UNIVERSITY OF BIRMINGHAM

The consequences of the aberrant expression of TAL1 and its isoforms in T cell acute lymphoblastic leukaemia

by

Shorog Al Omair

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

DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences

College of Medical and Dental Sciences

University of Birmingham

November 2019

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties.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

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

Abstract

Most of the T-ALL cases aberrantly express the bHLH transcription factor TAL1. To further understand the role of TAL1 in T-ALL, we examined TAL1 protein interactions, genome-wide distribution, and assessed their mediated regulation of cellular processes. We identified proteins interactions networks associated with TAL1 and its interacting partner HEB, in T-ALL cells, including major regulators of T cell development. Our characterisation of the genomic occupancy of TAL1 interacting partners identified DNA methylation as a modulator of TAL1 transcriptional complex. Furthermore, TAL1 has four isoforms with described structural and functional differences in haematopoietic cells. Although mouse models indicate that long and short isoforms contribute to leukaemia, differential functional characteristics of TAL1 isoforms have not been investigated. We constructed four FLAG-tagged TAL1 isoforms constructs. We used the constructs to stably overexpress different TAL1 isoforms in T-ALL cells and human CD34+ cells undergoing T cell differentiation In-vitro. Our study found a differential genomic distribution of TAL1 depending on isoform overexpression and consequent effects on differentiation markers. We demonstrated that TAL1 overexpression resulted in an increase in early T cell differentiation markers which was not isoform-specific. Finally, our work highlights molecular mechanisms of TAL1 role in T-cell acute lymphoblastic leukaemia at the isoform level.

Acknowledgements

I would like to express my deepest gratitude for my supervisor's Dr Maarten Hoogenkamp and Dr Vesna Stanulovic for their guidance and outstanding supervision. They provided support and excellent technical training during my studies. I would like to thank my family, especially my parents, for always believing in me and for being a source of support and inspiration. Special thank you goes to my husband Abdulaziz for his patience and support during this challenging journey. The work I present today would not have been possible without the collaborative attitude of colleagues. I would like to thank Dr Douglas Ward for running our samples on mass spectrometry, Dr Sylvia Miller, for her help with Mass spectrometry analysis and Dr Ana Cumano for providing us with OP9-DL4 cells. I would like to thank members of our group Sarah Binhassan and Fatima Suliman. Finally, I would like to acknowledge King Faisal University in Saudi Arabia for funding my studies.

Content

List of chapters

List of illustrations

List of tables

List of abbreviations

5mC	5-Methylcytosine
HEBalt	Alternative HEB
bHLH	Basic helix loop helix
BHLH	Basic positively charged region and the helix loop helix domain
BP	Biological processes
CIAP	Calf Intestinal Alkaline Phosphatase
HEBcan	Canonical HEB
ChIP-seq	Chromatin Immunoprecipitation Sequencing
CHTOP	Chromatin target of PRMT1 protein
co-IP	Co-immunoprecipitation
CMP	Common myeloid progenitors
CPEB1	Cytoplasmic Polyadenylation Element Binding Protein 1
°C	Degree Celsius
DL	Delta-like
DCs	Dendritic cells
DHS	DNaseI hypersensitivity
DN	Double-negative
ETP	Early thymic progenitors
ETP	Early T-lineage progenitor
ES	Embryonic stem cells
ERH	Enhancer of rudimentary homolog
FACS	Fluorescence activated cell sorting
FLT3L	FMS-like tyrosine kinase three ligand
HSCs	Haematopoietic stem cells
HCL	Hierarchical clustering
IC	Input control samples
ISP	Intermediate single positive
IP	Immunoprecipitation
JDP2	Jun dimerization protein 2
LMPP	Lymphoid-primed multipotent progenitors
MEP	Megakaryocyte-erythroid progenitors
MeDIP	Methylated DNA immunoprecipitation
µl	Micro litre
MLP	Multilineage progenitor
MPPs	Multipotent progenitors
NICD	NOTCH intracellular domain
ORF	Open reading frame
PCI	Phenol/chloroform/isoamyl alcohol
PB	PiggyBac
RBPJ	Recombining binding protein suppressor of hairless
RT	Room temperature
SIL	SCL-interrupting locus
SP	Single-positive cells
SSBPs	Single-stranded DNA binding proteins
SCF	Stem cell growth factor
T-ALL	T-cell acute lymphoblastic leukaemia
TECs	Thymic epithelium cells
uORF	Upstream open reading frame

CHAPTER ONE

INTRODUCTION

Haematopoiesis

Haematopoietic stem cells (HSCs) generate progenitors that further differentiate into all different types of blood cells (Busch et al., 2015). This hierarchy of haematopoiesis gives rise to fully differentiated, functional blood cells, as shown in **figure 1.1**. The continuous flow of haematopoiesis starts with HSCs, and multipotent progenitors (MPPs) followed with more specified progenitors like CMP (Common myeloid progenitors) and LMPP (lymphoid-primed multipotential progenitor). Subsequently, CMP progenies diverge into MEP (Megakaryocyte-erythroid progenitors) and GMP. MEPs maintain the production of megakaryocytes, and erythrocytes; while GMPs generate neutrophils, monocytes, and eosinophils. Dendritic cells (DCs) can develop from GMP and LMMPs (Laurenti & Göttgens, 2018). Progenitors with lymphoid potential include LMPPs (Goardon, Nicolas et al., 2011) and the common lymphoid progenitors CLPs, which develop into natural killer cells, and lymphocytes (Cumano et al., 2019; Saran et al., 2010). The intermittent influx of lymphoid-potent progenitors seeds the thymus and initiates T cell differentiation (Petrie, 2003). The various types of T cells that develop in the thymus are TCRαβ T cells including $CD4^+$ helper and $CD8^+$ cytotoxic T cells, TCRγδ T cells, regulatory T cells (Treg), and NKT cells, which are T cells with some NK characteristics (Bendelac, Savage, & Teyton, 2007; Petrie & Zúñiga-Pflücker, 2007).

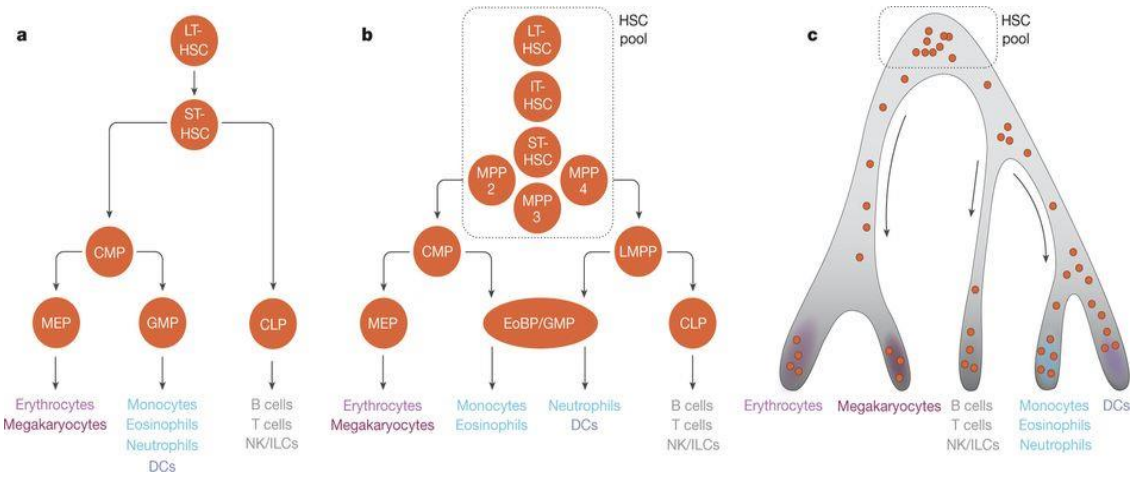


Figure 1.1 Schematic diagrams of the Hierarchy of Haematopoiesis

The figure shows the change in the view of haematopoiesis over time; (A) The hierarchy around 2000, (B) the period between 2005-2015, and (C) the continuum view of haematopoiesis introduced in 2016 (c).

T cell development

The pre-thymic progenitors relocate into the thymus. In the thymus, they become early thymic progenitors (ETP), where they interact with the signals of thymic epithelium cells (TECs) (Rothenberg, Moore, & Yui, 2008). Pre-thymic progenitors express Interleukin 7 receptor (IL7R), stem cell growth factor receptor (SCFR/KIT), FMS-like receptor tyrosine kinase 3 (FLT3), and NOTCH1 receptors (Heinzel, Benz, Martins, Haidl, & Bleul, 2007). The thymic environment provides their ligands, Interleukin 7 (IL7), stem cell growth factor (SCF), FMS-like tyrosine kinase three ligand (FLT3L), and Notch ligands (Moore & Zlotnik, 1997; Petrie & Zúñiga-Pflücker, 2007).

Notch signalling in the development of T cell

Notch signalling is highly conserved and has an essential role in regulating cell fate specification (Radtke, Fasnacht, & MacDonald, 2010). Four Notch receptors NOTCH1-4 have been described in mammalian cells, and many of their ligands have been identified including Delta-like (DL) 1, DL3, DL4, Jagged 1 and Jagged2 (Bray, 2006). Specifically, NOTCH1 receptors on ETPs have been shown indispensable for inducing early T cell development and inhibiting all other lineage potentials (Maillard, Fang, & Pear, 2005). DL1 and DL4 have both been considered major Notch1 ligands in the thymic environment. Stromal cell line OP9 overexpressing DL1 or DL4 can support T cell differentiation *in vitro* (Besseyrias et al., 2007; De Smedt, Hoebeke, & Plum, 2004). However, unlike DL1, knocking-in DL4 in thymic epithelium cells led to complete suppression of T cells (Koch, U. et al., 2008). Therefore, the activation of NOTCH1 by DL4 is vital for the specification of T cell lineage.

Modulators of Notch signalling

The mechanism of Notch signalling involves several modulators of its cascade. Upon ligand binding to the receptor, NOTCH1 is cleaved by the proteases ADAM (Selkoe & Kopan, 2003) and γ-secretase (Fortini, 2002). Then, NOTCH intracellular domain (NICD) heterodimerises with RBPJ (Recombining binding protein suppressor of hairless) within the nucleus and recruits MAML1 (mastermind-like 1) and the acetyltransferase p300 to regulate gene expression (Kopan & Ilagan, 2009). The PEST domain at the c-

terminus of NOTCH1 controls the activated intracellular domain (NICD). The PEST domain of the activated NICD binds E3-ubiquitin ligase, like FBXW7, which induces its degradation (Thompson et al., 2007).

T cell differentiation

The landscape of T cell development has been further defined in mice compared to the human, as illustrated in **figure 1.2**. In the mouse, differentiating T cells are categorised into four types, using the expression of CD4 and CD8, double-negative (DN) intermediate single positive (ISP), and double-positive stages (DP), and single-positive cells (SP), as reviewed in (Koch & Radtke, 2011). The four double negative stages are distinguished further using CD44 and CD25 (Petrie & Zúñiga-Pflücker, 2007; Rothenberg et al., 2008). ETP/DN1 differentiate into DN2 cells, which start migrating through the thymus as DN2a and transform into DN2b cells. DN2b cells with low KIT expression mark the initiation of T cell commitment, while DN3 cells in the subcapsular zone denote the completion of T cell commitment (Ceredig & Rolink, 2002; Rothenberg et al., 2008).

β-selection further divides DN3 cells into DN3a and DN3b (Rothenberg et al., 2008). The rearrangement of the T cell receptor gamma delta or beta (TCR-γ/δ or TCR-β) happens in the DN3a stage. The transition into DN3b indicates the assembly of the pre-TCR receptor, which consists of a fully rearranged TCR-β, CD3 and pre-TCR-α. DN3b cells with pre-TCR receive signals that allow them to pass β-selection and promote differentiation into DN4 (Taghon, Yui, Pant, Diamond, & Rothenberg, 2006).

The signal through the pre-TCR receptor induces CD4 and CD8 expression. The transient ISP stage is characterised by CD8 expression in the absence of CD4, preceding the double-positive stage (Xiong, Armato, & Yankee, 2010). The DP cells are wholly dependent on the completely rearranged TCR signalling, instead of Notch1 signalling. At this point, DP cells undergo further selection in their way to become fully functioning mature T cells (Borowski et al., 2002; Rothenberg et al., 2008).

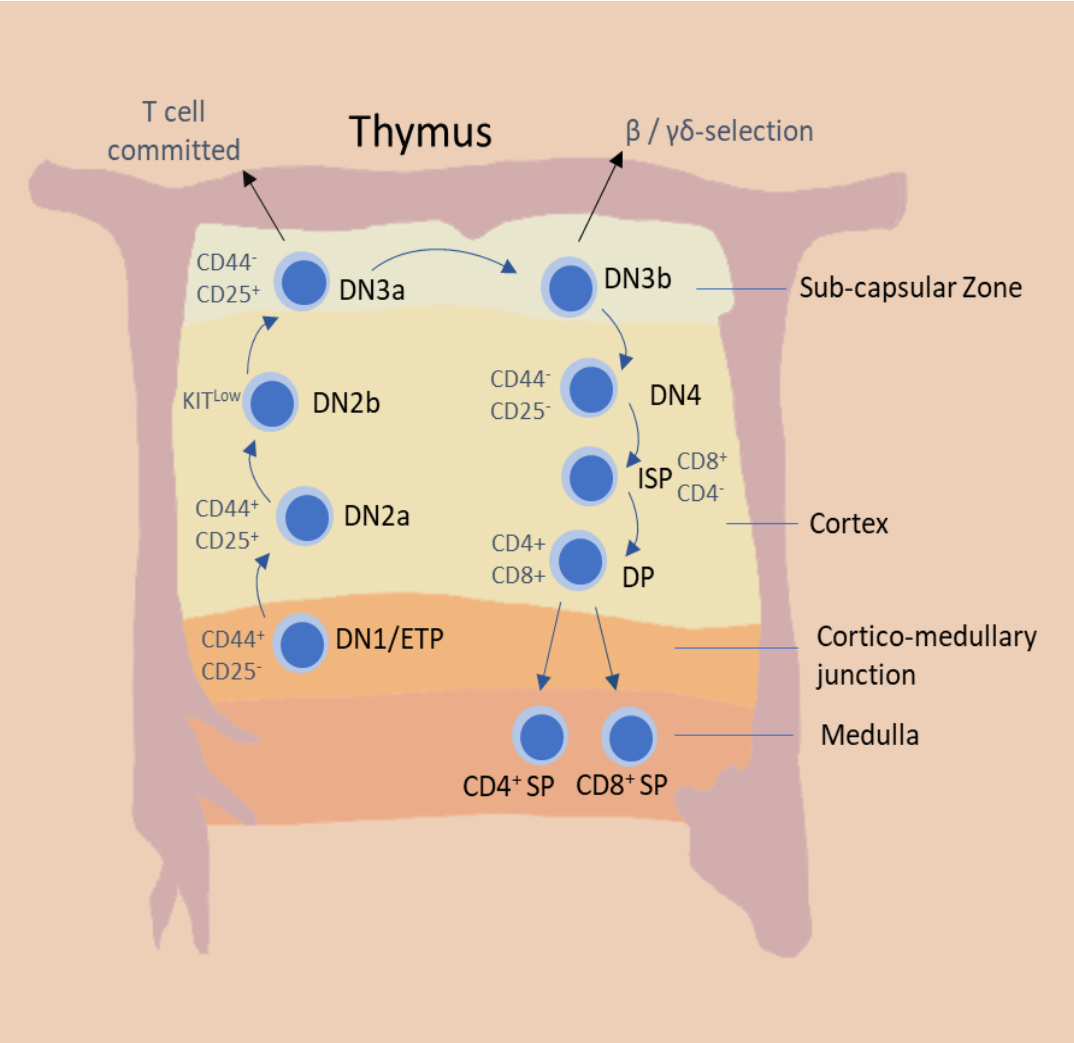


Figure 1.2 Illustration of mouse T cell differentiation stages in the thymus

Mouse T cell progenitors differente as they migrate through the thymus. Four CD4/CD8 double-negative (DN1-4) T cell progenitors can be distinguished on basis of CD44 and CD25 expression. KIT expression further divides the DN2 cells population into DN2a (KIT^{high}) and DN2b (KIT^{low}). T cell commitment occurs at the DN3a stage and Beta-/gamma delta TCR rearrangement at DN3b, which can be seprated on basis of their size and CD27 expression. An intermediate single-positive (ISP) with CD8 expression ($CD4^+$ ISP in human) precedes the transition into the CD4/CD8 double-positive stage and final differentiation into either $CD4^+$ or $CD8^+$ single-positive T cells.

Human T cell differentiation stages

Early T cell differentiation in human has been classified using different surface markers than for mice. Three double negative stages of T cell differentiation are commonly described (Awong et al., 2009; Blom & Spits, 2006). To identify pre-thymic progenitors in human, CD34 was first considered since it is known to be a marker of development and found in haematopoietic progenitors. The $CD34^{+}$; $CD7^{++}$; $CD45RA^{+}$ progenitors were found in cord blood and differentiated into T cell in the thymus (Haddad et al., 2004). Consistent with this, cells with $CD34^{+}CD38^{low}$ immunophenotype were shown to be an ETPs as they exhibited the least TCR rearrangements (Dik et al., 2005; Res et al., 1996).

The surface expression of CD5 and CD1a further categorise the next stages. Consequently, progenitors with a $CD7^{+}$; $CD5^{+}$; $CD1a^{-}$ phenotype represent the pro-T1 stage and the following stage of $CD7^{+}$; $CD5^{+}$; $CD1a^{+}$ is the ProT2 (Awong et al., 2009). The CD1a positive stage marks the human T cell commitment, as the cells lose their ability to differentiate into other lineages under normal conditions (Spits, 2002). The ISP stage in human carries the expression of CD4, which is different from the mouse ISP cells, which express CD8 (**figure 1.3**). Finally, markers that depict late human T cell

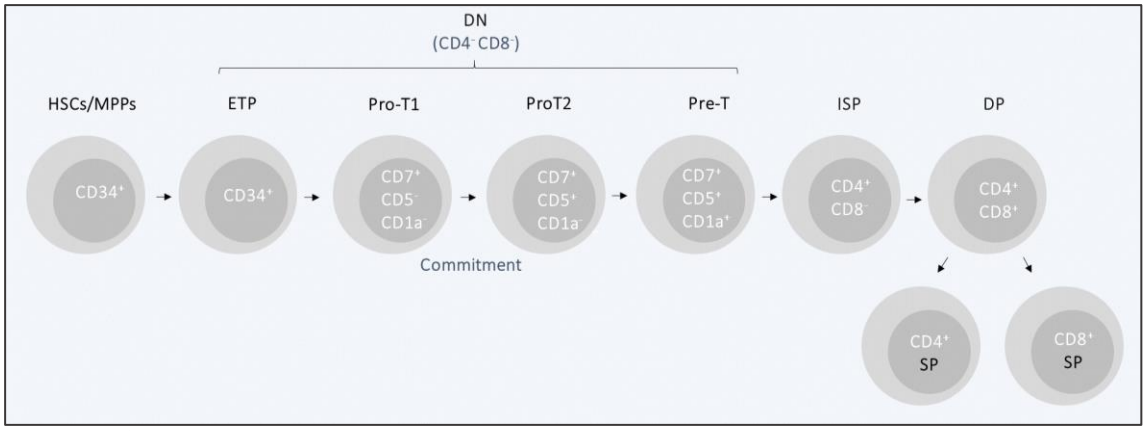


Figure 1.3 Schematic representation of human T cell differentiation stages
HSCs/multipotent progenitors reach the thymus, at which point they are called early thymus progenitors (ETPs). The next CD4/CD8 double negative stages are categorised using CD7, CD5 and CD1a. ETP differentiate into $CD7^{+}$; $CD5^{-}$; $CD1a^{-}$ (pro-T1), then $CD7^{+}$; $CD5^{+}$; $CD1a^{-}$ (ProT2) cells. Acquisition of CD1a expression coincides with T cell lineage commitment. Human intermediate single positive (ISP) cells express CD4, which will further differentiate into CD4/CD8 double positive cells and ultimately give rise to CD4 or CD8 single positive cells.

differentiation are the same as in mouse and include CD2, CD4, CD8, and CD3 (Awong et al., 2009).

Transcriptional regulation of T cell differentiation:

Transcription factor networks regulate the progression of T cell differentiation. Yui and Rothenberg (2014) classified the major transcriptional regulators of mouse T cell development into three phases (**figure 1.4**). First, the transcription factors that are expressed in pre-thymic progenitors and suppressed before T cell commitment. Transcription factors like *Hoxa9*, *Meis1* are suppressed before the ETP progress into DN2a stage, while *Lmo2* and *Mef2c* are next to be downregulated at the DN2a stage. Other transcription factors remain expressed until the DN2b stage, like *Spi1, Hhex, Gfi1b*, *Erg* and the basic helix loop helix class II proteins *Tal1*, *Lyl1*, and *N-Myc* (Heng et al., 2008, Yui, Rothenberg, 2014). Even though the expression of phase 1 transcription factors is restricted, there is evidence supporting that they are essential in early T cell differentiation. For instance, *Lyl1*, *Bcl11a*, and *Spi1* are required for T cell production (Zohren et al., 2012, Yu et al., 2012, Champhekar et al., 2015).

The second phase, as described by Yui and Rothenberg (2014), includes T cell-associated transcription factors supporting T cell commitment. *Gata3* and *Tcf7* are the first T cell transcription factors to be expressed starting at ETP stage, thus permitting their collaboration with the first phase regulators. Some genes of phase 1 are substituted with other members of the same family as *Ets1* and *Ets2* replacing *Spi1*, and *Runx1* dominating over *Runx2* and *Runx3* in committed T cell progenitors. Inhibition of phase 1 genes is essential for T cell commitment and is facilitated by T cell-specific genes, including *Gata3, Tcf7*, and *Bcl11b*. Also, with the suppression of class II HLH like Tal1 and Lyl1, E-protein dimers are formed.

Furthermore, after β-selection the gene expression of the final phase, as summarised in a review of Yui and Rothenberg (2014), is for transcription factors such as *Rorc*, *Ikzf3, Pou6f1*. Pre-TCR signalling activates proliferation, inhibits *Il7r*, and leads to the reduction of the expression of NOTCH1 target genes. Some of the phase 2 gene expression remains expressed and even increases during phase 3, like the expression of *Ets1/2, Tcf7, Lef1*, and *Tcf12*. Finally, the inhibition of genes of earlier phases is crucial

for normal T cell differentiation, as their aberrant expression may cause leukaemia (Yui, Rothenberg, 2014).

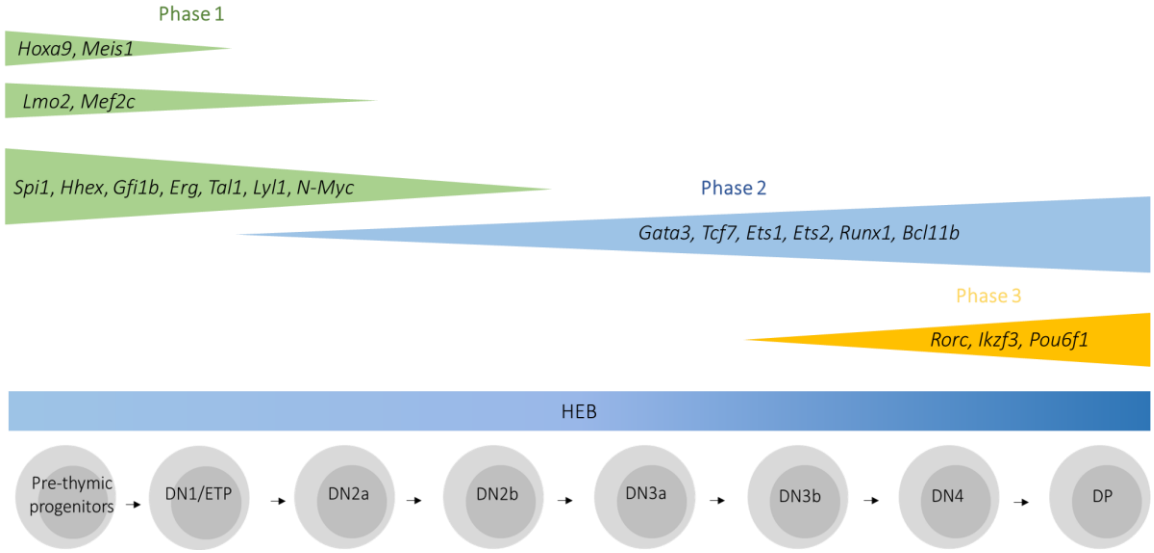


Figure 1.4 Three phases of major changes in transcription factors expression patterns occur during early T cell differentiation
Phase 1 (green) is characterised by the down-regulation of transcription factor genes that are typical for HSC/MPP cells. This occurs from the pre-thymic progenitors (*e.g. Hoxa9*, *Meis1*), through the DN2a stage (*e.g. Lmo2* and *Mef2c*). to the DN2b stage (*e.g. e.g. Spi1*, *Hhex*, *Gfi1b*, *Erg, Tal1*, *Lyl1*, and *N-Myc*). **Phase 2** (blue) is typified by the upregulation of T cell associated genes (e.g. *Gata3, Tcf7, Ets1, Ets2, Runx1,* and *Bcl11b*), which is initiated as early as the ETP/DN1 stage. Finally, **Phase 3** (orange) shows increasing expression of transcription factors such as *Rorc, Ikzf3, Pou6f1*, from DN3b, coinciding with T cell commitment. HEB (HEBcan) expression follows the pattern of the phase 3 genes. It should be noted that the expression patterns of the transcription factors within each group are similar, but not identical.

TAL1 in haematopoiesis:

Knowledge about TAL1 expression and function in normal haematopoiesis is essential for understanding the roles of TAL1 in T-ALL. TAL1 is first expressed during development in the yolk sac in cells that are destined to initiate primitive haematopoiesis (Kallianpur, Jordan, & Brandt, 1994a). Then, TAL1 is detected in all definitive haematopoietic organs including AGM, the fetal liver, bone marrow and spleen (Lécuyer & Hoang, 2004). TAL1 expression is high in HSCs and early multipotent progenitors (Kallianpur et al., 1994). Then, TAL1 is generally suppressed in differentiated haematopoietic cells, except for erythrocyte, megakaryocyte and mast cells, where TAL1 contributes to lineage differentiation and remains expressed at high levels (Hoang, Lambert & Martin, 2016a).

The significance of TAL1 in normal haematopoiesis was first recognised after studies showed that TAL1 knockout mice died due to a severe defect in haematopoiesis (Robb et al., 1995; Shivdasani, Mayer, & Orkin, 1995). Furthermore, $TAL1^{-/-}$ mouse embryonic cells could not differentiate into any of the haematopoietic lineages in chimeric mice or using *in vitro* haematopoietic differentiation studies (Endoh, Ogawa, Orkin, & Nishikawa, 2002; Porcher et al., 1996; Robb et al., 1996). Accumulating evidence demonstrated the importance of TAL1 in the haematopoietic and endothelial specification in mesoderm and more specifically for the formation of the hemangioblast (Gering, Rodaway, Gottgens, Patient, & Green, 1998; Lécuyer & Hoang, 2004).

Subsequently, the question of whether TAL1 is necessary for adult HSCs was addressed using overexpression and knockout studies. Overexpression of TAL1 in HSCs improved the HSC reconstitution of the irradiated mouse (Kunisato et al., 2004; Reynaud et al., 2005). Nevertheless, conditional knockout studies showed that TAL1 was dispensable for adult HSC maintenance (Curtis, D. J. et al., 2004; Mikkola et al., 2003). Several theories were suggested to explain this controversy as reviewed in (Hoang, Lambert, & Martin, 2016a).

One of these theories is that the functional roles of TAL1 in adult HSCs are redundant with its homolog LYL1. HSCs isolated from TAL1-LYL1 conditional double-knockout mice failed to reconstitute an irradiated recipient and to contribute to long term

haematopoiesis (Souroullas, Salmon, Sablitzky, Curtis, & Goodell, 2009). However, unlike TAL1, LYL1 knockout is not lethal during embryonic development and cannot rescue TAL1 embryonic lethality. LYL1 null mice showed a decrease in B cells, indicating that LYL1 is essential for the B lineage (Capron et al., 2006). In $TAL1^{-/-}$ embryonic stem cells (ES), it was shown that LYL1 could not compensate for TAL1 role in maintaining haematopoiesis (Chan et al., 2007). Therefore, while there are some overlapping functions between TAL1 and LYL1, there are distinct roles as demonstrated by expression and functional analyses.

Although LYL1 expression can explain the dispensable role of TAL1 in the maintenance of adult HSCs, there is also evidence that TAL1 may not be dispensable under all circumstances. One study hypothesised that the necessity of TAL1 in HSCs might become evident after a more extended period of transplantation. They showed that $TAL1^{+/-}$ HSCs had a reduced repopulating ability compared to $TAL1^{+/+}$ using serial transplantation because of the regulation of *Cdkn1a and Id1* expression by TAL1 (Lacombe, J. et al., 2010).

Determining the functional involvement of TAL1 in haematopoietic lineage choice and differentiation was challenging because of the unattainability of TAL1 knockout mouse models. Therefore, conditional silencing of the *Tal1* gene using an inducible cre/loxP system was applied to study TAL1 in adult haematopoiesis. It showed that TAL1 deletion affected the erythroid and megakaryocytic lineages but not the myeloid and lymphoid (Hall et al., 2003; Mikkola et al., 2003).

Correspondingly, overexpression in primary HSCs supported TAL1 implications in lineage differentiation and its tendency to shift differentiation to favour the erythroid and megakaryocytic lineages. However, the overexpression of TAL1 did not interfere with the specification of the myeloid lineage (Elwood, Zogos, Pereira, Dick, & Begley, 1998; Ravet et al., 2004). TAL1 deletion hindered the expansion of monocyte progenitors, demonstrating TAL1 functional contribution to myeloid lineage (Dey, Soumyadeep, Curtis, Jane, & Brandt, 2010). At the myeloid and lymphoid lineage, transduction of wild type TAL1 into HSCs resulted in a significant increase of myeloid differentiation while dominant-negative TAL1, lacking the bHLH domain, caused an increase in thymocyte differentiation, using the multilineage progenitor (MLP) assay (Kunisato et al., 2004).

Within the lymphoid specification, overexpression of TAL1 has been shown to reduce the specification of B cell lineage in the early pro-B cells but did not interfere with subsequent differentiation (Herblot, Aplan, & Hoang, 2002).

TAL1 transcriptional complex in haematopoiesis:

TAL1, as a basic helix loop helix class II protein, forms a heterodimer with the helix loop helix class I proteins (E-proteins) including HEB, E2A (E12 and E47), and E2-2. The heterodimer binds DNA at E-box motifs with high affinity (Hsu, Cheng, Chen, & Baer, 1991; Hsu et al., 1994; Massari & Murre, 2000). However, the binding of TAL1 and an E-protein heterodimer does not activate transcription on its own. TAL1 can activate gene expression as apart of multi-protein complex primarily consist of LIM only protein, LDB1, and a GATA protein.

The TAL1 pentameric complex was first described in the erythroid lineage and assembled on an E-box-GATA motif (Wadman, Isobel et al., 1994). LMO2 binding to TAL1 strengthens the heterodimer but reduced its binding affinity to the E-box motif. On the other hand, LMO2 links the heterodimer to other transcription factors like LDB1 and GATA proteins (El Omari et al., 2013) The association with LDB1 stabilises the complex and links promoters to distant enhancers via chromatin looping, achieved by multimerizing with other LDB1 proteins (Krivega, Dale, & Dean, 2014). LMO2 interacts with TAL1 at the loop-helix two domains, and this interaction is essential for the nucleation of the complex (Lecuyer et al., 2007). TAL1 binding to LMO2 protects LMO2 from proteasomal degradation. The single-stranded DNA-binding proteins (SSBPs) stabilise LMO2 and its transcriptional complex via its interaction with LDB1 (Gungor et al., 2007).

Characteristics of TAL1 transcriptional complex

The pentameric complex, composed of a TAL1 E-protein heterodimer, an LMO protein, LDB1, and a GATA protein, can associate with additional proteins to differentially regulate gene expression, as shown in **figure 1.5**. First, ChIPseq studies revealed the binding of the TAL1 pentameric transcription complex to DNA regions with proximity to RUNX1, and the ETS transcription factors ERG and FLI1 (Wilson et al., 2010a). Also, the ability of TAL1 to act as a transcriptional activator or suppressor can be a consequence of the transcription complex diversity in interacting proteins. The association of a repressor like ETO2 to TAL1 complex was shown to inhibit TAL1 target genes in the erythroid lineage (Cai et al., 2009, Goardon, Nicolas et al., 2006). TAL1 can also associate with coactivators such as P300 and PCAF or with corepressors like mSin3A, HDAC, and LSD1 (Huang, S. & Brandt, 2000; Huang, Suming, Qiu, Stein, & Brandt, 1999; Huang, Suming, Qiu, Shi, Xu, & Brandt, 2000; Xu, Z., Meng, Cai, Koury, & Brandt, 2006).

Genome-wide studies of TAL1 DNA binding significantly contributed to the knowledge of the TAL1 transcriptional complex. Most of the promoters bound with TAL1 were also bound with HEB or E2A in the Jurkat cell line and primary human T-ALL cells. They showed that TAL1 binding to the promoter region of a gene that was also bound with E2A or HEB could result in either activation or suppression of the gene (Palomero et al., 2006). TAL1 ChIPseq studies were also valuable in identifying determining factors that control the genomic distribution of TAL1 and how it varies based on lineage or differentiation stage. The expression of TAL1 partners in the transcriptional complex was demonstrated to determine TAL1 binding sites. Differential expression of GATA proteins determines TAL1 binding to target genes. TAL1 binding to GATA 1 instead of GATA2 would alter its genomic distribution (Wu et al., 2014). RUNX1/3 and ETS were also shown to be a determinant of TAL1 binding sites in the genome of T-ALL cells (Palii et al., 2011a). These studies feature the value of identifying the characteristics of the TAL1 complex in understanding the diverse roles of TAL1 in normal haematopoiesis and in T-ALL.

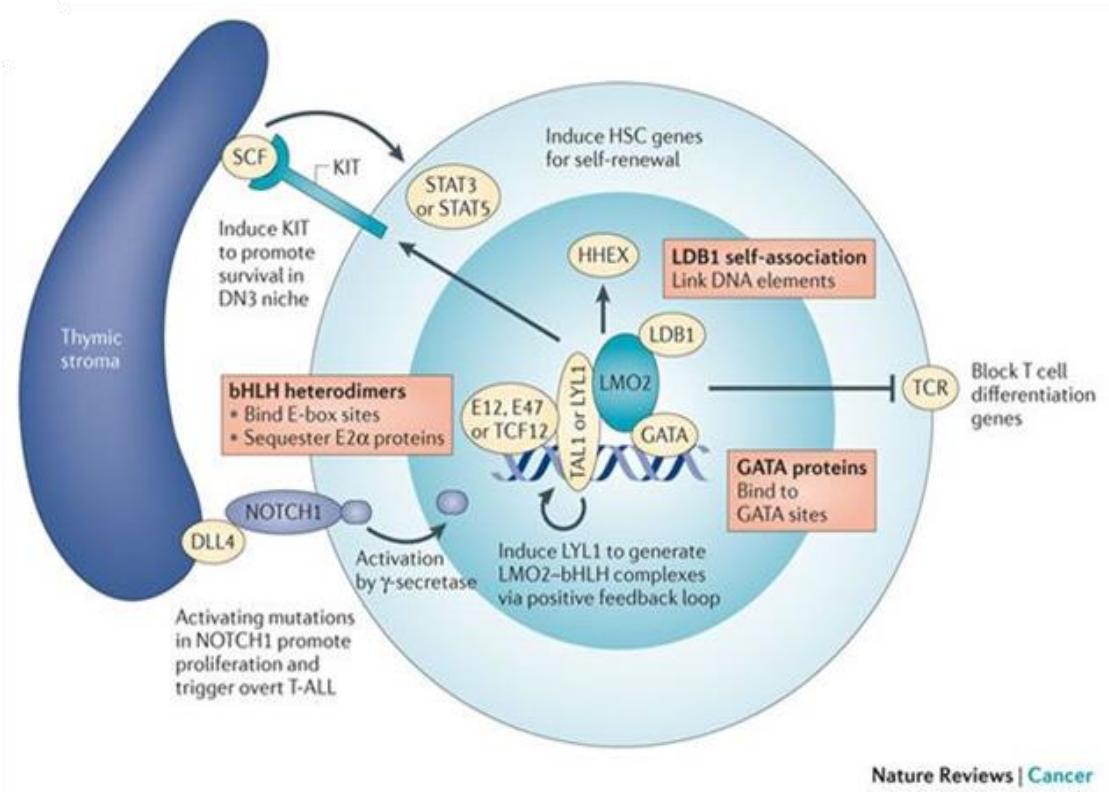


Figure 1.5 TAL1-transcription complex and self-renewal roles

The figure shows the transcription complex containing TAL1 or LYL1, LMO2, GATA, and E-proteins, including E2A (E12, and E47), or TCF12 (HEB) and its role in activating HSCs self-renewal genes, blocking T cell differentiation genes, inducing cKit for DN3 survival. The interaction between the thymic Notch ligand DL4 and the Notch1 receptor on T cell progenitors activates the transcriptional activity of the complex. Figure from Matthews, 2013.

The requirement of the TAL1 DNA binding domain

In haematopoiesis, a mutant TAL1 that was unable to bind DNA was able to reconstitute haematopoiesis in TAL1 null embryonic cells (Porcher et al., 1996). Furthermore, embryos of a murine model, expressing the TAL1 DNA binding mutant, developed further but died at E14.5 during the definitive haematopoiesis stage at the foetal liver (Kassouf, Chagraoui, Vyas, & Porcher, 2008). They demonstrated that TAL1 was dispensable for the specification of HSCs while being required for the maturation of erythroid cells. Similarly, the TAL1 DNA binding domain was not necessary for the regulation of KIT (Lecuyer et al., 2002). On the other hand, the TAL1 DNA binding ability was essential for its role in lineage differentiation, especially during erythroid and megakaryocytic differentiation. Likewise, the regulation of the erythroid gene GPA by TAL1 was shown to require the TAL1 DNA binding domain (Lécuyer & Hoang, 2004; Porcher et al., 1996).

It was suggested that the TAL1 DNA binding mutant could bind DNA and mediate many of the TAL1 functions in haematopoietic development via association with LMO2 and its DNA binding protein interactions, such as GATA (Stanulović, Cauchy, Assi, & Hoogenkamp, 2017). Stanulovic *et al.* showed that at genomic regions occupied with TAL1, LMO2, and LDB1, the GATA motif was highly enriched and much more so than the E-box motif. They proposed that the higher enrichment of GATA indicates its significance in linking the complex to the DNA more than TAL1 direct DNA binding to E-box motifs. They hypothesised that even though the DNA mutant of TAL1 can bind DNA through the other LMO2 complex members, it could be binding with lower affinity and thereby less effective in regulating transcription of some TAL1 target genes like in the erythroid lineage.

Overview of E-proteins

Mammalian E-proteins include E2A, which has two alternatively spliced isoforms E12 and E47, in addition to E2-2 and HEB (Kee, 2009). E2A, E2-2, and HEB are also referred to as TCF3, TCF4, and TCF12, respectively. HEB has two isoforms, termed canonical HEB (HEBcan) and alternative HEB (HEBalt), which are produced due to the availability of two transcription start sites (Wang et al., 2006). E-proteins regulate transcription as

dimers, which can bind DNA at E-box motifs. The E-protein dimer functions mainly as a transcriptional activator by interacting with coactivators, such as p300 (Massari & Murre, 2000). E-proteins can also heterodimerise with class II (basic helix loop helix) bHLH proteins such as TAL1. Heterodimerization of the widely expressed E-proteins with the tissue-specific class II bHLH proteins has been implicated in cell specification and maturation. For example, heterodimers of E2A and the myogenic bHLH II MYOD initiate the myogenic program.

E-protein induction of cell differentiation is associated with a proliferation arrest. E-protein-mediated induction of differentiation on account of proliferation requires a regulatory mechanism that resumes cell proliferation in response to activation signals. Upregulation of Inhibitor of DNA binding proteins (ID) expression is one of the regulatory mechanisms of E-protein function. ID proteins contain an HLH without the basic region, which is critical for protein interactions and binding DNA. Consequently, ID proteins form inactive heterodimers with E-proteins and prevent the formation of DNA binding E-protein homo- and heterodimers (Benezra, Davis, Lockshon, Turner, & Weintraub, 1990; Kee, 2009).

E-protein and ID proteins have been shown to regulate cell cycle inhibitors oppositely and the progression of the cell cycle in many mammalian cells, as reviewed in (Wang, L., Baker, 2015). Expression of ID proteins increases in actively dividing cells mainly to relieve the E-proteins-mediated suppression of the cell cycle (Ling, Kang, & Sun, 2014). E-protein binding to ID proteins results in the inhibition of E-protein-mediated transcriptional activation of cell cycle inhibitors. In T cell development, the inhibition of E-protein function by ID proteins manifests in the switching of E proteins on and off in resting versus proliferating T cell progenitors (Engel, Murre, 2001a). Finally, in addition to ID proteins, E-proteins can be inhibited by the E-box-binding transcription factor ZEB, ubiquitination in response to Notch signalling, suppression via calmodulin and by interacting with TAL1 (Braunstein & Anderson, 2012).

E-proteins in haematopoiesis:

E2A was the only mammalian E-protein to be implicated in HSC self-renewal and to impact myeloid, and erythroid progenitors (Semerad, Mercer, Inlay, Weissman, & Murre,

2009). Recently, downregulation of HEB in hESCs using CRISPR/Cas9 revealed a requirement of HEB in the specification of mesoderm and haematopoiesis (Li et al., 2017). All three mammalian E-proteins were shown to be necessary for the lymphoid lineage (Welinder et al., 2011, Borghesi et al., 2005, Barndt, R. J., Dai & Zhuang, 2000a, Bouderlique et al., 2019). E2A was found to be necessary for the generation of Lymphoid-primed multipotent progenitors (LMPP) (Dias, Månsson, Gurbuxani, Sigvardsson, & Kee, 2008). Target genes of E2A in LMPP include lymphoid related genes, including *Rag1, Notch1, and Il7ra.* Taken together, E2A is recognised as one of the transcription factors controlling T cell differentiation.

In T cells, there is considerable evidence supporting a significant role for E-proteins. The dimer of E2A and HEB was demonstrated to bind E box motifs in T cells (Engel & Murre, 2001; Sawada & Littman, 1993a). First, mutational studies verified the significant role played by E-proteins in establishing T cell development. $E2A^{-/-}$ mice developed lymphoma and showed a partial block at the DN1 stage with a ten-fold reduction in thymic cellularity (Bain et al., 1997a). The partial block might be due to redundancy with HEB in T cells. HEB-null mice showed a more significant impact on T cell development with a decrease in T cells and a later block at the ISP stage (Barndt, R., Dai, & Zhuang, 1999). A dominant-negative HEB was constructed by mutating the basic region of HEB without changing the helix-loop-helix domain, thus leading to the inactivation of HEB and the formation of inactive HEB-E2A heterodimers. The dominant-negative HEB was introduced into cells to study the overlapping roles of HEB and E2A. The study showed that the inactivation of both HEB and E2A resulted in an earlier block in T cell differentiation, as DN3 cells, and a profound reduction in T cell numbers (Barndt, R. J., Dai, & Zhuang, 2000).

E2A levels do not change significantly in ETP, DN2, DN3, and in part of the DN4 population. Then, E2A expression is downregulated in DP and SP T cells (Yui & Rothenberg, 2014). In contrast to E2A, the expression of HEB is differentially regulated during T cell development. Specifically, the HEBalt isoform is expressed at its highest in early T cell progenitors and fully inhibited during the transition into the DP stage. Whereas HEBcan is expressed throughout T cell development with a strong upregulation in DP cells (Wang et al., 2006).

Targets of E-proteins in T cells

E2A was shown to regulate RAG1 and RAG2 in CLP. Several T cell targets of E2A have been described, including Hes1, Deltex1, pTcra, Tcrb and Gfi1b (Kee, 2009).Gfi1b and Gfi1 were shown to be regulated with E2A (Schwartz, Engel, Fallahi-Sichani, Petrie, & Murre, 2006; Xu, W. & Kee, 2007). The regulation of Gfi1 expression by E-proteins was confirmed in murine models, where both *Tcf3* and *Tcf12* were disrupted (Jones & Zhuang, 2007). Furthermore, E2A was shown to regulate *Hes1* and *pTα* in collaboration with NOTCH1. Like E2A, HEB was implicated in the regulation of *pTα* (Takeuchi et al., 2001) *and Tcrb* (Barndt et al., 2000), in addition to CD4 expression (Sawada & Littman, 1993b). Both E2A and HEB were required to block proliferation before *pTα* is expressed. The diverse roles of E2A and HEB represent the vital involvement of E-proteins in the generation of T cells (Wojciechowski, Lai, Kondo, & Zhuang, 2007a).

LYL1 in the haematopoietic system

TAL1 is closely related to other bHLH transcription factors (Giroux et al., 2007). The amino acid sequence of LYL1 is similar to TAL1, especially within the basic helix-loop-helix region, with more than 85% of homology (Baer, 1993). LYL1 was the first bHLH transcription factor to be identified in T-ALL, where the translocation of t(7;19)(q35;p13) was detected leading to the overexpression of LYL1 (Mellentin, Smith, & Cleary, 1989). The expression of LYL1 in haematopoietic cells is detected in HSCs and all haematopoietic lineages (Capron et al., 2006; Visvader, Begley, & Adams, 1991). In T cells, LYL1 is downregulated as T cell differentiation progresses, while its overexpression was shown to induce B and T cell leukaemia in mice (Zhong, Jiang, Hiai, Toyokuni, & Yamada, 2007). Both TAL1 and LYL1 are found ectopically expressed in T-ALL. Moreover, LYL1 has been associated with ETP-ALL in gene expression profiling studies of paediatric T-ALL patient samples.

LYL1 protein interactions with E-proteins and LMO proteins were described before ChIP-sequencing analyses of haematopoietic transcription factors revealed the overlap between LYL1 binding with TAL1, GATA proteins, ETS, and RUNX1 in the HPC7 cell line (Wilson et al., 2010b). The overlap between TAL1 and LYL1 suggests that either can associate in the transcriptional complex with LMO, GATA, LDB1, ETS, and

RUNX1, as shown in the HPC7 cell line and primary erythroid and megakaryocytic cells from mouse HSCs (Pimkin et al., 2014; Wilson et al., 2010b). ChIP-sequencing analysis in HPC7 cells showed that only 30% of TAL1 binding sites were occupied with LYL1 (Wilson et al., 2010). The overlapping and differential binding of TAL1 and LYL1 to their target genes clarifies the parallel and distinct roles.

T-cell Acute Lymphoblastic Leukaemia

T-cell acute lymphoblastic leukaemia (T-ALL) is a heterogenous haematological cancer characterised by the transformation of T cell progenitors (Aifantis, Raetz, & Buonamici, 2008). The malignant transformation of thymocytes is marked with a differentiation block, uncontrolled proliferation, and survival (Teitell & Pandolfi, 2009). In T-ALL malignant lymphoblasts infiltrate the bone marrow, expressing immature T cell markers. T-ALL patients typically present with high white blood count, neutropenia, anaemia, thrombocytopenia, mediastinal thymic masses, and central nervous system meningeal infiltration (Belver & Ferrando, 2016). The prognosis of T-ALL has remarkably improved in children to a 5-year survival around 90%, whereas in adults it barely approaches 40% (Pui & Evans, 2006; Vora et al., 2013). Therefore, there is a pressing need for T-ALL dedicated research considering the relapse rate, the high level of toxicity, and the remaining patients with initial poor prognosis (Girardi, Vicente, Cools, & De Keersmaecker, 2017).

Gene expression and immunophenotypes define T-ALL molecular subgroups, as reviewed in (Durinck et al., 2015). First, an Early T-lineage progenitor (ETP) subgroup, which is the T-ALL that exhibits gene expression and immunophenotypes of $CD4^-$ $CD8^-$ as in early T cell progenitors. Then early cortical and late cortical subgroups. ETP represent up to 50% of adult T-ALL and is associated with poor prognosis. On the other hand, the late cortical subgroup, with the phenotype of $CD1a^+$ $CD4^+$ $CD8^+$, is linked to a more favourable outcome. The third subgroup is the late cortical subgroup, with a relatively mature immunophenotype of $CD4^+$ $CD8^+$ $CD3^+$.

Genetic and molecular pathogenies of T-ALL

In T-ALL genetic abnormalities accumulate, altering the healthy T cell development (Belver & Ferrando, 2016). Several genetic and molecular mechanisms underlying the malignant transformation have been described. Cytogenetic translocations can be found in 50% of T-ALL patients (Aifantis et al., 2008). T cell differentiation is under the control of a network of specific transcription factors (Rothenberg, Ungerbäck, & Champhekar, 2016). The transcriptional regulators orchestrate a series of stage-specific changes in cellular gene expression (Yui & Rothenberg, 2014). Therefore, in T-ALL, the translocations commonly associate oncogenic transcription factor genes to the TCR genes, located on chromosome 7 and chromosome 14 (Aifantis et al., 2008). These translocations can be a result of aberrant V(D)J rearrangements (Larmonie et al., 2013). In addition to translocations, other genetic aberrations include activating deletions, mutations, and duplications.

Common genetic aberrations detected in T-ALL highlight major players in T cell transformation. Genetic alterations in T-ALL generally target genes related to transcription factors, epigenetic factors, cell cycle proteins, signalling pathways, and ribosomal proteins (Belver & Ferrando, 2016). NOTCH1 signalling is a major oncogenic pathway in T-ALL, as more than 60% of patients have activating mutations of NOTCH1 (Weng et al., 2004). Activation of NOTCH signalling can be due to the mutations in the heterodimerization domain, PEST domain, or FBXW7, which facilitates NOTCH1 proteasomal degradation (Weng et al., 2004).

As indicated above, transcription factors are frequent targets of T-ALL genetic aberrations. Transcription factors that are abnormally expressed in T-ALL include the bHLH class II transcription factors like TAL1, LYL1, BHLH1, and LMO proteins, including LMO1 and LMO2. Mostly, the oncogenic transcription factor TAL1 is activated by mutations in T-ALL (Porcher, C., Chagraoui & Kristiansen, 2017). Other aberrantly expressed transcription factors include the homeobox transcription factors, and MYB (Belver & Ferrando, 2016). The genetic aberrations found in T-ALL disrupt the regulatory networks of haematopoiesis and result in the malignant transformation of thymocytes.

TAL1 in T-ALL

Most of the T-ALL patients have an abnormal expression of TAL1. Mostly TAL1 is overexpressed because of a 1p32 small interstitial deletion (SIL-TAL). The deletion fuses the constitutively active SIL promoter and 5'UTR to the complete TAL1 open reading frame, giving rise to SIL-TAL1 (Belver & Ferrando, 2016; Brown et al., 1990; Ferrando et al., 2002). Additionally, mutations upstream of the TAL1 gene were recently found to result in the creation of MYB binding sites and consequently upregulate TAL1 expression (Mansour et al., 2014). Finally, there are additional unknown trans-acting factors that result in TAL1 overexpression (Ferrando et al., 2002).

TAL1 mouse models

Animal models of T-ALL have significantly contributed to our understanding of T-ALL mechanisms. The assumption of an oncogenic role for TAL1 in T-ALL has been controversial in mouse model studies. Most mouse models of TAL1 showed that TAL1 overexpression did not lead to leukaemia on its own, whereas few showed that TAL1 could induce leukaemia after a long latency or in cooperation with another oncogene, as reviewed in (Aifantis et al., 2008). Mouse studies of a *Tal1* transgene under the control of the CD2 enhancer showed that TAL1 transgenic mice did not develop leukaemia (Curtis, David J., Robb, Strasser, & Begley, 1997; Larson et al., 1996; Robb, Rasko, Bath, Strasser, & Begley, 1995). However, using Lck-Tal1, two transgenic lines developed lymphoma with an incidence of 60% and 66.7% (Condorelli et al., 1996). Similarly, another study demonstrated lymphoma in Lck-Tal1 mice with 28% incidence rate (Kelliher, Seldin, & Leder, 1996). Since the SIL-TAL deletion often induces TAL1 overexpression, mouse models with the *Tal1* gene fused to the *Sil* regulatory region were studied. The SIL-TAL transgenic mice did not develop leukaemia during the study period of 12 months (Aplan et al., 1997a). The 12 months duration was possibly not enough to rule out the development of T-ALL after long latency. Moreover, SIL-TAL transgenic mouse models were generated (Cheng, Zhang, Slape, & Aplan, 2007). They proposed that using SIL-TAL would mimic the level and timing of overexpression in TAL1 in T-ALL patients, which could explain the inability of TAL1 to cause lymphoma in some of the transgenic models. The conditional SIL-TAL transgenic mice had similar survival rates to the controls within the 19 months of the study period. However, they showed an

increase in $CD4^-$ $CD8^-$ thymocyte with a reduction in $CD4^+$ and $CD8^+$ T cells in the spleen (Cheng et al., 2007).

Other mouse model studies that did not show leukaemia were able to show the requirement of additional oncogenic events to induce T-ALL as seen in patients. The CD2-TAL1 induced leukaemia in collaboration with LMO2 (Larson et al., 1996). Secondly, leukaemia in Lck-Tal1 models was accelerated by Casein kinase II (CKII) (Kelliher, Seldin & Leder, 1996b). SIL-TAL1 transgenic mice developed aggressive leukaemia after they were crossed with mice having LMO1 overexpression (Aplan et al., 1997b).

TAL1 and T cell differentiation

TAL1 overexpression was shown to mediate changes in T cell differentiation. The effect on T cell differentiation markers was studied in TAL1 transgenic mouse models. First, Lck-Tal1 transgenic mice showed a three-fold increase in DN cells and ISP CD8 cell numbers (O'Neil, Shank, Cusson, Murre, & Kelliher, 2004). Also, TAL1 overexpression caused a 50% decrease in the overall thymocyte count, three to four-fold reduction in DP cells, and a decrease in CD4 single-positive cells. The effect of TAL1 on T cell differentiation was augmented when it was combined with E-protein inhibition (O'Neil et al., 2004) or with LMO1/2 overexpression (Aplan et al., 1997; Gerby et al., 2014a; Herblot, Steff, Hugo, Aplan, & Hoang, 2000; Tremblay et al., 2010). TAL1 and LMO1/2 promote leukaemia by inducing self-renewal characteristics in thymocyte progenitors as demonstrated using serial transplantations. Additionally, NOTCH1 collaborates with TAL1 and LMO1/2 in inducing self-renewal characteristics during T cell reprogramming (Gerby et al., 2014b; Tatarek et al., 2011a). Finally, the changes in T cell differentiation in murine models of T-ALL oncogenes depict the mechanisms of their oncogenic role during T cell transformation.

TAL1 contribution to T-ALL

There are two aspects of the contribution of TAL1 to T-ALL pathogenesis. The two proposed models are not necessarily mutually exclusive. First, the TAL1 transcriptional complex directly regulates gene expression and facilitates thymocyte transformation. LMO proteins LMO1 or LMO2, are found aberrantly expressed in 80% of TAL1 overexpressing T-ALL (Ferrando et al., 2002; Wadman et al., 1994). In T-ALL, the aberrantly expressed TAL1 and LMO1/2 interact with E-proteins and can assemble on a bipartite Ebox-E-box motif (Grutz et al., 1998; Lécuyer & Hoang, 2004). Also, TAL1, LMO1/2 and LDB1 were shown to associate with GATA3, RUNX1, and ETS, which are normally expressed in T cells (Sanda et al., 2012a).

Several direct TAL1 target genes in T-ALL were identified, including TALL1, RALDH2, and NKX3.1 (Kusy et al., 2010; Ono, Fukuhara, & Yoshie, 1997; Ono, Fukuhara, & Yoshie, 1998). The involvement of TALL1, RALDH2 in T-ALL pathogenesis is not well understood, but NKX3.1 was shown to be required for proliferation in TAL1 overexpressing T-ALL. Moreover, the list has significantly expanded with TAL1 target genes identified using genome-wide assays complemented with TAL1 knockdown in T-ALL cell lines. Overall, genome-wide analyses showed that TAL1-mediated regulation of gene expression inhibits T cell differentiation and enforces antiapoptotic, and self-renewal programs (Palii et al., 2011; Palomero et al., 2006; Sanda et al., 2012).

The second mechanism for TAL1 contribution in T-ALL is the formation of heterodimers with E-proteins as an indirect consequence of TAL1 aberrant expression. E-proteins regulate T cell differentiation by activating T cell-specific genes (Kee, 2009). Consequently, TAL1 heterodimerization with E-proteins disrupts the normal gene regulation mediated by E-proteins (Chervinsky et al., 1999; Goardon et al., 2002; Herblot et al., 2000; O'Neil et al., 2004; Sanda et al., 2012). Specifically, T cell differentiation genes regulated by E-protein dimers are suppressed, resulting in a block of differentiation. The inhibition can be explained by TAL1 sequestering E-proteins from binding its target genes in a DNA binding-independent manner. Therefore, TAL1 DNA independent inhibitory effects on E-proteins function could include relocating E-proteins to the binding sites of the TAL1 complex and possibly interfering with the recruitment of chromatin modifiers by E-proteins. However, there is also evidence that TAL1/E-protein

heterodimers bind E-protein target genes but alter the E-protein mediated transcriptional regulation.

E-protein knockout models supported TAL1 inhibition of E-proteins function. E-protein null models showed a block of differentiation at the DN/DP stage before developing T-ALL (Braunstein & Anderson, 2012; Wojciechowski, Lai, Kondo, & Zhuang, 2007b). E2A knockout models developed tumours within three to nine months in αβ thymocytes (Bain et al., 1997; Yan et al., 1997) and the ectopic expression TAL1 accelerated the disease in E-protein knockout mice (O'Neil et al., 2004). Furthermore, it was later shown that the *pTα* gene is regulated by E-proteins E2A and HEB and suppressed by TAL1 (Tremblay, Herblot, Lecuyer, & Hoang, 2003). On the other hand, HEB null mice showed abnormal T cell differentiation with an increase in ISP cells. (Braunstein & Anderson, 2010; Wojciechowski et al., 2007). While TAL1 was shown to inhibit the *pTα* gene, the preTCR, and TCR signalling was found to be required in T-ALL, as shown in TAL1 LMO1 mouse models.

The requirement of the TAL1 DNA binding domain in T-ALL was investigated using a DNA binding mutant of TAL1. The study showed that the DNA binding mutant developed leukaemia (Draheim et al., 2011; O'Neil, Billa, Oikemus, & Kelliher, 2001). DNA independent functions of TAL1 could include sequestering, relocating E-proteins by the TAL1 transcriptional complex, thus inhibiting E-protein mediated transcription.

TAL1 isoforms

TAL1 isoforms were first studied when a subtype of T-ALL was found to be expressing only the short isoform of TAL1 (Aplan et al., 1997). Post-transcriptional regulation of *Tal1* results in four isoforms. The four isoform results from four different translational start sites. An upstream open reading frame (uORF) was shown to regulate the choice of the translational start site and thus controlling TAL1 isoforms (Calkhoven et al., 2003). They showed that the initiation factor 2 (eIF2) controlled the TAL1 isoform ratio and can be modulated by translation initiation drugs, the mTOR inhibitor Rapamycin and the eIF2α-kinases inhibitor 2-amino purine (2AP). They showed that Rapamycin suppressed the short TAL1 isoforms while 2AP suppressed the long isoforms.

Furthermore, alternative splicing was found to be another mechanism of controlling TAL1 isoforms. The shorter isoforms can be produced via exon skipping or alternative spicing, and this can be regulated by the splicing factor, SF3B1 (Jin et al., 2017a). They demonstrated that RPM15 recruits SF3B1 to TAL1 transcripts and mediates splicing, producing the shorter isoforms. Furthermore, they showed that PRMT1 facilitates the ubiquitination of RPM15 and thereby regulates TAL1 splicing (Zhang et al., 2015).

Knowledge about cellular mechanisms regulating TAL1 expression and function links the upstream regulatory network to functional roles of TAL1 protein. TAL1 protein structure hints at functional differences of TAL1 isoforms. The TAL1 protein contains an N-terminal proline-rich domain for transcription activation (transactivation domain), a positively charged basic region for DNA binding and a helix loop helix domain for protein interactions. Two phosphorylation sites were located within the transactivation domain, as shown in **figure 1.6**. First, ser122 can be phosphorylated to promote transcriptional activation by ERK (Tang, Arbiser, & Brandt, 2002; Wadman, I. A., Hsu, Cobb, & Baer, 1994). ERK-mediated phosphorylation of TAL1 was shown to be activated by NOTCH3 through induction of pre-TCR expression. Remarkably, in T-ALL patients NOTCH3 is overexpressed in almost all cases, even though it is rarely mutated (Bellavia et al., 2002; Bernasconi-Elias et al., 2016; Talora et al., 2006). The second is the ser172 residue, which can be phosphorylated by cAMP-dependent Protein Kinase (PKA) affecting TAL1 DNA binding to E-box motifs (Prasad & Brandt, 1997). Another phosphorylation site within TAL1 that has been described, at Thr90 by Akt, has been shown to alter TAL1

suppression of the EPB42 gene (Palamarchuk et al., 2005). Differences in phosphorylation sites and regulatory signals indicate differential regulation of TAL1 isoforms and suggest probable effects on their function.

The TAL1 isoforms have been implicated in haematopoietic lineage determination. The ectopic expression of isoforms of TAL1 in mouse bone marrow cells, which were cultured in supplemented methylcellulose, showed that the long isoform forced higher numbers of megakaryocyte colonies. On the other hand, with the shorter isoform, multipotent cells had a higher tendency to form erythroid colonies (CFU-E), as demonstrated by higher B-globin mRNA expression (Calkhoven et al., 2003). Induction of the erythroid lineage by the TAL1 truncated isoform was consistent with experiments studying the effects of overexpressing TAL1 isoforms in human CD34. After ten days of culture, the cells with the short isoform had higher expression of the erythroid precursor marker CD71 compared to the long isoform (Jin et al., 2017a). Interestingly, TAL1 isoform differential regulation of human T-cell differentiation or within the myeloid lineage has not been investigated yet.

Nevertheless, TAL1 long and short isoforms were co-expressed with LMO1 in double transgenic mice, and the data showed they both similarly induced leukaemia (Aplan et al., 1997; Ellsworth & Aplan, 1999). Subsequently, the short isoform vector, pSIL/TSCL, was used in investigating TAL1 role in several T-ALL mouse studies (Gerby et al., 2014; Tatarek et al., 2011b; Tremblay et al., 2010). Even though the TAL1 long and short isoforms both appear to retain the ability to cause T-ALL leukaemia, the cellular mechanism of TAL1 isoforms might still be different in protein-protein interaction, DNA binding, regulation of transcription and gene expression. The potential differences may be valuable for novel therapies or as prognostic markers.

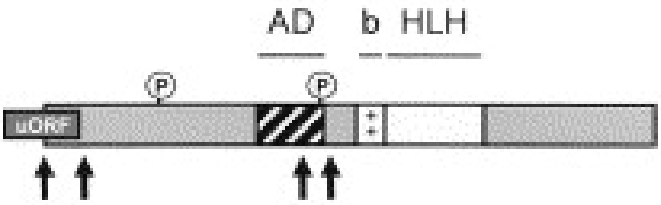


Figure 1.6: TAL1 (SCL) protein structure:
Illustration of TAL1 protein showing an Upstream Open Reading Frame (uORF), a putative proline-rich transactivation domain (AD), basic positively charged region and the helix loop helix domain (bHLH). Arrows points translation sites of TAL1 different isoforms. Figure adapted from Lécuyer, Eric (2004).

Methods for studying transcription factors (TF)

Transcription factors (TF) bind DNA at specific sequences or cis-regulatory elements and facilitate gene transcription (Lee & Young, 2013) via association with the transcription apparatus, including general transcription factors and RNA polymerase II, and the mediator complex (Istrail & Davidson, 2005). TF mostly modulate gene expression by binding promoters and enhancers. In addition to the recruitment or modulation of transcription apparatus, transcription factors can regulate gene expression through the recruitment of chromatin modifiers or by physically linking distal cis-regulatory elements in distant enhancers to proximal promoters via DNA-looping (Lee & Young, 2013). One primary approach to understanding the role of a given transcription factor is identifying its specific binding sites within the genome, either by assaying specific sites or with genome-wide approaches. Several techniques have been applied to study transcription factors binding to DNA.

A popular in vitro approach for the identification of TF binding at smaller DNA regions (typically 20-100bp) is the electrophoretic mobility-shift assay (EMSA). This is based on the fact that protein binding to DNA affects its migration through the gel (Watson, 2004). In this assay, labelled DNA is incubated with a protein of interest. The protein binding to the DNA fragment slows down the migration of the DNA fragment. Incubation with specific antibodies can create a further reduction in migration of the protein/DNA complex (supershift), providing identification of the DNA binding protein. The second type of techniques to study DNA-protein interactions is DNA footprinting. The principle of DNA footprinting is based on that proteins binding shields the DNA from nucleases such as DNAse I or chemical treatments (Lambert et al., 2018; Watson, 2004). This can be performed *in vitro* or *in vivo* and can identify TF binding over regions of 200-300bp per assay. Chromatin immunoprecipitation (ChIP) is used to identify proteins binding to DNA in the cell. Proteins are cross-linked to DNA using formaldehyde followed with cell lysis and sonication to produce 200-300 bp DNA fragments. Then, the DNA fragments bound with the protein of interest are immunoprecipitated using a specific antibody. The immunoprecipitated DNA can be further analysed using qPCR, microarray, or next-generation sequencing (Lambert et al., 2018; Watson, 2004). Recent advances in DNA sequencing permitted the study of transcription factors binding sites genome-wide using

chromatin immunoprecipitation sequencing (ChIP-seq). The reduction of DNA sequencing cost and the increase in capacity and availability has enhanced the efficiency of ChIP-seq applications (Mundade, Ozer, Wei, Prabhu, & Lu, 2014).

ChIP-seq applications mainly require two types of controls. The first type of control is using PCR primers targeting a region that is known to be bound by the protein of interest (positive control) in the cells used and primers for a genomic region that is known not to be bound by the protein (negative control). The second type of control is the input control, which is a sample of the DNA before performing the immunoprecipitation (Watson, 2004). Finally, digital footprinting which uses a bioinformatics approach to improve the identification of transcription factors footprint based on DNAse1 treatment, followed by genome-wide sequencing, has been described (Piper et al., 2013).

Aims and Objectives:

We aimed to study how the overexpression of TAL1 and its isoforms contributes to T-ALL pathogenesis.

- **Aim1:** Characterise the TAL1 transcriptional complex containing HEB and how it contributes to T-ALL pathogenesis.
    - To identify TAL1 protein interactions in T-ALL using co-immunoprecipitation and mass spectrometry.
    - To characterise the genomic occupancy of the TAL1 transcriptional complex in T-ALL by comparing ChIPseq data of members of TAL1 transcriptional complex.
- **Aim2:** The differential roles and characteristics of TAL1 isoforms in T-ALL.
    - To construct plasmids expressing a FLAG-tagged TAL1 isoform with the fluorescent protein tdTomato.
    - To evaluate differences between TAL1 isoforms in normal T cell development and T-ALL in:
        - Genomic distribution.
        - Regulation of gene expression.
        - Effect on T cells differentiation.

CHAPTER TWO

Material and Methods

Tomato-T2A-FLAG-TAL1 plasmids

PCR was performed using human TAL1 cDNA clone TalM1/pTM3326 (Hsu et al., 1994) as a template, kindly provided by Professor T.H. Rabbitts (MRC Weatherall Institute of Molecular Medicine, UK) to synthesise four isoforms of TAL1 using oligonucleotides designed to amplify TAL1 with the addition of an *EcoR*I site and FLAG sequence to the forward primers and a *BamH*I site added to the reverse primer. Primers sequences are listed in **table 2.1**.

The PCR mix included 10 µM of the forward and the reverse primers (Sigma), a 1 ng of the template, 1.1x of ReddyMix PCR Master Mix (Thermofisher), 5% DMSO. The program used for the PCR reaction in the thermocycler starts with two minutes at 94 °C for initial denaturation followed with 11 cycles of denaturation at 94 °C for 20 sec, annealing at 50 °C for 30 sec, and extension at 72 °C for one minute. After that, 24 cycles of denaturation at 94 °C for 20 sec, then annealing at 65 °C for 30 sec, and extension at 72 °C. The final step was an extension at 72 °C for two minutes. The PCR products were cloned into pDrive using UA cloning with QIAGEN PCR Cloning Kit. TAL1 was then sub-cloned into PLVX Tet one puro (Clontech), using *EcoR*I and *BamH*I.

Restriction enzymes reaction

Restriction enzyme sub-cloning was performed by incubating with Thermofisher FastDigest enzymes and 1X of the FastDigest buffer with plasmid overnight in a water bath at 37 °C. The digested plasmid was loaded in 1% agarose gel with 0.2 µg/mL EtBr. DNA bands were visualised using Gel Doc™ XR+ System (Bio-Rad). DNA was isolated using gel purification method with NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel).

DNA insert and vector were ligated using 2X Rapid Ligation Buffer (Promega) at room temperature (RT) for at least six hours or overnight. The ligation reaction was transformed into Alpha-Select Gold Efficiency competent cells (Bioline). Minipreps were prepared to isolate DNA from colonies and select clones carrying plasmids containing the insert as described before (Zhou, Yang, & Jong, 1990). Clones were checked by restriction enzymes and were verified by Sanger sequencing (Source BioScience). The validated

plasmid clone was propagated for maxi-prep plasmid purification using NucleoBond Xtra Maxi kit (Machery-Nagel).

The colour tag, tdTomato, followed by the peptide cleavage sequence T2A, was cloned in front of TAL1 open reading frame (ORF) 2, 3, and 4 PLVX vectors using NEBuilder HiFi DNA Assembly cloning kit, following the manufacturer's protocol. HiFi assembly primers were designed using NEBuilder Assembly Tool, and the restriction enzyme *EcoR*I was used to linearise the plasmid. A summary of the cloning steps is shown in **figure 2.1.**

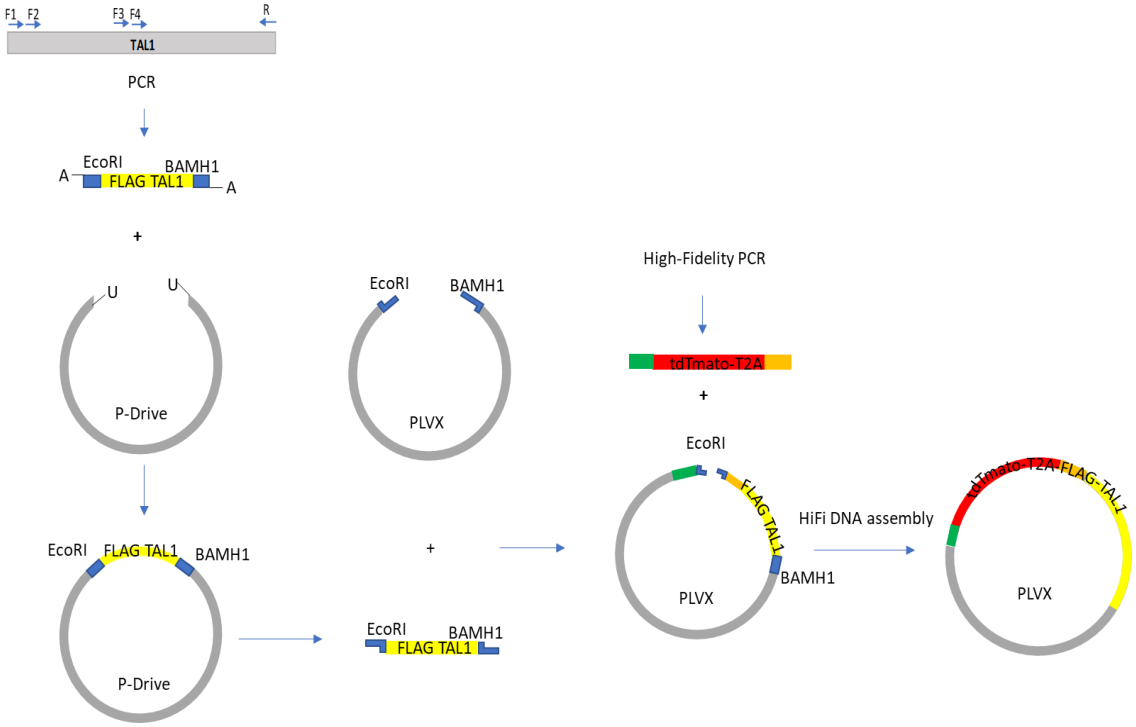


Figure 2.1. Illustration of the cloning process for TAL1 plasmids

The top left corner shows the design of PCR primers to amplify EcoR1-FLAG-TAL1-*BamHI* for the four isoforms of TAL1, and the UA cloning process used to make FLAG-TAL1-Pdrive plasmid followed with sub-cloning into PLVX and the addition of MyrtdTomato via HiFi DNA assembly.

TAL1 piggyBac plasmids:

The piggyBac mediated transposition plasmid PBCAG-eGFP was purchased from Addgene (plasmid # 40973). The plasmid was digested with *Age*I and *Bgl*II to remove eGFP and replace it with Tomato-T2A. First, Tomato-T2A was extracted from the Tomato-T2A-FLG-TAL1 PLVX plasmid using *Age*I and *Bgl*II and cloned to make PBCAG Tomato-T2A vector. Then, The FLAG TAL1 ORF1 was cut out using two *BamH*I sites from FLAG TAL ORF1 P-drive, whereas ORF2, ORF3, and ORF4 were cut out of their PLVX vectors using *Bgl*II and *BamH*I enzymes. The PBCAG Tomato-T2A vector was linearized using in the *Bgl*II site. At the end of the restriction reaction incubation, Calf Intestinal Alkaline Phosphatase (CIAP) from Promega was added to the restriction reaction and incubated for 30 minutes at 37 °C. FLAG TAL1 ORFs were ligated to the *Bgl*II- linearized vector (**Figure 2.2**).

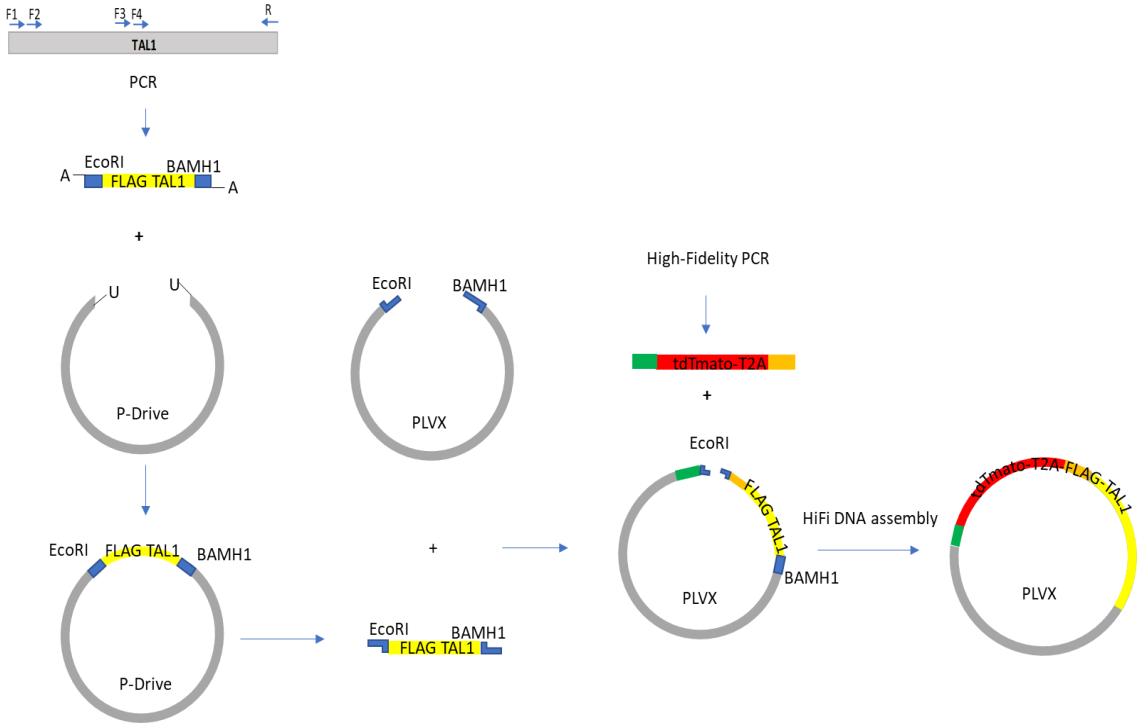


Figure 2.2 Schematic illustration of the cloning steps of PBCAG-tdTomato-TAL1 plasmids

The middle series of steps shows how tdTomato-T2A was moved from PLVX into PBCAG after removing eGFP to make tdTomato-T2A-PBCAG. The bottom centre section shows the subcloning of FLAG-TAL1-ORF1 into tdTomato-T2A-PBCAG from P-drive while ORF2, 3, and 4 subcloning from PLVX is shown at the top.

To correct a deletion in the FLAG sequence of PBCAG Tomato-T2A-FLAG-TAL1 ORF1, we amplified Tomato-T2A-FLAG sequence flanked with *EcoR*I, ligated into the pGEM-T vector via TA cloning. *EcoR*I sites were used to move Tomato-T2A-FLAG from the pGEM-T vector into PBCAG Tomato-T2A-FLAG-TAL1 ORF1, correcting its faulty Tomato-T2A-FLAG (**Figure 2.3**). The features of the generated plasmids are shown in the plasmid maps for ORF1 and ORF4 (**Figure 2.4**).

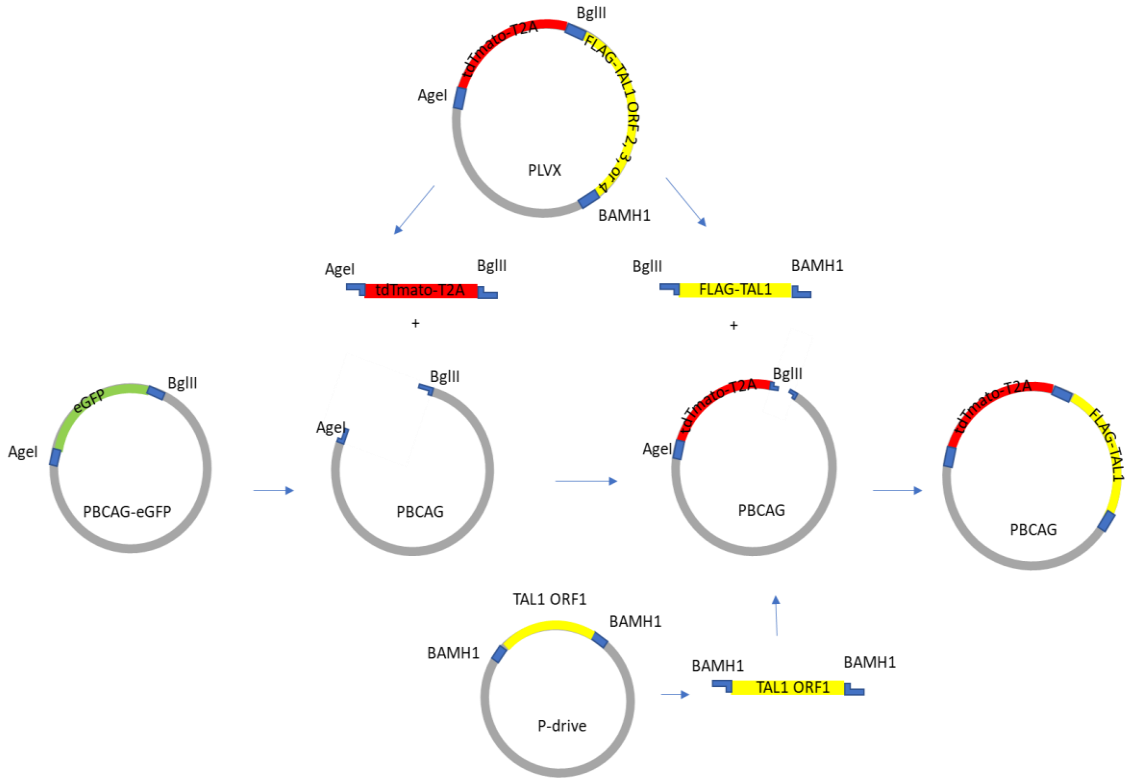


Figure 2.3 Schematic illustration demonstrating the process of cloning the TAL1-ORF1-PBCAG plasmid

tdTomato-T2A-FLAG sequence was amplified to rectify a deletion found in the TAL1-ORF1-PBCAG plasmid. In the top left corner, the PCR was cloned into pGEM-T easy using the TA cloning followed with subcloning into TAL1 ORF1 PBCAG plasmid correcting its faulty FLAG tag.

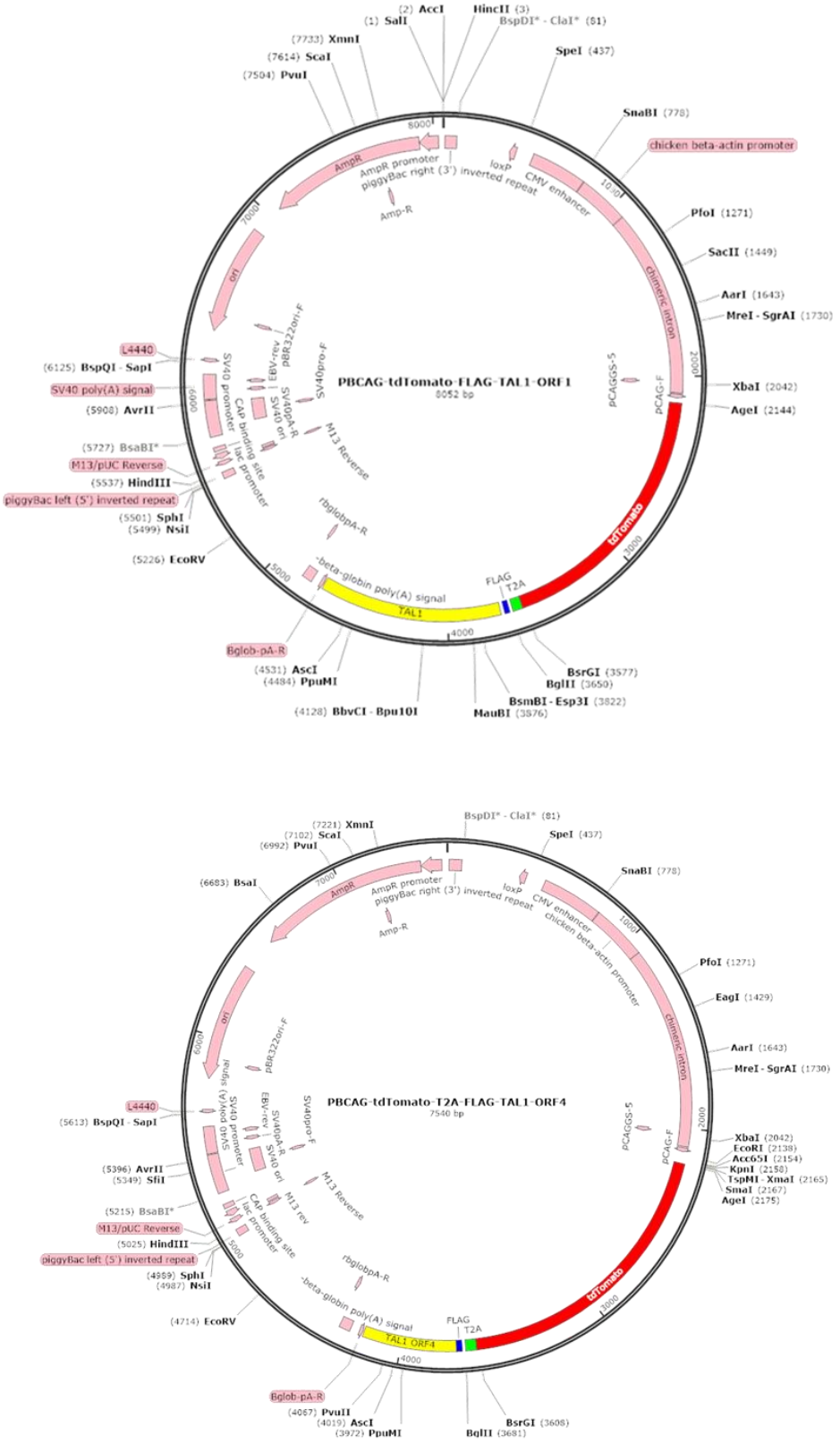


Figure 2.4 Maps of tdTomato-T2A-FLAG-TAL1

Maps of tdTomato-T2A-FLAG-TAL1 ORF1 (top) and ORF4 (bottom) illustrating key features of the constructed plasmids and unique restriction enzyme sites.

Table 2.1 Sequences of primers and oligos

Oligos /primers name	Sequence
*EcoR*I FLAG TAL1 ORF1 FWD	GAATTCATGGACTACAAAGACGATGACGACAAGACCGAGCGGCCGCCGAGC
*EcoR*I FLAG TAL1 ORF2 FWD	GAATTCATGGACTACAAAGACGATGACGACAAGGCCCCCCCGCACCTGGTC
*EcoR*I FLAG TAL1 ORF3 FWD	GAATTCATGGACTACAAAGACGATGACGACAAGGTGCAGCTGAGTCCTCCCGC
EcoRI FLAG TAL1 ORF1 FWD	GAATTCATGGACTACAAAGACGATGACGACAAGGAGATTACTGATGGTCCCCACACCAAAGTTG
*BamH*I TAL1 REV	GGATCCTCACCGAGGGCCGGCTCC
Tomato PLVX HiFi assembly FWD	CACTTCCTACCCTCGTAAAGCTCGAGCCTCTAGAGCCAC
Tomato PLVX HiFi assembly REV	CATCGTCTTTGTAGTCCATGGTGATATTCAGATCTCCTGGGC
*EcoR*I Tomato T2A FLAG FWD	GAATTCACCGGTCGCCTCCA
*EcoR*I Tomato T2A FLAG REV	GAATTCCCTTGTCGTCATCGT

Cell line culture

T-ALL cell lines ARR, DU.528, HSB-2, and CCRF-CEM were cultured in RPMI-1640 medium, supplemented with 10% fetal bovine serum (FBS; Gibco), 2 mM glutaMAX (Gibco), 1% penicillin/streptomycin (Gibco) and 0.075 mM α-monothioglycerol (MTG; Sigma-Aldrich). Every other day the cells were spun down and resuspended in fresh media, maintaining cell numbers between 0.3 - $2x10^6$. OP9-DL4, kindly provided by Professor A. Cumano (Pasteur Institute, Paris), were cultured in α-MEM medium (Gibco), supplemented with 20% FBS, 1% GlutaMAX, and 1% Penicillin/Streptomycin. Every four days, the cells were washed with PBS, trypsinised and split to maintain the culture at less than 40% confluency. HEK-293 cells were cultured in DMEM, supplemented with 10% FBS, 2 mM GlutaMAX, and 1% Penicillin/Streptomycin.

Cord blood processing and CD34 isolation

Upon receipt, cord blood was diluted with an equal volume of PBS, 20 ml suspension was carefully layered on 15 ml Lymphoprep™ (StemCell Technologies) and centrifuged at 750 g for 30 minutes without the brake. Mononuclear cells were collected in a 50 ml tube and washed twice with 3X PBS at 500 g for 10 minutes. Red blood cells were lysed using 1x Lysing buffer (BD Biosciences) at RT for 10 minutes. Human $CD34^+$ cells were isolated using a human CD34 MicroBead Kit (Miltenyi Biotec), according to the manufacturer's protocol. Briefly, per cord blood sample, the mononuclear cells were resuspended in MACS buffer (PBS, 0.5% BSA, 2 mM EDTA) up to 6×10^9 cells/ml and incubated with 20% FcR Blocking Reagent and CD34 MicroBeads for 30 minutes at 4 °C. Cell-bead mixture was washed with ice-cold MACs buffer and loaded onto LS columns (Miltenyi Biotec) mounted on a magnetic rack (QuadroMACS™ Separator from Miltenyi Biotec). Cells were washed on the column three times with MACS buffer. The CD34 enriched cells were eluted in MACS buffer using a plunger.

Electroporation

After isolation, $CD34^+$ cells were pelleted at 200g for 10 minutes. Using the human CD34+ Cell Nucleofector Kit (VPA-1003; Lonza), $CD34^+$ cells were resuspended at 1-$3x10^5$ cells/100 µl Nucleofector Solution (Lonza), containing 4 µg PB vector and since

the PB mediated transposition requires transposase, 2 µg transposase plasmid, and electroporated by an Amaxa™ Nucleofector™ II (Lonza), using the U-008 program. Cells were then diluted in culture media, which was pre-incubated in the tissue culture incubator.

T-ALL cells were pelleted with centrifugation at 90g for 10 minutes following Amaxa Nucleofector kit protocol C (VCA-1004). Cells were resuspended at $2x10^6$ cells/100 µl Nucleofector Solution containing 8 µg of PB plasmid and 2 µg transposase. The cells were electroporated with program X001 settings using the cuvette provided.

Alternatively, T-ALL cell lines were electroporated using a Gene Pulser Xcell (Bio-Rad) device. Cells were pelleted at 300g for 5 minutes and resuspended at 10^7 cells/ml in Opti-MEM I (brand). 10 µg of PB plasmid and 1 µg of transposase were added to the cell suspension. In 4 mm cuvettes, the cells were electroporated at 300 V and 1 mF. Electroporated cells were immediately transferred to a six-well plate with full culture media, which was pre-incubated in the tissue culture incubator. After 24 hours, fluorescent microscopy was used to determine transfection efficiency.

In vitro T cell differentiation

$CD34^+$ cells were co-cultured on 80% confluent OP9-DL4 cells. The OP9-DL4 media was supplemented with 5 ng/ml human FLT3L (Peprotech) and 5 ng/ml human IL7 (Peprotech). The medium of the OP9-DL4 co-culture was changed every two to three days, and at every other medium change, the T cells were removed from the OP9-DL4 by pipetting vigorously without extensively disrupting the OP9-DL4 layer. Then the collected cell suspension was passed through a 70 µm cell strainer to remove any OP9-DL4 clumps and transferred onto a new OP9-DL4 monolayer.

Flow cytometry

The expression of surface markers for haematopoietic differentiation was monitored using conjugated antibodies. Cells were washed with PBS and resuspended in MACS buffer, with 1:200 of the chosen conjugated antibody. After 30 min incubation on ice, the cells were washed with MACS buffer and analysed using either a CyanADP (Beckman Coulter) or BD LSRFortessa (BD Biosciences) flow cytometer. In HEB knockdown

experiments cells were fixed with 4% formaldehyde and permeabilised with 0.5% Triton X-100 and 1% FBS in PBS for 30 minutes before incubating with the primary antibody overnight. After washing with PBS, the samples were incubated with the secondary antibody for 30 minutes followed with two washes with PBS. Cells were analysed on a CyanADP or BD LSRFortessa flow cytometer.

Fluorescence-activated cell sorting (FACS)

Cells from the in vitro human T-cell differentiation were sorted at specific time points for RNA isolation and subsequent RNA sequencing. $CD34^+$ cells were sorted by FACS after an initial enrichment by MACS purification as described above. Human T cells were sorted at specific stages of differentiation. Cells with the $CD7^+$; $CD5^-$; $CD1^-$ immunophenotype were sorted at day 7, $CD7^+$; $CD5^+$; $CD1^-$ at day 14, and $CD7^+$; $CD5^+$; $CD1^+$; $CD3^-$ at day 21 Purification of cell populations was performed using a BD FACSAria Fusion (BD Biosciences) cell sorter. T-ALL cell lines transfected with the TAL1 ORF PiggyBac constructs expressed TdTomato, which was used to sort stably transfected cells seven days post-transcription. TAL1 stable cell lines were sorted a second time to achieve a pure population before RNA and chromatin isolation.

RNA isolation and RNAseq

FACS sorted cells were lysed, and total RNA was extracted using a NucleoSpin RNA kit (Machery-Nagel), according to the manufacturer's protocol. A TruSeq Stranded mRNA Sample Preparation Kit (Illumina) was used to prepare RNAseq libraries, which were sequenced on a Nextseq 500 sequencer (Illumina).

Chromatin Immunoprecipitation (ChIP) and ChIPseq:

T-ALL cells were crosslinked with formaldehyde at a final concentration of 1% for 12 minutes at RT. To stop crosslinking glycine was added at a final concentration of 0.4M. Fixed cells were pelleted by centrifugation at 300 g for 5 minutes at 4 °C and washed with cold PBS. Nuclei were isolated by resuspending the cells in Buffer A (10mM HEPES pH 8.0, 10mM EDTA, 0.5mM EGTA, 0.25% Triton X-100, 1:1000 dilution of

protease inhibitor cocktail (Roche), 0.1 mM PMSF) at $5x10^6$ cells/ml and incubating for 10 minutes on a rotating wheel at 4 °C, followed by centrifugation at 500g for 5 minutes at 4 °C . The nuclei were further purified by incubating with Buffer B (10mM HEPES pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.01% Triton X-100, PIC 1:1000 dilution, 0.1mM PMSF) at $5x10^6$ cells/ml for 10 minutes on a rotating wheel at 4 °C, followed by and centrifugation at 500g for 5 minutes at 4 °C. The cross-linked chromatin was sonicated at $3.3X10^7$ cells/ml in IP Buffer I (25mM Tris HCl, pH 8.0, 150mM NaCl, 2mM EDTA, 1% Triton X-100, 0.25% SDS, PIC 1:1000 dilution 0.1mM PMSF) for 20 cycles of 30 sec on 30 sec off in an ice water bath, using a Bioruptor sonicator (Diagenode). After sonication, the chromatin samples were centrifuged at 16000g for 10 minutes at 4°C. The supernatant was collected and diluted with two volumes of ice-cold IP buffer II (25mM Tris-HCl, pH 8.0, 150mM NaCl, 2 mM EDTA, 1% Triton X-100, 7.5% Glycerol, PIC 1:1000 dilution, 0.1mM PMSF). Chromatin samples were frozen for storage at -80 °C.

Proteins crosslinked to chromatin were isolated using protein G Dynabeads (Life Technologies) coated specific antibodies (1.5 µg antibody/10 µl of beads; see **table 2.2**). Beads were washed 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), ChIP buffer II (20 mM Tris, pH 8.0, 500 mM NaCl, 2 mM EDTA, pH 8.0, 1 % TritonX100, 0.1 % SDS), and with TE/NaCl buffer (TE, pH 8.0, and 50 mM NaCl), twice for each buffer. The bound chromatin was eluted from the beads using elution buffer (100 mM NaHCO3, 1 % SDS, 0.2 M NaCl, 250 µg proteinase K). Eluates were incubated overnight at 65 °C to reverse crosslinking, after which chromatin DNA was purified using Ampure PCR purification beads (Beckman Coulter) and analysed by qPCR. ChIPseq library preparation was conducted following the Illumina protocol and sequenced using an Illumina Nextseq 500 sequencer.

Methylated DNA immunoprecipitation (MeDIP)

T-ALL cells were lysed in lysis buffer (20mM Tris pH8, 4 mM EDTA, 20 mM NaCl, 1% SDS). Genomic DNA was collected using Phenol/Chloroform DNA extraction. Briefly, 10 million cells were washed with PBS, lysed with 2X lysis buffer made up of 20 mM Tris pH 8, 4mM EDTA, 20mM NaCl, 1% SDS at $2X10^6$ cells/ml. Lysed cells were incubated overnight with Proteinase k at 200 mg/ml. Phenol was added and mixed by

6inversion, then the sample was spun down at 16000 g for 10 minutes, and phenol was discarded. After that, phenol/chloroform/isoamyl alcohol (PCI) was added, mixed by inversion. The sample was spun for 5 minutes at 16000 g. The aqueous solution was collected, and DNA was precipitated using 1 volume of 0.2M NaCl and Isopropanol.

DNA was resuspended in 1X TE Buffer and transferred to Bioruptor Microtubes and sonicated in Biorupor Pico (Diagenode) to an average size of 300bp. 60 µl of Protein G coated Dynabeads (Life Technologies) were incubated with 18 µg of 5-Methylcytosine (D3S22, Cell Signaling) antibody. DNA samples were incubated with the antibody-coated beads, and the unbound DNA was washed with IP buffer (100mM Na-Phosphate, 1.4M NaCl, 0.5% Triton X-100). The sample was then resuspended in Digestion buffer (50mM Tris pH8, 10mM EDTA, 0.5% SDS). AMPure XP beads (Beckman Coulter) was used to isolate pulled-down DNA. MeDIP library preparation was conducted following the Illumina protocol and sequenced using an Illumina Nextseq 500 sequencer.

Genome-wide data analysis

The analyses of the Chip-seq, MeDIPseq and RNAseq data was done on usegalaxy.org and usegalaxy.eu (Blankenberg et al., 2010; Giardine et al., 2005; Goecks, Nekrutenko, Taylor, & Galaxy Team, 2010). Chip-seq and MeDIPseq reads were mapped to the human reference genome hg38 using HISAT2 (Kim, Langmead & Salzberg, 2015a), merged and sorted with BAMtools (Barnett, Garrison, Quinlan, Strömberg, & Marth, 2011) and the peaks were called with MACS 1.0.1 or MACS2 (Zhang, Y. et al., 2008). The produced BedGraph files were uploaded to the UCSC genome browser at genome.ucsc.edu (Kent et al., 2002). The MEME-ChIP tool (Machanick & Bailey, 2011) was used to analyse binding sites motifs. Heat map analysis of data was done using Easeq 1.5.0.0 software (Lerdrup, Johansen, Agrawal-Singh, & Hansen, 2016). For heat map analysis, the BAM file data were plotted on a reference interval file created by setting the start at 500 bp before the peak summit and the end after 500 bp of the summit. Functional annotation of significantly enriched cis-regulatory regions was performed using the bioinformatics tools GREAT (McLean et al., 2010) and DAVID 6.8 (Sherman & Lempicki, 2009).

RNAseq reads were aligned to human reference genome hg38 using HISAT2 tool (Kim, Langmead, & Salzberg, 2015). Transcripts were assembled using Cufflinks 2.0.0, and

Cuffmerg 2.1.1 was used to merge the data of the duplicate samples, and then Cuffdiff 2.0.1 was applied to identify significantly differentially expressed genes (Trapnell et al., 2012). Hierarchical clustering (HCL) and heat map analyses were generated using MultiExperiment Viewer (MeV) 4.9.0 software (Howe, Sinha, Schlauch, & Quackenbush, 2011). The clustering applied Pearson`s correlation and complete linkage method.

Protein assays

Nuclear extract was prepared by washing the cells once with PBS and once with hypotonic buffer (10mM HEPES pH7.6, 10mM KCl, 1.5mM MgCl2) before resuspending the cells at 10^7 cell/ml of hypotonic buffer followed by 30 minutes on ice. The lysate was spun at 500g for 5 minutes at 4°C to pellet the nuclei. The pellet was then resuspended in Hypertonic buffer (20mM HEPES pH7.6, 420mM NaCl, 1.5mM MgCl2, 0.2mM EDTA, 0.5% NP40, 20% Glycerol) at $3x10^8$ cells/ml, incubated on ice for 20 minutes and pelleted at 20000 xg at 4°C for 10 minutes. The supernatant was collected and diluted with 1.8 volume of No-salt buffer (20mM HEPES pH7.6, 1.5mM MgCl2, 0.2mM EDTA, 0.5% NP40, 20% Glycerol), and stored at -20°C until required.

Total protein samples were prepared by washing the cells with PBS and resuspended in IP cell lysis buffer (150mM NaCl, 20mM Tris PH 8, 2mM EDTA, 0.5% (v/v) NP-40+ 1:100 PIC) at 250 µl / 10 cm dish for HEK cells and 200 µl/$1x10^7$ for T-ALL cells. After 10 minutes on ice, cell lysates were spun at 22kg for 10 minutes at 4°C. After centrifugation, the supernatant was collected and stored at -20°C until required. **Table 2.2** summarises a list of antibodies used in the study.

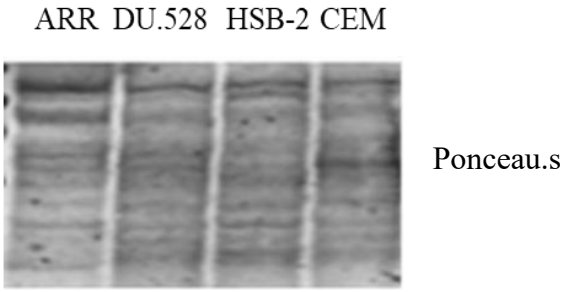


Figure 2.5 Ponceau.s staining of Western blot
The figure illustrates equal loading for 100 µg nuclear extract per sample. Nuclear extracts were isolated from ARR, DU.528, HSB-2, and CCRF-CEM.

Immunoprecipitation

30 µl Protein G coated Dynabeads (Life Technologies) were washed twice with PBS, using a magnetic rack, and incubated in PBS with 3 µg of the antibody for 30 minutes at 4°C. 300 µg of nuclear extract was incubated with the prepared beads for 2 hours on a rotating wheel at 4°C. The beads were subsequently washed twice with ice-cold wash buffer (20mM HEPES pH7.6, 250mM NaCl, 1.5mM MgCl2, 0.2mM EDTA, 0.5% NP40, 20% Glycerol). Beads were resuspended in 45 µl 4X LDS sample loading buffer and 5 µl 10X DTT sample reducing agent (both from Life Technologies) and boiled at 95°C for 5 minutes. For protein pull-down experiments followed by mass spectrometry, the same procedure was followed, except with 100 µl beads, 10ug antibody and 1 mg nuclear extract.

Crosslink Immunoprecipitation:

100 µl Protein G coated Dynabeads (Thermo Fisher Scientific) were washed twice with PBS and crosslinked with 20 µg of antibody overnight at 4°C using Pierce Crosslink Immunoprecipitation Kit following the manufacturer's protocol. 2 mg of nuclear extract was incubated with the antibody -crosslinked beads for 2 hours on a rotating wheel at

4°C. The beads were subsequently washed twice with cold wash buffer. The proteins pulled down with the beads were eluted from the beads in the elution Buffer provided with the Pierce Crosslink Immunoprecipitation Kit. The eluates were mixed with 1X LDS sample loading buffer and 1X DTT sample reducing agent (both from Life Technologies) and boiled at 95°C for 5 minutes.

Western Blots

Protein lysates were separated using 4-12% gradient Bis-Tris Plus Bolt Mini Gels (Novex by Life Technologies) at 200 Volt for 30-35 minutes. Western blot was prepared using iBlot™ Transfer Stack, nitrocellulose (Thermo Fisher Scientific) and incubated with the primary antibodies at 1 µg/ml for one hour and then with the conjugated secondary antibodies IRDye 680RD or 800RD (Li-Cor) for 30 minutes at a dilution of 1:2000. The membrane was washed one time after the primary antibody and three times after the secondary antibody with 10 minutes incubation for each wash. Blots were scanned for fluorescence intensity using an Odyssey CLx infrared imaging system (Li-Cor).

Mass spectrometry

The pulldown samples separated in the gel were stained with Coomassie blue staining dye for 15 minutes and de-stained overnight with 10% acetic acid (**Figure 2.6**). Stained lanes were cut into 12 pieces. For in-gel digestion of proteins, the gel slices were immersed in the first solution (50% acetonitrile, 50mM ammonium bicarbonate) for 30 minutes, followed by the second solution (100mM iodoacetamide in 50% acetonitrile, 50mM ammonium bicarbonate) in the dark for 30 minutes and a washed with (10% acetonitrile, 50mM ammonium bicarbonate) for 30 minutes. The gel slices were then dried overnight using a vacuum centrifuge. Gel slices were incubated in 20 µg of trypsin (Promega) in 500 µl of 10% acetonitrile, 50 mM ammonium bicarbonate overnight at room temperature.

The peptides from the digested proteins were extracted from the gel slices in three steps first with incubating the gel slices for one hour in a 100 µl of (1% formic acid in 10% acetonitrile), (2% formic acid in 60% acetonitrile), and (1% formic acid in 10%

acetonitrile). The fractions collected from each incubation were pooled, and peptides were dried in a vacuum centrifuge overnight. Finally, the pellets were reconstituted in 20 µl of 1% formic acid, of which 10 µl was sent to the mass spectrometry facility of the School of Cancer Sciences, University of Birmingham. The protein-protein network in the MS data was demonstrated using STRING v11 (Szklarczyk et al., 2018).

Immunohistochemistry

$1x10^6$ cells were fixed with formaldehyde in a final concentration of 4% for ten minutes and then permeabilised with PBS, containing 0.5% Triton X-100 and 1% FBS for 30 minutes. Fixed cells were then incubated overnight with the primary antibody at 1 µg /ml at RT. Cells were washed with PBS and incubated with the secondary antibody, for 30 minutes at a working concentration of 1:1000, followed by two washes with PBS. CYTOSPIN 3 (Thermo Fisher Scientific) was used to concentrate cells on slides at 500g for 5 min. Slides were then analysed using Confocal Microscope Zeiss LSM 880 with Airyscan.

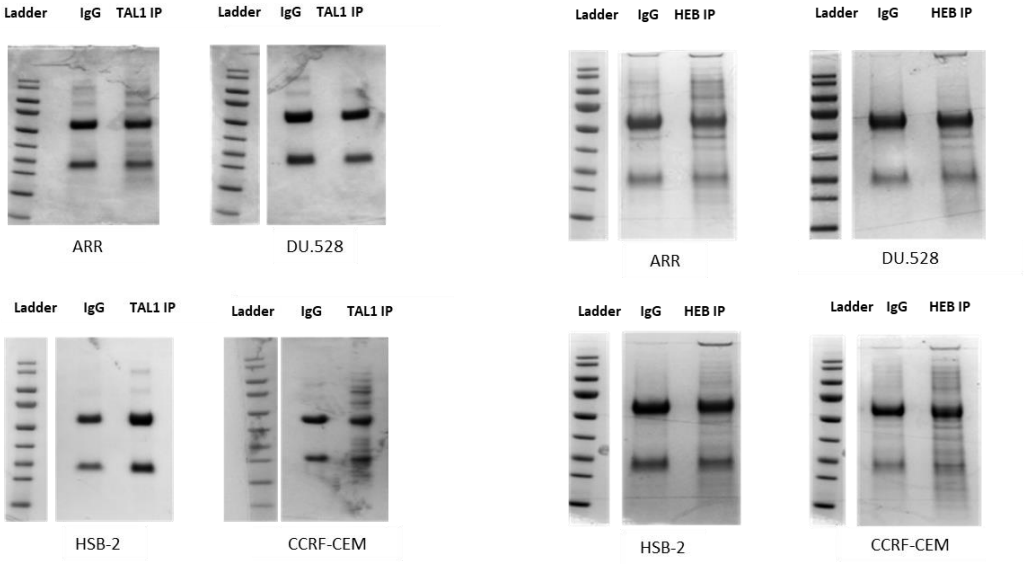


Figure 2.6 Coomassie Blue stained SDS-Page gels of TAL1 immunoprecipitation

Visualisation of immunoprecipitation samples prepared using goat- anti-humanTAL1 and rabbit-anti-human HEB antibodies for mass spectrometry analysis in T-ALL cell lines. Immuno-precipitate with goat IgG and rabbit IgG was used as a control for TAL1 and HEB mass spectrometry, respectively.

Table 2.2 List of antibodies

Antibody	Reference number	Vendor
aTAL1	sc-12984	Santa Cruz
aLMO2	AF2726	R&D
aTCF12	sc-357	Santa Cruz
aLDB1	Ab96799	Abcam
aLYL1	sc-374164	Santa Cruz
aFLAG	FG4R	ebiosciences
aRUNX1	ab23980	Abcam
aFli1	Sc365294	Santa Cruz
aGATA2	AF2046	R&D
aCD34 PE	4H11	ebiosciences
aCD34 FITC	130-008-12	Milteny Biotec
aCD7 eFlour450,	124-1D1	ebiosciences
aCD5 APC	L17F12	ebiosciences
aCD1a PE-Cy7	HI149	ebiosciences
aCD4 APC	OKT4	ebiosciences
aCD3 PE	HIT3a	ebiosciences
aCD3 eFlour450	OKT3	ebiosciences
aCD8 PE	HIT8a	ebiosciences

Statistical analyses

Gene set enrichment analyses were used to determine the statistical significance of genes or proteins enrichment in a term or a pathway compared to the background and reported as false discovery rate adjusted P-value (q-value). The significance of a motif identified in the *de-novo* motif analysis was demonstrated using E-values, as reported by the motif discovery or enrichment program (MEME, or DREME). The standard error of means (SEM) was used to show variation in the means of qPCR enrichment.

CHAPTER THREE

CHARACTERIZATION OF TAL1 PROTEIN-PROTEIN INTERACTIONS IN T-ALL

TAL1 gene code for a bHLH class II transcription factor and has four major isoforms. TAL1 overexpression is a recurrent event in T-ALL. TAL1 oncogenic contribution to T-ALL revolve around TAL1 protein interactions as it assembles in the pentameric transcriptional complex containing LMO, LDB1, GATA, and an E-protein such as E2A and HEB. TAL1 isoforms and their differential roles and characteristics in T-ALL have not been investigated. To examine the oncogenic role of TAL1 and its isoforms in T cell lymphoblastic leukaemia, we used four T-ALL cell lines, ARR, DU.528, HSB-2, and CCRF-CEM. Three of the cell lines, naming DU.528, HSB-2, and CCRF-CEM, carry deletions that fuse the regulatory region of the SCL-interrupting locus (SIL) to TAL1 gene causing its overexpression, while ARR displays features of the myeloid and lymphoid lineages (Burger, Hansen-Hagge, Drexler, & Gramatzki, 1999; ROBERTS, WIERSMA, URIBE, & WEINBERG, 1992).

To characterise the T-ALL cell lines in our study, we evaluated TAL1 protein expression and the immunophenotype of the four cell lines. We isolated nuclear extract samples from the T-ALL cells and analysed using western blot. There was an aberrant expression of TAL1 protein in the four T-ALL cell lines, as shown in TAL1 western blot (**Figure 3.1**). In addition, TAL1 western blot shows that the proportional expression of the different isoforms of TAL1 is different between the four cell lines. In particular, DU.528 mainly expresses the short TAL1 isoforms, whereas the other T-ALL cell lines predominantly express the longest two isoforms.

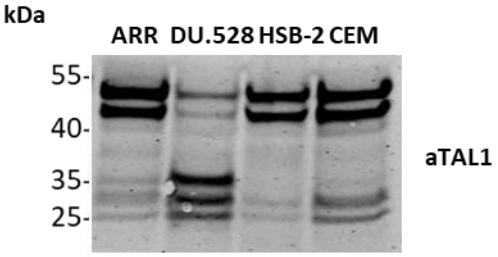
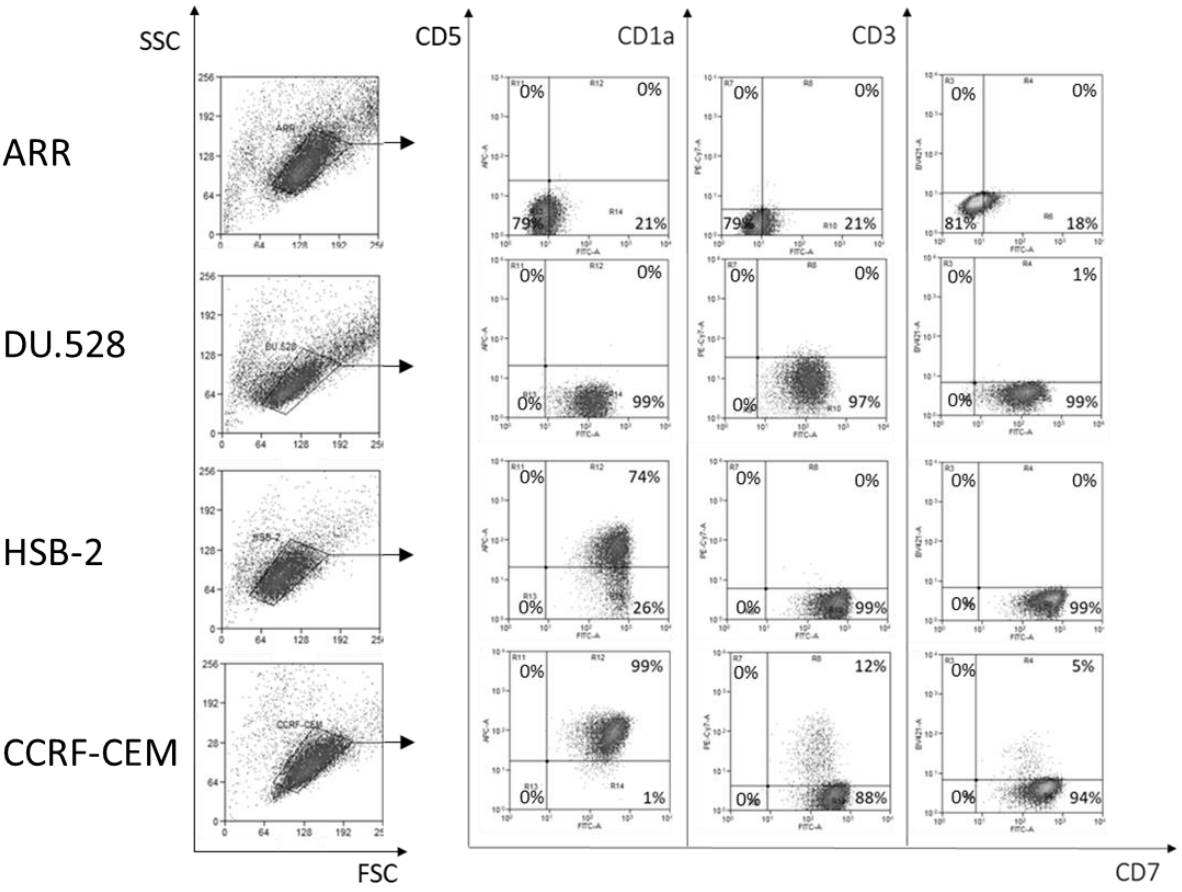


Figure 3.1: The aberrant expression of TAL1 protein in T-ALL cell lines

Nuclear protein samples were isolated from T-ALL cell lines, ARR, DU.528, HSB-2, and CCR-CEM. Western blot analysis of 100 µg of protein was loaded on polyacrylamide gels and blotted with 1 µg/ml TAL1 antibody to show TAL1 protein expression.



ARR

	Count	% Hist	% All
	12667	100.00	46.23
A	0	0.00	0.00
B	3	0.02	0.01
C	9991	78.87	36.46
D	2673	21.10	9.76

	Count	% Hist	% All
	12667	100.00	46.23
A	36	0.28	0.13
B	35	0.28	0.13
C	9955	78.59	36.33
D	2641	20.85	9.64

	Count	% Hist	% All
	12667	100.00	46.23
A	35	0.28	0.13
B	37	0.29	0.14
C	10282	81.17	37.53
D	2313	18.26	8.44

DU.528

	Count	% Hist	% All
	9314	100.00	37.62
A	1	0.01	0.00
B	2	0.02	0.01
C	59	0.63	0.24
D	9252	99.33	37.37

	Count	% Hist	% All
	9314	100.00	37.62
A	0	0.00	0.00
B	160	1.72	0.65
C	29	0.31	0.12
D	9125	97.97	36.86

	Region	Count	% Hist
	Total	9314	100.00
A	R3	4	0.04
B	R4	108	1.16
C	R5	28	0.30
D	R6	9174	98.50

HSB-2

	Count	% Hist	% All
	9539	100.00	44.55
A	0	0.00	0.00
B	7093	74.36	33.13
C	0	0.00	0.00
D	2446	25.64	11.42

	Count	% Hist	% All
	9539	100.00	44.55
A	0	0.00	0.00
B	34	0.36	0.16
C	0	0.00	0.00
D	9505	99.64	44.40

	Count	% Hist	% All
	9539	100.00	44.55
A	0	0.00	0.00
B	20	0.21	0.09
C	0	0.00	0.00
D	9519	99.79	44.46

CCRF-CEM

	Count	% Hist	% All
	10828	100.00	49.90
A	1	0.01	0.00
B	10672	98.56	49.18
C	1	0.01	0.00
D	154	1.42	0.71

	Count	% Hist	% All
	10828	100.00	49.90
A	0	0.00	0.00
B	1291	11.92	5.95
C	2	0.02	0.01
D	9535	88.06	43.94

	Count	% Hist	% All
	10828	100.00	49.90
A	0	0.00	0.00
B	544	5.02	2.51
C	1	0.01	0.00
D	10283	94.97	47.39

Figure 3.2. T-ALL cell lines are blocked at different stages of differentiation

Live T-ALL cells were gated through forward scatter / side scatter plots. Cells were stained using antibodies against CD7, CD5, CD1a, and CD3. Cells incubated with isotype-matched non-specific IgG antibodies served as a gating to divide the plot into quadrants. The statistics of the quadrants A-D for each T cell markers histogram are shown in the lower panel (A, top left, B, top right, c, lower left, D, lower right). The immunophenotype was confirmed in two similar experiments.

Although they all express TAL1, the four T-ALL cell lines are blocked at different stages of T cell differentiation. In order to study the effects on T cell differentiation markers further in the study, we needed to examine their expression in the four T-ALL cell lines. **Figure 3.2** shows flow cytometry assays of the T-ALL cell lines, using antibodies recognising cell surface markers that are commonly used to determine the differentiation stage of T-cell progenitors. The expression patterns showed a block at an early stage ($CD7^{low}$, $CD5^{-}$) for ARR cells, followed by DU.528 ($CD7^{+}$, $CD5^{-}$, $CD1a^{-}$), and HSB-2 ($CD7^{+}$, $CD5^{+}$, $CD1^{-}$), and CCRF-CEM ($CD7^{+}$, $CD5^{+}$, $CD1a^{+/-}$, $CD3^{+/-}$, $CD4^{+}$).

TAL1 Protein interactions in T-ALL

To explore TAL1 protein interactions in T-ALL, we performed co-immunoprecipitation (co-IP) assays followed by mass spectrometry. Nuclear extracts from T-ALL cell lines were incubated with magnetic beads pre-coated with anti-TAL1 or control IgG antibodies. The pulldown samples were run on polyacrylamide gels. Whole lanes were isolated and divided into gel slices, within which proteins were digested with trypsin. After extraction, the resulting peptides were analysed using mass spectrometry. We first subtracted the resulting proteins detected in the IgG control from the list of TAL1 co-immunoprecipitated proteins. **Figure 3.3** shows a Venn diagram comparing the results of the different T-ALL cell lines. Twelve proteins were shared between the four cell lines, whereas forty-nine proteins were identified in the TAL1 mass spectrometry of ARR, DU.528 and CCRF-CEM. The proteins shown in the table were first analysed using CRAPome database to filter probable false positives. Proteins matching CRAPome databases with a cut-off of >20/ 411 (Num of Expt. (found/total)) were excluded. The remaining proteins were sorted according to the number of peptides detected. **Table 3.1** lists top10 proteins found in the TAL1 mass spectrometry for each of the T-ALL cell lines.

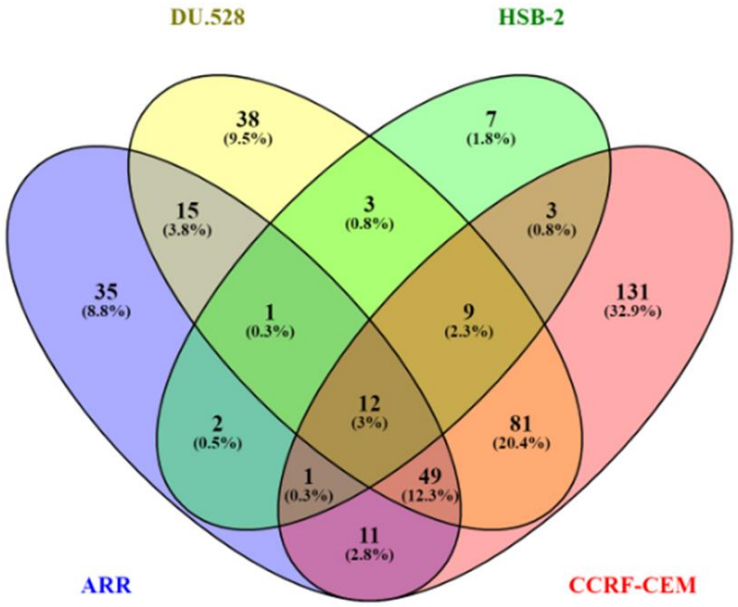


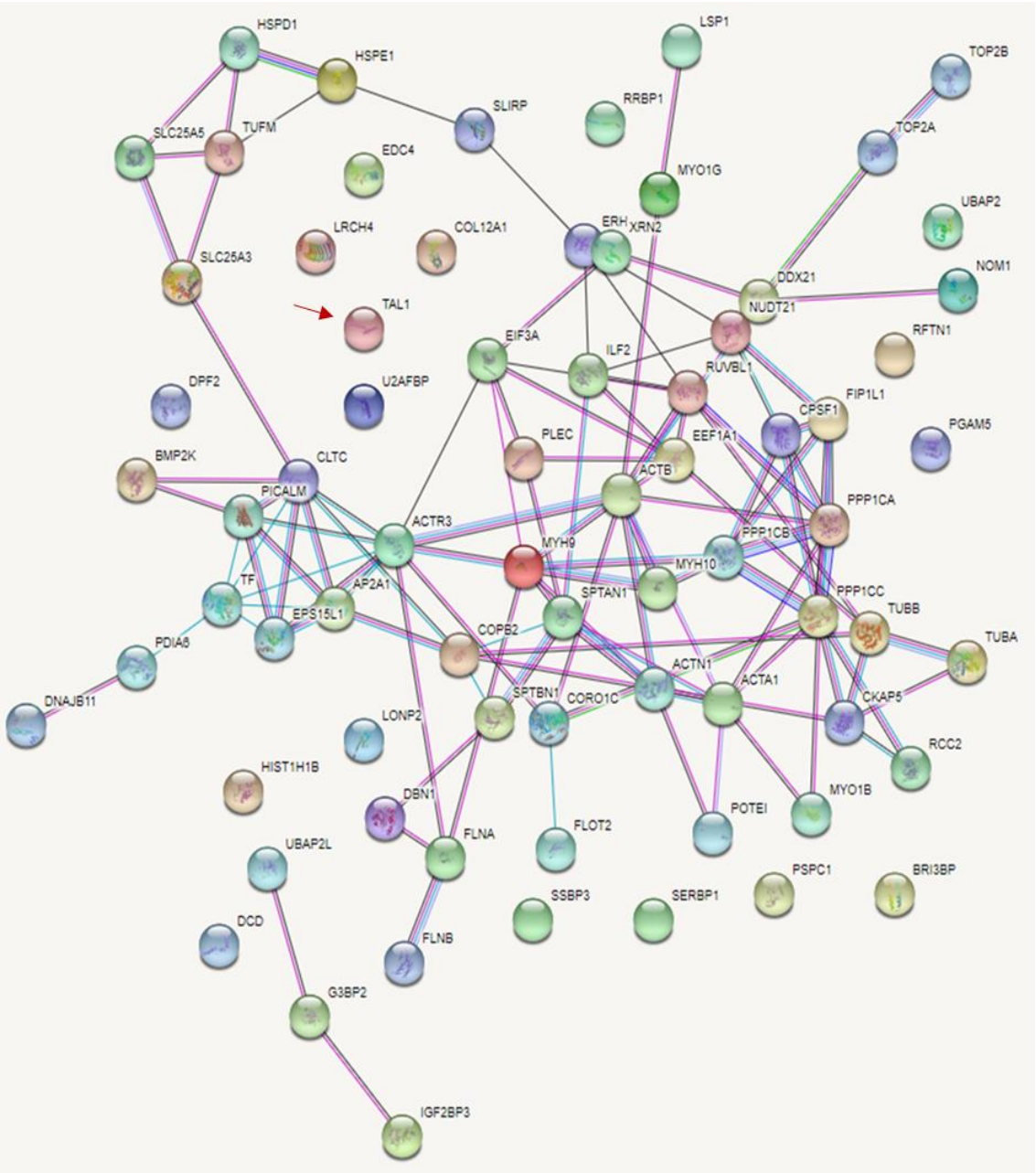
Figure 3.3 Venn-diagram of TAL1 mass spectrometry in T-ALL cell lines

1 mg of nuclear extracts were immunoprecipitated with protein G coated beads and 10 µg TAL1 antibody. Pulldown samples were analysed using mass spectrometry. Four-way comparison of proteins in TAL1 co-immunoprecipitation (co-IP) as identified using mass spectrometry.

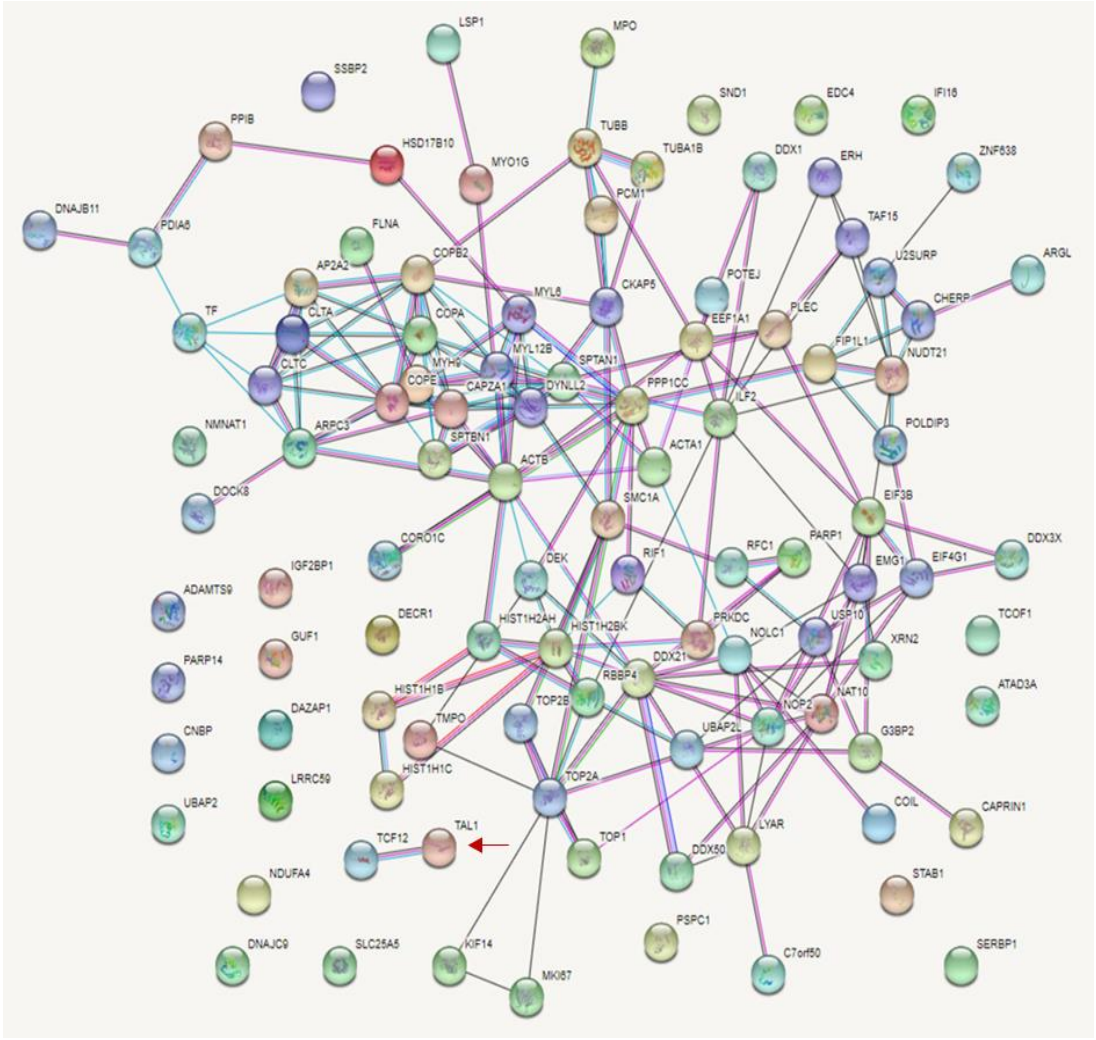
	#Peptides	Protein	Mapped Gene Symbol
ARR	11	Fibronectin	FN1
	10	Serotransferrin	TF
	8	Lymphocyte-specific protein 1	LSP1
	5	Unconventional myosin-Ig	MYO1G
	5	POTE ankyrin domain family member I	POTEI
	3	Actin-related protein 2/3 complex subunit 1B	ARPC1B
	3	Dedicator of cytokinesis protein 8	DOCK8
	3	Flotillin-1	FLOT1
	3	Leucine-rich repeat and calponin homology domain-containing protein 4	LRCH4
DU.528	12	Unconventional myosin-Ig	MYO1G
	6	2 4-dienoyl-CoA reductase mitochondrial	DECR1
	5	Stabilin-1	STAB1
	3	H/ACA ribonucleoprotein complex subunit 3	NOP10
	2	Dedicator of cytokinesis protein 8	DOCK8
	2	Myeloperoxidase	MPO
	2	Transcription factor 12	TCF12/HEB
	1	Lymphocyte-specific protein 1	LSP1
	1	Nicotinamide mononucleotide adenylyltransferase 1	NMNAT1
	1	Single-stranded DNA-binding protein 2	SSBP2
HSB-2	10	Dedicator of cytokinesis protein 8	DOCK8
	6	Unconventional myosin-Ig	MYO1G
	3	T-cell acute lymphocytic leukemia protein 1	TAL1
	1	Breast cancer type 1 susceptibility protein	BRCA1
	1	Translation factor GUF1, mitochondrial	GUF1
	1	Mitotic spindle assembly checkpoint protein	MAD1L1
	1	Shugoshin-like 2	SGOL2
	1	Sterol regulatory element-binding protein 1	SREBF1
	1	Transcription factor 12	TCF12/HEB
	1	Tigger transposable element-derived protein 6	TIGD6
CCRF-CEM	4	Gamma-interferon-inducible protein 16	IFI16
	4	H/ACA ribonucleoprotein complex subunit 3	NOP10
	2	2,4-dienoyl-CoA reductase, mitochondrial	DECR1
	2	Heparan sulfate 2-O-sulfotransferase 1	HS2ST1
	2	Tyrosine-protein kinase Lck	LCK
	2	DNA-directed RNA polymerase I subunit RPA1	POLR1A
	2	Titin	TTN

We used the STRING database to assess known protein interactions between proteins we identified in TAL1 MS. We illustrate a network of proteins identified in TAL1 mass spectrometry in ARR, DU.528, HSB-2, and CCRF-CEM, in **figure 3.4 a, b, c, and d** respectively. Nodes represent proteins, while known interactions are indicated as connecting lines. The list of TAL1 mass spectrometry proteins was curated by subtraction of the IgG results, as well as ribosomal proteins and splicing factors for clarity. The arrow in the TAL1 mass spectrometry network (**Figure 3.4 a-d**) indicates TAL1 protein. In DU.528 and HSB-2, the E-protein HEB (TCF12) was detected, as shown in **figure 3.4 b and c**.

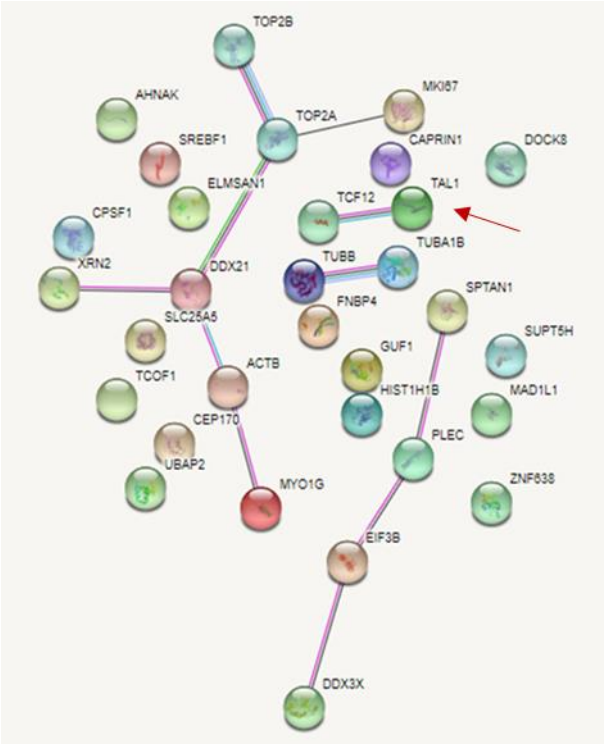
A


HSPD1
HSPE1
LSP1
TOP2B
SLIRP
RRBP1
SLC25A5
TUFM
TOP2A
EDC4
MYO1G
UBAP2
LRCH4
COL12A1
ERH
XRN2
SLC25A3
NOM1
DDX21
TAL1
NUDT21
RFTN1
EIF3A
ILF2
RUVBL1
DPF2
U2AFBP
FIP1L1
CPSF1
PGAM5
PLEC
EEF1A1
BMP2K
CLTC
ACTB
PICALM
PPP1CA
ACTR3
PPP1CB
MYH9
MYH10
TF
AP2A1
SPTAN1
PPP1CC
TUBB
EPS15L1
COPB2
PDIA6
TUBA
ACTN1
ACTA1
DNAJB11
SPTBN1
CORO1C
CKAP5
LONP2
RCC2
HIST1H1B
MYO1B
DBN1
POTEI
FLNA
FLOT2
UBAP2L
PSPC1
BRI3BP
SSBP3
SERBP1
DCD
FLNB
G3BP2
IGF2BP3

B


LSP1
MPO
SSBP2
SND1
EDC4
IFI16
PPIB
TUBB
HSD17B10
TUBA1B
DDX1
ERH
ZNF638
MYO1G
PCM1
DNAJB11
PDIA6
TAF15
FLNA
U2SURP
ARGL
AP2A2
COPB2
POTEJ
CKAP5
MYL6
CHERP
TF
CLTA
COPA
EEF1A1
PLEC
MYL12B
FIP1L1
NUDT21
CLTC
MYH9
SPTAN1
COPE
CAPZA1
DYNLL2
PPP1CC
LF2
NMNAT1
ARPC3
POLDIP3
SPTBN1
ACTA1
ACTB
DOCK8
SMC1A
EIF3B
CORO1C
PARP1
DDX3X
RFC1
DEK
RIF1
EMG1
EIF4G1
IGF2BP1
ADAMTS9
HIST1H2AH
HIST1H2BK
PRKDC
USP10
TCOF1
DECR1
NOLC1
XRN2
GUF1
DDX21
PARP14
RBBP4
ATAD3A
HIST1H1B
TOP2B
NAT10
NOP2
DAZAP1
TMPO
UBAP2L
G3BP2
CNBP
HIST1H1C
TOP2A
LRRC59
COIL
CAPRIN1
UBAP2
TAL1
TCF12
TOP1
LYAR
DDX50
NDUFA4
STAB1
PSPC1
DNAJC9
SLC25A5
KIF14
C7orf50
SERBP1
MKI67

C


TOP2B
MKI67
AHNAK
TOP2A
SREBF1
CAPRIN1
DOCK8
ELMSAN1
TCF12
TAL1
CPSF1
TUBA1B
XRN2
DDX21
TUBB
SPTAN1
FNBP4
SLC25A5
SUPT5H
GUF1
ACTB
TCOF1
HIST1H1B
MAD1L1
CEP170
PLEC
UBAP2
MYO1G
ZNF638
EIF3B
DDX3X

D

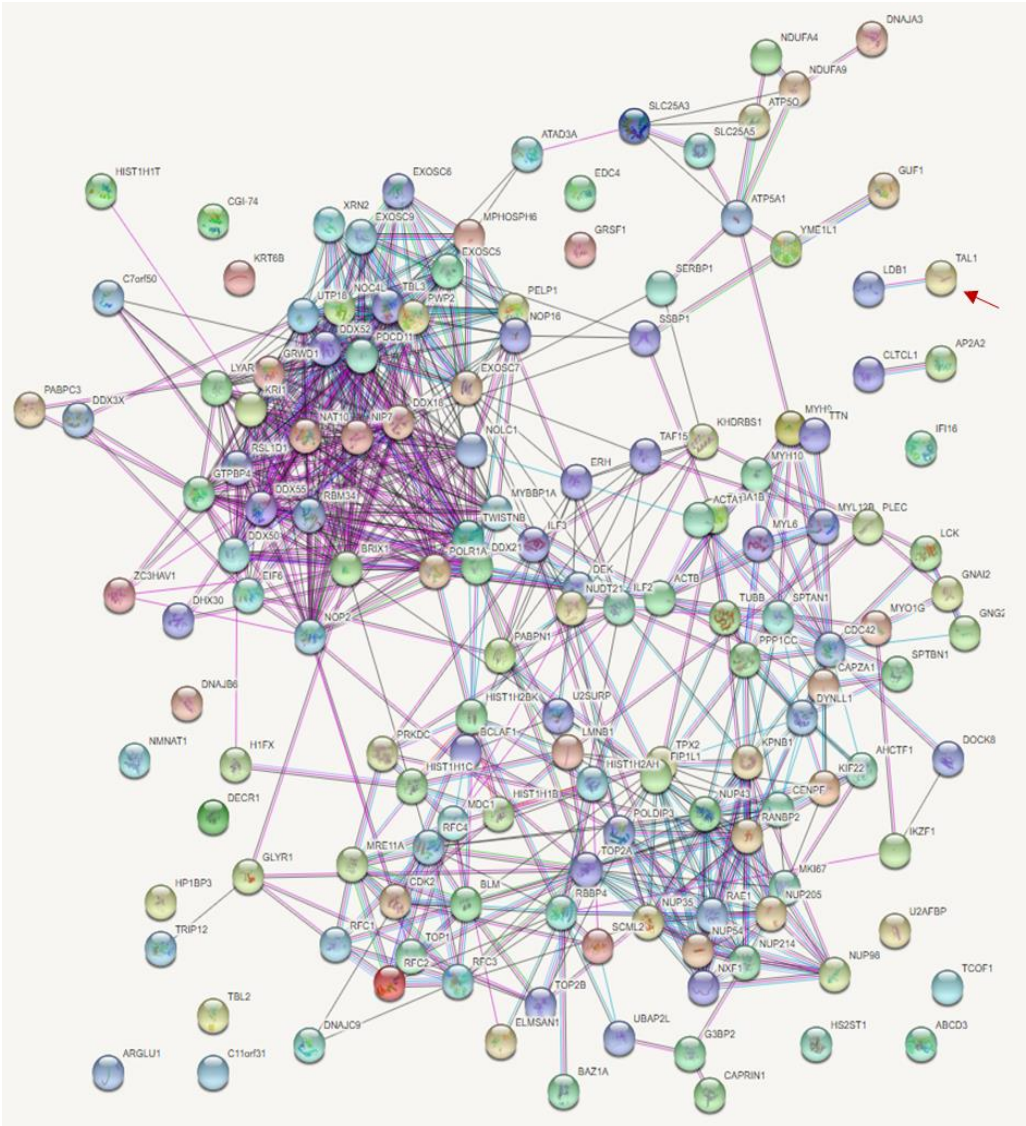


Figure 3.4 Networks of known interactions between proteins in TAL1 MS

Proteins detected using mass spectrometry analysis of TAL1 pulldowns were interrogated for known protein interactions in ARR (a), DU.528 (b), HSB-2 (c), and CCRF-CEM (d). Nodes represent proteins, lines signify known interactions and the arrow points to TAL1. Ribosomal and splicing factors were not included in the network analysis.

The same lists of proteins identified in TAL1 mass spectrometry were analysed after excluding ribosomal proteins and splicing factors to identify the top five significantly enriched pathways. In ARR, proteins involved in cell division, and proteins related to cellular localisation and organisation were found significantly enriched. In DU.528 TAL1 mass spectrometry we found regulation of gene expression and biological process, nucleic acid metabolism, cellular localisation and cell division. Biological processes related to the regulation of chromosome organisation, mitotic cell cycle, chromosome segregation and regulation of gene expression were statistically significant in HSB-2 TAL1 mass spectrometry. Lastly, terms associated with the proteins found in the CCRF-CEM TAL1 pulldown were ribosome, cellular localisation, nucleic acid metabolism, gene expression and cell cycle.

Looking at the combined TAL1 mass spectrometry data collected in the four T-ALL cell lines, we identified the top significantly enriched pathways, as shown in **figure 3.5**. They include transcript metabolism and transport, cell cycle, as well as RHO-GTPase.

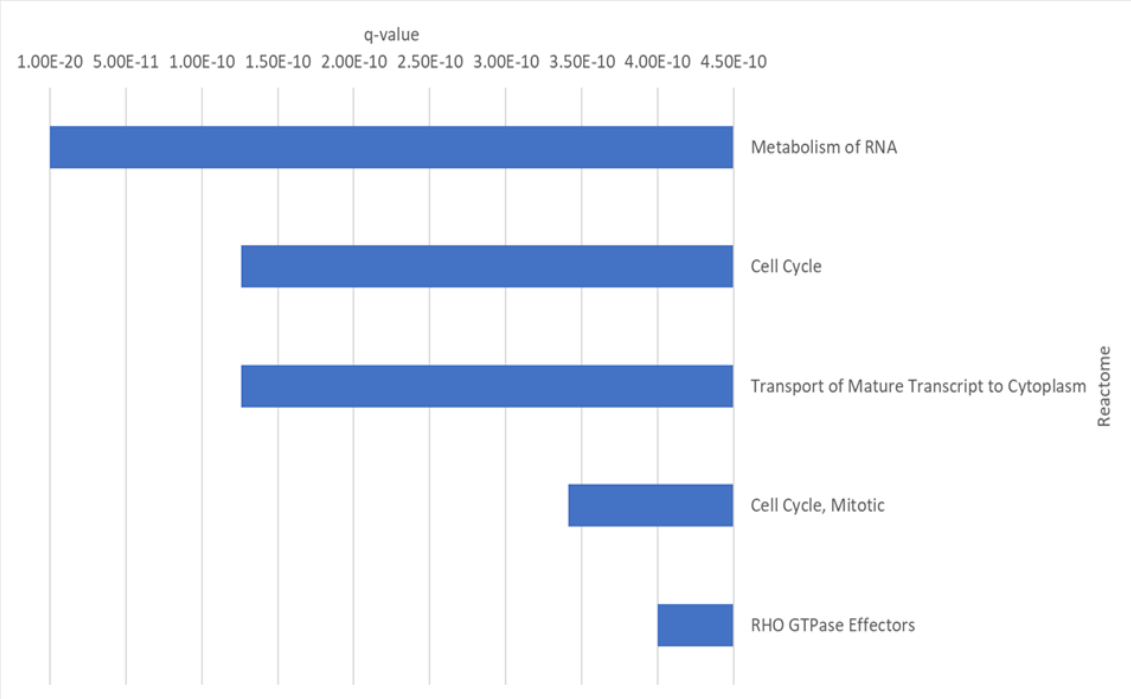


Figure 3.5 The top five significantly enriched pathways in TAL1 mass spectrometry of T-ALL cell lines.

Bar chart showing q-values for the top five significantly enriched pathways identified for proteins detected in the combined list TAL1 mass spectrometry analysis of the four T-ALL cell lines.

HEB association with TAL1 in T-ALL

TAL1 as class II bHLH protein is known to interact with E-proteins. In mammalian cells, three E-proteins have been described, including E2A, HEB, and E2-2. Therefore, the three E-protein are potential partners of TAL1. The fact that the E-protein HEB (TCF12) was detected in DU.528 and HSB-2 TAL1 mass spectrometry prompted us to investigate the heterodimer of TAL1 and HEB further. We proceeded to confirm the HEB and TAL1 interaction that the TAL1 MS first identified in the four T-ALL cell lines by demonstrating HEB protein in TAL1 pulldown and vice versa.

First, we incubated nuclear extract samples of the four T-ALL cell lines with magnetic beads pre-coated with TAL1 or IgG antibodies. Unbound proteins were washed, and pulldown proteins were eluted. The proteins in the TAL1 and IgG pulldown samples in addition to input control samples (IC) were run on an acrylamide gel and analysed using western blots probed with HEB antibody (**Figure 3.6 A**). The bands detected in the input controls confirm that all four T-ALL cell lines express HEB. In TAL1 pulldown lanes, we show HEB protein bands clearly in ARR and CCRF-CEM. The HEB protein bands in the TAL1 IP of DU.528 and HSB-2 was barely detectable. The lack of signal in the IgG lanes validate the specificity of HEB enrichment in TAL1 pulldown samples and thus demonstrate TAL1 and HEB interaction.

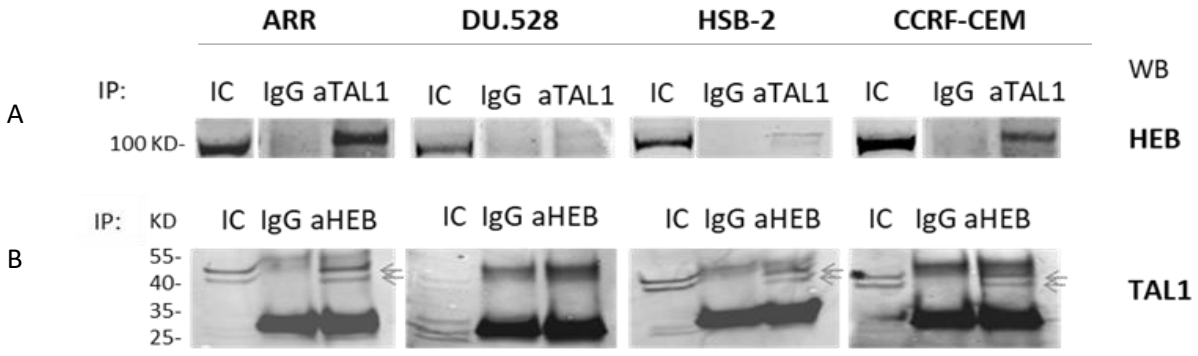


Figure 3.6 TAL1 and HEB co-immunoprecipitation

30 µg T-ALL nuclear lysates were co-immunoprecipitated using 3 µg aTAL1, aHEB or IgG protein G coated beads. **A,** the input control (IC), IgG and TAL1 pulldown samples (aTAL1) were blotted with aHEB. **B,** Input samples and pulldown samples were blotted with aTAL1.

Secondly, in the reverse experiment, we used magnetic beads pre-coated with HEB or IgG antibodies, and only proteins that are bound to the antibody-coated beads were eluted. The input control samples (IC), and HEB and IgG pulldown samples were analysed using western blots probed with TAL1 antibody (**figure 3.6 B**). In ARR, HSB-2, and CCRF-CEM western blots we found bands between 40-50 KD in HEB pulldown and not in the IgG, as indicated using arrows. The specific bands correspond to TAL1 long isoforms (**figure 3.6 B**). In DU.528, no bands of TAL1 long isoforms were found in HEB pull down, and this was expected because DU.528 favours the short isoforms as shown in the input lane and **figure 3.1**. The presence of TAL1 short isoforms in HEB IP could not be assessed because of the strong signal of the band between 25-35 kD in the IgG and pulldown lanes. The obscuring unspecific band corresponds to the light chain of the antibody used in the pulldown.

To minimise the obscuring light chain signal and evaluate TAL1 short isoforms interaction with TAL1 partners in the transcriptional complex, we used crosslink immunoprecipitation assay. We cross-linked HEB or IgG antibodies to the beads. Then, we incubated the nuclear extract samples of the T-ALL cell lines with the antibody-crosslinked beads and washed the unbound proteins. The assay included the use of an optimised elution buffer to collect the bound proteins while avoiding antibody contamination. As shown in **figure 3.7**, the eluted pulldown samples, IgG and input control samples were run on acrylamide gels and blotted using TAL1 antibody. The unspecific band of the light chain was considerably reduced. Looking at the specific TAL1 bands that are enriched in the pulldown compared to the IgG control, we found bands of TAL1 long isoforms as in the previous HEB co-immunoprecipitation. Additionally, we found bands corresponding to TAL1 short isoforms in DU.528 and CCRF-CEM in as indicated with arrows. In the crosslink co-immunoprecipitation assay, all bands were weaker compared to the previous co-immunoprecipitation assay, possibly suggesting a suboptimal elution.

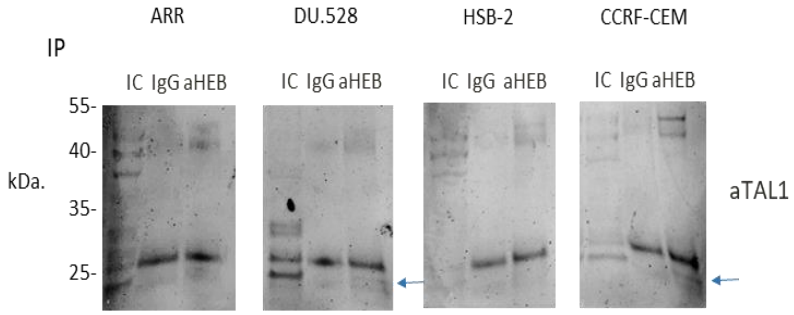


Figure 3.7: Cross-link immunoprecipitation using aHEB in T-ALL

T-ALL nuclear extracts were incubated with protein G coated beads previously cross-linked with aHEB or IgG. Input control (IC), IgG control (IgG), and HEB pulldown (aHEB) samples were separated on gel and blotted using TAL1 antibody. The arrows points to TAL1 short isoform in aHEB lane.

Furthermore, we investigated the interaction between TAL1 short isoforms and LMO2, an obligate partner of TAL1. We incubated T-ALL nuclear extract samples with the beads crosslinked with LMO2 or IgG antibodies, washed unbound proteins, and eluted LMO2 pulldown proteins. Then we blotted pulldown and input control samples using TAL1 antibody (**Figure 3.8**). We detected one band of TAL1 short isoform in ARR and in CCRF-CEM. LDB1 and LMO2 western blots were used to as controls. Although the LMO2 pulldown lanes of LMO2 western blot did not show signal in DU.528 and in HSB-2 there was only a faint signal, LDB1 western showed a relative enrichment in LMO2 pulldown compared to the IgG. The enrichment of LDB1 without LMO2 could suggest that it is harder to retrieve LMO2 protein from the crosslinked antibody. The presence of TAL1 short isoforms indicates that they retain the ability to interact with LMO2 and HEB.

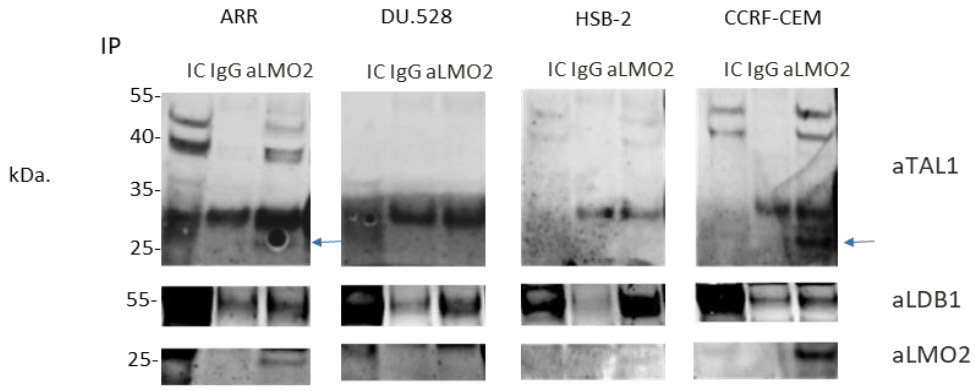


Figure 3.8 Cross-link immunoprecipitation using aLMO2 in T-ALL

2 mg T-ALL nuclear extracts were incubated with protein G coated beads previously cross-linked with 20 µg aLMO2 or IgG antibody. Input control (IC), IgG control (IgG), and HEB pulldown (aHEB) samples were blotted using TAL1, LDB1, and LMO2 antibodies.

HEB colocalisation with TAL1 and LMO2 in the T-ALL cells was visualised using immunocytochemistry. The T-ALL cells were fixed, permeabilised and stained with HEB, TAL1, and LMO2 as primary antibodies followed with their corresponding secondary antibodies. The fluorescent stain DAPI was used as a nuclear stain. Detection of the fluorescent signal of the conjugated secondary antibodies and DAPI was performed using confocal microscopy. **Figure 3.9** shows staining for LMO2 (green), TAL1 (purple) and HEB (red), while the nuclei are highlighted in blue, using DAPI. Nuclear staining was present for all three factors in all four T-ALL cell lines. In addition, HEB (red) was also observed outside the overlapping area, as shown by the red staining in the overlay image, where co-localisation appears in pink. The results of the co-immunoprecipitation and the immunocytochemistry assay demonstrate HEB interaction with TAL1 and LMO2. The results show that HEB assembles with TAL1 and LMO2 in the nucleus, suggesting that their transcriptional complex could bind cis-regulatory regions and regulate gene expression in T-ALL.

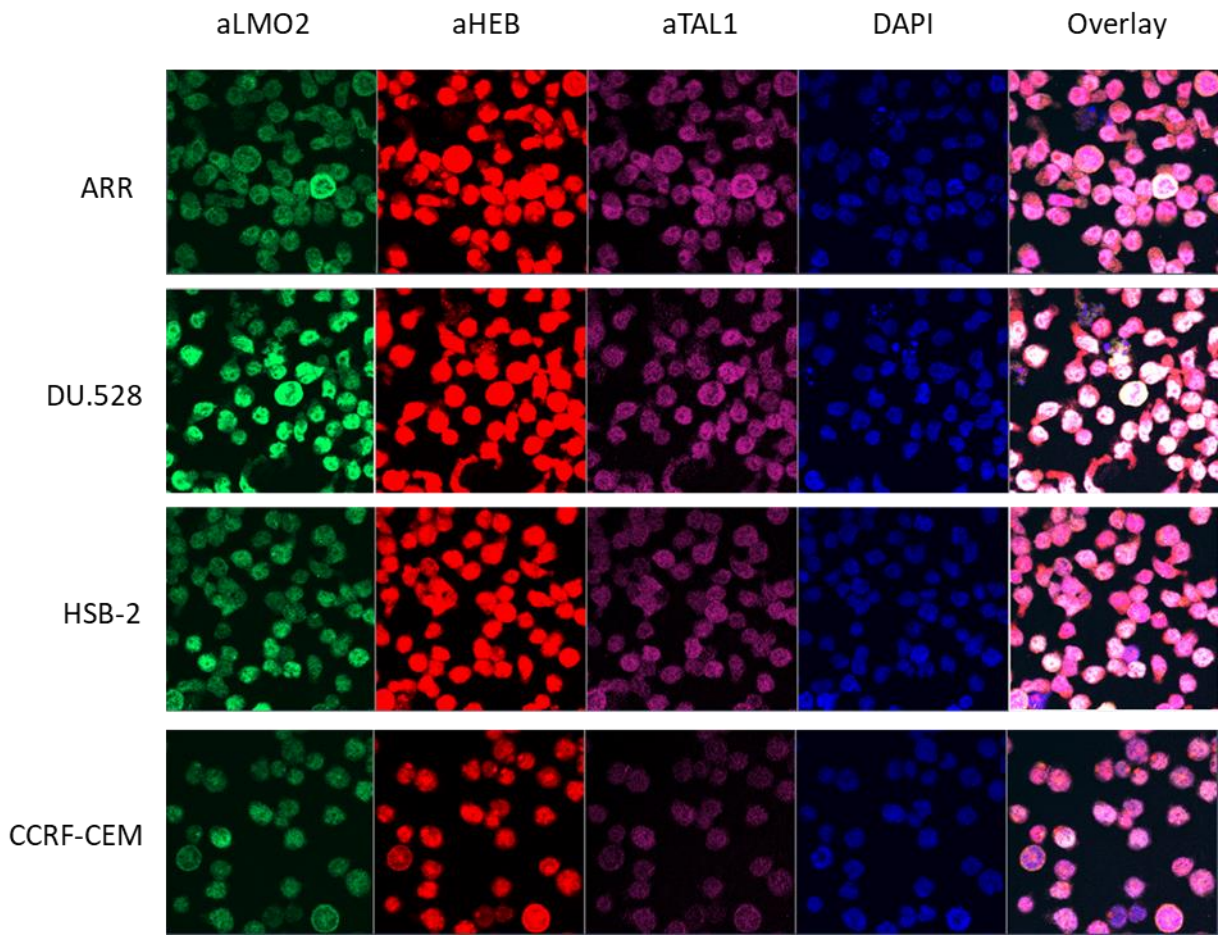


Figure 3.9 Immunocytochemistry of HEB, TAL1, and LMO2 in T-ALL cell lines

T-ALL cells were fixed and permeabilised before incubation with 1 µg/ml aLMO2, aHEB, and aTAL1. Fluorescence of their corresponding secondary antibody was detected to show the protein expression of LMO2, HEB, and TAL1 in T-ALL cells.

Discussion:

TAL1/HEB interaction in T-ALL cell lines

E-proteins (E2A, HEB, and E2-2) are known to dimerise with TAL1. The dimerisation of TAL1 with E-proteins is an essential component of the TAL1 pentameric complex, which mainly include LMO, GATA, RUNX, and ETS (Lécuyer & Hoang, 2004; Sanda et al., 2012b) . TAL1 was also demonstrated to induce leukaemia by sequestering E-proteins and thereby disrupting their transcriptional program (Chervinsky et al., 1999; O'Neil et al., 2004). Therefore, the TAL1/ E-protein heterodimer is central for understanding the mechanisms of the oncogenic role TAL1 plays in T-ALL.

TAL1 mass spectrometry was the starting point of our investigation of the role of TAL1 in T-ALL. Focusing on the TAL1 transcriptional complex with HEB, and not the other E-proteins, was initiated by TAL1 mass spectrometry results. We found HEB in the mass spectrometry analysis of TAL1 co-immunoprecipitation (co-IP) samples in DU.528 and HSB-2 (**table 3.1**). We validated the mass spectrometry findings with TAL1 co-IP followed with HEB western blot, as well as by the reverse experiment (**Figure 3.6**). Additionally, we showed the presence of HEB protein in the TAL1 pulldown from ARR and CCRF-CEM samples as well. DU528 and HSB-2 had minimal HEB signal in the TAL1 pulldown. In the reverse experiment, we demonstrated the presence of the long TAL1 isoforms in HEB pulldown samples in ARR, HSB-2 and CCRF-CEM. We did not find enrichment of TAL1 long isoforms in DU.528, which was expected because DU.528 mainly expresses the short isoforms of TAL1, as shown in the input control.

Showing the short TAL1 isoforms in the HEB co-immunoprecipitation, performed as described above, was not feasible due to the location of the strong signal given by the light chain on the western blot. To reduce this issue, we used cross-linked co-immunoprecipitation where the antibody is crosslinked to the magnetic beads. In crosslinked HEB pulldown, all signal was clearly less after crosslinking, compared to the non-cross-linked experiment (**Figure 3.7**). We found detectable levels of the shortest isoform of TAL1 in the HEB pulldown samples of DU.528, CCRF-CEM and ARR, although the last was very faint. HSB-2 had very low expression levels of the TAL1 short

isoforms, as shown in the input, and therefore these were not visible in the pulldown sample. Taken together, this showed that also the short isoforms of TAL1 could interact with HEB.

In co-IP experiments where LMO2 antibody was crosslinked to beads, we demonstrated the interaction with the short isoform of TAL1 in ARR and CCRF-CEM, even though the quality of western blots needs improving (**Figure 3.8**). LMO2 western blot used as a control for the pulldown did not show enrichment of LMO2 in DU.528 and HSB-2 LMO2 pulldown samples. Another control we used was the LIM binding domain1 (LDB1), which is an LMO2 interacting partner. DU.528 showed a relatively higher signal of LDB1 in LMO2 pulldown compared to the IgG control, while HSB-2 showed enrichment of LDB1 signal in LMO2 pulldown over the IgG control. The detection of the short TAL1 isoforms in DU.528 and HSB-2 LMO2 pulldown was not conclusive. The results of the assay could be enhanced with optimisation, especially for elution of the proteins from the crosslinked beads. Nevertheless, the experiments supported our pursuit of isoform-specific roles in T-ALL.

In order to examine the colocalisation of HEB, TAL1, and LMO2 in the T-ALL cell lines, we performed immunocytochemistry assay (**Figure 3.9**). We showed that HEB, TAL1 and LMO2 colocalise in the nucleus, by comparison with the nuclear stain DAPI. The immunocytochemistry results showed that TAL1 and LMO2 protein signal was mostly contained within the nucleus as expected for transcription factors. Nevertheless, low levels of TAL1 and LMO2 were also present outside the nucleus. We observed a strong signal for HEB outside the nucleus, possibly because the HEB signal was higher than that for TAL1 and LMO2. Interestingly, our immunocytochemistry results showed TAL1, LMO2, and HEB overlapped with DAPI staining of mitotic chromatin as well.

Screening for potential TAL1 protein-protein interactions in T-ALL

The peptides detected in TAL1 mass spectrometry identified a network of TAL1 protein interactions (**Figure 3.4**). In addition to HEB, we detected additional proteins known to interact with TAL1 complex in the analysis of TAL1 MS. We detected LDB1 and its co-factors SSBPs which are known to block LDB-1 degradation. We found LDB1 and SSBP1 in CCRF-CEM, SSBP2 in DU.528, and SSBP3 in ARR. LDB-1 binding to SSBPs is known to protect LDB-1 and LMO2 from degradation (Chen et al., 2002; Gungor et al., 2007). We compared TAL1 mass spectrometry results after filtering proteins found in the IgG and CRAPome database and sorting using the number of peptides found (**Table 3.1**). This narrowed down the results and the TAL1, and two of its known associates HEB, and SSBP2 remained in the top10. This indicates that the proteins retrieved using this filtration process could be true TAL1 interacting proteins.

We used another approach to represent the functional roles of TAL1 in TAL1 MS data. We excluded the most abundant proteins found in TAL1 MS which were related to ribosome and spliceosome. We showed a network of proteins detected in TAL1 pulldown samples of the four T-ALL cell lines. Consequently, we were able to find that proteins related to transcript metabolism and transport, cell cycle and Rho GTPases effectors, were significantly enriched in TAL1 MS when looking at the data of the four T-ALL cell lines together (**Figure 3.5**).

It was interesting to find the enrichment of the cell cycle term in TAL1 MS data. TAL1 has been shown to regulate cell cycle by regulating gene expression of cell cycle-related proteins in haematopoietic cells which is expected for an oncogene (Chagraoui et al., 2011; Dey, S., Curtis, Jane, & Brandt, 2010; Lacombe, Julie et al., 2010; Sincennes et al., 2016). However, direct protein-protein interaction between TAL1 and cell cycle-related proteins have not been shown before. TAL1 interacting partner LMO2 has recently been shown to interact with cell cycle and DNA replication proteins (Sincennes et al., 2016). Therefore, our result of protein-protein interaction between TAL1 and cell cycle proteins need to be verified and further investigated to understand the link between TAL1 and LMO2 and their direct involvement with DNA replication and cell division proteins. The direct participation of TAL1 in DNA replication and cell division would add to its oncogenic roles as a transcription factor to the induction of uncontrolled proliferation.

CHAPTER FOUR

THE DYNAMICS OF TAL1 TRANSCRIPTIONAL COMPLEX GENOMIC OCCUPANCY

HEB ChIPseq

After establishing the TAL1 and HEB interaction in T-ALL cell lines, we further examined the TAL1 transcriptional complex containing the TAL1 and HEB heterodimer with LMO2. We investigated the genomic distribution of the transcriptional complex using Chromatin Immunoprecipitation Sequencing (ChIP-seq). We performed ChIPseq using HEB antibody on chromatin samples isolated from T-ALL cell lines to study its distribution over the genome. Bioinformatic analyses of the HEB ChIP-seq data for ARR, DU.528, HSB-2, and CCRF-CEM identified 8615, 5899, 4980, and 15537 peaks, respectively. **Figure 4.1** shows the ChIP-seq tracks for HEB, TAL1, and LMO2 in the T-ALL cell lines. To validate HEB ChIPseq peaks, we chose two HEB ChIPseq peaks, one annotated to be associated with MYB as the nearest gene and one with SDE2. qPCR primers were designed to specifically amplify the peak region. Primers for CH18 were used as a negative control. qPCR measurements of two duplicate HEB ChIP samples were compared to input chromatin samples. Relative enrichment of HEB ChIPseq samples over the input was demonstrated in bar charts (**Figure 4.2**). The four MYB test samples showed higher enrichment of ChIP over input compared to the four CH18 controls. SDE2 test samples also showed higher enrichment over of ChIP over input compared to CH18 enrichment. HEB genomic distribution in the four T-ALL cell lines was assessed using intersection analysis (**Table 4.1).** The analysis determined the number of peaks with at least 10 bp overlap between HEB ChIPseq for the T-ALL cell lines. Notably, the SIL-TAL cell lines had more overlapping HEB ChIPseq peaks between them (1356) compared to their overlap with ARR (303 peaks).

Table 4.1: HEB ChIPseq peaks overlap in T-ALL cell lines

	ARR	DU.528	HSB-2	CCRF-CEM	
ARR	8615	543	484	934	303
DU.528		5899	1685	2740	1356
HSB-2			4980	2355	
CCRF-CEM				15537	

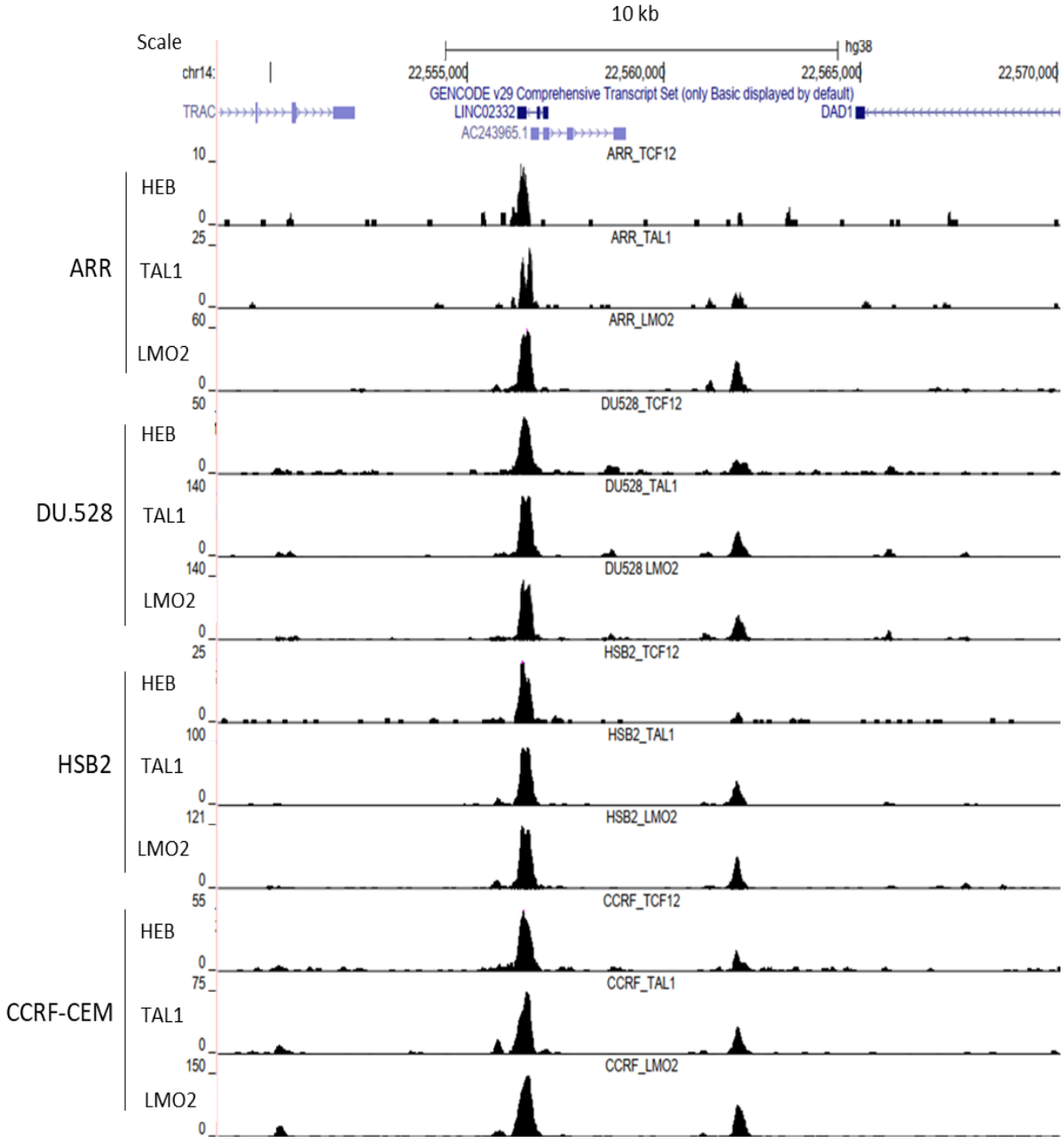


Figure 4.1 Visualisation of ChIPseq tracks is shown for HEB, TAL1, and LMO2 in T-ALL cell lines.

HEB ChIPseq data was collected using chromatin samples isolated from the T-ALL cell lines. HEB ChIPseq tracks were uploaded to UCSC in addition to TAL1 and LMO2 ChIPseq tracks for demonstration. Gencode hg38 genes and a 10 kb scale bar are indicated at the top.

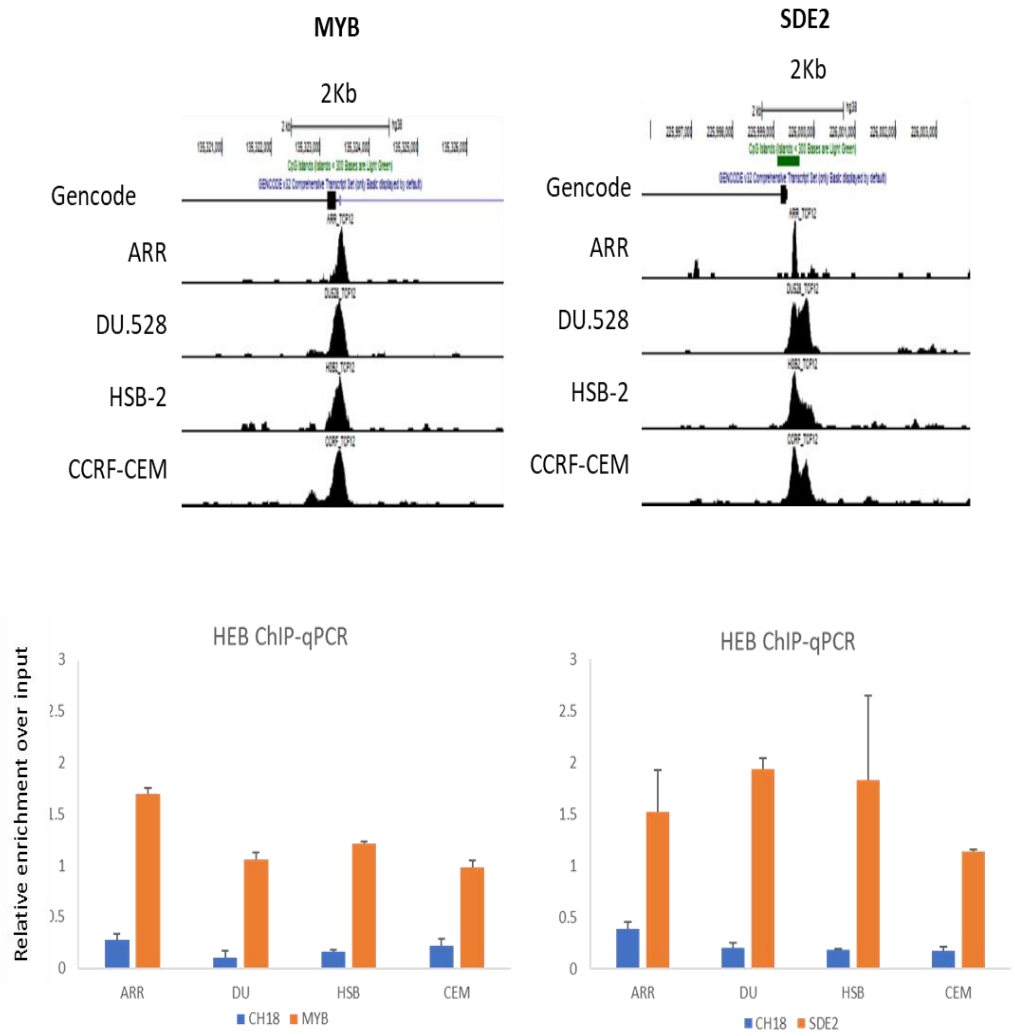


Figure 4.2 HEB ChIPseq validation:

ChIP-qPCR showing the HEB ChIP relative enrichment over input at two HEB ChIPseq peaks regions MYB and SDE2 in T-ALL cell lines. ChIPseq tracks of the HEB ChIPseq peaks are shown. Error bars show variation of duplicate qPCR measurements by calculating the standard error of means. Ch18 Primers are used as a negative control.

We searched for motifs enriched in HEB ChIPseq using *de novo* motif analysis. Significantly enriched motifs were identified for T-ALL cell lines (**Figure 4.3**). In ARR, the enriched motifs did not include E-boxes or any of the motifs usually expected for TAL1/HEB complex like RUNX or GATA. They were however found significantly enriched in DU.528. In HSB-2, RUNX and FLI1 motifs were found and in CCRF-CEM the E-box, ETS, and RUNX motifs were the three most significantly enriched motifs.

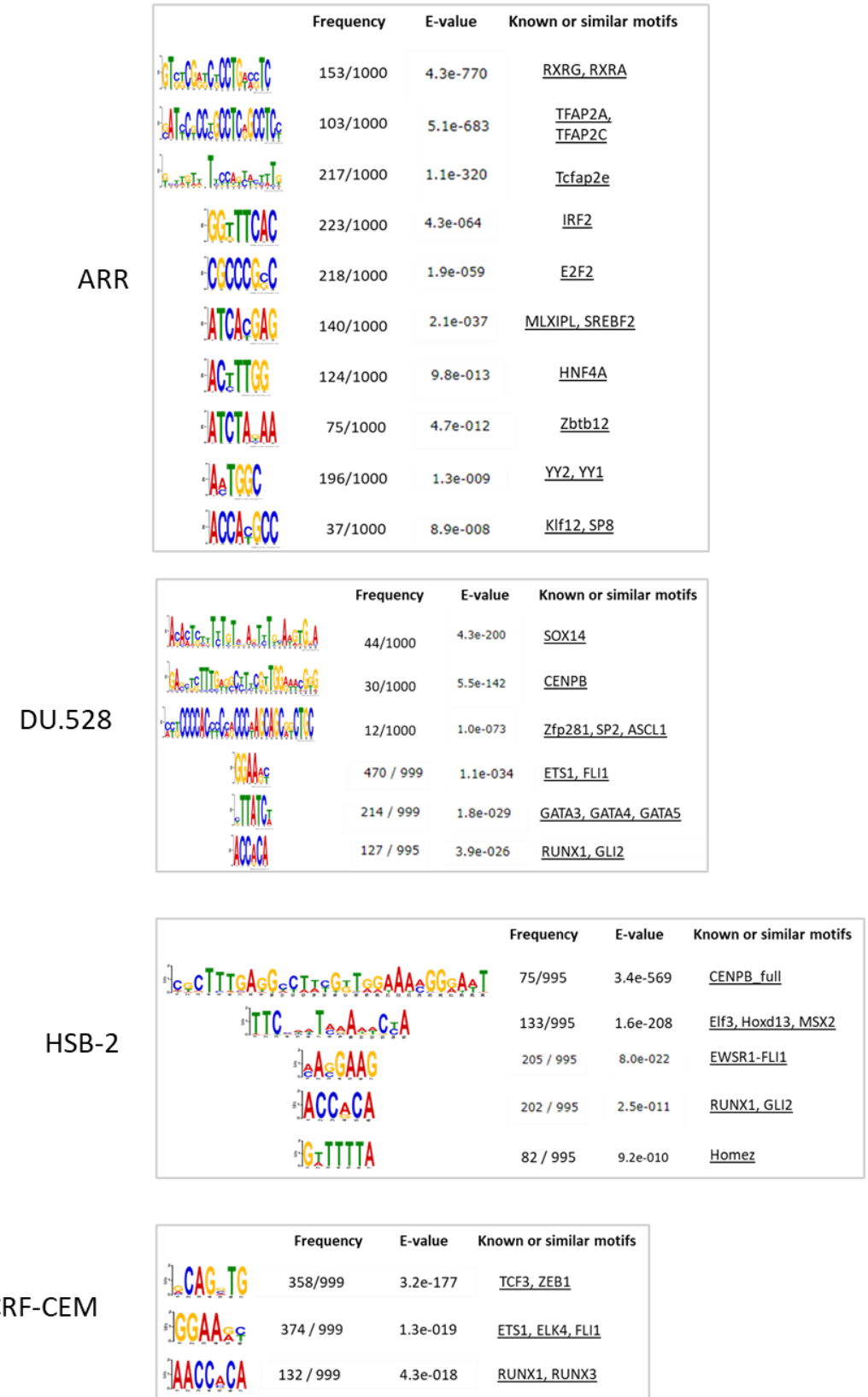


Figure 4.3 *De novo* motif analysis of HEB ChIPseq in T-ALL

Motifs significantly enriched in HEB regions in T-ALL cells are indicated with Expect values (E-values), the frequency the motif was found out of the total number of sequence analysed, and the names of known or similar motifs.

The genomic distribution of HEB, TAL1, and LMO2

HEB ChIPseq was subsequently compared to TAL1, and LMO2 binding patterns identified from ChIP-seq data that was produced by Sarah Binhassan in the Hoogenkamp group. We overlapped the HEB ChIP-seq peaks with those identified from TAL1 and LMO2 ChIPseq to evaluate the proportional co-occupancy of HEB and TAL1 and LMO2. **Figure 4.4** shows Venn diagrams, illustrating the results of the intersection between the HEB, TAL1, LMO2 data for each T-ALL cell line. The overlap between HEB, TAL1, and LMO2 was 1079, 3920, 2061, and 8915 peaks reflecting 12.5%, 66.4%, 41.4%, and 57.4% of HEB ChIP-seq peaks in ARR, DU.528, HSB-2, and CCRF-CEM respectively. The 12.5% intersection of HEB with TAL1 and LMO2 in ARR was considerably smaller than their overlap in the SIL-TAL cell lines (DU.528, HSB-2 and CCRF-CEM), where the average of the overlapping peaks was 55%.

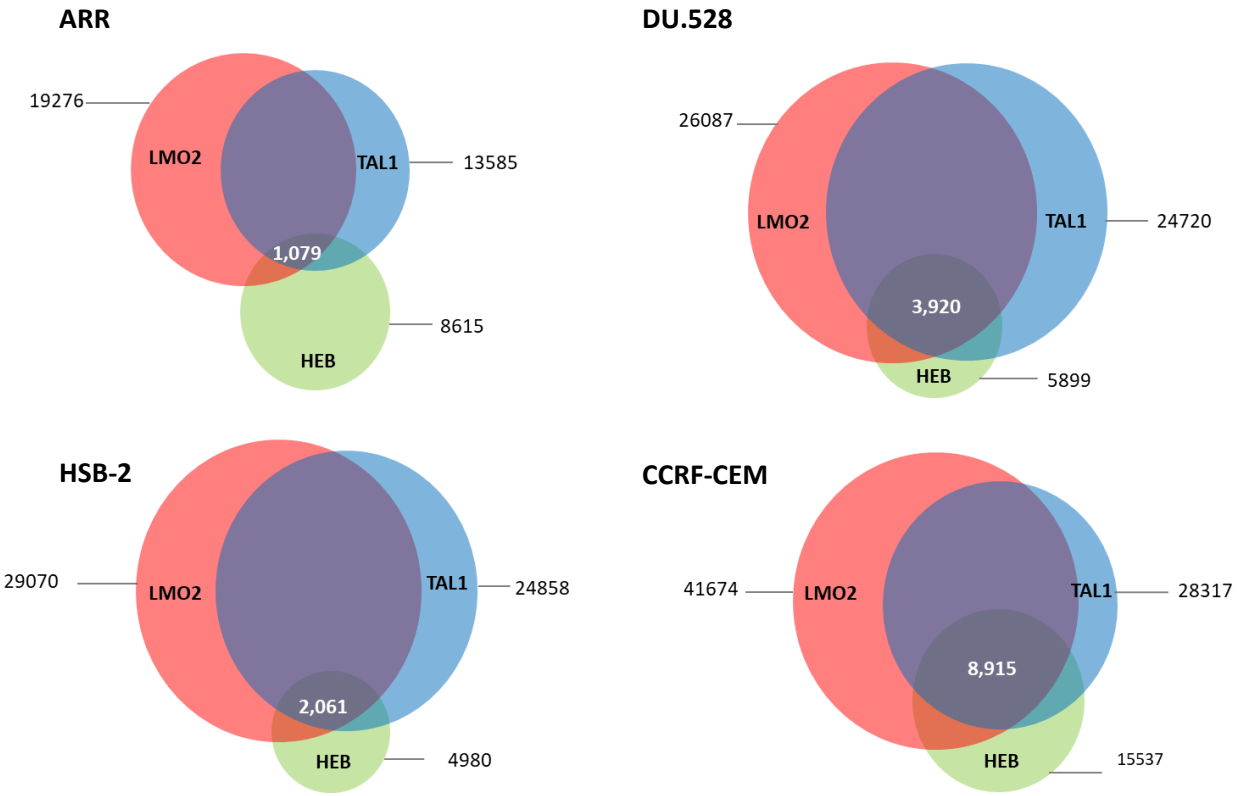


Figure 4.4 Venn diagrams of overlapping ChIPseq peaks in T-ALL cell lines

The overlap between TAL1, LMO2, and HEB ChIPseq peaks in T-ALL cell lines was evaluated using intersection analysis to identify peaks with at least 10 bp overlap. The number of peaks for each ChIPseq dataset is indicated outside the venn-diagrams and the number of the overlapping peaks is shown in the middle.

Table 4.2 The overlapping peaks of HTL ChIPseq in the four T-ALL cell lines

	ARR	DU.528	HSB-2	CCRF-CEM		
ARR	1078	408	332	572	233	All four cells
DU.528		3943	1113	1913	678	SIL-TAL only
HSB2			2060	1464		
CCRF-CEM				8920		

HEB-TAL1-LMO2 complex (HTL) versus HEB-only

Because the majority of HEB peaks overlapped with TAL1 and LMO2 together, we divided the HEB ChIP-seq data into two parts; HEB peaks overlapping with TAL1 and LMO2, which we named the HTL complex and the HEB-only, referring to HEB peaks that remained after subtracting peaks overlapping with TAL1 or LMO2. We compared the overlapping peaks of the HTL complex in the four T-ALL cell lines. **Table 4.2** summarises the comparison across the cell lines, which showed that HEB peaks overlapping with TAL1 and LMO2 were more comparable across the SIL-TAL cells and less with ARR.

De novo motif analysis was applied to identify potential interacting transcription factors with TAL1 transcriptional complex. **Figure 4.5** shows the motifs significantly enriched within the DNA sequences of the HTL complex ChIP-seq peaks. E-box, ETS, GATA, and RUNX motifs were the top motifs in the four T-ALL cell lines. Several other transcription factor motifs were found, such as for SP factors, ELK4, which is another ETS motif, FOS (AP1) and MEF2C.

ARR

Motif found	E-value	Known or similar motifs	Frequency
	6.00E-109	E-box	352/1079
	5.90E-89	ETS	258/1079
	1.20E-36	RUNX1	180/1079
	5.90E-25	GATA3	202 / 1079
	5.70E-03	Hdx	32 / 1079

DU.528

Motif found	E-value	Known or similar motifs	Frequency
	1.8e-165	ETS	1867 / 3920
	6.7e-154	RUNX1	1440 / 3920
	1.1e-105	GATA3	883 / 3920
	6.0e-039	E-box	1179 / 3920
	2.1e-019	Spdef	261/3920

HSB-2

Motif found	E-value	Known or similar motifs	Frequency
	1.3e-103	RUNX1	271/2061
	1.0e-082	ETS	212/2061
	9.2e-053	GATA3	420 / 2061
	5.2e-033	E-box	89/2061
	4.5e-028	ZNF384	20/2061
	5.5e-004	SP2	209 / 2061

CCRF-CEM

Motif found	E-value	Known or similar motifs	Frequency
	4.5e-305	E-box	4139 / 8915
	1.5e-275	ETS	3690 / 8915
	1.3e-203	RUNX1	2386 / 8915
	3.4e-112	GATA3	1405 / 8915
	1.5e-059	ELK4	354 / 8915
	3.3e-015	TBX5	6/8915
	8.8e-011	FOS	1058 / 8915
	1.9e-002	MEF2C	21 / 8915

Figure 4.5 De novo motif analysis of HEB-TAL1-LMO2 overlapping ChIPseq peaks in T-ALL

Motifs significantly enriched in the HEB-TAL1-LMO2 overlapping regions in T-ALL cells are indicated. The tables list Expect values (E-values), the frequency of the motifs reported and the names of known or similar motifs.

Furthermore, to delineate the significance of the genomic regions bound with TAL1, HEB and LMO2 functional annotation bioinformatics analysis was used. Since many of HTL ChIPseq peaks in DU.528, HSB-2, and CCRF-CEM (SIL-TAL cell lines) overlapped, we analysed the SIL-TAL overlapping HTL ChIPseq data set and the ARR data set. As shown in **figure 4.6**, the genomic regions of the HTL ChIPseq peaks in ARR were predicted to be regulating genes of cellular pathways including cytokines, interleukin, PI3 Kinase, histamine h1 receptor. The regions of the SIL-TAL dataset were found to be linked to regulating apoptosis, growth factors like PDGF, EGF and VEGF, angiogenesis, T and B cell activation, and the JAK/STAT signalling pathway. The distance between the HTL transcription factors binding and the nearest transcription start site was estimated to be mostly within 50-500 kb of the TSS for both the ARR and SIL-TAL datasets, as shown in **figure 4.7**.

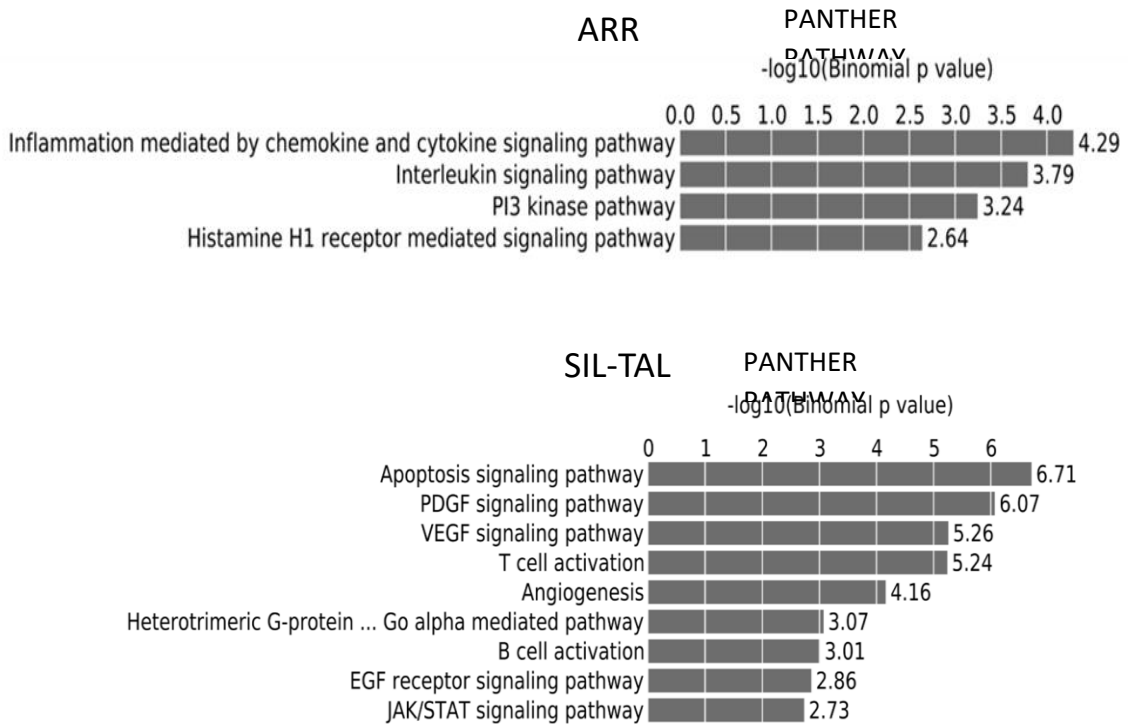


Figure 4.6 The annotation of the genes associated with HTL peaks

The genes associated with the HTL overlapping regions were analysed for significant enrichment. The significantly enriched pathways are shown for ARR and SIL-TAL T-ALL cell lines.

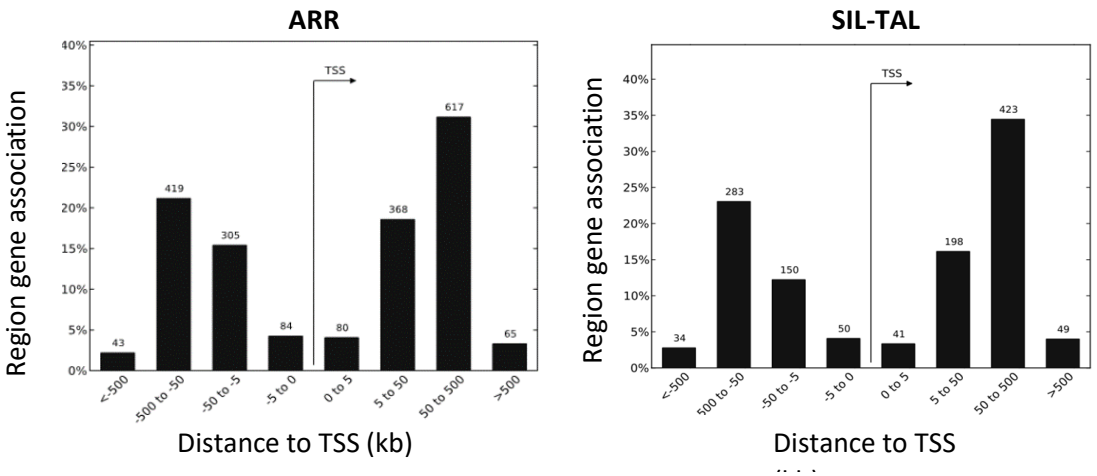


Figure 4.7 Region-gene association for the HTL ChIPseq peaks

Bar charts show the distance and orientation to the transcription start site for regions bound by HTL complex in ARR and SIL-TAL.

Characteristics of the genomic occupancy of HEB and TAL1 transcriptional complex in T-ALL

Many factors modulate the genomic distribution of TAL1 transcriptional complex such as which GATA is expressed (Wu et al., 2014), the association with coactivators (Huang & Brandt, 2000; Huang et al., 1999; Huang et al., 2000; Xu et al., 2006), suppressors (Cai et al., 2009; Goardon, Nicolas et al., 2006) or additional transcription factors such as RUNX, and ETS transcription factors (Palii et al., 2011). Validating TAL1 and HEB protein interaction in the T-ALL cell lines was as a setup for studying the genomic occupancy of TAL1 transcriptional complex containing HEB using ChIPseq. We analysed the overlapping HEB, TAL1, and LMO2 genomic occupancy in the four T-ALL cell lines. We showed that ARR had the least overlap between the three proteins, compared to the three SIL-TAL cell lines (**Figure 4.4**). HEB overlap with TAL1 ChIP was similar to HEB overlap with LMO2, *i.e.* nearly all of HEB overlap with TAL1 was shared with LMO2, as shown by the ChIPseq intersection analysis. Because of this, we divided the HEB genomic occupancy into two groups, HEB, TAL1 and LMO2 overlapping peaks (HTL) or nonoverlapping peaks (HEB-only). HTL peaks overlapped clearly more between the three SIL-TAL cell lines than with ARR. This difference points out the significance of recognising T-ALL heterogeneity.

The gene association analysis of the HTL peaks in the three SIL-TAL cell lines, and in ARR showed that the distribution of the HTL peaks around the transcriptional start sites of the associated genes was very similar in ARR and SIL-TAL cells. As per the analysis of HTL peaks, most peaks were within 50 to 500 KB from the transcription start site of the associated genes in T-ALL cells (**Figure 4.7**). This suggests that TAL1 transcriptional complex primarily bind enhancers distant from the transcription start sites. This is in line with studies showing TAL1 transcriptional complex induce chromatin looping via LDB1 (Krivega et al., 2014) .

The results of *De novo* Motif analysis of the HTL peaks were consistent with previously described motifs with proximity to TAL1 (**Figure 4.5**). Consistent with literature we showed that the five highly enriched motifs in HTL peaks were for TAL1 transcriptional partners E-box, ETS, GATA, and RUNX, in the four T-ALL cell lines (Palii et al., 2011; Palomero et al., 2006; Sanda et al., 2012; Wu et al., 2014). The motif for SP transcription

factors was enriched only in HSB-2. SP1 and TAL1 transcriptional complex was shown to overlap at the promoter of *c-kit* in the haematopoietic TF-1 cell line as demonstrated using ChIP (Lecuyer et al., 2002). ChIPseq analysis of TAL1 transcriptional complex in Jurkat reported the enrichment of SP1 motif in regions occupied with RUNX1 without TAL1 (Sanda et al., 2012).

In the *de novo* motifs (**Figure 4.5**) we showed additional motifs enriched in HTL occupied regions, including HDX in ARR; and ZNF384, JDP2/FOS, MEF2A/MEF2C in CCRF-CEM. The significance of these additional TAL1 overlapping motifs in T-ALL needs further evaluation of the expression of their transcription factors and the association with T-ALL oncogenesis. Oncogenic roles for JDP2 and MEF2C transcription factors, which we found in CCRF-CEM HTL peaks, have been suggested in T-ALL (Homminga et al., 2011; Mansour et al., 2018). Therefore, validating the novel association between TAL1 transcriptional complex and JDP2 or MEF2C could link two presumably unrelated mechanisms of T-ALL pathogenesis.

CHAPTER FIVE

DNA METHYLATION AND THE TAL1 TRANSCRIPTIONAL COMPLEX

We aimed to investigate the characteristics that distinguish the HEB-only regions from the HTL regions. Functional annotation of the HEB ChIPseq data, in general, showed that many of HEB ChIPseq in T-ALL overlapped with a CpG islands dataset. The percentage of HEB-ChIPseq peaks that overlapped with CpG islands represented 10% in SIL-TAL cells and 3% in ARR. The results indicate that HEB ChIPseq peaks overlapped with CpG islands. Therefore, it was intriguing to find out the methylation status of these CpG islands and how the methylation would affect the transcription factor-mediated regulation of gene expression in T-ALL. Genome-wide DNA methylation in T-ALL cell lines was performed by Methylated DNA Immunoprecipitation (MeDIP), which uses antibodies recognising 5-Methylcytosine (5mC). The MeDIP data was produced together with Liam Geraghty, as part of his undergraduate project.

Bioinformatical analysis of MeDIP-seq data identified 20,111 peaks in ARR, 19,264 in DU.528, 18,734 in HSB-2, and 10,823 in CCRF-CEM. The overlap between MeDIPseq peaks for the T-ALL cell lines was demonstrated using a Four-ellipse Venn Diagram (**Figure 5.1**). The figure shows that all four cell lines shared overlapping peaks and 821 peaks were common between the four T-ALL MeDIP datasets. We identified significantly enriched motifs in the top 1000 peaks of MeDIPseq data for each of the T-ALL cell lines. The peaks were sorted by the score, reflecting peak-signal enrichment (**Figure 5.2**).

The intersection of MeDIP peaks in T-ALL with the ChIPseq of HEB, TAL1, and LMO2 revealed that many of HEB-only peaks occurred at methylated regions, while the HEB/TAL1/LMO2 overlapping peaks did not. **Figure 5.3** presents a bar chart of HEB peaks at methylated sites in T-ALL cells. The bars were further segmented into the HTL peaks in red, and the HEB-only in blue. The bar chart illustrates how few of the HEB peaks that were found at methylated regions overlapped with TAL1 and LMO2.

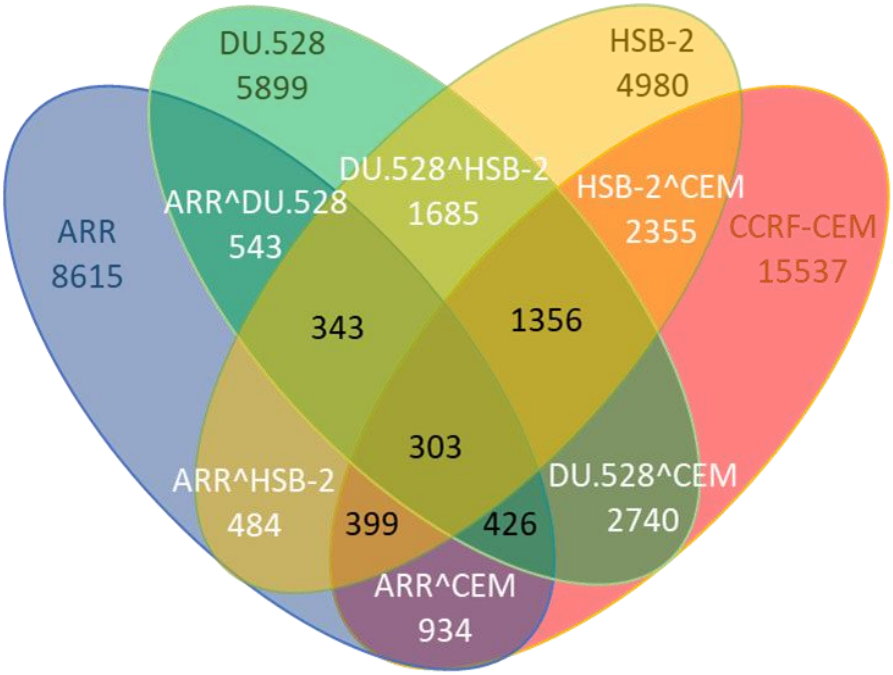


Figure 5.1 MeDIP genomic binding sites overlap in T-ALL.

Intersection analysis of MeDIP ChIPseq peaks in T-ALL cell lines. The number of overlapping peaks is indicated using the Four-ellipse Venn Diagram. Total number of peaks in ARR (20,111), DU.528 (19,264), HSB-2 (18,734), CCRF-CEM (10,823) are shown.

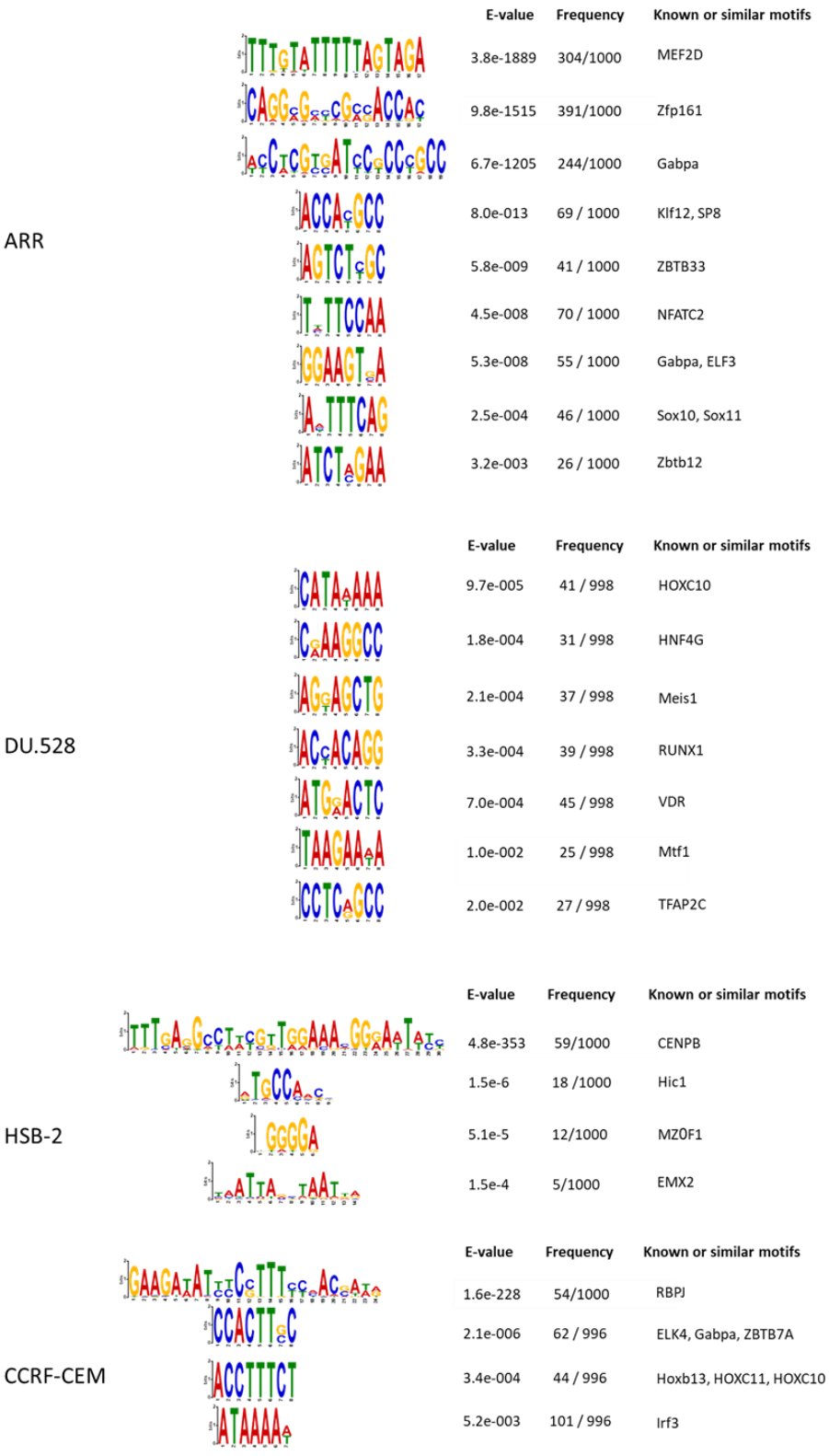


Figure 5.2 De novo motif analysis of MeDIP-seq peaks in T-ALL

Motifs significantly enriched in the top 1000 MeDIP-seq peaks are indicated. The tables list Expect values (E-values), the frequency the motifs were found and the names of known or similar motifs.

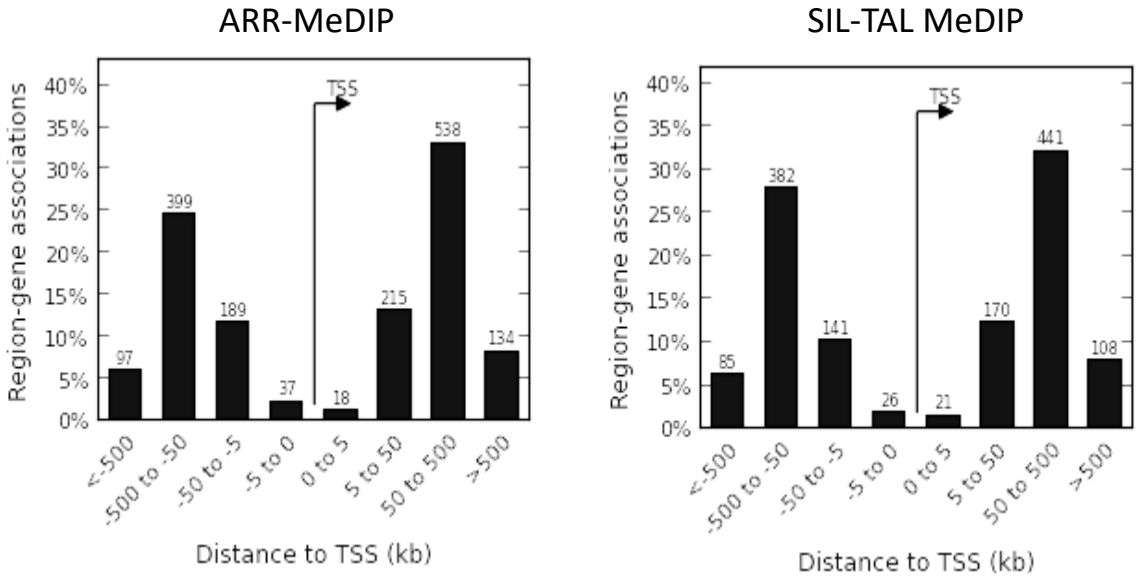


Figure 5.3 MeDIP peaks distance and orientation to the transcription start-site of associated genes

The distance to the transcription start-site are shown for methylated CpGs in ARR and SIL-TAL MeDIPseq top1000 peaks ordered based on score.

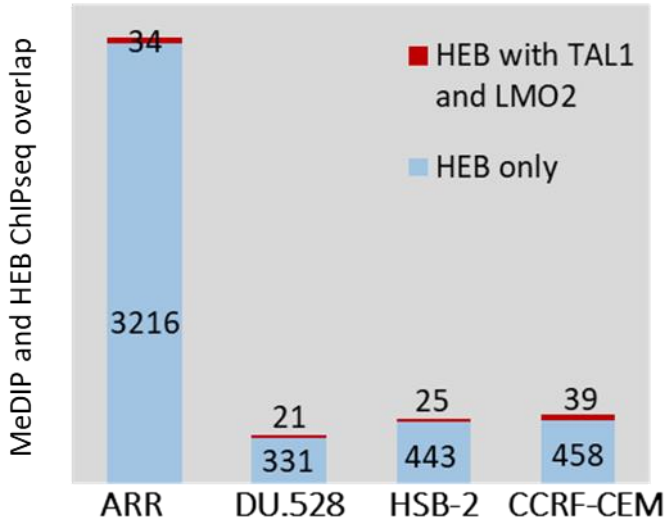


Figure 5.4 Majority of HEB-only peaks occur at methylated DNA regions

Bar chart showing the overlap between HEB ChIPseq and methylated CpGs in T-ALL**.** The proportion that intersects with TAL1 and LMO2is shown in red while HEB-only regions are shown in blue.

HEB peaks with the highest intensity overlap with highly methylated regions but not with TAL1 and LMO2

The analysis of the TAL1 transcriptional complex containing HEB was expanded by utilising a panel of genome-wide datasets produced by the Hoogenkamp lab, including TAL1, LMO2, LDB1, RUNX.1, GATA2, and LYL1 ChIPseq and DNAaseI. To visualise how they relate to each other at HEB regions, peak density heatmaps of the MeDIP and the ChIPseq of the TAL1 transcriptional complex were made. ChIPseq of LYL1, which is structurally and functionally closely related to TAL1, DNaseI hypersensitivity (DHS), and RNAseq were also incorporated in the panel of heatmaps.

To create the heatmaps using Easeq software, we sorted HEB peaks according to their significant enrichment as determined by (-10*log10 p-value). Then, we generated an interval file by setting the start of to 500 bp before the summit and the end at 500 bp after the summit of the peak. The datasets signal was blotted at the indicated intervals, and a window of two kilobases cantered around the summit of HEB peaks is shown in heat maps. The heatmaps in ARR (**Figure 5.5**) demonstrated that LDB1 and MeDIP largely overlapped with HEB. On the other hand, the pattern of TAL1 and LMO2 was different. These were spread across the HEB-occupied regions except at the highly enriched HEB peaks. Those strongest HEB-only peaks, lacking TAL1 and LMO2, co-localised with the

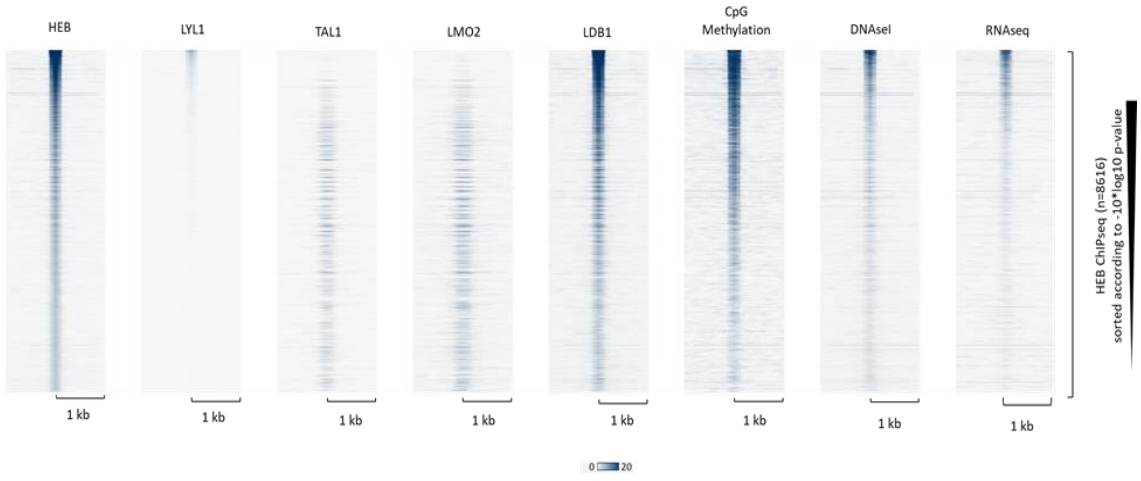


Figure 5.5 Heat maps of ChIPseq of HEB; LYL; TAL1; LMO2; and LDB1, MeDIP and RNAseq in ARR. The Y-axis represents individual HEB ChIPseq regions sorted in descending order according to fold enrichment. The X-axis measures a 2 kb area centred around the summit of HEB ChIPseq peaks. A bar showing the relationship between colouring and signal intensity is found at the bottom of the plot.

highest signals for methylated CpGs. Moreover, enrichment of RNAseq signal was observed for these HEB-only regions at methylated DNA as displayed in the RNAseq heat map.

The significance of the HEB-only peaks with CpG methylation in ARR was evaluated by integrating the enriched regions with RNAseq data of ARR. The RNAseq data we used in this analysis was produced by the Hoogenkamp lab. Since only ARR had many HEB-only peaks with CpG methylation, we retrieved the genes with differential expression between T-ALL cell lines. Then, we found the genes that are differentially expressed and associated with the HEB-only peaks with CpG methylation.

The gene ontology of the identified genes was determined using DAVID 6.8 functional annotation tool. The statistically significant pathways were identified and ranked based on the number of genes, as shown in **Figure 5.6**. First, sixty genes were implicated in metabolic pathways. Moreover, pathways in cancer, the small GTPase RAP1 and PI3K-AKT signalling pathway, and proteoglycan in cancer were also enriched.

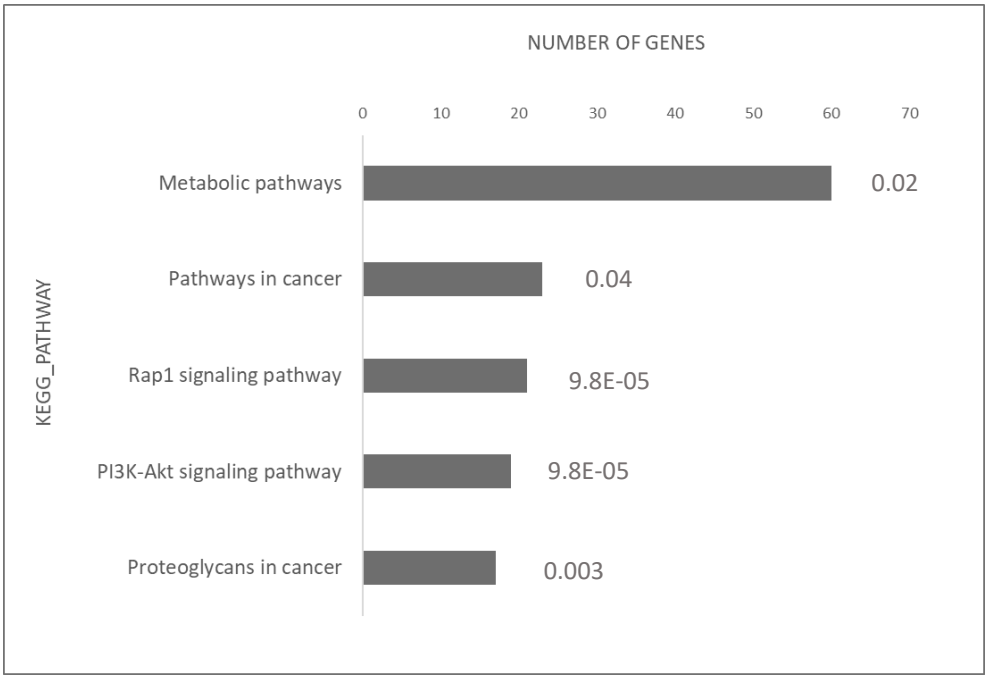


Figure 5.6 Functional annotation of genes differentially expressed in T-ALL cell lines and associated with HEB binding regions. The statistically significant pathways were identified and sorted according to genes number. P-values are indicated on the right. The top five are shown in the bar chart.

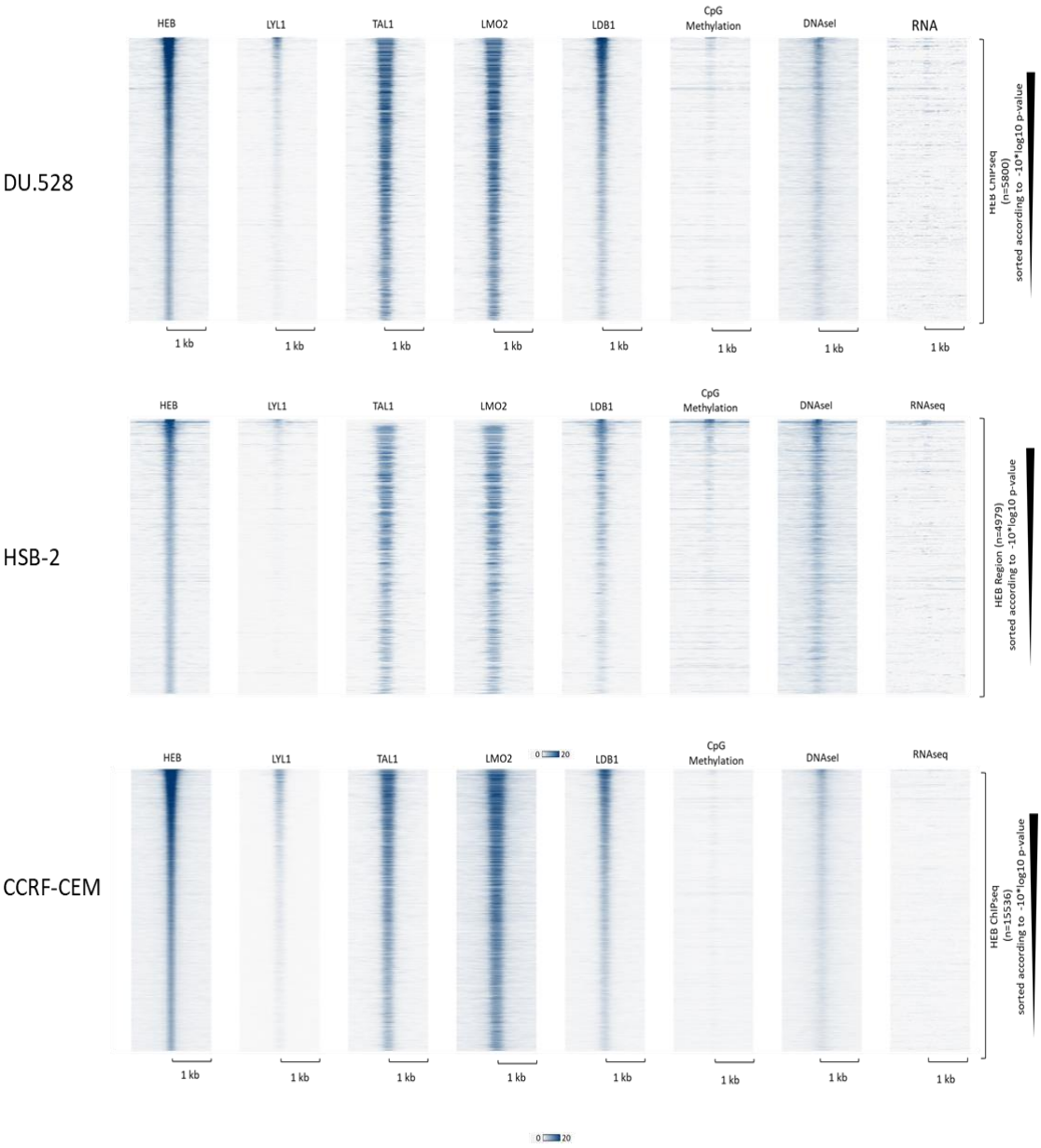


Figure 5.7 Heat maps of ChIPseq of HEB; LYL; TAL1; LMO2; and LDB1, MeDIP, DHS and RNAseq in the four T-ALL cell lines

The Y-axis represents individual HEB ChIPseq regions sorted in descending order according to fold enrichment of HEB ChIPseq peaks. The X-axis measures a 2 kb area centred around the summit of HEB ChIPseq peaks. A bar showing the relationship between colouring and signal intensity is found at the bottom of the plot.

In DU.528, HSB-2, and CCRF-CEM, there was very little methylation, as displayed in the MeDIP heatmap (**Figure 5.7**). Furthermore, in the SIL-TAL cells, the signal density of TAL1 and LMO2 ChIPseq correlated more with the descending order of HEB peaks compared to ARR. This indicates that HEB peaks with the highest intensity colocalise with TAL1 and LMO2 high-intensity peaks, and this was different from ARR. The RNAseq pattern seen in ARR at HEB peaks was not observed in the SIL-TAL cells. LYL1 overlapped with TAL1, LMO2, and LDB1 at the HEB-bound regions in DU.528, and CCRF-CEM. The low number of reads in LYL1 ChIP-seq for ARR and HSB-2 could be related to their lower expression levels of LYL1 protein, as shown using western blot in **Figure 5.8**.

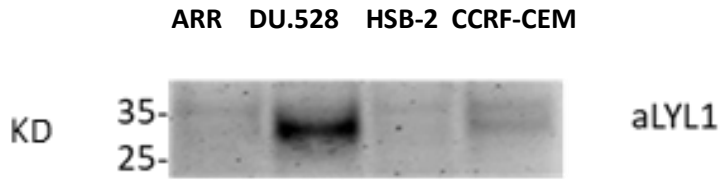


Figure 5.8. LYL1 Western blot analysis in T-ALL
Nuclear extracts from the four T-ALL cell lines was separated on 4-12% Bis-Tris gels and blotted with LYL1 antibody.

The dataset of the HEB-only regions with methylation was analysed using *de novo* motif analysis to identify transcription factor binding sites that could be interacting with E-protein homo/heterodimers in T-ALL cell lines. **Figure 5.9** shows motifs enriched in the HEB-only regions with methylation in the T-ALL cell lines. As ARR had a much higher number of HEB-only peaks with methylation (3216) than the other three cell lines, we selected the top 1000 peaks on the basis of the score of the peaks signal intensity and used them for *de novo* motif analysis. For the SIL-TAL cells, the number of HEB-only peaks with methylation was less than 500, so they were all included. In ARR, motifs identified included those for ESR1/RXRG, AP-2, MEF2D, and others, while GFI1, SOX18, and LHX6, had the highest statistical significance in DU.528, HSB-2, and CCRF-CEM. The motif for CPEB1 was significantly enriched in HSB-2 and CCRF-CEM.

ARR

Motif found	E-value	Known or similar motifs	Frequency
	7.0e-1750	ESR1, RXRG	289/1000
	1.5e-1433	TFAP2A	194/1000
	2.1e-1431	MEF2D	193/1000
	5.7e-447	NFATC2	385/1000
	5.7e-042	MLXIPL, SREBF2	166 / 1000
	1.2e-021	HNF4A	102 / 1000
	1.2e-012	Nkx2-6	77 / 1000
	6.4e-007	Sox3, SOX21, SRY	34 / 1000
	3.9e-006	Ddit3::Cebpa	45 / 1000
	1.1e-005	E2F2	46 / 1000

DU.528

Motif found	E-value	Known or similar motifs	Frequency
	3.5e-520	Gfi1	73/1000
	3.0e-007	Arid5a	1/1000
	5.3e-003	TBX4	30/1000
	6.7e-003	Hand1::Tcf3	35/1000
	1.5e-002	Elf5	14/1000

HSB-2

Motif found	E-value	Known or similar motifs	Frequency
	1.7e-013	SOX18	117 / 443
	2.2e-006	CPEB1	7/443
	2.1e-005	TBP	2/443
	1.0e-004	Hoxd9	11/443
	1.6e-004	FOXL1	12/443

CCRF-CEM

Motif found	E-value	Known or similar motifs	Frequency
	3.4e-557	LHX6	158/458
	4.1e-009	Zfx	52 / 458
	5.9e-008	REL	45 / 458
	2.9e-005	Zfp652	36 / 458
	3.2e-003	CPEB1	29 / 458
	6.2e-003	PRDM1	3/458
	3.4e-002	Nkx2-2	12/458

Figure 5.9 Motifs significantly enriched in HEB-only regions with methylation in T-ALL

Motifs significantly enriched in the top 1000 HEB-only peaks are indicated. The tables list Expect values (E-values), the names of known or similar motifs, and the frequency the motifs were found.

Mass spectrometry of HEB co-immunoprecipitation

HEB protein interactions were investigated using co-immunoprecipitation and mass spectrometry. Firstly, to compare TAL1 and HEB networks, and secondly to identify protein interactions supporting specific roles of HEB, like its association with DNA methylation. HEB co-immunoprecipitation was performed in the four T-ALL cell lines followed with mass spectrometry analysis. After removal of the proteins identified in the IgG control, a four-way comparison between the T-ALL cell lines for proteins detected in HEB MS was demonstrated using Venn-diagram **(Figure 5.10).** Twenty-eight proteins were found common between all the T-ALL cell lines and another 28 were shared between the SIL-TAL cells.

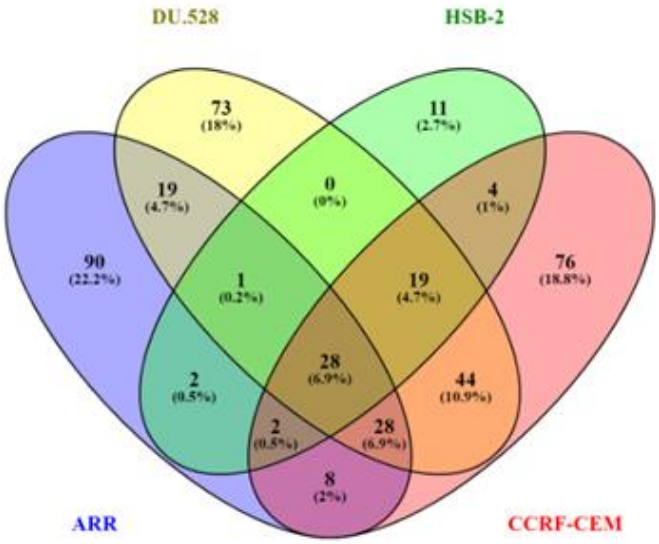


Figure 5.10 Venn-diagram of proteins pulled down in HEB immunoprecipitation (IP) in T-ALL cells

HEB immunoprecipitation (IP) was prepared using nuclear extract samples of the four T-ALL cell lines. Mass spectrometry was used to identify proteins in HEB pulldown. The figure shows four-way comparison of protein detected in HEB MS after the exclusion of proteins found in the IgG control.

The HEB mass spec data were filtered using the CRApome database to remove probable false positives that are commonly reported in mass spectrometry results. The protein was excluded if it was found in more than 20 reported experiments as a cut-off. The list of proteins remaining were compared and proteins shared between at least three T-ALL cell lines are demonstrated in **Table 5.1.** The remaining proteins were then sorted based on their abundance using the number of peptides detected in the mass spectrometry. The top ten proteins for each of the T-ALL cell lines are shown in **Table 5.2.**

Table 5.1 Proteins of HEB MS that were found shared between the T-ALL cell lines

Cell line	HEB MS proteins	
ARR, DU.528, and CCRF-CEM	RAC2	Rac Family Small GTPase 2
	ARHGEF2	Rho/Rac Guanine Nucleotide Exchange Factor 2
	GMFB	Glia Maturation Factor Beta
DU.528, HSB-2 and CCRF-CEM	MYO1G	Myosin IG
	CPT2	Carnitine O-palmitoyltransferase 2

Table 5.2 Top10 proteins found in HEB MS for T-ALL cell lines

ARR

Peptides	Protein	Mapped Gene Symbol
7	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3	PFKFB3
3	Transcription factor 12	TCF12
2	Magnesium transporter protein 1	MAGT1
2	Ras-related C3 botulinum toxin substrate 2	RAC2
1	Coronin-1A	ABCA1
1	Actin-related protein 2/3 complex subunit 5	ARHGEF2
1	ATP-binding cassette sub-family A member 1	ARPC5
1	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1	CORO1A
1	Rho guanine nucleotide exchange factor 2	DAD1

DU.528

Peptides	Protein	Mapped Gene Symbol
15	Actin-related protein 2/3 complex subunit	ARPC3
7	C2 domain-containing protein 2-like	C2CD2L
7	PDZ domain-containing protein GIPC1	GIPC1
6	Germ cell-specific gene 1 protein	GSG1
4	Leukocyte elastase inhibitor	SERPINB1
4	Succinyl-CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial	SUCLG1
4	Tubulin beta-3 chain	USP31
3	2,4-dienoyl-CoA reductase,	DECR1
3	Zinc finger protein Gfi-1	GFI1

HSB-2

#Peptides	Protein	Mapped Gene Symbol
2	Phospholipid transfer protein	C2CD2L
1	Coiled-coil domain containing 87	CCDC87
1	Transmembrane protein 133	TMEM133
1	Coiled-coil domain containing 186	C10orf118
1	Differentially expressed in FDCP 6 homolog; Phosphatidylinositol 3	DEF6
1	Brefeldin A-inhibited guanine nucleotide-exchange protein 2	ARFGEF2
1	Collagen alpha-3(VI) chain	COL6A3
1	Girdin	CCDC88A
1	Unconventional myosin-Ig	MYO1G
1	Vasodilator-stimulated phosphoprotein	VASP

CCRF-CEM

#Peptides	Protein	Mapped Gene Symbol
14	Arf-GAP domain and FG repeat-containing protein 1	AGFG1
13	Carnitine O-palmitoyltransferase 2, mitochondrial	CPT2
12	DNA mismatch repair protein Msh3	MSH3
11	DNA-binding protein Ikaros	IKZF1
11	Dynein heavy chain 8,	DNAH8
10	Dyslexia susceptibility 1 candidate gene 1 protein	DYX1C1
9	Glia maturation factor beta	GMFB
9	Galectin-9	LGALS9
9	F-box-like/WD repeat-containing protein	TBL1XR1

The list of all the proteins identified in HEB mass-spectrometry was analysed (Appendix), then after removing the ribosomal proteins and splicing factors was depicted in a proteins network. Interestingly, after having found the GFI1 motif enriched in the HEB-only peaks with (**Figure 5.9**), the mass spectrometry results of the HEB pulldown identified the presence of GFI1 in two of the T-ALL cell lines; ARR and DU.528. GFI1 was also one of the top 10 proteins found in DU.528 **Table 5.2**. Using STRING functional enrichment tool, we showed known interactions in HEB MS protein data in the four T-ALL cell lines combined as displayed in **figure 5.11**. The analysis identified the significantly enriched pathways represented in HEB mass spectrometry network analysis (**Figure 5.12**). The top three pathways were related to the cell cycle. Additionally, the Rho GTPases and cellular response to stress pathways were significantly enriched.

Since the TAL1 mass spectrometry was also significantly reflective of cell cycle and Rho-GTPases, the protein datasets were compared to each other, and 56 proteins were found in both of TAL1 and HEB mass spectrometry. Protein-protein interactions of the overlapping data are shown in **figure 5.13**. The comparison emphasised a list of proteins such as the interaction with the transcription factor Ikaros, which is known to regulate T cell development. The mass spectrometry data from the HEB immunoprecipitation was also used to search for proteins associated with DNA methylation. Using the ARR-exclusive HEB mass spectrometry data, DNA methylation was one of the enriched pathways. In **figure 5.14**, the red nodes represent the two proteins that are known to be directly associated with DNA methylation, CHTOP and ERH. Protein ERH was also found in TAL1 mass spectrometry of ARR, DU528, and CCRF-CEM. However, this finding needs to be supported with additional evidence because CHTOP and ERH both were flagged up as potential false positive based on the CRAPome database. The number of experiments CHTOP was found according to CRAPome was 51/411 and for ERH was 150/411.

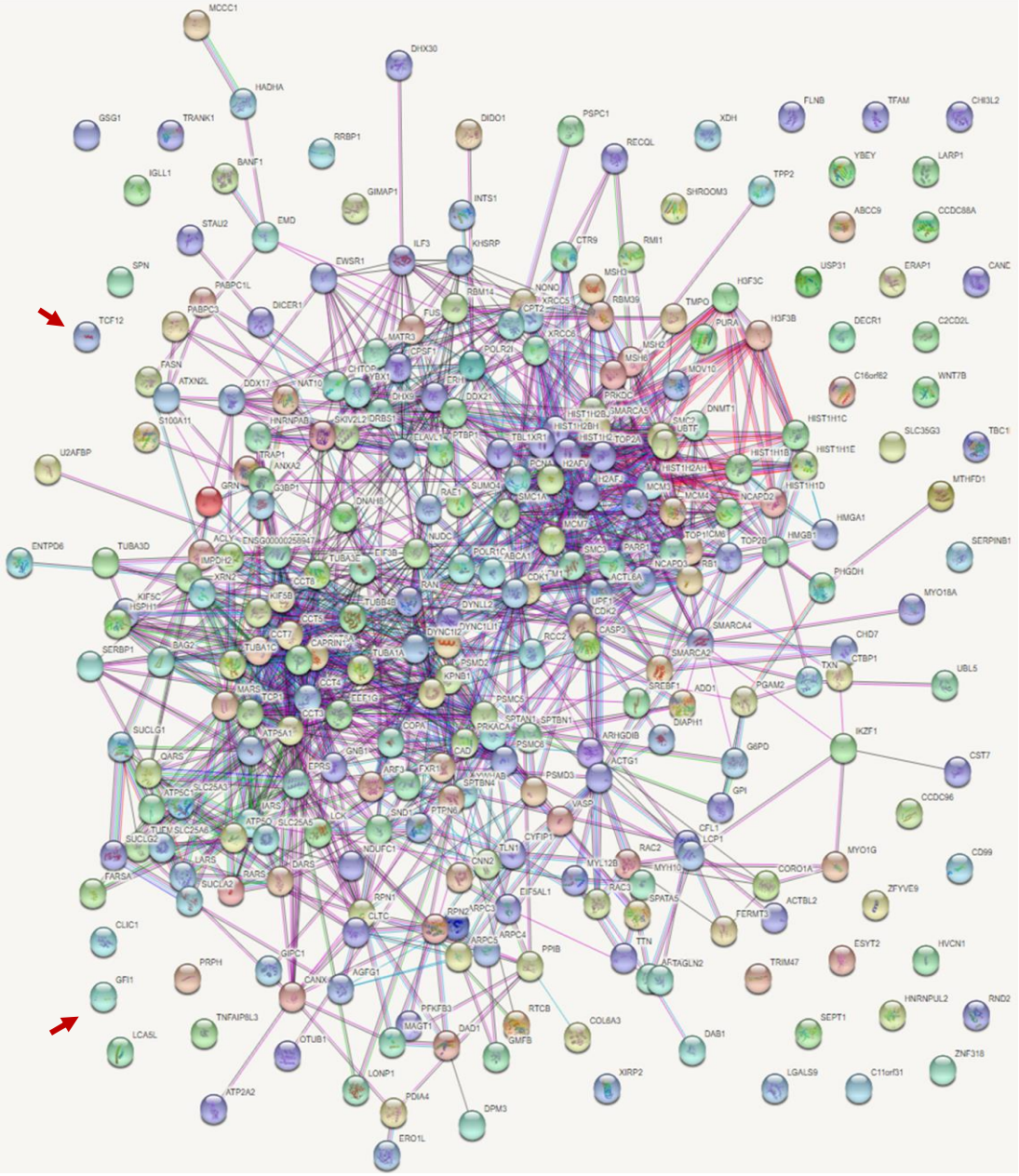


Figure 5.11 A network of known protein-protein interactions in HEB MS in T-ALL cell lines.

Nodes and connecting lines represent proteins and known interactions, respectively. Data represent the proteins identified in HEB MS of the four T-ALL cell lines together. Ribosomal and splicing factors were excluded in the network analysis.

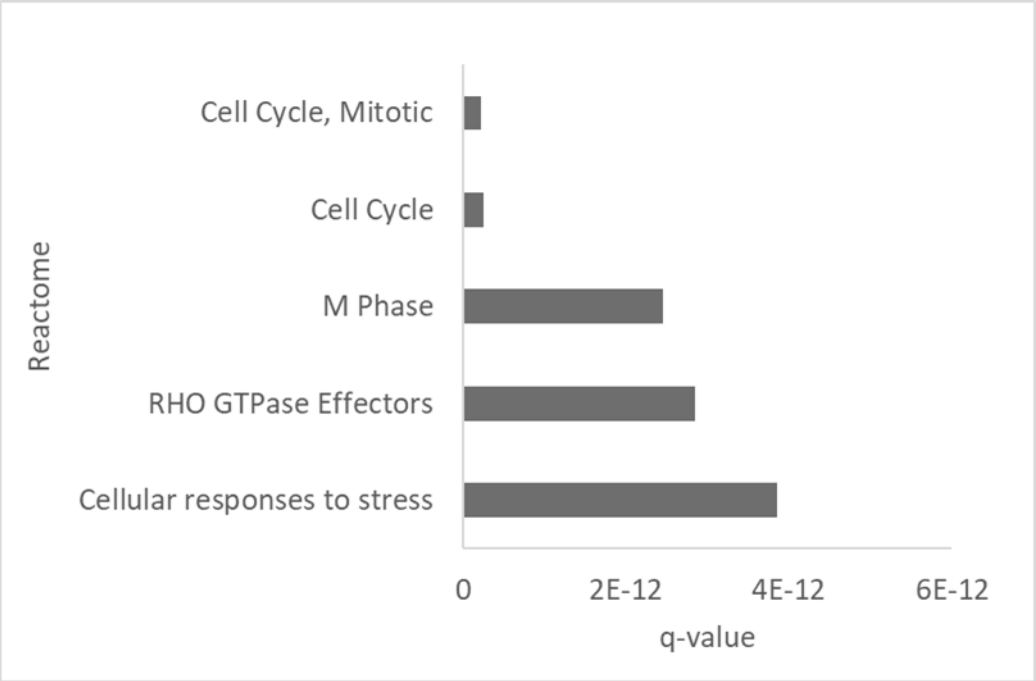


Figure 5.12 A Bar chart shows the top five statistically significant pathways enriched in HEB MS

Functional annotation of proteins detected in HEB mass spectrometry analysis in T-ALL cell lines identified significantly represented pathways. Statistical significance of the top five pathways are shown using q-values (P-value after adjustment for false discovery rate).

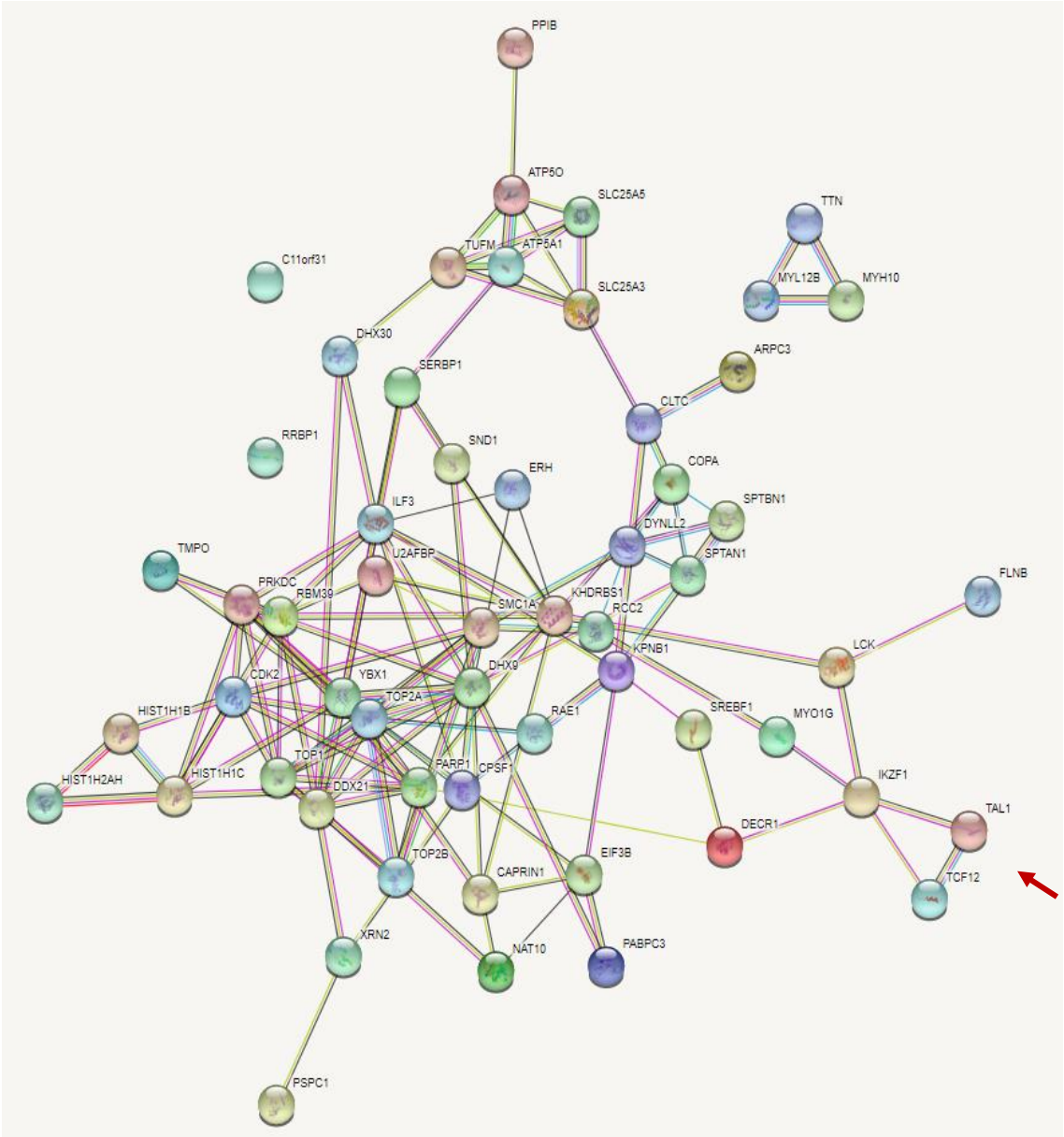


Figure 5.13: A network of protein-protein interactions detected in both of TAL1 and HEB pulldowns using mass spectrometry analysis in T-ALL

Nodes and connecting lines represent proteins and known interactions, respectively. The combined data of the four T-ALL cell lines was used in the comparison between TAL1 and HEB pulldown proteins. Ribosomal and splicing factors were excluded from the network analysis. The arrow indicates TAL1 and HEB.

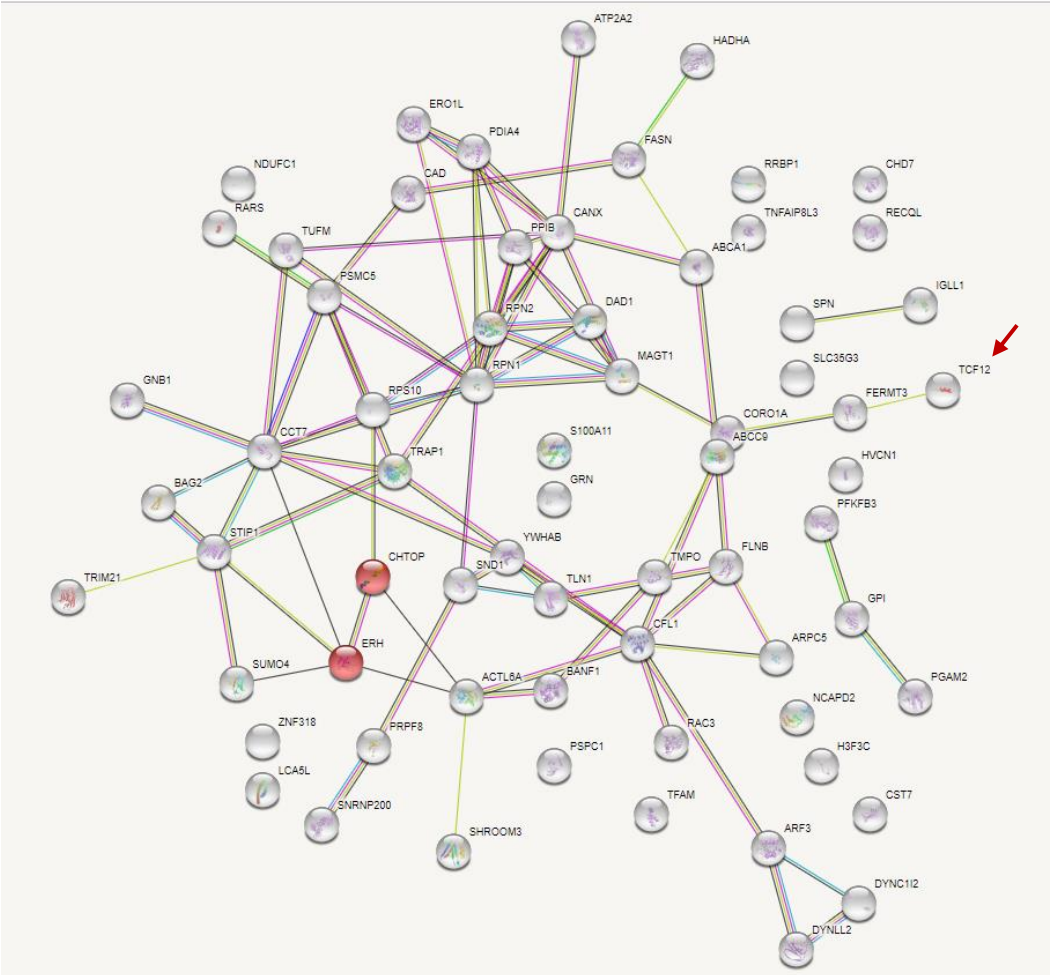


Figure 5.14 Network of protein interactions in proteins detected in HEB mass spectrometry and exclusively in ARR.

Proteins implicated in DNA methylation are highlighted in red and the arrow points to HEB (TCF12).

Discussion:

DNA methylation and the transcriptional complex of TAL1, LMO2, and HEB

In addition to the TAL1 transcriptional complex-mediated regulation of T-ALL genes, the change in the E-proteins transcriptional program is a fundamental mechanism of T-ALL pathogenesis (Engel & Murre, 2001). In order to gain further insight into this, we analysed the HEB-only regions which did not overlap with TAL1 and LMO2. This analysis aimed to find characteristics of HEB-only regions that would be distinct from the HTL complex genomic occupancy. We hypothesised that DNA methylation status would be an essential determining factor of the genomic occupancy of TAL1 transcriptional complex. Mechanisms of how DNA methylation is related to aberrant transcriptional regulation in cancer has been described. Recently, the TAL1 expression level in paediatric T-ALL patients correlated with (the CpG Island Methylator Phenotype) CIMP- subgroup, defined by lower DNA methylation and worse prognosis (Haider et al., 2019). How DNA methylation relates to the genomic distribution of the TAL1 complex, has not been described.

Using MeDIP-seq analysis, we mapped DNA methylation in the four T-ALL cell lines. Then, we overlapped the MeDIP-seq data with the ChIPseq of TAL1 partners HEB, LMO2, LDB1 and its homolog LYL1 (**Figure 5.8**). We showed that HEB regions overlapped with more highly methylated regions except in the HEB, TAL1 and LMO2 overlapping regions. In the HTL overlapping peaks, there was minimal overlap with MeDIP peaks, indicating hypomethylation of regions bound with the HTL complex. In ARR, the number of HEB-only regions was higher, so HEB overlap with the highly methylated regions was evident. HEB overlap with methylated DNA also contained LDB1 and LYL1. The results of ChIPseq overlap analysis suggest that a transcriptional complex of LYL1, HEB, and LDB1 colocalise with methylated CpG. However, we do not know if HEB is inducing DNA methylation or recruited to methylated DNA.

De novo motif analysis of methylated HEB-only peaks showed that they were different from motifs enriched in the HTL regions (**Figure 5.9**), except for the motif for MEF2 in ARR and the E-box motif in DU.528. We found a motif for CPEB1 (Cytoplasmic Polyadenylation Element Binding Protein 1) in two cell lines HSB-2 and CCRF-CEM data of methylated HEB-only regions; no other motif was shared between the four T-ALL cell lines.

HEB mass spectrometry showed a potential association with proteins known to bind 5hmC

In order to look for HEB interacting proteins that were associated with DNA methylation, we performed HEB pulldowns followed by mass spectrometry. Because HEB significantly overlapped with DNA methylation in ARR, we looked at the HEB MS proteins that were unique to ARR. Annotation of ARR unique proteins in HEB MS showed that DNA methylation was one of the enriched processes. In the ARR-unique HEB MS data, two proteins were present which have previously been associated with DNA methylation (**Figure 5.14**); Chromatin target of PRMT1 protein (CHTOP) and Enhancer of rudimentary homolog (ERH). CHTOP and ERH, in addition to Protein arginine methyltransferases, PRMT1 and PRMT5, have been shown to form the methylosome complex in glioblastoma to associate with 5hmC and induce methylation of H4R3 and activate the expression of oncogenic genes (Takai et al., 2014). We found ERH in TAL1 MS, but CHTOP was only in ARR HEB MS. The MS results indicating an interaction between HEB and CHTOP require confirmation by co-IP experiments followed by western blotting and with other techniques. Our results indicating HEB association with the methylosome complex are suggestive of a mechanism of how DNA methylation relates to the genomic distribution of TAL1 and HEB transcriptional complex in T-ALL.

Functional annotation of methylated HEB-only regions

We showed the significance of the highly methylated HEB-only peaks in ARR by integrating the ChIPseq and RNAseq analyses. First, we identified the genes with differential expression in the four T-ALL cell lines that were simultaneously associated with HEB-only regions with methylation. Then, we restricted the analysis to differentially expressed genes because the HEB-only peaks that are highly methylated were found more in ARR compared to the SIL-TAL cell lines. Gene ontology of the identified genes showed terms that were related to metabolic pathways and pathways in cancer (**Figure 5.6**). Our results show that DNA methylation status is linked to the genomic distribution of TAL1 transcriptional complex and describe a novel aspect of TAL1 contribution to T-ALL pathogenesis. The results provide support for implicating DNA methylation as a feature distinguishing TAL1 and LYL1 genomic distribution in T-ALL.

CHAPTER SIX

CHARACTERISATION OF T-ALL CELL LINES STABLY EXPRESSING TAL1 ISOFORMS.

TAL1 gene code for four TAL1 isoforms. Structural differences between TAL1 isoform could result in functional differences between the isoforms. We hypothesised that the characteristics of the TAL1 complex could differ in its distribution in the genome, its interactions, or effect, depending on which TAL1 isoforms is in the complex. In order to examine TAL1 isoforms in T-ALL and characterise potential isoform-specific roles, we needed to clone TAL1 isoforms. We designed four constructs for the fluorescent protein tdTomato and FLAG-tagged TAL1 isoforms with T2A between the two sequences, as shown in **figure 6.1**.

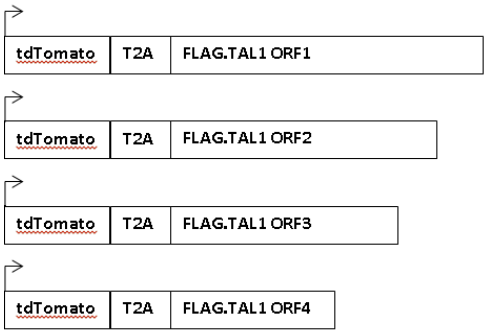


Figure 6.1 tdTomato T2A FLAG TAL1 constructs.
Diagram shows the design of the four FLAG-tagged TAL1 cDNA constructs containing the red fluorescent protein tdTmato separated by the T2A peptide cleavage sequence.

TAL1 isoform constructs were initially cloned into doxycycline-inducible lentiviral vectors (PLVX Tet one puro). Even though the constructed inducible lentiviral vectors of TAL1 isoforms were efficient in HEK cells, neuroblastoma cell lines, and in stimulated human CD34+ progenitors, we were not able to transduce them into T-ALL cell lines. Despite considerable efforts in optimising the transduction and generation of more concentrated viruses, the T-ALL cells were not accessible, possibly because of the lack of appropriate receptor. As an alternative approach, we tested a piggyBac transposition system in T-ALL. The PB vector (PB-CAG eGFP) was efficiently electroporated into T-ALL cell lines as shown by stable expression of the GFP fluorescent protein. Consequently, we subcloned our TAL1 isoforms constructs into the PB-CAG vector after removal of the eGFP cDNA. We generated four PB plasmid for the fluorescent protein

tdTomato and FLAG-tagged TAL1 isoforms with T2A between the two sequences. To test the constructs, we transfected HEK-293 cells with the constructs and collected total proteins 24 hours after the transfection. The protein samples of the transfected and untransfected cells were run on Bis-Tris polyacrylamide gels and blotted with either FLAG or TAL1 antibodies. **Figure 6.2** shows the four isoforms of TAL1 in HEK-293 cells transfected with tdTommato-T2A-FLAG-TAL1-ORF1, ORF2, ORF3, and ORF4 (lane 2-4) of FLAG and TAL1 western blot, while no FLAG or TAL1 was seen in untransfected HEK-293 (-ve Ctl) (lane 1). Therefore, the results confirm the correct expression by the constructs.

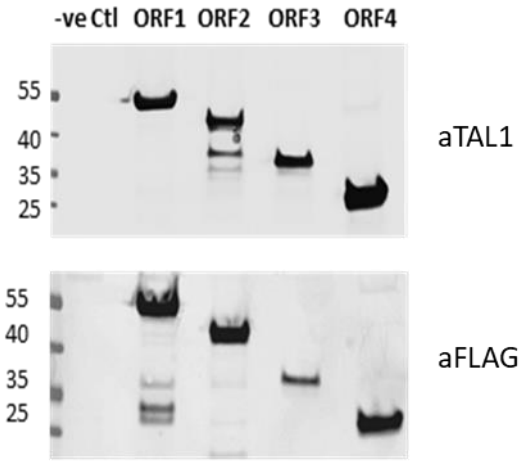


Figure 6.2 Testing PB TAL1 isoforms plasmids using TAL1 and FLAG western blots

Total protein samples were isolated from HEK-293 cells 24 hours post-transfection with PB-CAG-tdTomatoT2A-FLAG-TAL1 ORF1-4. Protein samples were separated on 4-12% Bis-Tris gel along with a sample of untransfected HEK-293 cells as a negative control (-ve Ctl). The gel was blotted with TAL1 and FLAG antibody.

After validating the constructed plasmids in HEK-293 cells, the PiggyBac system, consisting of a PB-CAG plasmid and a transposase expression plasmid, was used in T-ALL cells. The PiggyBac system facilitated the incorporation of the tdTomato-T2A-FLAG-TAL1-ORF1 or ORF4 sequence into the genome of T-ALL cells. Using electroporation and fluorescence-activated cell sorting (FACS), we established a set of four stable T-ALL cell lines for each isoform. Sorting was performed on the basis of the fluorescent protein tdTomato expressed in T-ALL cell lines transfected with the

construct, tdTommato-T2A-FLAG-TAL1-ORF1, and another set of T-ALL cells transfected with TAL1 ORF4 construct. The cell sorting was repeated once more to achieve high purity of the cell population in the process of generating stable cell lines expressing TAL1 isoforms.

The expression of the TAL1 protein was assessed using immunocytochemistry and western blots to validate the stable cell lines. Immunocytochemistry assays were prepared using FLAG and TAL1 antibodies for the stable cell lines (**Figure 6.3**). The cells were fixed and permeabilised before antibody staining; DAPI was used to highlight the nucleus of the cells, as shown in blue. TAL1 protein, as predicted, was demonstrated in both the transfected and untransfected cells. Additionally, higher TAL1 expression can be seen in some of the transfected cells compared to untransfected. The fluorescence of tdTomat was detected in ARR, DU.528, HSB-2, and CCRF-CEM cells transfected with the tdTomato-T2A-FLAG-TAL1-ORF1 or ORF4 constructs, but not in the untransfected control. The FLAG tag was also detected in the transfected T-ALL cells, but not in the untransfected control cells, except in the untransfected DU.528 where unspecific FLAG signal was observed.

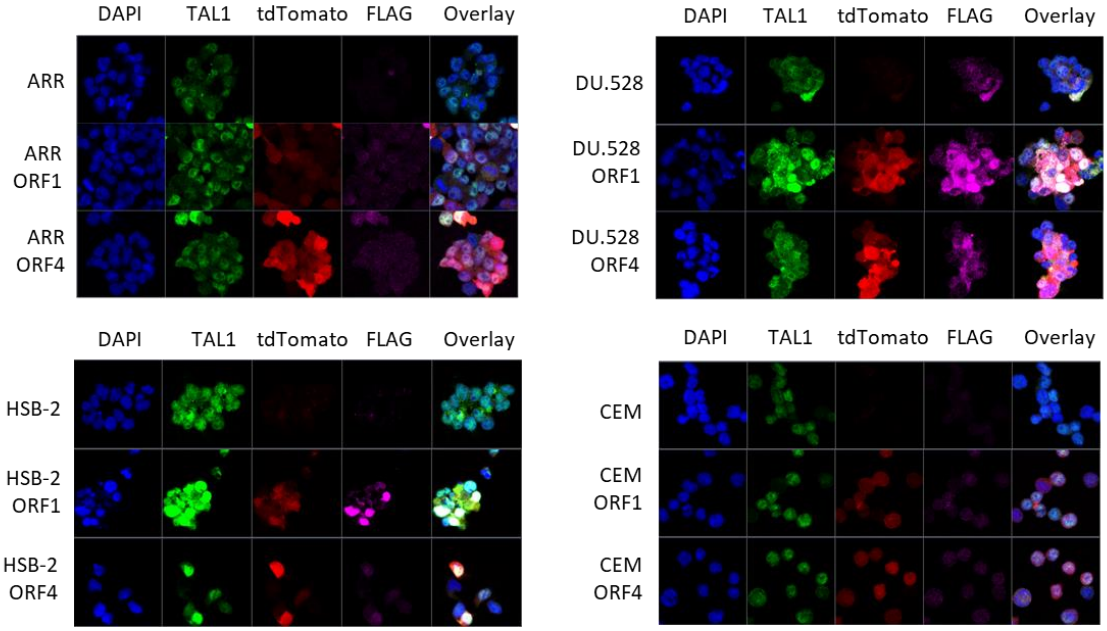


Figure 6.3 Immunocytochemistry of TAL1, TdTomato, and FLAG in the stable TAL1 ORF1/4 T-ALL cells
Protein expression of TAL1, TdTomato, and FLAG was measured using Immunocytochemistry in T-ALL cell lines and T-ALL cell lines stably expressing tdTomato-T2A-FLAG-TAL1 ORF1 or ORF4. DAPI was used as a counterstain for the nucleus.

Furthermore, the protein expression of FLAG-tagged isoforms was examined using western blots of FLAG and TAL1 proteins in TAL1-ORFs T-ALL cell lines. Total protein samples were analysed using western blots with TAL1 antibody (**Figure 6.4 A**). First, TAL1 western blots demonstrated the overexpression of TAL1-ORF1 and ORF4 in their respective TAL1-ORFs cell lines. Other isoforms detected in TAL1 western blots were expected because of isoforms endogenously expressed in the wild-type T-ALL cell as well as shorter isoforms generated from the TAL1 ORF1 construct.

Secondly, the samples were blotted with FLAG antibody, and a clear band at the predicted size of the longest isoform of TAL1 (55-40 KDa), was detected in all of the TAL1-ORF1 T-ALL cell lines, ARR, DU.528, HSB-2, and CCRF-CEM (**Figure 6.4 B**). The FLAG western blot of TAL1-ORF4 T-ALL cell lines showed a less strong band with the size of TAL1-ORF4 protein, seen below the 35 kDa mark in the ladder. The results show that DU.528 had relatively higher expression of FLAG-tagged TAL1-ORF4 compared to the other TAL1 ORF4 T-ALL cell lines. In FLAG westerns an unspecific band was found between 40 and 35 kDa.

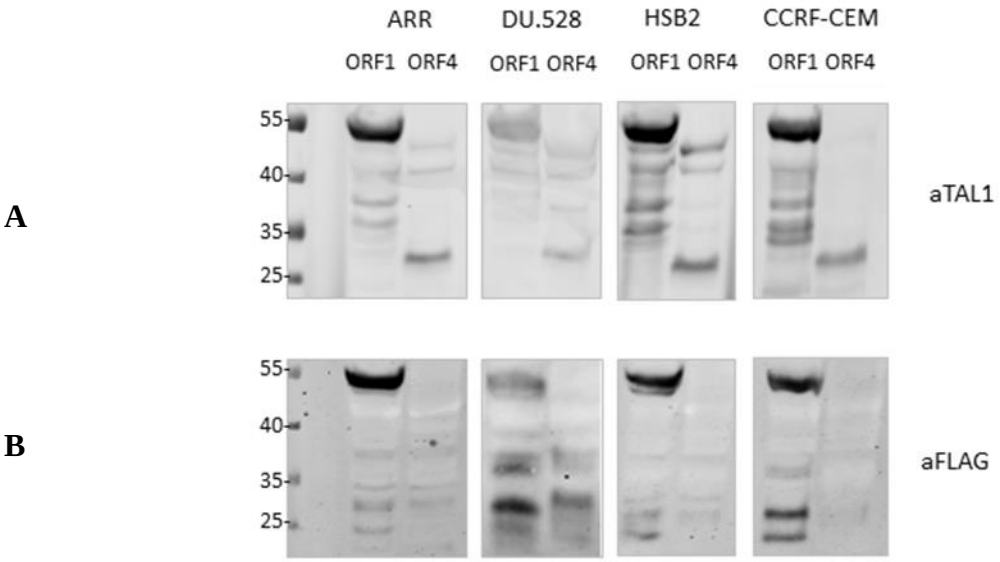
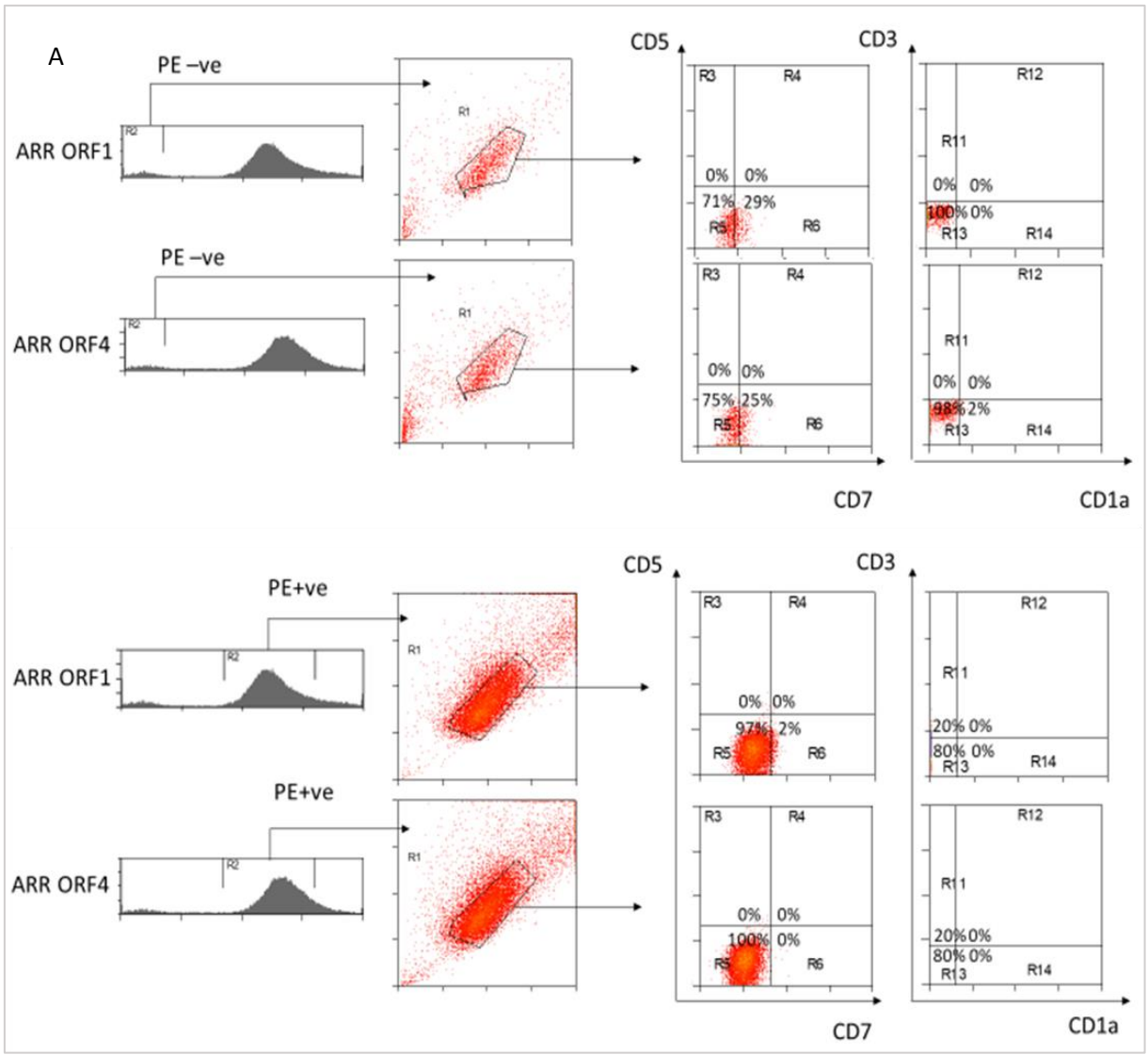


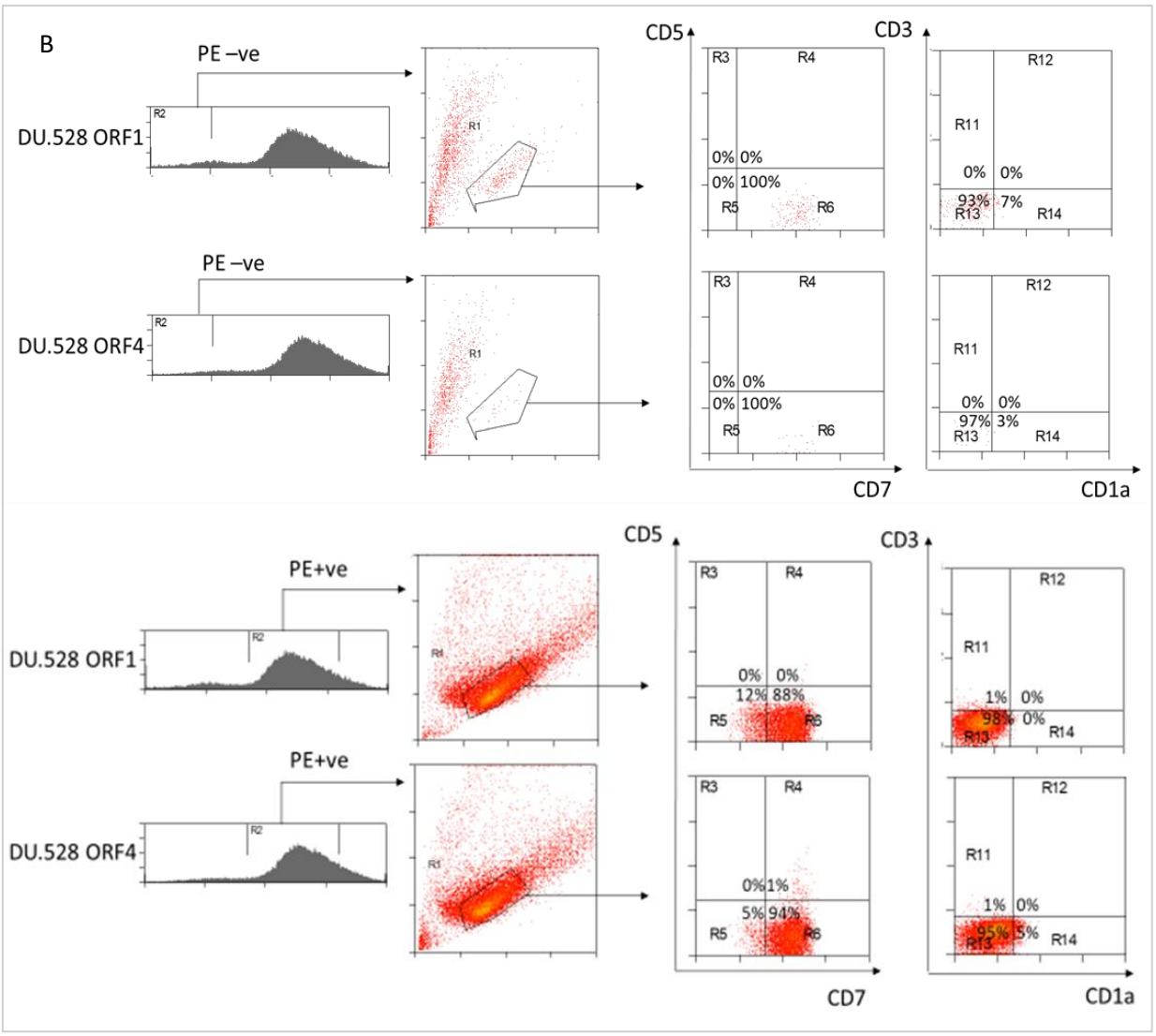
Figure 6.4: Western blots of TAL1 and FLAG in TAL1 ORF1/4 T-ALL stable cell lines

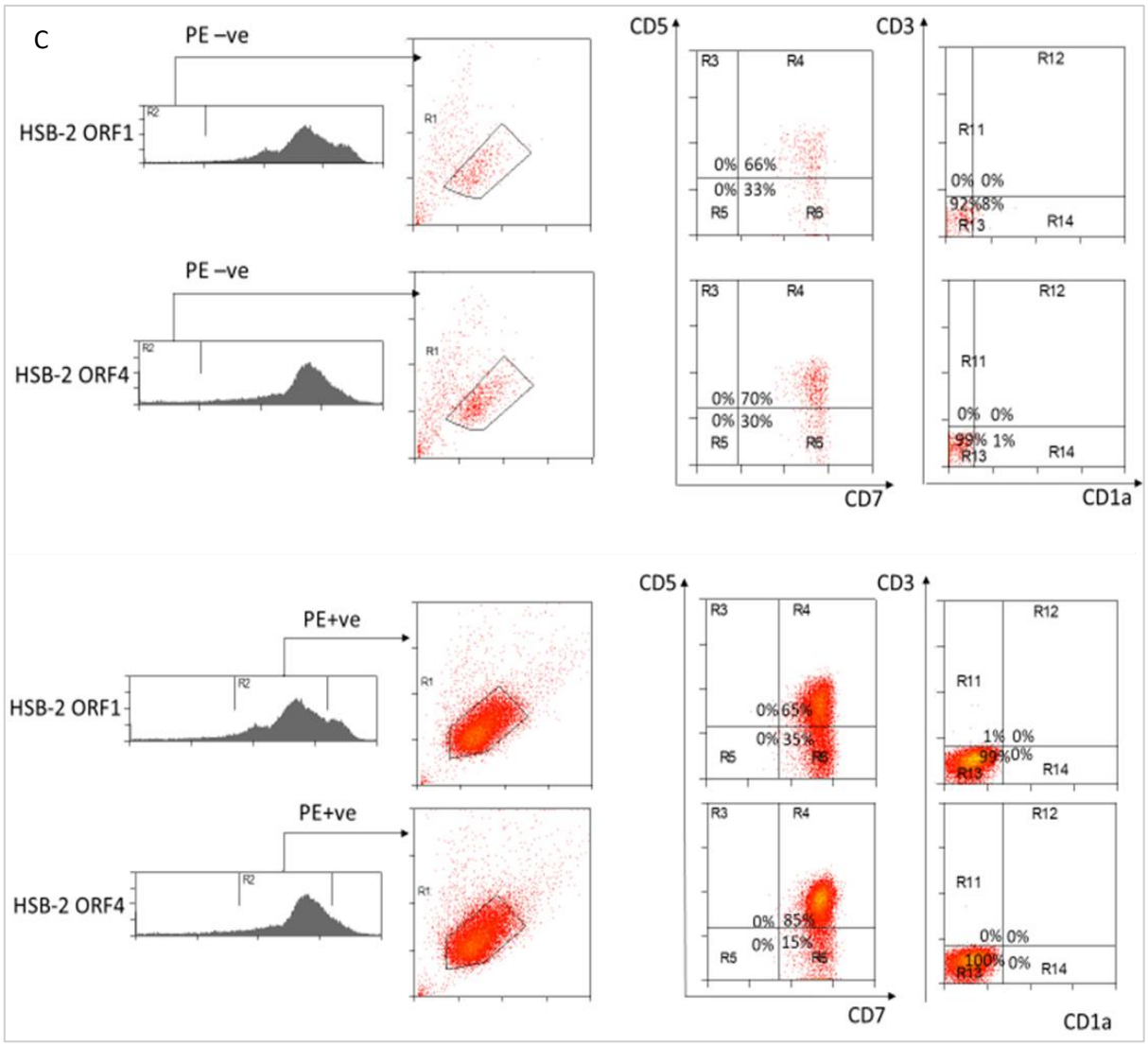
Total protein samples were isolated from TAL1 ORF1 or ORF4 stable T-ALL cell lines. Cells were transfected with PB-CAG-tdTomatoT2A-FLAG-TAL1 ORF1 or ORF4 then sorted twice to make stable cell lines. Protein samples were blotted using FLAG and TAL1 antibodies.

TAL1 short isoform induced the differentiation markers in SIL-TAL TALL cells.

To see whether the introduction of overexpression of ORF1/4 had an impact on their differentiation stage, we evaluated several T cell surface markers, as shown in **figure 6.5 A-D.** Firstly, there was a marginal reduction (2%) in CD7 in ARR overexpressing the short isoform ORF4 compared to ORF1. (**figure 6.5, A**), but not in the ARR cells that were not overexpressing the construct (PE-ve), which were used as a control. CD5, CD1a and CD3 were negative in ARR ORF1 and ORF4. In the SIL-TAL1 positive cell lines, we observed that the short isoform of TAL1 induced a further differentiated immunophenotype compared to cells transfected with the long isoforms. For DU.528 (**figure 6.5, B**), ORF4 transfected cells induced the presence of CD5 positive cells (1%) while ORF1-overexpressing and PE- cells did not. HSB-2 cells transfected with ORF4 had a higher percentage of CD7 CD5 double-positive cells (85% compared to 65%) in ORF1 HSB-2. The HSB-2 PE-ve controls also showed some increase in CD7 CD5 double-positive cells (**figure 6.5, C**). In CCRF-CEM ORF4, cells had 10% more CD1a positive cells, and 4% more CD1a CD3 double-positive cells compared to the ORF1 cells. In the CCRF-CEM PE-ve controls the CD1a did not change and all cells were negative for CD3 (**figure 6.5, D**). Taken together, the immunophenotype results of T-ALL cells transfected with the short isoform, ORF4, showed signs of a more differentiated phenotype compared to ORF1 cells.


A
PE –ve
ARR ORF1
PE –ve
ARR ORF4
CD5
CD3
CD7
CD1a
PE+ve
ARR ORF1
PE+ve
ARR ORF4
CD5
CD3
CD7
CD1a


B
PE –ve
DU.528 ORF1
PE –ve
DU.528 ORF4
PE+ve
DU.528 ORF1
PE+ve
DU.528 ORF4
CD5
CD3
CD7
CD1a
R1
R2
R3
R4
R5
R6
R11
R12
R13
R14
0% 0%
0% 100%
93% 7%
97% 3%
12% 88%
1% 0%
98% 0%
0% 1%
5% 94%
95% 5%


C
PE –ve
HSB-2 ORF1
R2
R1
CD5
CD3
R3
R4
0% 66%
0% 33%
R5
R6
R12
R11
0% 0%
92% 8%
R13
R14
PE –ve
HSB-2 ORF4
0% 70%
0% 30%
99% 1%
CD7
CD1a
PE+ve
0% 65%
0% 35%
1% 0%
90% 0%
0% 85%
0% 15%
100% 0%

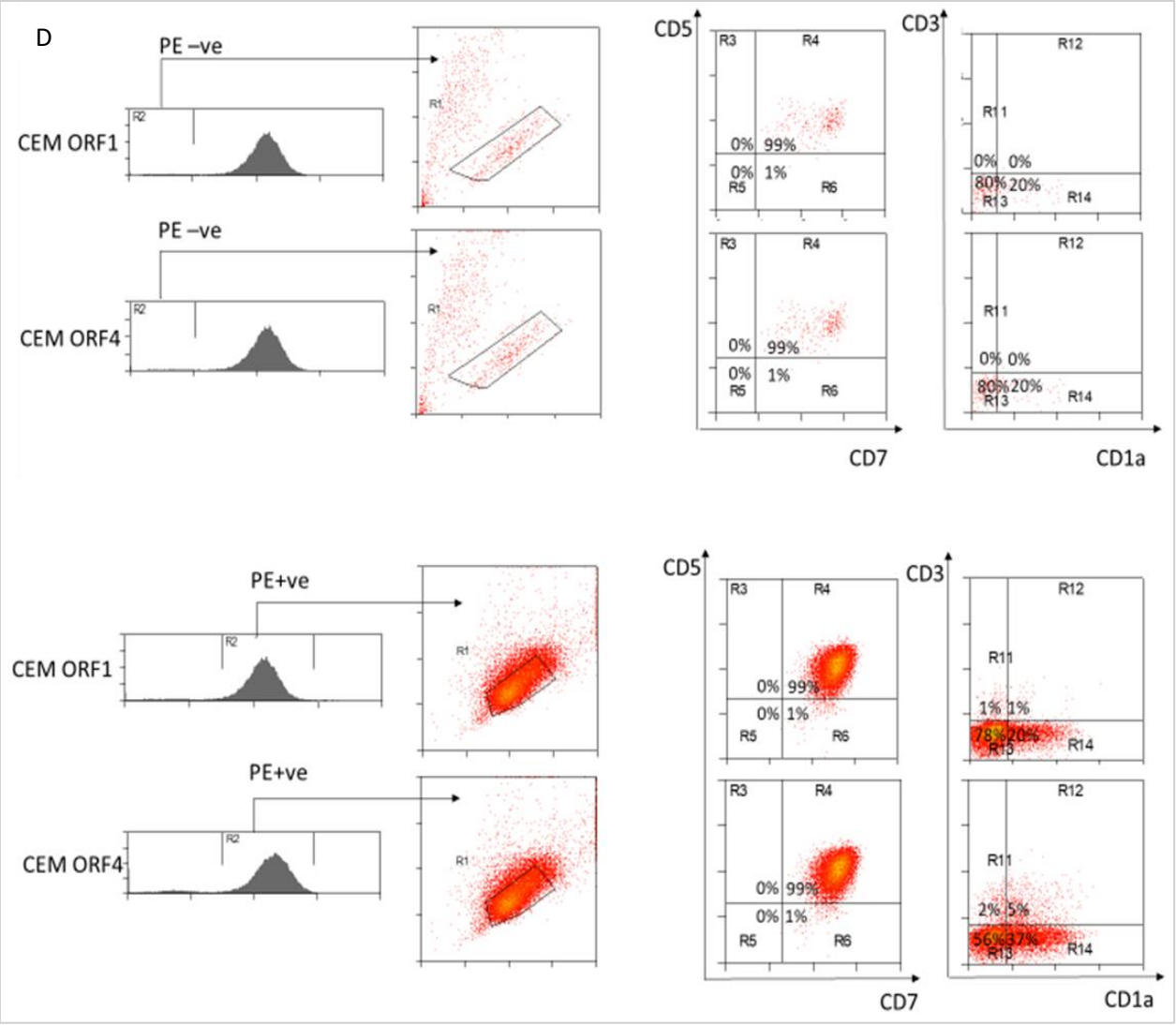


Figure 6.5 A-D TAL1-ORF4 induces differentiation in T-ALL cell lines

Live cells from the T-ALL cell lines (ARR (A), DU.528 (B), HSB-2 (C), CCRF-CEM (D)) stably expressing tdTomato-T2A-FLAG-TAL1 ORF1 or ORF4 were gated through PE plot to measure tdTomato fluorescence. PE –ve or PE+ population were gated through forward scatter / side scatter plots. Cells were stained using antibodies against CD7, CD5, CD1a, and CD3. Cells incubated with isotype-matched non-specific IgG antibodies served as a gating to divide the plot into quadrants.

TAL1 occupancy in TAL1 long and short isoform stable T-ALL cell lines:

To further investigate TAL1 short isoform ability to bind DNA and regulate transcription as part of the transcriptional complex, we evaluated the genomic occupancy of TAL1 isoforms. Chromatin was isolated from the TAL1-ORF1, and TAL1-ORF4 T-ALL cell lines and both TAL1 and FLAG antibodies were used to obtain ChIPseq data, but FLAG Ab gave no signal. TAL1 occupancy determined in ORF1, and ORF4 T-ALL cells had a significant overlap with the ChIPseq data collected from untransfected T-ALL cells. On the other hand, distinct TAL1 peaks were found when comparing TAL1-ORF1 to ORF4. **Figure 6.6** shows the overlap between TAL1 ChIPseq datasets in untransfected T-ALL cell lines (TAL1), TAL1-ORF1 (ORF1), and TAL1-ORF4 (ORF4) T-ALL cell lines.

In ARR, we identified 7748 TAL1 peaks in ORF1, and 17941 peaks in ORF4 cells. Only 1961 peaks were shared between ARR ORF1 and ORF4 ChIPseq. They both partially overlapped with the untransfected ARR. The three-factor overlap between TAL1 ChIPseq datasets in ARR was 1171. In SIL-TAL cells, the three TAL1 ChIPseq datasets overlapped more in SIL-TAL compared to ARR. In DU.528, we found 16721 TAL1 peaks for ORF1 cells and 17143 for ORF4 cells, and 7784 peaks shared between them. Almost half of the ORF1 and ORF4 datasets overlapped with the untransfected DU.528 dataset. The three-factor overlap in DU.528 was 5445 peaks. Interestingly, we observed a higher percentage of overlap between TAL1 ChIPseq datasets in HSB-2. An overlap of 53% of ORF1 dataset (39553) with ORF4 (31837) was shown. The majority of ORF1 and ORF4 data overlapped substantially with TAL1 ChIPseq in untransfected HSB-2, and the overlap between the three datasets was 13417. Similarly, we found a significant overlap of 24974 between ORF1 (56717) and ORF4 (33193) in CCRF-CEM. Most of the TAL1 peaks in untransfected cells were shared with ORF1 and ORF4, making the three-factor overlap in CCRF-CEM 15639 peaks.

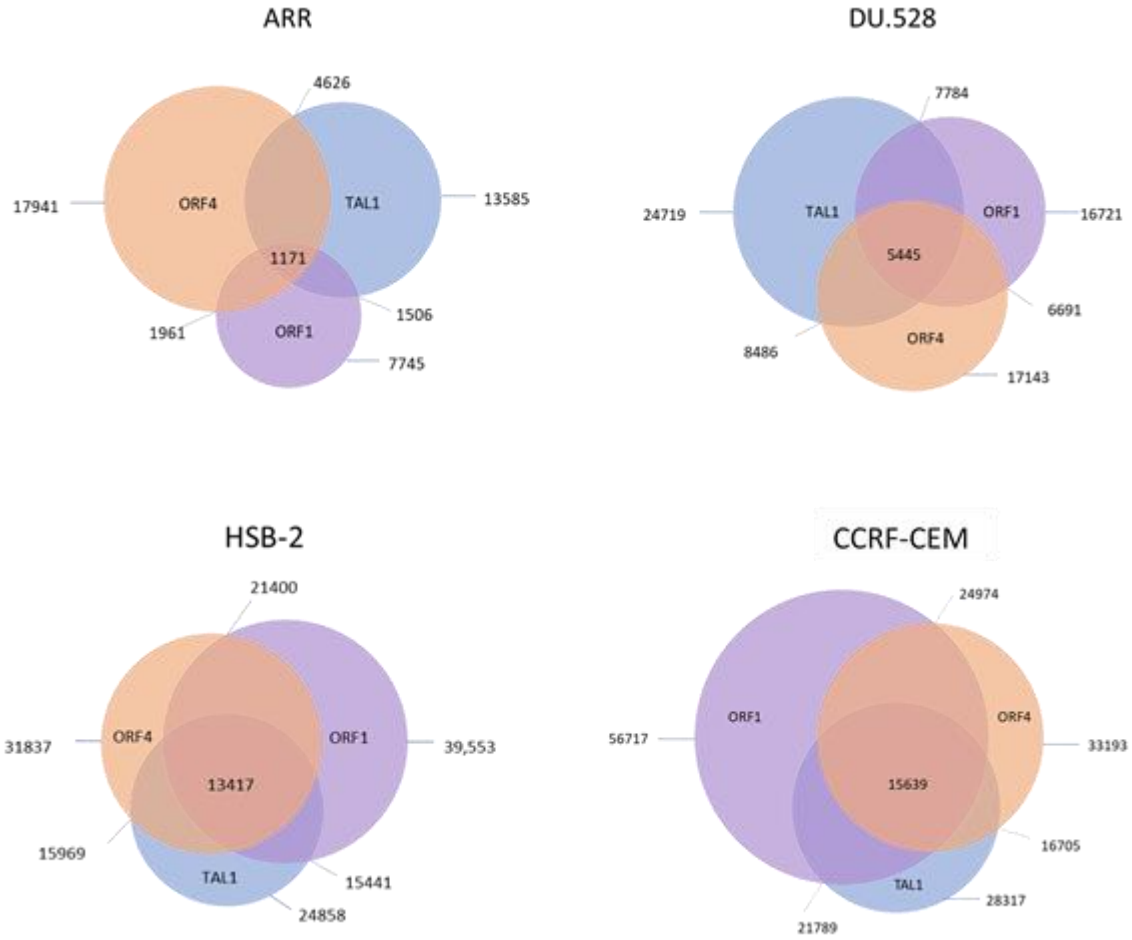
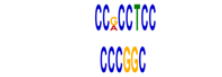
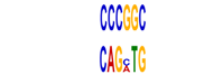
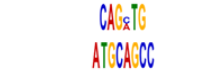


Figure 6.6 The overlap between ChIPseq peaks of TAL1 datasets in T-ALL
TAL1 ChIPseq peaks in untransfected T-ALL cell lines (blue), and TAL1 ORF1 (purple), ORF4 (orange) stable T-ALL lines. Grey lines indicate the number of overlapping peaks and blue lines indicate the total number of peaks in the ChIPseq dataset.

For Further analyses of TAL1 ChIPseq datasets we divided TAL1 ChIPseq data analysis into three groups; (i) peaks found in ORF1 dataset and not overlapping with ORF4 (ORF1-only), (ii) the converse (ORF4-only), and (iii) the overlapping peaks (Common). We searched for motifs enriched in each group using *de novo* motif analysis on sequences collected from the top thousand peaks of each group. The Common group in the four T-ALL cell lines (**Figure 6.7 A-D**) showed enrichment of GATA and E-box motifs, while ETS and RUNX motifs were additionally detected in the SIL-TAL cells, DU.528, HSB-2, and CCRF-CEM. Motifs for Sp1, and KLF4 which belong to KLF/SP family (Presnell, Schnitzler, & Browne, 2015), were found significant in ARR and CCRF-CEM ORF1 and ORF4 overlapping regions. The *De novo* motif analysis of the Common group in ARR returned eight additional motifs, including ZFP161 (Zinc finger), TFAP2A, MEF2C, EGR1 (Zinc finger), STAT6, MLXIPL, NR2F2, and YY2. In DU.528, we also found TBX21 and NKX2-2 while the RARA motif was enriched in HSB-2.

Using the *De novo* motif analysis of ORF1-only and ORF4-only groups, we identified motifs significantly enriched in sequences extracted from ORF1-only regions (**Figure 6.8**) and motifs in ORF4-only regions (**Figure 6.9**). We compared the motifs in the Common, ORF1-only, and ORF4-only regions to identify the motifs that are unique to ORF1-only. Many motifs of ORF1-only and ORF4-only were shared between each other or with the analysis of the Common regions. The ORF1-only motifs that were shared include TFAP2, RUNX, GATA, E-box, the AAGG motif (known for ETS; ELF; STAT3; STAT6; SPI), SP/KLF, RARA/RARG/RXRA, the poly-A motif for ZNF384/SRF/FOXC1, and CTCF. On the other hand, we found four motifs that were unique for ORF1-only, TEAD2 in ARR, MEIS/FOSL1 in DU.528, CEBP in HSB-2, and FOX transcription factors in CCRF-CEM. Moreover, we identified motifs that were exclusive for ORF4-only regions. First, our analysis found a similar motif for FOSL1/JUND, JDP2, BATF::JUN in three T-ALL cell lines ARR, DU.528, and HSB-2. ZIC2 in ARR and NFAT and MZF1 in HSB-2 were also motifs that were found significant exclusively in the ORF4-only regions.

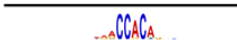
ARR

Motif found	E-value	Known or similar motifs	Frequency
	8.6e-650	Zfp161	118/1000
	1.3e-548	TFAP2A, TFAP2C	74/1000
	2.1e-204	MEF2C	48/1000
	2.6e-043	GATA	328 / 1000
	3.6e-021	Egr1	94 / 1000
	1.1e-019	Stat6	247 / 1000
	3.1e-011	MLXIPL, SREBF2	49 / 1000
	2.4e-004	SP2, SP1, Egr1	47 / 1000
	3.2e-004	Nr2f2	74 / 1000
	3.7e-003	TAL1::TCF3 (E-box)	191 / 1000
	1.1e-002	YY2	25 / 1000

DU.528

Motif found	E-value	Known or similar motifs	Frequency
	3.2e-271	GATA	492/997
	3.5e-230	RUNX1	661/997
	1.2e-162	ETS1, ERG	414/997
	4.8e-072	TBX21	17/997
	1.1e-070	Nkx2-2	19/997
	9.2e-008	ZNF238, ZBTB18 (E-box)	198 / 997

HSB-2

Motif found	E-value	Known or similar motifs	Frequency
	1.4e-350	RUNX1	781/994
	1.9e-178	GATA	361/994
	9.7e-093	ETS1, ELF5	342/994
	2.1e-013	Tcfe2a, TAL1::TCF3, TWIST1 (E-box)	172 / 994
	1.1e-002	RARA, RORA	68 / 994

CCRF-CEM

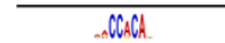
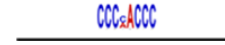
Motif found	E-value	Known or similar motifs	Frequency
	2.4e-285	RUNX	647/994
	7.8e-222	GATA	433/994
	4.7e-146	MESP1, TCF3	423/994
	3.5e-078	ELK4, ETS1	360/994
	1.2e-002	KLF5, Ascl2, Klf1	28 / 994

Figure 6.7 Motifs significantly enriched in TAL1-ORF1 and TAL1-ORF4 overlapping regions (Common) in T-ALL cells

Chromatin was isolated from the T-ALL cell lines overexpressing either TAL1 ORF1 or ORF4. TAL1 ChIPseq data were analysed to identify the top 1000 TAL1 ORF1 ORF4 common peaks for De novo motif analysis. The tables list significantly enriched motifs, Expect values (E-values), the names of known or similar motifs, and the frequency the motifs were found.

ARR

Motif found	E-value	Known or similar motifs
	1.1e-286	TFAP2A
	9.5e-225	TEAD2

DU.528

Motif found	E-value	Known or similar motifs
	1.6e-184	RUNX1, RUNX3
	1.2e-147	CTCF, CTCFL
	3.4e-042	ETS1
	1.1e-029	Meis2, FOSL1, MEIS3
	6.3e-010	ZNF238, ZBTB18 (E-box)
	1.9e-002	GATA

HSB-2

Motif found	E-value	Known or similar motifs
	1.0e-317	ETS1
	3.4e-100	RUNX1
	5.7e-043	ZNF384, Mtf1, Zfp105
	2.3e-017	GATA
	2.8e-006	ZNF238, ZBTB18 (E-box)
	8.9e-003	RARG, RARA
	1.5e-002	CEBPA, CEBPB, CEBPE

CCRF-CEM

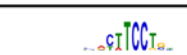
Motif found	E-value	Known or similar motifs
	9.0e-253	ELK4, ETS1
	6.8e-126	RARG
	1.0e-123	FOXI1, FOXO6
	6.3e-120	RUNX1, RUNX3
	1.1e-014	GATA3, GATA4
	5.6e-007	KLF5, Zfp740, Ascl2
	1.1e-006	TAL1::TCF3, ZNF238

Figure 6.8 De novo motif analysis of TAL1 ORF1-only ChIPseq peaks in T-ALL cell lines

Chromatin was isolated from the T-ALL cell lines overexpressing either TAL1 ORF1 or ORF4. TAL1 ChIPseq data were analysed to identify the top 1000 TAL1 ORF1 peaks after excluding the peaks overlapping with ORF4 peaks for *De novo* motif analysis. The tables list significantly enriched motifs, Expect values (E-values), the names of known and similar motifs.

ARR

Motif found	E-value	Known or similar motifs
	1.1e-286	Zfx, TFAP2A
	2.5e-118	RUNX
	7.8e-097	Rxra
	1.3e-052	GATA
	1.5e-022	Elf3, Stat4
	3.5e-005	ZBTB7A, ELK4, Gabpa
	1.9e-003	GATA
	5.4e-003	Zic2
	1.0e-002	KLF5, SP1, SP2
	1.9e-002	FOSL1, JUND, Jundm2

DU.528

Motif found	E-value	Known or similar motifs
	3.2e-205	ELF5 (ETS-like)
	5.1e-114	RUNX1
	4.7e-033	ZNF384, Srf, FOXC1
	1.9e-020	GATA
	9.4e-005	FOSL1, JUNB, FOSL2
	3.2e-003	NFAT5, NFATC1
	1.5e-002	MZF1

HSB-2

Motif found	E-value	Known or similar motifs
	1.4e-156	SPIB, SPIC
	4.4e-145	GATA
	1.1e-123	RUNX3
	1.2e-007	JDP2, BATF::JUN

CCRF-CEM

Motif found	E-value	Known or similar motifs
	9.0e-150	CTCF, CTCFL, CTCF_full
	7.5e-087	ELK4, ZBTB7A, Gabpa
	1.6e-057	RUNX1, RUNX3
	9.9e-021	Klf1

Figure 6.9 Motifs significantly enriched in TAL1-ORF4-only regions in T-ALL cells

Chromatin was isolated from the T-ALL cell lines overexpressing TAL1 ORF1 or ORF4. TAL1 ChIPseq data were analysed to identify the top 1000 TAL1 ORF4 peaks after excluding the peaks overlapping with ORF1 peaks. The tables list significantly enriched motifs, Expect values (E-values), and the names of known or similar motifs,.

TAL1 complex binding at isoform-specific regions compared to overlapping regions

We demonstrated the distribution of the three TAL1 ChIPseq datasets (untransfected, ORF1, and ORF4), and how it related to signal intensity. Additionally, we investigated the overlap between other members of the TAL1 transcriptional complex with TAL1 occupied regions in untransfected and isoform overexpressing T-ALL cells. RUNX1 and GATA2 ChIPseq were also included because they were enriched in the *de novo* motif analysis of the four T-ALL cell lines. LYL1, a homologue of TAL1, was of interest because we previously observed in DU.528, where the short isoforms are dominant, that LYL1 was expressed at much higher levels compared to the other T-ALL cell lines (**Figure 5.8**).

Heat maps of TAL1 ChIPseq datasets and other members of the TAL1 transcriptional complex are shown in **Figure 6.10**. Firstly, datasets included in the signal intensity heatmaps were TAL1 ChIPseq in untransfected cells, ORF1, and ORF4-overexpressing cells. Then, heat maps LYL1, HEB, LMO2, and LDB1, and RUNX1 ChIPseq, in addition to MeDIP, and DNAseI. To create the heat maps matrix, we identified ORF1-only and ORF-4 only by excluding any overlapping regions. Then, as shown on the Y-axis, the ORF1-only were ordered in a descending order based on signal intensity while ORF4 peaks were sorted in ascending order and finally, the overlapping peaks in between them based on the ratio of ORF1 to ORF4 signal (Common).

The TAL1 ChIPseq in untransfected ARR cells showed that the high-intensity signal was in the ORF1-ORF4 common region and also in the ORF4-only area, whereas only low intensity was found in the ORF1 top section. TAL1 ChIPseq in ORF1-ARR signal was enriched in the ORF1-only order whereas the TAL1 ChIPseq of ORF4-cells was prominent in the ORF4-only part of the heat map, as expected. We found that the strongly enriched peaks of the three TAL1 ChIPseq datasets were in the overlap region.

The intensity of LYL1 ChIPseq in ARR was low in all part of the TAL1-ORFs matrix. Nevertheless, LYL1 overlapped with the ORF1-ORF4 overlapping region and the highest peaks of ORF4-only dataset seen in the bottom of the heat map. The distribution was similar in HEB, LMO2, LDB1, and RUNX1 ChIPseq within the matrix of TAL1 ORFs. They all overlapped with the TAL1 ORF1-only, ORF4-only, and the ORF1-ORF4 overlapping region. The heatmap of MeDIP displayed a relatively lower signal, which however colocalised with the ORF1-ORF4 overlapping peaks and the ORF4-only peeks without clear signal in the ORF1-only part.

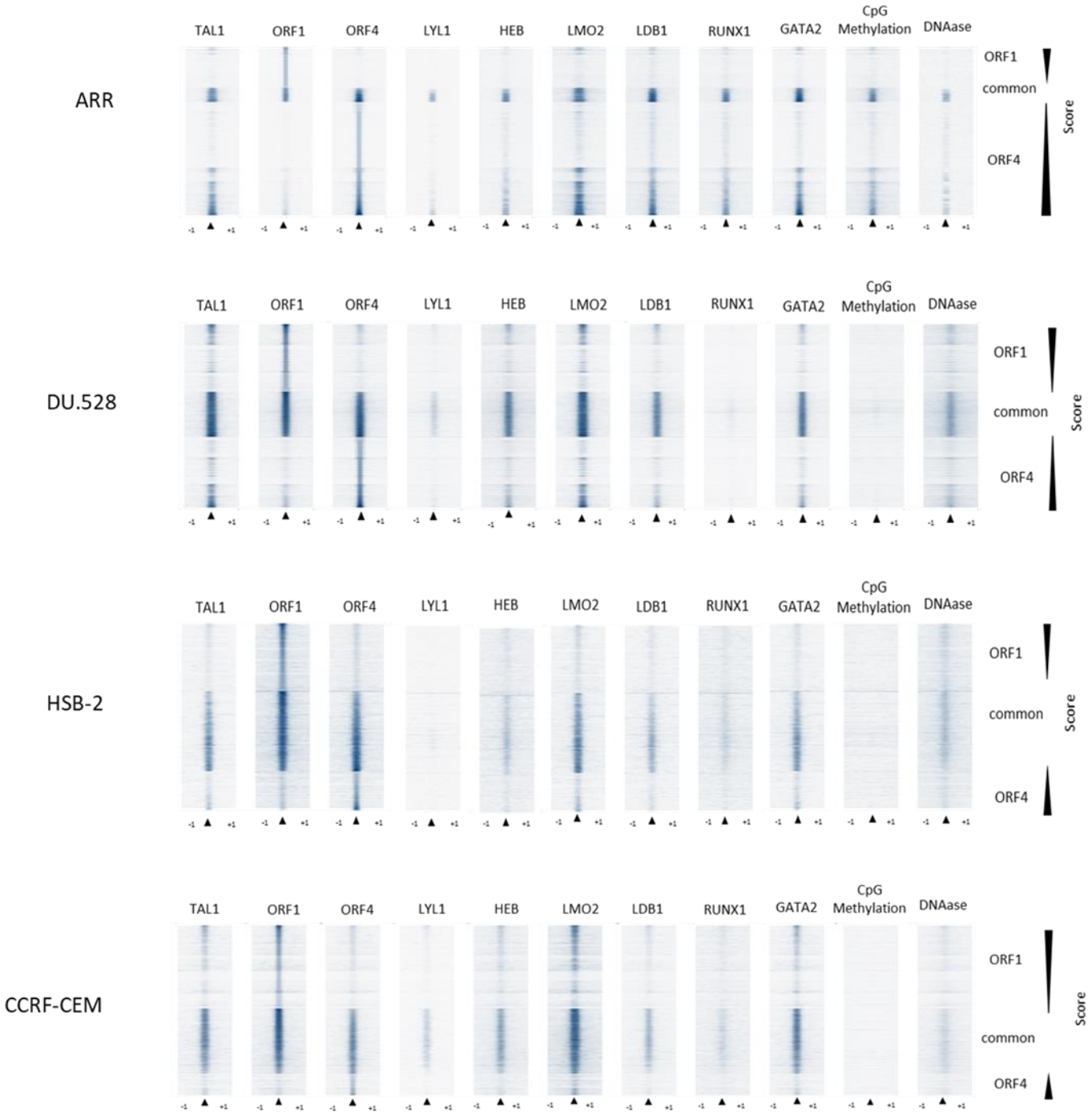


Figure 6.10 Heat maps of TAL1 ChIPseq in T-ALL; TAL1-ORF1 cells, TAL1-ORF4 cells and several TAL1 partners

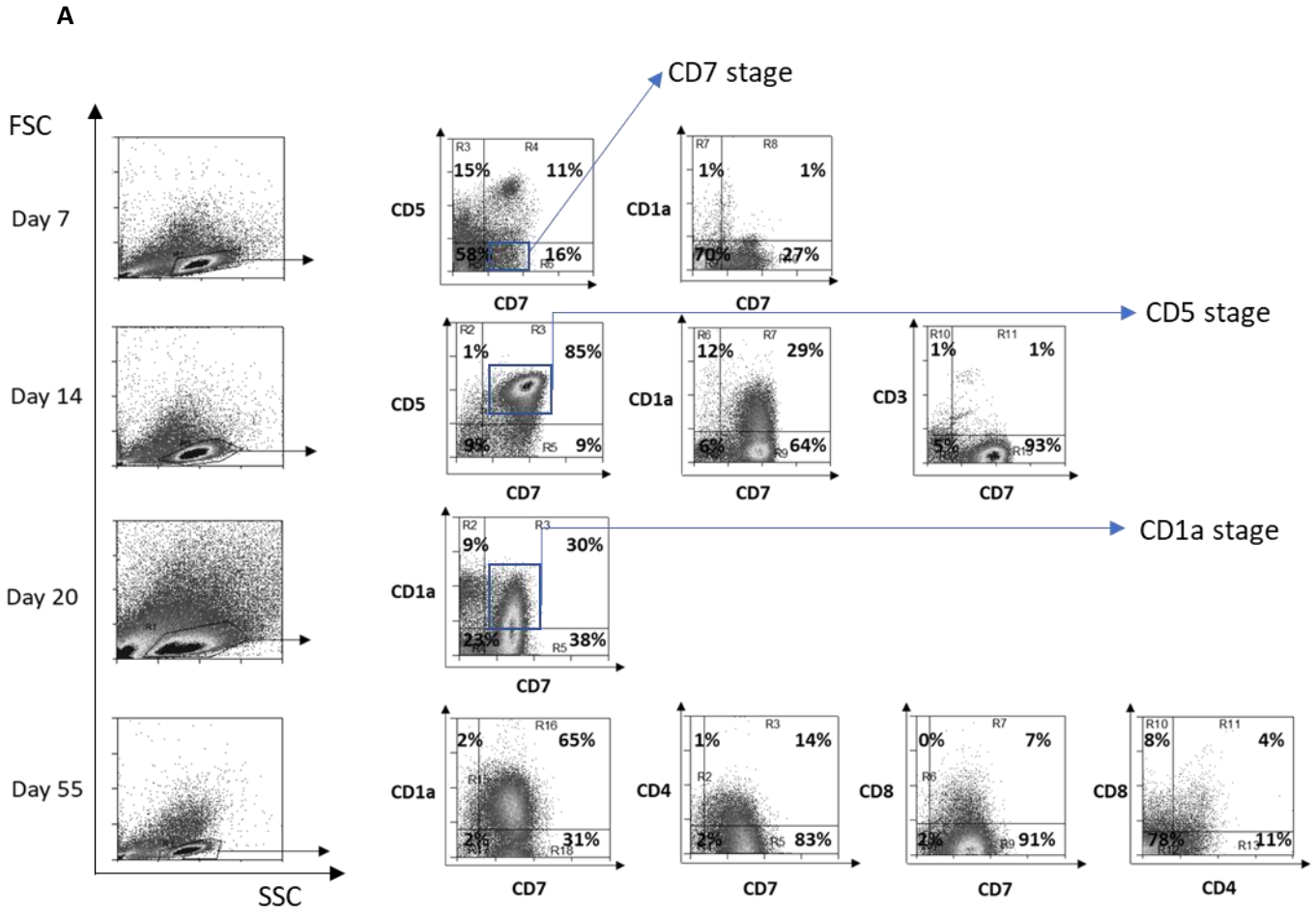
Heat maps of TAL1 ChIPseq datasets generated from untransfected and transfected cells; LYL1; HEB; LMO2, LDB1, RUNX1, GATA2 ChIPseq, MeDIP and RNAseq in the indicated T-ALL cell lines. The Y-axis represents individual TAL1 ChIPseq regions in TAL1-ORF1 cells sorted in descending order according to score. The X-axis measures a 2 kb area centred around the summit of TAL1 ChIPseq peaks.

In vitro T cell differentiation system as a model to study TAL1 and its isoform in T-ALL

We aimed to study the effects of the overexpression of TAL1 and its isoforms on T cell differentiation and how TAL1 facilitates T-ALL oncogenic transformation of differentiating thymocytes. We wanted to monitor the effect of the TAL1 aberrant expression and potential differences between its isoforms in a more dynamic model that mirrors the *in vivo* transformation T cell progenitors.

Therefore, we worked on establishing the *in vitro* human T cell differentiation system in our lab. The system uses human CD34 progenitors isolated from cord blood and initiates T cell differentiation *in vitro* while supplementing the culture with hIL7 and hFLT3 ligand. The progenitors must be cultured on OP9 cells expressing NOTCH ligand DL1 or DL4. Initially, we attempted to set up the differentiation system using OP9-DL1 cells but with little success. After switching to OP9-DL4 cells, we were able to induce T cell differentiating from $CD34^+$ *in vitro* with reproducibility. We used T cell differentiation markers to monitor the progress of the *in-vitro* T cell differentiation system, as displayed in **Figure 6.11**. The analysis identified the day at which most of the culture reaches a specific stage of T cell differentiation. The timing of T cell differentiation stages in our culture was useful for collecting samples representing four stages of T cell development via cell sorting.

RNA-seq data was generated from differentiating T-cell progenitors at four different stages, day 0 $CD34^+$ progenitors, $CD7^+$ $CD5^-$ $CD1a^-$ cells, $CD7^+$ $CD5^+$ $CD1a^-$ cells, and the $CD7^+$ $CD5^+$ $CD1a^+$ $CD3^-$. We collected twelve independent samples representing triplicates from each of the four different stages. Additionally, the Hoogenkamp lab produced RNAseq data from two T-ALL patients. The RNAseq tracks, shown in **Figure 6.12 A-E** were visualised using the UCSC genome browser. The figure shows five screenshots of RNAseq tracks at the *CD34*, *CD7*, *TAL1*, *LMO2*, and *HEB* genes. TAL1 and LMO2 were expressed in $CD34^+$ $CD7^-$ samples and $CD34^+$ $CD7^+CD5^-$ samples but not expressed in $CD34^+$ $CD7^+$ $CD5^+$ cells.


A
FSC
SSC
Day 7
Day 14
Day 20
Day 55
CD7 stage
CD5 stage
CD1a stage
CD5
CD1a
CD3
CD4
CD8
CD7
15%
11%
58%
16%
1%
1%
70%
27%
1%
85%
9%
9%
12%
29%
6%
64%
1%
1%
5%
93%
9%
30%
23%
38%
2%
65%
2%
31%
1%
14%
2%
83%
0%
7%
2%
91%
8%
4%
78%
11%

B

Day 7

Count	% Hist	% All
39446	100.00	39.45
6008	15.23	6.01
4262	10.80	4.26
22793	57.78	22.79
6383	16.18	6.38

Count	% Hist	% All
39446	100.00	39.45
408	1.03	0.41
507	1.29	0.51
27959	70.88	27.96
10572	26.80	10.57

Count	% Hist	% All
39446	100.00	39.45
890	2.26	0.89
286	0.73	0.29
29330	74.35	29.33
8940	22.66	8.94

Day 14

Count	% Hist	% All
61427	100.00	61.43
807	1.31	0.81
52346	85.22	52.35
3075	5.01	3.08
5199	8.46	5.20

Count	% Hist	% All
61427	100.00	61.43
370	0.60	0.37
17765	28.92	17.77
3756	6.11	3.76
39536	64.36	39.54

Count	% Hist	% All
57117	100.00	57.12
652	1.14	0.65
526	0.92	0.53
2978	5.21	2.98
52961	92.72	52.96

Day 21

Count	% Hist	% All
129097	100.00	10.06
12028	9.32	0.94
38866	30.11	3.03
29668	22.98	2.31
48535	37.60	3.78

Day 55

Count	% Hist	% All
67658	100.00	67.66
1245	1.84	1.24
44167	65.28	44.17
1482	2.19	1.48
20764	30.69	20.76

Count	% Hist	% All
67658	100.00	67.66
432	0.64	0.43
9579	14.16	9.58
1079	1.59	1.08
56568	83.61	56.57

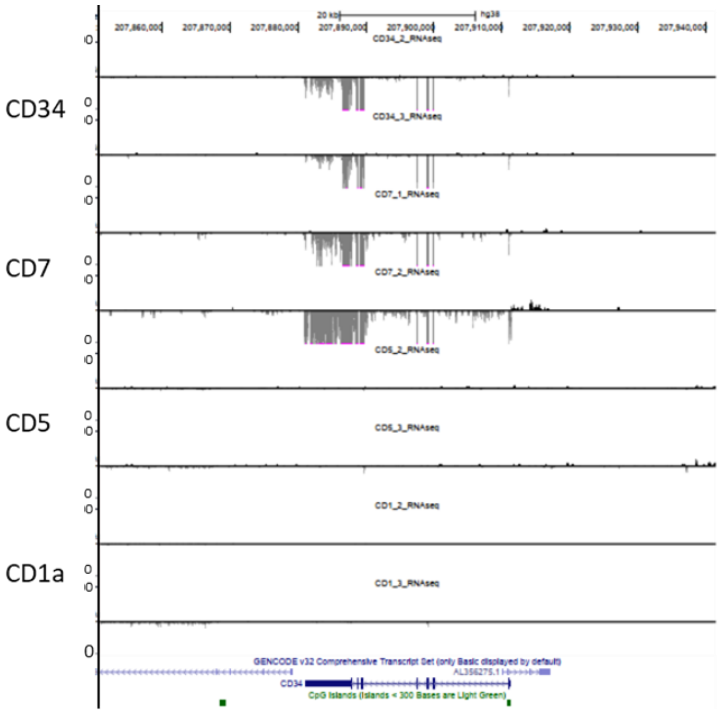
Count	% Hist	% All
67658	100.00	67.66
95	0.14	0.10
4584	6.78	4.58
1264	1.87	1.26
61715	91.22	61.71

Count	% Hist	% All
67658	100.00	67.66
5529	8.17	5.53
2473	3.66	2.47
52469	77.55	52.47
7187	10.62	7.19

Figure 6.11 Expression of T cell CD markers to monitor the progress of *In-vitro* human T cells differentiation.

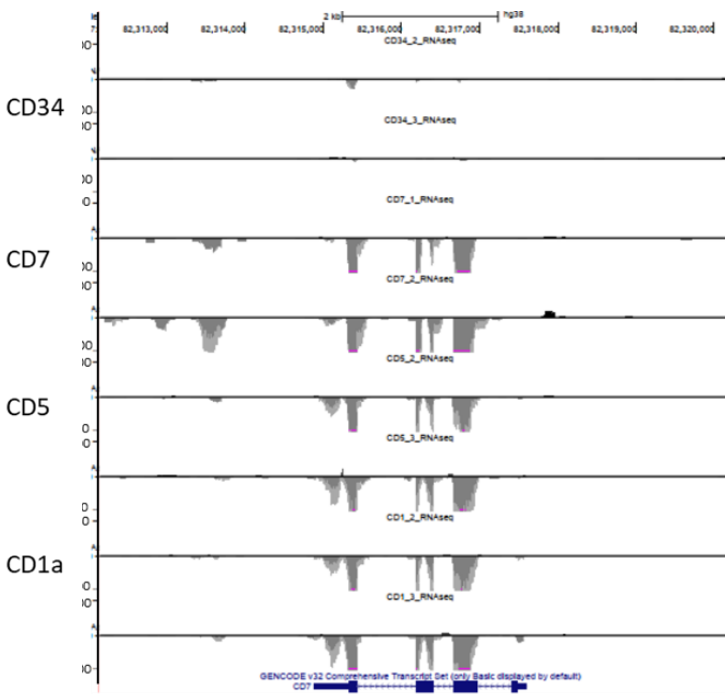
CD markers used to monitor the progress of *In-vitro* human T cells differentiation. Cord blood CD34 cells were cultured with OP9-DL4 cells in the presence of hIL7 and hFLT3 ligand. Live T cells were gated through forward scatter / side scatter plot. the expression of T cell markers CD7, CD5, CD1a, CD4, and CD8 was used to evaluate the progress of T cell differentiation on day 7, 14, 20, and 55 of culture. Gating in blue indicate T cell progenitor sorted for RNA isolation at four stages, stage CD34 (day 0), CD7 (day 7), CD5 (day 14), CD1a (day 20). FACS statistics are shown in the same order of the CD markers plots and quadrants.

A


CD34
CD7
CD5
CD1a
CD34_2_RNAseq
CD34_3_RNAseq
CD7_1_RNAseq
CD7_2_RNAseq
CD5_2_RNAseq
CD5_3_RNAseq
CD1_2_RNAseq
CD1_3_RNAseq
GENCODE v32 Comprehensive Transcript Set (only Basic displayed by default)
CD34
CpG Islands (Islands < 300 Bases are Light Green)

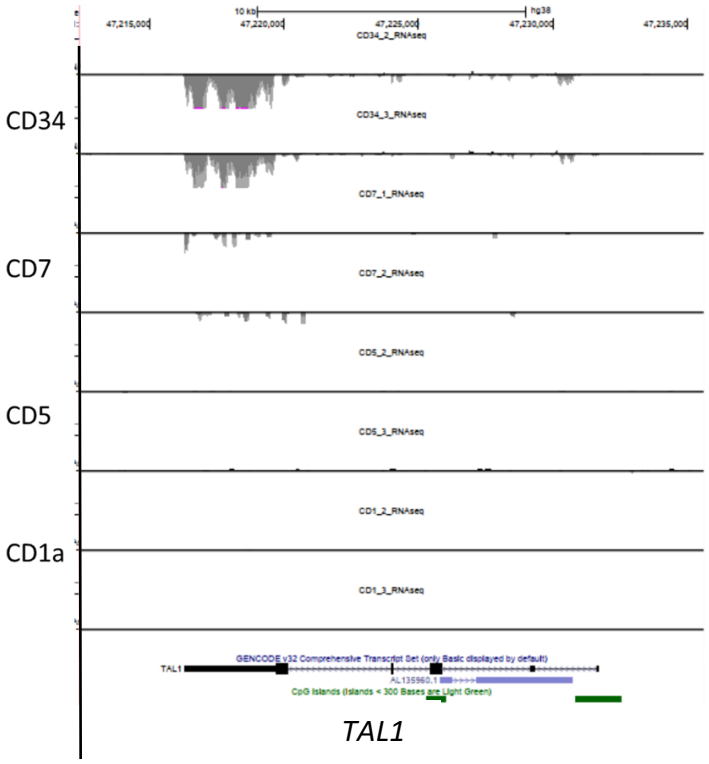
CD34

B


CD34
CD7
CD5
CD1a
CD34_2_RNAseq
CD34_3_RNAseq
CD7_1_RNAseq
CD7_2_RNAseq
CD5_2_RNAseq
CD5_3_RNAseq
CD1_2_RNAseq
CD1_3_RNAseq
GENCODE v32 Comprehensive Transcript Set (only Basic displayed by default)
CD7

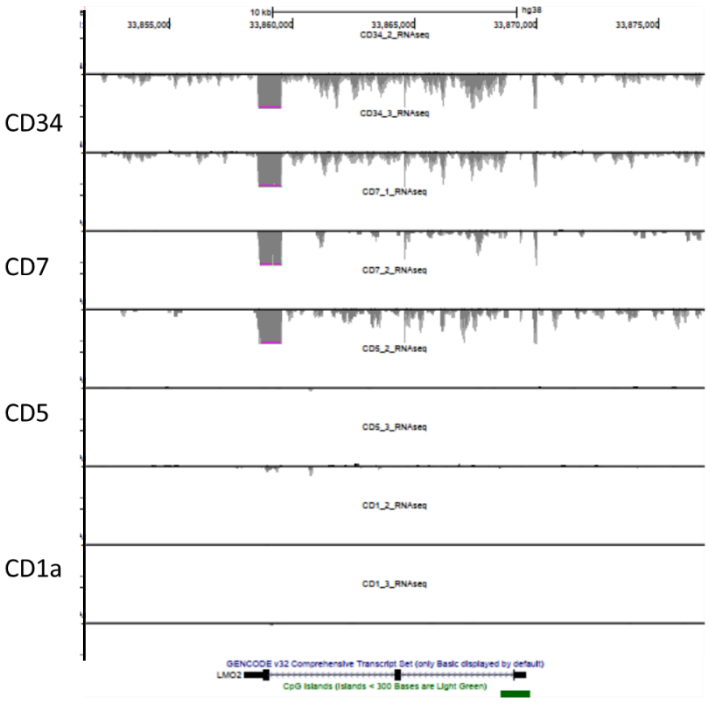
CD7

C


CD34
CD7
CD5
CD1a
TAL1
GENCODE v32 Comprehensive Transcript Set (only Basic displayed by default)
CpG Islands (Islands < 300 Bases are Light Green)

TAL1

D


CD34
CD7
CD5
CD1a
GENCODE v32 Comprehensive Transcript Set (only Basic displayed by default)
LMO2
CpG Islands (Islands < 300 Bases are Light Green)

LMO2

E

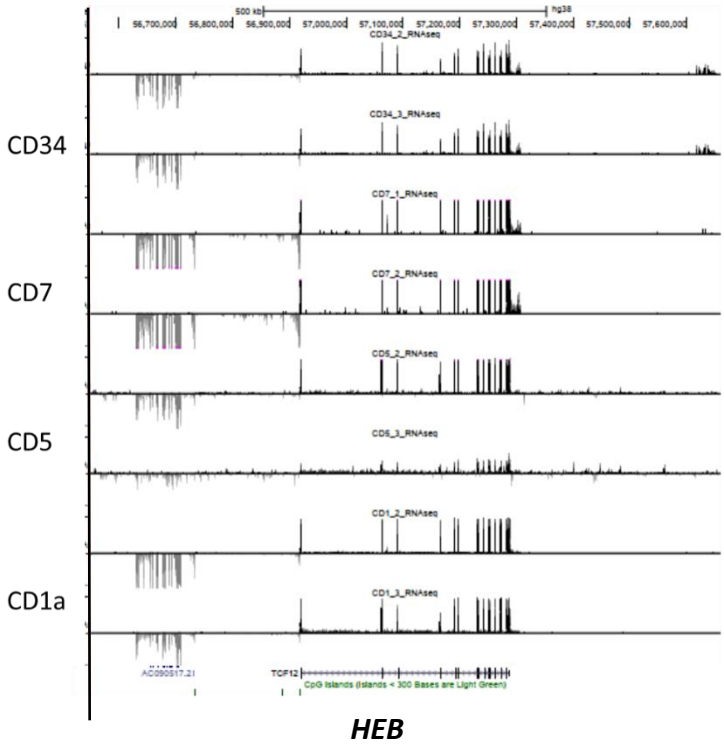


Figure 6.12 A-E: Visualisation of RNAseq tracks in human T cell progenitors.

Five screenshots of RNAseq datasets representing human T cells at four stages of differentiation; **CD34 stage** (CD34+ cord blood-derived mononuclear cells), **CD7 stage** (CD7+ CD5- CD1a-), **CD5 stage** (CD7+ CD5+ CD1a-) and **CD1a stage** (CD7+ CD5+ CD1a+). Replicates from each of the four different stages represents biological replicates. Gencode hg38 genes are shown at the bottom with a 10 kb scale bar. A-E corresponds to the area of CD34, CD7, CD5, TAL1, LMO2, and HEB.

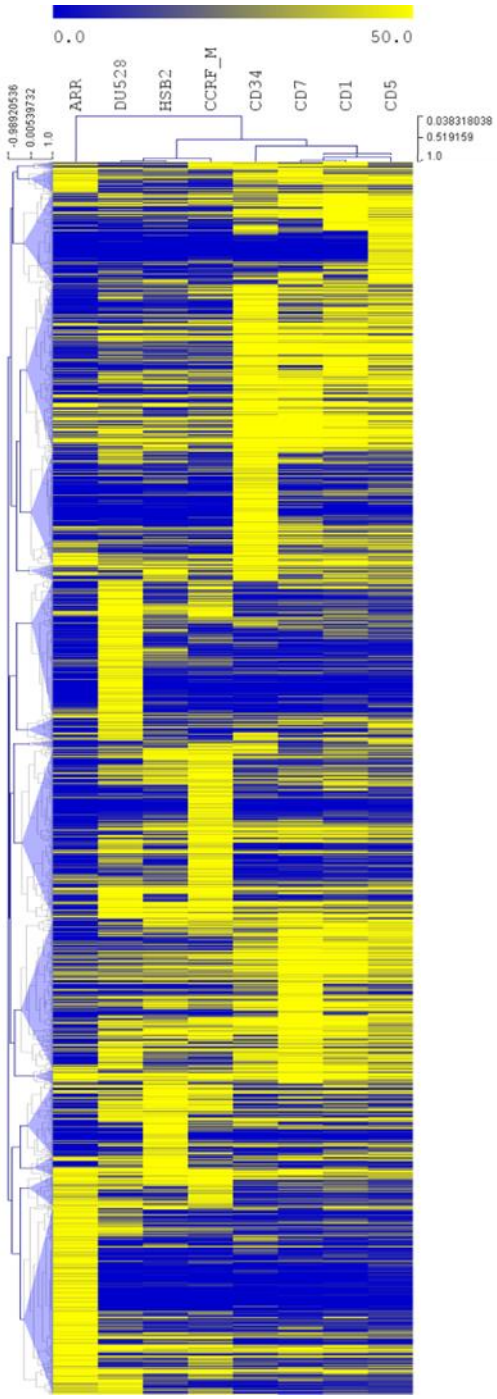


Figure 6.13 Heat map showing hierarchical clustering of differentially expressed genes in RNAseq of T cell progenitors and T-ALL cell lines

Heat map showing clustering of genes with significantly differential expression in four stages of human T-cell differentiation and the indicated T-ALL cell lines. Cluster 4 out of 16 is indicated with the red arrow.

Identification of genes differentially expressed in developing T cells and T-ALL cell lines

Our analysis of RNAseq data of four T cell stages and the four T-ALL cell lines revealed genes that were differentially expressed with statistical significance. Using hierarchical clustering analysis, we recognised 16 clusters of differentially expressed genes. Cluster four included genes that were significantly differentially expressed T cell progenitors compared to T-ALL cells, as shown in **Figure 6.13**. We applied a functional annotation analysis of cluster four genes, which was done to describe biological processes regulated by the genes of cluster four. We found that the genes were significantly associated with biological processes including translation; positive regulation of apoptosis, protein ubiquitination; regulation of GTPases; base-excision repair; reactive oxygen metabolic processes, and metabolism of reactive oxygen (**Figure 6.14**). These results of gene ontology of aberrantly expressed genes in T-ALL compared to normal T cell progenitors are indicative of oncogenic mechanisms in T-ALL cells.

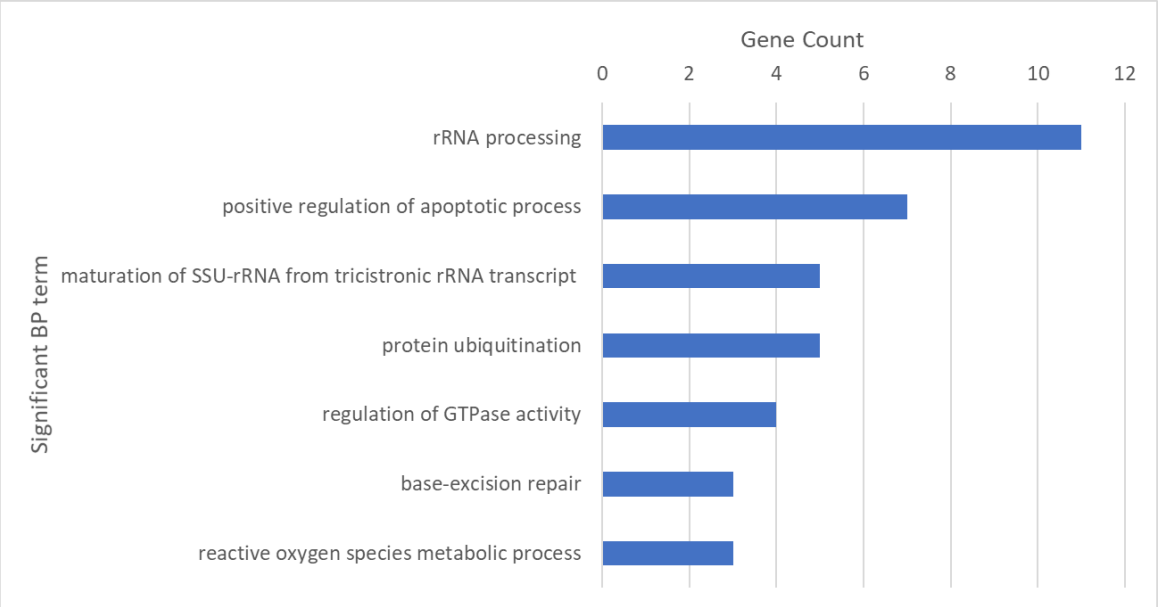


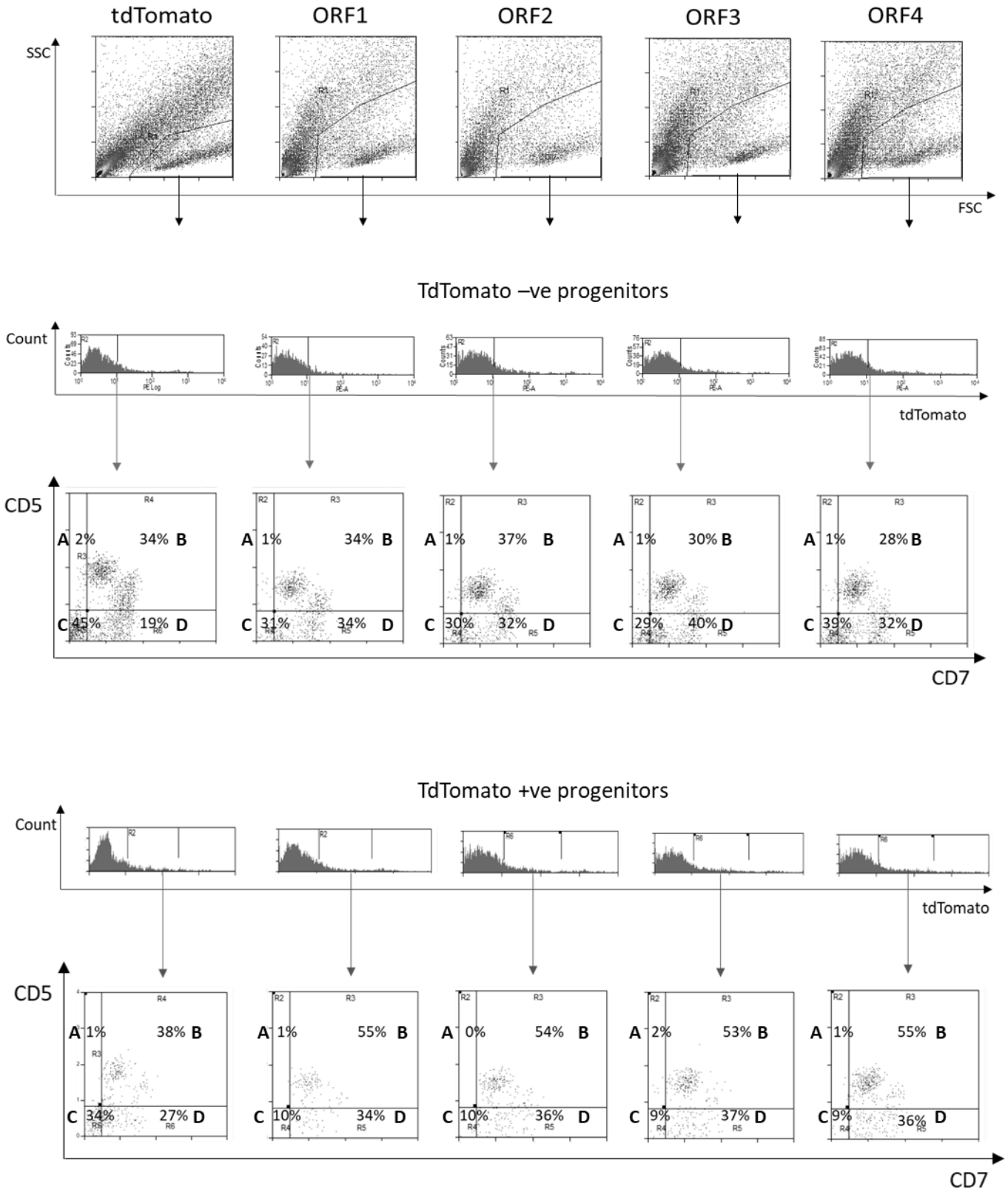
Figure 6.14 Bar chart shows the statistically significant pathways enriched in the gene ontology of genes in cluster 4

Cluster 4 represent genes with differential gene expression between T cell progenitors and T-ALL cell lines. Redundancy in the significant Biological processes (BP) was removed and the number of genes corresponding to each term of BP is shown.

TAL1 overexpressing progenitors progressed into $CD7^+$ $CD5^+$ T cell progenitors faster than wild type cells.

Next, we asked whether the overexpression of a specific TAL1 isoform in CD34 progenitors would influence human T cell differentiation. The immunophenotype of human T cells was assayed seven days after electroporation of CD34 cells with the TAL1 ORF1/ORF4 PB plasmids or with the tdTomato vector control plasmid and incubation in the in vitro T cell differentiation co-culture. The co-culture of electroporated cells was started with tdTomato-positive CD34 cells which were sorted twenty-four hours after electroporation. Initially, we planned to monitor the coculture and analyse after two or three weeks of culture to have more cells enough to sample.

However, while monitoring the culture, we observed using fluorescent microscopy that the number of tdTomato-positive cells decreased while negative cells expanded. T cells culture transfected with tdTomato control plasmid also showed a reduction in tdTomato-positive cells with the progress of the culture. Therefore, we had to assess CD7 and CD5 within a week in the *in vitro* T cell culture. **Figure 6.15** showed that both the differentiation markers, CD7 and CD5, were not different when comparing T cells expressing different TAL1 isoforms. However, we found that all of the T cells that were overexpressing any of the TAL1 isoforms (tdTomato +) displayed a higher percentage of CD7 CD5 double-positive cells compared to tdTomato negative cells and compared to the Tomato control. Our analysis of tdTomato positive and negative cells came from the same culture, and therefore they were cultured under the same conditions and this further support our findings.


tdTomato
ORF1
ORF2
ORF3
ORF4
SSC
FSC
TdTomato –ve progenitors
Count
tdTomato
CD5
A 2% 34% B
C 45% 19% D
A 1% 34% B
C 31% 34% D
A 1% 37% B
C 30% 32% D
A 1% 30% B
C 29% 40% D
A 1% 28% B
C 39% 32% D
CD7
TdTomato +ve progenitors
Count
tdTomato
CD5
A 1% 38% B
C 34% 27% D
A 1% 55% B
C 10% 34% D
A 0% 54% B
C 10% 36% D
A 2% 53% B
C 9% 37% D
A 1% 55% B
C 9% 36% D
CD7

	TdTomato-ve			TdTomato+ve		
TdTomato	Count	% Hist	% All	Count	% Hist	% All
Total	2682	100.00	5.99	491	100.00	1.10
A	46	1.72	0.10	6	1.22	0.01
B	917	34.19	2.05	187	38.09	0.42
C	1201	44.78	2.68	165	33.60	0.37
D	518	19.31	1.16	133	27.09	0.30
ORF1	Count	% Hist	% All	Count	% Hist	% All
Total	1408	100.00	9.58	187	100.00	1.27
A	14	0.99	0.10	1	0.53	0.01
B	479	34.02	3.26	103	55.08	0.70
C	438	31.11	2.98	18	9.63	0.12
D	477	33.88	3.24	65	34.76	0.44
ORF2	Count	% Hist	% All	Count	% Hist	% All
Total	1619	100.00	8.58	350	100.00	1.85
A	20	1.24	0.11	1	0.29	0.01
B	592	36.57	3.14	189	54.00	1.00
C	489	30.20	2.59	34	9.71	0.18
D	518	32.00	2.74	126	36.00	0.67
ORF3	Count	% Hist	% All	Count	% Hist	% All
Total	1777	100.00	6.50	414	100.00	1.51
A	18	1.01	0.07	7	1.69	0.03
B	541	30.44	1.98	219	52.90	0.80
C	513	28.87	1.88	35	8.45	0.13
D	705	39.67	2.58	153	36.96	0.56
ORF4	Count	% Hist	% All	Count	% Hist	% All
Total	2002	100.00	6.39	558	100.00	1.78
A	20	1.00	0.06	5	0.90	0.02
B	555	27.72	1.77	305	54.66	0.97
C	789	39.41	2.52	49	8.78	0.16
D	638	31.87	2.04	199	35.66	0.63

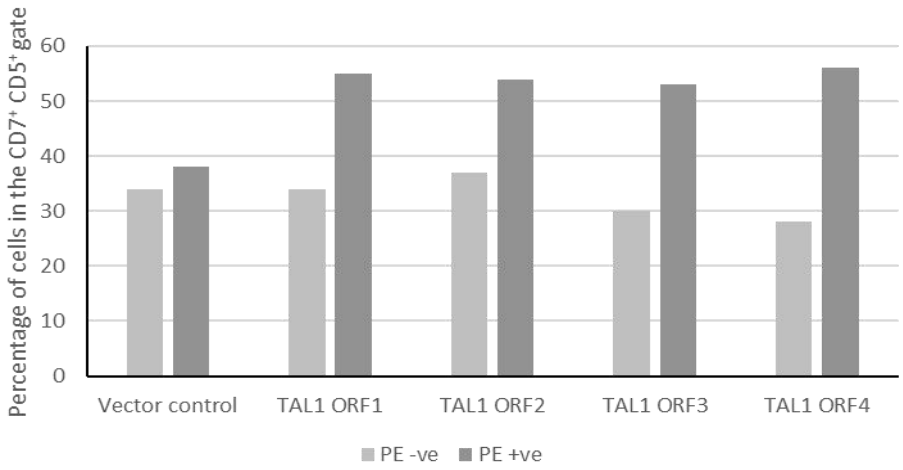


Figure 6.15 FACS analysis of T cell differentiation markers CD7, and CD5 in human T cells transfected with TAL1 isoforms

CD7 and CD5 expression seven days after electroporation of cord blood $CD34^+$ cells either with tdTomato-T2A-FLAG-TAL1 ORF1/ORF4 or with tdTomato vector control and culture in the in-vitro T cell differentiation system.

Discussion:

TAL1 isoforms in T-ALL

The size and structural differences between the TAL1 isoforms could influence their functional characteristics. Several variations in protein interactions and phosphorylation sites have been described for TAL1 isoforms. Differential roles of TAL1 isoforms have been demonstrated in haematopoietic progenitors and during erythropoiesis (Calkhoven et al., 2003; Jin et al., 2016). Overexpression of different *Tal1* isoforms in haematopoietic progenitors showed that they support the erythroid lineage over the megakaryocyte lineage. In erythropoiesis, the short isoform was shown to induce erythropoiesis because unlike the long isoforms, the short isoforms do not interact with suppressor ETO2. TAL1 overexpression is a recurrent event in T-ALL, and several mechanisms of oncogenic roles in T-ALL pathogenesis have been studied. In mouse studies, a short isoform of TAL1 has been shown to cause T-ALL (Aplan et al., 1997c; Gerby et al., 2014; Tatarek et al., 2011; Tremblay et al., 2010). However, the differential roles of TAL1 isoforms in their oncogenic contribution to T-ALL remain to be determined.

To study the consequences of TAL1 overexpression in T-ALL at the isoform level by constructing four plasmids expressing different isoforms of TAL1 protein (**Figure 6.1**). We constructed tdTomato-T2A-FLAG-TAL1 isoform ORF1, ORF2, ORF3, and ORF4 recombinant cDNA sequences individually and cloned each into doxycycline-inducible lentiviral vectors (Tet one puro) and subsequently to the PiggyBac-CAG vector. Including the T2A sequence permitted us to co-express tdTomato fluorescent protein and TAL1 as one transcript. The resulting fusion protein will be cleaved at the T2A sequence separating tdTomato and TAL1 proteins after translation. The FLAG tag was added to gain the ability to evaluate each of the transfected TAL1 isoforms individually.

We wanted to test the plasmids in a cell line that does not express indigenous TAL1. For this, we transfected HEK-293 cells with TAL1 plasmids. Then, we evaluated the protein expression of the different TAL1 isoforms (**Figure 6.2**). The HEK-293 cells do not naturally express TAL1, as shown in the untransfected samples. Western blot analysis of FLAG tag and TAL1 protein expression validated the efficacy of the plasmid by showing the overexpression of specific isoforms of TAL1 in the transfected cells, with molecular weights consistent with the transfected isoforms. The sequence of the long TAL1 isoforms can generate the shorter isoforms. Therefore, the presence of additional bands in the TAL1 western of the longer isoforms, like the ORF2 lysate, was expected. Nevertheless, our results showed that the TAL1

isoform in the construct was the dominant isoform in terms of expression, indicating that the plasmids are useful in overexpressing a particular isoform.

Similarly, the FLAG western results showed that the transfected cells expressed the FLAG-tagged TAL1 isoform with correct sizes; however, additional bands were also detected. In the expression of FLAG-tagged TAL1, alternative start sites would not be an issue, but alternative splicing could still produce shorter isoforms from the transfected FLAG-tagged TAL1 transcript. It is also likely to be some unspecific signal from the Ab.

TAL1-ORF1 and TAL1-ORF4 stable T-ALL cell lines:

In our experience, the transduction of lentiviral vectors, the PiggyBac (PB) transposase system is an efficient tool to establish stable T-ALL cell lines. The PB transposon system consisted of the validated tdTomato-T2A-FLAG-TAL1 isoform PB-CAG plasmids and a Transposase plasmid. The T-ALL cell lines were transfected and then sorted based on tdTomato expression twice to achieve a pure population of stably transfected cells.

We demonstrated that the stable TAL1-ORF1/4 T-ALL cell lines overexpress the transfected ORF1 or ORF4 TAL1 isoform using TAL1 western blot (**Figure 6.4**). There was an evident overexpression of the transfected isoform over the endogenous isoforms already present in the T-ALL cell lines. The overexpression of the TAL1-ORF4 transgene was lower than ORF1. DU.528 stable TAL1-ORF1/4 cell lines showed a lower signal of TAL1, yet the dominant isoform corresponded to the transfected ORF. Moreover, we showed the FLAG-tag in all the TAL1-ORF1 T-ALL cells. FLAG-tag levels in TAL1-ORF4 overexpressing cells was detectable only in DU.528.

The expression of FLAG-tagged TAL1-ORF1/4 transgenes was evaluated further using immunocytochemistry of the stable cell lines. The immunostaining signal of FLAG-tag and TAL1 protein was comparable to the western blots, especially concerning the lower expression of the FLAG-TAL1-ORF4 transgene in ARR, HSB-2, CCRF-CEM (**Figure 6.3**). The fluorescence of tdTomato in these three TAL1-ORF4 cell lines was not lower than ORF1 cells. Thus, the difference in the expression of FLAG-TAL1 ORF4 transgene is most likely a result of posttranslational mechanisms suppressing TAL1 short isoforms, such as degradation. Consistent with this notion, the three indicated cell lines favour the long isoforms of TAL1. Nevertheless, all TAL1 westerns showed the transfected isoform as dominant over the TAL1 isoforms endogenously expressed in the cells. Collectively, the results show that we

successfully established T-ALL cell lines with stable overexpression of either ORF1 or ORF4 TAL1 isoforms.

Overexpression of TAL1 isoforms affect T cell differentiation markers:

Next, we wanted to investigate the effect of overexpression of the full length or short isoform of TAL1 on the differentiation stage of the T-cell lines. After stable integration of the TAL1-ORF1 and TAL1-ORF4 constructs in our T-ALL cell lines, we measured the cell surface expression of CD7, CD5, CD1a, and CD3 (**Figure 6.5**). FACS analysis showed that the T-ALL cell lines overexpressing ORF4 displayed a more differentiated immunophenotype compared to ORF1 overexpressing cells, except for the early T-cell T-ALL cell line, ARR. ARR initially has very little CD7 and does not express CD5, CD1a, and CD3. The increase in the differentiation immunophenotype was best portrayed by the CD marker correlated with the furthest differentiation stage for each of the cell lines before transfection. Accordingly, we showed an increase in cells with CD7 expression in DU.528, CD5 in HSB-2, and CD3 in CCRF-CEM overexpressing ORF4 compared to ORF1. The results provide preliminary evidence supporting a differential role of TAL1 isoforms in the regulation of T cell differentiation in T-ALL. The experiment overexpresses TAL1 isoforms with the presence of the endogenous TAL1 isoforms. Also, the immunophenotype of the T-ALL cell lines is usually stable. Therefore, any change in the expression of the CD markers would be of interest. Nevertheless, statistical analysis of multiple repeats should be completed to validate this conclusion.

Regulation of T cell differentiation markers by TAL1 was shown using an in *vitro* T cell differentiation culture of $CD34^+$ transfected with FLAG-TAL1 ORF1, ORF2, ORF3, and ORF4. We showed that TAL1 overexpressing cells had more CD7 and CD5-expressing cells compared to the tdTomato vector control. Although the results did not show a differential effect between different isoforms, it did support TAL1-mediated regulation of the development of early T cell progenitors (**Figure 6.15**). In TAL1 transgenic mouse models, an increase in the DN and ISP T cell progenitors has been reported (O'Neil et al., 2004). Specifically, O`neil et al. showed that TAL1 transgenic mice have a block of differentiation with an increase in the DN2 and DN3. Our results indicate the need for further characterisation of the effect of TAL1 overexpression on T cell differentiation using human *in vitro* as a model of mimicking TAL1 pre-leukemic effects and how it could induce leukaemia.

In the wild-type T cell differentiation culture, we were able to produce RNAseq datasets representing four stages of T cell differentiation (**Figure 6.12**). We identified genes differentially expressed in normal T cell development, compared to the four T-ALL cell lines, from data previously generated. The results of the hierarchical clustering identified a cluster of genes with differential expression in normal differentiating T cell progenitors and T-ALL cells (**Figure 6.13**). Annotation of the genes revealed significantly overrepresented biological processes (**Figure 6.14**). Many of the significantly enriched biological processes were also shown to be significant in TAL1 and HEB MS and ChIPseq analysis like apoptosis, GTPases, metabolic processes. The results of gene ontology are consistent with the abnormal cancer-related processes (Hanahan & Weinberg, 2011). Malignant cells proliferate uncontrollably and inhibit apoptosis. Additionally, the metabolic pathways must be tailored to maintain the energy and metabolites required for cell division. Our results highlight several genes to be differentially expressed in T-ALL cells, which could represent downstream targets of T-ALL oncogenicity.

Genomic occupancy of TAL1 in T-ALL cell lines overexpressing TAL1 isoforms:

We showed differences in the genomic occupancy of TAL1 ChIP-seq in TAL1-ORF1 and TAL1-ORF4 T-ALL cells which presented further evidence for differential roles of TAL1 isoforms in T-ALL. Both ChIPseq datasets were generated using TAL1 antibody, so our analysis evaluated changes in TAL1 genomic occupancy upon the overexpression of ORF1 or ORF4. We showed overlapping and distinct TAL1 ChIPseq of ORF1 and ORF4 overexpressing T-ALL cell lines (**Figure 6.6**). ARR had the most noticeable difference between ORF1 and ORF4 TAL1 ChIPseq results, whilst DU.528, HSB-2, and CCRF-CEM showed a greater overlap between ORF1 and ORF4 binding.

Also, we showed how ORF1 and ORF4 TAL1 ChIPseq datasets relate to the TAL1 ChIPseq from untransfected T-ALL cells. TAL1-ORF1 and TAL1-ORF4 ChIPseq datasets overlapped partially with the TAL1 ChIPseq of untransfected T-ALL cells. The peaks of the untransfected TAL1 ChIPseq could be representing the binding of all TAL1 isoform. Therefore, the portion of TAL ChIPseq that does not overlap with TAL1-ORF1 and TAL1-ORF4 datasets could be bound with TAL1 isoforms ORF2 and ORF3. On the other hand, the TAL1-ORF1 and TAL1-ORF4 peaks that were not present in TAL1 ChIPseq of untransfected cells could indicate the introduction of additional isoform-specific peaks as a result of overexpression.

No differential binding of TAL1 interacting partners to TAL1-ORF1 or TAL1-ORF4 occupied regions was observed in the ChIPseq intensity heatmaps. We visualised the datasets of TAL1 interacting partners in the matrices of ORF1-only, Common, and ORF4-only TAL1 ChIPseq peaks. We compared the distribution of ChIPseq datasets of TAL1 transcriptional complex. Using signal intensity heat maps the HEB ChIPseq, and MeDIP datasets were analysed at TAL1/isoform occupied regions. To further examine the heatmaps comparison also included a panel of genome-wide datasets, including TAL1, LMO2, LDB1, RUNX.1, GATA2, and LYL1 ChIPseq and DNAaseI, all from the untransfected T-ALL cell lines. The aim was to identify isoform-specific patterns in the overlap between the analysed datasets. Therefore, we concluded that the overlap between the indicated datasets that were collected in the untransfected T-ALL did not show differences in their overlap with TAL1 at the ORF1-only and ORF4-only regions. The heatmaps comparison did not address the effect of isoform overexpression on the distribution of TAL1 interacting partners in transfected cells. However, it showed that the peaks with the highest enrichment were the overlapping peaks between the three TAL1 ChIPseq datasets (Common). We analysed the characteristics of the overlapping (Common) and the distinct (ORF1-only, and ORF4-only) peaks. We found enriched motifs specific for the ORF1-only or ORF4-only peaks and motifs of the overlapping regions (**Figure 6.7-6.9**). Our results showed that the JDP2 motif was enriched only in the ORF4-only peaks in ARR, DU.528, and HSB-2. The JDP2 motif enrichment in ORF4-only regions verified our finding of JDP2 (Jun dimerization protein 2) motif in regions occupied by HEB, TAL1, LMO2 in CCRF-CEM.

Our attempts to produce FLAG ChIPseq and RNAseq data for TAL1-ORF1/4 T-ALL cell lines were not successful. We tried to prepare libraries from T-ALL cell lines overexpressing different isoforms twice and for unknown reasons both attempts have failed. Further work on FLAG-ChIP such as testing other antibodies that work in ChIP and troubleshoot the difficulty in generating RNAseq datasets from the isolated samples will greatly demonstrate the effects of differential characteristics of TAL1 isoforms on gene expression.

CHAPTER SEVEN

OVERALL DISCUSSION

TAL1 is aberrantly expressed in T-ALL mostly as a result of the *SIL-TAL* deletion (Belver & Ferrando, 2016; Brown et al., 1990; Ferrando et al., 2002). Our investigations described novel molecular characteristics of TAL1/HEB transcriptional complex in T-ALL. Also, we compared the TAL1 oncogenic role in T-ALL at the isoform level. Two cellular models were applied to study the oncogenic impact of TAL1 aberrant expression in T-ALL. We utilised four human T-ALL cell lines; ARR, DU.528, HSB-2, and CCRF-CEM as well as an *in vitro* human T cell differentiation system. TAL1 exerts its oncogenic role in T-ALL using two main mechanisms. First, the disassociation of E-proteins homo/heterodimers because of TAL1 aberrant expression and the consequent disruption of the E-protein mediated transcriptional regulation, which is vital for normal T cell development (Kee, 2009).

TAL1/HEB transcriptional complex: Why HEB?

The aberrantly expressed TAL1 can heterodimerise with any of the mammalian E-proteins (E2A, HEB, and E2-2) normally expressed in T cells (O'Neil, Shank, Cusson, Murre, & Kelliher, 2004). In T cell development, the expression pattern of E-proteins varies, resulting in the formation of different E-proteins heterodimers such as E2A/HEB. The differences between E-proteins were observed in their knockout models. $E2A^{-/-}$ models were blocked partially at the DN1 stage, $E2\text{-}2^{-/-}$ showed a block at DN3 stage, and HEB had more substantial block but later at the ISP stage (Bain et al., 1997b; Barndt et al., 1999; Wikström, Forssell, Penha-Goncalves, Bergqvist, & Holmberg, 2008). Consequently, the study of the oncogenic effects of TAL1 heterodimerisation with E-proteins requires the specification of which E-protein is in the complex. We screened four human T-ALL cell lines overexpressing TAL1. In two of the T-ALL cell lines (DU.528 and HSB-2) HEB was detected in the mass spectrometry analysis of TAL1 co-immunoprecipitation. Before proceeding further, we also confirmed that HEB interacts with TAL1 in the other two cell lines.

Characteristics of TAL1 complex

Mouse models overexpressing TAL1 alone do not develop leukaemia except with long latency or combined with other oncogenes (Aifantis et al., 2008) indicating that TAL1 oncogenic role in T-ALL relies on its protein-protein interactions. TAL1/E-protein dimer take part in a transcriptional complex mainly containing LMO, LDB1, GATA, RUNX, and ETS (Matthews, Lester, Joseph, & Curtis, 2013). We showed the genomic occupancy of TAL1/LMO2/HEB transcription complex by overlapping the ChIPseq data obtained in T-ALL cell lines. A similar investigation was done in Jurkat and CCRF-CEM in which ChIPseq data of TAL1, HEB, and LMO1/2 were overlapped in addition to GATA3 and RUNX.1 (Sanda et al., 2012). They demonstrated that almost all (99%) of HEB peaks were occupied with TAL1. Our intersection analysis also showed the majorities of HEB overlapped with TAL1 in the T-ALL cell lines except in ARR. The four T-ALL cell lines have aberrant expression of TAL1 protein. The T cell surface markers of these cell lines indicate that they represent leukemic cells that have been blocked at different stages of T-cell differentiation as shown by (Burger et al., 1999; Sanda et al., 2012) and confirmed in this study. The difference observed in the TAL1/LMO2/HEB overlap between ARR, and the SIL-TAL cell lines is an example of how the complex shift its genomic occupancy. We studied factors regulating the shift in the complex genomic occupancy by analysing characteristics of the genomic regions bound with the TAL1/LMO2/HEB complex.

TAL1 association with DNA binding proteins regulate its genomic distribution

First, it has been shown that interactions with DNA binding transcription factors regulate the genomic distribution of the complex in haematopoietic cells and T-ALL cells (Palii et al., 2011b; Wilson et al., 2010; Wu et al., 2014). Known characteristics of TAL1 complex like regulating genes by binding distant enhancers via LDB1 (Krivega et al., 2014) was also evident in the gene association results of HEB, TAL1, LMO2 (HTL). The analysis of the total HEB ChIPseq showed enrichment of the motifs for known TAL1 partners (RUNX, ETS, GATA, and E-box) but not in all four cell lines. Structural studies showed that the association between E-proteins and RUNX/ETS/GATA require the link via LMO protein (El Omari et al., 2013). Similarly, our analysis of HEB ChIPseq regions after overlapping with TAL1 and LMO2 neatly demonstrated enrichment of RUNX, ETS, GATA, and E-box as the top four highly enriched motifs in the four T-ALL cell lines.

HEB-occupied regions that are not shared with TAL1/LMO2 (HEB-only)

The analysis of the HTL overlap drives attention to the HEB-only section of the data, which is not shared by TAL1and LMO2. We found that ARR had a greater proportion of HEB-only peaks. In T cells E-proteins dimers such as E2A/HEB normally form and in transformed thymocytes, the aberrantly expressed TAL1 sequester E-proteins (Wang & Baker, 2015). Therefore, HEB-only regions in T-ALL cell lines could be representing regions bound with E-protein dimers that remained in the transformed thymocytes. To confirm this hypothesis, genomic occupancy of other E-proteins E2A, and E2-2 at the HEB-only regions should be assessed in future studies. Another possible explanation that the HEB-only regions are occupied with TAL1 closely related protein LYL1. LYL1 shares 85% of homology at the helix-loop-helix domain, which is the place for dimerisation with E-proteins (Baer, 1993; Giroux et al., 2007). Visualisation of LYL1 enrichment signal at HEB ChIPseq tracks showed LYL1 at the HEB-only regions. TAL1 and LYL1 are both abnormally expressed in T-ALL and gene expression analysis linked LYL1 to ETP-ALL, which is a high-risk group with an immunophenotype of $CD5^-$ $CD1a^-$ (Jain et al., 2016). ARR cell line, where we found the most HEB-only regions, demonstrate the ETP-ALL immunophenotype. Therefore, it can be inferred that a complex containing LYL1 and HEB occupy the HEB-only regions. demonstrating LYL1 and HEB dimerisation and identifying other members of the complex remain to be studied.

Another factor that was suggested to be a determinant of the genomic distribution of TAL1 transcriptional complex is histone modifiers (Wu et al., 2014). DNA methylation, which is another epigenetic regulator, has never been associated with the regulation of TAL1 genomic occupancy. Nevertheless, the interactive relationship between DNA methylation and transcription factor binding to DNA is not a new concept (Zhu, Wang, & Qian, 2016). We compared the DNA methylation status of genomic regions occupied with TAL1 transcriptional complex to the HEB-only regions. We were able to show that DNA methylation as determined using MeDIP occurs at HEB-only regions and rarely in the HTL regions. ARR has the most HEB-only and showed more overlap with methylated CpG regions compared to the SIL-TAL. The number of the MeDIPseq peaks was comparable across the four T-ALL cell lines, and the lack of DNA methylation signal in SIL-TAL cell lines was observed only when overlapped with HEB regions. In haematopoiesis, it has been shown that the genome of the lymphoid lineage shows more DNA methylation compared to the myeloid lineage (Ji et al., 2010) and TAL1 was shown to favour the myeloid over lymphoid lineage (Kunisato et al., 2004). These studies support an inverse relationship between TAL1 and DNA methylation. More

importantly, TAL1 expression correlated with lower DNA methylation in T-ALL patients (Haider et al., 2019). In order to identify the mechanism that link HEB but not TAL1 to DNA methylation. We searched proteins detected in TAL1 pulldown and exclusively in ARR since it is the cell line with the most methylated HEB regions. Our analysis highlighted the chromatin target of PRMT1 (CHTOP) a protein known to bind methylated DNA (Takai et al., 2014).

Integration of TAL1/HEB mass spectrometry, motif analysis in T-ALL identified potential interaction with major regulators of T cell development.

Differentially regulated transcription factors notably control haematopoiesis. T cell development is regulated by a precisely regulated expression of transcription factors. Several stage-dependent transcription factors control T cell differentiation (Yui & Rothenberg, 2014). The expression of Growth factor independence 1 (GFI1) is maintained and suppressed at the SP stage (Yücel, Karsunky, Klein-Hitpass, & Möröy, 2003).

We compared the data obtained using mass spectrometry of HEB in T-ALL to motifs identified in HEB-only regions. We found GFI1 in ARR and DU.528. GFI1 motif was significantly enriched in DU.528 HEB-only regions. GFI1 is vital in haematopoiesis with evidence of involvement in T-ALL pathogenesis (Möröy, Vassen, Wilkes, & Khandanpour, 2015). *GFI1* is a known target of E2A and HEB (Jones & Zhuang, 2007). Human GFI1 has not been shown to have protein-protein interaction with any of the E-proteins, but their homologous proteins interact in other species (Jafar-Nejad, Tien, Acar, & Bellen, 2006; Ravasi et al., 2010).

The zinc-finger transcription factor IKAROS (Ikzf1) is a major regulator of T cell proliferation and checkpoints between DN3-DP stage (Yui & Rothenberg, 2014). We compared TAL1 and HEB MS datasets and identified the shared proteins. Analysis of the TAL1/HEB shared proteins MS dataset highlighted an interaction with IKAROS, a known T cell regulator. Similar to GFI1, the interaction between IKORAS and TAL1/HEB has been indicated before homologues proteins in other species (Jafar-Nejad et al., 2006; Klein & Campos-Ortega, 1997). IKAROS was found in HEB and TAL1 MS for CCRF-CEM only. Still, between the four T-ALL cell lines we analysed CCRF-CEM is the latest in term of T cell differentiation and that fits with the DN3-DP window of IKAROS (Burger et al., 1999) . Thus, TAL1/HEB could be cooperating with existing transcription factors in a subtype of T-ALL with more differentiated immunophenotype.

TAL1 isoforms in T-ALL:

Post-translational regulation or splicing produce four isoforms of TAL1 (Calkhoven et al., 2003; Jin et al., 2016). In haematopoiesis, TAL1 isoforms were shown to modulate lineage choice. The two longer isoforms favoured the megakaryocytic while the shorter isoforms induced the erythroid lineage (Calkhoven et al., 2003). A subtype of T-ALL expressed only the short isoform of TAL1 (Aplan et al., 1997). A double transgenic mouse overexpressing the long or the short isoform of TAL1 with LMO1 showed they both similarly induced leukaemia (Ellsworth & Aplan, 1999). In more recent studies investigating TAL1 oncogenic model, the short isoform was used (Gerby et al., 2014a; Tatarek et al., 2011a; Tremblay, Herblot, Lecuyer, & Hoang, 2003).

We produced ARR, DU.528, HSB-2, and CCRF-CEM cell lines overexpressing human TAL1 long (ORF1) or short isoforms (ORF4). TAL1 structure is mainly composed of an activation domain and a b-HLH domain. The short isoforms also lack the activation domain, in addition to phosphorylation sites (Lécuyer & Hoang, 2004). To examine potential differences in TAL1 oncogenic role at the isoform levels, we designed TAL1 constructs overexpressing TAL1 longest and shortest isoforms. We produced ARR, DU.528, HSB-2, and CCRF-CEM cell lines overexpressing FLAG-TAL1 long or short isoforms. Human TAL1 long and short isoforms have been cloned before and used to study isoform differences in erythroid lineages (Jin et al., 2017b) but not in T-ALL.

Production of T-ALL stably overexpressing the short or the long isoform as the dominant isoform.

Our results showed that the cell lines overexpressed the expected isoforms and that it became the dominant isoform. This was not a given because the same cellular processes that make TAL1 short isoforms endogenously can still produce the short isoforms from the long isoform construct. Even though we were able to overexpress ORF1 in DU.528 which originally overexpressed the short isoforms, it was less than ORF1 in the other T-ALL cell lines suggesting underlying regulatory mechanisms that favour the production of the short isoform from ORF1 in DU.528. TAL1 cDNA we cloned did not include the uORF which has been shown to control the choice of TAL1 isoforms (Calkhoven et al., 2003). The mechanism they demonstrated involved eIF2-mediated translation which can be modulated using the inhibitors Rapamycin and 2AP. It would be interesting to apply these inhibitors in future studies of TAL1 isoforms in T-ALL. The specificity of the inhibitors should be taken into consideration because

the inhibitors act on upstream regulatory processes regulating TAL1 isoforms in addition to other cellular proteins.

Analysis of the genomic occupancy of TAL1 at the isoform level show similarities more than differences

After demonstrating that all the cell lines predominantly overexpressed TAL1 ORF1 or ORF4, the cell lines became a valuable cellular model to study TAL1 oncogenic role in T-ALL at the isoform levels. We hypothesised that TAL1 isoform would have differences in their oncogenic contribution to T-ALL. The genomic distribution of TAL1 in the TAL1-ORF1/4 cell lines shared the strongest peaks, but there were considerable isoform-specific peaks. Finding isoform-specific TAL1 ChIPseq peaks means that the regulation of the genomic distribution of TAL1 isoforms could be different which is in line with our hypothesis.

Since protein-protein interactions of TAL1 could explain the difference in the genomic occupancy between isoforms, we compared motifs enriched in TAL1 ChIPseq for ORF1 and ORF4 cells. Contrary to the hypothesis, most of the motifs were shared between the two isoforms. Few ORF1-only motifs and ORF4-only motifs were significantly enriched. Like the motif analysis, the visualisation of TAL1 partners at ORF1-only, or ORF4-only regions did not show a notable difference. From the results, we could conclude that both isoforms bind at similar regions with TAL1 partners.

Even though no differences were observed between the isoforms in the motifs or overlap with TAL1 partners, our results do not rule out differences in protein-protein interactions between the isoforms and non-DNA binding proteins such as cofactors or regulators of TAL1 phosphorylation sites. As shown in the erythroid lineage the interaction with ETO2, a suppressor of erythroid differentiation, is lost in TAL1 short isoform (Jin et al., 2017). Using co-immunoprecipitation of FLAG-TAL1-ORF1/ORF4 in future studies could aid in identifying such protein interactions. Known cofactors such as p300 and P/CAF, corepressors like mSin3A, HDAC, and LSD1. (Huang & Brandt, 2000; Huang et al., 1999; Huang et al., 2000; Xu, Z., Huang, Chang, Agulnick, & Brandt, 2003), and regulators of TAL1 phosphorylation like ERK1 should be investigated (Huang & Brandt, 2000; Huang et al., 1999; Huang et al., 2000; Lécuyer & Hoang, 2004; Xu et al., 2003).

On the other hand, comparing the effects of TAL1 isoforms on T cell differentiation markers in T-ALL cell lines demonstrated preliminary evidence for differences between isoforms. TAL1-ORF4 induces differentiation in T-ALL cell lines. Interruption of normal T cell differentiation is a characteristic of T-ALL malignancy (Aifantis et al., 2008). Inducing differentiation in the T-ALL cell lines if validated indicate that the short isoform is less oncogenic. This fits with the fact that the DU.528 cell line which favour the short isoform express more LYL1 proteins compared to the other T-ALL cell lines. One can suggest that the malignant DU.528 cells overexpressed LYL1 to compensate for the inferior TAL1 isoform oncogenic effects. However, one of the main mechanisms of TAL1 oncogenic role in T-ALL is the sequestration of E-proteins. Based on the protein structure of TAL1, the helix-loop-helix domain is retained in TAL1 short isoforms. Therefore, it can heterodimerise with E-proteins and induce T-ALL. Further investigation of the TAL1 isoforms in T-ALL is required to understand its involvement in T-ALL and how it relates to LYL1.

In the *in vitro* T cell differentiation system, we observed an increase in the percentage of early T cell progenitors in the four TAL1 isoforms. It is unclear if the increase would be maintained beyond what we observed seven days after transfection. TAL1 has been shown to induce the self-renewal capacity of T cell progenitors (Gerby et al., 2014c; Tatarek et al., 2011). Based on the literature, a block in differentiation should not be expected by overexpressing TAL1 alone at least not before a long latency period (Aifantis et al., 2008). Thus, the increase shown in our human T cell progenitors after the overexpression of TAL1 isoforms could indicate an enhancement of the self-renewal characteristics of differentiating thymocytes. Optimising the *in vitro* protocol to obtain a higher number of stably expressing cells in culture, enough for RNAseq sample preparation or testing TAL1 isoform using in Vivo Xenograft Mice would provide great insight into how continued TAL1 expression facilitates T-ALL development.

TAL1 effect on gene expression:

Consistent with the shift in its genomic occupancy, TAL1 regulation of gene expression varies depending on cell type (Lécuyer & Hoang, 2004). TAL1 oncogenic role is associated with TAL1 transcriptional complex and its regulation of gene expression as shown by identifying individual TAL1 targets genes such as TALL1, RALDH2, and NKX3.1 (Kusy et al., 2010; Ono et al., 1997; Ono et al., 1998). Previously described genome-wide analysis of TAL1 described its mediated regulation of gene expression. In haematopoiesis, the expression of TAL1 is high in HSCs and multipotent progenitors and generally suppressed as cells

differentiate except for erythroid, megakaryocytic and mast cells (Hoang, Lambert, & Martin, 2016b; Kallianpur, Jordan, & Brandt, 1994b). In haematopoietic stem cells and early progenitors, TAL1 was shown to regulate genes supporting cell survival and proliferation (Lacombe et al., 2010; Reynaud et al., 2005) while in differentiating cells it regulates lineage-specific genes such as erythroid related genes (Kassouf et al., 2010; Palii et al., 2011; Palomero et al., 2006). In T-ALL, the TAL1 has been shown to regulate cell survival and inhibit apoptosis, which is in line with its role in HSCs/MPPs. The inhibition of T cell-specific genes normally activated by E-proteins dimers has been shown (Palii et al., 2011; Palomero et al., 2006; Sanda et al., 2012). In agreement with the literature, our gene association analysis of regions occupied with TAL1 LMO2 and HEB showed association with proliferative growth factor genes and cell-specific pathway such as genes related to T cell activation in SIL-TAL cells. We compared the transcriptome data obtained from four stages of T cell differentiation to that from four T-ALL cell lines. The differentially expressed genes identified represent pathways abnormally regulated in TAL1 overexpressing T-ALL cells. Some of the genes differentially expressed were consistent with TAL1 known targets like apoptotic pathways. Others have not been described as TAL1 regulated pathways. Their role in T-ALL specifically and whether they are targets of TAL1 remains to be examined.

To conclude, this study provided a comprehensive analysis of the TAL1/HEB transcriptional complex and its oncogenic contribution to T-ALL. Our results introduced DNA methylation as a determining factor of the genomic occupancy of TAL1 transcriptional complex containing HEB. The study provided preliminary evidence for novel differences between TAL1 long and short isoforms.

ARR TAL1 Mass spectrometry	PAIRB_HUMAN HNRPL_HUMAN LAP2B_HUMAN CH60_HUMAN TOP2A_HUMAN PSPC1_HUMAN EF1A1_HUMAN RCC2_HUMAN XRN2_HUMAN CPSF1_HUMAN RL5_HUMAN RS9_HUMAN RS21_HUMAN RS3A_HUMAN RL8_HUMAN SFRS1_HUMAN U2AF1_HUMAN RS7_HUMAN CA077_HUMAN RS30_HUMAN THOC4_HUMAN RS16_HUMAN SLIRP_HUMAN SFRS3_HUMAN CH10_HUMAN RL27_HUMAN RS6_HUMAN TAL1_HUMAN RL31_HUMAN RL18_HUMAN RUXF_HUMAN MPCP_HUMAN RL29_HUMAN DCD_HUMAN RL17_HUMAN RL10L_HUMAN SNR40_HUMAN RS29_HUMAN PGAM5_HUMAN H15_HUMAN RL19_HUMAN RS11_HUMAN RL26L_HUMAN ILF2_HUMAN EFTU_HUMAN H3L_HUMAN RS8_HUMAN ERH_HUMAN RS25_HUMAN COCA1_HUMAN SSBP3_HUMAN RS4X_HUMAN DJB11_HUMAN ACTS_HUMAN ACTB_HUMAN MYO1B_HUMAN NOM1_HUMAN POTEI_HUMAN FLNA_HUMAN MYH10_HUMAN SPTN1_HUMAN MYH9_HUMAN CLH1_HUMAN EDC4_HUMAN EIF3A_HUMAN SPTB2_HUMAN TBA1B_HUMAN SEPT2_HUMAN LONP2_HUMAN HSP7C_HUMAN RS5_HUMAN RS2_HUMAN ACTN1_HUMAN PABP1_HUMAN UBP2L_HUMAN MYO1G_HUMAN REQU_HUMAN PP1B_HUMAN PP1G_HUMAN PP1A_HUMAN HNRPR_HUMAN AP2A1_HUMAN HNRH1_HUMAN RS24_HUMAN HNRH3_HUMAN PRP8_HUMAN LSP1_HUMAN FLNB_HUMAN PDIA6_HUMAN G3BP2_HUMAN IF2B3_HUMAN PICAL_HUMAN U520_HUMAN SEPT7_HUMAN BMP2K_HUMAN RBM39_HUMAN PLEC_HUMAN	**ARR HEB Mass spectrometry**	YG001_HUMAN S35G3_HUMAN LV102_HUMAN FUS_HUMAN ZN318_HUMAN RL26_HUMAN LONM_HUMAN TP8L3_HUMAN RL12_HUMAN XRCC5_HUMAN XRCC6_HUMAN RL18_HUMAN DHSA_HUMAN HVCN1_HUMAN RS16_HUMAN PRKDC_HUMAN ARF3_HUMAN RS17L_HUMAN MRRP3_HUMAN MYH10_HUMAN RL10A_HUMAN NDUC1_HUMAN LEUK_HUMAN RS11_HUMAN HNRPU_HUMAN TOP2A_HUMAN PDIA4_HUMAN RL27_HUMAN IF5AL_HUMAN PARP1_HUMAN G6PI_HUMAN FAS_HUMAN GDIR2_HUMAN RL32_HUMAN PPIB_HUMAN RL14_HUMAN CALX_HUMAN ABCA1_HUMAN TCPQ_HUMAN NONO_HUMAN RL23_HUMAN ZFYV9_HUMAN AIFM1_HUMAN BAG2_HUMAN 1433B_HUMAN GBB1_HUMAN THIO_HUMAN TBA1C_HUMAN PYR1_HUMAN GFI1_HUMAN BAF_HUMAN TCPE_HUMAN TOP1_HUMAN DHX9_HUMAN SHRM3_HUMAN TCPZ_HUMAN RL13A_HUMAN RS26L_HUMAN MAGT1_HUMAN F263_HUMAN RAC3_HUMAN CLIC1_HUMAN TCPD_HUMAN TB10C_HUMAN UBF1_HUMAN VASP_HUMAN DDX5_HUMAN HMGA1_HUMAN ILF3_HUMAN TCPH_HUMAN RL19_HUMAN RPN2_HUMAN ATPA_HUMAN DC1I2_HUMAN GMFB_HUMAN HGB1A_HUMAN FXR1_HUMAN RAN_HUMAN RRBP1_HUMAN S10AB_HUMAN IGLL1_HUMAN EFTU_HUMAN H3C_HUMAN SND1_HUMAN ML12B_HUMAN GRN_HUMAN ECHA_HUMAN RL28_HUMAN U520_HUMAN HTF4_HUMAN CLH1_HUMAN ARHG2_HUMAN PTBP1_HUMAN SYDC_HUMAN ACLY_HUMAN SYRC_HUMAN HNRPM_HUMAN

	COPB2_HUMAN LRCH4_HUMAN SRP14_HUMAN UBAP2_HUMAN RRBP1_HUMAN CKAP5_HUMAN TOP2B_HUMAN DDX21_HUMAN TRFE_HUMAN COR1C_HUMAN RL4_HUMAN C1QBP_HUMAN RFTN1_HUMAN TBB5_HUMAN RUVB1_HUMAN ADT2_HUMAN SRSF1_HUMAN FLOT2_HUMAN BRI3B_HUMAN ARP3_HUMAN DREB_HUMAN CPSF5_HUMAN RL21_HUMAN FIP1_HUMAN SRSF3_HUMAN SRSF7_HUMAN EP15R_HUMAN CN166_HUMAN SNUT2_HUMAN		STIP1_HUMAN RL40_HUMAN MOV10_HUMAN ARPC5_HUMAN AT2A2_HUMAN C2C2L_HUMAN TBB4B_HUMAN RECQ1_HUMAN LAP2A_HUMAN H15_HUMAN ABCC9_HUMAN DAD1_HUMAN URP2_HUMAN SERA_HUMAN ACL6A_HUMAN SRSF7_HUMAN RL6_HUMAN CYTF_HUMAN CHD7_HUMAN PGAM2_HUMAN RO52_HUMAN DDX3Y_HUMAN PRP8_HUMAN CHTOP_HUMAN CALM_HUMAN TFAM_HUMAN ARPC4_HUMAN SYEP_HUMAN XRN2_HUMAN IMB1_HUMAN RL13_HUMAN SUMO4_HUMAN RL23A_HUMAN TLN1_HUMAN RPN1_HUMAN TAGL2_HUMAN DYL2_HUMAN PSPC1_HUMAN PRS8_HUMAN FUBP2_HUMAN HNRPQ_HUMAN DDX17_HUMAN ERH_HUMAN COF1_HUMAN EF1G_HUMAN MPCP_HUMAN RL22L_HUMAN PSMD3_HUMAN RS10_HUMAN EWS_HUMAN COR1A_HUMAN ERO1A_HUMAN YBOX1_HUMAN KU70_HUMAN TBB2C_HUMAN TBB3_HUMAN FLNB_HUMAN RAC2_HUMAN RS17_HUMAN LCA5L_HUMAN CND1_HUMAN RL18A_HUMAN TCPG_HUMAN TRAP1_HUMAN
DU.528 TAL1 Mass spectrometry	TBA1B_HUMAN H2B1K_HUMAN ACTB_HUMAN ACTS_HUMAN SPTN1_HUMAN THOC4_HUMAN ROA0_HUMAN MYH9_HUMAN RL23_HUMAN COPE_HUMAN RL19_HUMAN PABP1_HUMAN RL18_HUMAN FLNA_HUMAN MYO1G_HUMAN ZN638_HUMAN NOLC1_HUMAN FMR1_HUMAN ARPC3_HUMAN NH2L1_HUMAN RS30_HUMAN RS7_HUMAN FIP1_HUMAN GUF1_HUMAN RL32_HUMAN RL5_HUMAN COPB2_HUMAN RS26_HUMAN RS9_HUMAN RS17L_HUMAN U5S1_HUMAN RL28_HUMAN CKAP5_HUMAN LRC59_HUMAN CAPR1_HUMAN	**DU.528 HEB Mass spectrometry**	ACTG_HUMAN TOP2A_HUMAN DHX9_HUMAN TOP2B_HUMAN TBB2C_HUMAN H12_HUMAN RL7A_HUMAN C1TC_HUMAN RL26_HUMAN RL3_HUMAN RL6_HUMAN TBB3_HUMAN IF4A3_HUMAN RL28_HUMAN RS3A_HUMAN RL18_HUMAN TOP1_HUMAN PRKDC_HUMAN RS11_HUMAN MYH10_HUMAN U5S1_HUMAN ILEU_HUMAN RS2_HUMAN SYEP_HUMAN RL12_HUMAN SMC3_HUMAN H15_HUMAN MATR3_HUMAN DHX30_HUMAN K1967_HUMAN RAC2_HUMAN TBA1C_HUMAN RL13_HUMAN RL35A_HUMAN ANXA2_HUMAN

	RL38_HUMAN NOP10_HUMAN ATS9_HUMAN KI67_HUMAN PRP19_HUMAN TRA2B_HUMAN RL8_HUMAN SRSF9_HUMAN PM14_HUMAN RL29_HUMAN ARGL1_HUMAN DHX15_HUMAN SND1_HUMAN RBP56_HUMAN RL14_HUMAN DHX9_HUMAN MYL6_HUMAN IF16_HUMAN TOP2B_HUMAN SRSF1_HUMAN RL21_HUMAN H15_HUMAN ERH_HUMAN RS4X_HUMAN H12_HUMAN RL27_HUMAN PSPC1_HUMAN GAR1_HUMAN RS11_HUMAN KIF14_HUMAN RL35A_HUMAN HNRPR_HUMAN TOP2A_HUMAN PAIRB_HUMAN SEPT9_HUMAN HNRPC_HUMAN SRSF3_HUMAN PDIP3_HUMAN CLCA_HUMAN RL24_HUMAN SRP09_HUMAN LAP2A_HUMAN RS16_HUMAN RL18A_HUMAN PRKDC_HUMAN RL26L_HUMAN PARP1_HUMAN SEPT1_HUMAN NMNA1_HUMAN NAT10_HUMAN DDX50_HUMAN DDX21_HUMAN RL11_HUMAN POTEJ_HUMAN PCBP2_HUMAN DKC1_HUMAN DDX1_HUMAN RS5_HUMAN RL3_HUMAN COIL_HUMAN RU2A_HUMAN R39L5_HUMAN TOP1_HUMAN G3BP2_HUMAN SMD2_HUMAN FRG1_HUMAN RL10A_HUMAN SEPT2_HUMAN RS8_HUMAN SF3B1_HUMAN PERM_HUMAN NDUA4_HUMAN RL10L_HUMAN DOCK8_HUMAN CNBP_HUMAN H2A1H_HUMAN HCD2_HUMAN RL34_HUMAN PPIB_HUMAN RIF1_HUMAN COPA_HUMAN TBB5_HUMAN RBMX_HUMAN PP1G_HUMAN SEPT7_HUMAN NEP1_HUMAN RS25_HUMAN SRSF7_HUMAN IF2B1_HUMAN SEPT6_HUMAN RFC1_HUMAN DECR_HUMAN HSP7C_HUMAN RS27A_HUMAN HNRH1_HUMAN RL17_HUMAN RS29_HUMAN RL36_HUMAN UBAP2_HUMAN RUXG_HUMAN U520_HUMAN		RL4_HUMAN ML12B_HUMAN RL7_HUMAN RL27_HUMAN ROAA_HUMAN GBLP_HUMAN PCBP1_HUMAN CDC2_HUMAN RENT1_HUMAN YBOX1_HUMAN TCPD_HUMAN RL13A_HUMAN ATPO_HUMAN TCPG_HUMAN ARHG2_HUMAN COPA_HUMAN SMCA4_HUMAN HNRPF_HUMAN H33_HUMAN RS17_HUMAN ELAV1_HUMAN RS15A_HUMAN RL18A_HUMAN RL19_HUMAN MY18A_HUMAN PCNA_HUMAN TPP2_HUMAN SUCB1_HUMAN HNRPQ_HUMAN RL14_HUMAN DDX5_HUMAN DNMT1_HUMAN TCPQ_HUMAN RL21_HUMAN TITIN_HUMAN ESYT2_HUMAN DICER_HUMAN SMC2_HUMAN SMC1A_HUMAN SFRS7_HUMAN PTN6_HUMAN SYMC_HUMAN MOV10_HUMAN RL10A_HUMAN GMFB_HUMAN TAGL2_HUMAN RL36L_HUMAN PSAL_HUMAN GIMA1_HUMAN CPSF1_HUMAN PAP1L_HUMAN HMGLX_HUMAN RS16_HUMAN DDX23_HUMAN PRS10_HUMAN SEPT1_HUMAN MPCP_HUMAN MSH6_HUMAN TCPE_HUMAN MCM3_HUMAN MCM6_HUMAN GDIR2_HUMAN TCPZ_HUMAN RL5_HUMAN MCM7_HUMAN PARP1_HUMAN MCM4_HUMAN RS27_HUMAN CYFP1_HUMAN DDX17_HUMAN DDX21_HUMAN RL23A_HUMAN SMCA5_HUMAN ADT3_HUMAN HNRPU_HUMAN ROA0_HUMAN SFRS3_HUMAN SUCB2_HUMAN RUXF_HUMAN ERAP1_HUMAN LV102_HUMAN CTR9_HUMAN YBEY_HUMAN RL30_HUMAN RL32_HUMAN SPTN4_HUMAN RS17L_HUMAN WNT7B_HUMAN SMCA2_HUMAN IF5AL_HUMAN GSG1_HUMAN HV302_HUMAN SRBP1_HUMAN H2AV_HUMAN ADT2_HUMAN RLA2_HUMAN RL9_HUMAN INT1_HUMAN LARP1_HUMAN DIAP1_HUMAN HNRPD_HUMAN

C

	RS3A_HUMAN CG050_HUMAN CN166_HUMAN RS24_HUMAN SPF27_HUMAN RL30_HUMAN SRSF4_HUMAN RL4_HUMAN UBP2L_HUMAN RS2_HUMAN PAR14_HUMAN NOP2_HUMAN CLH1_HUMAN RLA1_HUMAN ADT2_HUMAN RL15_HUMAN RBM10_HUMAN RBM39_HUMAN RL9_HUMAN RUXE_HUMAN HNRPU_HUMAN SPTB2_HUMAN HNRPF_HUMAN SF3B5_HUMAN KV303_HUMAN ILF2_HUMAN R13AX_HUMAN YBOX1_HUMAN DAZP1_HUMAN CALM_HUMAN STAB1_HUMAN ML12B_HUMAN CPSF5_HUMAN UBP10_HUMAN ATD3A_HUMAN DEK_HUMAN CHERP_HUMAN RL37A_HUMAN DDX3X_HUMAN SR140_HUMAN CAZA1_HUMAN TRFE_HUMAN DYL2_HUMAN COR1C_HUMAN PCM1_HUMAN FINC_HUMAN DNJC9_HUMAN SF3B3_HUMAN RBBP4_HUMAN EF1A1_HUMAN XRN2_HUMAN SMC1A_HUMAN SRP14_HUMAN TCOF_HUMAN RL31_HUMAN NHP2_HUMAN RSMB_HUMAN RS23_HUMAN LSP1_HUMAN PDIA6_HUMAN AP2A2_HUMAN PLEC_HUMAN RUXF_HUMAN RU2B_HUMAN IF4G1_HUMAN EDC4_HUMAN DJB11_HUMAN LYAR_HUMAN HNRH3_HUMAN EIF3B_HUMAN HTF4_HUMAN SSBP2_HUMAN		ARPC3_HUMAN KAPCA_HUMAN CNDD3_HUMAN AIFM1_HUMAN RL23_HUMAN THIO_HUMAN U2AF1_HUMAN GFI1_HUMAN AT5SL_HUMAN UBL5_HUMAN PERI_HUMAN SYIC_HUMAN THOC4_HUMAN RS26L_HUMAN RL11_HUMAN CLIC1_HUMAN CTBP1_HUMAN CPT2_HUMAN GIPC1_HUMAN ILF3_HUMAN ATPA_HUMAN FXR1_HUMAN RAN_HUMAN RL40_HUMAN SUCA_HUMAN RS21_HUMAN RAE1L_HUMAN TBB4B_HUMAN CDK1_HUMAN SRSF5_HUMAN UBP31_HUMAN OTUB1_HUMAN SYLC_HUMAN PTBP1_HUMAN CD99_HUMAN SPTN1_HUMAN C2C2L_HUMAN SYQ_HUMAN PURA_HUMAN RPAC1_HUMAN SPAT5_HUMAN CNN2_HUMAN CASP3_HUMAN RM21_HUMAN HMGB1_HUMAN G6PD_HUMAN PLSL_HUMAN SRSF7_HUMAN DDX3Y_HUMAN CAND1_HUMAN RL36A_HUMAN ARPC4_HUMAN LONM_HUMAN DPM3_HUMAN NUDC_HUMAN SYDC_HUMAN XRN2_HUMAN SRP72_HUMAN IMB1_HUMAN RL36_HUMAN SMD2_HUMAN KIF5C_HUMAN MSH2_HUMAN SRSF3_HUMAN RL24_HUMAN MYO1G_HUMAN ACLY_HUMAN CLH1_HUMAN DECR_HUMAN EF1G_HUMAN EIF3B_HUMAN PSMD3_HUMAN SRSF1_HUMAN EWS_HUMAN UBF1_HUMAN NAT10_HUMAN
HSB-2 TAL1 Mass spectrometry	TAL1_HUMAN CE170_HUMAN ACTB_HUMAN CPSF1_HUMAN SPT5H_HUMAN DDX21_HUMAN RL11_HUMAN TOP2A_HUMAN EIF3B_HUMAN HTF4_HUMAN MYO1G_HUMAN MD1L1_HUMAN XRN2_HUMAN DHX9_HUMAN H15_HUMAN RS23_HUMAN K1C16_HUMAN SRBP1_HUMAN GUF1_HUMAN KI67_HUMAN DOCK8_HUMAN TCOF_HUMAN PLEC_HUMAN	**HSB-2 HEB Mass spectrometry**	MYH10_HUMAN HNRPM_HUMAN DHX9_HUMAN TOP1_HUMAN H12_HUMAN RL18_HUMAN RL13A_HUMAN TBB2C_HUMAN RL14_HUMAN RL18A_HUMAN RL12_HUMAN H15_HUMAN RL10A_HUMAN H33_HUMAN MOV10_HUMAN TBA3E_HUMAN HNRH1_HUMAN DDX17_HUMAN LV102_HUMAN RL32_HUMAN RL28_HUMAN PRKDC_HUMAN GRDN_HUMAN

	ZN638_HUMAN AHNK_HUMAN ADT2_HUMAN TBB5_HUMAN CAPR1_HUMAN SRRM2_HUMAN EMSA1_HUMAN LAP2B_HUMAN SON_HUMAN TOP2B_HUMAN FNBP4_HUMAN DDX3X_HUMAN TBA1B_HUMAN SPTN1_HUMAN UBAP2_HUMAN		XDH_HUMAN ADDA_HUMAN RS11_HUMAN H2A1H_HUMAN MCM7_HUMAN MATR3_HUMAN XRCC6_HUMAN RL24_HUMAN RS15A_HUMAN CO6A3_HUMAN LARP1_HUMAN DIDO1_HUMAN DDX21_HUMAN RPB9_HUMAN RS26L_HUMAN TB10C_HUMAN CPT2_HUMAN VASP_HUMAN H13_HUMAN MCM3_HUMAN CJ118_HUMAN ILF3_HUMAN RL19_HUMAN HNRPU_HUMAN ATPA_HUMAN H90B2_HUMAN TRNK1_HUMAN PARP1_HUMAN C2C2L_HUMAN RL21_HUMAN C1TC_HUMAN RL27A_HUMAN MCM4_HUMAN RL6_HUMAN ATPO_HUMAN MCM6_HUMAN RBM14_HUMAN ML12B_HUMAN SMD2_HUMAN RL13_HUMAN SRSF3_HUMAN K1967_HUMAN MYO1G_HUMAN YBOX1_HUMAN
CCRF-CEM TAL1 Mass spectrometry	DDX21_HUMAN H12_HUMAN H15_HUMAN TOP1_HUMAN H1X_HUMAN HNRPR_HUMAN HSP7C_HUMAN IF16_HUMAN HNRPU_HUMAN DECR_HUMAN RS3A_HUMAN GBLP_HUMAN ILF3_HUMAN DDX3Y_HUMAN RS4X_HUMAN NOP2_HUMAN KHDR1_HUMAN PABP3_HUMAN THOC4_HUMAN HNRL1_HUMAN CA077_HUMAN MPCP_HUMAN XRN2_HUMAN MYO1G_HUMAN DHX15_HUMAN U2AF1_HUMAN SFRS7_HUMAN ATD3A_HUMAN K2C6B_HUMAN HNRPL_HUMAN RL8_HUMAN SFRS1_HUMAN NOP56_HUMAN TITIN_HUMAN ERH_HUMAN RUXGL_HUMAN GUF1_HUMAN H2A1H_HUMAN RBMX_HUMAN RL13A_HUMAN TOP2A_HUMAN RL10_HUMAN RL15_HUMAN SSBP_HUMAN RS9_HUMAN RS8_HUMAN RL19_HUMAN SRSF2_HUMAN RBM39_HUMAN ACTS_HUMAN ACTB_HUMAN RL35A_HUMAN RS29_HUMAN ARGL1_HUMAN H2B1K_HUMAN	**CCRF-CEM HEB Mass spectrometry**	DHX9_HUMAN HNRPM_HUMAN PTBP1_HUMAN TOP1_HUMAN RS3A_HUMAN RL7A_HUMAN H12_HUMAN H14_HUMAN SYDC_HUMAN H13_HUMAN NONO_HUMAN TBB2C_HUMAN H2B1H_HUMAN RL13_HUMAN PRKDC_HUMAN RL7_HUMAN C1TC_HUMAN H2B1J_HUMAN ARHG2_HUMAN MYH10_HUMAN SERA_HUMAN MATR3_HUMAN H33_HUMAN DDX17_HUMAN COPA_HUMAN TBA3C_HUMAN HNRPU_HUMAN U5S1_HUMAN RU17_HUMAN K1967_HUMAN ADT2_HUMAN RL12_HUMAN ACTBL_HUMAN RS2_HUMAN PCNA_HUMAN PCBP1_HUMAN RL28_HUMAN H15_HUMAN TBB3_HUMAN RL14_HUMAN ATPA_HUMAN RL35A_HUMAN TOP2A_HUMAN RCC2_HUMAN DDX21_HUMAN HNRPC_HUMAN RS15A_HUMAN RL13A_HUMAN U2AF1_HUMAN RS17_HUMAN RL18_HUMAN ILF3_HUMAN RS11_HUMAN PRP6_HUMAN MCM6_HUMAN

	RL11_HUMAN		GBLP_HUMAN
	RL4_HUMAN		RL19_HUMAN
	RS7_HUMAN		H2AJ_HUMAN
	RS25_HUMAN		DDX5_HUMAN
	RL14_HUMAN		RL30_HUMAN
	RS11_HUMAN		ROA0_HUMAN
	TOP2B_HUMAN		RBM14_HUMAN
	RL3_HUMAN		SMC3_HUMAN
	DKC1_HUMAN		PRP19_HUMAN
	PLEC_HUMAN		RL10A_HUMAN
	DHX9_HUMAN		RL3_HUMAN
	RL31_HUMAN		HNRH1_HUMAN
	RL38_HUMAN		RS16_HUMAN
	MYL6_HUMAN		RL4_HUMAN
	RL34_HUMAN		THOC4_HUMAN
	RL26L_HUMAN		DHX30_HUMAN
	RL36_HUMAN		HMGA1_HUMAN
	FBRL_HUMAN		SFRS1_HUMAN
	RL1D1_HUMAN		TCPA_HUMAN
	NOG1_HUMAN		UBF1_HUMAN
	SRRM2_HUMAN		FUS_HUMAN
	RAE1L_HUMAN		RL21_HUMAN
	SRSF6_HUMAN		DYH8_HUMAN
	SRSF5_HUMAN		YBOX1_HUMAN
	SRSF4_HUMAN		RL6_HUMAN
	RL18_HUMAN		HNRPQ_HUMAN
	RL30_HUMAN		DDX3Y_HUMAN
	NOLC1_HUMAN		G3BP1_HUMAN
	GAR1_HUMAN		ATPG_HUMAN
	DDX18_HUMAN		TCPG_HUMAN
	ILF2_HUMAN		ML12B_HUMAN
	R39L5_HUMAN		PSMD3_HUMAN
	SRSF3_HUMAN		CDC2_HUMAN
	RL28_HUMAN		SFRS2_HUMAN
	SF3B5_HUMAN		SF3B1_HUMAN
	RS23_HUMAN		MOV10_HUMAN
	RL27_HUMAN		PARP1_HUMAN
	SRRM1_HUMAN		MYO1G_HUMAN
	NAT10_HUMAN		TBL1R_HUMAN
	PABP1_HUMAN		LEG9_HUMAN
	DHX30_HUMAN		IMA2_HUMAN
	HNRH1_HUMAN		RL36A_HUMAN
	HNRPF_HUMAN		DC1L1_HUMAN
	HNRH2_HUMAN		MCM4_HUMAN
	rndMIB1_HUMAN		RL27_HUMAN
	RS16_HUMAN		RL18A_HUMAN
	TCOF_HUMAN		SMCA5_HUMAN
	H1T_HUMAN		SYFA_HUMAN
	RS30_HUMAN		RL36L_HUMAN
	RL17_HUMAN		RL5_HUMAN
	KI67_HUMAN		XRN2_HUMAN
	NOP10_HUMAN		PAIRB_HUMAN
	SNUT1_HUMAN		TITIN_HUMAN
	RS27A_HUMAN		MCM7_HUMAN
	NUP43_HUMAN		RL32_HUMAN
	RL18A_HUMAN		CYFP1_HUMAN
	U520_HUMAN		NAT10_HUMAN
	SRSF9_HUMAN		TCPE_HUMAN
	RS17L_HUMAN		CAPR1_HUMAN
	LMNB1_HUMAN		KINH_HUMAN
	MYH9_HUMAN		XIRP2_HUMAN
	SF3B1_HUMAN		H2B1O_HUMAN
	MYH10_HUMAN		LV102_HUMAN
	IF4A3_HUMAN		CTR9_HUMAN
	SRSF1_HUMAN		CA173_HUMAN
	BCLF1_HUMAN		ENTP6_HUMAN
	RL9_HUMAN		RND2_HUMAN
	ELYS_HUMAN		DYXC1_HUMAN
	RL36A_HUMAN		RS17L_HUMAN
	RL5_HUMAN		WNT7B_HUMAN
	SELH_HUMAN		XRCC6_HUMAN
	SPTB2_HUMAN		CP062_HUMAN
	RL24_HUMAN		XRCC5_HUMAN
	PRP8_HUMAN		PABP3_HUMAN
	RBM10_HUMAN		PM14_HUMAN
	CAZA1_HUMAN		TCPD_HUMAN
	SMD2_HUMAN		TRA2B_HUMAN
	DDX50_HUMAN		KHDR1_HUMAN
	SPTN1_HUMAN		GSG1_HUMAN
	NDUA4_HUMAN		H2A1H_HUMAN
	MBB1A_HUMAN		TBB4B_HUMAN
	HS2ST_HUMAN		SRBP1_HUMAN
	CAPR1_HUMAN		MCCA_HUMAN
	RRP5_HUMAN		RLA2_HUMAN
	TRA2B_HUMAN		RL24_HUMAN
	MPH6_HUMAN		MSH3_HUMAN
	EXOS9_HUMAN		INT1_HUMAN
	EMSA1_HUMAN		RMI1_HUMAN
	RL23_HUMAN		RL26L_HUMAN
	DOCK8_HUMAN		SRSF2_HUMAN
	PABP2_HUMAN		LARP1_HUMAN
	RM34_HUMAN		ZFYV9_HUMAN
	LC7L2_HUMAN		STAU2_HUMAN
	PAIRB_HUMAN		RL23_HUMAN
	TRIPC_HUMAN		TBA1A_HUMAN
	RL21_HUMAN		RL37_HUMAN
	ML12B_HUMAN		TRI47_HUMAN
	RUXE_HUMAN		DAB1_HUMAN
	RL10A_HUMAN		PERI_HUMAN
	RBMX2_HUMAN		EMD_HUMAN
	SNR40_HUMAN		SYEP_HUMAN

f

	HP1B3_HUMAN RBP56_HUMAN IKZF1_HUMAN EXOS6_HUMAN DDX23_HUMAN BRX1_HUMAN HNRPC_HUMAN RMTL1_HUMAN LC7L3_HUMAN EXOS5_HUMAN NOP58_HUMAN SRP09_HUMAN LYAR_HUMAN LAP2B_HUMAN RS6_HUMAN RBP2_HUMAN TBL3_HUMAN ABCD3_HUMAN BLM_HUMAN CLH2_HUMAN SON_HUMAN PRP19_HUMAN DEK_HUMAN NXF1_HUMAN FRG1_HUMAN CG050_HUMAN DDX3X_HUMAN SNUT2_HUMAN RS15_HUMAN RL29_HUMAN UTP18_HUMAN RFC1_HUMAN RS2_HUMAN PDIP3_HUMAN PM14_HUMAN PCBP1_HUMAN CD2B2_HUMAN BAZ1A_HUMAN RS24_HUMAN DYL1_HUMAN RU2A_HUMAN DDX47_HUMAN THOC2_HUMAN PWP2_HUMAN PCBP2_HUMAN PP1G_HUMAN SPF27_HUMAN ADT2_HUMAN DNJB6_HUMAN TPX2_HUMAN NH2L1_HUMAN RL32_HUMAN FIP1_HUMAN PR38A_HUMAN SMD1_HUMAN U5S1_HUMAN CENPF_HUMAN TBA1B_HUMAN SF3B3_HUMAN CRNL1_HUMAN MRE11_HUMAN ACINU_HUMAN RS5_HUMAN NU205_HUMAN NUP54_HUMAN SF3B4_HUMAN NUP98_HUMAN TBL2_HUMAN PELP1_HUMAN KRR1_HUMAN SCML2_HUMAN RFC3_HUMAN ZCCHV_HUMAN IMA2_HUMAN GLYR1_HUMAN RUXF_HUMAN TRA2A_HUMAN SR140_HUMAN NIP7_HUMAN SRSF7_HUMAN EXOS7_HUMAN RBBP4_HUMAN CDK2_HUMAN NMNA1_HUMAN RPA1_HUMAN NDUA9_HUMAN TBB5_HUMAN RFC2_HUMAN GRSF1_HUMAN KV303_HUMAN PRP6_HUMAN FMR1_HUMAN RS26_HUMAN KRI1_HUMAN IMB1_HUMAN RU17_HUMAN KIF22_HUMAN SRP14_HUMAN DNJA3_HUMAN LDB1_HUMAN ROA0_HUMAN		RS26L_HUMAN CPT2_HUMAN SPTB2_HUMAN CCD96_HUMAN MCM3_HUMAN RBM39_HUMAN SRSF1_HUMAN GMFB_HUMAN EWS_HUMAN FXR1_HUMAN SELH_HUMAN SF3B2_HUMAN SRSF5_HUMAN RAC2_HUMAN UBP31_HUMAN OTUB1_HUMAN SRP68_HUMAN ATX2L_HUMAN RS12_HUMAN AGFG1_HUMAN SK2L2_HUMAN RENT1_HUMAN HS105_HUMAN RU2A_HUMAN G6PD_HUMAN SRSF7_HUMAN ATPO_HUMAN IMDH2_HUMAN ROAA_HUMAN CDK2_HUMAN ELAV1_HUMAN DDX46_HUMAN RU2B_HUMAN RSMN_HUMAN SYMC_HUMAN RTCB_HUMAN LCK_HUMAN SMD2_HUMAN NOP56_HUMAN RL23A_HUMAN SRSF3_HUMAN SNUT1_HUMAN CH3L2_HUMAN IKZF1_HUMAN HNRL2_HUMAN SMC2_HUMAN FUBP2_HUMAN RB_HUMAN MPCP_HUMAN RL22L_HUMAN PSMD2_HUMAN NOP58_HUMAN ROA3_HUMAN

	YBOX1_HUMAN DDX52_HUMAN RLA1_HUMAN CPSF5_HUMAN GNAI2_HUMAN CALM_HUMAN RNPS1_HUMAN GRWD1_HUMAN GBG2_HUMAN RPA43_HUMAN SRS11_HUMAN U2AF2_HUMAN PRKDC_HUMAN RL37A_HUMAN AP2A2_HUMAN CDC42_HUMAN ATPO_HUMAN NOP16_HUMAN DDX55_HUMAN HNRH3_HUMAN G3BP2_HUMAN NUP53_HUMAN YMEL1_HUMAN THOC6_HUMAN RFC4_HUMAN LCK_HUMAN LSM2_HUMAN NHP2_HUMAN DNJC9_HUMAN NU214_HUMAN UBP2L_HUMAN C1QBP_HUMAN NOC4L_HUMAN RBM34_HUMAN MDC1_HUMAN ATPA_HUMAN RM01_HUMAN EDC4_HUMAN TADBP_HUMAN IF6_HUMAN		

References

Aifantis, I., Raetz, E., & Buonamici, S. (2008). Molecular pathogenesis of T-cell leukaemia and lymphoma. Nature Reviews Immunology, 8(5), 380-390.

Aplan, P. D., Jones, C. A., Chervinsky, D. S., Zhao, X., Ellsworth, M., Wu, C., et al. (1997a). An scl gene product lacking the transactivation domain induces bony abnormalities and cooperates with LMO1 to generate T-cell malignancies in transgenic mice. The EMBO Journal, 16(9), 2408-2419.

Aplan, P. D., Jones, C. A., Chervinsky, D. S., Zhao, X., Ellsworth, M., Wu, C., et al. (1997b). An scl gene product lacking the transactivation domain induces bony abnormalities and cooperates with LMO1 to generate T-cell malignancies in transgenic mice. The EMBO Journal, 16(9), 2408-2419.

Aplan, P. D., Jones, C. A., Chervinsky, D. S., Zhao, X., Ellsworth, M., Wu, C., et al. (1997c). An scl gene product lacking the transactivation domain induces bony abnormalities and cooperates with LMO1 to generate T-cell malignancies in transgenic mice. The EMBO Journal, 16(9), 2408-2419.

Awong, G., Herer, E., Surh, C. D., Dick, J. E., La Motte-Mohs, R. N., & Zuniga-Pflucker, J. C. (2009). Characterization in vitro and engraftment potential in vivo of human progenitor T cells generated from hematopoietic stem cells. Blood, 114(5), 972-982.

Baer, R. (1993). TAL1, TAL2 and LYL1: A family of basic helix-loop-helix proteins implicated in T cell acute leukaemia. Seminars in Cancer Biology, 4(6), 341-347.

Bain, G., Engel, I., Robanus Maandag, E. C., te Riele, H. P., Voland, J. R., Sharp, L. L., et al. (1997a). E2A deficiency leads to abnormalities in alphabeta T-cell development and to rapid development of T-cell lymphomas. Molecular and Cellular Biology, 17(8), 4782-4791.

Bain, G., Engel, I., Robanus Maandag, E. C., te Riele, H. P., Voland, J. R., Sharp, L. L., et al. (1997b). E2A deficiency leads to abnormalities in alphabeta T-cell development and to rapid development of T-cell lymphomas. Molecular and Cellular Biology, 17(8), 4782-4791.

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. Molecular and Cellular Biology, 20(18), 6677-6685.

Barndt, R., Dai, M. F., & Zhuang, Y. (1999). A novel role for HEB downstream or parallel to the pre-TCR signaling pathway during alpha beta thymopoiesis. Journal of Immunology (Baltimore, Md.: 1950), 163(6), 3331-3343.

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: A C API and toolkit for analyzing and managing BAM files. Bioinformatics, 27(12), 1691-1692.

Bellavia, D., Campese, A. F., Checquolo, S., Balestri, A., Biondi, A., Cazzaniga, G., et al. (2002). Combined expression of pTalpha and Notch3 in T cell leukemia

identifies the requirement of preTCR for leukemogenesis. Proceedings of the National Academy of Sciences of the United States of America, 99(6), 3788-3793.

Belver, L., & Ferrando, A. (2016). The genetics and mechanisms of T cell acute lymphoblastic leukaemia. Nature Reviews Cancer, 16(8), 494.

Bendelac, A., Savage, P. B., & Teyton, L. (2007). The biology of NKT cells. Annu.Rev.Immunol., 25, 297-336.

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(1), 49-59.

Bernasconi-Elias, P., Hu, T., Jenkins, D., Firestone, B., Gans, S., Kurth, E., et al. (2016). Characterization of activating mutations of NOTCH3 in T-cell acute lymphoblastic leukemia and anti-leukemic activity of NOTCH3 inhibitory antibodies. Oncogene, 35(47), 6077.

Besseyrias, V., Fiorini, E., Strobl, L. J., Zimber-Strobl, U., Dumortier, A., Koch, U., et al. (2007). Hierarchy of Notch–Delta interactions promoting T cell lineage commitment and maturation. Journal of Experimental Medicine, 204(2), 331-343.

Blankenberg, D., Kuster, G. V., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., et al. (2010). Galaxy: A web-based genome analysis tool for experimentalists. Current Protocols in Molecular Biology, 89(1), 19.10. 1-19.10. 21.

Blom, B., & Spits, H. (2006). Development of human lymphoid cells. Annu.Rev.Immunol., 24, 287-320.

Borowski, C., Martin, C., Gounari, F., Haughn, L., Aifantis, I., Grassi, F., et al. (2002). On the brink of becoming a T cell. Current Opinion in Immunology, 14(2), 200-206.

Braunstein, M., & Anderson, M. K. (2010). Developmental progression of fetal HEB−/− precursors to the pre-T-cell stage is restored by HEBAlt. European Journal of Immunology, 40(11), 3173-3182.

Braunstein, M., & Anderson, M. K. (2012). HEB in the spotlight: Transcriptional regulation of T-cell specification, commitment, and developmental plasticity. Clinical and Developmental Immunology, 2012

Bray, S. J. (2006). Notch signalling: A simple pathway becomes complex. Nature Reviews Molecular Cell Biology, 7(9), 678.

Brown, L., Cheng, J., Chen, Q., Siciliano, M., Crist, W., Buchanan, G., et al. (1990). Site-specific recombination of the tal-1 gene is a common occurrence in human T cell leukemia. The EMBO Journal, 9(10), 3343-3351.

Burger, R., Hansen-Hagge, T. E., Drexler, H. G., & Gramatzki, M. (1999). Heterogeneity of T-acute lymphoblastic leukemia (T-ALL) cell lines: Suggestion for classification by immunophenotype and T-cell receptor studies. Leukemia Research, 23(1), 19-27.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., et al. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature, 518(7540), 542.

Calkhoven, C. F., Muller, C., Martin, R., Krosl, G., Pietsch, H., Hoang, T., et al. (2003). Translational control of SCL-isoform expression in hematopoietic lineage choice. Genes & Development, 17(8), 959-964.

Capron, C., Lecluse, Y., Kaushik, A. L., Foudi, A., Lacout, C., Sekkai, D., et al. (2006). The SCL relative LYL-1 is required for fetal and adult hematopoietic stem cell function and B-cell differentiation. Blood, 107(12), 4678-4686.

Ceredig, R., & Rolink, T. (2002). A positive look at double-negative thymocytes. Nature Reviews Immunology, 2(11), 888.

Chagraoui, H., Kassouf, M., Banerjee, S., Goardon, N., Clark, K., Atzberger, A., et al. (2011). SCL-mediated regulation of the cell-cycle regulator p21 is critical for murine megakaryopoiesis. Blood, the Journal of the American Society of Hematology, 118(3), 723-735.

Chan, W. Y., Follows, G. A., Lacaud, G., Pimanda, J. E., Landry, J. R., Kinston, S., et al. (2007). The paralogous hematopoietic regulators Lyl1 and scl are coregulated by ets and GATA factors, but Lyl1 cannot rescue the early scl-/- phenotype. Blood, 109(5), 1908-1916.

Chen, L., Segal, D., Hukriede, N. A., Podtelejnikov, A. V., Bayarsaihan, D., Kennison, J. A., et al. (2002). Ssdp proteins interact with the LIM-domain-binding protein Ldb1 to regulate development. Proceedings of the National Academy of Sciences of the United States of America, 99(22), 14320-14325.

Cheng, Y., Zhang, Z., Slape, C., & Aplan, P. D. (2007). Cre-/IoxP-mediated recombination between the SIL and SCL genes leads to a block in T-cell development at the CD4-CD8-to CD4 CD8 transition. Neoplasia, 9(4), 315-321.

Chervinsky, D. S., Zhao, X. F., Lam, D. H., Ellsworth, M., Gross, K. W., & Aplan, P. D. (1999). Disordered T-cell development and T-cell malignancies in SCL LMO1 double-transgenic mice: Parallels with E2A-deficient mice. Molecular and Cellular Biology, 19(7), 5025-5035.

Condorelli, G. L., Facchiano, F., Valtieri, M., Proietti, E., Vitelli, L., Lulli, V., et al. (1996). T-cell-directed TAL-1 expression induces T-cell malignancies in transgenic mice. Cancer Research, 56(22), 5113-5119.

Cumano, A., Berthault, C., Ramond, C., Petit, M., Golub, R., Bandeira, A., et al. (2019). New molecular insights into immune cell development. Annual Review of Immunology, 37, 497-519.

Curtis, D. J., Hall, M. A., Van Stekelenburg, L. J., Robb, L., Jane, S. M., & Begley, C. G. (2004). SCL is required for normal function of short-term repopulating hematopoietic stem cells. Blood, 103(9), 3342-3348.

Curtis, D. J., Robb, L., Strasser, A., & Begley, C. G. (1997). The CD2-scl transgene alters the phenotype and frequency of T-lymphomas in N-ras transgenic or p53 deficient mice. Oncogene, 15(24), 2975.

De Smedt, M., Hoebeke, I., & Plum, J. (2004). Human bone marrow CD34 progenitor cells mature to T cells on OP9-DL1 stromal cell line without thymus microenvironment. Blood Cells, Molecules, and Diseases, 33(3), 227-232.

Dey, S., Curtis, D. J., Jane, S. M., & Brandt, S. J. (2010). The TAL1/SCL transcription factor regulates cell cycle progression and proliferation in differentiating murine bone marrow monocyte precursors. Molecular and Cellular Biology, 30(9), 2181-2192.

Dey, S., Curtis, D. J., Jane, S. M., & Brandt, S. J. (2010). The TAL1/SCL transcription factor regulates cell cycle progression and proliferation in differentiating murine bone marrow monocyte precursors. Molecular and Cellular Biology, 30(9), 2181-2192. doi:10.1128/MCB.01441-09

Dias, S., Månsson, R., Gurbuxani, S., Sigvardsson, M., & Kee, B. L. (2008). E2A proteins promote development of lymphoid-primed multipotent progenitors. Immunity, 29(2), 217-227.

Dik, W. A., Pike-Overzet, K., Weerkamp, F., de Ridder, D., de Haas, E. F., Baert, M. R., et al. (2005). New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. The Journal of Experimental Medicine, 201(11), 1715-1723.

Draheim, K., Hermance, N., Yang, Y., Arous, E., Calvo, J., & Kelliher, M. (2011). A DNA-binding mutant of TAL1 cooperates with LMO2 to cause T cell leukemia in mice. Oncogene, 30(10), 1252.

Durinck, K., Goossens, S., Peirs, S., Wallaert, A., Van Loocke, W., Matthijssens, F., et al. (2015). Novel biological insights in T-cell acute lymphoblastic leukemia. Experimental Hematology, 43(8), 625-639.

El Omari, K., Hoosdally, S. J., Tuladhar, K., Karia, D., Hall-Ponselé, E., Platonova, O., et al. (2013). Structural basis for LMO2-driven recruitment of the SCL: E47bHLH heterodimer to hematopoietic-specific transcriptional targets. Cell Reports, 4(1), 135-147.

Ellsworth, K. W. G., & Aplan, P. D. (1999). Disordered T-cell development and T-cell. Mol.Cell.Biol, 19(7), 5025.

Elwood, N. J., Zogos, H., Pereira, D. S., Dick, J. E., & Begley, C. G. (1998). Enhanced megakaryocyte and erythroid development from normal human CD34(+) cells: Consequence of enforced expression of SCL. Blood, 91(10), 3756-3765.

Endoh, M., Ogawa, M., Orkin, S., & Nishikawa, S. (2002). SCL/tal-1-dependent process determines a competence to select the definitive hematopoietic lineage prior to endothelial differentiation. The EMBO Journal, 21(24), 6700-6708.

Engel, I., & Murre, C. (2001). The function of E-and id proteins in lymphocyte development. Nature Reviews Immunology, 1(3), 193.

Ferrando, A. A., Neuberg, D. S., Staunton, J., Loh, M. L., Huard, C., Raimondi, S. C., et al. (2002). Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell, 1(1), 75-87.

Fortini, M. E. (2002). Signalling: Γ-secretase-mediated proteolysis in cell-surface-receptor signalling. Nature Reviews Molecular Cell Biology, 3(9), 673.

Gerby, B., Tremblay, C. S., Tremblay, M., Rojas-Sutterlin, S., Herblot, S., Hébert, J., et al. (2014a). SCL, LMO1 and Notch1 reprogram thymocytes into self-renewing cells. PLoS Genetics, 10(12), e1004768.

Gerby, B., Tremblay, C. S., Tremblay, M., Rojas-Sutterlin, S., Herblot, S., Hébert, J., et al. (2014b). SCL, LMO1 and Notch1 reprogram thymocytes into self-renewing cells. PLoS Genetics, 10(12), e1004768.

Gerby, B., Tremblay, C. S., Tremblay, M., Rojas-Sutterlin, S., Herblot, S., Hébert, J., et al. (2014c). SCL, LMO1 and Notch1 reprogram thymocytes into self-renewing cells. PLoS Genet, 10(12), e1004768.

Gering, M., Rodaway, A. R., Gottgens, B., Patient, R. K., & Green, A. R. (1998). The SCL gene specifies haemangioblast development from early mesoderm. The EMBO Journal, 17(14), 4029-4045.

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: A platform for interactive large-scale genome analysis. Genome Research, 15(10), 1451-1455.

Girardi, T., Vicente, C., Cools, J., & De Keersmaecker, K. (2017). The genetics and molecular biology of T-ALL. Blood, 129(9), 1113-1123.

Giroux, S., Kaushik, A., Capron, C., Jalil, A., Kelaidi, C., Sablitzky, F., et al. (2007). Lyl-1 and tal-1/scl, two genes encoding closely related bHLH transcription factors, display highly overlapping expression patterns during cardiovascular and hematopoietic ontogeny. Gene Expression Patterns, 7(3), 215-226.

Goardon, N., Schuh, A., Hajar, I., Ma, X., Jouault, H., Dzierzak, E., et al. (2002). Ectopic expression of TAL-1 protein in ly-6E.1-htal-1 transgenic mice induces defects in B- and T-lymphoid differentiation. Blood, 100(2), 491-500.

Goecks, J., Nekrutenko, A., Taylor, J., & Galaxy Team. (2010). Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology, 11(8), R86.

Grutz, G. G., Bucher, K., Lavenir, I., Larson, T., Larson, R., & Rabbitts, T. H. (1998). The oncogenic T cell LIM-protein Lmo2 forms part of a DNA-binding complex specifically in immature T cells. The EMBO Journal, 17(16), 4594-4605.

Gungor, C., Taniguchi-Ishigaki, N., Ma, H., Drung, A., Tursun, B., Ostendorff, H. P., et al. (2007). Proteasomal selection of multiprotein complexes recruited by LIM homeodomain transcription factors. Proceedings of the National Academy of Sciences of the United States of America, 104(38), 15000-15005.

Haddad, R., Guardiola, P., Izac, B., Thibault, C., Radich, J., Delezoide, A. L., et al. (2004). Molecular characterization of early human T/NK and B-lymphoid progenitor cells in umbilical cord blood. Blood, 104(13), 3918-3926.

Haider, Z., Larsson, P., Landfors, M., Köhn, L., Schmiegelow, K., Flægstad, T., et al. (2019). An integrated transcriptome analysis in t-cell acute lymphoblastic leukemia links DNA methylation subgroups to dysregulated TAL1 and ANTP homeobox gene expression. Cancer Medicine, 8(1), 311-324.

Hall, M. A., Curtis, D. J., Metcalf, D., Elefanty, A. G., Sourris, K., Robb, L., et al. (2003). The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. Proceedings of the National Academy of Sciences of the United States of America, 100(3), 992-997.

Heinzel, K., Benz, C., Martins, V. C., Haidl, I. D., & Bleul, C. C. (2007). Bone marrow-derived hemopoietic precursors commit to the T cell lineage only after arrival in the thymic microenvironment. Journal of Immunology (Baltimore, Md.: 1950), 178(2), 858-868.

Herblot, S., Aplan, P. D., & Hoang, T. (2002). Gradient of E2A activity in B-cell development. Molecular and Cellular Biology, 22(3), 886-900. doi:10.1128/MCB.22.3.886-900.2002

Herblot, S., Steff, A., Hugo, P., Aplan, P. D., & Hoang, T. (2000). SCL and LMO1 alter thymocyte differentiation: Inhibition of E2A-HEB function and pre-tα chain expression. Nature Immunology, 1(2), 138.

Hoang, T., Lambert, J., & Martin, R. (2016a). SCL/TAL1 in hematopoiesis and cellular reprogramming. Current topics in developmental biology (pp. 163-204) Elsevier.

Hoang, T., Lambert, J., & Martin, R. (2016b). SCL/TAL1 in hematopoiesis and cellular reprogramming. Current topics in developmental biology (pp. 163-204) Elsevier.

Howe, E. A., Sinha, R., Schlauch, D., & Quackenbush, J. (2011). RNA-seq analysis in MeV. Bioinformatics, 27(22), 3209-3210.

Hsu, H. L., Cheng, J. T., Chen, Q., & Baer, R. (1991). Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins. Molecular and Cellular Biology, 11(6), 3037-3042.

Hsu, H. L., Huang, L., Tsan, J. T., Funk, W., Wright, W. E., Hu, J. S., et al. (1994). Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. Molecular and Cellular Biology, 14(2), 1256-1265.

Huang, S., & Brandt, S. J. (2000). mSin3A regulates murine erythroleukemia cell differentiation through association with the TAL1 (or SCL) transcription factor. Molecular and Cellular Biology, 20(6), 2248-2259.

Huang, S., Qiu, Y., Shi, Y., Xu, Z., & Brandt, S. J. (2000). P/CAF-mediated acetylation regulates the function of the basic helix–loop–helix transcription factor TAL1/SCL. The EMBO Journal, 19(24), 6792-6803.

Huang, S., Qiu, Y., Stein, R. W., & Brandt, S. J. (1999). p300 functions as a transcriptional coactivator for the TAL1/SCL oncoprotein. Oncogene, 18(35), 4958.

Istrail, S., & Davidson, E. H. (2005). Logic functions of the genomic cis-regulatory code. Proceedings of the National Academy of Sciences of the United States of America, 102(14), 4954-4959.

Jafar-Nejad, H., Tien, A. C., Acar, M., & Bellen, H. J. (2006). Senseless and daughterless confer neuronal identity to epithelial cells in the drosophila wing margin. Development (Cambridge, England), 133(9), 1683-1692.

Jain, N., Lamb, A. V., O'Brien, S., Ravandi, F., Konopleva, M., Jabbour, E., et al. (2016). Early T-cell precursor acute lymphoblastic leukemia/lymphoma (ETP-ALL/LBL) in adolescents and adults: A high-risk subtype. Blood, 127(15), 1863-1869.

Ji, H., Ehrlich, L. I., Seita, J., Murakami, P., Doi, A., Lindau, P., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature, 467(7313), 338-342.

Jin, S., Su, H., Tran, N., Song, J., Lu, S. S., Li, Y., et al. (2017a). Splicing factor SF3B1K700E mutant dysregulates erythroid differentiation via aberrant alternative splicing of transcription factor TAL1. PloS One, 12(5), e0175523.

Jin, S., Su, H., Tran, N., Song, J., Lu, S. S., Li, Y., et al. (2017b). Splicing factor SF3B1K700E mutant dysregulates erythroid differentiation via aberrant alternative splicing of transcription factor TAL1. PloS One, 12(5), e0175523.

Jin, S., Tran, N., 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.

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(6), 860-870.

Kallianpur, A. R., Jordan, J. E., & Brandt, S. J. (1994a). The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis. Blood, 83(5), 1200-1208.

Kallianpur, A. R., Jordan, J. E., & Brandt, S. J. (1994b). The SCL/TAL-1 gene is expressed in progenitors of both the hematopoietic and vascular systems during embryogenesis. Blood, 83(5), 1200-1208.

Kassouf, M. T., Chagraoui, H., Vyas, P., & Porcher, C. (2008). Differential use of SCL/TAL-1 DNA-binding domain in developmental hematopoiesis. Blood, 112(4), 1056-1067.

Kassouf, M. T., Hughes, J. R., Taylor, S., McGowan, S. J., Soneji, S., Green, A. L., et al. (2010). Genome-wide identification of TAL1's functional targets: Insights into its mechanisms of action in primary erythroid cells. Genome Research, 20(8), 1064-1083.

Kee, B. L. (2009). E and ID proteins branch out. Nature Reviews.Immunology, 9(3), 175-184.

Kelliher, M. A., Seldin, D., & Leder, P. (1996). Tal-1 induces T cell acute lymphoblastic leukemia accelerated by casein kinase IIalpha. The EMBO Journal, 15(19), 5160-5166.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. Genome Research, 12(6), 996-1006.

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. Nature Methods, 12(4), 357.

Klein, T., & Campos-Ortega, J. A. (1997). Klumpfuss, a drosophila gene encoding a member of the EGR family of transcription factors, is involved in bristle and leg development. Development (Cambridge, England), 124(16), 3123-3134.

Koch, U., & Radtke, F. (2011). Mechanisms of T cell development and transformation. Annual Review of Cell and Developmental Biology, 27, 539-562.

Kopan, R., & Ilagan, M. X. G. (2009). The canonical notch signaling pathway: Unfolding the activation mechanism. Cell, 137(2), 216-233.

Krivega, I., Dale, R. K., & Dean, A. (2014). Role of LDB1 in the transition from chromatin looping to transcription activation. Genes & Development, 28(12), 1278-1290.

Kunisato, A., Chiba, S., Saito, T., Kumano, K., Nakagami-Yamaguchi, E., Yamaguchi, T., et al. (2004). Stem cell leukemia protein directs hematopoietic stem cell fate. Blood, 103(9), 3336-3341.

Kusy, S., Gerby, B., Goardon, N., Gault, N., Ferri, F., Gerard, D., et al. (2010). NKX3.1 is a direct TAL1 target gene that mediates proliferation of TAL1-expressing human T cell acute lymphoblastic leukemia. The Journal of Experimental Medicine, 207(10), 2141-2156.

Lacombe, J., Herblot, S., Rojas-Sutterlin, S., Haman, A., Barakat, S., Iscove, N. N., et al. (2010). Scl regulates the quiescence and the long-term competence of hematopoietic stem cells. Blood, 115(4), 792-803.

Lacombe, J., Herblot, S., Rojas-Sutterlin, S., Haman, A., Barakat, S., Iscove, N. N., et al. (2010). Scl regulates the quiescence and the long-term competence of hematopoietic stem cells. Blood, 115(4), 792-803.

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., et al. (2018). The human transcription factors. Cell, 172(4), 650-665.

Larmonie, N. S., Dik, W. A., Meijerink, J. P., Homminga, I., van Dongen, J. J., & Langerak, A. W. (2013). Breakpoint sites disclose the role of the V(D)J recombination machinery in the formation of T-cell receptor (TCR) and non-TCR associated aberrations in T-cell acute lymphoblastic leukemia. Haematologica, 98(8), 1173-1184.

Larson, R., Lavenir, I., Larson, T., Baer, R., Warren, A., Wadman, I., et al. (1996). Protein dimerization between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. The EMBO Journal, 15(5), 1021-1027.

Laurenti, E., & Göttgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. Nature, 553(7689), 418.

Lecuyer, E., Herblot, S., Saint-Denis, M., Martin, R., Begley, C. G., Porcher, C., et al. (2002). The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. Blood, 100(7), 2430-2440.

Lecuyer, E., Lariviere, S., Sincennes, M. C., Haman, A., Lahlil, R., Todorova, M., et al. (2007). Protein stability and transcription factor complex assembly determined by the SCL-LMO2 interaction. The Journal of Biological Chemistry, 282(46), 33649-33658.

Lécuyer, E., & Hoang, T. (2004). SCL: From the origin of hematopoiesis to stem cells and leukemia. Experimental Hematology, 32(1), 11-24.

Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. Cell, 152(6), 1237-1251.

Lerdrup, M., Johansen, J. V., Agrawal-Singh, S., & Hansen, K. (2016). An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nature Structural & Molecular Biology, 23(4), 349.

Li, Y., Brauer, P. M., Singh, J., Xhiku, S., Yoganathan, K., Zúñiga-Pflücker, J. C., et al. (2017). Targeted disruption of TCF12 reveals HEB as essential in human mesodermal specification and hematopoiesis. Stem Cell Reports, 9(3), 779-795.

Ling, F., Kang, B., & Sun, X. (2014). Id proteins: Small molecules, mighty regulators. Current topics in developmental biology (pp. 189-216) Elsevier.

Machanick, P., & Bailey, T. L. (2011). MEME-ChIP: Motif analysis of large DNA datasets. Bioinformatics, 27(12), 1696-1697.

Maillard, I., Fang, T., & Pear, W. S. (2005). Regulation of lymphoid development, differentiation, and function by the notch pathway. Annu.Rev.Immunol., 23, 945-974.

Mansour, M. R., Abraham, B. J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A. D., et al. (2014). Oncogene regulation. an oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science (New York, N.Y.), 346(6215), 1373-1377.

Massari, M. E., & Murre, C. (2000). Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. Molecular and Cellular Biology, 20(2), 429-440.

Matthews, J. M., Lester, K., Joseph, S., & Curtis, D. J. (2013). LIM-domain-only proteins in cancer. Nature Reviews Cancer, 13(2), 111-122.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., et al. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature Biotechnology, 28(5), 495.

Mellentin, J. D., Smith, S. D., & Cleary, M. L. (1989). Lyl-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell, 58(1), 77-83.

Mikkola, H. K., Klintman, J., Yang, H., Hock, H., Schlaeger, T. M., Fujiwara, Y., et al. (2003). Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. Nature, 421(6922), 547-551.

Moore, T. A., & Zlotnik, A. (1997). Differential effects of flk-2/flt-3 ligand and stem cell factor on murine thymic progenitor cells. Journal of Immunology (Baltimore, Md.: 1950), 158(9), 4187-4192.

Möröy, T., Vassen, L., Wilkes, B., & Khandanpour, C. (2015). From cytopenia to leukemia: The role of Gfi1 and Gfi1b in blood formation. Blood, the Journal of the American Society of Hematology, 126(24), 2561-2569.

Mundade, R., Ozer, H. G., Wei, H., Prabhu, L., & Lu, T. (2014). Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. Cell Cycle, 13(18), 2847-2852.

O'Neil, J., Billa, M., Oikemus, S., & Kelliher, M. (2001). The DNA binding activity of TAL-1 is not required to induce leukemia/lymphoma in mice. Oncogene, 20(29), 3897.

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(6), 587-596.

Ono, Y., Fukuhara, N., & Yoshie, O. (1997). Transcriptional activity of TAL1 in T cell acute lymphoblastic leukemia (T-ALL) requires RBTN1 or -2 and induces TALLA1, a highly specific tumor marker of T-ALL. The Journal of Biological Chemistry, 272(7), 4576-4581.

Ono, Y., Fukuhara, N., & Yoshie, O. (1998). TAL1 and LIM-only proteins synergistically induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by acting as cofactors for GATA3. Molecular and Cellular Biology, 18(12), 6939-6950.

Palamarchuk, A., Efanov, A., Maximov, V., Aqeilan, R. I., Croce, C. M., & Pekarsky, Y. (2005). Akt phosphorylates Tal1 oncoprotein and inhibits its repressor activity. Cancer Research, 65(11), 4515-4519.

Palii, C. G., Perez-Iratxeta, C., Yao, Z., Cao, Y., Dai, F., Davison, J., et al. (2011a). Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. The EMBO Journal, 30(3), 494-509.

Palii, C. G., Perez-Iratxeta, C., Yao, Z., Cao, Y., Dai, F., Davison, J., et al. (2011b). Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. The EMBO Journal, 30(3), 494-509.

Palomero, T., Odom, D. T., O'Neil, J., Ferrando, A. A., Margolin, A., Neuberg, D. S., et al. (2006). Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. Blood, 108(3), 986-992.

Petrie, H. T. (2003). Cell migration and the control of post-natal T-cell lymphopoiesis in the thymus. Nature Reviews Immunology, 3(11), 859.

Petrie, H. T., & Zúñiga-Pflücker, J. C. (2007). Zoned out: Functional mapping of stromal signaling microenvironments in the thymus. Annu.Rev.Immunol., 25, 649-679.

Pimkin, M., Kossenkov, A. V., Mishra, T., Morrissey, C. S., Wu, W., Keller, C. A., et al. (2014). Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. Genome Research, 24(12), 1932-1944.

Piper, J., Elze, M. C., Cauchy, P., Cockerill, P. N., Bonifer, C., & Ott, S. (2013). Wellington: A novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Research, 41(21), e201.

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F. W., & Orkin, S. H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell, 86(1), 47-57.

Prasad, K. S., & Brandt, S. J. (1997). Target-dependent effect of phosphorylation on the DNA binding activity of the TAL1/SCL oncoprotein. The Journal of Biological Chemistry, 272(17), 11457-11462.

Presnell, J. S., Schnitzler, C. E., & Browne, W. E. (2015). KLF/SP transcription factor family evolution: Expansion, diversification, and innovation in eukaryotes. Genome Biology and Evolution, 7(8), 2289-2309.

Pui, C., & Evans, W. E. (2006). Treatment of acute lymphoblastic leukemia. New England Journal of Medicine, 354(2), 166-178.

Radtke, F., Fasnacht, N., & MacDonald, H. R. (2010). Notch signaling in the immune system. Immunity, 32(1), 14-27.

Ravasi, T., Suzuki, H., Cannistraci, C. V., Katayama, S., Bajic, V. B., Tan, K., et al. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. Cell, 140(5), 744-752.

Ravet, E., Reynaud, D., Titeux, M., Izac, B., Fichelson, S., Romeo, P. H., et al. (2004). Characterization of DNA-binding-dependent and -independent functions of SCL/TAL1 during human erythropoiesis. Blood, 103(9), 3326-3335.

Res, P., Martinez-Caceres, E., Cristina Jaleco, A., Staal, F., Noteboom, E., Weijer, K., et al. (1996). CD34+CD38dim cells in the human thymus can differentiate into T, natural killer, and dendritic cells but are distinct from pluripotent stem cells. Blood, 87(12), 5196-5206.

Reynaud, D., Ravet, E., Titeux, M., Mazurier, F., Renia, L., Dubart-Kupperschmitt, A., et al. (2005). SCL/TAL1 expression level regulates human hematopoietic stem cell self-renewal and engraftment. Blood, 106(7), 2318-2328.

Robb, L., Elwood, N., Elefanty, A., Köntgen, F., Li, R., Barnett, L., et al. (1996). The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. The EMBO Journal, 15(16), 4123-4129.

Robb, L., Lyons, I., Li, R., Hartley, L., Kontgen, F., Harvey, R. P., et al. (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. Proceedings of the National Academy of Sciences of the United States of America, 92(15), 7075-7079.

Robb, L., Rasko, J. E., Bath, M. L., Strasser, A., & Begley, C. G. (1995). Scl, a gene frequently activated in human T cell leukaemia, does not induce lymphomas in transgenic mice. Oncogene, 10(1), 205-209.

ROBERTS, W., WIERSMA, S., URIBE, L., & WEINBERG, K. (1992). The arr cell-line-a model of acute leukemias expressing both t-lymphoid and myeloid-associated genes. Clinical Research, , 40. (1) pp. A30.

Rothenberg, E. V., Moore, J. E., & Yui, M. A. (2008). Launching the T-cell-lineage developmental programme. Nature Reviews Immunology, 8(1), 9-21.

Rothenberg, E. V., Ungerbäck, J., & Champhekar, A. (2016). Chapter four-forging T-lymphocyte identity: Intersecting networks of transcriptional control. Advances in Immunology, 129, 109-174.

Sanda, T., Lawton, L. N., Barrasa, M. I., Fan, Z. P., Kohlhammer, H., Gutierrez, A., et al. (2012a). Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer Cell, 22(2), 209-221.

Sanda, T., Lawton, L. N., Barrasa, M. I., Fan, Z. P., Kohlhammer, H., Gutierrez, A., et al. (2012b). Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer Cell, 22(2), 209-221.

Saran, N., Lyszkiewicz, M., Pommerencke, J., Witzlau, K., Vakilzadeh, R., Ballmaier, M., et al. (2010). Multiple extrathymic precursors contribute to T-cell development with different kinetics. Blood, 115(6), 1137-1144.

Sawada, S., & Littman, D. R. (1993a). A heterodimer of HEB and an E12-related protein interacts with the CD4 enhancer and regulates its activity in T-cell lines. Molecular and Cellular Biology, 13(9), 5620-5628.

Sawada, S., & Littman, D. R. (1993b). A heterodimer of HEB and an E12-related protein interacts with the CD4 enhancer and regulates its activity in T-cell lines. Molecular and Cellular Biology, 13(9), 5620-5628.

Schwartz, R., Engel, I., Fallahi-Sichani, M., Petrie, H. T., & Murre, C. (2006). 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 of the United States of America, 103(26), 9976-9981.

Selkoe, D., & Kopan, R. (2003). Notch and presenilin: Regulated intramembrane proteolysis links development and degeneration. Annual Review of Neuroscience, 26, 565-597.

Semerad, C. L., Mercer, E. M., Inlay, M. A., Weissman, I. L., & Murre, C. (2009). E2A proteins maintain the hematopoietic stem cell pool and promote the maturation of myelolymphoid and myeloerythroid progenitors. Proceedings of the National Academy of Sciences of the United States of America, 106(6), 1930-1935.

Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44-57.

Shivdasani, R. A., Mayer, E. L., & Orkin, S. H. (1995). Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature, 373(6513), 432.

Sincennes, M. C., Humbert, M., Grondin, B., Lisi, V., Veiga, D. F., Haman, A., et al. (2016). The LMO2 oncogene regulates DNA replication in hematopoietic cells. Proceedings of the National Academy of Sciences of the United States of America, 113(5), 1393-1398.

Souroullas, G. P., Salmon, J. M., Sablitzky, F., Curtis, D. J., & Goodell, M. A. (2009). Adult hematopoietic stem and progenitor cells require either Lyl1 or scl for survival. Cell Stem Cell, 4(2), 180-186.

Spits, H. (2002). Development of alphabeta T cells in the human thymus. Nature Reviews.Immunology, 2(10), 760-772.

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(17), 9874-9888.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2018). STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research, 47(D1), D607-D613.

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(1), 53-64.

Takai, H., Masuda, K., Sato, T., Sakaguchi, Y., Suzuki, T., Suzuki, T., et al. (2014). 5-hydroxymethylcytosine plays a critical role in glioblastomagenesis by recruiting the CHTOP-methylosome complex. Cell Reports, 9(1), 48-60.

Takeuchi, A., Yamasaki, S., Takase, K., Nakatsu, F., Arase, H., Onodera, M., et al. (2001). E2A and HEB activate the pre-TCR alpha promoter during immature T cell development. Journal of Immunology (Baltimore, Md.: 1950), 167(4), 2157-2163.

Talora, C., Cialfi, S., Oliviero, C., Palermo, R., Pascucci, M., Frati, L., et al. (2006). Cross talk among Notch3, pre-TCR, and Tal1 in T-cell development and leukemogenesis. Blood, 107(8), 3313-3320.

Tang, T., Arbiser, J. L., & Brandt, S. J. (2002). Phosphorylation by mitogen-activated protein kinase mediates the hypoxia-induced turnover of the TAL1/SCL transcription factor in endothelial cells. The Journal of Biological Chemistry, 277(21), 18365-18372.

Tatarek, J., Cullion, K., Ashworth, T., Gerstein, R., Aster, J. C., & Kelliher, M. A. (2011a). Notch1 inhibition targets the leukemia-initiating cells in a Tal1/Lmo2 mouse model of T-ALL. Blood, 118(6), 1579-1590.

Tatarek, J., Cullion, K., Ashworth, T., Gerstein, R., Aster, J. C., & Kelliher, M. A. (2011b). Notch1 inhibition targets the leukemia-initiating cells in a Tal1/Lmo2 mouse model of T-ALL. Blood, 118(6), 1579-1590.

Teitell, M. A., & Pandolfi, P. P. (2009). Molecular genetics of acute lymphoblastic leukemia. Annual Review of Pathological Mechanical Disease, 4, 175-198.

Thompson, B. J., Buonamici, S., Sulis, M. L., Palomero, T., Vilimas, T., Basso, G., et al. (2007). The SCFFBW7 ubiquitin ligase complex as a tumor suppressor in T cell leukemia. The Journal of Experimental Medicine, 204(8), 1825-1835.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nature Protocols, 7(3), 562.

Tremblay, M., Herblot, S., Lecuyer, E., & Hoang, T. (2003). Regulation of pT alpha gene expression by a dosage of E2A, HEB, and SCL. The Journal of Biological Chemistry, 278(15), 12680-12687.

Tremblay, M., Tremblay, C. S., Herblot, S., Aplan, P. D., Hebert, J., Perreault, C., et al. (2010). Modeling T-cell acute lymphoblastic leukemia induced by the SCL and LMO1 oncogenes. Genes & Development, 24(11), 1093-1105.

Visvader, J., Begley, C. G., & Adams, J. M. (1991). Differential expression of the LYL, SCL and E2A helix-loop-helix genes within the hemopoietic system. Oncogene, 6(2), 187-194.

Vora, A., Goulden, N., Wade, R., Mitchell, C., Hancock, J., Hough, R., et al. (2013). Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): A randomised controlled trial. The Lancet Oncology, 14(3), 199-209.

Wadman, I. A., Hsu, H. L., Cobb, M. H., & Baer, R. (1994). The MAP kinase phosphorylation site of TAL1 occurs within a transcriptional activation domain. Oncogene, 9(12), 3713-3716.

Wadman, I., Li, J., Bash, R. O., Forster, A., Osada, H., Rabbitts, T. H., et al. (1994). Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia. The EMBO Journal, 13(20), 4831-4839.

Wang, D., Claus, C. L., Vaccarelli, G., Braunstein, M., Schmitt, T. M., Zuniga-Pflucker, J. C., et al. (2006). The basic helix-loop-helix transcription factor HEBAlt is expressed in pro-T cells and enhances the generation of T cell precursors. Journal of Immunology (Baltimore, Md.: 1950), 177(1), 109-119.

Watson, J. D. (2004). Molecular biology of the gene Pearson Education India.

Weng, A. P., Ferrando, A. A., Lee, W., Morris, J. P., Silverman, L. B., Sanchez-Irizarry, C., et al. (2004). Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science (New York, N.Y.), 306(5694), 269-271.

Wikström, I., Forssell, J., Penha-Goncalves, M. N., Bergqvist, I., & Holmberg, D. (2008). A role for E2-2 at the DN3 stage of early thymopoiesis. Molecular Immunology, 45(11), 3302-3311.

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., et al. (2010a). Combinatorial transcriptional control in blood stem/progenitor cells: Genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell, 7(4), 532-544.

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., et al. (2010b). Combinatorial transcriptional control in blood stem/progenitor cells: Genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell, 7(4), 532-544.

Wojciechowski, J., Lai, A., Kondo, M., & Zhuang, Y. (2007a). E2A and HEB are required to block thymocyte proliferation prior to pre-TCR expression. Journal of Immunology (Baltimore, Md.: 1950), 178(9), 5717-5726.

Wojciechowski, J., Lai, A., Kondo, M., & Zhuang, Y. (2007b). E2A and HEB are required to block thymocyte proliferation prior to pre-TCR expression. Journal of Immunology (Baltimore, Md.: 1950), 178(9), 5717-5726.

Wu, W., Morrissey, C. S., Keller, C. A., Mishra, T., Pimkin, M., Blobel, G. A., et al. (2014). Dynamic shifts in occupancy by TAL1 are guided by GATA factors and drive large-scale reprogramming of gene expression during hematopoiesis. Genome Research, 24(12), 1945-1962.

Xiong, J., Armato, M. A., & Yankee, T. M. (2010). Immature single-positive CD8 thymocytes represent the transition from notch-dependent to notch-independent T-cell development. International Immunology, 23(1), 55-64.

Xu, W., & Kee, B. L. (2007). Growth factor independent 1B (Gfi1b) is an E2A target gene that modulates Gata3 in T-cell lymphomas. Blood, 109(10), 4406-4414.

Xu, Z., Huang, S., Chang, L. S., Agulnick, A. D., & Brandt, S. J. (2003). Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation. Molecular and Cellular Biology, 23(21), 7585-7599.

Xu, Z., Meng, X., Cai, Y., Koury, M. J., & Brandt, S. J. (2006). Recruitment of the SWI/SNF protein Brg1 by a multiprotein complex effects transcriptional repression in murine erythroid progenitors. The Biochemical Journal, 399(2), 297-304.

Yan, W., Young, A. Z., Soares, V. C., Kelley, R., Benezra, R., & Zhuang, Y. (1997). High incidence of T-cell tumors in E2A-null mice and E2A/Id1 double-knockout mice. Molecular and Cellular Biology, 17(12), 7317-7327.

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. The Journal of Experimental Medicine, 197(7), 831-844.

Yui, M. A., & Rothenberg, E. V. (2014). Developmental gene networks: A triathlon on the course to T cell identity. Nature Reviews Immunology, 14(8), 529-545.

Zhang, L., Tran, N., Su, H., Wang, R., Lu, Y., Tang, H., et al. (2015). Cross-talk between PRMT1-mediated methylation and ubiquitylation on RBM15 controls RNA splicing. Elife, 4, e07938.

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(48), 6937.

Zhou, C., Yang, Y., & Jong, A. Y. (1990). Mini-prep in ten minutes. BioTechniques, 8(2), 172-173.

Zhu, H., Wang, G., & Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. Nature Reviews Genetics, 17(9), 551.