUMAN SAPS1_HUMAN K0020_HUMAN
CD3E_HUMAN RAB7A_HUMAN TCPG_HUMAN
CSK21_HUMAN P5CR1_HUMAN FGGY_HUMAN
PP1B_HUMAN PPIL1_HUMAN
C170L_HUMAN SMRD2_HUMAN
KIF2A_HUMAN ATF5_HUMAN
NCOA5_HUMAN DJC15_HUMAN

SRP14_HUMAN
SC61B_HUMAN
FADS2_HUMAN
1433T_HUMAN
ATPO_HUMAN
ARPC4_HUMAN
NMNA1_HUMAN
SMRD2_HUMAN
PR38B_HUMAN
MD13L_HUMAN
SLTM_HUMAN
PRP4B_HUMAN
OPA1_HUMAN
DNJB6_HUMAN
PR38A_HUMAN
MED20_HUMAN
ISY1_HUMAN
PDLI2_HUMAN
RSBNL_HUMAN
LARP4_HUMAN
WAPL_HUMAN
YTHD1_HUMAN
CLAP2_HUMAN
PRPF3_HUMAN
UTP18_HUMAN
NHP2_HUMAN
ITM2A_HUMAN
SDC10_HUMAN
STAU2_HUMAN
BIRC6_HUMAN
PHF5A_HUMAN
UIF_HUMAN
RPN2_HUMAN
LC7L3_HUMAN
BAG2_HUMAN
AZI1_HUMAN
TCPQ_HUMAN
NPM3_HUMAN
STAT1_HUMAN
ABCF2_HUMAN
GAPR1_HUMAN
NMT1_HUMAN
SRBD1_HUMAN
DNJA2_HUMAN
SP130_HUMAN
MAGB5_HUMAN
SRP54_HUMAN
KIF4A_HUMAN
AGO1_HUMAN
TAF6L_HUMAN
RM19_HUMAN
ANR27_HUMAN
TB182_HUMAN
CTF18_HUMAN
SHKB1_HUMAN
DIMT1_HUMAN
MMTA2_HUMAN
RFC5_HUMAN
RPOM_HUMAN
IGLL1_HUMAN
PSIP1_HUMAN
PPIB_HUMAN
MED1_HUMAN
RL28_HUMAN
SRPK1_HUMAN
IMP3_HUMAN
NIPBL_HUMAN
IMB1_HUMAN
RCOR1_HUMAN
NOP14_HUMAN
TRRAP_HUMAN
R3HD1_HUMAN
TFCP2_HUMAN
MD12L_HUMAN
COPG_HUMAN
SMC2_HUMAN
IF4E_HUMAN
NOP2_HUMAN
RL1D1_HUMAN
DDX55_HUMAN
DYH10_HUMAN
TFP11_HUMAN

NP1L1_HUMAN
ANR17_HUMAN
DAXX_HUMAN
TADA1_HUMAN
RAD54_HUMAN
ZC3HD_HUMAN
RBBP5_HUMAN
ZN207_HUMAN
AP2A2_HUMAN
WDR82_HUMAN

220

LYAR_HUMAN
AQR_HUMAN
WIZ_HUMAN
TF3C5_HUMAN
PDS5A_HUMAN
GNL3_HUMAN
RAC2_HUMAN
SMN_HUMAN
RSRC1_HUMAN
RL18A_HUMAN
SSRA_HUMAN
URP2_HUMAN
NELFB_HUMAN
ABCE1_HUMAN
SRBS1_HUMAN
COPB_HUMAN
TRI27_HUMAN
E2AK2_HUMAN
CU070_HUMAN
STRAP_HUMAN
INT1_HUMAN
PRP31_HUMAN
RBM33_HUMAN
ANLN_HUMAN
AHSP_HUMAN
SRP19_HUMAN
DHX35_HUMAN
KRR1_HUMAN
TCRG1_HUMAN
GRWD1_HUMAN
WDR5_HUMAN
G3P_HUMAN
AP3B1_HUMAN
H1X_HUMAN
DNJC7_HUMAN
GCFC_HUMAN
SMHD1_HUMAN
DDX54_HUMAN
SRP09_HUMAN
DRG1_HUMAN
TSR1_HUMAN
PCBP2_HUMAN
ORC1_HUMAN
SERA_HUMAN
GOGA2_HUMAN
RMTL1_HUMAN
MYH10_HUMAN
ASH2L_HUMAN
MYPT1_HUMAN
RBM28_HUMAN
SYF1_HUMAN
RT07_HUMAN
DGC6L_HUMAN
PKCB1_HUMAN
NOL6_HUMAN
ALDOA_HUMAN
RL15_HUMAN
TCPZ_HUMAN
POP1_HUMAN
LSM3_HUMAN
DNJA3_HUMAN
SSBP_HUMAN
TBL2_HUMAN
RAB7A_HUMAN
SP16H_HUMAN
KC1D_HUMAN
SSX7_HUMAN
OSBL8_HUMAN
NOM1_HUMAN
SFRS8_HUMAN
ARHG8_HUMAN
BA2L1_HUMAN
LENG8_HUMAN
PNMA5_HUMAN
COPB2_HUMAN
STAT2_HUMAN
RUXF_HUMAN
DDX20_HUMAN
ICLN_HUMAN
PURA_HUMAN
ALKB5_HUMAN
CT043_HUMAN

CALU_HUMAN
THOC6_HUMAN
PUM1_HUMAN
GDE_HUMAN
MSI2H_HUMAN
CRNL1_HUMAN
CSTF2_HUMAN
CCYL2_HUMAN
CC072_HUMAN
NCLN_HUMAN
PHF3_HUMAN
CATG_HUMAN
SNTB2_HUMAN
MED10_HUMAN
DP13B_HUMAN

HSB2			CCRFCEM		
GFI1 Mass spectrometry	**Kme1 Mass spectrometry**	**Common between GFI1 and Kme1**	**GFI1 Mass spectrometry**	**Kme1 Mass spectrometry**	**Common between GFI1 and Kme1**
KI67_HUMAN	AHNK_HUMAN	KI67_HUMAN	MYH9_HUMAN	ACACA_HUMAN	MYH9_HUMAN
PABP1_HUMAN	ACACA_HUMAN	PABP1_HUMAN	KI67_HUMAN	CHD4_HUMAN	KI67_HUMAN
TBA1C_HUMAN	DYHC1_HUMAN	TBA1C_HUMAN	MYH10_HUMAN	MYH9_HUMAN	MYH10_HUMAN
PABP4_HUMAN	PRKDC_HUMAN	PABP4_HUMAN	TOP2A_HUMAN	SMCA5_HUMAN	TOP2A_HUMAN
PRKDC_HUMAN	DDX21_HUMAN	PRKDC_HUMAN	U520_HUMAN	AHNK_HUMAN	U520_HUMAN
RS8_HUMAN	HS90B_HUMAN	RS8_HUMAN	TOP2B_HUMAN	TOP2A_HUMAN	TOP2B_HUMAN
DDX17_HUMAN	HSP71_HUMAN	DDX17_HUMAN	PRP8_HUMAN	KU70_HUMAN	PRP8_HUMAN
AHNK_HUMAN	TBA1C_HUMAN	AHNK_HUMAN	DHX9_HUMAN	U520_HUMAN	DHX9_HUMAN
RL7_HUMAN	TBA1A_HUMAN	RL7_HUMAN	DDX21_HUMAN	MTA2_HUMAN	DDX21_HUMAN
HNRPC_HUMAN	KU70_HUMAN	HNRPC_HUMAN	DDX17_HUMAN	HS90A_HUMAN	DDX17_HUMAN
HNRCL_HUMAN	HS90A_HUMAN	MYO1G_HUMAN	PABP1_HUMAN	HSP71_HUMAN	PABP1_HUMAN
MYO1G_HUMAN	MYO1G_HUMAN	PLEC1_HUMAN	MDC1_HUMAN	DDX17_HUMAN	U5S1_HUMAN
PLEC1_HUMAN	PABP1_HUMAN	RS13_HUMAN	U5S1_HUMAN	TBB5_HUMAN	NAT10_HUMAN
RS13_HUMAN	TBB3_HUMAN	RS3_HUMAN	SPTB2_HUMAN	TOP2B_HUMAN	CHD4_HUMAN
RS3_HUMAN	KU86_HUMAN	RLA0_HUMAN	NAT10_HUMAN	DDX21_HUMAN	HNRPR_HUMAN
RLA0_HUMAN	TOP2B_HUMAN	ILF2_HUMAN	SPTA2_HUMAN	BAZ1A_HUMAN	TOP1_HUMAN
ILF2_HUMAN	TOP2A_HUMAN	RL7A_HUMAN	CHD4_HUMAN	KU86_HUMAN	TBB5_HUMAN
RL7A_HUMAN	PABP4_HUMAN	RBM14_HUMAN	HNRPR_HUMAN	SP16H_HUMAN	MYO1G_HUMAN
RBM14_HUMAN	SMCA5_HUMAN	U520_HUMAN	TOP1_HUMAN	RBBP4_HUMAN	LAP2A_HUMAN
ZN638_HUMAN	DDX17_HUMAN	H15_HUMAN	TBB5_HUMAN	SMCA4_HUMAN	KU70_HUMAN
U520_HUMAN	TCPG_HUMAN	HSP71_HUMAN	MYO1G_HUMAN	TBA1B_HUMAN	HNRPK_HUMAN
H15_HUMAN	RL7_HUMAN	RS16_HUMAN	LAP2A_HUMAN	HNRPK_HUMAN	SF3B1_HUMAN
H12_HUMAN	RFC1_HUMAN	NUCL_HUMAN	KU70_HUMAN	BC11B_HUMAN	IF4G1_HUMAN
NBN_HUMAN	ECHA_HUMAN	RS3A_HUMAN	ACTG_HUMAN	SSRP1_HUMAN	DDX5_HUMAN
HSP71_HUMAN	RL18_HUMAN	RL18_HUMAN	DDX3X_HUMAN	MTA1_HUMAN	SR140_HUMAN
RS16_HUMAN	RS8_HUMAN	DDX21_HUMAN	HNRPK_HUMAN	LAP2A_HUMAN	LAP2B_HUMAN
MAP1A_HUMAN	GRP78_HUMAN	RS18_HUMAN	SF3B1_HUMAN	U5S1_HUMAN	ILF2_HUMAN
NUCL_HUMAN	RS13_HUMAN	RL6_HUMAN	IF4G1_HUMAN	DDX5_HUMAN	SF3B3_HUMAN
RS3A_HUMAN	COPA_HUMAN	RS4X_HUMAN	DDX5_HUMAN	KIF4A_HUMAN	CDC5L_HUMAN
RL18_HUMAN	RL7A_HUMAN	RALY_HUMAN	SR140_HUMAN	SF3B3_HUMAN	TBA1B_HUMAN
DDX21_HUMAN	DHX9_HUMAN	RS2_HUMAN	LAP2B_HUMAN	RBBP7_HUMAN	ILF3_HUMAN
RS18_HUMAN	EF1G_HUMAN	RL15_HUMAN	PABP4_HUMAN	PABP1_HUMAN	HNRPM_HUMAN
MRE11_HUMAN	SYLC_HUMAN	RL13A_HUMAN	ILF2_HUMAN	PRP8_HUMAN	PRP6_HUMAN
RL6_HUMAN	TOP1_HUMAN	RL27A_HUMAN	SF3B3_HUMAN	EF1G_HUMAN	SNUT1_HUMAN
H2B1H_HUMAN	ENPL_HUMAN	DHX9_HUMAN	CDC5L_HUMAN	LAP2B_HUMAN	PRP19_HUMAN
RS4X_HUMAN	RLA0_HUMAN	HNRPK_HUMAN	TBA1B_HUMAN	TOP1_HUMAN	MATR3_HUMAN
RALY_HUMAN	TCPQ_HUMAN	PRP8_HUMAN	ILF3_HUMAN	CNOT1_HUMAN	RL4_HUMAN
RS2_HUMAN	RL14_HUMAN	RL14_HUMAN	HNRPM_HUMAN	EFTU_HUMAN	PTBP1_HUMAN
EIF3A_HUMAN	IKZF1_HUMAN	HNRPU_HUMAN	PRP6_HUMAN	K2C3_HUMAN	RSSA_HUMAN
RLA2_HUMAN	EF2_HUMAN	TOP2A_HUMAN	SNUT1_HUMAN	SATB1_HUMAN	SMCA4_HUMAN
RL15_HUMAN	RS3_HUMAN	YBOX1_HUMAN	FLNA_HUMAN	RSSA_HUMAN	DDX50_HUMAN
RL13A_HUMAN	UBA1_HUMAN	HNRPM_HUMAN	PRP19_HUMAN	DHX15_HUMAN	DHX30_HUMAN
RL27A_HUMAN	DNJC9_HUMAN	KU70_HUMAN	HNRPQ_HUMAN	POGZ_HUMAN	EIF3A_HUMAN
DHX9_HUMAN	PARP1_HUMAN	DNJC9_HUMAN	MATR3_HUMAN	UHRF1_HUMAN	GBLP_HUMAN
PRP16_HUMAN	RS18_HUMAN	ADT2_HUMAN	RL4_HUMAN	ILF2_HUMAN	RS3_HUMAN
BIRC6_HUMAN	YBOX1_HUMAN	GRP78_HUMAN	PTBP1_HUMAN	K2C6C_HUMAN	SF3B2_HUMAN
HNRPK_HUMAN	RS3A_HUMAN	TOP2B_HUMAN	RSSA_HUMAN	IKZF1_HUMAN	DHX15_HUMAN
PRP8_HUMAN	CDC2_HUMAN	CAPR1_HUMAN	SMCA4_HUMAN	RS8_HUMAN	RBM14_HUMAN
RL14_HUMAN	RL13A_HUMAN	TOP1_HUMAN	DDX50_HUMAN	SNUT1_HUMAN	RL7_HUMAN
DDX3X_HUMAN	HM20A_HUMAN	RS14_HUMAN	EIF3L_HUMAN	SR140_HUMAN	NOP56_HUMAN
HNRPU_HUMAN	KI67_HUMAN	RL8_HUMAN	DHX30_HUMAN	BUB3_HUMAN	UN84B_HUMAN
TFG_HUMAN	RS2_HUMAN	CDC2_HUMAN	EIF3A_HUMAN	RBM14_HUMAN	RL7A_HUMAN
TOP2A_HUMAN	NSF_HUMAN	PHB2_HUMAN	GBLP_HUMAN	SF3B1_HUMAN	HNRH1_HUMAN
YBOX1_HUMAN	LDHA_HUMAN	RS19_HUMAN	DDX18_HUMAN	HNRH1_HUMAN	MBB1A_HUMAN
HNRPM_HUMAN	RL4_HUMAN	HS90B_HUMAN	G3BP1_HUMAN	RS3_HUMAN	RL3_HUMAN
KU70_HUMAN	HNRPU_HUMAN	LDHA_HUMAN	RS3_HUMAN	PUR6_HUMAN	IF4A3_HUMAN
RS15A_HUMAN	CH60_HUMAN	RL11_HUMAN	SF3B2_HUMAN	K1C16_HUMAN	RS8_HUMAN
DNJC9_HUMAN	RL11_HUMAN	RS20_HUMAN	DHX15_HUMAN	EF2_HUMAN	RBM39_HUMAN
ADT2_HUMAN	EF1D_HUMAN	HNRPD_HUMAN	RBM14_HUMAN	TPX2_HUMAN	CRNL1_HUMAN
GRP78_HUMAN	RS4X_HUMAN	RL24_HUMAN	SRRM2_HUMAN	PRDX4_HUMAN	MTA2_HUMAN
TOP2B_HUMAN	AP3M1_HUMAN	RL5_HUMAN	RL7_HUMAN	PRP6_HUMAN	HNRPU_HUMAN
ADT1_HUMAN	RL6_HUMAN	RL22_HUMAN	NOP56_HUMAN	HCFC1_HUMAN	RL10A_HUMAN
CAPR1_HUMAN	PUR6_HUMAN	C1QBP_HUMAN	RL5_HUMAN	HNRPM_HUMAN	CHERP_HUMAN
TOP1_HUMAN	EF1B_HUMAN	1433Z_HUMAN	UN84B_HUMAN	PRP6_HUMAN	RIF1_HUMAN
RS14_HUMAN	CAZA1_HUMAN	RL26L_HUMAN	RL7A_HUMAN	RUVB2_HUMAN	RL17_HUMAN
RL8_HUMAN	HNRPG_HUMAN	KU86_HUMAN	HNRH1_HUMAN	RFIP1_HUMAN	NOP2_HUMAN
CDC2_HUMAN	HNRPD_HUMAN	TFR1_HUMAN	NUFP2_HUMAN	HDAC1_HUMAN	EDC4_HUMAN
PHB2_HUMAN	PGK1_HUMAN	ATPG_HUMAN	MBB1A_HUMAN	H2A1J_HUMAN	SFRS1_HUMAN
RS19_HUMAN	ALDOA_HUMAN	HEM6_HUMAN	RL3_HUMAN	PARP1_HUMAN	MTA1_HUMAN
UN84B_HUMAN	H33_HUMAN	PGAM5_HUMAN	IF4A3_HUMAN	KIF22_HUMAN	RU2A_HUMAN
HS90B_HUMAN	HNRPC_HUMAN	H33_HUMAN	RS8_HUMAN	HNRPC_HUMAN	RL6_HUMAN
PP1A_HUMAN	DOCK8_HUMAN	ILF3_HUMAN	RBM39_HUMAN	WIZ_HUMAN	RS16_HUMAN
LDHA_HUMAN	RFA1_HUMAN	LAP2A_HUMAN	CRNL1_HUMAN	UBA1_HUMAN	MRE11_HUMAN
RL11_HUMAN	SYIC_HUMAN	PCBP1_HUMAN	MTA2_HUMAN	RALY_HUMAN	NOP58_HUMAN
RS20_HUMAN	H15_HUMAN	RL19_HUMAN	MAP1A_HUMAN	K22O_HUMAN	ARHG2_HUMAN

HNRPD_HUMAN	CYFP2_HUMAN	MYL6_HUMAN
RL24_HUMAN	RL17_HUMAN	RL17_HUMAN
RL5_HUMAN	RL15_HUMAN	RL10A_HUMAN
PCF11_HUMAN	WIZ_HUMAN	RL23A_HUMAN
PQBP1_HUMAN	RL38_HUMAN	PRDX1_HUMAN
RL22_HUMAN	H2AY_HUMAN	HNRPR_HUMAN
C1QBP_HUMAN	ILF2_HUMAN	RL4_HUMAN
KV101_HUMAN	BUB3_HUMAN	CAZA1_HUMAN
1433Z_HUMAN	HDAC1_HUMAN	DYHC1_HUMAN
RL26L_HUMAN	RS16_HUMAN	PEBB_HUMAN
WDR82_HUMAN	HNRPK_HUMAN	ROA1_HUMAN
KU86_HUMAN	TCPD_HUMAN	RSMB_HUMAN
TFR1_HUMAN	SAHH_HUMAN	DHX30_HUMAN
STAT1_HUMAN	RL21_HUMAN	2ABA_HUMAN
ATPG_HUMAN	HEM6_HUMAN	MPCP_HUMAN
HEM6_HUMAN	C1QBP_HUMAN	RL38_HUMAN
PGAM5_HUMAN	KIF4A_HUMAN	AIFM1_HUMAN
FUS_HUMAN	RL23A_HUMAN	SSBP2_HUMAN
H33_HUMAN	COPB_HUMAN	CNN2_HUMAN
1B08_HUMAN	HNRPR_HUMAN	SHKB1_HUMAN
ILF3_HUMAN	RL26L_HUMAN	AT1A3_HUMAN
LAP2A_HUMAN	HNRPR_HUMAN	PCM1_HUMAN
EST1A_HUMAN	PEBB_HUMAN	HS90A_HUMAN
SR140_HUMAN	TADBP_HUMAN	EIF3D_HUMAN
PP1G_HUMAN	TFR1_HUMAN	CAN1_HUMAN
EIF3G_HUMAN	RL24_HUMAN	COPA_HUMAN
DDX18_HUMAN	RO52_HUMAN	ANM1_HUMAN
HNRDL_HUMAN	RL19_HUMAN	R39L5_HUMAN
PCBP1_HUMAN	HNRPM_HUMAN	SET_HUMAN
RL19_HUMAN	RN213_HUMAN	NSUN2_HUMAN
EXOC3_HUMAN	PCM1_HUMAN	RL21_HUMAN
RM15_HUMAN	RL27A_HUMAN	HDAC1_HUMAN
MYL6_HUMAN	PGAM5_HUMAN	PDLI2_HUMAN
RL17_HUMAN	PRDX1_HUMAN	PCNA_HUMAN
RL10A_HUMAN	RBM14_HUMAN	RS28_HUMAN
RL35A_HUMAN	PCNA_HUMAN	COX2_HUMAN
H2A1H_HUMAN	TCPA_HUMAN	ROA2_HUMAN
CDK9_HUMAN	CAPR1_HUMAN	DDX42_HUMAN
ABT1_HUMAN	RSSA_HUMAN	SYFA_HUMAN
UBP2L_HUMAN	IKZF2_HUMAN	RS11_HUMAN
PERI_HUMAN	CHD4_HUMAN	NDUA4_HUMAN
RL23A_HUMAN	SP16H_HUMAN	RL3_HUMAN
MCM7_HUMAN	1433Z_HUMAN	DOCK8_HUMAN
MMTA2_HUMAN	ILF3_HUMAN	CHD4_HUMAN
RS5_HUMAN	RL8_HUMAN	RN213_HUMAN
RM48_HUMAN	BC11B_HUMAN	SF3B1_HUMAN
EIF3I_HUMAN	U520_HUMAN	BUB3_HUMAN
PRDX1_HUMAN	TCPH_HUMAN	NXF3_HUMAN
HNRPR_HUMAN	MPCP_HUMAN	
WBP11_HUMAN	CNOT1_HUMAN	
RL35_HUMAN	MTA2_HUMAN	
RL4_HUMAN	URP2_HUMAN	
CAZA1_HUMAN	TIM44_HUMAN	
DYHC1_HUMAN	LDHB_HUMAN	
DIDO1_HUMAN	GCP3_HUMAN	
CLP1_HUMAN	RBM39_HUMAN	
OSBL5_HUMAN	RL5_HUMAN	
KRIT1_HUMAN	SSRP1_HUMAN	
PEBB_HUMAN	TCPZ_HUMAN	
ELAV1_HUMAN	SYQ_HUMAN	
RM46_HUMAN	EDC4_HUMAN	
SVIL_HUMAN	CNO6L_HUMAN	
NFM_HUMAN	EIF3F_HUMAN	
MTMR5_HUMAN	NUCL_HUMAN	
EIF3B_HUMAN	NDUA4_HUMAN	
CQ085_HUMAN	LAP2A_HUMAN	
ROA1_HUMAN	EIF3D_HUMAN	
SAFB1_HUMAN	NSUN2_HUMAN	
ATPA_HUMAN	CNOT2_HUMAN	
RSMB_HUMAN	2ABA_HUMAN	
RS26_HUMAN	SHKB1_HUMAN	
UIF_HUMAN	TCPB_HUMAN	
CHD1_HUMAN	AP1M1_HUMAN	
1433T_HUMAN	RL3_HUMAN	
RS17_HUMAN	RS14_HUMAN	
DHX30_HUMAN	RL10_HUMAN	
NIPS1_HUMAN	ACLY_HUMAN	
NIPS2_HUMAN	ADT2_HUMAN	
CU070_HUMAN	SMC4_HUMAN	
RL31_HUMAN	COF1_HUMAN	
EWS_HUMAN	ATPO_HUMAN	
SAFB2_HUMAN	RCN1_HUMAN	

HNRPU_HUMAN	IF4A3_HUMAN	YBOX1_HUMAN
RL10A_HUMAN	RUNX1_HUMAN	UBP2L_HUMAN
CHERP_HUMAN	ARI1A_HUMAN	DDX6_HUMAN
RENT1_HUMAN	HNRPG_HUMAN	FIP1_HUMAN
RIF1_HUMAN	KIF2C_HUMAN	CHD3_HUMAN
RL17_HUMAN	CBX3_HUMAN	H15_HUMAN
NOP2_HUMAN	DDX3Y_HUMAN	RS18_HUMAN
EDC4_HUMAN	CHD3_HUMAN	FUS_HUMAN
SFRS1_HUMAN	EF1D_HUMAN	XRN2_HUMAN
MTA1_HUMAN	RL13_HUMAN	ZCCHV_HUMAN
G3BP2_HUMAN	MYH10_HUMAN	FXR1_HUMAN
RU2A_HUMAN	P66B_HUMAN	RL30_HUMAN
RL6_HUMAN	ADT2_HUMAN	RL18_HUMAN
RS16_HUMAN	HDAC2_HUMAN	RL13_HUMAN
MRE11_HUMAN	EHMT2_HUMAN	FBRL_HUMAN
MOV10_HUMAN	ANXA2_HUMAN	KIF2C_HUMAN
NOP58_HUMAN	IKZF2_HUMAN	PR40A_HUMAN
ARHG2_HUMAN	ZN644_HUMAN	KU86_HUMAN
YBOX1_HUMAN	PTBP1_HUMAN	ACINU_HUMAN
UBP2L_HUMAN	H2AV_HUMAN	CAPR1_HUMAN
DDX6_HUMAN	EF1B_HUMAN	CPSF1_HUMAN
FIP1_HUMAN	RBM39_HUMAN	PP1A_HUMAN
PRP16_HUMAN	RL6_HUMAN	RUVB1_HUMAN
CHD3_HUMAN	HNRPU_HUMAN	KIF22_HUMAN
H15_HUMAN	RIF1_HUMAN	RS9_HUMAN
CCAR1_HUMAN	P66A_HUMAN	UBF1_HUMAN
RS18_HUMAN	RL4_HUMAN	TR150_HUMAN
TP53B_HUMAN	ACINU_HUMAN	RL22_HUMAN
FUS_HUMAN	RL18_HUMAN	HP1B3_HUMAN
XRN2_HUMAN	FUS_HUMAN	ELAV1_HUMAN
DIDO1_HUMAN	KI20B_HUMAN	RBM10_HUMAN
RS17_HUMAN	EHMT1_HUMAN	TPX2_HUMAN
ZCCHV_HUMAN	COPB_HUMAN	SSRP1_HUMAN
FXR1_HUMAN	HNRPF_HUMAN	RS15A_HUMAN
RL30_HUMAN	KI67_HUMAN	NUMA1_HUMAN
RL18_HUMAN	6PGD_HUMAN	H2A1J_HUMAN
RL13_HUMAN	RL7_HUMAN	PININ_HUMAN
RRP5_HUMAN	ENOA_HUMAN	PCBP1_HUMAN
TRIPC_HUMAN	HLTF_HUMAN	HNRPG_HUMAN
FBRL_HUMAN	SMRC2_HUMAN	HNRPC_HUMAN
KIF2C_HUMAN	GBLP_HUMAN	SF3A3_HUMAN
PR40A_HUMAN	H2AY_HUMAN	SMCA5_HUMAN
KU86_HUMAN	SK2L2_HUMAN	HNRPF_HUMAN
EIF3D_HUMAN	ACL6A_HUMAN	PDIP3_HUMAN
NOG1_HUMAN	CDC5L_HUMAN	THOC1_HUMAN
COPA_HUMAN	SMRC1_HUMAN	CBX3_HUMAN
RL1D1_HUMAN	DDX50_HUMAN	SF3A1_HUMAN
ACINU_HUMAN	1433T_HUMAN	BC11B_HUMAN
CAPR1_HUMAN	GCP3_HUMAN	RL27_HUMAN
CPSF1_HUMAN	NOP56_HUMAN	RALY_HUMAN
EIF3F_HUMAN	ATD3A_HUMAN	HDAC2_HUMAN
PP1A_HUMAN	ENPL_HUMAN	EWS_HUMAN
RL13A_HUMAN	RS20_HUMAN	HDAC1_HUMAN
RUVB1_HUMAN	H2A1A_HUMAN	TFR1_HUMAN
KIF22_HUMAN	HNRPR_HUMAN	PCM1_HUMAN
RS9_HUMAN	RL15_HUMAN	ZN638_HUMAN
P80C_HUMAN	K1C13_HUMAN	SK2L2_HUMAN
UBF1_HUMAN	SNW1_HUMAN	ATPA_HUMAN
CG050_HUMAN	PROF1_HUMAN	CSTF3_HUMAN
TPR_HUMAN	PININ_HUMAN	DDX23_HUMAN
DDX47_HUMAN	DDX23_HUMAN	RL10_HUMAN
TR150_HUMAN	CARF_HUMAN	ATD3A_HUMAN
ROA3_HUMAN	RFC2_HUMAN	ML12B_HUMAN
RL22_HUMAN	REQU_HUMAN	RUNX1_HUMAN
GLYR1_HUMAN	SAFB1_HUMAN	ZC11A_HUMAN
RS2_HUMAN	PEBB_HUMAN	RBBP4_HUMAN
ACTC_HUMAN	RL14_HUMAN	RL15_HUMAN
PP1B_HUMAN	HIRA_HUMAN	C1QBP_HUMAN
PRKDC_HUMAN	CSTF3_HUMAN	UHRF1_HUMAN
STAU1_HUMAN	RL7A_HUMAN	H1X_HUMAN
HP1B3_HUMAN	RS18_HUMAN	SAFB1_HUMAN
EIF3C_HUMAN	MRE11_HUMAN	KHDR1_HUMAN
LYAR_HUMAN	TADBP_HUMAN	RS20_HUMAN
ELAV1_HUMAN	MATR3_HUMAN	BUB3_HUMAN
RBM10_HUMAN	PCM1_HUMAN	RU17_HUMAN
TPX2_HUMAN	THOC1_HUMAN	RLA1_HUMAN
HNRPL_HUMAN	UBIQ_HUMAN	PARP1_HUMAN
SNUT2_HUMAN	ZN828_HUMAN	SFRS6_HUMAN
SSRP1_HUMAN	ZC11A_HUMAN	RS12_HUMAN
RS15A_HUMAN	ML12B_HUMAN	RL14_HUMAN
IF16_HUMAN	TFR1_HUMAN	RFA1_HUMAN
NUMA1_HUMAN	RL11_HUMAN	RL24_HUMAN

2ABA_HUMAN	PHB2_HUMAN	SRPK1_HUMAN	RBM25_HUMAN	ADT2_HUMAN
MPCP_HUMAN	SF3B1_HUMAN	RL32_HUMAN	CAPR1_HUMAN	THOC2_HUMAN
DIMT1_HUMAN	RGPA2_HUMAN	H2A1J_HUMAN	YBOX1_HUMAN	SNW1_HUMAN
U5S1_HUMAN	RCD1_HUMAN	PININ_HUMAN	K2C4_HUMAN	NOLC1_HUMAN
CDK12_HUMAN	ANM1_HUMAN	PCBP1_HUMAN	DOCK2_HUMAN	RS6_HUMAN
RL30_HUMAN	PDLI2_HUMAN	HNRPG_HUMAN	RL17_HUMAN	PCBP2_HUMAN
RM50_HUMAN	PSMD3_HUMAN	HNRPC_HUMAN	RU2A_HUMAN	RL23A_HUMAN
RL38_HUMAN	SERPH_HUMAN	H2B1N_HUMAN	PR40A_HUMAN	RFC4_HUMAN
RL23_HUMAN	RALY_HUMAN	SF3A3_HUMAN	ARHG2_HUMAN	HCFC1_HUMAN
AIFM1_HUMAN	PCBP1_HUMAN	PCF11_HUMAN	H33_HUMAN	RL21_HUMAN
EPN4_HUMAN	GNAT1_HUMAN	SMCA5_HUMAN	SF3A3_HUMAN	CARF_HUMAN
PRKRA_HUMAN	ROA2_HUMAN	DDX1_HUMAN	NUMA_HUMAN	H2A1A_HUMAN
SNF5_HUMAN	RS11_HUMAN	HNRPF_HUMAN	ADNP_HUMAN	RL36_HUMAN
SSBP2_HUMAN	IF4A1_HUMAN	PDIP3_HUMAN	DNJC9_HUMAN	ARI1A_HUMAN
CNN2_HUMAN	EHMT2_HUMAN	THOC1_HUMAN	NKX25_HUMAN	DSRAD_HUMAN
STAU2_HUMAN	IDHP_HUMAN	CBX3_HUMAN	ELAV1_HUMAN	RL27A_HUMAN
CALU_HUMAN	HSP74_HUMAN	SF3A1_HUMAN	TFCP2_HUMAN	RL19_HUMAN
ZC3H4_HUMAN	6PGD_HUMAN	RL23_HUMAN	MPCP_HUMAN	IKZF1_HUMAN
GTSE1_HUMAN	NP1L1_HUMAN	PP1G_HUMAN	CDC73_HUMAN	KIF2A_HUMAN
TITIN_HUMAN	ATPG_HUMAN	ESF1_HUMAN	YMEL1_HUMAN	NCBP1_HUMAN
K6PL_HUMAN	P66B_HUMAN	BC11B_HUMAN	CRNN_HUMAN	SMRD2_HUMAN
GBLP_HUMAN	UBR1_HUMAN	EIF3B_HUMAN	RL10A_HUMAN	DNJC9_HUMAN
SHKB1_HUMAN	CAND1_HUMAN	RL27_HUMAN	RUVB1_HUMAN	PM14_HUMAN
ZC3HD_HUMAN	RA1L3_HUMAN	CD2AP_HUMAN	RL19_HUMAN	SMRD1_HUMAN
ZCHC8_HUMAN	UAP56_HUMAN	RA1L3_HUMAN	PHB2_HUMAN	CHD1_HUMAN
H1X_HUMAN	COX2_HUMAN	RALY_HUMAN	RL36_HUMAN	SFRS9_HUMAN
PCY1A_HUMAN	KDM3B_HUMAN	CN166_HUMAN	TRA2B_HUMAN	SMD2_HUMAN
MRT4_HUMAN	PHF14_HUMAN	HDAC2_HUMAN	RSMB_HUMAN	SAFB2_HUMAN
TRA2A_HUMAN	CAN1_HUMAN	DDX42_HUMAN	UBF1_HUMAN	NKX25_HUMAN
PABP2_HUMAN	MCM4_HUMAN	EWS_HUMAN	FIP1_HUMAN	SPF27_HUMAN
RT29_HUMAN	PURA_HUMAN	HDAC1_HUMAN	PRP19_HUMAN	HSP71_HUMAN
AT1A3_HUMAN	PRP8_HUMAN	LAR4B_HUMAN	CRNL1_HUMAN	P66B_HUMAN
PCM1_HUMAN	SYSC_HUMAN	ZFR_HUMAN	ISY1_HUMAN	SFRS7_HUMAN
HS90A_HUMAN	CISY_HUMAN	TFR1_HUMAN	ZMYM4_HUMAN	ZC3HE_HUMAN
SRP09_HUMAN	SYEP_HUMAN	PCM1_HUMAN	RFC4_HUMAN	IF4A1_HUMAN
EIF3D_HUMAN	MYOM3_HUMAN	ZN638_HUMAN	HP1B3_HUMAN	PGAM5_HUMAN
SK2L2_HUMAN	HIRA_HUMAN	SK2L2_HUMAN	AINX_HUMAN	PUF60_HUMAN
CAN1_HUMAN	COPZ1_HUMAN	YLPM1_HUMAN	RUXG_HUMAN	SF13A_HUMAN
MCM6_HUMAN	RENT1_HUMAN	NCOA5_HUMAN	SF3B4_HUMAN	H2AV_HUMAN
COPA_HUMAN	SAMD9_HUMAN	ATPA_HUMAN	RL27A_HUMAN	RL8_HUMAN
PCBP2_HUMAN	RL22_HUMAN	CSTF3_HUMAN	FBRL_HUMAN	SMRC2_HUMAN
PSDE_HUMAN	FPPS_HUMAN	DDX23_HUMAN	RFC3_HUMAN	RED_HUMAN
ANM1_HUMAN	NDUC1_HUMAN	RL10_HUMAN	KCRB_HUMAN	ACL6A_HUMAN
RM19_HUMAN	ADT3_HUMAN	ATD3A_HUMAN	RS16_HUMAN	TADBP_HUMAN
R39L5_HUMAN	SYYC_HUMAN	ML12B_HUMAN	RS9_HUMAN	EXOSX_HUMAN
KCMF1_HUMAN	ARPC4_HUMAN	RUNX1_HUMAN	KIF2A_HUMAN	SMRC1_HUMAN
ERH_HUMAN	GCN1L_HUMAN	HNRL1_HUMAN	CAZA1_HUMAN	KIF14_HUMAN
SRPK1_HUMAN	SUZ12_HUMAN	PHF3_HUMAN	CALL5_HUMAN	RBM25_HUMAN
KIF14_HUMAN	CF174_HUMAN	ZC11A_HUMAN	LSM2_HUMAN	TRA2B_HUMAN
RS24_HUMAN	MEP50_HUMAN	GNAI2_HUMAN	EWS_HUMAN	SMD1_HUMAN
RRAGC_HUMAN	RS28_HUMAN	SRP68_HUMAN	IF4A1_HUMAN	RAD50_HUMAN
SET_HUMAN	SYRC_HUMAN	RBBP4_HUMAN	SAFB2_HUMAN	THOC6_HUMAN
HNRH3_HUMAN	RAN_HUMAN	RL15_HUMAN	UN84B_HUMAN	LC7L2_HUMAN
MED4_HUMAN	RS19_HUMAN	PRP4B_HUMAN	TR150_HUMAN	NH2L1_HUMAN
CRNL1_HUMAN	WDR1_HUMAN	IF4G3_HUMAN	MBB1A_HUMAN	RUVB2_HUMAN
FRG1_HUMAN	FAS_HUMAN	C1QBP_HUMAN	XRN2_HUMAN	CAZA1_HUMAN
NSUN2_HUMAN	SLAI2_HUMAN	NU133_HUMAN	PGK1_HUMAN	BAZ1A_HUMAN
HNRL1_HUMAN	GFOD1_HUMAN	UHRF1_HUMAN	SAHH_HUMAN	TRA2A_HUMAN
RL21_HUMAN	RSMB_HUMAN	H1X_HUMAN	ILF3_HUMAN	SFRS3_HUMAN
HP1B3_HUMAN	PDIP3_HUMAN	SAFB1_HUMAN	PP1A_HUMAN	HNRL2_HUMAN
NUMA1_HUMAN	AT1A3_HUMAN	LCK_HUMAN	DPM1_HUMAN	PEBB_HUMAN
ABLM3_HUMAN	STAT4_HUMAN	CPSF3_HUMAN	PRDX1_HUMAN	RPN1_HUMAN
ARHG2_HUMAN	GSTP1_HUMAN	RS7_HUMAN	KDM1_HUMAN	RL11_HUMAN
RFC3_HUMAN	NAT10_HUMAN	PAIRB_HUMAN	FXR1_HUMAN	IMA2_HUMAN
MCM2_HUMAN	GMDS_HUMAN	KHDR1_HUMAN	RL8_HUMAN	MGN_HUMAN
K1522_HUMAN	OSR1_HUMAN	RS20_HUMAN	CHRC1_HUMAN	ABCD3_HUMAN
ABLM1_HUMAN	BAZ1A_HUMAN	BUB3_HUMAN	SF3A1_HUMAN	DPM1_HUMAN
UTP18_HUMAN	TRP13_HUMAN	RU17_HUMAN	SMC1A_HUMAN	SMHD1_HUMAN
RL36_HUMAN	GCP6_HUMAN	RLA1_HUMAN	WDR5_HUMAN	MPCP_HUMAN
CYFP1_HUMAN	RL29_HUMAN	PARP1_HUMAN	RCOR1_HUMAN	TFCP2_HUMAN
TRY2_HUMAN	RPB2_HUMAN	SFRS6_HUMAN	SMD2_HUMAN	CNOT1_HUMAN
HDAC1_HUMAN	ACOD_HUMAN	RS12_HUMAN	SFRS9_HUMAN	NONO_HUMAN
PDLI2_HUMAN	SMD3_HUMAN	HNRH2_HUMAN	RAD50_HUMAN	UBIQ_HUMAN
RHG25_HUMAN	KCTD5_HUMAN	RL14_HUMAN	RFA1_HUMAN	CTF18_HUMAN
COR1B_HUMAN	BACH_HUMAN	RFA1_HUMAN	PM14_HUMAN	H33_HUMAN
VIR_HUMAN	SYTC_HUMAN	RL24_HUMAN	CDK13_HUMAN	ZCH18_HUMAN
DCAF7_HUMAN	EIF3L_HUMAN	WBP11_HUMAN	SUZ12_HUMAN	RFC2_HUMAN
PCNA_HUMAN	SYFA_HUMAN	STAU2_HUMAN	HDGR2_HUMAN	RL35A_HUMAN
RS28_HUMAN	NOX5_HUMAN	ATX2L_HUMAN	RFC5_HUMAN	EXOS5_HUMAN
RT22_HUMAN	GMEB1_HUMAN	SNR40_HUMAN	RAN_HUMAN	DOCK2_HUMAN
TP53B_HUMAN	KIF22_HUMAN	ADT2_HUMAN	UAP56_HUMAN	ZN828_HUMAN
UBS3B_HUMAN	PLEC1_HUMAN	RBP56_HUMAN	H1X_HUMAN	PB1_HUMAN

225

COX2_HUMAN   COPD_HUMAN
SN12L_HUMAN   AIFM1_HUMAN
SMD1_HUMAN   SPNS1_HUMAN
ROA2_HUMAN   PPIA_HUMAN
YTDC1_HUMAN   CHRC1_HUMAN
DDX42_HUMAN   SET_HUMAN
SYFA_HUMAN   SSBP2_HUMAN
RS11_HUMAN   AEBP2_HUMAN
IMA3_HUMAN   CNN2_HUMAN
THIO_HUMAN   ROA1_HUMAN
XRN2_HUMAN   PR40A_HUMAN
SRRM2_HUMAN   FCGRN_HUMAN
SFRS7_HUMAN   PLXB3_HUMAN
KHDR1_HUMAN   MYL6_HUMAN
FXL18_HUMAN   MYL10_HUMAN
NDUA4_HUMAN   C170L_HUMAN
NOP2_HUMAN   PA2G4_HUMAN
BIG2_HUMAN   R39L5_HUMAN
CAPZB_HUMAN   RS20_HUMAN
QN1_HUMAN   HELC1_HUMAN
EP400_HUMAN   RL10A_HUMAN
RL3_HUMAN   DHX30_HUMAN
MED1_HUMAN   PTHB1_HUMAN
TRA2B_HUMAN   IPO7_HUMAN
CARF_HUMAN   DDX42_HUMAN
ZO1_HUMAN   AIMP2_HUMAN
ESF1_HUMAN   MAGI2_HUMAN
ROA0_HUMAN   FTSJ1_HUMAN
RT33_HUMAN   ZN274_HUMAN
DOCK8_HUMAN   NXF3_HUMAN
CHD4_HUMAN
TRIPB_HUMAN
OTOF_HUMAN
CL073_HUMAN
CPT1A_HUMAN
MTMRC_HUMAN
CDC5L_HUMAN
SHPRH_HUMAN
NO66_HUMAN
RH10L_HUMAN
PPME1_HUMAN
MINT_HUMAN
UH1BL_HUMAN
RBM27_HUMAN
NDUB4_HUMAN
SOS1_HUMAN
MTMR1_HUMAN
MOV10_HUMAN
AP1G1_HUMAN
ECT2L_HUMAN
RN213_HUMAN
PHLA1_HUMAN
KI21B_HUMAN
MED12_HUMAN
OPA1L_HUMAN
ATP5I_HUMAN
ATF6B_HUMAN
SF3B1_HUMAN
MA13P_HUMAN
DJC15_HUMAN
BUB3_HUMAN
ATRN_HUMAN
FGGY_HUMAN
FBSP1_HUMAN
INADL_HUMAN
PYR1_HUMAN
COR2A_HUMAN
NXF3_HUMAN
ACOC_HUMAN

VIR_HUMAN   PLST_HUMAN   POGZ_HUMAN
THOC2_HUMAN   LSM8_HUMAN   PHB2_HUMAN
HELC1_HUMAN   NOP58_HUMAN   SATB1_HUMAN
ELYS_HUMAN   RL29_HUMAN   EXOS6_HUMAN
SNW1_HUMAN   PGAM5_HUMAN   CDC73_HUMAN
TRAP1_HUMAN   PHF14_HUMAN   P66A_HUMAN
NOLC1_HUMAN   TPIS_HUMAN   HLTF_HUMAN
RSMN_HUMAN   RL3_HUMAN   ABCF1_HUMAN
RS6_HUMAN   ZMYM2_HUMAN   SNF5_HUMAN
NUP54_HUMAN   TIM44_HUMAN   SF3B4_HUMAN
PCBP2_HUMAN   NONO_HUMAN   HNRDL_HUMAN
U2AF2_HUMAN   LC7L2_HUMAN   PSIP1_HUMAN
RL23A_HUMAN   DPY30_HUMAN   SMC1A_HUMAN
LYRIC_HUMAN   RL22_HUMAN   MSH6_HUMAN
RRP1B_HUMAN   NAT10_HUMAN   RFC3_HUMAN
RFC4_HUMAN   LEG9_HUMAN   CD2B2_HUMAN
NU153_HUMAN   CHD1_HUMAN   TSR1_HUMAN
HCFC1_HUMAN   RL23A_HUMAN   XRN1_HUMAN
TOE1_HUMAN   IMB1_HUMAN   VPS35_HUMAN
RS24_HUMAN   EIF3A_HUMAN   EXOS9_HUMAN
HSP76_HUMAN   SMHD1_HUMAN   PELP1_HUMAN
ZC3H4_HUMAN   EIF3H_HUMAN   RFC5_HUMAN
UBP10_HUMAN   SF3B2_HUMAN   1433T_HUMAN
SRP72_HUMAN   EXOS9_HUMAN   EIF3H_HUMAN
RL21_HUMAN   TCF20_HUMAN   AINX_HUMAN
CARF_HUMAN   ALDOA_HUMAN   COPB_HUMAN
H2A1A_HUMAN   ZCCHV_HUMAN   PRDX1_HUMAN
ARFG3_HUMAN   TCPG_HUMAN   SEC13_HUMAN
SPF45_HUMAN   TBL1R_HUMAN   MBD3_HUMAN
RL36_HUMAN   TMM33_HUMAN   TCF7_HUMAN
DECR_HUMAN   SMAL1_HUMAN   ZN207_HUMAN
ADT3_HUMAN   EXOSX_HUMAN   SP16H_HUMAN
ARI1A_HUMAN   AP2A1_HUMAN   H2AY_HUMAN
DSRAD_HUMAN   HNRDL_HUMAN   DNJA3_HUMAN
PRP4_HUMAN   RCN2_HUMAN   PHF14_HUMAN
CPSF5_HUMAN   MSH6_HUMAN   TMM33_HUMAN
CHD7_HUMAN   THOC6_HUMAN   TIM44_HUMAN
RBP2_HUMAN   SC22B_HUMAN   RL29_HUMAN
RL27A_HUMAN   CD2B2_HUMAN   CPT1A_HUMAN
RL19_HUMAN   NOLC1_HUMAN   UAP56_HUMAN
IF6_HUMAN   SFRS6_HUMAN   TCOF_HUMAN
FL2D_HUMAN   R13AX_HUMAN   SSBP2_HUMAN
NU214_HUMAN   GNAT1_HUMAN   RCOR1_HUMAN
ROA0_HUMAN   UBP2L_HUMAN   KI20B_HUMAN
THOC4_HUMAN   EDC4_HUMAN   EFTU_HUMAN
E2AK2_HUMAN   PDLI2_HUMAN   WDR5_HUMAN
RBM5_HUMAN   SMRD2_HUMAN   MED1_HUMAN
IKZF1_HUMAN   SEC13_HUMAN   LEO1_HUMAN
KIF2A_HUMAN   CD11A_HUMAN   SLIRP_HUMAN
TOX4_HUMAN   ATPA_HUMAN   LAS1L_HUMAN
NCBP1_HUMAN   SFRS3_HUMAN   LS14B_HUMAN
DIMT1_HUMAN   BACH_HUMAN   SC61B_HUMAN
DDX52_HUMAN   SNF5_HUMAN   GRWD1_HUMAN
SMRD2_HUMAN   RL27_HUMAN   YMEL1_HUMAN
DNJC9_HUMAN   TOIP1_HUMAN   LEG9_HUMAN
ZCHC3_HUMAN   HNRL2_HUMAN   PHF5A_HUMAN
PM14_HUMAN   SPB3_HUMAN   SMCE1_HUMAN
SMRD1_HUMAN   PSIP1_HUMAN   SLBP_HUMAN
CHD1_HUMAN   CABIN_HUMAN   RAN_HUMAN
SFRS9_HUMAN   RL24_HUMAN   REQU_HUMAN
SMD2_HUMAN   LEO1_HUMAN   R39L5_HUMAN
PRPF3_HUMAN   DHX30_HUMAN   RUXG_HUMAN
SAFB2_HUMAN   AZI1_HUMAN   IKZF2_HUMAN
NHP2_HUMAN   SC61B_HUMAN   AQR_HUMAN
ARPC4_HUMAN   KIF14_HUMAN   SIN3A_HUMAN
NKX25_HUMAN   R39L5_HUMAN   LSM2_HUMAN
SPF27_HUMAN   CLPX_HUMAN   ADNP_HUMAN
HSP71_HUMAN   EXTL2_HUMAN   KDM1_HUMAN
P66B_HUMAN   THOC2_HUMAN   CHRC1_HUMAN
SFRS7_HUMAN   EXOS5_HUMAN   ZMYM2_HUMAN
BAT2_HUMAN   PCBP1_HUMAN   LSM3_HUMAN
GAR1_HUMAN   DSRAD_HUMAN   RAE1L_HUMAN
CQ085_HUMAN   RS6_HUMAN   IMB1_HUMAN
NUSAP_HUMAN   SMCE1_HUMAN   MBD2_HUMAN
ZC3HE_HUMAN   NCBP1_HUMAN   GCP2_HUMAN
IF4A1_HUMAN   PDIP3_HUMAN   PROF1_HUMAN
HCD2_HUMAN   ANM1_HUMAN   HM20A_HUMAN
PGAM5_HUMAN   CLUS_HUMAN   COX2_HUMAN
PAF1_HUMAN   RBM10_HUMAN   TOIP1_HUMAN
PUF60_HUMAN   KHDR1_HUMAN   LR16C_HUMAN
SF13A_HUMAN   DHX9_HUMAN   EF2_HUMAN
DHX8_HUMAN   GSTO1_HUMAN   RCN2_HUMAN

H2AV_HUMAN	TCOF_HUMAN
MMTA2_HUMAN	UBN1_HUMAN
RL8_HUMAN	GCP2_HUMAN
SMRC2_HUMAN	LSM3_HUMAN
RCC2_HUMAN	MGN_HUMAN
RED_HUMAN	GSTP1_HUMAN
RS27_HUMAN	ZCH18_HUMAN
ACL6A_HUMAN	RNZ2_HUMAN
FRG1_HUMAN	PB1_HUMAN
TADBP_HUMAN	SSBP2_HUMAN
SRP14_HUMAN	IF4G1_HUMAN
FXR2_HUMAN	RED_HUMAN
EXOSX_HUMAN	NH2L1_HUMAN
EIF3I_HUMAN	ZN592_HUMAN
PRP31_HUMAN	SIN3A_HUMAN
MTMR5_HUMAN	CNOT3_HUMAN
SYF1_HUMAN	ABCD3_HUMAN
SMRC1_HUMAN	DNJA3_HUMAN
KIF14_HUMAN	CPT1A_HUMAN
GPTC4_HUMAN	PCBP2_HUMAN
RBM25_HUMAN	GCP6_HUMAN
BLM_HUMAN	SF13A_HUMAN
TRA2B_HUMAN	RL35A_HUMAN
RL18A_HUMAN	KV305_HUMAN
RS5_HUMAN	TCPD_HUMAN
SMD1_HUMAN	C1QBP_HUMAN
RAD50_HUMAN	RLA1_HUMAN
NUP50_HUMAN	BOLA2_HUMAN
THOC6_HUMAN	BAZ1B_HUMAN
LC7L2_HUMAN	ZN207_HUMAN
NH2L1_HUMAN	TSR1_HUMAN
RL35_HUMAN	XPC_HUMAN
SFRS4_HUMAN	SLIRP_HUMAN
RUVB2_HUMAN	TCF7_HUMAN
IMA1_HUMAN	PHF5A_HUMAN
UTP18_HUMAN	TRA2A_HUMAN
PDS5A_HUMAN	CNO6L_HUMAN
CHD6_HUMAN	F128A_HUMAN
CAZA1_HUMAN	MA7D3_HUMAN
SMU1_HUMAN	LR16C_HUMAN
RL26L_HUMAN	RCN1_HUMAN
CAPZB_HUMAN	RPN1_HUMAN
RPB1_HUMAN	ZN638_HUMAN
BUD13_HUMAN	RAE1L_HUMAN
BAZ1A_HUMAN	SMRD1_HUMAN
TRA2A_HUMAN	COF1_HUMAN
CTR9_HUMAN	VPS35_HUMAN
RGAP1_HUMAN	CNOT2_HUMAN
SFRS3_HUMAN	ZC3HE_HUMAN
ARPC3_HUMAN	PI4KA_HUMAN
HNRL2_HUMAN	TCPQ_HUMAN
PEBB_HUMAN	IMA2_HUMAN
MED10_HUMAN	RL21_HUMAN
RPN1_HUMAN	TF3C3_HUMAN
CL043_HUMAN	CTF18_HUMAN
DCP1A_HUMAN	MBD3_HUMAN
RL22L_HUMAN	MYO1G_HUMAN
DHX36_HUMAN	SFRS7_HUMAN
LARP4_HUMAN	LG3BP_HUMAN
NO66_HUMAN	IYD1_HUMAN
TPM4_HUMAN	DESP_HUMAN
LBR_HUMAN	PELP1_HUMAN
RL11_HUMAN	STRAP_HUMAN
IMA2_HUMAN	MBD2_HUMAN
TEX10_HUMAN	HM20A_HUMAN
MED4_HUMAN	NOP2_HUMAN
PQBP1_HUMAN	RS12_HUMAN
SYFA_HUMAN	RU17_HUMAN
MGN_HUMAN	SMD1_HUMAN
DKC1_HUMAN	AQR_HUMAN
ABCD3_HUMAN	DDX6_HUMAN
DOCK8_HUMAN	MTF2_HUMAN
DPM1_HUMAN	TCPA_HUMAN
SMHD1_HUMAN	DPOE3_HUMAN
ARP3_HUMAN	FPPS_HUMAN
RRBP1_HUMAN	ESYT1_HUMAN
NUP53_HUMAN	COX2_HUMAN
CENPF_HUMAN	LS14B_HUMAN
TAF10_HUMAN	GRWD1_HUMAN
MTA3_HUMAN	CPSF1_HUMAN
FNBP4_HUMAN	SPF27_HUMAN
MPCP_HUMAN	SFRS1_HUMAN

227

FMR1_HUMAN
TFCP2_HUMAN
ABT1_HUMAN
CNOT1_HUMAN
NONO_HUMAN
SFRS2_HUMAN
SAMD9_HUMAN
IMA3_HUMAN
UBIQ_HUMAN
RB15B_HUMAN
IF2B3_HUMAN
NOL6_HUMAN
SRRM1_HUMAN
RM11_HUMAN
NUP62_HUMAN
CTF18_HUMAN
GNL3_HUMAN
RM12_HUMAN
RNPS1_HUMAN
RPR1B_HUMAN
ZC3HD_HUMAN
H33_HUMAN
CALM_HUMAN
LARP1_HUMAN
TAF9_HUMAN
ZCH18_HUMAN
CDC2_HUMAN
PR38B_HUMAN
RFC2_HUMAN
RL35A_HUMAN
EXOS5_HUMAN
RBM15_HUMAN
PO210_HUMAN
RS27L_HUMAN
PWP2_HUMAN
CDK12_HUMAN
SRRT_HUMAN
DOCK2_HUMAN
ZN828_HUMAN
PURA_HUMAN
EXOS2_HUMAN
KIF23_HUMAN
KRR1_HUMAN
RBM4_HUMAN
RBM27_HUMAN
PB1_HUMAN
POGZ_HUMAN
RSBNL_HUMAN
PHB2_HUMAN
SATB1_HUMAN
STRBP_HUMAN
AR6P4_HUMAN
EXOS6_HUMAN
CDC73_HUMAN
FA98A_HUMAN
SRP09_HUMAN
ARPC5_HUMAN
IF2P_HUMAN
NU155_HUMAN
PERI_HUMAN
NDUS3_HUMAN
EIF3G_HUMAN
GNAI3_HUMAN
P66A_HUMAN
CP088_HUMAN
EMD_HUMAN
SC23B_HUMAN
AATF_HUMAN
WDR33_HUMAN
HLTF_HUMAN
ABCF1_HUMAN
DDX41_HUMAN
SNF5_HUMAN
ABCF2_HUMAN
SF3B4_HUMAN
HNRDL_HUMAN
UBP7_HUMAN
CA077_HUMAN
CEP55_HUMAN
MRT4_HUMAN
NU107_HUMAN
B3GT6_HUMAN

SCML2_HUMAN
ABCF1_HUMAN
XRN1_HUMAN
MED1_HUMAN
HMMR_HUMAN
RL30_HUMAN
SLBP_HUMAN
RS15A_HUMAN
RL10_HUMAN
PUF60_HUMAN
EXOS6_HUMAN
CHERP_HUMAN
TCPZ_HUMAN
LAS1L_HUMAN
FHR5_HUMAN
UCHL1_HUMAN

228

IF4E_HUMAN
CP080_HUMAN
HNRH3_HUMAN
PIHD1_HUMAN
SH21A_HUMAN
STAG2_HUMAN
PSIP1_HUMAN
RAD18_HUMAN
CSTFT_HUMAN
SMC1A_HUMAN
PCNA_HUMAN
RBM8A_HUMAN
MSH6_HUMAN
NPA1P_HUMAN
WDR82_HUMAN
CS043_HUMAN
SDC10_HUMAN
AP2B1_HUMAN
RFC3_HUMAN
EIF3K_HUMAN
SON_HUMAN
ZN281_HUMAN
ALKB5_HUMAN
MYPT1_HUMAN
PRC1_HUMAN
PFD2_HUMAN
DREB_HUMAN
PARP2_HUMAN
ARP5L_HUMAN
CD2B2_HUMAN
SFRS5_HUMAN
DRG1_HUMAN
EIF3M_HUMAN
PLK1_HUMAN
MED17_HUMAN
RRMJ3_HUMAN
DDX55_HUMAN
TSR1_HUMAN
YTHD2_HUMAN
XRN1_HUMAN
VPS35_HUMAN
CV028_HUMAN
ZC3H1_HUMAN
EXOS9_HUMAN
DJC13_HUMAN
PABP2_HUMAN
CPSF2_HUMAN
EIF3E_HUMAN
PELP1_HUMAN
CCNK_HUMAN
NUP93_HUMAN
ITM2A_HUMAN
SPF30_HUMAN
SENP3_HUMAN
CH033_HUMAN
PAPD1_HUMAN
PP1RA_HUMAN
RUXF_HUMAN
TIF1B_HUMAN
BRD7_HUMAN
ARC1B_HUMAN
ARFG2_HUMAN
CDCA2_HUMAN
TOP3A_HUMAN
CBX5_HUMAN
TAF6_HUMAN
OGT1_HUMAN
RCD1_HUMAN
RFC5_HUMAN
1433T_HUMAN
EIF3H_HUMAN
BAG2_HUMAN
TFB1M_HUMAN
AINX_HUMAN
CPSF6_HUMAN
ASH2L_HUMAN
TBL2_HUMAN
COPB_HUMAN
RAGP1_HUMAN
DICER_HUMAN
ANM5_HUMAN
CE170_HUMAN

229

NCOR1_HUMAN
KRI1_HUMAN
CEBPZ_HUMAN
ZN326_HUMAN
PRDX1_HUMAN
MA7D1_HUMAN
GLD2_HUMAN
ANR17_HUMAN
SEC13_HUMAN
SF01_HUMAN
HELZ_HUMAN
MBD3_HUMAN
ROD1_HUMAN
MED30_HUMAN
TCF7_HUMAN
EXOS7_HUMAN
COPB2_HUMAN
ZN207_HUMAN
RBM6_HUMAN
CN021_HUMAN
NCBP2_HUMAN
COPD_HUMAN
SP16H_HUMAN
KI18A_HUMAN
ARL4C_HUMAN
TFAM_HUMAN
RRP8_HUMAN
ARPC2_HUMAN
H2AY_HUMAN
FUBP3_HUMAN
DNJA3_HUMAN
PHF14_HUMAN
AIM1_HUMAN
TM209_HUMAN
CS029_HUMAN
VAV_HUMAN
BCLF1_HUMAN
HAX1_HUMAN
MSH2_HUMAN
TRI26_HUMAN
DHX16_HUMAN
LSM12_HUMAN
PFD6_HUMAN
RENT2_HUMAN
TMM33_HUMAN
TIM44_HUMAN
ATAD5_HUMAN
RL29_HUMAN
MK67I_HUMAN
LRC59_HUMAN
NGDN_HUMAN
PPIL3_HUMAN
UXT_HUMAN
AIFM1_HUMAN
GLE1_HUMAN
PDIA6_HUMAN
CPT1A_HUMAN
PHTNS_HUMAN
SPC24_HUMAN
DDX24_HUMAN
LRRC1_HUMAN
NIPS2_HUMAN
TTF1_HUMAN
SC24B_HUMAN
SFR11_HUMAN
MED12_HUMAN
KDM3B_HUMAN
UBAP2_HUMAN
UAP56_HUMAN
TERF2_HUMAN
TCOF_HUMAN
SSBP2_HUMAN
MCM7_HUMAN
SRPK2_HUMAN
NSMA3_HUMAN
RPB4_HUMAN
3MG_HUMAN
CC124_HUMAN
PRAF2_HUMAN
RCOR1_HUMAN
GMFG_HUMAN
KI20B_HUMAN

230

TDIF1_HUMAN
EXOS4_HUMAN
HAUS1_HUMAN
PR38A_HUMAN
PIAS1_HUMAN
SF04_HUMAN
CD043_HUMAN
THIO_HUMAN
COR1A_HUMAN
EFTU_HUMAN
EAF6_HUMAN
CPSF4_HUMAN
NOP16_HUMAN
WDR5_HUMAN
RBBP5_HUMAN
MED1_HUMAN
MOES_HUMAN
XRCC1_HUMAN
NIPBL_HUMAN
COR1B_HUMAN
APC7_HUMAN
LEO1_HUMAN
SLIRP_HUMAN
EDC3_HUMAN
MSL1_HUMAN
LAS1L_HUMAN
RPB2_HUMAN
LS14B_HUMAN
CAF1B_HUMAN
LYRM7_HUMAN
FUBP2_HUMAN
NELFB_HUMAN
CU070_HUMAN
SAMD1_HUMAN
SPA5L_HUMAN
RMI1_HUMAN
ETV6_HUMAN
NIPS1_HUMAN
SIN3B_HUMAN
RBM34_HUMAN
UT14A_HUMAN
SFRS8_HUMAN
SC61B_HUMAN
THA11_HUMAN
NUP98_HUMAN
GRWD1_HUMAN
YTHD1_HUMAN
REN3B_HUMAN
THOC3_HUMAN
RU2B_HUMAN
YMEL1_HUMAN
HMGA1_HUMAN
RPB9_HUMAN
IF4B_HUMAN
CQ049_HUMAN
CHK1_HUMAN
MED27_HUMAN
TF3C1_HUMAN
TTK_HUMAN
OAS3_HUMAN
RPAB3_HUMAN
STK3_HUMAN
CN043_HUMAN
PPIG_HUMAN
DNJA1_HUMAN
PPIL2_HUMAN
KTN1_HUMAN
RBMS1_HUMAN
LEG9_HUMAN
RPA34_HUMAN
CT117_HUMAN
SFRIP_HUMAN
LMO7_HUMAN
RM53_HUMAN
PYR1_HUMAN
PLRG1_HUMAN
CCD61_HUMAN
PATZ1_HUMAN
CWC22_HUMAN
RM46_HUMAN
GCFC_HUMAN
SFR12_HUMAN

231

EXOS1_HUMAN
PKCB1_HUMAN
T4AF1_HUMAN
DHX40_HUMAN
GPKOW_HUMAN
NU205_HUMAN
PRP17_HUMAN
PAWR_HUMAN
MPP8_HUMAN
PHF5A_HUMAN
PAPOG_HUMAN
MED20_HUMAN
AP2A2_HUMAN
NDEL1_HUMAN
SMCE1_HUMAN
NOC4L_HUMAN
TIM13_HUMAN
RT16_HUMAN
2ABA_HUMAN
SLBP_HUMAN
NACA_HUMAN
RAN_HUMAN
CKAP5_HUMAN
REQU_HUMAN
CS066_HUMAN
EDF1_HUMAN
EST1A_HUMAN
PCID2_HUMAN
TAF5_HUMAN
HBS1L_HUMAN
CENPQ_HUMAN
CK059_HUMAN
RT28_HUMAN
SI1L1_HUMAN
BRMS1_HUMAN
WRN_HUMAN
R39L5_HUMAN
RPRD2_HUMAN
CT004_HUMAN
LMBL3_HUMAN
WDR6_HUMAN
NDUA4_HUMAN
RS29_HUMAN
LDB1_HUMAN
COPG2_HUMAN
SPAS2_HUMAN
CX056_HUMAN
MED11_HUMAN
RL34_HUMAN
PEO1_HUMAN
SFR14_HUMAN
RUXG_HUMAN
CND3_HUMAN
F120A_HUMAN
IKZF2_HUMAN
CSK21_HUMAN
CYFP2_HUMAN
SEH1_HUMAN
PPIL1_HUMAN
GBB2_HUMAN
BRX1_HUMAN
K0020_HUMAN
RAD21_HUMAN
MRM1_HUMAN
TITIN_HUMAN
AQR_HUMAN
TR112_HUMAN
SRBD1_HUMAN
SLTM_HUMAN
SIN3A_HUMAN
SSRD_HUMAN
TFAP4_HUMAN
RM19_HUMAN
LSM2_HUMAN
PCAT1_HUMAN
Z3H7A_HUMAN
PRKRA_HUMAN
ADNP_HUMAN
GTSF1_HUMAN
PMF1_HUMAN
DDX46_HUMAN
KDM1_HUMAN

232

RL40_HUMAN
LRCH4_HUMAN
CHRC1_HUMAN
INO80_HUMAN
MEP50_HUMAN
IWS1_HUMAN
CGBP1_HUMAN
WDR36_HUMAN
RFIP2_HUMAN
MYST2_HUMAN
ABLM3_HUMAN
ELMO1_HUMAN
U640_HUMAN
NP1L1_HUMAN
NSD3_HUMAN
TADA1_HUMAN
RM03_HUMAN
RM40_HUMAN
CT043_HUMAN
ZGPAT_HUMAN
ZMYM2_HUMAN
CCD12_HUMAN
CNOT7_HUMAN
GNL3L_HUMAN
RBMX2_HUMAN
VAPA_HUMAN
PPIH_HUMAN
PERQ2_HUMAN
THOC7_HUMAN
NU160_HUMAN
NU188_HUMAN
TFPT_HUMAN
CSTF1_HUMAN
SCC4_HUMAN
ECHA_HUMAN
RM38_HUMAN
LSM3_HUMAN
CLP1_HUMAN
POP7_HUMAN
WDR61_HUMAN
NDUAD_HUMAN
SF3A2_HUMAN
RB12B_HUMAN
CTBL1_HUMAN
RAE1L_HUMAN
MED22_HUMAN
MARK2_HUMAN
TBL3_HUMAN
RPOM_HUMAN
NOL7_HUMAN
CC019_HUMAN
RBM16_HUMAN
RM48_HUMAN
INT3_HUMAN
PLPL6_HUMAN
MSI2H_HUMAN
CSK_HUMAN
EBP2_HUMAN
CI114_HUMAN
LSM7_HUMAN
GPTC1_HUMAN
RRP7A_HUMAN
UCHL5_HUMAN
AAAS_HUMAN
CAF1A_HUMAN
RM18_HUMAN
EMSY_HUMAN
RAC2_HUMAN
DHX35_HUMAN
VDAC2_HUMAN
TRPT1_HUMAN
PPIA_HUMAN
THOC5_HUMAN
PUM2_HUMAN
BOREA_HUMAN
KRIT1_HUMAN
NOM1_HUMAN
RPAB1_HUMAN
GTPB1_HUMAN
EIF3J_HUMAN
RT27_HUMAN
ACTZ_HUMAN

233

CO6A1_HUMAN
CCNL2_HUMAN
CDC23_HUMAN
KDM5A_HUMAN
RFA3_HUMAN
NOL11_HUMAN
RT29_HUMAN
IMP3_HUMAN
PDRG1_HUMAN
TE2IP_HUMAN
VRK3_HUMAN
TIAR_HUMAN
4F2_HUMAN
RRP15_HUMAN
IMB1_HUMAN
RPP30_HUMAN
PP4C_HUMAN
TF2H4_HUMAN
SAS10_HUMAN
BUD31_HUMAN
SMC2_HUMAN
DNLI4_HUMAN
ENY2_HUMAN
MBD2_HUMAN
GCP2_HUMAN
ABCE1_HUMAN
TAF6L_HUMAN
NUPL2_HUMAN
CD3E_HUMAN
DDX28_HUMAN
DNJC2_HUMAN
PROF1_HUMAN
DCAF7_HUMAN
HM20A_HUMAN
NSUN2_HUMAN
TCB_HUMAN
MPH6_HUMAN
HAKAI_HUMAN
KV303_HUMAN
TIM14_HUMAN
IMA4_HUMAN
APC5_HUMAN
PFD4_HUMAN
SSRG_HUMAN
ZCRB1_HUMAN
COX2_HUMAN
CCNG1_HUMAN
TOIP1_HUMAN
NIP7_HUMAN
TLN2_HUMAN
K1267_HUMAN
STK6_HUMAN
GATD1_HUMAN
RT17_HUMAN
EP400_HUMAN
LR16C_HUMAN
TAF13_HUMAN
TMOD3_HUMAN
NOC3L_HUMAN
EF2_HUMAN
RP1_HUMAN
RT22_HUMAN
SAM13_HUMAN
OTOAN_HUMAN
FGGY_HUMAN
DOCK7_HUMAN
Z512B_HUMAN
RT09_HUMAN
MED28_HUMAN
MCM6_HUMAN
SSRA_HUMAN
DNJC1_HUMAN
FMNL2_HUMAN
AP2M1_HUMAN
RCN2_HUMAN
VAMP8_HUMAN
AKAP8_HUMAN
DERL1_HUMAN
NOL9_HUMAN
GHSR_HUMAN
IF4G2_HUMAN
CNT3B_HUMAN

234

DLRB2_HUMAN
RAVR1_HUMAN
TRI75_HUMAN
RM47_HUMAN
F120C_HUMAN
RING1_HUMAN
FRM4A_HUMAN
DAXX_HUMAN
MED8_HUMAN
RCL1_HUMAN
ZN598_HUMAN
PURB_HUMAN
CO044_HUMAN
PTHB1_HUMAN