Investigations of the gene regulatory protein, ZBTB7A, and its role in red blood cells

Author:
Norton, Laura

Publication Date:
2015

DOI:
https://doi.org/10.26190/unsworks/2822

Investigations of the gene regulatory protein, ZBTB7A, and its role in red blood cells

Laura Norton

A thesis submitted for the degree of Doctor in Philosophy

(Biochemistry and Molecular Genetics)

School of Biotechnology and Biomolecular Sciences

The University of New South Wales

March 2015

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **Norton**

First name: **Laura** Other name/s:

Abbreviation for degree as given in the University calendar: **PhD**

School: **School of Biotechnology and Biomolecular Sciences** Faculty: **Science**

Title: **Investigations of the gene regulatory protein, ZBTB7A, and its role in red blood cells**

Abstract 350 words maximum: (PLEASE TYPE)

Haemoglobin, the oxygen carrying protein in erythrocytes, is composed of two alpha- and two beta-like globin subunits. Its composition changes during gestation to meet varying oxygen demands of the developing embryo. A developmental switch occurs at birth, where foetal *γ-globin* expression is replaced by adult *β-globin*. Mutations in the *γ-globin* promoter can impair the switch and thereby cause elevated levels of foetal haemoglobin in adulthood, known as hereditary persistence of foetal haemoglobin (HPFH). This condition alleviates symptoms of adult *β-globin* disorders such as *β*-thalassemia and sickle cell disease. Reactivation of *γ-globin* expression currently offers a promising therapeutic approach for these diseases.

A group of HPFH mutations clustered approximately 200 basepairs upstream of the *γ-globin* gene (-200 site) are believed to disrupt binding of transcriptional repressors of *γ-globin* expression. We have previously demonstrated that ZBTB7A, a transcriptional repressor highly expressed in erythroid cells and required for normal erythroid development, is able to bind this site *in vitro*, and binding is abolished by mutations in the -200 site. The purpose of this study was to further analyse the binding of ZBTB7A *in vivo*, as well as determine how it is regulated in erythroid cells.

Here we use ChIP-Seq to demonstrate that ZBTB7A binds the *γ-globin* -200 site in cultured cells. Previous results suggested that *Zbtb7a* is regulated by the master regulator of erythropoiesis, KLF1, however, whether this effect was direct or mediated through other KLF1 target genes was unknown. We use a *Klf1*$^{-/-}$ mouse model, and the KLF1 inducible cell line, K1ER, to show that *Zbtb7a* is directly activated by KLF1. We also find that *ZBTB7A* has a novel transcription start site in both humans and mice that is highly expressed in erythroid cells and is induced upon erythroid cell differentiation.

These findings suggest ZBTB7A plays a role in haemoglobin switching. As ZBTB7A is an important factor in erythroid maturation and survival, completely abolishing its expression may be detrimental. However, it may be possible to target the erythroid-specific transcript to reduce *ZBTB7A* expression in erythroid cells alone, which may be sufficient to boost foetal haemoglobin as a potential therapeutic for *β*-haemoglobinopathies.

FOR OFFICE USE ONLY Date of completion of requirements for Award:

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

Originality Statement

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed 

Date 31/03/15

Table of Contents

Acknowledgements

Firstly, I would like to thank my supervisor, Merlin, for the opportunity to complete my PhD in the Crossley lab. I have thoroughly enjoyed my time in your lab. You have been a wonderful mentor, and have instilled in me a deep love of science, and for this you have my sincerest thanks.

Thank you also to Ali, who has been an incredible inspiration and friend to me over the years. Without your support and guidance, it is uncertain if this thesis would have ever been completed. I'm glad you kindly waited till near the end to move to the US!

Also, a big thanks must go to Richard for his supervision throughout my PhD, and his constant reminder to "manage my expectations". Your dry wit and quick quips will not soon be forgotten. And also to Kate, you may have come into my project near its completion, but you have been a tremendous help in proof reading, and managing stress levels over the past year. A special thank you also to our past lab manager Bev and current manager Dixie. You are both incredibly kind and patient, for which I am especially grateful.

Thanks also to the other members of the Crossley lab, past and present. Cassie and Crisbel (the infamous Casbel), for answering my endless questions and sharing protocols. Jon, who, as well as helping with the bioinformatics in this thesis, for always playing devil's advocate and providing a counter argument to…well, everything. Becky, for lamenting the complexities of transcription factor binding with me over (the occasional) beer/coffee, and always putting things into perspective! And to Cath, Gabby, Alex, Jasmin, Lu and Tanit, without whom my experience in the Crossley lab would have been severely lacking. You have all been great sounding boards for ideas and solutions to the everyday problems faced in the lab. Thank you for the laughs and the amazing cakes shared over the years!

Lastly, of course, I have to thank my family. Mom and Dad, where do I even begin? Your constant love and support has undeniably been instrumental in finishing this project. Even though you don't quite understand what it is that I have been doing over the last four years, I know you are proud - I mean, you tell practically everyone you meet! Thanks for always believing in me and being the best parents anyone could hope for.

And the final thanks to my husband, Ben. While I would not necessarily recommend getting married during one's PhD, I am glad we did! Thank you so much for trying to understand the ins and outs of my project (and succeeding pretty damn well) and for your all your love and support over the duration of this seemingly never-ending ride. You have had to deal with the emotional side of this process, which I do not wish upon anyone. Getting here hasn't been easy, but I know you will agree with me when I say it has been well worth it!

Publications arising from this candidature

Journal Articles

Funnell APW*, **Norton LJ*,** Mak KS, Burdach JG, Artuz CM, Twine NA, Wilkins MR, Power CA, Hung TT, Perdomo J, Koh P, Bell-Anderson KS, Orkin SH, Fraser ST, Perkins AC, Pearson RC, Crossley M (2012). The CACCC-binding protein KLF3/BKLF represses a subset of KLF1/EKLF target genes and is required for proper erythroid maturation in vivo. *Molecular and Cellular Biology*, 32 (16) 3281-3292. (*Joint first authors)

Norton LJ, Funnell AP, Pearson RC, Crossley M (2011). Cellular Reprogramming toward the Erythroid Lineage. *International Journal of Cell Biology*, Article ID 501464.

Conference abstracts

Norton LJ, Funnell APW, Martyn, GE, Pearson, RCM, Quinlan, KGR, Crossley, M (February 2015). The role of novel transcription factors in globin switching. *Poster presented at the 36th Lorne Genome Conference, Lorne, VIC.*

Norton LJ, Mak K, Funnell APW, Burdach JG, Twine NA, Pearson RCM, Crossley M (February 2014). Regulatory networks in the Krüppel-like factor family. *Abstract presented at the Asian Conference on Transcription, Melbourne, VIC.*

Norton LJ, Funnell APW, Mak K, Burdach JG, Twine NA, Pelka GJ, Radziewic T, Power M, Tam PP, Pearson RCM, Crossley M (February 2014). Krüppel-like factor networks in erythropoiesis. *Abstract presented at Lorne Genome Conference, Lorne, VIC.*

Norton LJ, Funnell APW, Mak K, Burdach JG, Twine NA, Pelka GJ, Radziewic T, Power M, Tam PP, Pearson RCM, Crossley M (November 2013). Investigating the erythroid roles of the transcription factors Klf3 and Klf8. *Poster presented at the BABS Research Symposium 2013, Sydney, NSW.*

Norton LJ, Funnell APW, Mak K, Burdach JG, Twine NA, Pelka GJ, Radziewic T, Power M, Tam PP, Pearson RCM, Crossley M (September 2013). Investigating the erythroid roles of the transcription factors Klf3 and Klf8. *Poster presented at the EMBO Meeting 2013, Amsterdam, NE.*

Norton LJ, Pearson RCM, Crossley M (February 2013). A potential role for CtBP in human foetal globin. *Poster Presented at the Lorne Genome Conference, Lorne, VIC.*

Norton LJ, Funnell APW, Mak K, Pearson RCM, Crossley M (July 2012). Klf3/BKlf represses a subset of Klf1/EKlf target genes to ensure normal erythropoiesis. *Poster presented at the Hemoglobin Switching Conference, Monterey, USA.*

Norton LJ, Funnell APW, Mak K, Pearson RCM, Crossley M (June 2012). Klf3/BKlf represses a subset of Klf1/EKlf target genes to ensure normal erythropoiesis. *Poster presented at ASMR's 20th NSW Scientific Meeting, Sydney, NSW.*

Norton LJ, Mak K, Funnell APW, Pearson, RCM, Crossley M (February 2012). The Role of Klf3 in Haematopoiesis. *Poster presented at Lorne Genome Conference, Lorne, VIC.*

Norton LJ, Mak K, Funnell APW, Pearson RCM, Crossley M (February 2011). Investigating the Role of Klf3 in Erythropoiesis at the Physiological and Molecular Level. *Poster Presented at the Lorne Genome Conference, Lorne, VIC.*

Abstract

Haemoglobin, the oxygen carrying protein in erythrocytes, is composed of two alpha- and two beta-like globin subunits. Its composition changes during gestation to meet varying oxygen demands of the developing embryo. A developmental switch occurs at birth, where foetal *γ-globin* expression is replaced by adult *β-globin*. Mutations in the *γ-globin* promoter can impair the switch and thereby cause elevated levels of foetal haemoglobin in adulthood, known as hereditary persistence of foetal haemoglobin (HPFH). This condition alleviates symptoms of adult *β-globin* disorders such as *β*-thalassemia and sickle cell disease. Reactivation of *γ-globin* expression currently offers a promising therapeutic approach for these diseases.

A group of HPFH mutations clustered approximately 200 basepairs upstream of the *γ-globin* gene (-200 site) are believed to disrupt binding of transcriptional repressors of *γ-globin* expression. We have previously demonstrated that ZBTB7A, a transcriptional repressor highly expressed in erythroid cells and required for normal erythroid development, is able to bind this site *in vitro*, and binding is abolished by mutations in the -200 site. The purpose of this study was to further analyse the binding of ZBTB7A *in vivo*, as well as determine how it is regulated in erythroid cells.

Here we use ChIP-Seq to demonstrate that ZBTB7A binds the *γ-globin* -200 site in cultured cells. Previous results suggested that *Zbtb7a* is regulated by the master regulator of erythropoiesis, KLF1, however, whether this effect was direct or mediated through other KLF1 target genes was unknown. We use a *Klf1*$^{-/-}$ mouse model, and the KLF1 inducible cell line, K1ER, to show that *Zbtb7a* is directly activated by KLF1. We also find that *ZBTB7A* has a novel transcription start site in both humans and mice that is highly expressed in erythroid cells and is induced upon erythroid cell differentiation.

These findings suggest ZBTB7A plays a role in haemoglobin switching. As ZBTB7A is an important factor in erythroid maturation and survival, completely abolishing its expression may be detrimental. However, it may be possible to target the erythroid-specific transcript to reduce *ZBTB7A* expression in erythroid cells alone, which may be sufficient to boost foetal haemoglobin as a potential therapeutic for *β*-haemoglobinopathies.

List of Abbreviations

α	alpha
aa	Amino acid
ADD2	Adducin 2
AGRF	Australian Genome Research Facility
AHSP	α-Haemoglobin stabilising protein
ALAD	δ-Aminolevulinate dehydratase
ANK1	Ankyrin 1, Erythrocytic
APS	Ammonium persulfate
ATP	Adenosine triphosphate
β	beta
BAC	Bacterial artificial chromosome
BCL11A	B cell lymphoma 11A
BFU	Burst-forming unit
BCL2L11	Bcl-2 Interacting Mediator of cell death
BLAST	Basic Local Alignment Search Tool
bp	Base pair(s)
BRE	TFIIB recognition element
BSA	Bovine serum albumin
Ca1	Carbonic anhydrase I
CBP	CREB-binding protein
CentriMo	Local Motif Enrichment Analysis
CD34	Cluster of differentiation 34 antigen
cDNA	Complementary DNA
CFU	Colony-forming unit
ChIP	Chromatin immunoprecipitation
ChIP-qPCR	Chromatin immunoprecipitation-quantitative polymerase chain reaction
ChIP-Seq	Chromatin immunoprecipitation Sequencing
CHX	Cycloheximide
CRISPR	Clustered regularly interspaced short palindromic repeats
CTBP	C-terminal binding protein
DAVID	The Database for Annotation, Visualization and Integrated Discovery

δ	delta
DBTSS	Database of Transcriptional Start Sites
DDBJ	DNA Data Bank of Japan
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethylsulfoxide
DNA	2'deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
dsDNA	Double stranded DNA
DPE	Downstream promoter element
DREME	Discriminative Regular Expression Motif Elicitation
DTT	Dithiothreitol
ε	epsilon
E10.5 (etc)	Mouse age, embryonic day 10.5, 10.5 days post coitum
EDTA	Ethylenediamine-tetraacetic acid disodium dihydrate
EGTA	Ethylene glycol-bis[2-aminoethylether]-N,N,N¢,N¢-tetraacetic acid
EMSA	Electrophoretic mobility shift assay
ENCODE	Encyclopedia of DNA elements
EPB4.2	Erythrocyte Membrane Protein Band 4.2
EPO	Erythropoietin
ER	Oestrogen receptor
E-RC1	Eklf coactivator-remodeling complex 1
ES	Embryonic stem cell(s)
EtOH	Ethanol
FANTOM5	Functional Annotation of the Mammalian Genome
FASN	Fatty acid synthase
FBI1	Factor that Binds to Inducer of Short Transcripts Protein 1
FCS	Foetal calf serum
G1E	Gata-1 erythroid
γ	gamma
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GATA-1	GATA-binding transcription factor 1
GENCODE	Encylopedia of genes and gene variants

GEO	Gene expression omnibus
GSH	Glutathione, reduced
GST	Glutathione S-transferase
GTF	General transcription factor
h	hour(s)
HAT	Histone acetyl transferase
HbA	Adult haemoglobin protein
HBA1	Human adult haemoglobin 1
HBA2	Human adult haemoglobin 2
Hba-a1	Mouse adult α-like globin, α1
Hba-a2	Mouse adult α-like globin, α2
Hba-x	Mouse embryonic α-like globin, ζ
HBB	Human adult β-globin
Hbb-b1	Mouse adult β-like globin, β-major
Hbb-b2	Mouse adult β-like globin, β-minor
Hbb-bh1	Mouse embryonic β-like globin, βh1
Hbb-y	Mouse emrbyonic β-like globin, εy
HBD	Human adult β-like globin, δ
HBE	Human embryonic β-like globin, ε
HbF	Foetal haemoglobin protein
HbF	Foetal haemoglobin
HBG1	Human foetal β-like globin, Aγ
HBG2	Human foetal β-like globin, Bγ
HbS	Sickle haemoglobin protein
HBZ	Human embryonic α-like globin, ζ
HEPES	N-[2-hydroxyethyl]piperazine-N¢-[2-ethanesulfonic acid]
HIPK	Homeodomain-interacting protein kinase
HMG	High Mobility Group box
HMT	Histone methyl transferase
HOMER	Hypergeometric optimisation of motif enrichment
HPFH	Hereditary persistence of foetal haemoglobin
HRP	Horseradish peroxidase
HS	Hypersensitive site

HSC	Haematopoietic stem cell
Hsp90	Heat shock protein 90
HUGO	Human genome organisation
IgG	Immunoglobulin G
IP	Immunoprecipitation
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
K	Lysine
K1ER	Klf1-oestrogen receptor cell line
kb	Kilobase pairs
kDa	Kilodalton(s)
KLF	Krüppel-like factor
KLF1	Erythroid Krüppel-like factor
KLF1-ER	Erythroid Krüppel-like factor-oestrogen receptor fusion protein
KLF2	Lung Krüppel-like factor
KLF3	Basic Krüppel-like factor
KLF11	Krüppel-like factor 11
KO	Knockout
LB	Luria-Bertani
LCR	Locus control region
LDB1	Lim domain binding 1
LMO2	Lim-only protein 2
LRF	Leukaemia/Lymphoma-Related Factor
LSD1	Lysine-specific demethylase 1A
MEL	Mouse erythroleukaemia cell line
MEME	Multiple Em for Motif Elicitation
MEP	Megakaryocytic-erythroid progenitor(s)
MFSD7B	Major facilitator superfamily domain containing 7B
MGI	Mouse genome informatics
min	minute(s)
MOPS	3-[N-Morpholino]propanesulfonic acid
MQW	Milli-Q water
mRNA	Messenger RNA
NCBI	National Centre for Biotechnology Information

NCoR	Nuclear Receptor Corepressor
Neo	Neomycin resistance gene
NK	Natural killer cell(s)
NTP	Nucleoside triphosphate
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBGD	Porphobilinogen deaminase
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PGK	Phosphoglycerate kinase
PI	Pre-immune serum
PIC	Pre-initiation complex
PMSF	Phenylmethylsulfonyl fluoride
POK	POZ and Krüppel type
POKEMON	POZ and Krüppel Erythroid Myeloid Ontogenic Factor
POZ	Pox virus and Zinc finger domain
PSG	Penicillin, streptomycin and glutamine solution
qPCR	Quantitative real time RT-PCR
R	Arginine
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
RNase A	Ribonuclease A
RPMI	Roswell Park Memorial Institute Media
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription-polymerase chain reaction
RUNX1	Runt-related transcription factor 1
SCD	Sickle cell disease
SCF	Stem cell factor
SCL/SCL/TAL1	Stem cell leukaemia transcription factor/T-Cell Acute Lymphocytic Leukaemia 1
SDS	Sodium dodecyl sulfate
SGPL1	Sphingosine-1-Phosphate Lyase 1
SLC25A38	Solute Carrier Family 25, Member 38
SOX6	SRY (Sex Determining Region Y)-Box 6

SREBP-1	Sterol regulatory element-binding broteins-1
SSTAR	Semantic catalogue of samples, transcription initiation and Regulators
SUMO	Small ubiquitin-like modifier
SWI/SNF	Switching mating type/sucrose non-fermenting
TAF	TBP-associated factors
Tamoxifen	4-hydroxytamoxifen
TBE	Tris-Borate-EDTA buffer
TBP	TATA-binding protein
TBST	Tris-buffered saline Tween™-20
TER-119	TER-119 erythroid antigen
TFII	General transcription factor of RNA polymerase II
TPO	Thrombopoietin
Trim10	Tripartite motif containing 10
TRIS	Tris-hydroxymethyl-methylamine
Triton X-100	T-octylphenoxypolyethoxyethanol
TSS	Transcription start site
TTS	Transcript termination site
Tween™-20	Polyoxyethylenesorbitanmonolaurate
Ubc9	Ubiquitin-conjugating enzyme 9
UCSC	University of California Santa Cruz
UROS	Uroporphyrinogen III synthase
UTR	Untranslated region
WT	Wild-type
YAC	Yeast artificial chromosome
ζ	zeta
ZBTB7A	Zinc finger and BTB domain containing 7A
ZBTB32	Zinc finger and BTB domain containing 32
Zn Fingers	Zinc-fingers

CHAPTER 1 – GENERAL INTRODUCTION

1.1 Mammalian Gene Expression

The fact that the majority of an organism's somatic cells, despite their vast morphological and functional differences, are genetically identical and arise from a single cell is a remarkable realization. The question of how this amazing diversity is achieved has been a key driver of molecular biology based research for many years. This has led to our current understanding of how differential expression of specific genes, both spatially and temporally, influences development and cellular differentiation. The proportion and absolute quantity of genes encoding transcription factors, proteins that regulate the expression of other genes, in an organism's genome underpin and reflect this complexity [1]. Gene regulation can be controlled at various stages between the initial transcription of a gene and the production of the final translated protein. The first point of regulation is at the transcriptional level, whereby the frequency of synthesis of a pre-mRNA molecule can be increased or decreased. At the post-transcriptional level, the pre-mRNA molecule can subsequently be processed through several different mechanisms, mainly via RNA splicing or selective degradation in response to specific sequences coded in the 3' untranslated regions (UTRs) of transcripts [2]. Lastly, post-translational modifications, such as glycosylation, phosphorylation and acetylation, add further dimensions to gene expression and protein function and stability [3].

However, despite these various mechanisms of control, the primary regulation of gene expression is achieved at the transcriptional level [4] where transcription factors and their co-regulators influence the frequency of transcriptional initiation of genes by binding to specific *cis*-acting DNA sequences known as regulatory elements. There are three different types of regulatory sequences involved in the control of gene expression, namely promoters, enhancers and silencers.

1.2 DNA regulatory elements

1.2.1 Promoter Regions

The basal promoter of a gene is found approximately 100 base pairs (bp) upstream, and straddling the transcription start site (TSS). Many promoters contain sequences such as TATA-boxes, which are important for the assembly of general transcription factors (GTFs) and the binding of RNA polymerase. Interestingly, regulatory elements such as TATA-boxes do not appear necessary for the initiation of transcription [5], making the exact start points of transcription and basal promoter regions difficult to identify by sequence examination alone.

In addition to basal promoters, individual genes are also regulated by promoter-proximal elements, which are other *cis*-acting DNA sequences capable of influencing the frequency of transcriptional initiation. They are usually found between 100 and 200 bp upstream of the TSS and are not directly bound by the GTFs. These elements instead interact with sequence specific, regulatory transcription factors. Promoter-proximal elements are often involved in orchestrating the tissue-specific expression of a gene. For instance, GATA- and CACCC-sequences, are often co-located in control regions of haematopoietic genes [6]. The combination of the basal promoter region of a gene and any promoter-proximal elements is often loosely referred to as the "promoter" of the gene.

Many genes may be regulated by more than one promoter; these additional promoters are termed alternative promoters. By having multiple promoters, genes can be regulated in a temporal-, tissue- or cell-specific manner [7]. An example of such is the human porphobilinogen deaminase (PBGD) gene, which contains a housekeeping, and a tissue-specific promoter [8]. The upstream housekeeping promoter is active in all tissues and therefore drives ubiquitous expression of this gene. The downstream promoter is erythroid-specific and induces expression of a different isoform of the PBGD protein, which is unique to erythroid cells. From this example, it can be seen that the use of alternative promoters is a vital part of differential gene expression in cells.

1.2.2 Enhancers and silencers

Enhancers and silencers are additional examples of elements to which regulatory transcription factors can bind, and can be both *cis*- and *trans*-acting. Enhancers are involved in the activation of transcription while silencers mediate repression. Like proximal promoters, they are often involved in controlling the specificity of gene expression, and different enhancers and silencers may be active in different cell types, or at various stages of development. Where they differ from proximal regulatory elements is that they are involved in long-range interactions, and are able to exert their effects on transcriptional frequency independently of their position and orientation [9]. They can be found kilobases (kb) up- or downstream of their target genes, or even within the gene's introns, making them yet more difficult to identify. Enhancers and silencers are typically larger than the proximal promoters, and are capable of simultaneously binding numerous regulatory transcription factors. A well-studied example of these interactions occurs at the *α-globin* locus [10]. This locus is regulated in an erythroid-specific manner by four distal elements, lying 10 to 15 kb upstream of the locus. It has been shown that one of these elements is the most important, and removal

of this particular element (DNAse I Hypersensitive site (HS) -40) results in a loss of all the looping interactions, as well as *α-globin* expression [11].

1.3 Gene Regulators

1.3.1 Transcription Factors

Transcription factors and their co-regulators, which can be either proteins or RNA [12], are crucial in orchestrating the ordered temporal and spatial expression of genes in an organism. These molecules intricately interact with one another in regulatory complexes to control gene expression programs through the activation and/or repression of their target genes. Transcription factors are able to bind specific sequences of DNA in control regions, such as promoters, enhancers and silencers. These proteins recognise their specific binding sequences via one or more DNA-binding domains, which include classes such as zinc finger, leucine zipper and basic helix-loop-helix [13]. Zinc fingers are the most common class of DNA-binding domain in eukaryotes [14] The majority of Zinc finger proteins consists of a C-terminal C2H2 zinc finger domain that binds DNA, as well as modular N-terminal domain which confers functionality, for instance gene activation or silencing.

1.3.1.1 The role of transcription factors in transcriptional regulation

Once bound to specific regions of DNA in promoters, enhancers and/or silencers, transcription factors influence the expression of their target genes by interacting with other proteins, known as co-regulators. These are proteins that are able to modify the levels of transcription but are not able to directly bind DNA. Transcription factors will recruit their co-regulators through the use of activation and repression domains, which are less well characterised. Activation domains have been broadly shown to be acidic, and/or glutamine- or proline-rich [15], while repression domains have often been described as being basic, and/or alanine- or proline-rich [16].

Transcriptional activators are able to increase the levels of gene expression in a number of ways. They may be involved in the recruitment and stabilisation of the basal transcription apparatus, which may occur directly, or indirectly by association with co-activators. They may also operate by tying distal enhancer elements to regions closer to the basal apparatus [17, 18]. In contrast, repressors interact with co-repressors, to reduce levels of gene expression. Another mechanism repressors may act by is physically obstructing interactions between activators and the basal transcriptional apparatus [19]. Genes can be both activated

and silenced by transcription factors interacting, either directly or indirectly, with histones, thereby altering the physical conformation of DNA to form euchromatin, loosely packaged DNA associated with active genes, and heterochromatin, tightly packaged DNA associated with silenced genes.

DNA activation or inactivation is highly dependent on specific residues in the tail of specific histones being either acetylated, usually associated with active euchromatin, or methylated, which can be associated with both active and inactive genes, and the degree and specific residues on which this occurs [20, 21]. Histone acetyl transferases (HATs), a subset of co-activators, act by adding acetyl groups to lysines found in histone tails, thereby reducing the strength of their ionic interaction with DNA. This promotes chromatin remodelling, which in turn allows greater access for both general and regulatory transcription factors. Histone methylation is mediated by histone methyltransferases (HMTs), which catalyse the addition of one, two or three methyl groups on lysine (K) and arginine (R) residues. Histone demethylases, such as LSD1/KDM1 are involved in removing methyl groups. Various methylation marks at different residues serve as markers for the recruitment of various proteins or protein complexes which regulate chromatin activation or inactivation.

The regulation of gene expression to ensure normal cellular development and differentiation is clearly complex. Past insights into this complexity have been gained by studying how gene regulatory proteins control specific biological processes. Following this approach, this thesis focuses on haematopoiesis, aiming in particular to provide a better understanding of the transcriptional regulation of red blood cell development.

1.4 Haematopoiesis

The various types of mature blood cells in the mammalian system are all derived from a multi-potent haematopoietic stem cell (HSC), which is able to differentiate down a number of different pathways. This process is collectively referred to as haematopoiesis [reviewed in 22]. Haematopoiesis can broadly be subdivided into two processes: lymphopoiesis and myelopoiesis (Figure 1.1). While the first gives rise only to lymphocytes, the latter generates all the other cells of the blood, including red blood cells, also known as erythrocytes. Differentiation along each of the haematopoietic lineages is defined by distinct sets of transcription factors, which are referred to as master regulators. Extracellular signals, for instance cytokines such as erythropoietin (Epo) can drive differentiation down a particular lineage, in this case erythropoiesis, by modifying transcription factor programs [23].

Mammalian haematopoiesis is summarised in Figure 1.1.

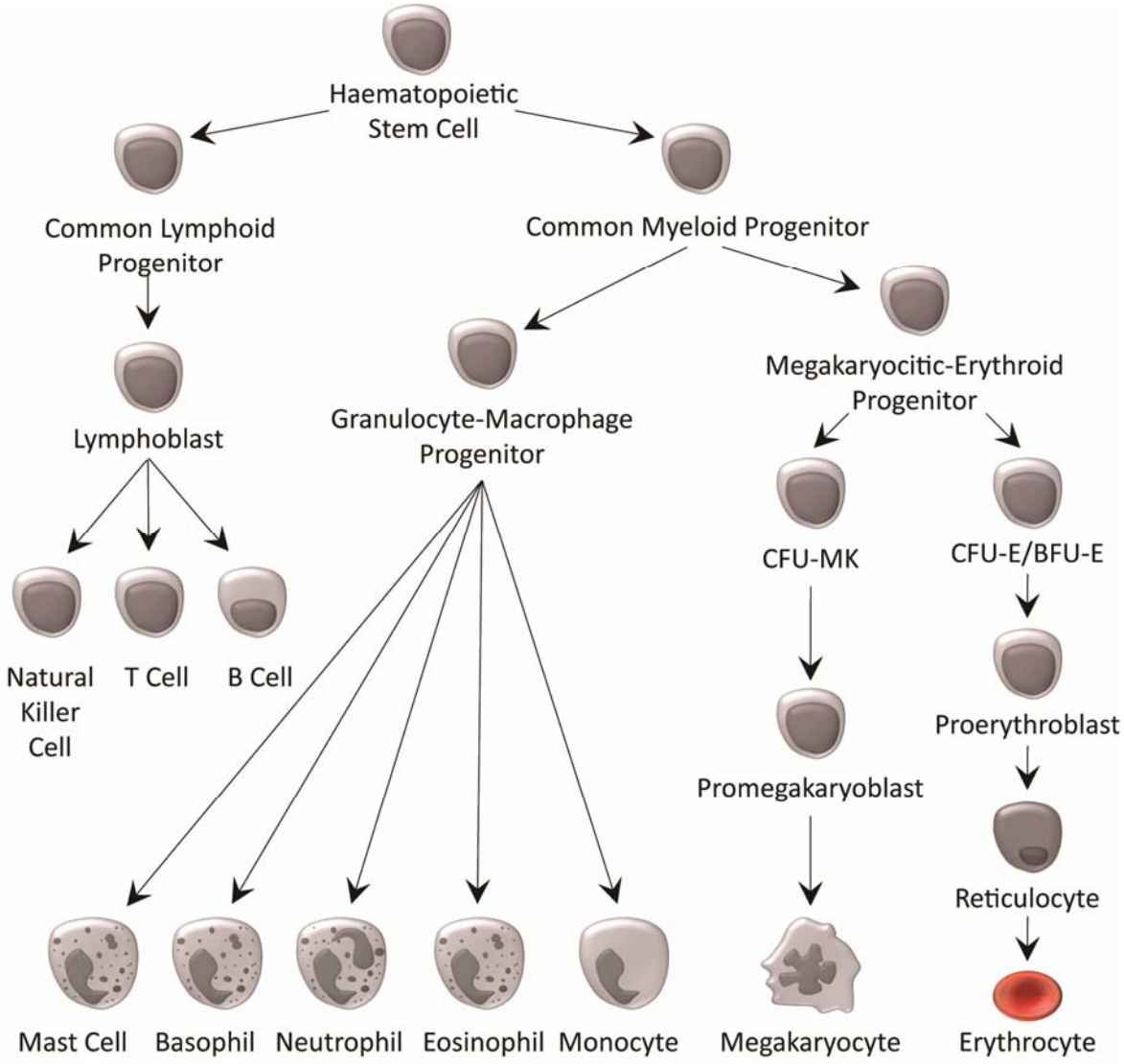


Figure 1.1. Haematopoiesis gives rise to a number of distinct cell lineages. A haematopoietic stem cell differentiates to yield myeloid and lymphoid progenitors. These progenitors in turn undergo distinct differentiation pathways to generate the various blood cells depicted. Megakaryocyte Colony forming unit (CFU-MK), erythroid Colony forming unit (CFU-E), erythroid burst forming unit (CFU-B). [Adapted from 24].

1.4.1 Lymphopoiesis

During lymphopoiesis, a common lymphoid progenitor differentiates into B and T lymphocytes, as well as natural killer (NK) cells. B cells, which develop in the bone marrow, play a role in the antibody-dependent humoral immune response, forming a crucial part of the adaptive immune system. T cells, on the other hand, mature in the thymus and are principally involved in cell mediated immunity.

1.4.2 Myelopoiesis

During myelopoiesis, a common myeloid progenitor can differentiate into a granulocyte-macrophage progenitor or a megakaryocytic-erythroid progenitor. Granulocyte-macrophage progenitors give rise to monocytes and granulocytes, which are further specialised into mast cells, neutrophils, eosinophils, monocytes, and basophils [25]. These myeloid cells are phagocytic and form an important part of the innate immune system [26].

The other branch of myelopoiesis can be further subdivided into erythropoiesis and megakaryopoiesis. Megakaryopoiesis yields large, polyploid cells known as megakaryocytes, which are responsible for producing platelets, required for blood clotting. Erythropoiesis gives rise to the erythrocytes; small enucleated cells that transport oxygen. These cells are both derived from the megakaryocytic-erythroid progenitors (MEPs) [27].

The main focus of this thesis will be the branch of erythropoiesis.

1.4.3 Erythropoiesis

Erythrocytes are the oxygen carrying cells of the body. Erythrocytes begin their maturation process as common megakaryocytic-erythroid progenitors (MEPs), where GATA-1, a master erythroid and megakaryocyte transcription factor, expression is high [28]. GATA-1 also regulates the expression of Epo and Epo receptor [29], which commit cells to the erythroid lineage. At this point they form small colonies known as erythroid colony forming units (CFU-E) and mature into erythroid burst forming units (BFU-E). The next stage of maturation is the proerythroblast followed by a number of different stages of erythroblasts with various chromatic states. It then begins to condense and extrude its nucleus at which point it is known as reticulocyte. Reticulocytes still contain RNA and nuclear fragments. Once these are all extruded from the cell it is a mature erythrocyte, with its well-known concave disc shape.

Mammalian erythropoiesis commences during early embryogenesis in two distinct waves. The first wave, known as primitive erythropoiesis, involves the transient generation of primitive erythroblast progenitors in blood islands located in the yolk sac, as early as embryonic day 7.5 (E7.5) in the mouse [30-32], and around day 17 of development in the human [33]. Coinciding with the onset of cardiac contractions, these large, nucleated erythroid cells enter the circulation at around E8.5 in the mouse and Day 21 in the human,

and continue to mature [33-36]. Some of these primitive cells reside temporarily within the foetal liver, where they enucleate before re-entering the circulation [37].

The second wave is known as definitive erythropoiesis; it begins in the foetal liver at around E9.5 in the mouse, and is seeded by progenitors from the embryonic yolk sac and aorta-gonad-mesonephros region [38-41]. In the human, hepatic colonisation by HSCs happens first at day 23, arterial cluster formation occurs at day 30, followed by a second round of hepatic colonisation [33]. Definitive erythrocytes then mature and enucleate in the foetal liver before being released into the bloodstream [42-45]. The liver remains the main site of erythropoiesis during murine foetal development until around E16, and in human foetal development till around 10.5 weeks [46], after which time blood cell development begins to move to the bone marrow. In both humans and mice, the bone marrow remains the major site of erythropoiesis throughout adult life [47], although in the mouse, some erythropoiesis also occurs in the adult spleen, particularly in response to anaemic stress [48].

1.5 Haemoglobin protein

Haemoglobin, the protein found in erythrocytes which binds to oxygen, consists of three major types: embryonic haemoglobin, foetal haemoglobin and adult haemoglobin [49]. Each of these proteins have different oxygen affinities in order to most effectively facilitate oxygen transfer. For example, foetal haemoglobin (HbF), which is the predominant form of haemoglobin in the foetus from around 10 weeks of gestation until birth, has a higher affinity for oxygen than adult haemoglobin (HbA). This property enables HbF to effectively take oxygen from the blood of the mother [49]. At around the time of birth, the adult globin genes are up-regulated and HbA is formed.

1.6 The Globin Loci

Erythrocytes contain haemoglobin, the oxygen and carbon dioxide transporting component of blood. This tetrameric metalloprotein is comprised of two α- and two β-like globin subunits, which are encoded by two gene loci in mammals: the *α-globin* and the *β-globin* locus. Each of these loci contains several genes, described below.

1.6.1 The alpha locus

The human *α-globin* locus is located on chromosome 16, and mouse on chromosome 11. This locus consists of three active genes: the embryonic *ζ-globin*, and the adult *α1-* and *α2- globins*.

1.6.2 The beta locus

The β-globin locus in the mouse, on chromosome 7, contains four expressed genes: two embryonic globins, *εy* and *βh1*, and two adult globins, *βmajor* and *βminor*. The human *β-globin* gene cluster, located on chromosome 11, consists of five transcribed genes: the embryonic globin (*ε-globin*), two foetal (*Gγ-* and *Aγ-globin*) and two adult globin genes (*β-* and *δ-globin*) (Figure 1.2). The *Gγ-* and *Aγ-globin* genes are a result of a gene duplication and only differ by a single glycine to alanine amino acid substitution at residue 136, and for simplicity, are collectively referred to as the "*γ-globin*" gene [50].

A summary of the *α-* and *β-like globin* genes discussed in this thesis, as well as their corresponding Human Genome Organisation (HUGO) or Mouse Genome Informatics (MGI) nomenclature, are provided in Table 1.1.

Table 1.1. Gene symbols for the mouse and human *α*- and *β*-like globins. The corresponding Mouse Genome Informatics (MGI) and Human Genome Organisation (HUGO) nomenclature is also provided.

Species	Locus	Conventional gene symbol	MGI or HUGO nomenclature
Mouse	*α*-locus	*ζ*	*Hba-x*
		α1	*Hba-a1*
		α2	*Hba-a2*
	β-locus	*εy*	*Hbb-y*
		βh1	*Hbb-bh1*
		β-major	*Hbb-b1*
		β-minor	*Hbb-b2*
Human	*α*-locus	*ζ*	*HBZ*
		α1	*HBA1*
		α2	*HBA2*
	β-locus	*ε*	*HBE*
		Aγ	*HBG1*
		Gγ	*HBG2*
		δ	*HBD*
		β	*HBB*

1.6.3 The Beta-Locus Control Region

An important regulatory element lies 6-22 kb upstream of the *β-globin* genes, known as the locus control region, or LCR. This region is involved in the regulation of the timed expression of *β-globin* genes during embryonic development through facilitating physical chromatin looping interactions [51]. The LCR contains five DNase I hypersensitive sites (HS), with HS1-4 exhibiting strong enhancer functions [51].

1.7 Globin Switching

The composition of haemoglobin changes during embryonic development, which coincides with the shift in the location of erythropoiesis, in order to meet the varying oxygen demands of the developing embryo [52]. This sequential activation and silencing of the various globin genes during these different stages is known as haemoglobin switching.

The switch at the alpha globin locus is similar in mouse and human. First, embryonic *ζ-globin* is exclusively expressed in primitive erythroblasts produced in the blood islands of the embryonic yolk sac. Then adult *α1* and *α2 globins* are expressed when primitive

erythropoiesis gives way to definitive erythropoiesis in the foetal liver and bone marrow. Adult *α-globin* expression is then maintained into adulthood [49, 53].

The switch at the beta globin locus is slightly different, and will be the focus of this thesis. In humans, there are two switches that occur. During primitive erythropoiesis, where erythrocytes are produced in the blood islands of the yolk sac, embryonic *ε-globin* is expressed. However, at around week 10 of gestation, *ε-globin* expression is down-regulated and the foetal *γ-globin* expression is increased in erythrocytes produced in the foetal liver [51]. At around the time of birth to about 12 months postnatal, erythropoiesis shifts again from the foetal liver and is completely taken over by the bone marrow, with a corresponding silencing of the *γ-globin* genes and activation of the adult *β*-, and to a lesser extent, *δ-globin* genes [51].

In contrast to humans, in mouse development, there is only one switch at the *β*-locus. During primitive erythropoiesis in the yolk sac of the embryo, the embryonic *β*-like globins *εy* and *βh1* are expressed. When erythropoiesis shifts from the yolk sac to the foetal liver, and later to the bone marrow of the adult mouse, the embryonic globins are superseded by adult *β-major* and *β-minor* expression [51]. The generation of mouse models expressing human globins have proven highly useful in furthering our understanding of the regulation processes involved in globin switching [50]. Figure 1.2 shows a diagram summarising the *β-globin* locus and globin switching, along with the locations of expression during development, and the corresponding predominant haemoglobin proteins that are present.

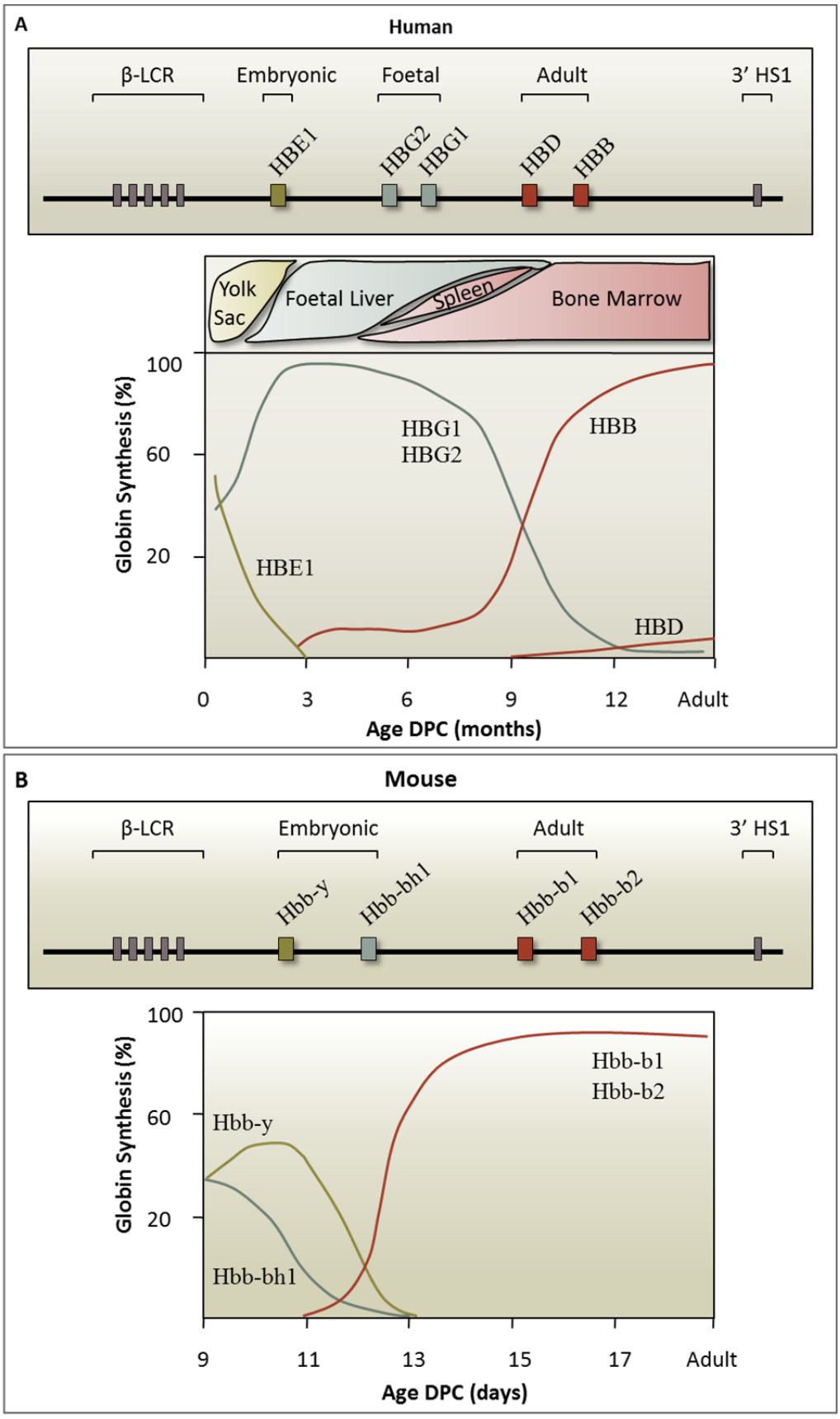


Figure 1.2. *β-globin* switching during development of human and mouse, including the organisation of the *β-globin* locus. This figure depicts the developmental switching of the *β*-like globin gene expression in human (top) and mouse (bottom) and the time at which this occurs days post coitum (DPC). The organisation of the genes at the *β-globin* locus appear on the chromosome in the same order that they are expressed during development.

The second switch of expression from foetal *γ-globin* to adult *β-globin* is of particular importance in the context of developing therapeutic strategies to treat *β*-haemoglobinopathies, a class of red blood cell diseases that are discussed in the following section.

1.8 Haemoglobinopathies

Haemoglobinopathies, diseases caused by the disruption of the adult globin genes, are some of the most common genetic diseases in the population, and include diseases such as thalassaemias and certain types of anaemia. Thalassaemia is caused by a reduction or complete abolition in the expression of one or more globin genes, resulting in an imbalance of α- or *β-globin* chains in erythrocytes [54, 55]. Sickle cell disease (SCD) is caused by a specific mutation in position 6 of the adult *β-globin* gene, which results in a single amino acid substitution of a glutamic acid to valine. This mutation results in polymerisation of *β-globin* chains, known as sickle haemoglobin (HbS), in circulating erythrocytes [56, 57]. This then results in a rigid and sickled cell phenotype [58, 59], which can trigger a number of acute conditions such as vaso-occlusion, splenic sequestration, and haemolytic anaemia [60]. Currently, the available treatments are costly, and often accompanied by unwanted side effects and risks.

The reactivation of the foetal *γ-globin* expression in adult patients poses an attractive therapeutic option. Residual production of foetal *γ-globin* persists naturally throughout life, allowing *γ*-globin chains to combine with adult *α*-globin chains to form HbF in the adult. Levels vary between individuals but in adults HbF is typically found to be around 1% of total haemoglobin [61, 62]. As only the adult *β-globin* gene is mutated in *β*-haemoglobinopathies, affected infants are protected from severe symptoms until around a year after birth, as the levels of HbF in circulation are still high [63]. Similarly, patients with *β*-thalassaemia who exhibit higher levels of HbF have a reduced disease phenotype [64]. Accordingly, drug treatments for these diseases, such as 5-azacitidine, hydroxyurea, and butyrate, all act by non-specifically reactivating foetal *γ-globin* gene expression by various mechanisms. The effects of these drug treatments are transient and thus require ongoing administration throughout the life of the patient. There is also a growing body of evidence that long-term administration of these drugs can have chronic side effects, which are consistent with their lack of specificity [59, 65]. Lastly, and perhaps most importantly from a potential therapeutic point of view, patients who have inherited alleles associated with increased levels of HbF, known as hereditary persistence of foetal haemoglobin (HPFH), are protected from these diseases throughout adulthood [66]. Together, these observations indicate that reactivation of

foetal *γ-globin* is able to compensate, at least in part, for the loss of adult *β-globin* function thereby ameliorating the symptoms of certain adult haemoglobinopathies.

1.9 Hereditary Persistence of Foetal Haemoglobin

HPFH is a benign condition classified by elevated HbF to levels greater than 1% of total haemoglobin during adult life [49, 66]. This condition has been shown to be associated with various point mutations found in the promoter region of the foetal *γ-globin* gene. In particular, a cluster of mutations located approximately 200 bp upstream of the TSS, known as the -200 region, has been reported in a number of different families with HPFH [49, 67-72]. Seven single-nucleotide HPFH mutations in this region, shown in Figure 1.3, are associated with various levels of elevated HbF, outlined in Table 1.2. The fact that these mutations show such dense clustering has led to the hypothesis that they may disrupt the binding of a transcriptional repressor, which under normal circumstances would be involved in silencing the foetal *γ-globin* gene. Understanding the mechanisms underlying these mutations, and how exactly they lead to elevated levels of HbF, is important for therapeutic reasons. If these naturally occurring mutations could be introduced into patients with various *β*-globinopathies, disease symptoms could theoretically be ameliorated. As such, determining the transcription factors involved is of considerable interest.

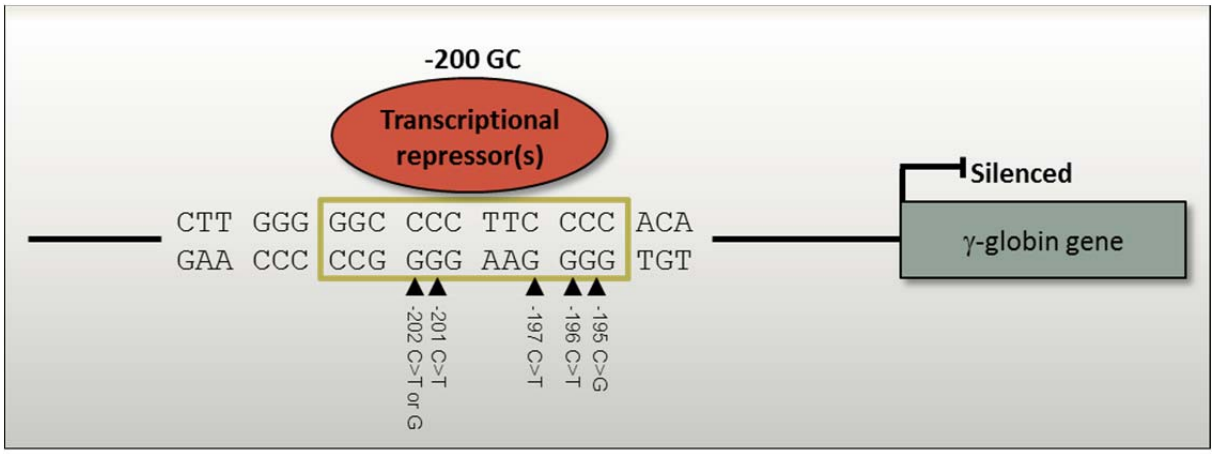


Figure 1.3. HPFH mutations at the -200 region of the foetal *γ-globin* promoter. The diagram displays the DNA sequence between -210 and -190 bp upstream of the transcription start site (+1). HPFH mutations are denoted by arrows. These HPFH mutations are thought to disrupt the binding of a transcriptional repressor(s) which silences the *γ-globin* gene during the foetal to adult haemoglobin switch.

Table 1.2. **The HPFH mutations in the *γ-globin* gene promoter are associated with different levels of HbF persistence in individuals.** HbF levels typically represent <1% of total haemoglobin in normal individuals. N/A indicates that data are unavailable for the genotype listed.

Gene	Position relative to TSS	Mutation	% HbF in homozygote	% HbF in heterozygote	Reference
Aγ	-202	C → T	25	2.5	[71]
Gγ	-202	C → G	N/A	20-23.5	[67]
Aγ	-201	C → T	N/A	10.2	[72]
Aγ	-197	C → T	N/A	6	[49]
Aγ	-196	C → T	N/A	38-40	[69, 70]
Gγ	-196	C → T	N/A	8.6	[72]
Aγ	-195	C → G	N/A	7	[68]

1.10 Transcription factors in globin switching

1.10.1 Klf Family

The Klfs are a family of transcription factors which consist of 18 members identified to date (Klf1-Klf18) [73-75], which are divided into groups based on functional similarities. These transcription factors are characterised by three highly conserved C-terminal C2H2 zinc finger motifs that function as the DNA-binding domain [76]. This domain recognises and binds to CACCC-boxes and other GC rich elements typically found at promoter and enhancer regions [77-79]. Members of this family have been shown to function as transcriptional activators and/or repressors [80].

Several Klfs have been shown to play important roles in haematopoietic cells. The most extensively studied Klf is arguably KLF1.

1.10.1.1 KLF1

Krüppel-like factor 1 (KLF1), the founding member of the family was first identified in 1993 [79]. The expression of *Klf1* is largely restricted to erythroid cells, although it has also been detected in other haematopoietic cells, such as macrophages and mast cells [79, 81]. KLF1 is a potent transcriptional activator, which binds to CACCC-box motifs and GC rich sequences [82] that are prevalent in the regulatory regions of many erythroid genes [79, 82-85]. The most notable CACCC-box lies within the adult *β-globin* promoter and KLF1 is crucial for the expression of this gene [82, 86]. KLF1 has been shown by chromatin immunoprecipitation (ChIP) to directly bind to the *β-globin* promoter *in vivo* [87, 88] and *Klf1* null mice die at around E15 of severe *β-globin* deficiency and anaemia [89, 90]. Many

other erythroid genes are also deregulated in the absence of *Klf1* [84, 87, 91, 92] including another member of the Klf family, the transcriptional repressor, Krüppel-like factor 3 (Klf3) [93]. Expression of *Klf3* has since been demonstrated to be directly activated by KLF1 [93, 94].

1.10.1.2 KLF2

KLF2 has been shown to play an important role in early erythropoiesis in the embryo [95]. KLF2 is closely related to KLF1, having approximately 88% amino acid sequence conservation in the zinc finger DNA-binding domain [96]. *Klf2* null mice die between E11 and E13 from anaemia and intra-embryonic haemorrhaging [97, 98]. KLF2 also plays a part in globin regulation, and *Klf2* null mice show decreased expression of the mouse embryonic globins *εy* and *βh1* in the yolk sacs of E12 embryos [95]. In addition, KLF2 has been shown to positively regulate human *ε-globin* in transgenic mice [95]. *Klf1* and *Klf2* double knockout mice show a further reduction in *εy* and *βh1* globin expression with more severe yolk sac and erythroid morphological abnormalities [99]. Hence, it is proposed that the two KLFs share functional redundancies in the regulation of embryonic globins and primitive erythropoiesis [100, 101].

1.10.1.3 KLF3

KLF3 is a potent transcriptional repressor that was first cloned from erythroid tissue in a screen designed to identify factors related to KLF1 [93]. KLF3 represses transcription by recruiting the co-repressor C-terminal Binding Protein (CtBP) [102, 103]. *Klf3* is widely expressed, but is particularly abundant in erythroid tissues [93]. It has a similar DNA-binding preference to KLF1, and the two factors have been shown to bind in vitro to many of the same erythroid promoter CACCC-boxes [93], including that of the *β-globin* gene. Moreover, as mentioned previously, *Klf3* expression is selectively reduced in *Klf1* null foetal liver [93] and KLF1 directly drives the expression of an erythroid-specific transcript of *Klf3* in primary erythroid tissue, namely *Klf3 1b* [94]. This transcript gives rise to the same protein as the ubiquitously expressed *Klf3* transcript. KLF1, KLF3 and another related family member, KLF8, have been shown to be involved in a regulatory loop that fine-tunes the expression of erythroid genes, including globins. That is, KLF1 is capable of activating a multitude of erythroid and non-erythroid genes, including KLF3 and KLF8, and KLF3 is capable of repressing a subset of non-erythroid genes to maintain the correct expression program in these cells [104].

1.10.2 BCL11A

B cell lymphoma 11A (BCL11A) is a multi-zinc finger transcriptional repressor with a well-established role in silencing the foetal *γ-globin* gene [105]. It was first identified in a genome-wide association study (GWAS) as a modulator of HbF expression [106-108]. It has subsequently been shown that a truncated variant of BCL11A is expressed in foetal and embryonic cells that robustly express *γ-globin*. BCL11A transcript levels are affected by genetic variation in an erythroid enhancer, resulting in increased HbF levels [109]. Furthermore, transgenic mice engineered with the human *β-globin* locus fail to completely silence the *γ-globin* genes in the absence of BCL11A [110]. Additionally, in a proof of principle study, inactivation of BCL11A in a SCD mouse model elevates HbF levels and corrects the haematologic and pathologic defects of SCD [111].

Importantly, KLF1 has been shown to drive the expression of BCL11A [112]. Mutations in the *KLF1* gene have also been found to be associated with HPFH, and it has been hypothesised that this is due to the down-regulation of BCL11A [113, 114], and studies in mice deficient in both *Klf1* and *Bcl11a* have supported this theory [115]. It has been reported in a recent study that co-expression of KLF1 and BCL11A in induced pluripotent stem cells and erythroid cells derived from cord blood, which usually express low levels of both of these genes and high levels of *γ-globin*, are able to strongly induce the expression of the adult *β-globin* gene [116].

1.10.3 SOX6

SOX6 is a member of the Sry-related HMG box transcription factor family. This transcription factor enhances erythroid differentiation in human erythroid progenitors [117]. It has also been shown to play an essential role in the silencing of *εy-globin* (the murine embryonic *γ-like globin* gene) during definitive erythropoiesis [118]. SOX6 binds directly to the *εy*-promoter and acts as a repressor of *εy-globin* expression [119]. A modest induction of *γ-globin* has also been shown in human erythroid cell knock down. This effect is greatly increased when combined with BCL11A knock down [120].

Interestingly, it has been demonstrated that a long range physical interaction exists between BCL11A and SOX6, which results in transcriptional silencing of the *γ-globin* genes [120]. Chromatin immunoprecipitation-chip (ChIP-chip) data revealed that SOX6 strongly binds the proximal promoters of the *Aγ-* and *Gγ-globin* genes in primary human erythroid progenitor cells, whereas BCL11A is not detected [120]. It has thus been postulated that SOX6 might recruit BCL11A to the *γ-globin* promoter to achieve silencing [120].

1.10.4 GATA-1

GATA-1 is another zinc finger transcription factor, which recognises GATA-motifs present in the regulatory regions of many erythroid genes [121-123]. It is one of the master regulators of both erythropoiesis and megakaryopoiesis. Erythroid maturation is affected in GATA-1 mutant mice and *Gata-1*$^{-/-}$ embryos die due to severe anaemia at around E10.5 [123-126]. It has also been shown that GATA-1 can bind to 3' enhancer elements and HS regions in the βLCR, as well as the *α-globin* and *β-globin* promoters in humans [127-129]. GATA-1 has been found to work in conjunction with other regulatory factors, such as NF-E2, Ikaros, Friend of GATA-1 (FOG-1) and the NuRD-complex in a stage specific manner to silence human *γ-globin* [130-132].

GATA-1, binds a region upstream of the *γ*-promoters in a FOG-1 dependent manner. This region is suggested to be necessary for *γ-globin* silencing in transgenic mice [133]. Ikaros recruits HDAC1 and Mi-2 (NuRD) to the *γ-globin* promoter and has been shown to enhance GATA-1 binding to various regulatory regions in the *β-globin* locus [134]. A recent study using transgenic mice containing the human *β-globin* locus has described a temporal mechanism of *γ-globin* silencing involving GATA-1, FOG-1 and Mi-2/NuRD complex. Here they show that GATA-1 is first recruited at E16 of development, followed then by FOG-1 and Mi-2 at E17 [135].

1.10.5 ZBTB7A

Another transcription factor which has been shown to play a role in erythropoiesis is the transcriptional repressor ZBTB7A (also known as LRF/POKEMON/FBI-1), a member of the BTB-ZF family. The C-terminus of this protein consists of four zinc fingers, and a BTB (Broad complex, Tramtrack, Bric a`brac and Zinc Finger) domain is located at the N-terminus is involved in protein-protein interactions [136]. ZBTB7A has a crucial anti-apoptotic role during terminal erythroid differentiation [137].The *Zbtb7a* knockout mouse is embryonic lethal at E15.5 due to severe anaemia, caused by an increase in apoptosis in late stage-erythroblasts [137]. For this reason, a conditional knockout of *Zbtb7a* in adult mouse long-term hematopoietic stem cells has been developed [138]. Microarray analysis of these cells revealed significant up-regulation of the murine foetal/embryonic stage haemoglobin gene *βh1*, by approximately 60-fold [138]. Interestingly, ZBTB7A has a DNA-binding consensus sequence, 5-G(A/G)GGG(T/C)(C/T)(T/C)(C/T)-3 [139] that closely resembles the -200 region of the *γ-globin* promoter. This, coupled with the fact that it is important in erythroid development, therefore establishes ZBTB7A as a promising candidate

transcriptional repressor that may operate through this region. In support of this, preliminary data from our laboratory has confirmed that ZBTB7A binds to the -200 region of the *γ-globin* promoter *in vitro*, and that this binding is disrupted by HPFH mutations in this region [140].

1.11 Aims and Hypothesis

The aims of this thesis were three fold. Firstly, we aimed to further explore the roles of ZBTB7A in erythropoiesis and foetal *γ-globin* regulation. This was done by analysing a combination of murine tissues, cell lines, and primary human haematopoietic stem cells (HSCs) from both foetal and adult sources, as well as analysis of a ZBTB7A ChIP-Seq. Given that we have hypothesised that ZBTB7A is involved in regulating the switch from foetal to adult globin, understanding how this gene is regulated is important. This led to our next set of aims.

Secondly, we aimed to investigate the regulation of ZBTB7A in an erythroid-specific context. Thirdly, we investigated whether *ZBTB7A*, like many other known erythroid genes such as *Klf3*, had an erythroid-specific transcript variant. We hypothesised that, like *Bcl11a*, KLF1 drives *Zbtb7a* in erythrocytes, and up-regulates the expression of an erythroid-specific transcript in these cells.

This hypothesis is particularly attractive in the light of previous work on KLF1 and BCL11A. KLF1 is known to drive *BCL11A* in erythroid cells. Patients with mutations in KLF1 accordingly display reduced levels of *BCL11A*, and since BCL11A represses foetal globin, the levels of HbF are elevated. Interestingly, foetal globin remains repressed to some degree in studies in *Bcl11a* transgenic mice [110, 115]. This implies that KLF1 is required to drive another distinct repressor of foetal globin expression. We believe ZBTB7A is that repressor.

CHAPTER 2 – MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals and reagents

Found below is a list of important chemicals and reagents used in the experimental work presented in this thesis, along with details of their suppliers or manufacturers. All chemicals and reagents used were of "molecular biology grade" unless specified otherwise.

- acetic acid (Asia Pacific Specialty Chemicals, Seven Hills, NSW, Australia)
- acrylamide (electrophoresis grade) (Sigma Chemical Company)
- adenosine triphosphate (ATP) (Sigma Chemical Company)
- adenosine 5′-[γ-32P] triphosphate ([γ-32P] ATP) (PerkinElmer Life Sciences, Boston, MA, USA)
- agarose (DNA grade) (Progen Industries, Toowong, QLD, Australia)
- agar (Amyl Media, Dandenong, VIC, Australia)
- albumin, bovine serum, fraction powder V (BSA) (Sigma Chemical Company)
- ampicillin sodium salt (Progen Industries)
- aprotinin (Sigma Chemical Company)
- β-mercaptoethanol (Sigma Chemical Company)
- calcium chloride (Sigma Chemical Company)
- casein peptone (Amyl Media)
- chloroform (Biolab Scientific, Clayton, VIC, Australia)
- deoxynucleotide triphosphates (dNTPs) (Sigma Chemical Company)
- diethylpyrocarbonate (DEPC) (Sigma Chemical Company)
- dimethylsulfoxide (DMSO) (Sigma Chemical Company)
- dipotassium hydrogen orthophosphate (Ajax Laboratory Chemicals, Auburn, NSW, Australia)
- dithiothreitol (DTT) (Sigma Chemical Company)
- DirectPCR (Tail) (Viagen Biotech Incorporation, Los Angeles, CA, USA)
- Dulbecco's modified Eagle medium (DMEM) (high and low glucose) (Gibco-BRL Life Technologies, Grand Island, NY, USA)
- Dynabeads Protein G (Life Technologies, CA, USA)
- ethanol (Ajax Finechem, Taren Point, NSW, Australia)
- ethidium bromide (Roche Molecular Biochemicals, Castle Hill, NSW, Australia)

- ethylenediaminetetraacetic acid (EDTA) (Ajax Laboratory Chemicals)
- ethylene glycol-bis[2-aminoethylether]-N,N,N′,N′-tetraacetic acid (EGTA) (Sigma Chemical Company)
- foetal calf serum (FCS) (Commonwealth Serum Laboratories, Parkville, VIC, Australia)
- formaldehyde (Sigma Chemical Company)
- formalin (Lomb Scientific, Taren Point, NSW, Australia)
- FuGENE® 6 transfection reagent (Roche Molecular Biochemicals)
- GeneRuler™ DNA ladder mix (Progen Industries)
- glutaraldehyde solution, 25% (Sigma Chemical Company)
- glutathione, reduced (GSH) (Roche Molecular Biochemicals)
- glycerol (Asia Pacific Specialty Chemicals)
- glycine (Ajax Laboratory Chemicals)
- heparin sodium salt (Sigma Chemical Company)
- N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid] (HEPES) (Roche Molecular Biochemicals)
- holo-human transferrin (Sigma Chemical Company)
- human inactivated plasma (Sigma Chemical Company)
- hydrocortisone (Sigma Chemical Company)
- isopropanol (Biolab Scientific)
- isoflurane – I.S.O. Veterinary Companies of Australia Pty Ltd, Kings Park, NSW, Australia
- Iscove's MDM (StemCell Technologies, Tullamarine, VIC, Australia)
- Jetstar Midi and Maxi Purification Systems (Genomed, Research Triangle Park, NC, USA)
- leupeptin (Sigma Chemical Company)
- lithium acetate dihydrate (Sigma Chemical Company)
- lithium chloride (Sigma Chemical Company)
- magnesium chloride (Sigma Chemical Company)
- methanol – Ajax Laboratory Chemicals
- 3-[N-Morpholino]propanesulfonic acid (MOPS) (Sigma Chemical Company)
- N-lauroylsarcosine (Sigma Chemical Company)
- NP-40 (Igepal CA-630) (Sigma Chemical Company)
- nucleoside triphosphates (NTPs) (Roche Molecular Biochemicals)

- t-octylphenoxypolyethoxyethanol (Triton X-100) (Sigma Chemical Company)
- penicillin, streptomycin and glutamine solution (Gibco-BRL Life Technologies)
- phenol:chloroform:isoamyl alcohol (25:24:1) (Progen Industries)
- phenylmethylsulfonyl fluoride (PMSF) (Sigma Chemical Company)
- phosphate-buffered saline (PBS) tablets (Sigma Chemical Company)
- poly(dI-dC) (Amersham Pharmacia Biotech, Little Chalfont, Buckinghamshire, UK)
- polyoxyethylenesorbitanmonolaurate (Tween™-20) (Sigma Chemical Company)
- potassium chloride (Sigma Chemical Company)
- potassium hydroxide (Sigma Chemical Company)
- Rainbow™ protein size standards (Amersham Pharmacia Biotech)
- skim milk powder (No Frills, Chullora, NSW, Australia)
- sodium acetate (Ajax Laboratory Chemicals)
- sodium azide (Asia Pacific Specialty Chemicals)
- sodium chloride (Ajax Laboratory Chemicals)
- sodium citrate (Asia Pacific Specialty Chemicals)
- sodium dihydrogen orthophosphate (Ajax Laboratory Chemicals)
- Sodium-deoxycholate (Sigma Chemical Company)
- sodium dodecyl sulfate (lauryl sulfate sodium salt) (SDS) (Sigma Chemical Company)
- sodium hydrogen carbonate (Asia Pacific Specialty Chemicals)
- sodium hydroxide (Ajax Laboratory Chemicals)
- recombinant human insulin (Sigma Chemical Company)
- RNase-Free DNase Set (Qiagen, Clifton Hill, VIC, Australia)
- RNeasy mini Plus kit (Qiagen)
- sodium acetate (Ajax Laboratory Chemicals)
- SYBR green PCR master mix (Applied Biosystems, Foster City, CA, USA)
- 4-hydroxytamoxifen (tamoxifen) (minimum 98% Z-isomer, remainder primarily E-isomer) (Sigma Chemical Company)
- tris-hydroxymethyl-methylamine (Tris) (Ajax Laboratory Chemicals)
- TRI-REAGENT™ (Sigma Chemical Company)
- QBSF60 (Quality Biologicals, Gaithersburg, MD, USA)

2.1.2 Recombinant Human Cytokines

- recombinant human erythropoietin (Epo) (Peprotech, Rocky Hill, NJ, USA)
- recombinant human Flt-3 (Peprotech)
- recombinant human Il-6 (Peprotech)
- recombinant human Il-3 (Peprotech)
- recombinant human stem cell factor (SCF) (Peprotech)
- recombinant human thrombopoietin (Tpo) (Peprotech)

2.1.3 Enzymes

- alkaline phosphatase (from calf intestine) (Roche Molecular Biochemicals)
- λ-protein phosphatase (New England Biolabs, Ipswich, MA, USA)
- Pfu and PfuUltra™ Hotstart DNA polymerase (Stratagene, La Jolla, CA, USA)
- proteinase K (Astral Scientific, Gymea, NSW, Australia)
- ribonuclease A (RNase A) (Roche Molecular Biochemicals)
- T4 DNA ligase (Roche Molecular Biochemicals)
- T4 polynucleotide kinase (New England Biolabs)
- Taq and REDTaq® DNA polymerase (Roche Molecular Biochemicals)
- Mango Taq (Bioline, Alexandria, NSW, Australia)
- type II restriction endonucleases (New England Biolabs or Roche Molecular Biochemicals)

2.1.4 Antibodies

Antibodies used for Western blots and chromatin immunoprecipitation are listed below. Rabbit anti-Eklf polyclonal antibody (αEklf) (raised against amino acids 1-114) has been described previously [93].

2.1.4.1 Primary Antibodies

- Anti-β-actin (αβ-actin) (Sigma Chemical Company)
- Anti-ZBTB7a/LRF (αZBTB7a/LRF) (Cat# sc-33683x, Santa Cruz at Thermo Fisher Scientific)
- Anti-KLF1 (αKLF1) (Cat# PIEPA5-18031, Thermo Fisher Scientific)
- Anti-rabbit IgG (Cat# sc-2027, Santa Cruz)
- Anti-goat IgG (Cat# sc-2028, Santa Cruz)

2.1.4.2 Secondary Antibodies

- Rabbit and Goat IgG (Cat # sc-2027 or sc-2028, Santa Cruz)
- ECL™ Anti – Mouse IgG – NA931V (GE Life Sciences, UK)
- ECL™ Anti – Rat IgG – NA 9350 (GE Life Sciences)
- Horseradish peroxidase-linked anti-Armenian hamster (Santa Cruz)

2.1.5 Oligonucleotides

Single stranded oligonucleotides were synthesised by Sigma Chemical Company. A list of the names and sequences of all oligonucleotides is found in Appendix I.

2.1.6 Commercial services and kits

All DNA sequencing reactions were conducted by The Australian Genome Research Facility Ltd, Sydney, NSW. Techniques that involved the use of a commercial kit were carried out as advised in the manufacturers' protocols. A list of commercial kits used is displayed below.

- 5' Rapid Amplification of cDNA Ends (RACE) (Clontech Laboratories Inc. Mountain View, CA, USA)
- cDNA synthesis was performed with the SuperScript III First-Strand Synthesis System for reverse transcription (RT)-PCR (Invitrogen, Groningen, The Netherlands)
- DNA-free™ (Ambion, Austin, TX, USA) for DNase-treatment of RNA
- Magnetic Cell Sorting mouse anti-TER119 Microbeads (Miltenyi Biotech, Auburn, CA, USA)
- PureLink HiPure Plasmid MaxiPrep Kit (Invitrogen)
- RNeasy mini kit (Qiagen, Clifton Hill, VIC, Australia) for purification of RNA samples
- Western Lightning™ Chemiluminescence Reagent Plus kit (PerkinElmer Life Sciences) for Western blot visualization
- Wizard SV Gel and PCR Clean-up System (Promega)
- Thermo Hybaid PCR Express (Integrated Sciences, Chatswood, NSW, Australia) and Mastercycler® (Eppendorf AG, Barkhausenweg, Hamburg, Germany) thermal cyclers were used for all PCRs
- TissueLyser II Adapter Set (Qiagen) for homogenisation and lysis of adipose tissue

2.1.7 Bacterial strains and culture medium

The bacterial strain used for all subcloning was Escherichia coli DH5α (supE44, ΔlacU169 [ϕ80lac ZΔM15], hsdR17, recA1, endA1, gyrA96, thi-1, relA1) (Biolone, Alexandria, NSW, Australia).

E. coli were cultured in Luria-Bertani (LB) broth or on LB-agar plates:

LB broth

- 10 g/L casein peptone
- 5 g/L yeast extract
- 10 g/L sodium chloride

LB was made up with MQW and was sterilised by autoclaving. LB-agar plates were prepared by adding 15 g/L bacteriological agar prior to autoclaving. Filter-sterilised ampicillin (50 mg/mL in MQW) was added to cooled, autoclaved broth to a final concentration of 50 μg/mL. In the preparation of agar plates, this step was performed immediately prior to pouring. Where appropriate, X-gal and IPTG were added to final concentrations of 40 mg/mL and 400 μM respectively. All media were stored at 4°C.

2.2 Methods

2.2.1 General methods

Routine molecular biological techniques were carried out as outlined in Sambrook et al. [141]. Page references for each technique are shown below.

- restriction endonuclease digestion of DNA: 5.24-5.32
- agarose gel electrophoresis: 6.1-6.20
- polyacrylamide gel electrophoresis: 6.36-6.43, 6.45, 18.47-18.55
- agarose gel DNA purification: 6.22-6.23
- calcium phosphate transfection of cells: 16.39-16.40
- DNA ligation: 1.63-1.69
- transformation of competent bacterial cells: 1.74, 1.76, 1.86
- phenol/chloroform extraction of DNA: E.3-E.4
- ethanol precipitation of DNA/RNA: E.10-E.14
- mini-preparations of plasmid DNA: 1.21-1.31
- polymerase chain reaction (PCR): 14.1-14.4, 14.14-14.21

- nuclear extracts from cultured cells: 17.8-17.10
- Western blots: 18.60-18.61, 18.64-18.66, 18.69-18.74
- electrophoretic mobility shift assays (EMSAs): 17.13-17.17
- RNA electrophoresis/formaldehyde agarose gels: 7.31-7.34

2.2.2 Mouse procedures

2.2.2.1 Husbandry

The ethics approval of animal use for this project was obtained from the Animal Care and Ethics Committee, University of Sydney (Approval No**. L02/7-2009/3/5079**). Animals were kept in a specific pathogen-free environment, in a stable temperature of 22°C to 24°C with a 12 h light dark cycle and were given free access to water and standard chow. Mice were housed at up to 6 animals per cage, with male and female mice caged separately except for breeding purpose.

2.2.2.2 Genotyping of mice

Pups were weaned at 3 weeks of age, and tail snips were taken under anaesthesia for genotyping purposes. Anaesthesia was performed by placing mice in an enclosed chamber containing cotton wool soaked with isoflurane. Anaesthesia was confirmed by observation of a slowed respiration rate and non-response to the standard toe pinch reflex test.

Tail biopsies from mice and embryo tail tissues were digested at 55°C in 150 µL DirectPCR (Tail) lysis buffer and 1 µL proteinase K (20 mg/mL) per reaction, with occasional mixing over a four hour period. Digestion was terminated at 85°C for 45 minutes and samples were spun at 13,000 g for 2 minutes to collect the supernatant containing genomic DNA. A single PCR was carried out allowing simultaneous detection of wild-type and neo sequences. 0.5 µL DNA extract was added to a 25 µL PCR mixture composed of 50 µM dNTPs, 1.0 µM *Klf1* WT primers, 1.0 µM *Klf1* KO primers, 1.25 units of RedTaq DNA Polymerase and 1x RedTaq buffer (see Appendix 1 for primers used for genotyping). The thermal cycling program for these PCRs was 94°C/2 min, followed by 36 cycles of 94°C/30 sec, 60°C/30 sec, 72°C/1 min and termination at 72°C/5 min. *Klf1* WT forward and reverse primers (A3510, A3511) lie in *Klf1* exon 2 and yield a product of 250 bp, indicating the *Klf1* WT genotype. The neo-*Klf1* specific primers (A3512 and A3591) recognise *Klf1* exon 2 and neo respectively, and amplify a neo-*Klf1* amplicon of approximately 350 bp from the *Klf1*

KO allele. PCR products were electrophoresed on a 2% agarose gel and visualised under UV light following ethidium bromide staining.

2.2.2.3 Timed mating of mice

To generate embryos of a specific age, heterozygote mice at breeding age were allowed to mate overnight with one male and one female per breeding box. Females were removed the following morning so that the age of the embryos was known at the time of tissue collection. Pregnant females were sacrificed at embryonic days E10.5, E13.5 and E14.5 by first anaesthetising in a closed isoflurane chamber and then culling by cervical dislocation. Foetal livers were dissected and peripheral blood was collected from the embryos for experimental analysis. The tail of each embryo was also collected and used for genotyping.

2.2.2.4 Adult mouse tissue collection

Adult tissue samples (up to 140 mg each) were obtained from 10- to 16-week-old male FVBN mice. Bone marrow cells were flushed out using a 26.5 gauge needle and homogenised by pipetting. Single cell suspensions of spleen were made by mechanical dissociation of through a 50 μM mesh cell strainer. Fat samples were immersed in QIAzol® Lysis Reagent, and all other samples in TR-REAGENT, and homogenised using a TissueLyser II. All erythroid samples, E10.5 yolk sac, E14.5 foetal liver and adult bone marrow and spleen, were sorted using Magnetic Cell Sorting anti-TER119 Microbeads as per the manufacturer's instructions (Miltenyi), and total RNA was extracted, or nuclear extracts prepared.

2.2.3 Mammalian cell culture

Mouse erythroleukaemia (MEL) and COS cells were cultured in Dulbecco modified Eagle medium (DMEM) (low glucose) supplemented with 10% (v/v) heat-inactivated FCS and 1% (v/v) penicillin/streptomycin/glutamine (PSG) solution. All cells were cultured at 37°C and 5% CO2.

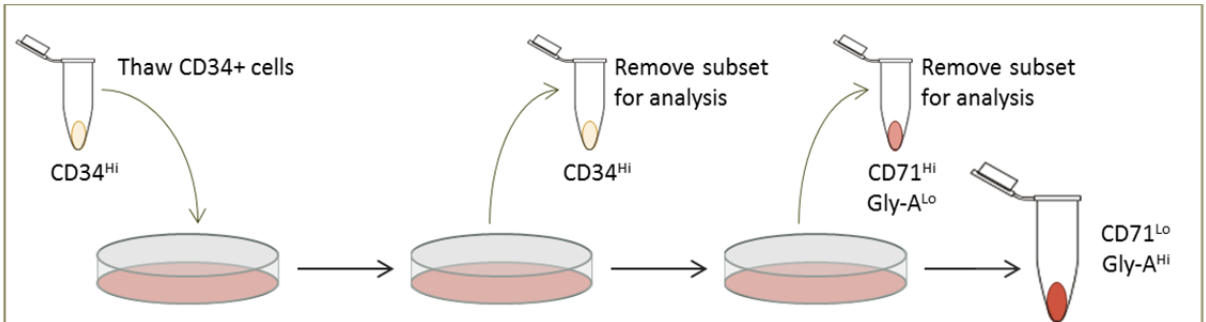
2.2.3.1 K1ER cell line

Klf1 Inducible Oestrogen Receptor (K1ER) erythroblast cells were provided by A. Perkins (Mater Research Institute, Brisbane, QLD, Australia) and were cultured as described previously [142]. Briefly, cells were maintained in high-glucose DMEM supplemented with 10% (v/v) heat-inactivated FCS and 1% (v/v) PSG at 37°C and 5% CO2. Subconfluent (30-50%) K1ER cells grown in 100 mm plates were differentiated by the addition of 1 mM 4-hydroxytamoxifen (tamoxifen) to a final concentration of 200 nM. Tamoxifen was made up in ethanol, so as a negative control, ethanol was added to K1ER cells to a final concentration of 0.02% (v/v). Total RNA was extracted (described in section 2.2.5) 48 hours after tamoxifen addition or at time 0, 0.25, 0.5, 1, 2, 3, 4, 6, 8 and 24 hours post-induction for the time course studies. Nuclear extracts were prepared from K1ER cells from 2 hours to up to 48 hour post addition of tamoxifen as described previously [143]. For the translation inhibition assays, cycloheximide (made up at 100 mg/mL in ethanol) was added to K1ER cells to a final concentration of 5 μg/mL and the cells were induced with tamoxifen (or ethanol as a negative control) 30 minutes later as described above. Due to the cytotoxicity of cycloheximide, RNA was extracted from the cells 6 hour after tamoxifen addition, rather than 48 hour after, to avoid cell death.

2.2.3.2 Primary human CD34+ cord blood and bone marrow cells – slow differentiation method

Human CD34+ cord blood cells were provided by K. Mackenzie (Children's Cancer Institute Australia, University of New South Wales, NSW, Australia) and were cultured and differentiated towards the erythroid lineage as previously described in [144]. Briefly, CD34+ cells were expanded at 4 x 10^4 cells/mL for 1 week in 20% foetal calf serum, 100 ng/ml stem cell factor (SCF), 100 ng/ml thrombopoietin (TPO) and 100 ng/ml FLT-3 ligand, plus 50 mg/ml gentamycin and 200mM glutamine in Isocove's modified Dulbecco's media (IMDM). Subsequently, cells were differentiated towards the erythroid lineage at 4 x 10^4 cells/mL in serum-free QBSF-60, substituted with 20 ng/mL SCF and 6 U/mL erythropoietin (EPO) for

another week, and in serum-free QBSF-60 substituted with 6 U/mL EPO for the final week of culture. Each week, cells were harvested for assessment of expansion and differentiation. A diagram of this protocol is shown in Figure 2.1.



Week:	Week 0	Week 1	Week 2	Week 3
Harvest:	Thaw cells	Harvest & resuspend	Harvest & resuspend	Final harvest
Media:	IMDM	QBSF-60	QBSF-60	
Cytokines:	SCF	SCF	SCF	
	TPO	EPO	EPO	
	Flt-3			
	Il-6			

Figure 2.1. Slow erythroid differentiation method of primary human CD34+ haematopoietic stem cells derived from cord blood or bone marrow. CD34+ cells were thawed and cultured in IMDM containing stem cell factor (SCF), thrombopoietin (TPO), Flt-3 ligand, and Interleukin-6 (Il-6) for seven days. At week 1, cells were harvested and a subset removed for analysis. Remaining cells were resuspended in QBSF-60 containing SCF and Erythropoietin (EPO). They were cultured for another seven days and then harvested at Week 2. Another subset was removed for analysis. Remaining cells were resuspended in QBSF containing SCF and EPO for a further seven days until week 3, after which they were all harvested for downstream experiments. Cell surface markers of cells at each stage are depicted in the figure.

2.2.3.3 Primary human CD34+ bone marrow cells – fast differentiation method

Human CD34+ bone marrow cells were purchased from Lonza (Gordon, NSW, Australia) and StemCell Technologies (Tullamarine, VIC, Australia) and were expanded and differentiated towards the erythroid lineage as previously described [145]. Briefly, CD34+ cells were cultured in erythroid differentiation medium (EDM) on the basis of IMDM supplemented with stabilized L-glutamine, 330 µg/mL holo-human transferrin, 10 µg/mL recombinant human insulin, 2 IU/mL heparin, and 5% solvent/detergent virus-inactivated plasma. The expansion procedure comprised 3 steps. In the first step (day 0 to day 7), 10^4 cells/mL were cultured in EDM in the presence of 10-6M hydrocortisone, 100 ng/mL SCF, 5 ng/mL IL-3, and 3 IU/mL EPO. On day 4, cell culture was diluted 1 in 4 in fresh medium containing SCF, IL-3, EPO, and hydrocortisone. In the second step (day 7 to day 11), the

cells were resuspended at 10^5/mL in EDM supplemented with SCF and EPO. In the third step (day 11 to day 18), the cells were cultured in EDM supplemented with EPO alone. Cell counts were adjusted to 7.5 x 10^5 to 1 x 10^6 and 5-10 x 10^6 cells/mL on days 11 and 14, respectively. Cells were harvested on days 7, 11 and 14.

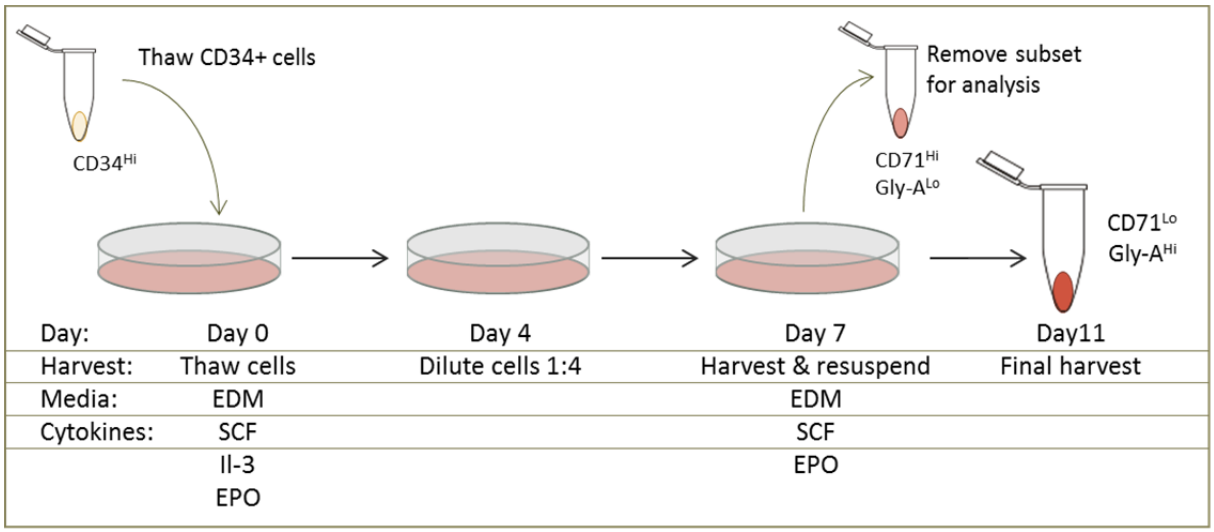


Figure 2.2. Fast erythroid differentiation method of primary human CD34+ haematopoietic stem cells derived from bone marrow. CD34+ cells were thawed and cultured in erythroid differentiation media (EDM) which consisted of IMDM with L-glutamine, holo-human transferrin, recombinant human insulin, heparin and human plasma. This media was supplemented with stem cell factor (SCF), Interleukin-3 (Il-3), Flt-3 ligand, and Erythropoietin (EPO) for four days, after which they were diluted 1 in 4 in the same media. At Day 7, cells were harvested and a subset removed for analysis. Remaining cells were resuspended in EDM supplemented with SCF and EPO. They were cultured for a further 2 days till Day 11 at which point they were all harvested for downstream experiments. Cell surface markers of cells at each stage are depicted in the figure.

Differentiation of cells by both methods was assessed by flow cytometry after staining cells fluorescein isothiocyanate (FITC)-conjugated Glycophorin A (GlyA). Aliquots were taken weekly for analysis by qPCR, Western blot and ChIP. The remaining cells were replated at 4 x 10^4 cells/mL in the appropriate cytokines. Cultures were expanded until the rate of death was greater than the rate of expansion. Population doublings (PDs) were calculated using the formula PDs=[log (number of cells harvested/initial cell number)/log 2].

2.2.4 Protein overexpression in COS cells

COS cells were grown in 100 mm plates to 50-80% confluence. Shortly prior to transfection, cells were supplied with fresh DMEM (supplemented with FCS and penicillin/streptomycin/glutamine). Cells were then transfected with 1-5 μg plasmid DNA using FuGENE® 6 as instructed by the manufacturer. Cells were harvested for nuclear

extracts 48 or 72 hour post-transfection as described previously [143]. When applicable, cells were provided with fresh medium 48 hour following transfection. 5 μg pCMV6-XL6-h*ZBTB7a* plasmid for overexpression of ZBTB7a, and 1-2 μg pMT3-*Klf1* plasmid for KLF1 overexpression, both for a 72-hour transfection period, were used for positive controls on Western blots. As a negative control, COS cells were mock-transfected with 1-2 μg empty pcDNA3 vector.

2.2.5 Western blotting

Nuclear extracts were performed as previously described [143]. Western blotting of nuclear extracts was performed as described previously [141, 146]. To detect mKlf1, 5 μL αmKlf1 A in 10 mL Tris buffered saline Tween-20 (TBST; 50 mM Tris, pH 7.4, 150 mM NaCl, 0.05% Tween-20) and 1 μL horseradish peroxidase-linked anti-rabbit antibody in 15 mL TBST were used. To detect mouse and human ZBTB7A, 7.5 ng of αZBTB7a in 10 mL TBST and 7.5 ng of horseradish peroxidase-linked anti-Armenian Hamster antibody in 15 mL TBST were used. Blots were probed for β-actin as a loading control, using 1 μL anti-β-actin in 10 mL TBST followed by 1 μL horseradish peroxidase-linked anti-mouse antibody in 15 mL TBST. Rainbow™ protein standards were loaded in each gel for size estimation. Densitometry of relative band intensities was performed using ImageJ Gel Analysis tool [147, 148]. Briefly, ImageJ was used to select and determine the background-subtracted density of the bands in the indicated blots.

2.2.6 RNA extraction and cDNA synthesis

Cell cultures or tissue samples were washed with PBS prior to RNA extraction. Adult tissue samples (up to 140 mg each) were obtained from 10- to 16-week-old male FVBN mice. Embryonic tissue was collected from FVBN embryos E10.5, E13.5 and E14.5.

Total RNA was extracted with TRI-REAGENT as per the supplier's protocol but with an additional centrifuge step at 12,000 g for 10 min at 4°C following homogenisation to reduce possible genomic DNA contamination. Cell preparations and embryonic tissues were lysed in TRI-REAGENT by pipetting while lysis of adult tissue samples was assisted by the use of the TissueLyser II laboratory homogenizer. To further reduce genomic and other contamination, RNA was cleaned by the use of RNeasy kits and was subsequently 'rigorously' DNase-treated with DNA-free kits as instructed by the supplier.

Approximately 5 μg of RNA was used for cDNA synthesis using SuperScriptTM III First Strand Synthesis System for qPCR Kit in which random hexamers were used as primers for

generation of first strand cDNA. For each sample, a negative control was also set up in the absence of SuperScript III reverse transcriptase (-RT) to check for genomic DNA contamination.

2.2.7 Quantitative real-time RT-PCR

Total RNA (up to 5 μg, assessed by Nanodrop) was used as a template for cDNA synthesis using the SuperScript III First-Strand Synthesis System kit. Reactions were primed with random hexamers rather than oligo(dT) primers to achieve more efficient reverse transcription from RNA templates with long 3′ untranslated regions (UTRs). The use of random hexamers also enabled cDNA synthesis from rRNA, thus allowing the use of 18S levels as standards for normalisation. Quantitative real-time PCRs (qPCR) (final volume 25 μL) were set up with SYBR Green PCR Master Mix and were run with the default cycle parameters of the ABI Prism 7000 SDS. Approximately 10 ng cDNA (assuming 100% reverse transcription efficiency) was employed in each qPCR in duplicate. Expression levels of genes of interest were normalised against 18S rRNA levels. As negative controls, duplicate minus RT and no-template reactions were always included. Data from qPCRs were analysed with ABI Prism 7000 SDS software.

2.2.7.1 q PCR primers

Primer Express™ software was used to design paired qPCR primers. Primer pairs were designed to straddle exon-exon junctions where possible to prevent amplification of any contaminating genomic DNA. Specificity of primers was verified by conducting genomic sequence searches using the Basic Local Alignment Search Tool [149]. For each primer set, optimal reaction concentrations were determined which provided high sensitivity and reproducibility but low amplification from possible primer dimerisation. For all reactions, final primer concentrations lay between 200 nM and 600 nM. Amplification efficiencies of primer pairs were tested using standard curves and were all found to be within an acceptable range close to 1. A list of all qPCR primers used is found in Appendix I.

2.2.8 5’ Rapid Amplification of cDNA Ends (RACE)

600 ng total RNA from un-induced and induced K1ER cells, and erythroid differentiated CD34+ cord blood and bone marrow at Week 1 and Week 3 of differentiation was used as template for first strand cDNA synthesis using the SMARTerTM RACE cDNA Amplification Kit. The RACE PCR was performed as directed by the supplier. RACE PCR primers are listed in Appendix I. Thermal cycler settings were 24 cycles of 94°C/30 sec,

64°C/30 sec and 72°C/1 min. Amplified products were resolved by electrophoresis through a 2% agarose gel and were purified using the Wizard SV Gel and PCR Clean-up System as per the manufacturer's manual.

Amplicons were then cloned into pUC19 vector using In-Fusion cloning, transformed into Stellar Competent cells, both supplied by Clontech, and minipreps were performed to extract the plasmids. Samples were sent for sequencing at the AGRF in Sydney, NSW, Australia.

2.2.9 Microarrays

10 µL of RNA extracted from E13.5 foetal liver cells was sent to the Ramaciotti Centre for Gene Function Analysis, University of New South Wales, at a concentration of 50 ng/µL. Quality of RNA was assessed by Nanodrop and Bioanalyzer. The labelling was done in accordance with Affymetrix IVT Express kit and hybridization in quadruplicate to Affymetrix Mouse Gene 1.0 ST Arrays (Affymetrix, Santa Clara, CA). Microarray data were analysed using Partek Genomics Suite version 6.6. Data were normalised using Robust Multi-Array Averaging with correction for GC content.

2.2.10 Chromatin immunopreciptiation

ChIP experiments were performed as previously described [150]. Approximately 5 x 10^7 cells were used for each IP. ChIP-qPCR experiments were conducted on K1ER cells induced with 200 nM Tamoxifen, harvested at 0, 2 and 8 hours post induction, and human cord blood CD34+ cells differentiated for two weeks towards the erythroid lineage. ChIP-seq was conducted on four K562 replicates, and duplicate human bone marrow CD34+ cells differentiated for 7 and 11 days towards the erythroid lineage. The antibodies used were anti-KLF1, ZBTB7A/LRF or anti-rabbit/anti-goat IgG as a negative control.

2.2.11 High-throughput sequencing

Library preparation was performed using the TruSeq DNA Sample Preparation Kit (Cat# FC-121-2001, Illumina, San Diego, CA) according to the manufacturer's instructions with minor modifications. Adapter sequences were diluted 1/40 before use and following adapter ligation, the library size extracted from the gel was 100-280 bp (excluding adapters) in line with the size of sonicated fragments. Libraries (4 inputs and 4 IP samples) were multiplexed into 4 lanes such that there were 2 samples per lane. Samples were sequenced using 50 bp single read chemistry on the HiSeq 2500 (Illumina, San Diego, CA). Sequencing was performed by the Ramaciotti Centre for Genomics, University of New South Wales, New South Wales, Australia.

2.2.12 Bioinformatics

Genomic DNA, cDNA and protein sequences were retrieved from GenBank and homology searches were performed using the BLAST algorithm at the National Centre for Biotechnology Information (NCBI) website [149]. Full-length *Zbtb7a* cDNA entries from GenBank and the DNA Data Bank of Japan (DDBJ) were obtained from nucleotide searches at the NCBI website. Additional start points of transcription and erythroid-specific transcript variants were derived from FANTOM 5. The UCSC genome browser was used to align 5'RACE sequenced products with mouse and human genomes.

2.2.12.1 Gene expression Microarray Analysis

Microarray data were analysed using Partek genomic suite v6.6 (Partek Inc., St. Louis, MO). Microarray CEL files were imported into Partek and normalized using the robust multi array average (RMA) algorithm. After confirming array quality (Affymetrix built-in controls and principal components analysis), differential gene expression was calculated and tested for significance using a 1-way analysis of variance (ANOVA). Gene expression P values were corrected for multiple testing using a false discovery rate (FDR) threshold of 0.05.

2.2.12.2 Alignment of ChIP-Seq reads

Quality control was performed using FastQC v0.10.1. Reads were quality filtered, trimmed and adapter sequences were removed using Trimmomatic v0.3.2 [151]. Reads were aligned to the hg19 *Homo sapiens* genome using Bowtie v2.2.1 [152] set to --very-sensitive.

2.2.12.3 IDR analysis

Pseudoreplicates were created using homer-idr v0.1 [153] for individual and combined IP samples. Peaks were then called using Homer v4.7.2 [154] using the permissive settings (-P .1 -LP .1 -poisson .1) on individual replicates, combined replicates, individual pseudoreplicates and combined pseudoreplicates against the combined input control. Peaks lists were then supplied to homer-idr to determine the IDR statistic for each peak generating a final peak list satisfying the thresholds set by homer-idr.

Reads were aligned to the hg19/NCBI build 37 *Homo sapiens* genome using Bowtie2 v2.0.0-beta7 [152]. In the first round, Bowtie2 was set to --very-sensitive and –D 40. Non-aligned reads were subjected to a second round of alignment where the read could be soft clipped by running Bowtie2 with the switch --very-sensitive-local. Resulting alignments

were sorted, merged and indexed using Samtools v0.1.18 [155]. HOMER was used to create bedgraph files using the makeUCSCfile program. These were viewed using IGV v2.3 [156].

2.2.12.4 Motif Discovery

The MEME-ChIP online tool was used for motif discovery [157]. This tool employs two motif discovery algorithms that have complementary characteristics, namely MEME [158] and DREME [159]. The top 800 peaks, all 164 basepairs in length, were selected from each ChIP-Seq dataset and genomic DNA sequences were extracted using Galaxy [160-162] in the FASTA format. These datasets were then individually uploaded into MEME-ChIP and were run using the JASPAR Vertebrates and UniPROBE Mouse database. 1st order model of sequences was used as the background model for all datasets. The expected motif site distribution in MEME was set to any number of occurrences, and MEME was instructed to search for 10 motifs between 5 and 30 wide. Default settings for DREME and CentriMo were used.

2.2.12.5 Published ChIP-Seq data sets

Published KLF1 ChIP-Seq datasets produced from murine KLF1-inducible K1ER cells by the Perkins Lab at the Mater Hospital in Brisbane, QLD, downloaded from GEO (Accession # GSE20478[84]), and human erythroid cell line K562s cells by the Gallagher Lab at Yale University School of Medicine, New Haven, Connecticut, downloaded from GEO (Accession # GSE43626 [85]) were used for analysis. The raw sequencing reads from these datasets were processed using the ChIP-seq pipeline described above to make bedgraph files for visualisation in IGV and to quantify sequencing tags at genomic locations of interest.

CHAPTER 3 – ZBTB7A IS A POTENTIAL REGULATOR OF GLOBIN SWITCHING

3.1 Chapter 3 Introduction

Foetal globin silencing has been studied for many years in relation to diseases of the adult *β-globin* gene such as SCD, in the hopes of reactivating foetal globin to alleviate symptoms. Many genes have been found to play a role in this process [163], and one very promising candidate is *Bcl11a*. BCL11A has been found under various circumstances to be involved in silencing *γ-globin* expression [105-107, 110]. However, the mechanism by which this occurs is largely unknown. BCL11A has never been found to bind directly to the *γ-globin* promoter [120], and knockout of this gene does not prevent all silencing of foetal globin [115]. This suggests that there are other important repressors, not yet discovered, involved in foetal globin silencing. Here, we make the case for another transcription factor with a role to play in this process, namely ZBTB7A.

As outlined in Chapter 1, ZBTB7A has been found to be important in erythropoiesis. Firstly, it has been found to be capable of binding many of the same target genes as the important erythroid transcription factor, GATA-1 [164]. Secondly, the *Zbtb7a* knockout mouse is embryonic lethal at E15.5 due to severe anaemia, which has been found to be caused by increased apoptosis of late stage-erythroblasts [137]. This is because ZBTB7A acts as a crucial anti-apoptotic factor during terminal erythroid differentiation [137]. Thirdly, conditional knockout of *Zbtb7a* in adult murine long-term hematopoietic stem cells showed significant up-regulation of the murine embryonic haemoglobin gene *βh1*, and orthologue of the human foetal *γ-globin* gene [165], by approximately 60-fold [138]. There is not yet published evidence of ZBTB7A regulating the human globin genes, however, the DNA-binding motif of the ZBTB7A protein, 5′-G(A/G)GGG(T/C)(C/T)(T/C)(C/T)-3′ [139], closely resembles the -200 region of the foetal *γ-globin* promoter, which contains a cluster of human foetal *γ-globin* gene up-regulating HPFH mutations, listed in Table 1.2 in Chapter 1. We hypothesised that ZBTB7A is the repressor that binds to the -200 region of the human foetal *γ-globin* gene promoter, and that HPFH mutations clustered at this region result in abolished binding of ZBTB7A and incomplete silencing of *γ-globin*.

3.1.1 ZBTB7A binds to the -200 region of the foetal globin promoter

Our lab has already begun investigations into whether ZBTB7A is capable of binding the -200 site of the *γ-globin* promoter *in vitro,* through the use of Electrophoretic Mobility

Shift Assays (EMSAs) (unpublished thesis [140]). In this project, G. Martyn from our lab investigated whether ZBTB7A can bind to the wild-type -200 region of the *γ-globin* gene, and also whether this binding is disrupted by any of the HPFH mutations discussed in Chapter 1. Figure 3.1 showed that ZBTB7A is indeed capable of binding to this site, and that binding is abolished by the HPFH mutations.

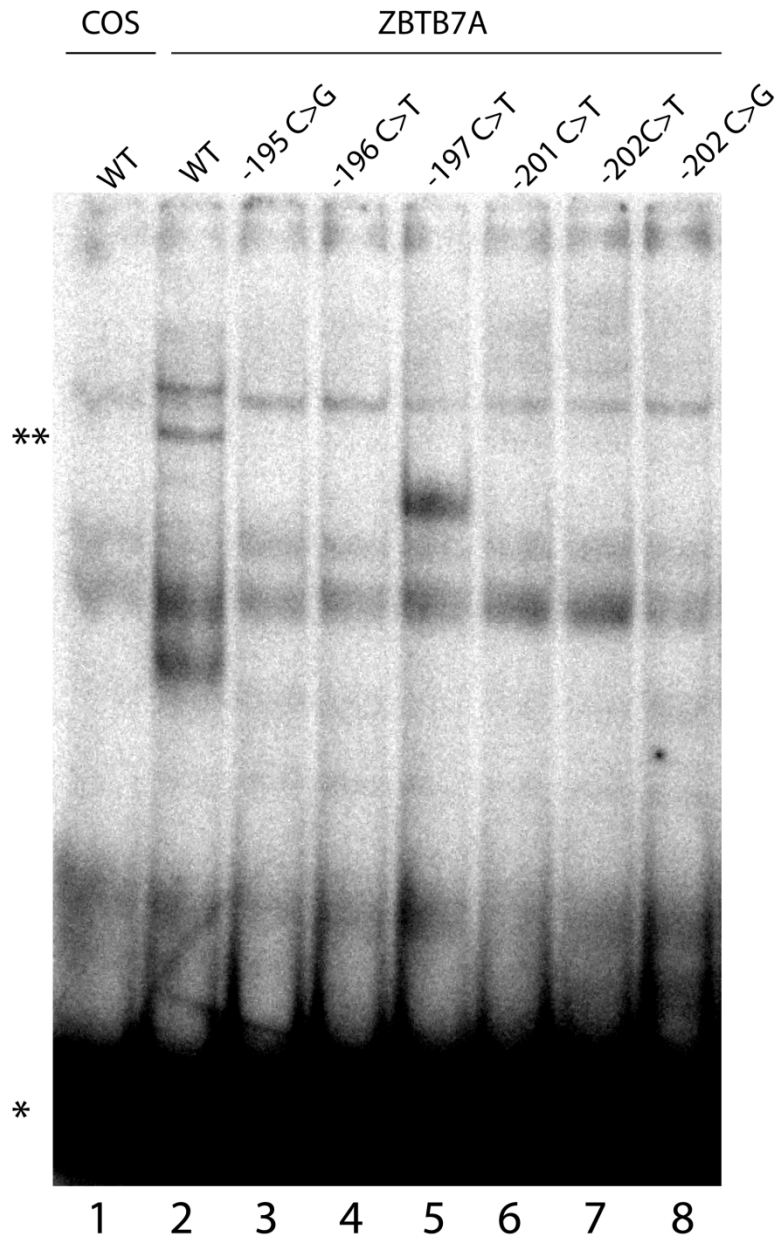


Figure 3.1. ZBTB7A binds the -200 region of the *γ-globin* promoter, and this is disrupted by 6 individual HPFH mutations. Lanes 1 and 2 contain radiolabelled wild-type (WT) probe (-209 to -187 bp) whilst lanes 3-8 contain HPFH mutant probes as indicated. Lane 1 contains nuclear extracts from Cos-7 cells transfected with pcDNA3 empty vector. Lanes 2-8 contain nuclear extracts from Cos-7 cells over-expressing ZBTB7A. Binding of ZBTB7A to the WT -200 *γ-globin* probe is seen in lane 2. Various HPFH mutations cause disruption of ZBTB7A binding to this sequence (lanes 3-8). Asterisks represent the unbound probe (*) and the ZBTB7A:probe complex (**). This result is representative of two EMSAs. This figure has been taken from the Honours thesis of G. Martyn, Figure 3.3, pg 33 [140].

This, taken together with the reported erythroid defects in the *Zbtb7a* conditional knockout mouse [137], the published binding motif of ZBTB7A [139], and the up-regulation of the mouse embryonic globin *βh1* in the *Zbtb7a* knockout cell line [138], provides considerable evidence that ZBTB7A is capable of aiding in the switch from foetal to adult globin expression through the repression the foetal *γ-globin* gene. This may be achieved by binding the -200 region of the *γ-globin* promoter.

In this chapter, we sought to further explore the role of *Zbtb7a* during erythropoiesis. Firstly, we investigated the expression of *Zbtb7a* in mouse and human erythroid cells, and compared this to the expression of the various *β*-like globin genes. And secondly, we explored the binding patterns of ZBTB7A through the use of ChIP-qPCR and ChIP-Seq experiments in human cells, and tested whether it binds to the -200 region *in vivo*.

3.2 *Zbtb7a* is expressed in erythroid cells in mouse

We began by investigating the expression of the *Zbtb7a* transcript in various mouse tissues in order to establish whether it is more highly expressed in erythroid tissues compared to non-erythroid tissues. In order to do this, we looked at a number of different non-erythroid tissues, namely, epididymal white adipose tissue, heart, liver, lung, kidney and brain, and erythroid tissues, both during development and from adult mice, namely, E10.5 yolk sac, E14.5 foetal liver as well as adult bone marrow and spleen. Erythroid tissues were sorted using the mature erythroid marker, TER119, present on the surface of erythroblasts, reticulocytes and mature erythroid cells. This enrichment step was performed because the whole tissues are made up of cells from various other sources, such as connective tissues and white blood cells and we were specifically interested in characterising the erythroid lineage cells only. Quantitative real-time RT-PCR (qPCR) was then performed across all tissues to investigate the expression levels of *Zbtb7a*. *Zbtb7a* transcripts were more highly expressed in erythroid cells than other tissue types, and were most highly expressed in the foetal liver at E14.5 of development (Figure 3.2). The difference in expression level between E14.5 foetal liver compared to all other tissues was found to be statistically significant. Transcript levels in yolk sac, bone marrow and spleen were also all statistically above the levels seen in non-erythroid tissues. All significance levels were calculated using a two tailed *t* test ($P < 0.05$).

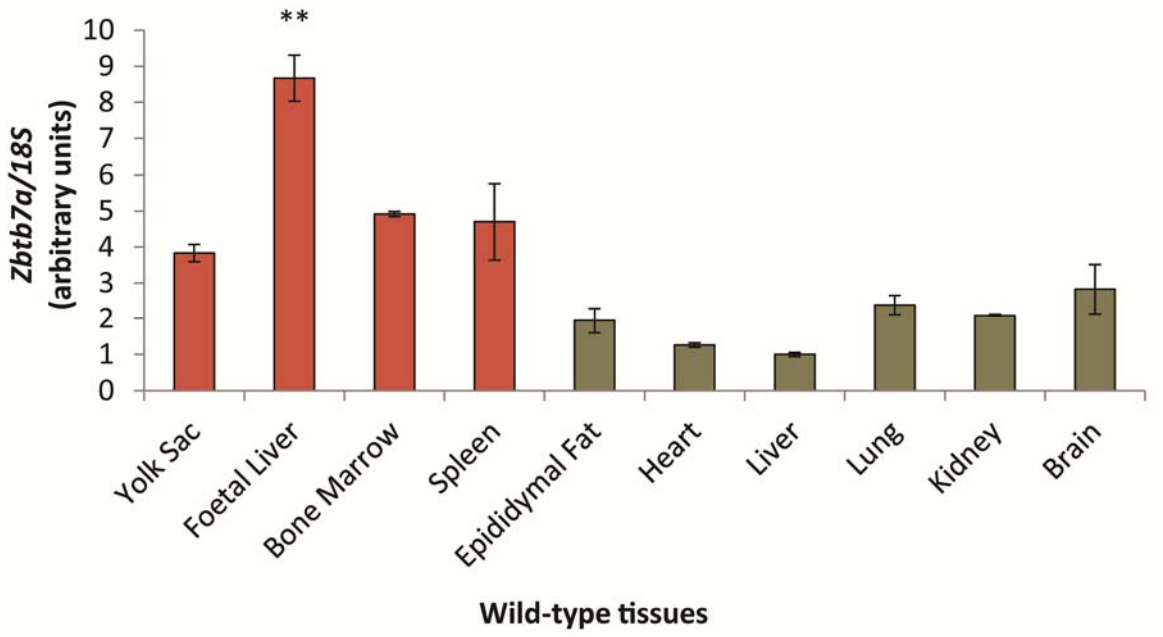


Figure 3.2. *Zbtb7a* transcripts are more highly expressed in erythroid tissues compared to non-erythroid tissues of wild-type mice. Yolk sac was harvested at E10.5, and foetal livers at E14.5. All adult tissues were harvested from mice aged between 10-12 weeks. Erythroid tissues were sorted with anti-TER119 Magnetic Microbeads. All other tissues were homogenized using a TissueLyserII. Whole RNA was extracted from the TER119+ erythroid cells, as well as the homogenized non-erythroid tissues. qPCR was used to investigate the expression of *Zbtb7a* and levels were normalised to *18S* rRNA levels. Error bars represent standard error of the mean (n = 3 per tissue). All tissues are statistically significantly different with a *P* value of < 0.05 when individually compared to E14.5 foetal liver (paired Student's *t* test).

Protein levels of ZBTB7A were also investigated in the erythroid tissues. Western blot analysis was performed on nuclear extracts from TER119+ erythroid cells from E10.5 yolk sac, E14.5 foetal liver and adult bone marrow and spleen (Figure 3.3). These results showed that the ZBTB7A protein, like the transcript, was also most highly expressed in E14.5 foetal liver. Densitometry analysis of bands on this blot revealed the difference again to be statistically significant between all tissues compared to E14.5 foetal liver. Interestingly, the magnitude of differences in protein expression appeared to be greater than the differences at the transcript level.

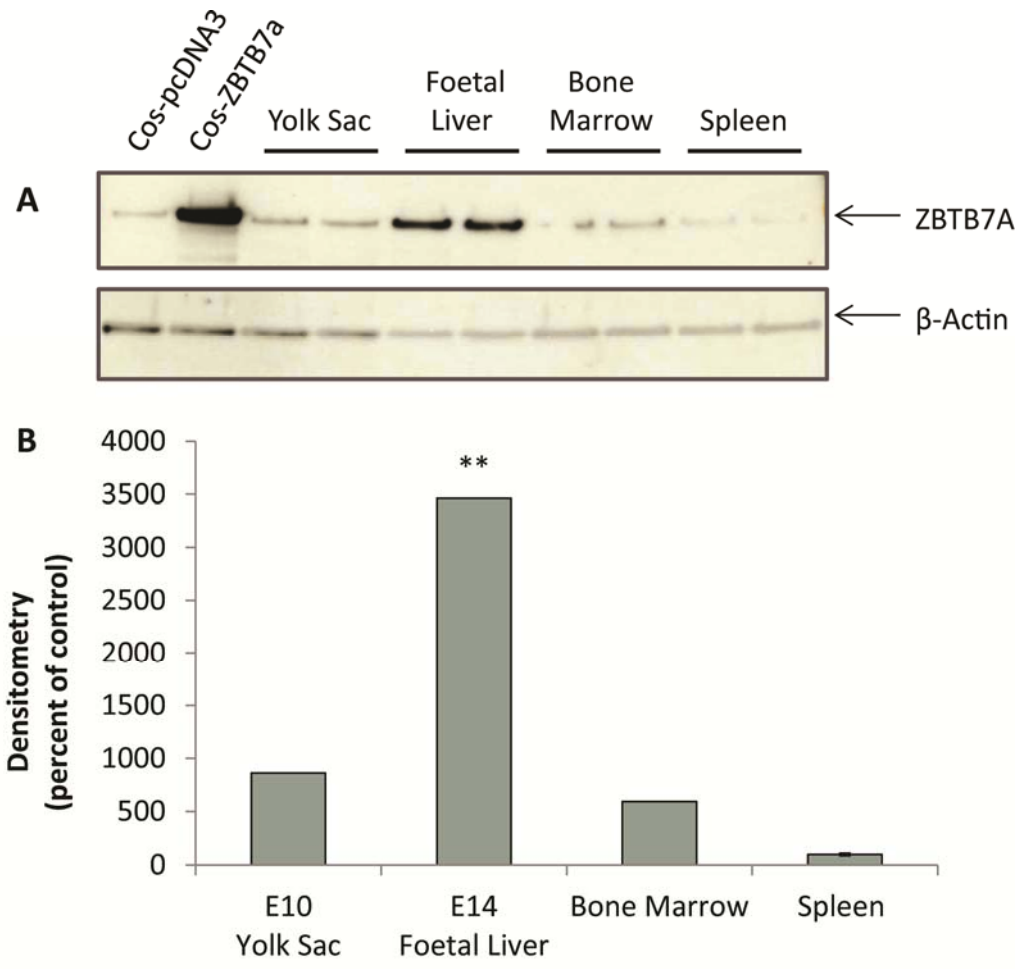


Figure 3.3. The ZBTB7A protein is most highly expressed in the foetal liver at E14.5. (A) Nuclear extracts were prepared from mock-transfected Cos-7 cells (lane 1), and Cos-7 cells overexpressing ZBTB7A (lane 2), as well as cells sorted with anti-TER119 Magnetic Microbeads from E10.5 yolk sac (lanes 3-4), E14.5 foetal livers (lane 5-6), adult bone marrow (lanes 7-8) and spleen (lanes 9-10). ZBTB7A protein levels were assessed by Western Blot. Blots were probed with αZBTB7A and β-Actin was used as a loading control. **(B)** Densitometry analysis of Western Blot in (A) performed using ImageJ analysis software. Error bars represent the standard error of the mean (n = 2 per tissue). **, $P < 0.01$ (paired Student's *t* test).

We also investigated the levels of the various genes at the *β-globin* locus in the TER119+ erythroid tissues (Figure 3.4). We found that the embryonic globins, *εy* and *βh1*, were both highly expressed early in development in the E10.5 yolk sac. Expression of these transcripts was significantly lower in E14.5 foetal liver, and not detectable in the adult tissues. On the other hand, the adult globins, *β-major* and *β-minor*, were not expressed in E10.5 yolk sac, but were expressed to a similar level in E14.5 foetal liver as well as the adult tissues. This adds evidence to the role of *Zbtb7a* in globin regulation, as the time point where it is most highly expressed correlates to the time point of embryonic globin silencing in mice.

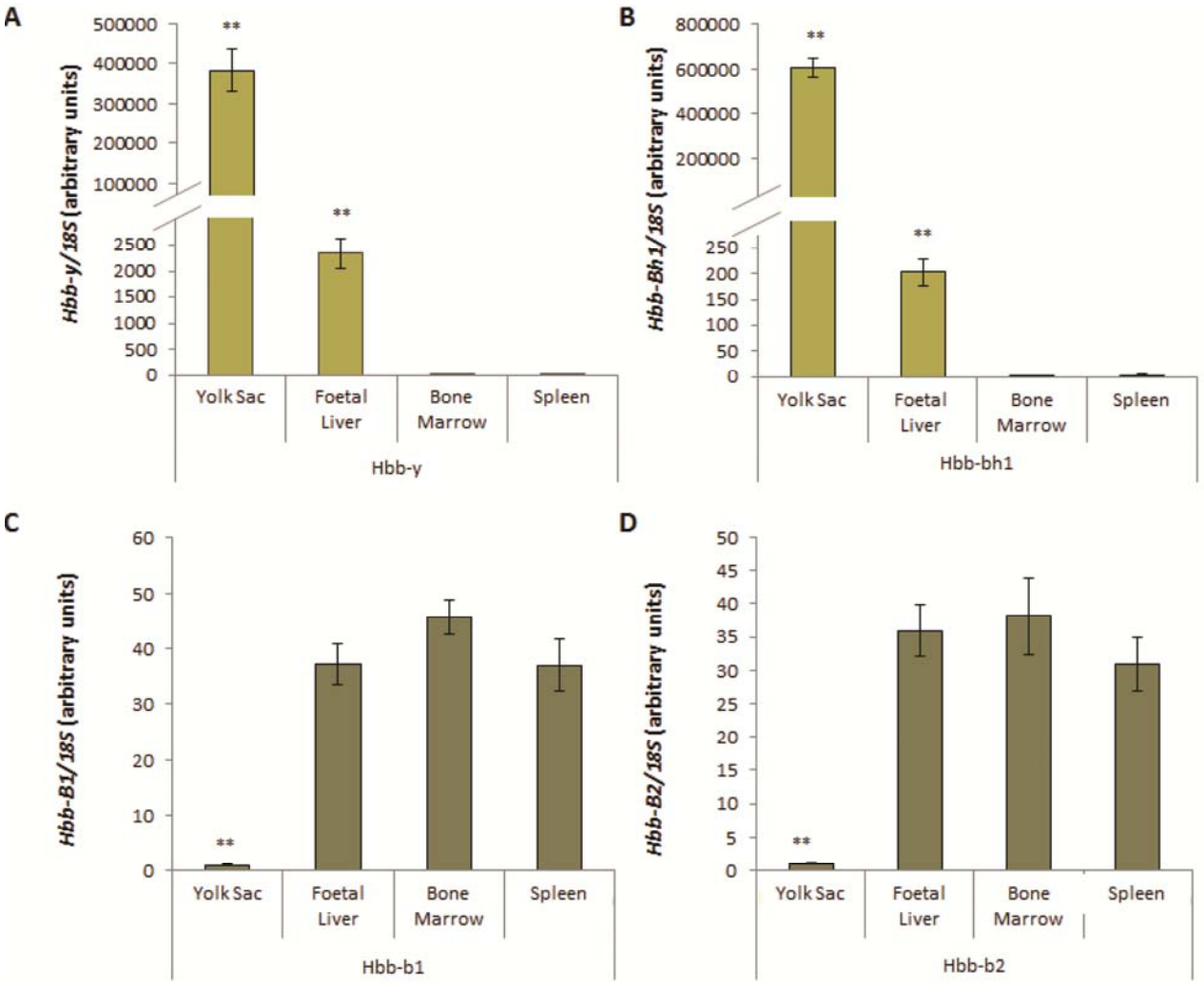


Figure 3.4. Embryonic and adult globin gene expression at various stages during development. Yolk sac was harvested at E10.5 and foetal livers from embryos at E14.5 embryos, and spleen and bone marrow from adult mice aged 10-12 weeks. These cells were then sorted with anti-TER119 Magnetic Microbeads. Whole RNA was extracted, and qPCR was used to investigate the expression of *Zbtb7a.* Levels were normalised to *18S* rRNA levels. Error bars represent standard error of the mean (n = 3 per tissue). **, $P < 0.01$ (paired Student's *t* test).

3.3 *ZBTB7A* is up-regulated during human erythroid differentiation in haematopoietic stem cells

Next we investigated whether *ZBTB7A* is up-regulated during erythroid differentiation. To achieve this, we used primary human CD34+ HSCs from the Mackenzie lab at the Children's Cancer Institute (NSW, Australia), which were harvested at the time of birth from cord blood. These foetal derived cells were differentiated towards the erythroid lineage using the protocol outline in Chapter 2 (Figure 2.1), and samples were collected for mRNA and protein analysis at weeks 1, 2 and 3 during differentiation. The first week of this culture is a stem cell expansion phase, where the CD34+ cells are cultured with cytokines involved in HSC

self-renewal, and at the end of the first week, these cells are still predominantly CD34+. Differentiation towards the erythroid lineage begins in the second week, and by the end of week 2, these cells resemble immature erythrocytes known as erythroblasts. The differentiation is continued into the third week, and by the end of week 3, they are equivalent to reticulocytes and are beginning to enucleate and terminally differentiate into mature erythrocytes [144].

We first investigated transcript levels of the *β-globin* genes to ensure differentiation was occurring as expected (Figure 3.5 A). In these samples it can be seen that foetal *γ-globin* was the most highly up-regulated globin gene, followed by adult *β-globin*. The embryonic globin, *ε-globin*, and other adult globin, *δ-globin*, were not significantly altered during this erythroid differentiation protocol. We also looked at the levels of various transcription factors during the differentiation period (Figure 3.5 B), namely *KLF1,* an important erythroid transcription factor, *KLF3,* an erythroid target of KLF1 [94, 166] also thought to play a role in globin switching [104], *BCL11A,* another KLF1 target gene [112] and established foetal globin silencer [62, 105-107, 110], and *ZBTB7A*. *KLF1* was found to be the most highly up-regulated transcription factor, followed by *KLF3*. *BCL11A* was up-regulated by 10-fold whereas *Zbtb7a* was up-regulated by 6-fold by the third week of differentiation.

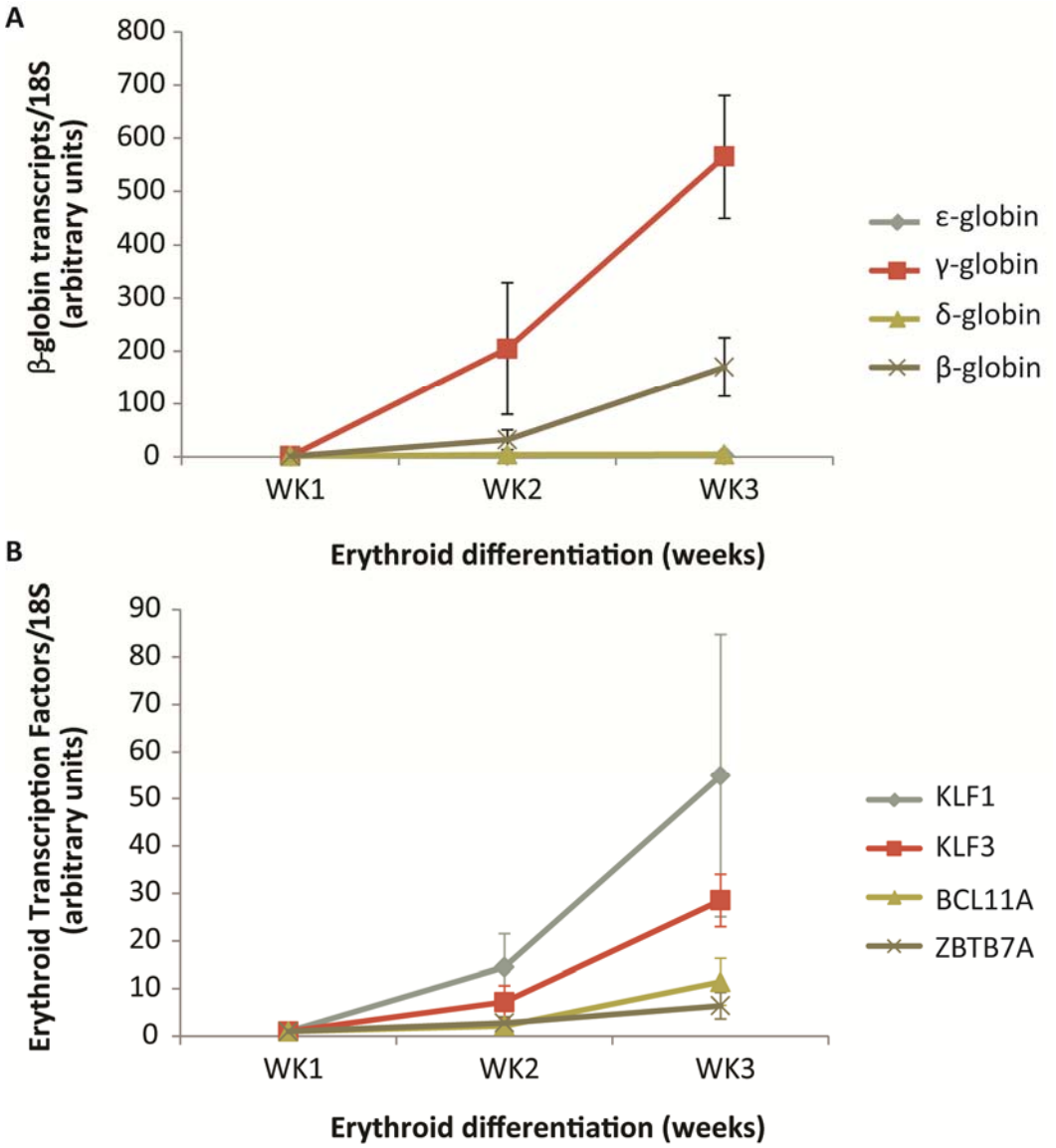


Figure 3.5. *ZBTB7A* transcript expression levels increase during erythroid differentiation of foetal human haematopoietic stem cells derived from cord blood. Expression of *β*-like globin genes **(A)** and erythroid transcription factors **(B)** during erythroid differentiation of CD34+ cord blood cells. RNA was extracted from CD34+ cord blood cells at week 1, 2 and 3 of differentiation. qPCR was used to investigate the expression of embryonic, foetal and adult *β-globin* genes (A), as well as erythroid transcription factors (B) *KLF1* and *KLF3*, the known foetal globin silencer *BCL11A* and *ZBTB7A*. Levels were normalised to *18S* rRNA levels. Error bars represent standard error of the mean (n = 5).

ZBTB7A protein expression was also investigated by Western blot analysis during the erythroid differentiation of these cells (Figure 3.6 A). There was a marked up-regulation of ZBTB7A at week 3 of differentiation compared to week 1. The levels of ZBTB7A at week 3 were significantly increased compared to both week 1 and week 2, using a two tailed *t* test ($P < 0.05$). Again, the assessment of protein level suggested that the difference in protein expression is greater than the difference in the up-regulation at the transcript level during differentiation (Figure 3.6 B).

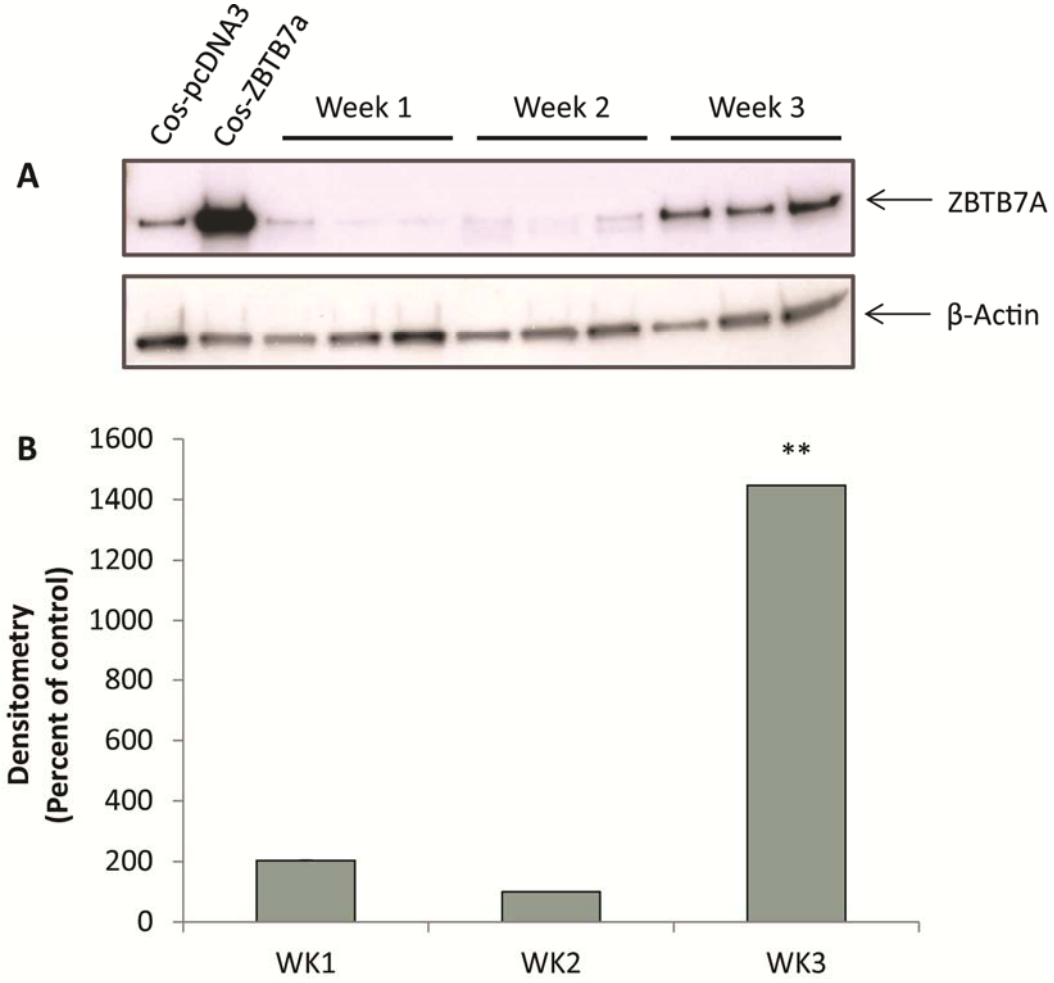


Figure 3.6. ZBTB7A protein expression levels increase during erythroid differentiation of foetal human haematopoietic stem cells derived from cord blood. **(A)** Nuclear extracts were prepared from mock-transfected Cos-7 cells (lane 1), and Cos-7 cells overexpressing ZBTB7A (lane 2), as well as CD34+ cells at week 1 (lanes 3-5), week 2 (lanes 6-8) and week 3 (lanes 9-10) of erythroid differentiation. ZBTB7A protein levels were assessed by Western Blot. Blots were probed with αZBTB7A and β-Actin was used as a loading control. **(B)** Densitometry analysis of Western Blot in (A) performed using ImageJ analysis software. Error bars represent the standard error of the mean (n = 3 per time point) **, $P < 0.01$ (paired Student's *t* test).

We repeated these experiments with cells derived from human adult bone marrow, purchased from Lonza and StemCell Technologies. These cells were differentiated following the same methodology as the cord blood cells and RNA was extracted at the same time points of week 1, 2 and 3 during erythroid differentiation. Experiments performed in these cells showed greater variability than those of cord blood, with a 10-fold difference in globin up-regulation seen between some experiments. The pattern of transcript expression changes, however, was similar in each of the three experiments. For this reason, data for each of the three experiments have been plotted separately, with *β*-like globin genes shown in the top graphs, and erythroid transcription factors in the bottom graphs (Figure 3.7). In all experiments the adult *β-globin* gene was the most highly up-regulated, followed by foetal

γ-globin. The other adult globin, *δ-globin*, was also slightly up-regulated, but as was the case with the cord blood samples, embryonic *ε-globin* expression was not increased during differentiation.

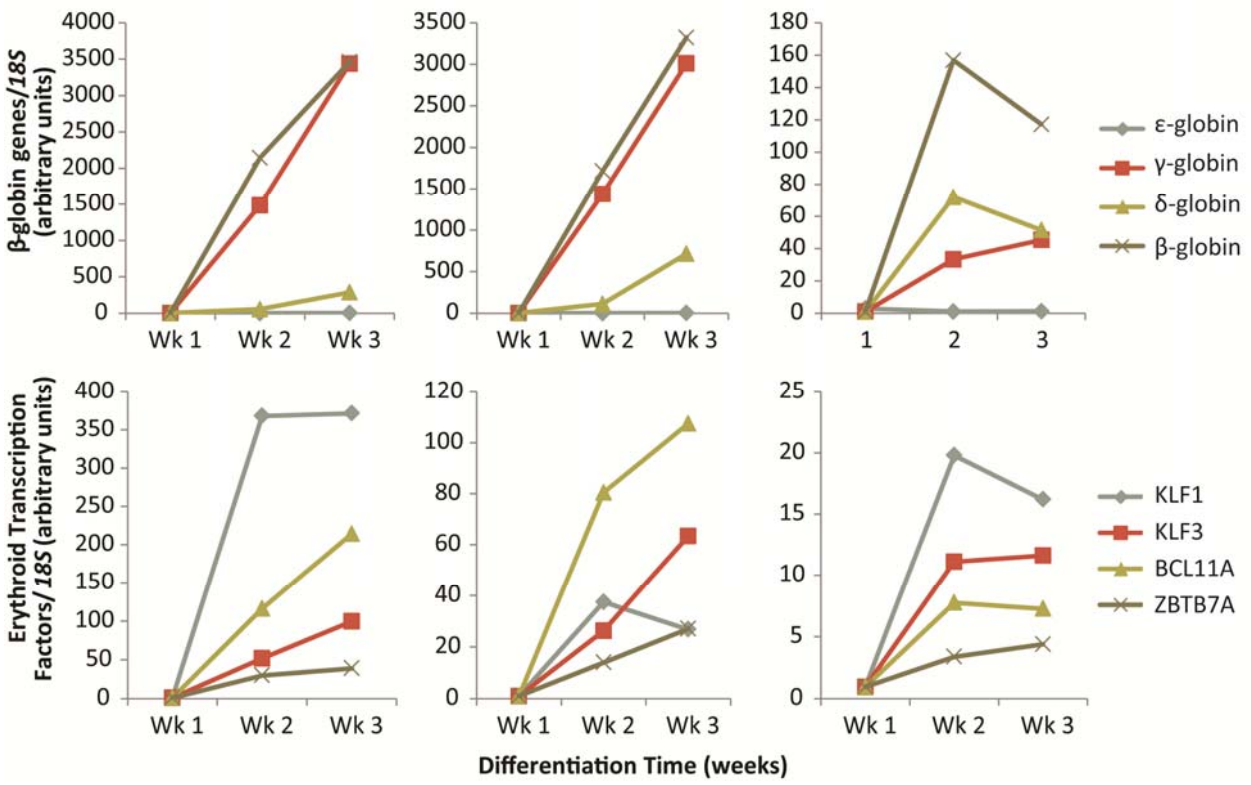


Figure 3.7. *ZBTB7A* transcript expression levels increase during erythroid differentiation of adult human haematopoietic stem cells derived from bone marrow. Expression of *β*-like globin genes **(A)** and erythroid transcription factors **(B)** during erythroid differentiation of CD34+ bone marrow cells. Three separate erythroid differentiations of CD34+ bone marrow cells were performed and RNA was extracted at week 1, 2 and 3 of differentiation. qPCR was used to investigate the expression of embryonic, foetal and adult *β-globin* genes (A), as well as erythroid transcription factors (B) *KLF1* and *KLF3*, the known foetal globin silencer *BCL11A* and *ZBTB7A*. Levels were normalised to *18S* rRNA levels. These three experiments are displayed on separate graphs as the level of up-regulation varied greatly between the samples.

It was not possible to obtain enough cells for protein extraction from these cultured human bone marrow samples, and therefore Western blot analysis was not performed on these cells.

3.4 ZBTB7A binding in the human genome

So far in this chapter, we have shown that ZBTB7A is capable of binding the -200 region of the *γ-globin* promoter *in vitro*, is expressed at the time of embryonic globin silencing in mouse and human cells, and is up-regulated at the transcript and protein level during human erythroid differentiation. We next sought to investigate the genomic occupancy of ZBTB7A *in vivo* in human erythroid cells using ChIP-qPCR and ChIP-Sequencing (ChIP-Seq). We

hypothesised that these experiments would provide further insights into the role of ZBTB7A in erythropoiesis.

We performed ChIP-qPCR and ChIP-Sequencing (ChIP-Seq) experiments in the human erythro-leukemia cell line, K562, which resemble foetal cells, as they express low levels of adult *β-globin* and higher levels of *γ-globin* [167]. We also examined adult primary erythroid cells, differentiated from CD34+ HSCs derived from bone marrow using the protocol described in Chapter 2 (Figure 2.2). ChIP experiments were performed and subsequent downstream analyses were carried out as outlined in Chapter 2.

3.4.1 ChIP-qPCR of ZBTB7A in human cells

Initially we performed qPCR to determine if the ChIP was successful and examine specific regions of interest in the genome. The regions that were investigated in the K562 experiment comprised a number of positive and negative controls, as well as various regions upstream of the *γ-globin* promoter, with the focus of the experiment being the -200 site of the *γ-globin* promoter. The promoter regions of *KLF1* and the apoptotic factor *BCL2L11*, which are known ZBTB7A target genes [137, 164] were used as positive controls to assess whether the ChIP was successful. Regions of the *β-globin* gene cluster, namely the embryonic *ε*-, and adult *β-globin* promoters, were used as negative controls for the experiment to determine levels that were essentially equivalent to background. Regions upstream of the *γ-globin* promoter, labelled -1.1 kb, -2.1 kb and -3.1 kb, were also used as negative controls. The particular region of interest was the -200 region of the *γ-globin* promoter. The ChIP-qPCR in the K562 cell line showed significant enrichment of ZBTB7A at the promoter regions for *KLF1* ($P < 0.05$) and *BCL2L11* ($P < 0.01$), suggesting successful pull-down of DNA fragments bound to ZBTB7A. Both *ε-globin* and adult *β-globin* showed little enrichment above IgG levels. Significant enrichment of ZBTB7A ($P < 0.01$) was detected using primers centred at the -200 site of the foetal *γ-globin* promoter, which was greater than the background level detected at regions upstream of the *γ-globin* promoter (-1.1 kb, -2.1 kb, -3.1 kb) (Figure 3.8). As the ChIP was successful, and we saw enrichment at the -200 region of the *γ-globin* promoter, we performed ChIP-Seq experiments in these cells, described in the following section.

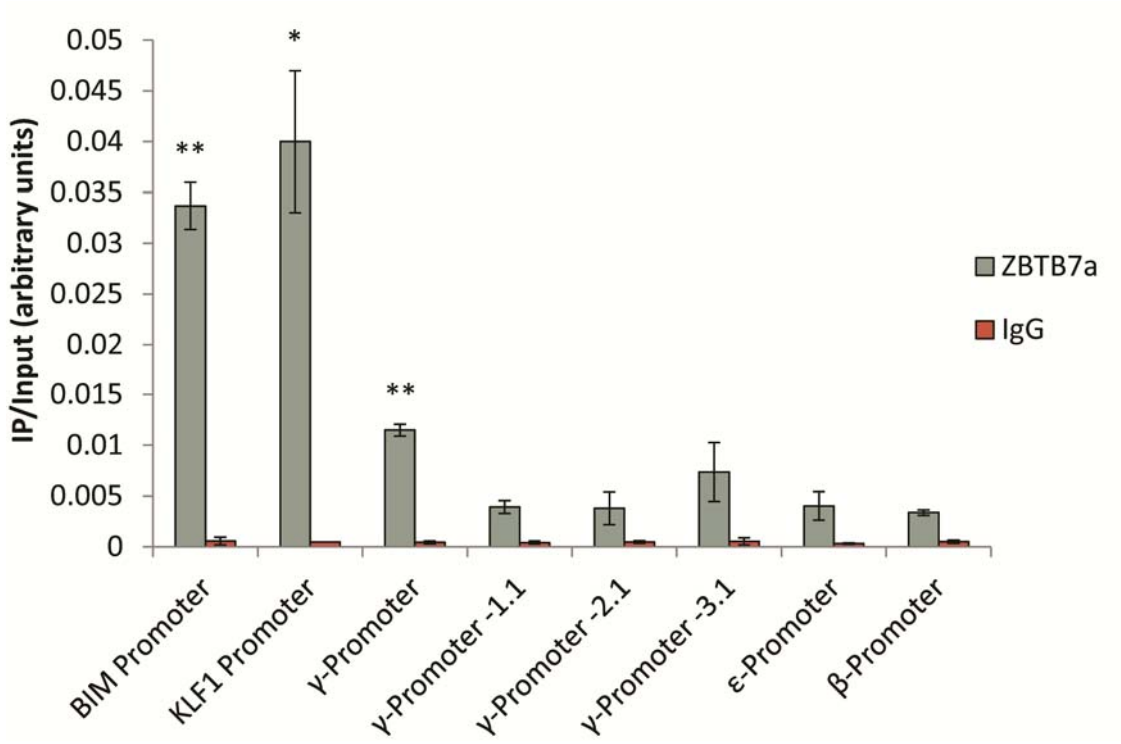


Figure 3.8. ChIP-qPCR analysis shows ZBTB7A binds to the -200 site of the *γ-globin* promoter in the K562 cell line. ChIP was performed using an anti-ZBTB7A antibody on K562 cells, and qPCR was used to analyse the enrichment of ZBTB7A at various loci. The tested genomic loci include the *BCL2L11* and *KLF1* proximal promoters as positive controls, as well as negative areas at the *ε-globin* and *β-globin* proximal promoters. Other controls include -1.1, -2.1 and -3.1 kb upstream of the *γ-globin* gene promoter. IP was normalised against input and error bars represent standard error of the mean (n = 2). *, $P < 0.05$, **, $P < 0.01$ (paired Student's *t* test).

We also performed ChIP-qPCR experiments in duplicate on CD34+ bone marrow cells differentiated towards the erythroid lineage using the fast method, described in Chapter 2 (Figure 2.2). We harvested cells at Day 7 (Figure 3.9 A) and Day 11 (Figure 3.9 B) of differentiation. In this experiment we only looked at a small subset of the targets investigated previously to test whether the ChIP was successful as the amount of material was limited. We examined enrichment of ZBTB7A at one positive control, *BCL2L11*, one negative control, *ε-globin*, and the -200 region of the *γ-globin* promoters. Here we saw significant enrichment at the promoter for *BCL2L11*, and no enrichment at the *ε-globin* promoter, indicating that the ChIP was successful in these cells. While the enrichment at the -200 region of the *γ-globin* promoter is marginal compared to the K562 ChIP-qPCR above, we were encouraged that the ChIP was successful and continued with a ChIP-Seq using these samples.

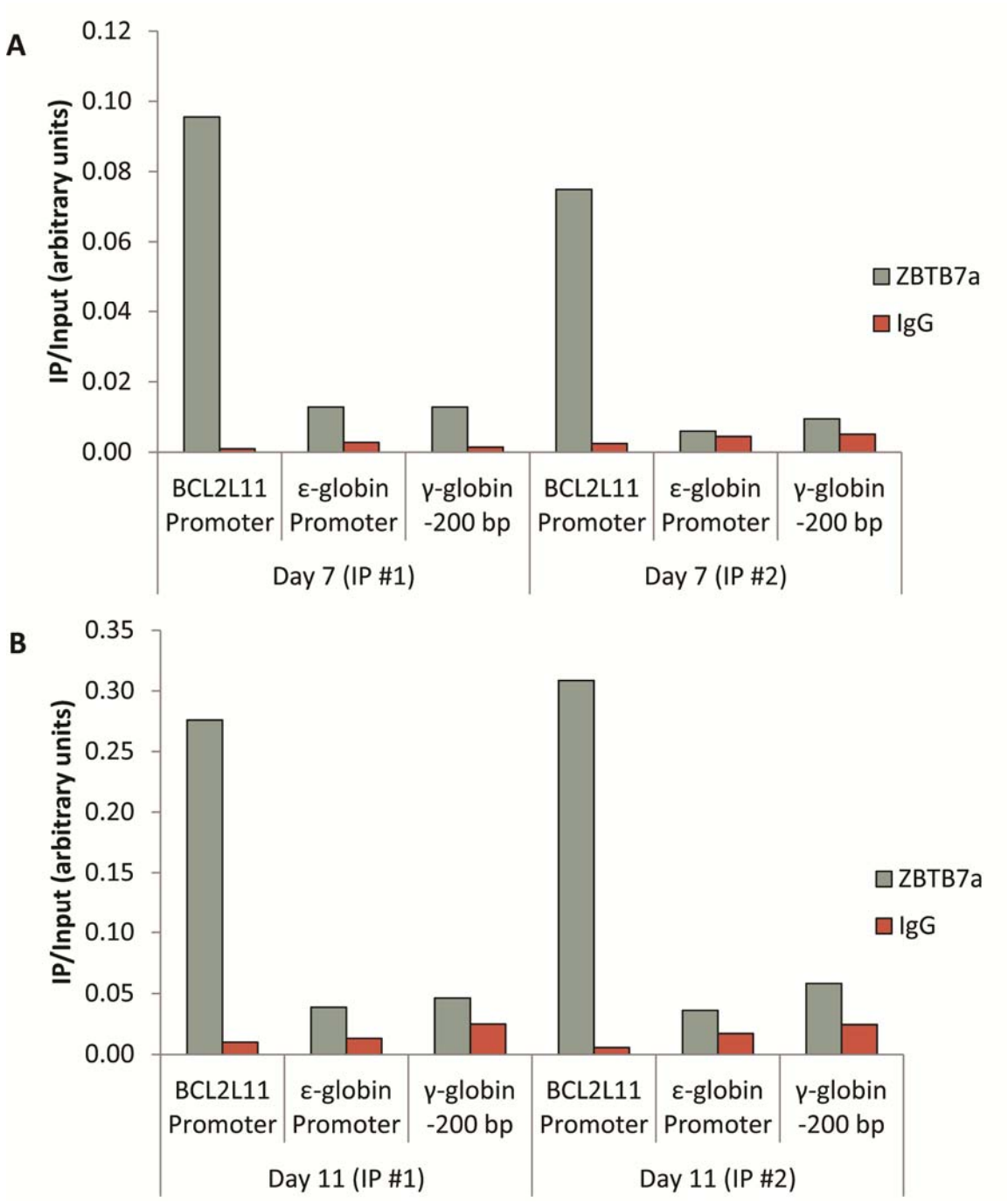


Figure 3.9. ChIP-qPCR in primary human adult erythroid cells demonstrated that the ChIP was successful but did not provide convincing evidence for ZBTB7A occupancy at the *γ-globin* promoter. CD34+ bone marrow cells were differentiated towards the erythroid lineage in duplicate, and harvested at **(A)** Day 7 and **(B)** Day 11. ChIP was performed using an anti-ZBTB7A antibody, and qPCR was used to analyse the enrichment of ZBTB7A at various loci. The tested genomic loci include the *BCL2L11* as a positive control, as well as a negative control at the *ε-globin* proximal promoters. The particular region of interest was the -200 region of the *γ-globin* promoter. IPs were normalised against inputs.

3.4.2 ChIP-Seq of ZBTB7A in human cells

ZBTB7A ChIP-Seq was performed in four replicates of K562 cells, and duplicates of primary human CD34+ bone marrow samples differentiated for 7 and 11 days down the erythroid lineage. The actual wetlab ChIP-Seq experiment in the case of the K562 cells was performed in our lab by A. Funnell. Alignments, trimming, peak calling, and annotation of peaks, were performed by J. Burdach, as outlined in Chapter 2 and conforming to the guidelines for ChIP-Seq analysis outlined by the ENCODE consortium [168].

Firstly, we determined broadly which genomic regions ZBTB7A was bound to in the each of the experiments (Figure 3.10). Across all the experiments, ZBTB7A was bound mainly to promoters, defined as 1,000 bp downstream (-1,000 bp) to 100 bp upstream (+100 bp) of the TSS. ZBTB7A was also found at many intronic and intergenic regions. Regions termed "other" include 3'- and 5'UTRs, exonic regions and transcriptional termination site regions. As intergenic regions are fairly abundant in the genome, we would expect a large number of binding sites to lie in these regions. Introns can also be very large, and the high number of binding sites in these regions is also to be expected. We found that, across all three datasets, at least 30% of ZBTB7A binding sites are found at promoter regions. According to GENCODE and the Human Genome Project, the human genome contains roughly 20,000 genes [169] spread over 3 billion bp [170]. The annotation used for this experiment dictates that promoter regions extend for 1, 100 bp (from -1, 000 bp to +100 bp of the TSS). In order to find the percentage of the genome devoted to promoter regions in this annotation, the length of the defined promoter region (1,100 bp) was multiplied by the number of genes, (20,000) and divided by the number of base pairs of the genome (3 billion). This means that 0.7% of the human genome is annotated as a promoter in this analysis. The large number of binding sites at promoter regions, therefore, is highly enriched above their representation in the genome. This is an encouraging finding - as ZBTB7A is a transcription factor, we would expect to find it at promoters of genes, which gives us confidence that we have successfully pulled down regions bound to ZBTB7A in all three datasets.

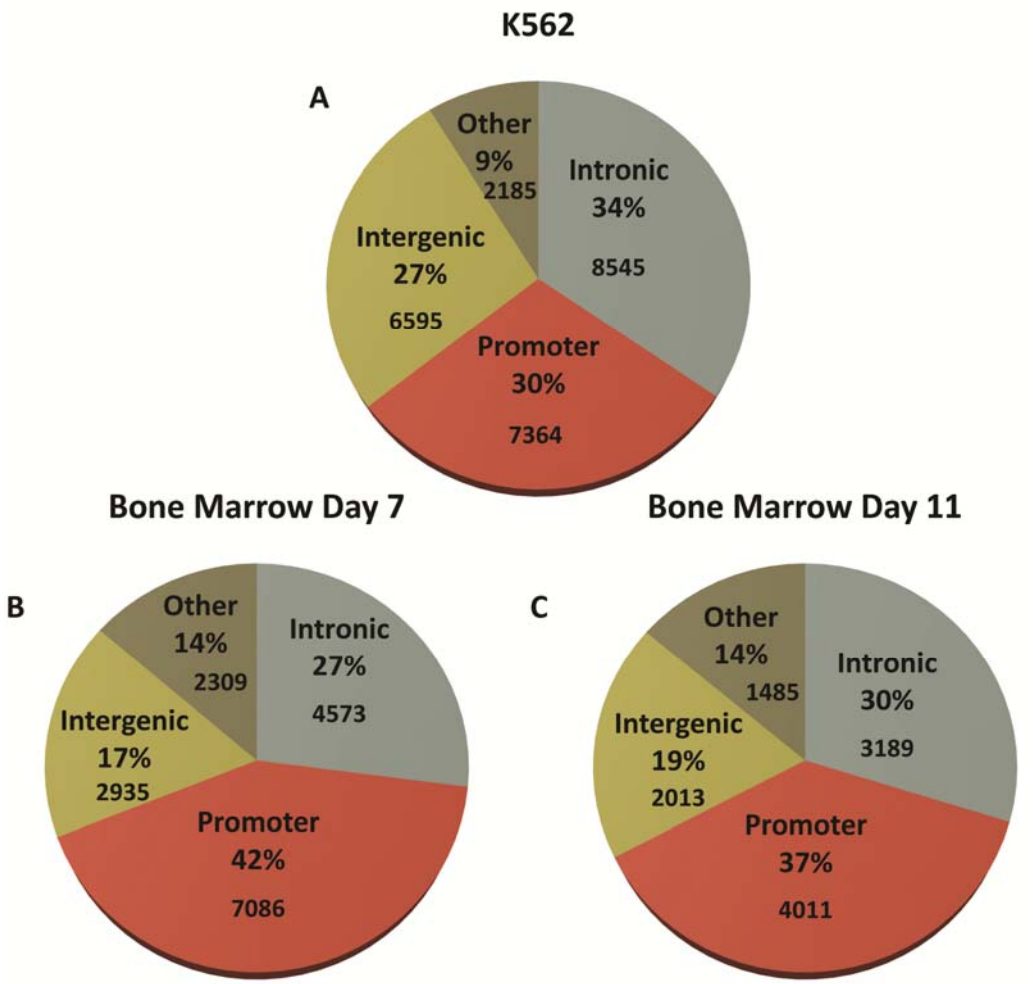
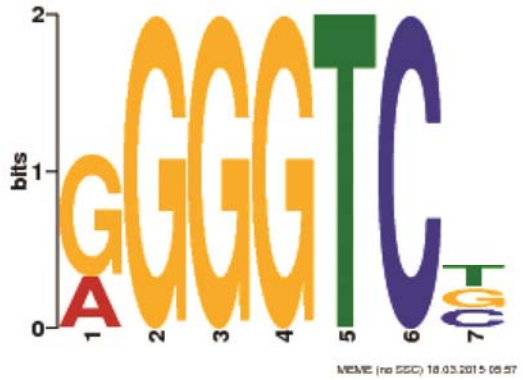


Figure 3.10. Percentage of ZBTB7A binding to regions of the genome in K562 and bone marrow cells. ChIP-Seq analysis of ZBTB7A binding was performed on four replicates of K562 cells. Annotation of peaks was performed using HOMER. ZBTB7A is found to bind mostly intronic regions, followed by promoters and intergenic regions. Other regions are composed of 3'- and 5'UTRs, exonic regions and TTSs.

Next we analysed the consensus binding motifs of ZBTB7A using the MEME-ChIP online tool [157-159]. We used the top 800 peaks from each dataset for this analysis. We found that the most enriched motif in the K562 experiment pulled out by DREME (Figure 3.11), was very similar to the already published ZBTB7A binding motif of 5-G(A/G)GGG(T/C)(C/T)(T/C)(C/T)-3 [139], which, as previously mentioned, closely resembles the -200 site of the *γ-globin* promoter. This motif we pulled out was found in 407 of the 800 peaks submitted. This motif was also highly enriched in the bone marrow samples, but was not found in as many of the top 800 peaks as it was for the K562 sample. It was found in 363 out of 800 peaks at Day 7, and 261 out of 800 at Day 11. The DREME analysis also pulled out a GATA-binding motif as the second most enriched motif in the K562 experiment, as well as bone marrow at Day 7. This motif was found to be the most highly enriched in ZBTB7A peaks in the Day 11 samples. This is interesting as ZBTB7A has been found to bind many of the same promoters as GATA-1 [164].



Sample	Peaks / 800	P Value	E Value
K562	488	4.9e-37	1.5e-32
BM Day 7	394	2.1e-28	6e-24
BM Day 11	270	9.6e-14	2.7e-9

Sample	Peaks / 800	P Value	E Value
K562	108	2e-20	5.7e-16
BM Day 7	95	9.61e-21	2.7e-16
BM Day 11	144	2.8e-25	8.1e-25

Figure 3.11. ZBTB7A DNA binding consensus motifs discovered using online MEME-ChIP tool. The locations of the top 800 peaks were uploaded into the online Galaxy online tool, and FASTA sequences were obtained. These sequences were then used for MEME-ChIP analysis. The peak count figure was determined by the software after disregarding sites that match previously found motifs. The *E* value is the number of candidate motifs tested multiplied by the *P* value, which was calculated by the software using Fisher's Exact Test for enrichment of the motif in the positive sequences.

While it is likely that ZBTB7A affects distally located genes from its intergenic binding sites, or even from intronic, exonic and UTR regions, genes where ZBTB7A is bound to the promoter are more likely to be under the direct regulation of ZBTB7A, and are certainly the easiest to test. We were interested where in these promoter regions ZBTB7A was mostly found in relation to the TSS of the associated genes. The data shown in Figure 3.12 indicated that ZBTB7A was found at or adjacent to the TSS in all three experiments, and appears to bind preferentially to regions located between 100 bp upstream and 50 bp downstream of the TSS.

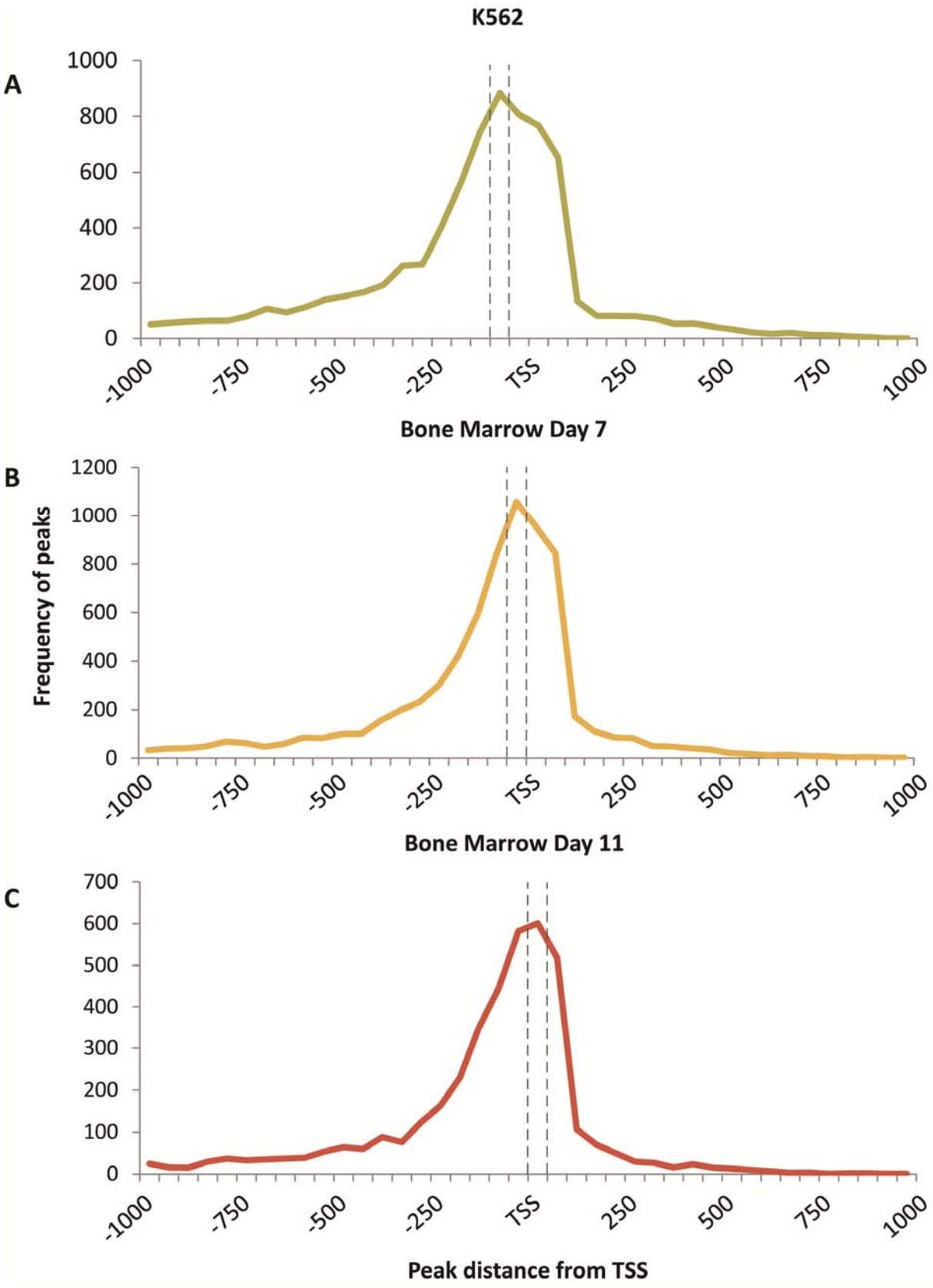


Figure 3.12. ZBTB7A largely binds regions close to TSS. ChIP-Seq analysis of ZBTB7A binding was performed on four replicates of K562 cells. Annotation of peaks was performed using HOMER. Distance to TSS was plotted against number of sites in each region of 50 bp from -1, 000 kb to 1, 000 kb around the TSS. ZBTB7A is primarily bound to regions close to the TSS, with a peak at 50-100 bp upstream.

Next, we examined the binding at key erythroid genes using the Interactive Genome Viewer (IGV) v2.3. We looked at genes which have previously been described as ZBTB7A targets, *KLF1* and the apoptotic factor *BCL2L11* (Figure 3.13), and additional interesting erythroid genes, namely *GATA-1* and *BCL11A*. As many factors are involved in regulating themselves, we also investigated the promoter of *ZBTB7A* (Figure 3.14). We found ZBTB7A binds to the promoter regions of both *KLF1* and *BCL2L11*, as previously described [137, 164]. But we also see evidence of ZBTB7A binding to a region slightly upstream of the promoter of *GATA-1,* which may be a distal enhancer/silencer site, and to the promoter of *BCL11A*, as well as its own promoter.

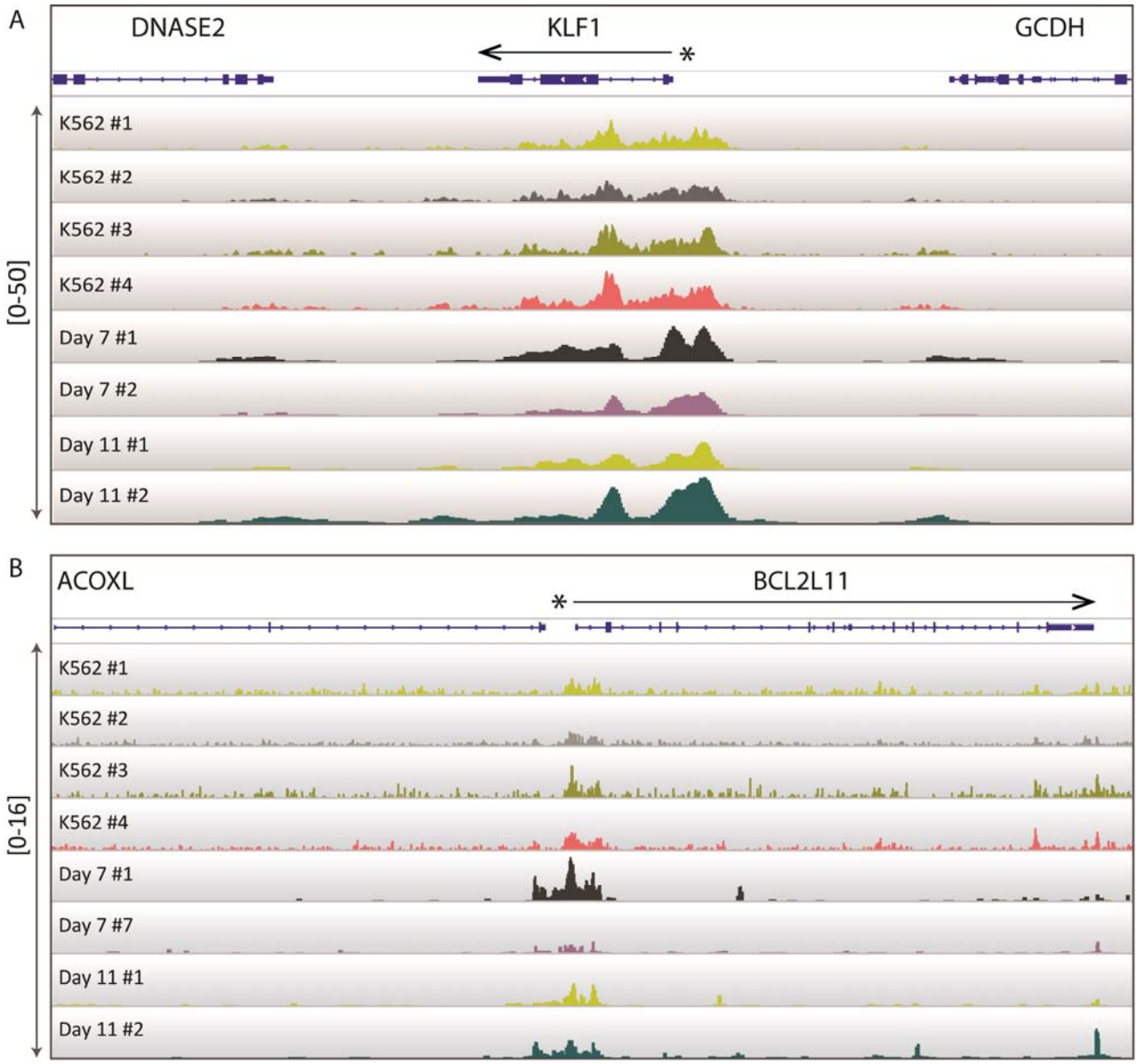


Figure 3.13. ZBTB7A binds to the promoters of *KLF1* and *BCL2L11* in human cells. ChIP-Seq peaks were viewed using IGV v2.3 and indicate that ZBTB7A binds to the promoters of both *KLF1* **(A)** and *BCL2L11* **(B)**. This binding is seen across all K562 and bone marrow samples. The asterisks indicate the positon of the promoters and the arrows indicate the direction of translation.

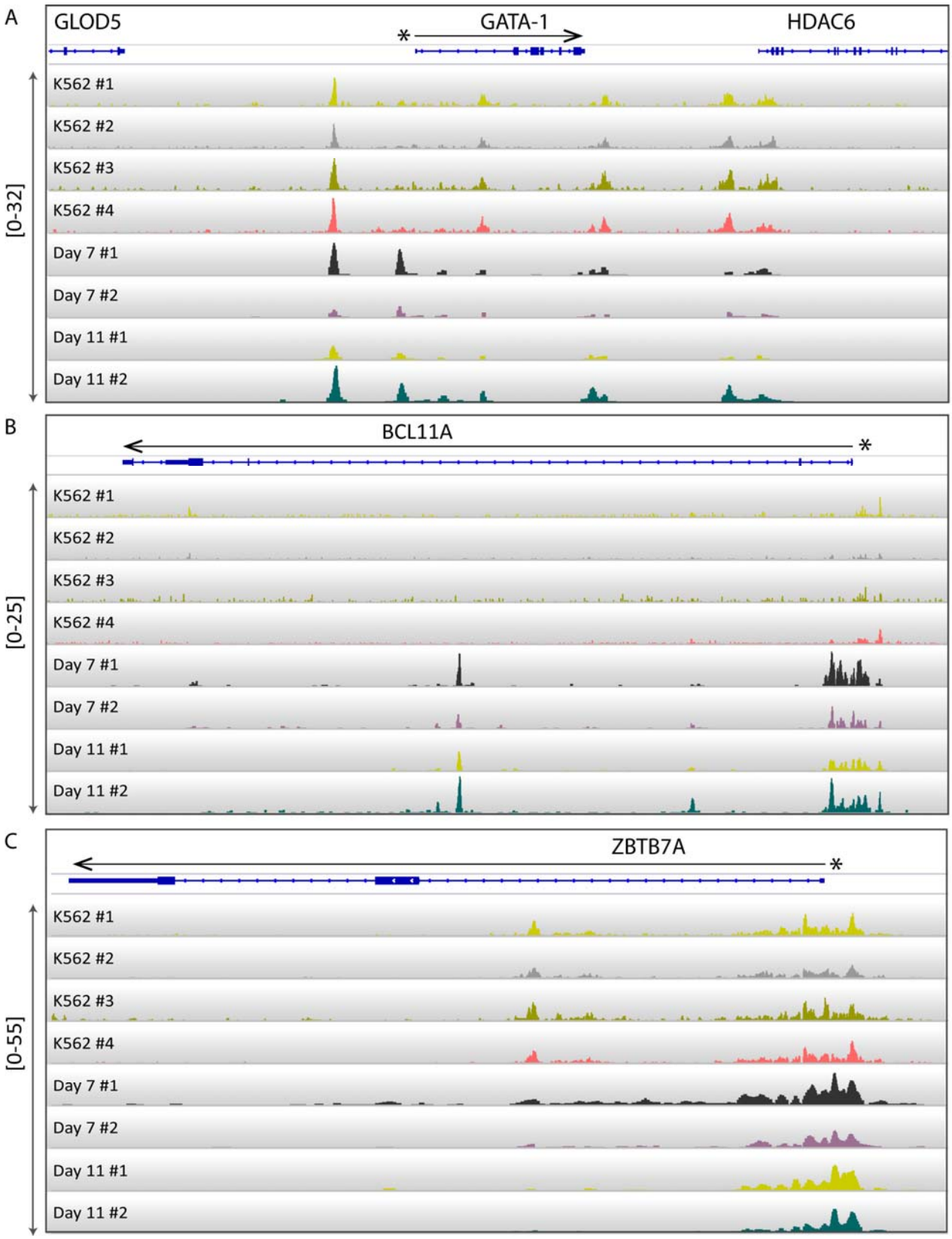


Figure 3.14. ZBTB7A binds to regulatory regions of *GATA-1*, *BCL11A* and *ZBTB7A* in human cells. ChIP-Seq peaks were viewed using IGV v2.3 and indicate that ZBTB7A binds to a region upstream of the promoter of *GATA-1* **(A)**, and at the promoters of *BCL11A* **(B)** and *ZBTB7A* **(C)**. This binding is seen across all K562 and bone marrow samples. The asterisks indicate the positon of the promoters and the arrows indicate the direction of translation.

We were chiefly interested in assessing the binding of ZBTB7A to the *β-globin* locus and particularly the -200 site of the *γ-globin* promoters. We found that ZBTB7A binding was present at a region upstream of the embryonic *ε-globin* gene in both K562 and bone marrow samples. There appeared to be more binding to this region in the bone marrow samples, and these displayed four discreet peaks, which were enriched in the Day 11 samples. This region is the *β*-LCR, consisting of five DNase hypersensitive sites, involved in facilitating long-range interactions to various regions of the *β*-globin locus to regulate these genes. The bone marrow samples also displayed a peak at the adult *β-globin* gene, which is also enriched in at the later time point. The K562 samples showed a peak at the -200 region of both *γ-globin* promoters, however this was not seen in the bone marrow samples (Figure 3.15).

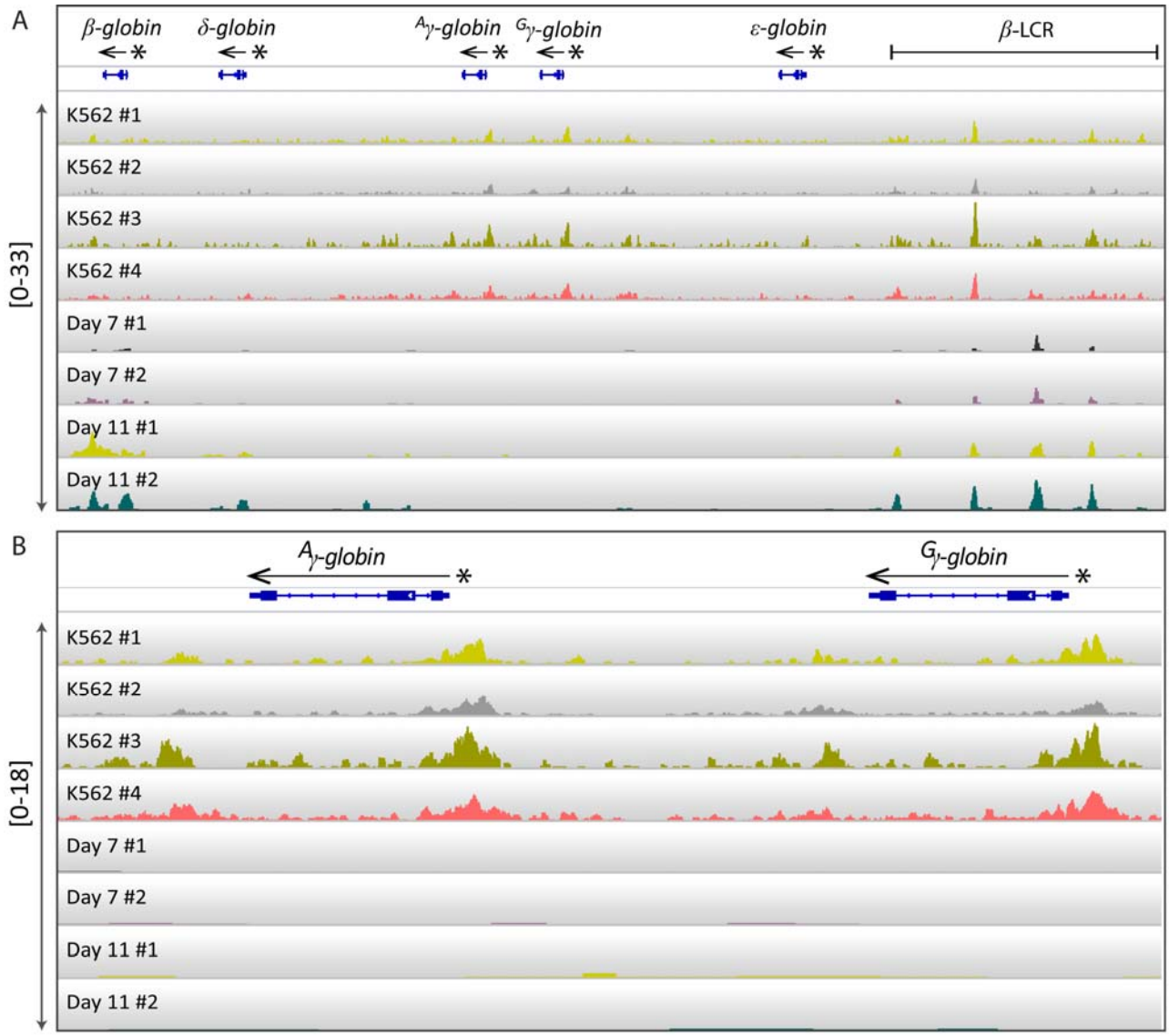


Figure 3.15. ZBTB7A binds to regulatory regions of the *β*-globin locus. ChIP-Seq data revealed that ZBTB7A binds to the DNase Hypersensitive sites of the β-Locus Control Region (LCR) across both K562 and bone marrow samples **(A)**, as well as the -200 site of the both the *γ-globin* promoters in all K562 replicates **(B)**. The asterisks indicate the positon of the promoters and the arrows indicate the direction of translation.

3.4.3 ChIP-qPCR of ZBTB7A in primary foetal erythroid cells

As the enrichment of ZBTB7A at the -200 region in the K562 cells is modest, it is difficult to conclude from the bone marrow experiments whether ZBTB7A is simply not bound to this region, or whether the experiment, for whatever reason, was not efficient enough to detect the binding. As the K562 cells are more foetal in nature, we hypothesised that we would be able to detect binding of ZBTB7A to the *γ-globin* locus more readily in foetal cells. Thus, we performed initial ChIP-qPCR experiments on foetal primary human erythroid cells, differentiated from CD34+ HSCs derived from cord blood (Figure 3.16), using the slow method described in Chapter 2 (Figure 2.1). A similar pattern of expression was seen as in the K562 experiments, with ZBTB7A enriched at the *KLF1* and *BCL2L11* promoters, as well as the -200 region of the foetal *γ-globin* promoter. Levels detected at -1.1 kb and -2.1 kb regions upstream of the *γ-globin* promoter, as well as the negative control region of the promoters of embryonic *ε-globin* and adult *β-globin* were at background levels. We saw greater enrichment at the -3.1 kb region. Obtaining enough cells to perform a ChIP-qPCR experiment using this method was challenging, and therefore only one replicate is shown. As this result is more comparable to the results obtained for the K562 ChIP-qPCR experiments, we plan to perform ChIP-Seq experiments in this cell type in the future.

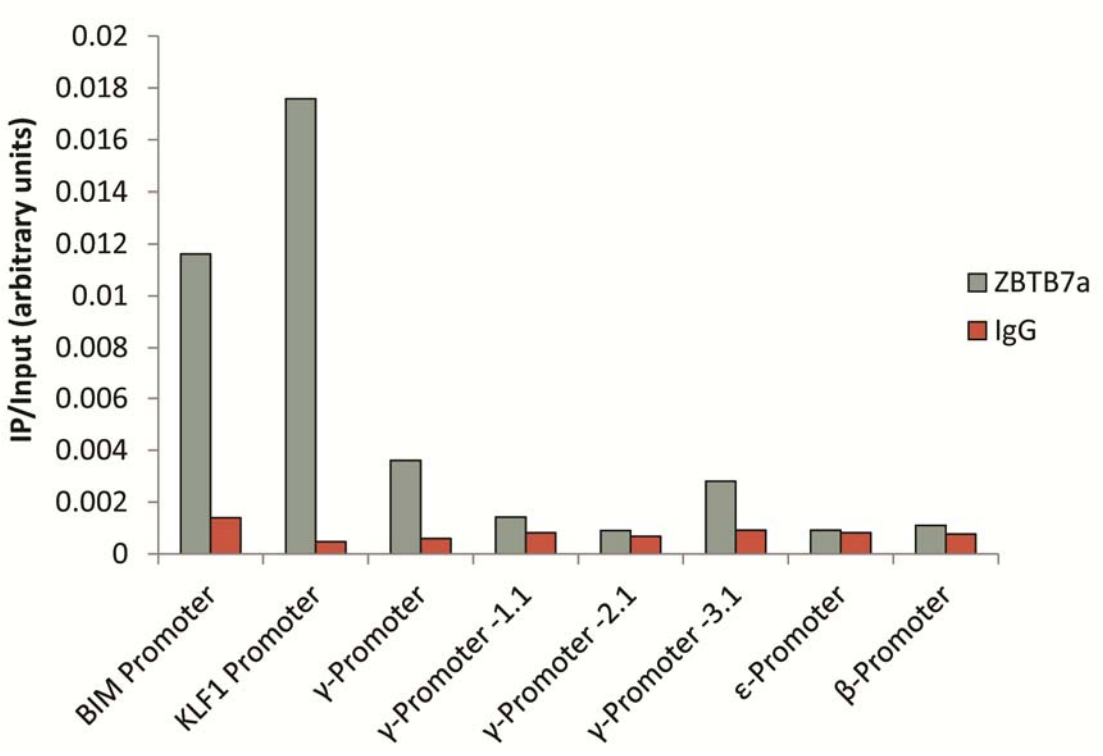


Figure 3.16. ChIP-qPCR analysis shows ZBTB7A binds to the -200 site of the *γ-globin* promoter in erythroid differentiated human primary haematopoietic stem cells derived from cord blood cells. The tested genomic loci include the *BCL2L11* and *KLF1* proximal promoters as positive controls, as well as negative areas at the *ε-globin* and *β-globin* proximal promoters. Other controls include -1.1, -2.1 and -3.1 kb downstream of the *γ-globin* gene promoter. IP was normalised against input. A representative plot is shown.

3.5 Chapter 3 Discussion

Previously published data has shown that ZBTB7A is important for normal erythroid development in mice [137, 164] and preliminary microarray experiments in HSCs have implicated it in repression of the embryonic *βh1-globin* gene [138], an orthologue of the human foetal *γ-globin* gene [165]. The DNA-binding consensus sequence of ZBTB7A [139] closely resembles the -200 region of the foetal *γ-globin* promoter. Previously, our lab provided the initial evidence of ZBTB7A binding to the -200 site of the foetal globin promoter in *in vitro* EMSA studies [140]. Here we investigated the role of ZBTB7A *in vivo* in the mouse as well as in human cells.

We found that *Zbtb7a* was most highly expressed in erythroid tissues, consistent with the view that it has an important role in these cells in particular. *Zbtb7a* expression was, however, also detected in the non-erythroid cells. *Zbtb7a* has been found to have roles in various cancers both as a proto-oncogene [171-175] and more recently as a tumour suppressor [176, 177]. This suggests the precise regulation of *Zbtb7a* levels is of critical importance in many different cell types. We also found that *Zbtb7a* was most highly expressed at E14.5 in the foetal liver, which correlates well with the embryonic lethality of the *Zbtb7a*$^{-/-}$ mice at day E15.5 due to severe anaemia. There was also a correlation between high *Zbtb7a* levels and embryonic globin gene silencing. Increased levels of *Zbtb7a* at E14.5 correspond to the time of down-regulation of the transcripts of the embryonic globins, *εy* and *βh1*. This provided anecdotal evidence of ZBTB7A's involvement in silencing these genes. As *Zbtb7a* expression was higher at the time of embryonic globin silencing, we suggest that its up-regulation, at least in part, is to assist in repressing these genes. Once they have been silenced, there would no longer be a need for the high levels of ZBTB7A. At later time points we saw *Zbtb7a* down-regulated to basal levels which may be required for normal erythroid development and maturation. The basal level in these tissues was still higher than expression seen in non-erythroid tissues, in line with an important erythroid-specific role for ZBTB7A.

We also found *ZBTB7A* expression was up-regulated during human erythroid differentiation in cells from both foetal and adult origin. In the foetal cells, derived from cord blood, we found a large up-regulation of the foetal *γ-globin*, as well as a modest up-regulation in the adult *β-globin* gene. As these cells were harvested at the time of birth from the umbilical cord, which is around the time that globin switching begins to occur in humans (Figure 1.2), we would expect to see the expression of both foetal and adult globins in the mixed population of cells. We also see a modest up-regulation of the foetal silencer *BCL11A*,

and *ZBTB7A*. The up-regulation of ZBTB7A seen at the protein level was significantly higher than that of the transcript. This is possibly indicative of the sensitivity of qPCR, where the low levels of transcript are readily detected. Protein detection by Western Blot is less robust. The high up-regulation of *γ-globin* in these cells suggests it is not being repressed, and may be explained by the low up-regulation seen in transcripts of the repressors *BCL11A* and *ZBTB7A*.

A similar pattern of globin and erythroid transcription factor expression was seen in the cells derived from bone marrow. However, in this case, as expected, the adult *β-globin* was most highly up-regulated. We also found a small up-regulation of the other adult globin gene, *δ-globin*, a gene whose expression is not normally as high as *β-globin*. It is interesting here that *γ-globin* was also highly up-regulated, in some experiments by the same amount as *β-globin.* This result fits with previous reports of variable expression in these cells [120], although in our experiments, the level of foetal globin was particularly high. The reason for this is unknown, and we can only speculate. There have been a number of studies published which suggest that when cells are stressed, they activate *γ-globin* [59, 178, 179], and perhaps this was occurring in these cells. This again may be accounted for by the observation that the expression of repressors *BCL11A* and *ZBTB7A* was low in these cells, resulting in foetal globin not being fully silenced in these cells.

We performed ChIP-qPCR in K562 cells. The enrichment seen at the -200 site, while statistically significant above background, was substantially lower than that seen at the positive controls of the *KLF1* and *BCL2L11* promoters. As the predominant function of ZBTB7A described in the literature is as a repressor, it is interesting that ZBTB7A was highly enriched at the *KLF1* promoter, which was expressed highly in these cells. However, while ZBTB7A has been found to bind to the *KLF1* promoter *in vivo* [164], it has not been shown to directly silence its expression. ZBTB7A has been shown to be capable of activating expression of some of its target genes, for instance in conjunction with SREBP-1 it activate the expression of the fatty acid synthase gene, FASN [180]. It is possible then, that ZBTB7A, acting with some unknown co-activator or other activating transcription factors, is enhancing the expression of KLF1 in these cells. We believe it is not involved in activation at the -200 site due to our *in vitro* studies where we found binding to this region is abolished by known HPFH mutations which lead to up-regulation of *γ-globin*. While the enrichment of ZBTB7A at the -200 site in K562 cells is modest, this ChIP-qPCR experiment has been consistently repeated a number of times in our lab and we propose it is indicative of ZBTB7A binding to this region.

We also investigated the genome-wide binding of ZBTB7A in primary human cells using ChIP-Seq experiments. Here we see that ZBTB7A binds mostly at promoters, introns, and intergenic regions, with strong enrichment at promoters compared to their representation in the genome. As ZBTB7A is a transcription factor, this enrichment at promoters is expected. The binding of ZBTB7A to non-promoter regions was also detected. While it may be playing a regulatory role from distal enhancers/silencers at intergenic and even intronic regions, this is much harder to assess as the understanding of long-range transcriptional regulation at a genome wide scale is limited. It would be interesting to look at these promoter peaks in conjunction with microarrays, or RNA-Seq experiments performed in *ZBTB7A*$^{-/-}$ cells compared to WT K562s to see if these promoters, and possibly intronic, peaks correlate with a change in expression of the genes.

In this experiment we also pulled out an enriched binding motif of ZBTB7A, which was found to be very similar to the one published which resembles the -200 region of the *γ-globin* promoter. Interestingly, a GATA site was also found to be highly enriched, and was the most prevalent motif in the bone marrow samples at Day 11. Perhaps as cells mature towards the erythroid lineage, ZBTB7A is recruited to GATA-sites by factors which are up-regulated in an erythroid-specific manner, such as SCL/TAL1, often found to bind near GATA-1 [181]. As GATA-sites are prevalent in the promoters of many erythroid genes, it may be through this mechanism that ZBTB7A is able to act at its erythroid targets to ensure proper erythropoiesis. We also see a number of peaks at genes which have been shown to be regulated by ZBTB7A such as *BCL2L11* and *KLF1*. We also find a peak at the promoter regions of both the *γ-globin* genes, but not any other regions of the *β-globin* locus in the K562 replicates. These peaks are quite small, which might be due to the fact that K562 cells express large amount of *γ-globin* and that the binding of repressors such as ZBTB7A is limited at the developmental time point captured in these cells. We found no enrichment of ZBTB7A at the -200 region in the bone marrow experiments, but did see peaks at other regions in the locus such as the LCR and adult *β-globin* gene. It is possible that ZBTB7A is involved in looping interactions to various regions in the *β*-locus to affect expression of different genes.

Initial analysis of the ZBTB7A ChIP-Seq data, revealed that the ChIP-Seq experiments performed in the CD34 bone marrow cells lower signal to noise ratios that those in the K562 experiment. This could be due to a number of factors. There may simply have not been as much material used in the CD34 experiments. While cell counts are carried out to ensure the same numbers of cells across experiments, these cells are differentiating towards the erythroid lineage, and are doing so quite rapidly, and some cells may have already enucleated

at the time points chosen, resulting in less genomic material in the samples. Another possibility is that ZBTB7A is not expressed as highly as in K562 cells at these time points. Finally, as explained below, it is possible that it is harder to detect a repressor at its target genes.

The reasons for this difficulty are two-fold. Firstly, if the gene is expressed, is it likely that the repressor would be bound to this region? On the other hand, if the gene is successfully silenced, the repressor may have initiated conformational changes in the chromatin, resulting in tightly packaged heterochromatin. In this case, if the repressor is no longer required for active repression of the gene, is it still likely to be bound to this site? As we see a small peak of binding to these regions while *γ-globin* is expressed in K562 cells, we believe that it is an indication that ZBTB7A is able to bind to this region, and under the right conditions, or the correct time point, binding will be significantly enriched. We were able to show binding of ZBTB7A to similar levels seen in K562 cells in primary erythroid cells derived from cord blood. We hypothesise that this is because switching is actively occurring in these cells and ZBTB7A is being recruited to the promoter of the foetal globin genes to aid in their repression. We therefore will be performing ChIP-Seq experiments in these cells in the future.

Our lab has determined the ability of ZBTB7A to bind to the -200 site of the *γ-globin* promoter *in vitro*, and shown that this binding is abolished by mutations that have been found to cause up-regulation of foetal globin, resulting in HPFH [140]. In this chapter we showed enrichment of ZBTB7A at the wild-type -200 site *in vivo* in both the K562 cell line, and primary human erythroid cells from cord blood. This, taken together with the previously published data, such as the capability of ZBTB7A to repress genes and its potential role in the silencing of the embryonic *βh1* in a conditional *Zbtb7a*$^{-/-}$ mouse cell line [138], provides considerable evidence that ZBTB7A is a regulator of globin switching. The next logical question is, of course, whether ZBTB7A is capable of silencing the expression of *γ-globin* in human cells *in vivo*. Our lab is currently collaborating with Stuart Orkin and Takahiro Maeda at Harvard University who have developed a conditional *Zbtb7a*$^{-/-}$ mouse expressing a YAC containing the human *β-globin* locus.

The following chapters of this thesis will investigate the regulation of *Zbtb7a* in an erythroid context. As we have added further weight to the role of ZBTB7A in erythropoiesis, and possibly haemoglobin regulation, understanding how ZBTB7A is activated in erythroid cells is important.

CHAPTER 4 – *ZBTB7A* IS A DIRECT TARGET OF KLF1

4.1 Chapter 4 Introduction

As discussed in Chapter 1, KLF1 is a master regulator of erythropoiesis, and *Klf1* null mice die at around embryonic day E14.5 due to severe anaemia. An image of a *Klf1* null embryo at embryonic day E15 compared to a heterozygous littermate can be seen in Figure 4.1 [90]. This image clearly shows the anaemic phenotype of the *Klf1*$^{-/-}$ embryo. Initially, it was thought that this was predominantly due to the loss of adult *β-globin* gene expression, but the deficit has proven to be more severe than the absence of adult *β-globin* alone [182], as correction of this globin imbalance fails to sufficiently rescue the phenotype [54, 183, 184]. KLF1-binding motifs are prevalent in the regulatory regions of many genes involved in erythropoiesis. Microarray analyses have uncovered many KLF1 target genes, which have been found to be involved in all manner of erythroid cell maintenance, including the erythroid cytoskeleton, cell membranes and transmembrane proteins, haeme-biosynthesis, transcription and cell-cycle factors, as well as the various blood group antigens [87].

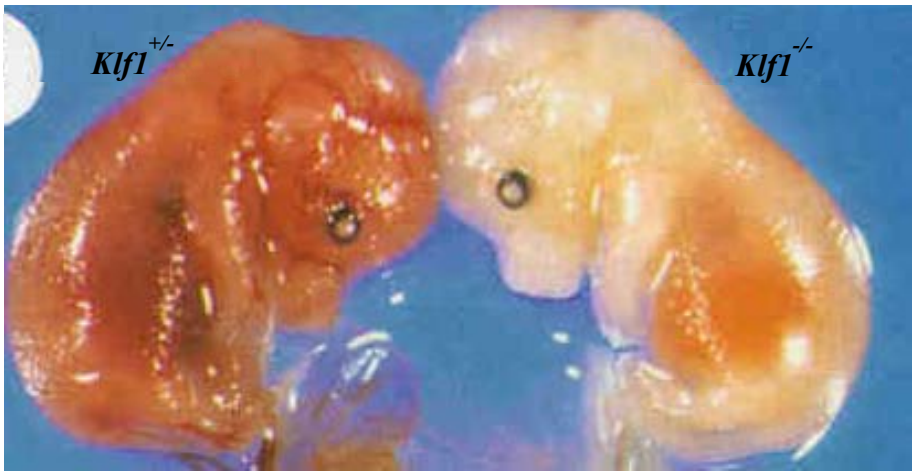


Figure 4.1. Phenotype of the *Klf1*$^{-/-}$ embryo at E15 compared to a heterozygous littermate [90]. Development of the Klf1 null embryos is normal until embryonic day E14.5, at which point the KO embryos are unable to produce mature definitive erythrocytes. At this time, the KO embryos succumb to fatal beta-thalassemia, hence the pail phenotype of the KO embryo.

It has long been known that KLF1 plays a critical role in regulating the developmental switch between foetal and adult haemoglobin expression by direct activation of adult *β-globin*. But more recently, it has also been found to be an important factor in the indirect repression of foetal *γ-globin* gene expression in adult erythroid progenitors, via direct activation of the known *γ-globin* transcriptional repressor, *Bcl11a* [185]. This is evidenced in a particular HPFH mutation within a Maltese family, which has been linked to a

heterozygous nonsense mutation, K288X, in *KLF1* that results in haploinsufficient expression of this gene. Genome-wide expression studies in these individuals have shown that one of the down-regulated genes is *BCL11A* [105, 106, 110], and consequently, the embryonic and foetal globins are up-regulated [185]. Also, it has been shown that *BCL11A* levels are directly proportional to altered KLF1 levels, and inversely proportional to *γ-globin* levels. ChIP analyses have also confirmed occupancy of KLF1 at the promoter region of the *BCL11A* gene in both humans and mice [112, 185]. As mentioned previously, BCL11A is not the only factor involved in *γ-globin* silencing. It therefore is reasonable to predict that KLF1 may activate the expression of other foetal globin silencers.

Microarrays, while identifying many erythroid genes regulated by KLF1, have also identified KLF1 target genes with no apparent erythroid function. For some of these genes, erythroid function is now being investigated. *Zbtb7a* is one such gene. This gene was identified as a potential KLF1 target in microarray analyses of *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ E14.5 foetal livers, as well as induced and un-induced K1ER cells, an inducible KLF1 cell line developed by rescuing immortalised E14.5 *Klf1* null foetal liver erythroblasts with a transgene encoding *Klf1* fused to the ligand binding domain of the oestrogen receptor (KLF1-ER) [142]. These cells are discussed further in following sections. The role of *Zbtb7a* in erythropoiesis has recently emerged and, as we have shown in the Chapter 3, there is evidence that ZBTB7A, like BCL11A, plays a role in foetal globin silencing.

It has not yet been examined whether *Zbtb7a* is, in fact, a direct target of KLF1, or whether it is regulated by other activating transcription factors which are downstream KLF1 targets. In this chapter, we aimed to test the hypothesis that *Zbtb7a* is directly activated by KLF1. As we know *Zbtb7a* plays an important role in erythropoiesis and potential role in globin switching, it is important to understand how *Zbtb7a* is regulated, especially in an erythroid context.

4.2 *Zbtb7a* is down-regulated in *Klf1*$^{-/-}$ animals at E14.5

Microarrays of *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ tissue have been previously performed on E14.5 foetal livers [87], and have revealed that *Zbtb7a* transcript levels are lower in the absence of *Klf1*. We performed qPCR and Western blot analysis on *Klf1*$^{+/+}$, *Klf1*$^{+/-}$ and *Klf1*$^{-/-}$ E14.5 foetal livers to validate this finding. A statistically significant down-regulation in *Zbtb7a* expression levels of approximately 3.5-fold was seen in the absence of KLF1 (Figure 4.2 A). It is worth noting that these embryos originated from different litters, which, despite the use of timed matings, may not be at exactly the same developmental time point. When individual samples

were compared, colour-coded according to litter, each *Klf1*$^{+/+}$ sample displayed consistently higher levels of *Zbtb7a* compared to their *Klf1*$^{-/-}$ littermates (Figure 4.2 B). ZBTB7A protein levels are also reduced in the *Klf1*$^{-/-}$ foetal livers, compared to their *Klf1*$^{+/+}$ and heterozygous littermates (Figure 4.3 A). Densitometry performed on these bands showed that the difference in protein level is similar to that of the transcripts (Figure 4.3 B).

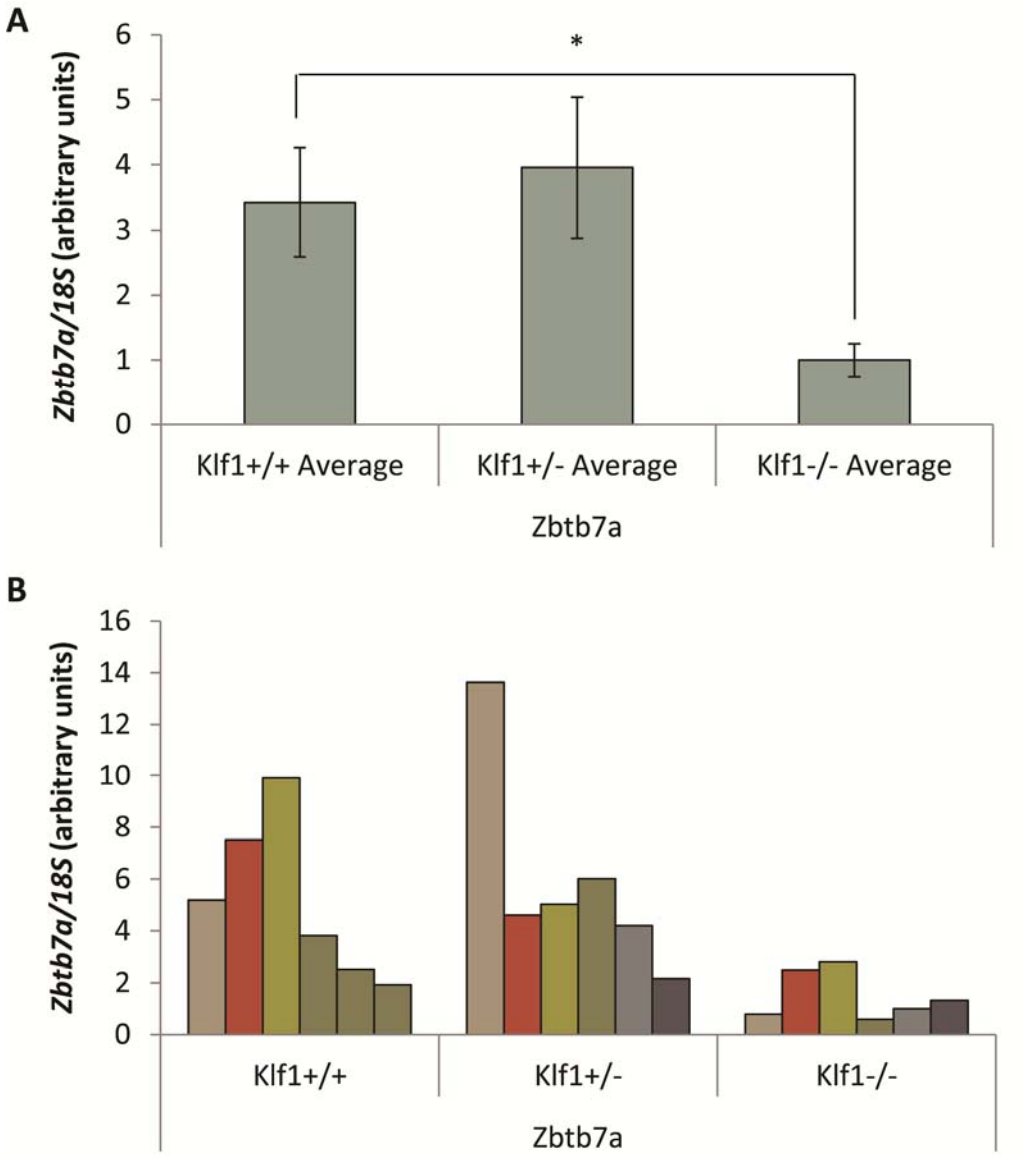


Figure 4.2. *Zbtb7a* transcript levels are depleted in E14.5 *Klf1* null foetal liver. **(A)** Whole RNA was extracted from E14.5 litter-matched *Klf1*$^{+/+}$, *Klf1*$^{+/-}$ and *Klf1*$^{-/-}$ foetal livers. Average mRNA levels of *Zbtb7a* were assessed by qPCR and have been normalised to *18S* rRNA levels. Error bars represent standard error of the mean (n = 5 per genotype). *, $P < 0.05$ (paired Student's *t* test, compared *Klf1*$^{+/+}$ to *Klf1*$^{-/-}$ samples). **(B)** *Zbtb7a* transcript levels assessed by qPCR for each sample, colour-coded according to litter.

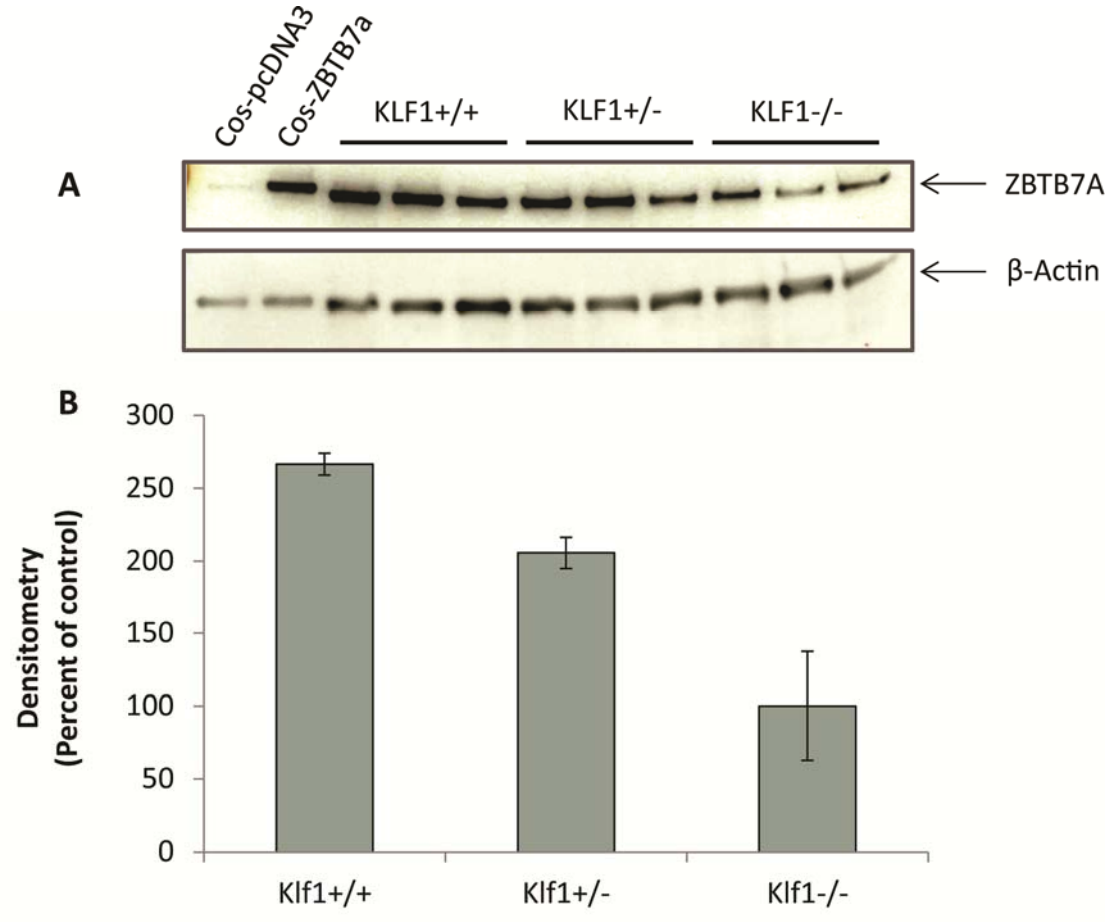


Figure 4.3. ZBTB7A protein levels are lower in E14.5 *Klf1* null foetal liver. **(A)** Nuclear extracts were prepared from mock-transfected Cos-7 cells (lane 1), and Cos-7 cells overexpressing ZBTB7A (lane 2), as well as E14.5 foetal livers from litter-matched $Klf1^{+/+}$, $Klf1^{+/-}$ and $Klf1^{-/-}$ embryos (lane 3-11). ZBTB7A protein levels were assessed by Western Blot. Blots were probed with αZBTB7A and β-Actin was used as a loading control. **(B)** Densitometry analysis of Western Blot in (A) performed using ImageJ analysis software. Error bars represent the standard error of the mean (n = 3 per genotype). $P = 0.06$ (paired Student's t test, compared $Klf1^{+/+}$ to $Klf1^{-/-}$ samples).

E14.5 is the time of embryonic lethality in the $Klf1^{-/-}$ mice [89, 90], and from our experience, $Klf1^{-/-}$ embryos have succumbed to severe anaemia at this time point. These mice were pale and small compared to their $Klf1^{+/+}$ and $Klf1^{+/-}$ littermates. Therefore changes in expression levels at this time point may be due to degrading RNA and protein species and not exclusively attributable to the absence of KLF1. Thus, we performed our own microarrays comparing $Klf1^{+/+}$ to $Klf1^{-/-}$ foetal livers at the earlier time point of E13.5, prior to $Klf1^{-/-}$ embryos succumbing to anaemia. At this time point, *Klf1* null embryos were more comparable to their WT and heterozygous littermates as they were still viable. RNA was extracted from $Klf1^{+/+}$ and *Klf1* null foetal livers, and was then subjected to whole transcript sense labelling and hybridisation to Mouse Gene ST 1.0 arrays (Affymetrix, CA). Microarray data were analysed using Partek genomic suite v6.5 (Partek Inc., MO).

Microarray CEL files were imported into Partek and normalized using the robust multi-array average (RMA) algorithm. Figure 4.4 shows a Principal Components Analysis plot, used to visualise each sample set as a whole. Each point on the plot represents a sample, red being the *Klf1*$^{-/-}$ samples, and grey the *Klf1*$^{+/+}$. The *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ samples tended to group together, indicating similar intensity values across the probe sets for the entire genome for each group, and a clear separation of *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ samples. There did not appear to be any outlier samples, although slightly greater variability was observed among the *Klf1*$^{+/+}$ samples.

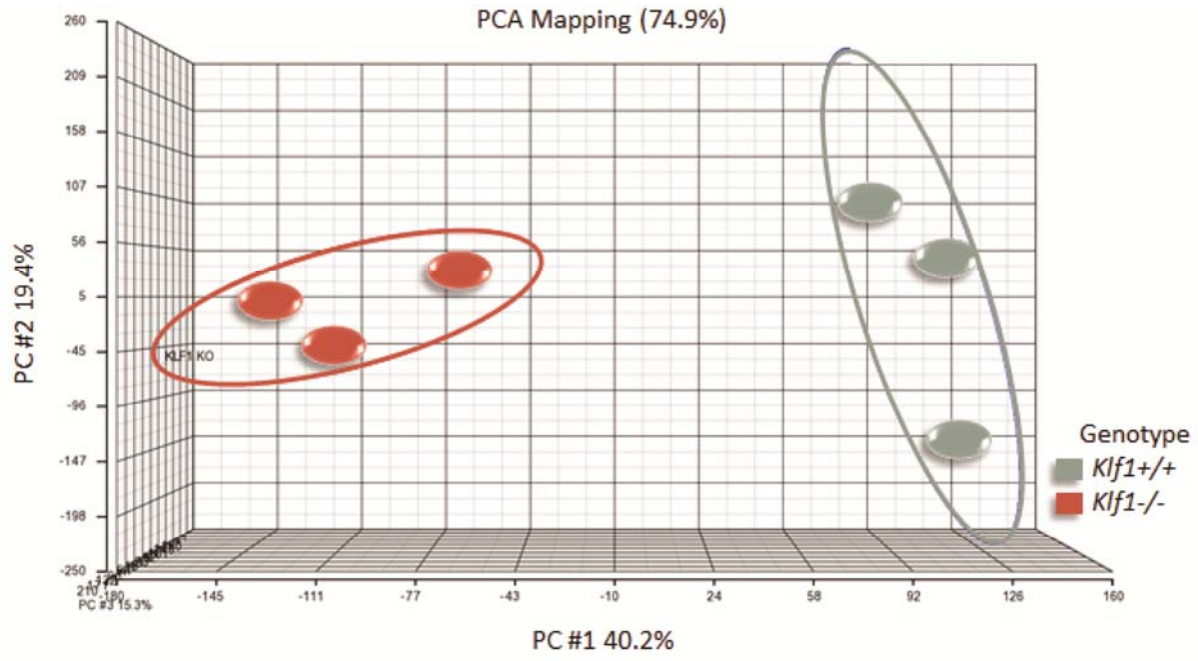


Figure 4.4. Principal component analysis of E13.5 foetal liver mircoarrays shows *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ samples group together. Whole RNA from *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ foetal livers was then subjected to whole transcript sense labelling and hybridisation to Mouse Gene ST 1.0 arrays (Affymetrix, CA). Data was imported into Partek software v6.5 for analysis. A PCA map was produced, showing the groupings of the data for all the probe sets across the entire mouse genome for each sample. *Klf1*$^{+/+}$ samples are shown in grey and *Klf1*$^{-/-}$ samples in red (n = 3 for each genotype).

After confirming array quality (Affymetrix built-in controls and principal components analysis), differential gene expression was assessed and tested for significance using a one-way analysis of variance (ANOVA). In total, there were over 28, 000 genes deregulated in the absence of *Klf1*. A volcano plot of these data is shown in Figure 4.5. A *P* value cut-off of less than 0.05 was applied to the total list of deregulated genes, and of these, transcripts deregulated by more than 2-fold were selected. In total, 503 probe sets showed altered expression between *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ at these cut-offs. A heat map was generated from these 503 significantly altered genes (Figure 4.6). The majority of genes that showed altered expression in the *Klf1*$^{-/-}$ samples were down-regulated: a total of 382 genes were down- and

121 were up-regulated. This heat map also assesses the similarity between the samples - the *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ samples group together very distinctly.

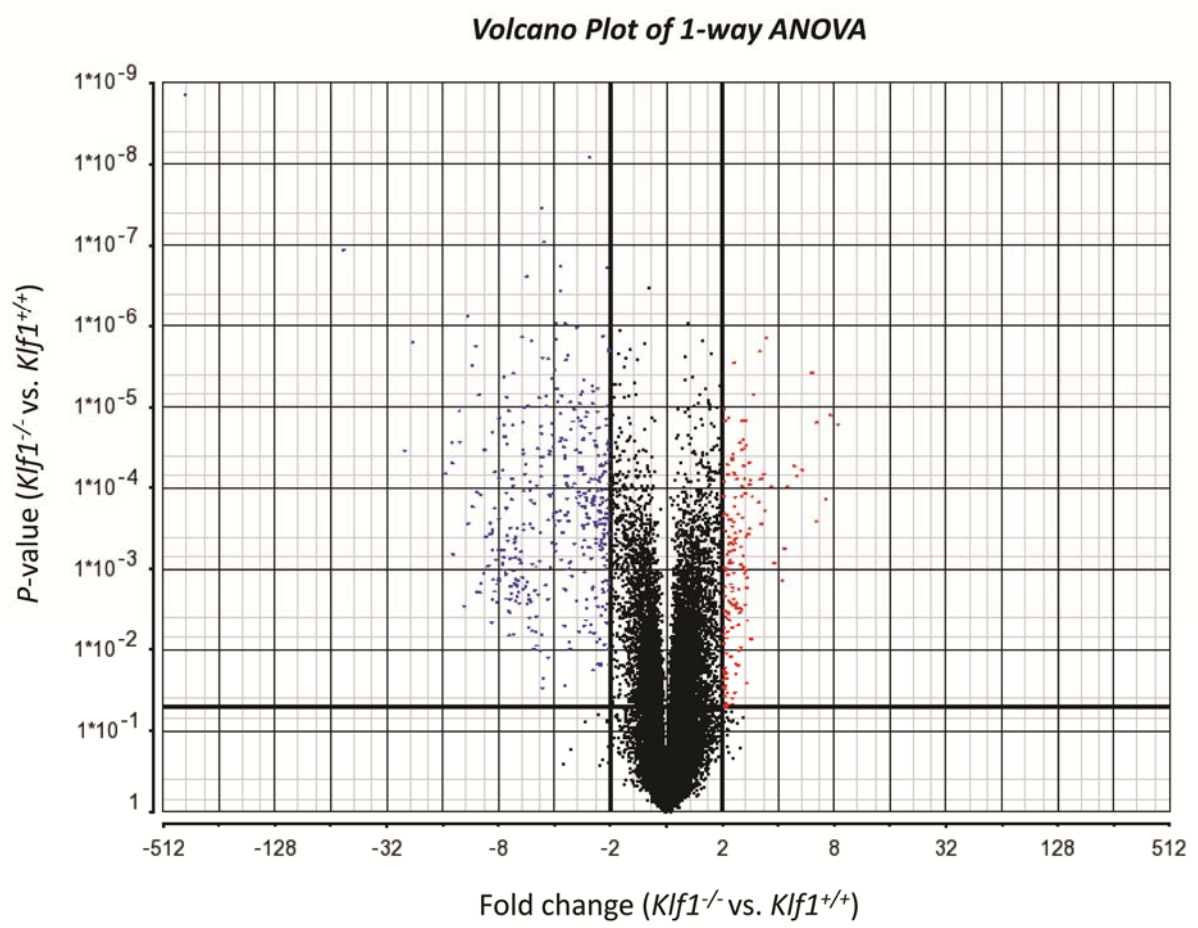


Figure 4.5. Effect of *Klf1* knockout on transcript expression in E13.5 foetal liver cells. Vertical cut-offs represent a fold change of 2.0 in transcript expression, and horizontal cut-offs represent a *P*-value of < 0.05 as determined by one-way ANOVA, generated by Partek Genomics Suit™ 6.6. Blue data points represent genes down-regulated in the *Klf1*$^{-/-}$, and red data points up-regulated, that satisfy both of these cut-offs. These genes were then used to create the gene list used for further analysis.

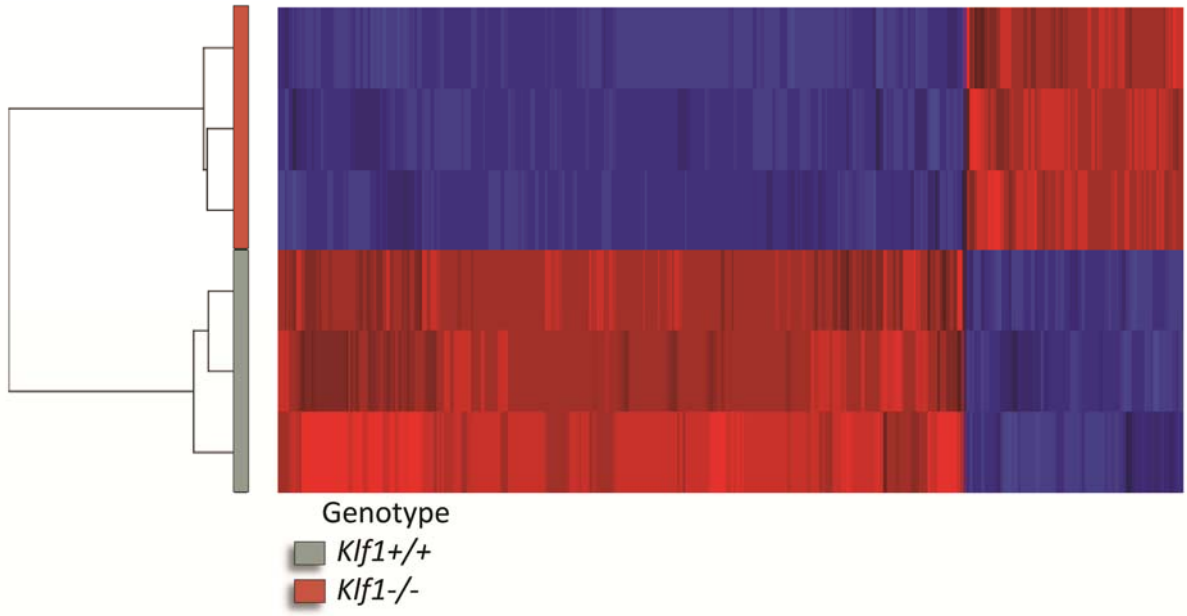


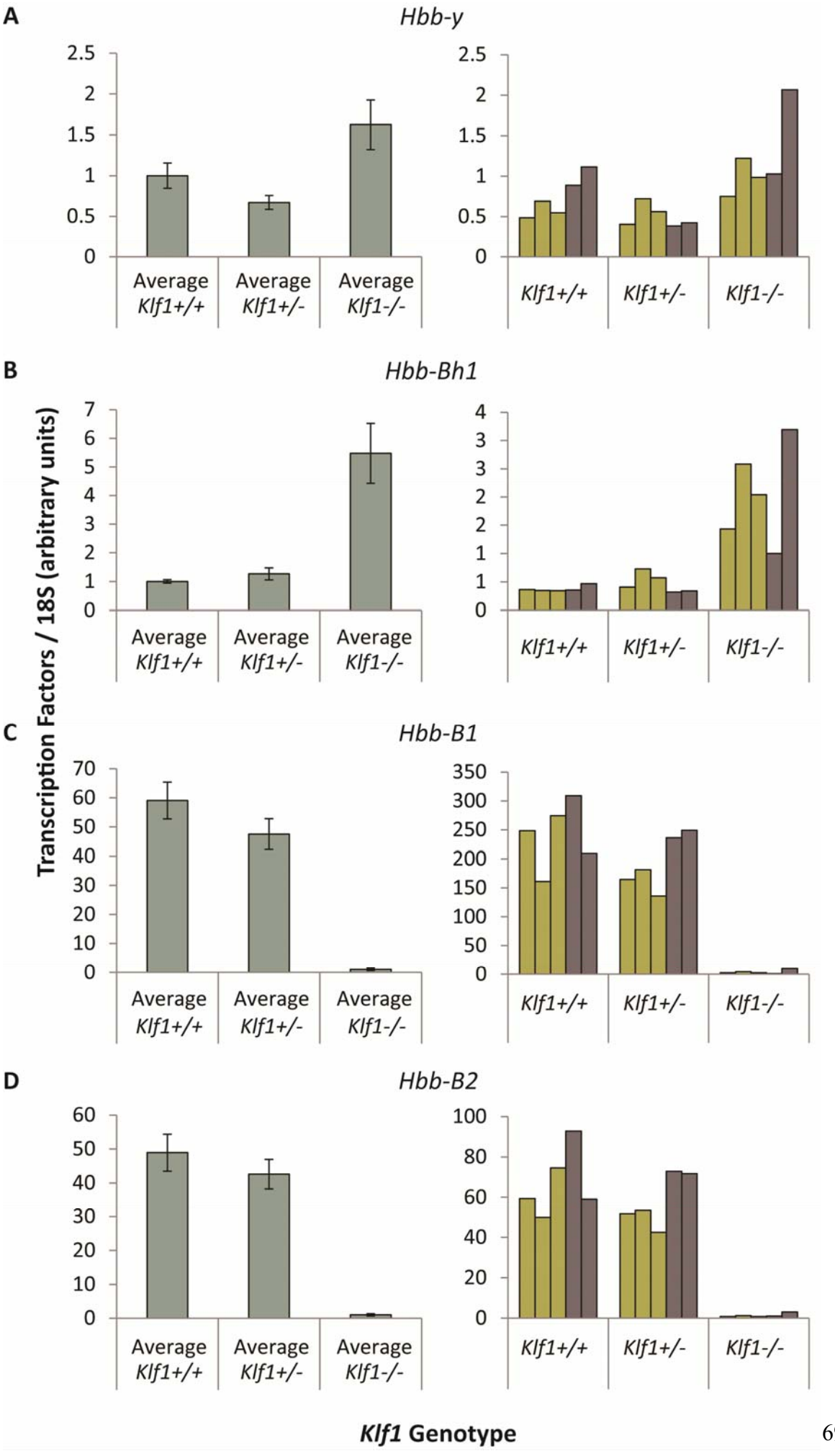
Figure 4.6. Gene expression changes in *Klf1*$^{+/+}$ versus *Klf1*$^{-/-}$ E13.5 foetal livers. The heat map displays results for all differentially expressed genes whose average expression changed by at least 2-fold and $P < 0.05$. Up-regulated genes are shown in red and down-regulated genes in blue. This map was generated by Partek Genomics Suit™ 6.6.

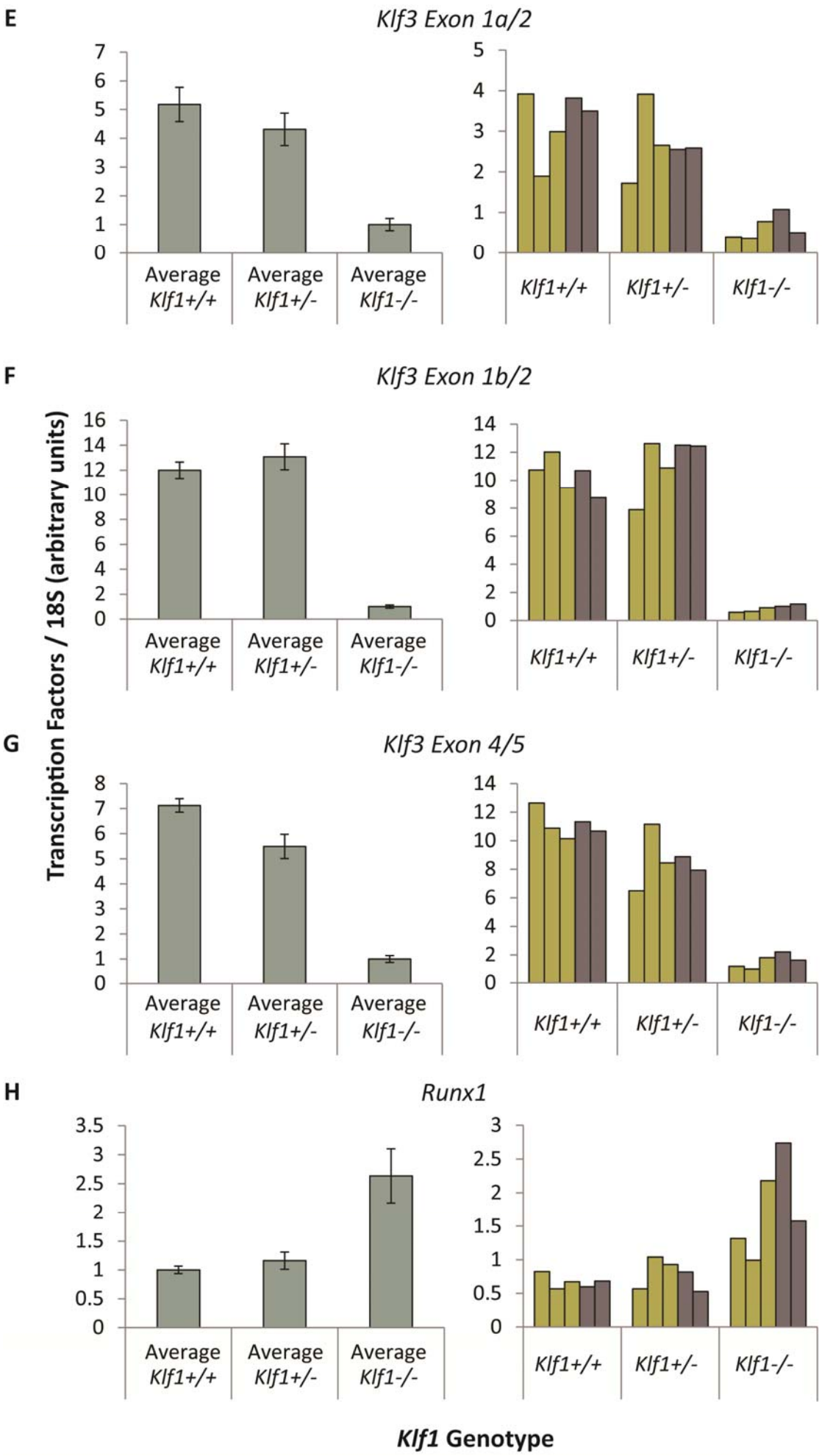
Many of the 503 probe sets found to have altered expression between these samples have not yet been annotated to specific transcripts. Of the 322 annotated probe sets found, 228 were down-regulated in the *Klf1*$^{-/-}$, and 94 up-regulated. Gene ontology performed on this list of 322 genes using the online DAVID Bioinformatics Resources 6.7 tool [186, 187] indicated the most enriched pathways to be those involved in erythroid cells, such as erythrocyte differentiation, homeostasis and development, as well as other more general haematopoietic pathways. Some genes found to be deregulated in the microarrays, such as *Klf3, Zbtb7a* and *Lmo2*, were not identified in the DAVID analysis as having a specific haematopoietic role, however their role in erythropoiesis has been established in previous studies and therefore, they were included in the table. A composite list of the most highly deregulated genes involved in haematopoiesis is shown in Table 4.1.

Table 4.1. Gene list of 20 of the most highly deregulated haematopoietic genes found by microarray analysis of *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ E14.5 foetal livers. First column is the gene name, column three is the *P*-value, column four shows the haematopoietic pathways identified by gene ontology software DAVID, and column five lists references relating to the role of each gene in haematopoiesis.

Gene Name	P Value	Fold Change	Pathway from DAVID	Reference
Ahsp	1.38×10^{-9}	-389.07	Erythroid differentiation & homeostasis	[188, 189]
Sox6	1.66×10^{-5}	-6.70	Erythrocyte development	[117-120, 190]
Klf3	1.37×10^{-6}	-5.99	Transcription	[91, 104, 191]
Zbtb32	9.20×10^{-5}	-4.03	Haematopoiesis, lymphopoiesis & immune development	[192-194]
Add2	1.49×10^{-6}	-3.93	Haematopoiesis, lymphopoiesis & immune development	[195, 196]
Epb4.2	6.48×10^{-6}	-2.85	Erythroid differentiation, development & homeostasis, haematopoiesis, lymphopoiesis & immune development	[197]
Ank1	6.06×10^{-5}	-2.83	Erythroid differentiation, development & homeostasis, haematopoiesis, lymphopoiesis & immune development	[198]
Klf1	2.82×10^{-4}	-2.75	Erythroid differentiation, development & homeostasis	[84, 89, 90, 101]
Sgpl1	4.75×10^{-4}	-2.25	Haematopoiesis, lymphopoiesis & immune development	-
Zbtb7a	4.89×10^{-4}	-2.15	Transcription	[136-138, 164]
Slc25a38	1.70×10^{-5}	-2.100	Erythrocyte differentiation, homeostasis, myelopoiesis	[199]
Hbb-b1	2.04×10^{-4}	-2.10	Haematopoiesis	-
Klf11	2.48×10^{-4}	-2.03	Haematopoiesis, lymphopoiesis & immune development	[200]
Lmo2	2.05×10^{-6}	-2.03	-	[201-204]
Bcl11a	4.54×10^{-5}	-2.02	Haematopoiesis, lymphopoiesis & immune development	[105-108]
Trim10	2.03×10^{-4}	-2.00	Erythroid differentiation & homeostasis	[205]
Mfsd7b	5.22×10^{-4}	+2.52	Erythroid differentiation & homeostasis, Haematopoiesis, lymphopoiesis & immune development	[206]
Hbb-bh1	9.58×10^{-5}	+3.68	-	-

As KLF1 is an important erythroid transcription factor, we would expect to find many genes involved in erythroid pathways affected, which was indeed the case. This gives us additional confidence in the microarray data. Some of the more interesting genes found in this list are discussion in Section 4.5. Next, we validated a number of the differentially expressed genes by qPCR, including the genes of the *β-globin* locus (Figure 4.7).


A
Hbb-y
2.5
2
1.5
1
0.5
0
Average Klf1+/+
Average Klf1+/-
Average Klf1-/-
Klf1+/+
Klf1+/-
Klf1-/-
B
Hbb-Bh1
Transcription Factors / 18S (arbitrary units)
7
6
5
4
3
2
1
0
C
Hbb-B1
70
60
50
40
30
20
10
0
350
300
250
200
150
100
50
D
Hbb-B2
60
50
40
30
20
10
0
100
80
60
40
20
Klf1 Genotype


E
Klf3 Exon 1a/2
F
Klf3 Exon 1b/2
G
Klf3 Exon 4/5
H
Runx1
Transcription Factors / 18S (arbitrary units)
Average Klf1+/+
Average Klf1+/-
Average Klf1-/-
Klf1+/+
Klf1+/-
Klf1-/-
Klf1 Genotype

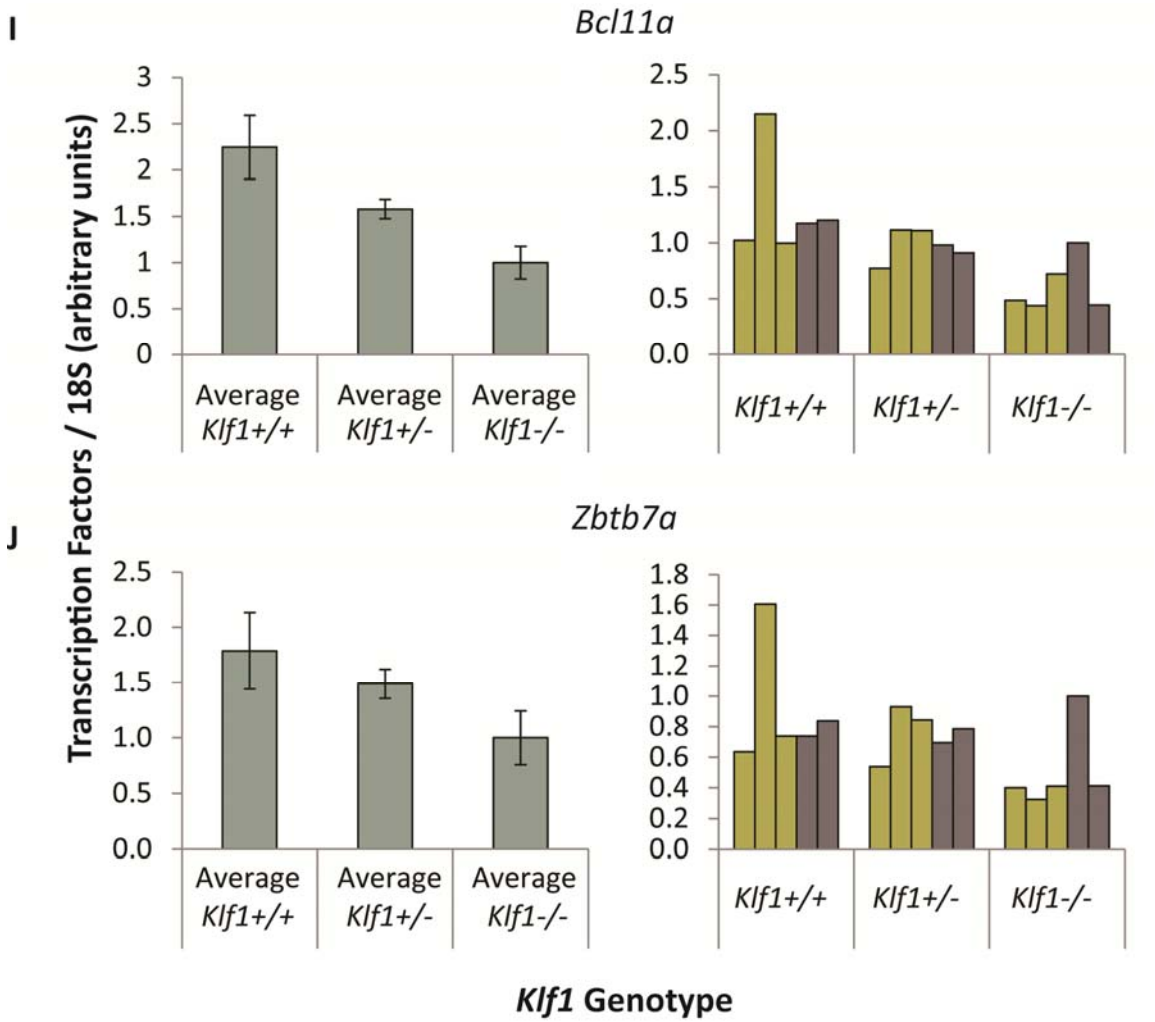


Figure 4.7. Validation of a subset of deregulated genes found in *Klf1* microarrays. **(A)** *embryonic εy-globin* **(B)** *embryonic βh1-globin* **(C)** *adult β-major-globin* **(D)** *adult β-minor-globin* **(E)** *Klf3 exon 4/5* **(F)** *Klf1 1a* **(G)** *Klf3 1b* **(H)** *Runx1* **(I)** *Bcl11a* **(J)** *Zbtb7a.* Left-hand side plots show the average levels of transcripts assessed by qPCR in E13.5 litter-matched $Klf1^{+/+}$, $Klf1^{+/-}$ and $Klf1^{-/-}$ foetal livers. Error bars represent standard error of the mean (n = 5 per genotype). Right-hand side plots show each individual sample, colour-coded by litter. All targets have been normalised to *18S* rRNA levels, and are represented as arbitrary units.

The fold-changes of genes that were found by qPCR validation were compared to the fold-change seen in the microarray analysis (Table 4.2).

Table 4.2. List of qPCR validated microarray genes. This table shows the fold-change of each gene validated by qPCR, their *P* value and their fold change found by microarray analysis.

Gene Name	*P*-value	Fold-change by Microarray	Fold-change by RT-PCR
Hbb-y	3.87E-01	+1.08	+1.8
Hbb-bh1	1.87E-04	+3.70	+6
Hbb-b1	7.59E-05	-2.15	-60
Hbb-b2	4.37E-05	-2.13	-50
Klf34/5	8.84E-06	-5.96	-7
Klf3 1a	-	-	-5
Klf3 1b	-	-	-12
Runx1	1.57E-04	+1.51	+2.8
Bcl11a	9.42E-05	-2	-2.4
Zbtb7a	1.57E-04	-2.13	-1.8

This comparison illustrates that results found by microarray and qPCR were, by and large, comparable. However, the example highlighted by *Klf3* demonstrates how the full scale of deregulation can be masked in both of these methods when looking at various parts if the transcript. We have previously described that *Klf3* has both a ubiquitous promoter and an erythroid-specific promoter, which is regulated by KLF1, (promoters termed *1a* and *1b* respectively). These transcripts are both located upstream of the translational start site in the 5'UTR region of the gene [94]. The microarray probe sets for *Klf3* lay downstream of the translational start site. In order to look at the expression level of both transcripts and validate the microarrays, we used primers that span the exon 4 and 5 boundary, seen in Figure 4.5 E, labelled as *Klf3* exon4/5. Here, a similar fold-change between the qPCR and microarray analysis was found. However, when the differentially expressed transcripts, *1a* and *1b*, were investigated individually, it was found that the erythroid transcript *1b* is significantly more down-regulated in the *Klf1*$^{-/-}$. This is expected as it is known that this transcript in particular is regulated by KLF1. From this it can be seen how the true scale of up- or down-regulation of particular transcripts which are differentially expressed in tissues may be masked in both microarrays and qPCR analyses.

Two other genes which showed a greater degree of down-regulation in the qPCR analysis were *β-major* and *β-minor*. The reason for this discrepancy is unclear. It may be due to a similar explanation as that provided for *Klf3,* where the probe sets recognised the entire gene and our primers lay in a region which was more differentially expressed in the absence of *Klf1*, although there is no particular evidence to support this theory. Another reason may be that these two genes are very highly expressed, and were readily detected even in the *Klf1*$^{-/-}$

foetal liver samples by qPCR (see Appendix II for primary amplification plots). It may be that the expression level is simply too high for an accurate fold change, and that these genes are expressed at the upper-limit of microarray detection.

Importantly for this thesis, *Zbtb7a* was found to be down-regulated to a similar level in both the microarrays and the qPCR analysis. The difference between the *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ genotypes is similar to what has been described before [87], and is relatively small. Microarray probe sets for *Zbtb7a* again are found across the entire gene, and our primers are specific to sites in exon 2. It is possible that greater down-regulation would be seen in more specific areas of the transcript, such as the 5'UTR, due to the differential expression of unknown variants not being detected in these experiments. This will be further investigated in Chapter 5.

Our microarray analysis adds yet more evidence to the hypothesis that *Zbtb7a* is a target of KLF1, as it has now been found in a number of *Klf1* microarrays, but the question still remains of whether or not *Zbtb7a* is *directly* activated by KLF1. This will be examined in the following sections.

4.3 ZBTB7A is up-regulated in response to KLF1 in tamoxifen-induced K1ER cells

The KLF1 Oestrogen Receptor Inducible cell line (K1ER) was engineered by immortalising E14.5 *Klf1* null foetal liver erythroblasts with the J2 retrovirus [142]. These cells were subsequently rescued by stable transfection of a transgene encoding *Klf1* fused to the ligand binding domain of the oestrogen receptor (KLF1-ER). KLF1-ER is constitutively expressed in these cells, and interactions between the oestrogen receptor (ER) and heat shock proteins such as HSP90, sequester KLF1 in the cytoplasm [207]. Treating these cells with the synthetic ligand 4-hydroxytamoxifen (hereafter, tamoxifen) results in dissociation of KLF1-ER from the heat shock protein complex and translocation into the nucleus where it is able to activate its target genes [142]. A mutant form of the ER has been used in this system, ensuring that KLF1-ER responds only to tamoxifen and not oestrogen, sometimes present in cell culture medium [207]. Induction of KLF1-ER by tamoxifen causes these cells to terminally differentiate. They significantly up-regulate the adult *β-major* globin, form haemoglobin and display reduced proliferation [142]. The K1ER system represents a good model for the study of *Klf1*-induced genes, and has been used to emphasise the importance of KLF1 in the late stages of definitive erythropoiesis. As such, K1ER cells have been used to validate several putative Klf1 target genes, including *Ahsp*, *dematin* and *Klf3* [87, 94, 189].

Lastly, the K1ER cells harbour a YAC containing the entire human *β-like globin* gene locus, which enables the use of these cells in the study of human *β-like globin* gene regulation.

Microarrays have also previously been performed on these cells, comparing un-induced to induced, and found that *Zbtb7a* is up-regulated in induced cells in response to KLF1-ER [87]. Again, we sought to validate this result and looked at transcript levels of *Zbtb7a* in un-induced cells treated with EtOH, and induced cells treated with tamoxifen at for 48 hours by qPCR. *Klf3 1b* transcript levels were used as a control to validate whether the induction was successful as this is an established KLF1 target [94] (Figure 4.8). *Klf3 1b* expression increased dramatically at 48 hours compared to the EtOH treated control, indicating that the induction was successful. *Zbtb7a* transcript levels showed a 4-fold increase upon induction. This level is comparable to what is seen in the foetal liver at both E13.5 and E14.5.

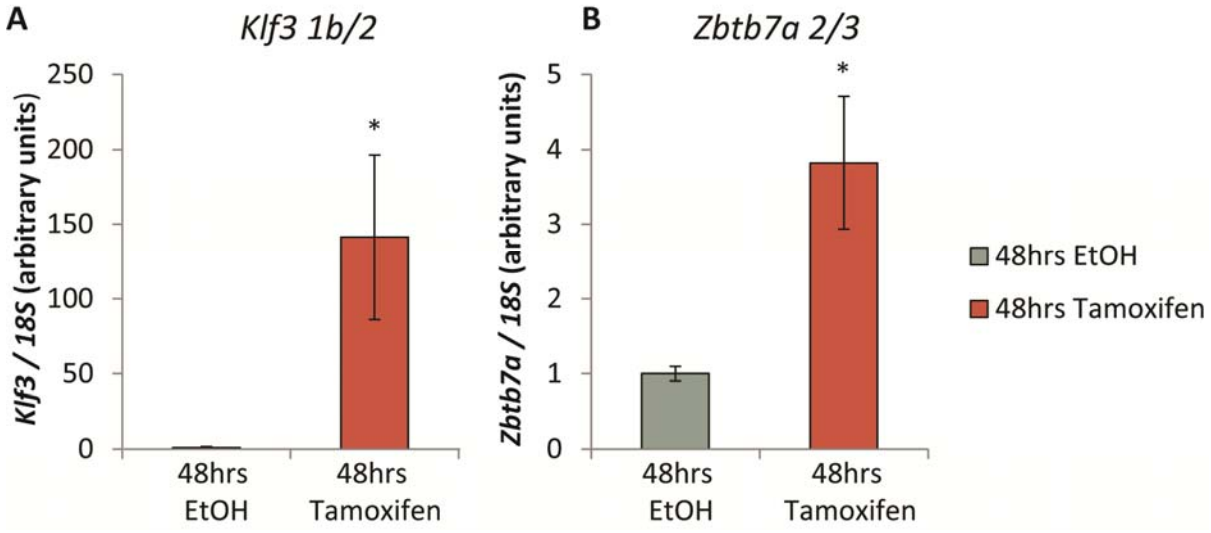


Figure 4.8. Up-regulation of *Klf3 1b* and *Zbtb7a* mRNA by KLF1-ER at 48 hours post-induction with tamoxifen. Cells were treated with 200 nM tamoxifen, or 0.02% EtOH for control cells, for 48 hours and RNA was extracted. *Klf3 1b* and *Zbtb7a* mRNA levels were determined by qPCR and have been normalised to *18S* rRNA. Error bars represent standard error of the mean (n = 6 per condition). *, $P < 0.05$ (paired Student's *t* test, compared Control to 48 hours).

We also performed a time-course of induction in these cells to further investigate this relationship, looking at 0, 2, 4, 6, 8, 24 and 48 hours. The induction was successful, as the cells turned red at 24 hours compared to the EtOH treated controls, indicating the production of haemoglobin in these cells (Figure 4.9 A). *Klf3 1b* expression was strongly induced, also demonstrating successful induction of these cells (Figure 4.9 B).

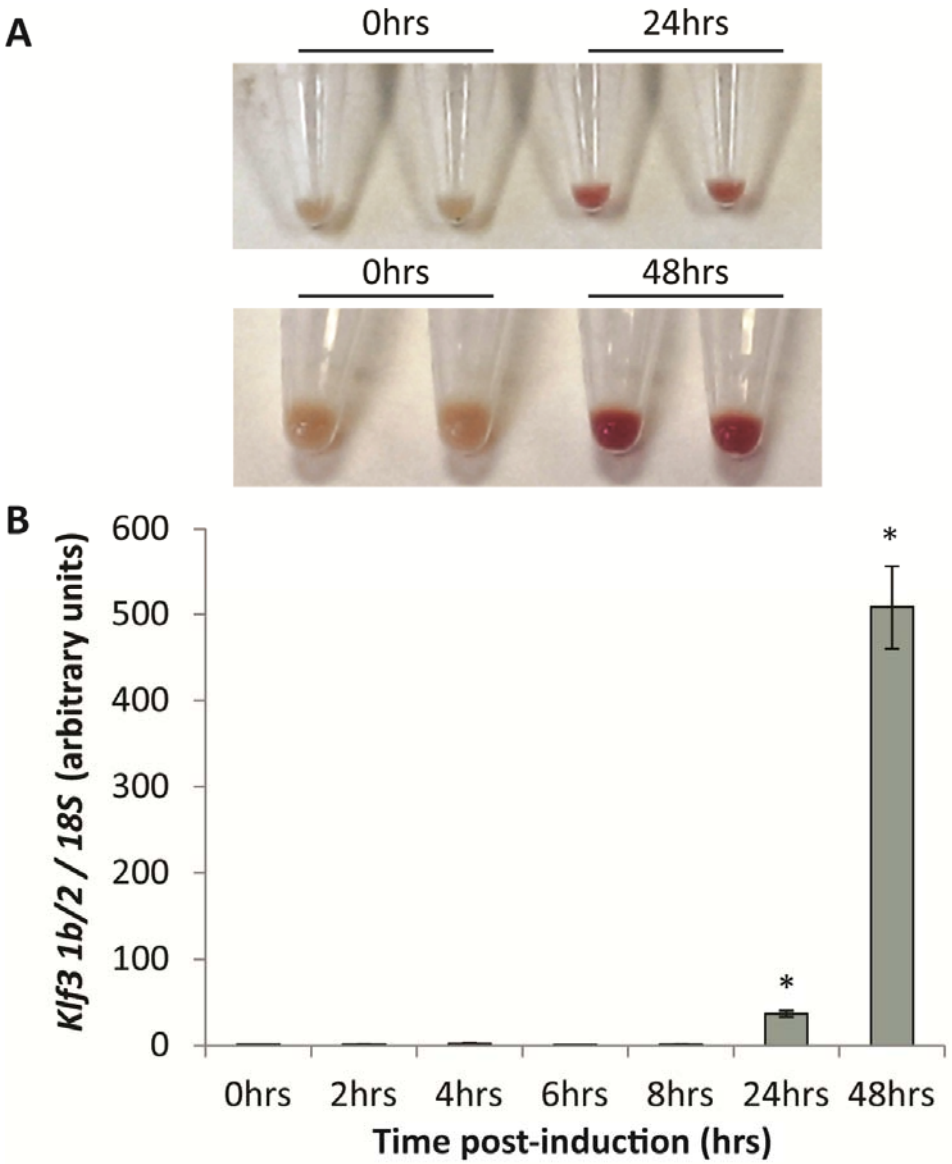


Figure 4.9. Successful K1ER tamoxifen induction time course. (A) K1ER cells treated with 0.02% EtOH or 200 nM tamoxifen shown 24 and 48 hours post-induction. **(B)** mRNA was extracted from these cells and *Klf3 1b* transcript levels were assessed by qPCR. Transcript levels were normalised to *18S* rRNA. Error bars represent standard error of the mean (n = 2 per condition).*, $P < 0.05$ (paired Student's *t* test, compared 0 hour time point to 24 and 48 hours).

We found that both *Zbtb7a* transcript (A) and protein levels (B) were up-regulated during the time course (Figure 4.9). Interestingly, we found the largest up-regulation of *Zbtb7*a transcript levels was at 2 hours post-induction, which then rapidly decreased again at the 4 hour time point to roughly 2-fold above the level at 0 hours. This was maintained until the final time point of 48 hours post-induction. The protein level followed a similar pattern; however the peak of the increase lagged behind the transcript time points.

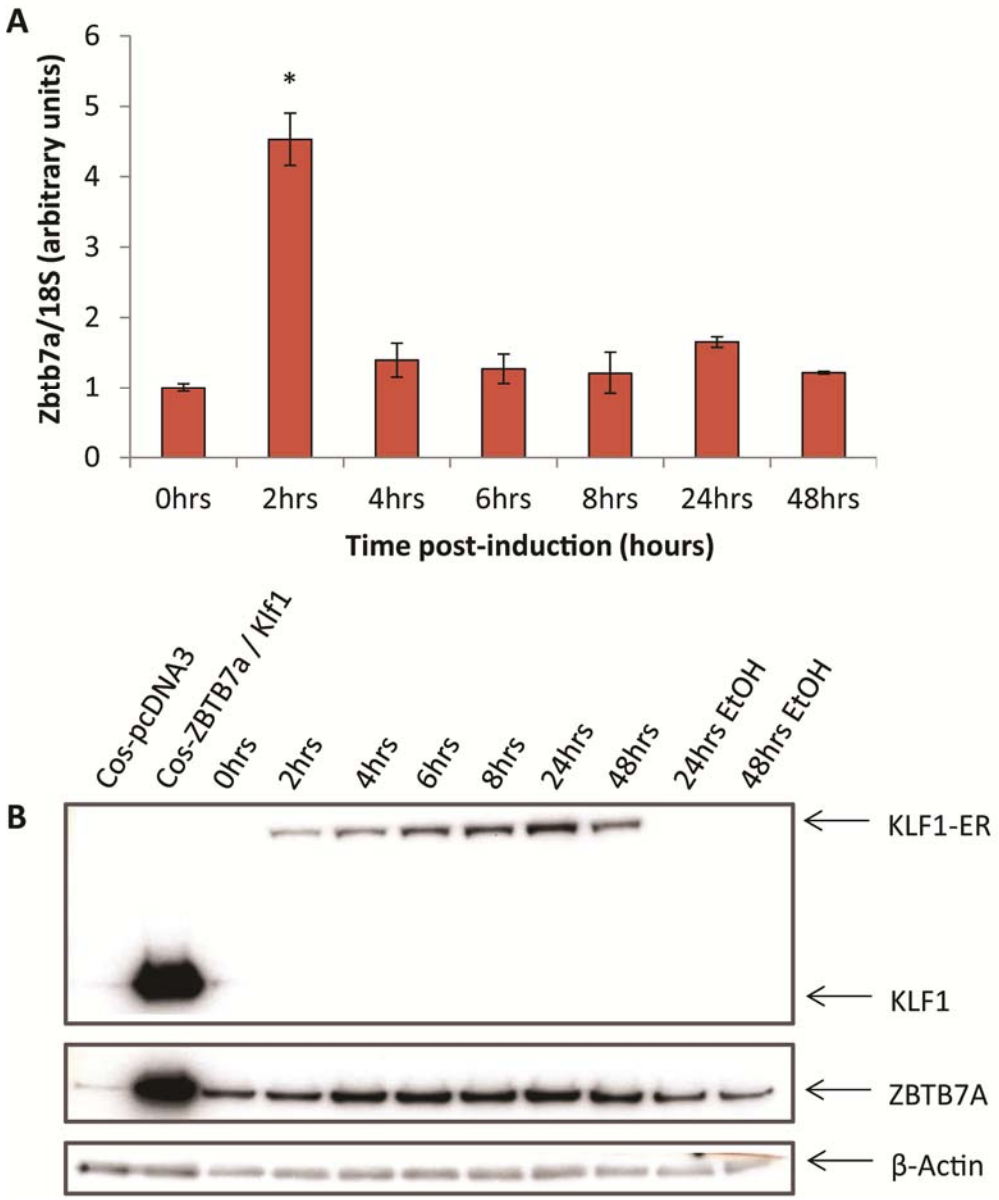


Figure 4.10. Zbtb7a is up-regulated upon induction of K1ER cells at both the transcript and protein level. (A) *Zbtb7a* levels were assessed by qPCR, and were normalised to *18S* rRNA. Error bars represent standard error of the mean (n = 2 per condition).*, $P < 0.05$ (paired Student's *t* test, compared 0 hour to 2 hour time point). **(B)** Nuclear extracts were prepared from mock-transfected Cos cells (lane 1), and Cos cells overexpressing ZBTB7A (lane 2), as well as K1ER cells at 0, 2, 4, 6, 8, 24 and 48 hours post-induction with 200 nM tamoxifen (lane 3-9) and K1ER cells at 24 and 48 hours post-treatment with 0.02% EtOH (lane 10 and 11). Blots were probed with αZBTB7A and β-Actin was used as a loading control. The top panel of the blot shows KLF1 overexpressed in Cos cells in lane 2, and KLF1-ER in lanes 3-11. As KLF1-ER is fused to the oestrogen receptor, it runs higher in the gel that endogenous KLF1.

As the increase in *Zbtb7a* was seen so shortly after induction, we decided to tease out the early time points in a shorter time course to determine how quickly *Zbtb7a* is targeted by KLF1-ER (Figure 4.11). *Klf3 1b* transcript levels had increased dramatically by 48 hours, again indicating successful induction of KLF1-ER in this time course. Interestingly, while the

levels of *Klf3 1b* showed a small increase early on, it was not until the later time point of 4 hours that the levels of this transcript began to significantly increase. *Zbtb7a* levels however, showed a statistically significant increase as early as 30 minutes post-induction, and peaked at around 1-3 hours, after which they gradually began to fall to a final fold change of 2 at the 48 hour time point compared to 0 hours.

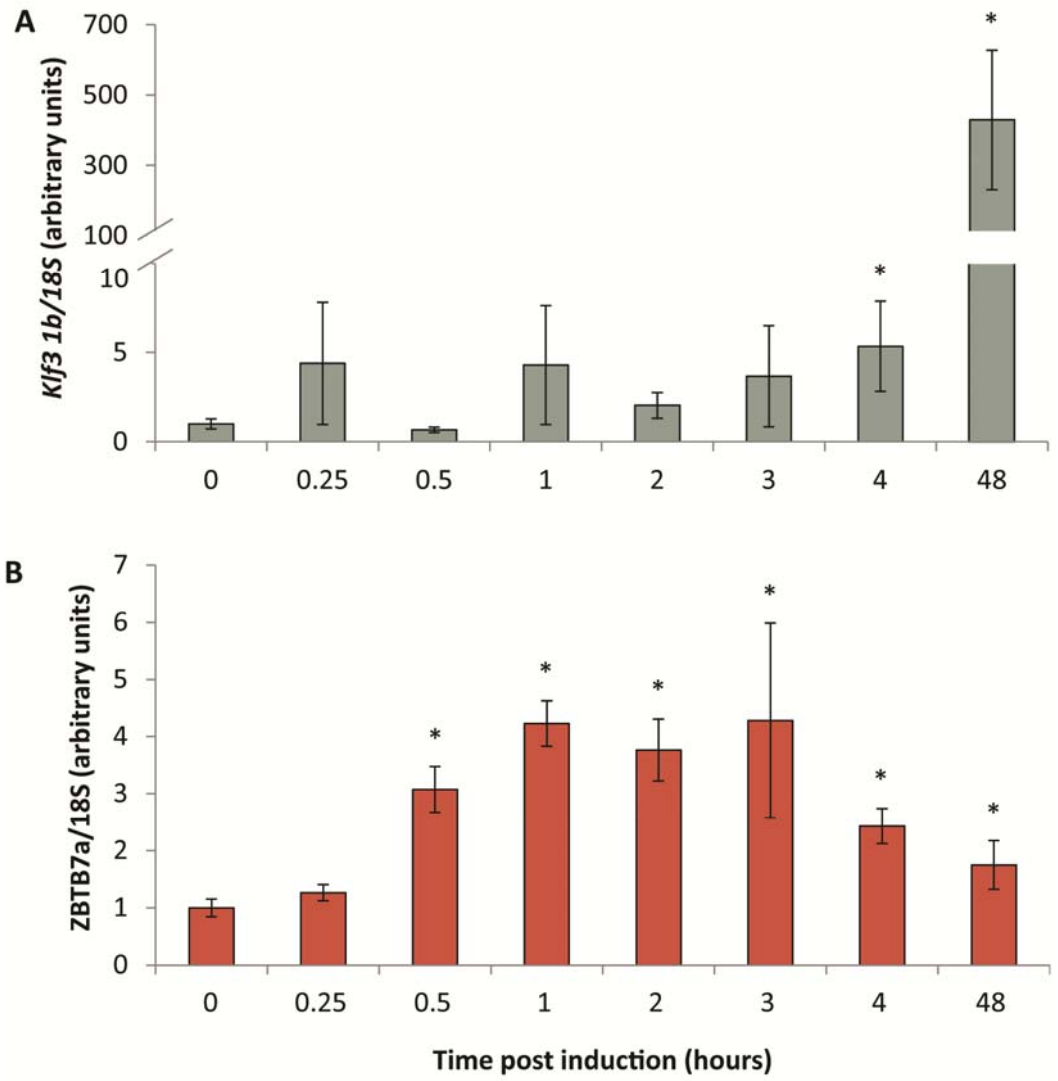


Figure 4.11. Transcript levels of *Klf3 1b* and *Zbtb7a* are up-regulated upon induction of K1ER cells compared to controls. Total RNA was extracted from, K1ER cells 0, 15 min, 30 min and 1, 2, 3, 4 and 48 h post induction with 200 nM tamoxifen. qPCR was used to determine mRNA levels of both *Klf3 1b* **(A)** and *Zbtb7a* **(B)**. Transcript levels have been normalised to *18S* rRNA. Errors bars represent standard error of the mean (n = 4). *, $P < 0.05$ (paired Student's *t* test, compared other time points to the 0 hour time point).

In the K1ER cell line, the KLF1-ER protein is constitutively expressed and is sequestered in the cytoplasm in the absence of tamoxifen. We were therefore able to exploit this constitutive expression by using the translation inhibitor cycloheximide to examine whether KLF1-ER directly or indirectly activates transcription of *Zbtb7a*. If KLF1-ER indirectly

activates *Zbtb7a* expression, for example, through up-regulation of other intermediary transcription factors, then an increase in *Zbtb7a* mRNA should not be observed when K1ER cells are induced in the presence of cycloheximide. In contrast, if KLF1-ER directly activates transcription of the *Zbtb7a* gene, then this would be observed in induced K1ER cells even in the presence of cycloheximide because no *de novo* protein synthesis would be required.

K1ER cells were exposed to cycloheximide (5 µg/mL) 30 minutes prior to induction with tamoxifen (200 nM), or ethanol (0.02% v/v) as a negative control. Total RNA was harvested from the cells every 2 hours to a maximum of 6 hours following induction, rather than 48 hours, due to the cytotoxicity of cycloheximide. *Zbtb7a* transcripts were indeed up-regulated to similar levels as previously seen during induction in the presence of cycloheximide (Figure 4.12). This suggests that KLF1-ER directly activates the *Zbtb7a* gene as no *de novo* protein synthesis was required for its activation.

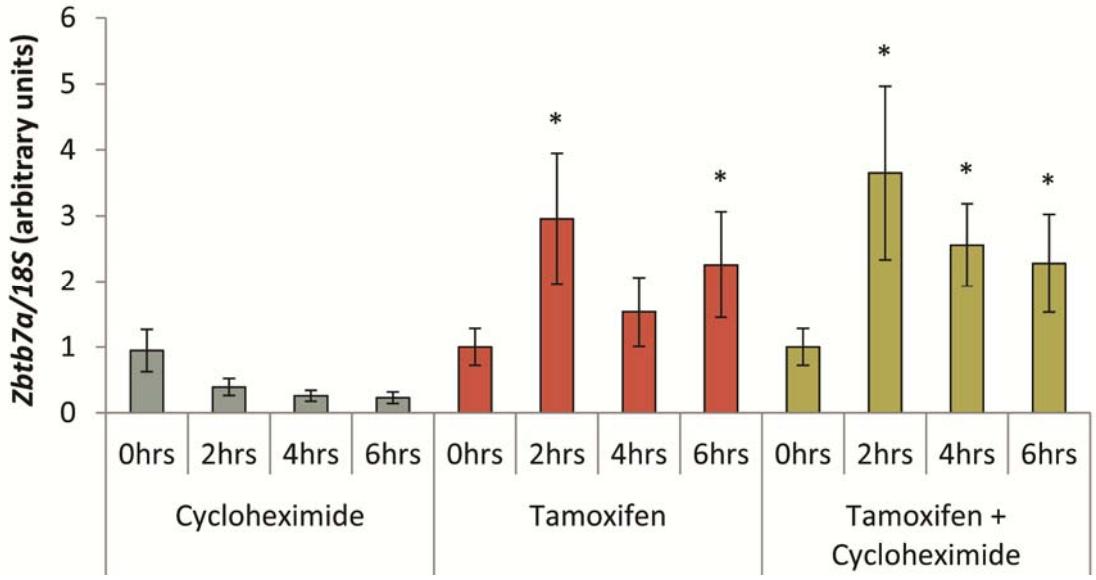


Figure 4.12. *Zbtb7a* is likely to be directly activated by KLF1-ER. K1ER cells were treated with 5 µg/mL. cycloheximide alone, 200 nM tamoxifen alone, and 5 µg/mL cycloheximide followed 30 minutes later by 200 nM tamoxifen, and were harvested at 0, 2, 4, and 6 hours after post-treatment. *Zbtb7a* mRNA levels were determined by qPCR and have been normalised to *18S* rRNA levels. Error bars represent the standard error of the mean (n = 4 for each treatment).

4.4 KLF1 is found at the promoter region of *Zbtb7a in vivo*

So far in this chapter, we have presented evidence to suggest that *Zbtb7a* is a direct target of KLF1. We next sought to determine whether KLF1 is able to bind to the promoter of *Zbtb7a*, or regulatory regions near this locus which may directly interact with proximal regulatory elements through mechanisms such a chromosomal looping. A number of KLF1

ChIP-Seq experiments have been published in human and mouse cells. These datasets were downloaded from NCBI gene expression omnibus (GEO) (mouse dataset Accession number: GSE20478 [84]; Human dataset accession number: GSE43626 [85]). Files were visualised using the Integrative Genomics Viewer (IGV v2.3) [156], and sequencing tags at genomic locations of interest surrounding the *Zbtb7a* locus were explored, shown in Figure 4.13 and Figure 4.14.

4.4.1 KLF1 binds the promoter of *Zbtb7a* in mouse

The published KLF1 ChIP-Seq performed in induced K1ER cells showed that KLF1 is enriched at various regions of the *Zbtb7a* locus [84] (Figure 4.13 A). A small peak is seen slightly upstream of the promoter, and there are peaks present in the intronic regions of the gene as well. The areas high-lighted in red were used to design probes for ChIP-qPCR analysis performed in our lab to validate these results.

We performed KLF1 ChIP-qPCR experiments using un-induced K1ER cells, and compared these to cells harvested at 2 and 8 hours post-induction (Figure 4.13 B). IgG IPs were also performed in these samples to establish a background level of DNA pull-down. In this experiment, positive regions of KLF1 occupancy were the *Klf3 1b* promoter and the promoter of the adult mouse *β-major* gene, Negative controls were 10 kb upstream of the *Klf3 1b* promoter, as well as regions surrounding the *Zbtb7a* gene, namely at +5.3 and +6.9 kb, and -7.7 kb from the *Zbtb7a* TSS. As these cells also express a YAC containing the human *β-globin* locus, we were also able to use the human adult *β-globin* promoter as a positive control and the embryonic *ε-globin* promoter as a negative control. The un-induced K1ER cells (0 hours) showed no enrichment of KLF1 at any of the loci tested, as no KLF1 was present in the nucleus of these cells. This also served as a good negative control and further defined background levels. We saw enrichment of KLF1 at the positive control regions of the *Klf3 1b* promoter, as well as both the human and mouse adult *β-globin* promoters. The negative regions showed little enrichment above the background levels defined by the 0 hour and IgG samples. We found significant enrichment of KLF1 in the first intron of *Zbtb7a,* as well as enrichment at the promoter region, but to a lesser extent. These results validate the published ChIP-Seq findings and verify that KLF1 binds to the *Zbtb7a* promoter in a mouse cell model. Interestingly, there was little difference in enrichment between the 2 and 8 hour time points, even though by 8 hours the expression levels of *Zbtb7a* have decreased substantially compared to the 2 hour time point.

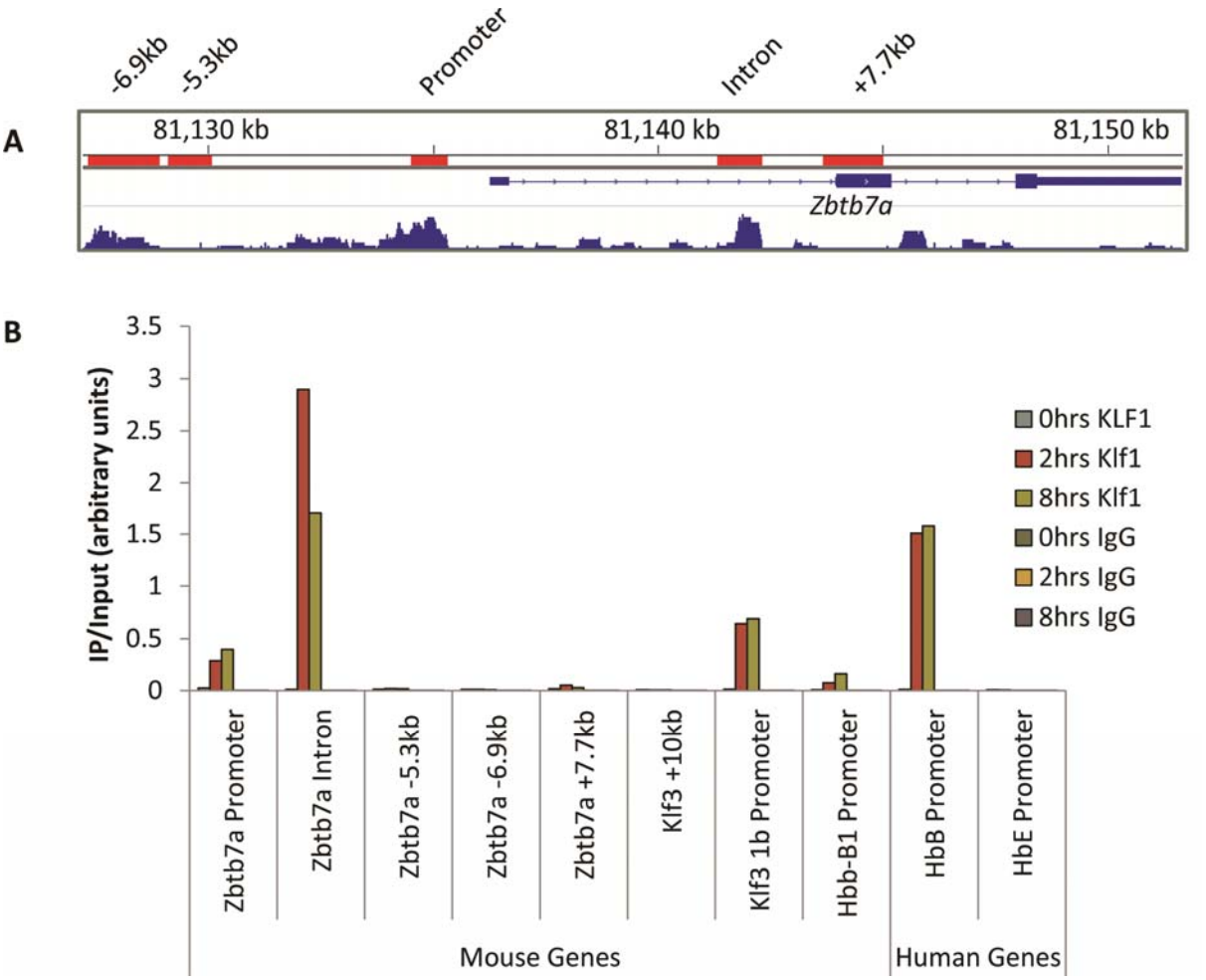


Figure 4.13. KLF1 ChIP-Seq tracks in mouse K1ER cells. (A) Data were sourced from NCBI gene expression omnibus (GEO). The raw sequencing reads from these datasets were processed using the ChIP-seq pipeline described in Chapter 2 to make bedgraph files for visualisation in IGV and to quantify sequencing tags at genomic locations of interest. Accession number: GSE20478 [84]. **(B)** ChIP-qPCR analysis of the relative enrichment of KLF1 at various genomic loci in K1ER cells. The tested genomic loci include the *Zbtb7a* proximal promoter and first intron, as well as negative areas at 6.9kb upstream of the TSS and 7.7 kb downstream. Other positive controls include the mouse *β-major globin* gene (*Hbb-b1*) promoter, the human adult *β-globin* gene promoter, and the *Klf3 1b* promoter region and negative regions of 10kb upstream of the *Klf3* TSS and the *ε-globin* promoter.

4.4.1 KLF1 binds the promoter of ZBTB7A in human erythroid cells

The published human KLF1 ChIP-Seq experiments were performed using the human erythro-leukaemia cell line, K562 (Figure 4.14 A). It is useful to note that in the human, the *ZBTB7A* gene is on the anti-sense strand, and hence runs in the opposite direction in this figure. The pattern of KLF1 occupancy in humans is similar to that which was seen in the mouse, with a small peak at the promoter of *ZBTB7A* and a few peaks in intronic regions. The human ChIP-Seq also showed enrichment at two regions far upstream of the TSS, at around +8.2- and +15.4 kb respectively, which may be potential enhancer sites. The red highlighted

areas again indicate the regions used for probe design. These probes were then used to analyse our own KLF1 ChIP experiments by qPCR, seen in Figure 4.14 B.

We tested the occupancy of KLF1 at the *ZBTB7A* promoter in K562 cells by ChIP-qPCR. These results generally validated the published ChIP-Seq findings, however some discrepancies were found. Our ChIP-qPCR demonstrated that KLF1 was highly enriched at the promoter and both of the peaks of intron 1 of *ZBTB7A*, identified from the published ChIP-Seq. The two upstream enhancer peaks seen in the ChIP-Seq experiment showed little enrichment in our ChIP-qPCR experiments. There was no enrichment at the negative control, the embryonic *ε-globin* promoter, but also no enrichment seen at the adult *β-globin* promoter, which is a typical positive control for KLF1. This may be due to the fact that these cells are not currently expressing high levels of the adult *β-globin* gene as they are more foetal in nature [167].

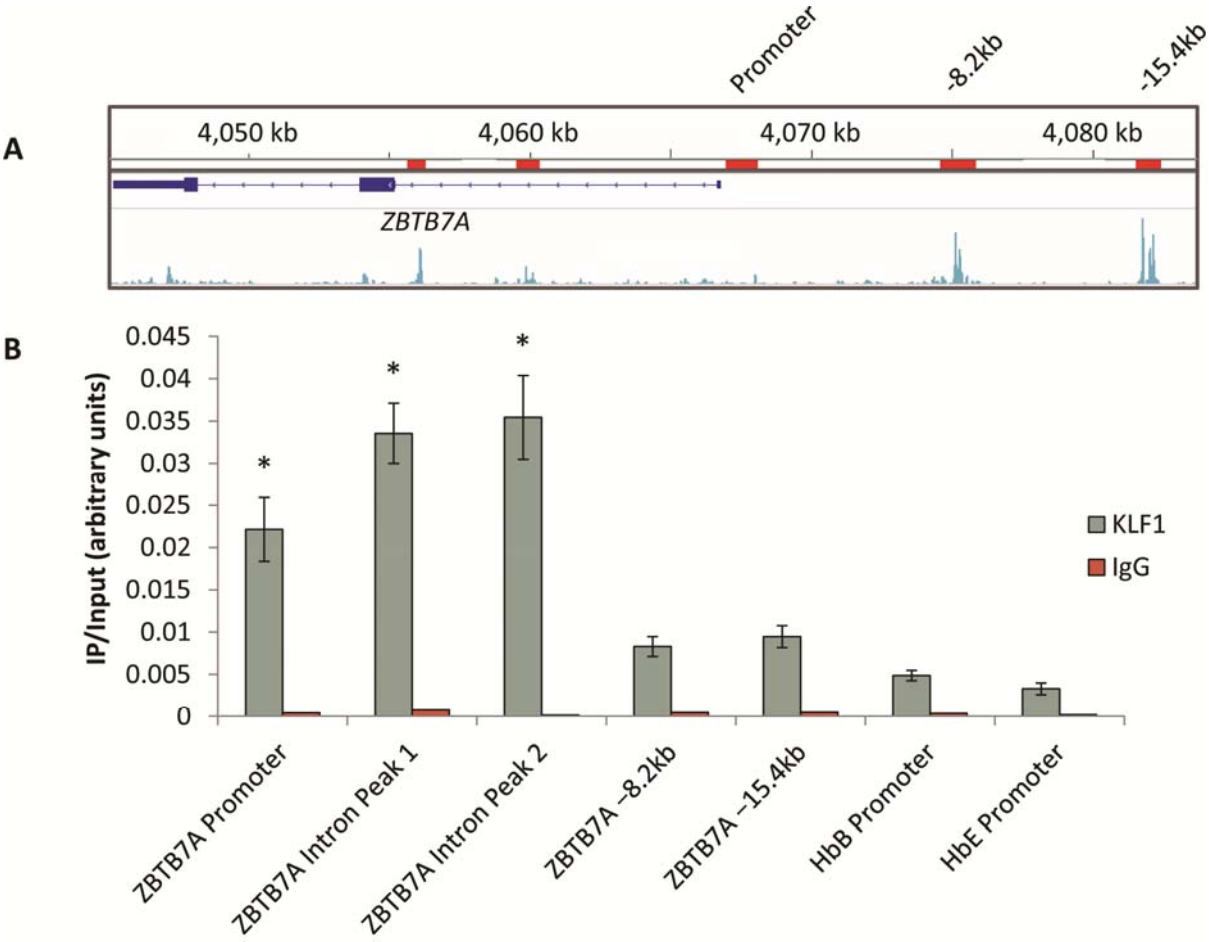


Figure 4.14. KLF1 ChIP-Seq tracks in human K562 cells. (A) Data were sourced from NCBI gene expression omnibus (GEO). The raw sequencing reads from these datasets were processed using the ChIP-Seq pipeline described in Chapter 2 to make bedgraph files for visualisation in IGV and to quantify sequencing tags at genomic locations of interest. Accession number: GSE43626 [85]. **(B)** ChIP-qPCR analysis of the relative enrichment of KLF1 at various genomic loci in K562 cells. The tested genomic loci include the *Zbtb7a* proximal promoter and the first intron, as well as negative areas at 8.2kb and 15.4kb upstream of the TSS. Other negative controls include the human adult *β-globin* gene promoter, and the embryonic *ε-globin* promoter, neither of which are expressed in this cell line.

4.5 Chapter 4 Discussion

In this chapter we have demonstrated that *Zbtb7a* expression is directly regulated by the master erythroid regulator, KLF1. It had previously been suggested that KLF1 activates *Zbtb7a,* as it has been shown to be down-regulated in the absence of KLF1 in *Klf1*$^{-/-}$ mouse experiments, as well as up-regulated when KLF1 is introduced into the nucleus in the KLF1 inducible cell line, K1ER [87]. However, it was not previously known whether KLF1 directly regulates *Zbtb7a*, or whether it is indirectly regulated by another KLF1 target gene. We have validated the results found in microarrays performed in both E14.5 foetal livers of *Klf1*$^{+/+}$ and *Klf1* null embryos, as well as induced and un-induced K1ER cells, performed by the Perkins lab [87], and we have investigated this relationship further to demonstrate that *Zbtb7a* is a *bona fide* direct target of KLF1. This further emphasises the importance of *Zbtb7a* as a transcriptional repressor in erythroid cells.

The majority of deregulated genes found in our microarray analysis were down-regulated in the *Klf1*$^{-/-}$ samples, consistent with KLF1's role as a potent activator. KLF1 has been shown be capable of acting as a repressor under certain conditions [80, 208, 209], and it may therefore be possible that some of the genes with increased expression in the absence of KLF1 may well be direct targets of this protein. However, as KLF1 acts predominantly as an activator, it is more likely that the majority of these are indirect targets, regulated instead by other repressive transcription factors which KLF1 directly activates. It should also be noted that some of the differentially expressed genes are likely to be a result of the phenotype, rather than as a direct effect of the loss of KLF1. For instance, some genes may be sensitive to hypoxia, and may be altered as a result of the anaemia seen in these animals.

As KLF1 is important in many aspects of erythroid biology, we would expect to see many erythroid genes differentially expressed in the *Klf1*$^{-/-}$ samples, which is indeed the case. Some of the more interesting genes involved in these cells found to be differentially expressed are discussed below.

The most highly down-regulated gene on the list was found to be *Ahsp* (*α*-haemaglobin stabilising protein). This is an erythroid-specific protein that forms a stable complex with free *α*-globin peptide chains, which protects cells from the oxidative effects of the precipitation of free *α*-globin chains in the cytoplasm [188]. It is critical in preventing reticulocytosis, a common symptom of anaemia, and ensuring the normal maturation of erythrocytes.

Sox6 was also down-regulated in the absence of KLF1. As discussed in Chapter 1, SOX6 is involved in looping interactions between BCL11A and regions of the *β-globin* locus,

resulting in the silencing of foetal *γ-globin* and [120]. It also plays a role in the silencing of the embryonic *εy-globin* in mouse [118, 119]. *Sox6* has been found to be important in erythroid development and survival, as *Sox6*$^{-/-}$ mice display delayed development of erythrocytes, which are often misshapen and have a shortened life-span [190]. *Sox6* has not previously been reported as a KLF1 target gene. *Bcl11a* also appears on this list, however its fold change is much less than that of *Sox6* [105-107, 112, 120]. *Bcl11a* is a reported KLF1 target [112, 185], and its role in erythropoiesis has already been extensively discussed in this thesis.

Klf3 was also found to be down-regulated in the absence of *Klf1*. Our lab has studied this gene and its involvement in erythropoiesis quite extensively [91, 104]. These mice have been found to have a compensated anaemia, where they displayed decreased mean cell haemoglobin, but had increased numbers of erythrocytes to compensate for this defect. Their erythrocytes also had a reduced life-span in circulation. *Klf3* shares many of the same targets as *Klf1,* so it may also be regulating erythroid genes which play a role in cell stability and survival [91, 104, 166]. Together with *Klf8, Klf3* may have a role in regulating globin gene expression [104]. We have also found the *Klf3* has a role in lymphopoiesis through the regulation of Marginal Zone B cells in the spleen [191].

Lmo2 was not found annotated in DAVID as a gene involved in haematopoietic pathways, but its role in erythropoiesis is well established. This protein is a non-DNA-binding component of a protein complex composed of multiple master regulators of erythropoiesis, including GATA-1, SCL/TAL1 and LDB1. Knock-down of *Lmo2* in human K562 cell lines showed reduced chromatin occupancy of these transcription factors at their target genes [204]. [49]. *Lmo2*$^{-/-}$ mouse embryos had a complete lack of erythropoiesis in the yolk sac, which results in early embryonic lethality [201]. Erythroid differentiation has also been shown to be blocked in embryonic stem cells lacking *Lmo2* [201, 202]. Previous microarrays have also found *Lmo2* to be down-regulated in the absence of *Klf1* [87].

Klf1 also appears on this list, but somewhat counterintuitively, it is only down-regulated 2.5-fold in the *Klf1*$^{-/-}$. This has to do with how the mouse was originally generated, where a neomycin resistance cassette was inserted into the *Klf1* gene to replace a portion of the zinc fingers of the protein found at the C-terminus, and thereby abolish its DNA-binding capacity [90]. For this reason, parts of the gene are still transcribed. This microarray utilises 25 probes covering the *Klf1* transcript. As the DNA-binding zinc fingers of KLF1 are highly homologous to other members of the family, it is likely that most of these probes are found in regions that are unique to *Klf1*, predominantly the N-terminus of the transcript. This portion

of the transcript is still transcribed in the *Klf1*$^{-/-}$ mice, and thus, the down-regulation seen in these animals is either due to the fact that some probe sets are found in regions that are not transcribed in the *Klf1*$^{-/-}$ samples, or simply due to the fact that KLF1 regulates itself.

Interestingly, the adult *β-globin* gene *β-major* is not the most differentially expressed gene, or even in the top 100, even though this is one of KLF1's most well-known target genes, and one of the major reasons of embryonic lethality of the *Klf1* knockout [82, 86]. Also interesting to note, the embryonic globin *βh1* is up-regulated in the absence of *Klf1*. This is likely due to the fact KLF1 is no longer able to activate repressors of *βh1*, such as *Bcl11a* [110], and potentially *Zbtb7a* [138].

Zbtb7a was also found to be down-regulated in the *Klf1*$^{-/-}$ samples. Interestingly, the down-regulation seen at this gene was found to be greater than that of either *β-major* or *Bcl11a*, both of which have been shown to be directly activated by KLF1. We have already discussed evidence that *Zbtb7a* plays a role erythropoiesis in this thesis, using our own data (Chapter 3) and that of others [137, 138, 164]. It is interesting to note that the qPCR validation of the microarray results indicated that the magnitude of differences between *Klf1*$^{+/+}$ and *Klf1*$^{-/-}$ samples at E13.5 was less than that seen at E14.5. There are two possible explanations for this. Firstly, as these *Klf1* null embryos have already succumbed to anaemia at this time point, it is possible that the RNA species as a whole have degraded more in these samples. Another possibility is by E14.5, the up-regulation of *Zbtb7a* is further increased in the *Klf1*$^{+/+}$ samples. The effect is likely due, in part, to both of these factors. Finally, while *Zbtb7a* expression was found to be decreased in the *Klf1*$^{-/-}$ samples at both E13.5 and E14.5, the transcript was still present at detectable levels. This may be due to the fact that it is also regulated by GATA1 [137], but may also be due to other KLF proteins, such as KLF2 which is closely related to KLF1 [96], that can potentially compensate for the loss of KLF1 at some target genes.

The DAVID analysis also identified genes involved in other haematopoietic pathways. *Klf1* was long believed to be exclusively expressed in erythroid cells. In the last ten years, however, there has been evidence to suggest *Klf1* is expressed in macrophages, and regulates the expression of an important factor in the activation of innate immunity and development of Th1 responses, namely IL-12 p40, suggestive of a role in inflammation pathways [81]. It has also been found to play a role in megakaryopoiesis [210]. As the foetal liver represents a mixed population of haematopoietic cells, this data may give useful insights into the function of *Klf1* in other cell types. The non-erythroid genes identified here may be evidence of other roles of Klf1 in the haematopoietic system. For instance, *Zbtb32*, also known as *Plzp* and

Rog, appears to be involved in various branches of lymphopoiesis. Deregulation has been shown to lead to increased T cell proliferation [192], as well as appearing to de-sensitise T cells against activation [193]. It has also been shown to play a role in the increased production of natural killer cells in response to viral infection [194]. While expanding the role of *Klf1* in haematopoiesis is interesting, the relevance of these genes to this thesis is limited, and will not be discussed further.

We also found *Zbtb7a* expression was up-regulated upon induction of the KLF1-inducible cell line, K1ER. This happened soon after induction at around 2 hours, suggesting KLF1-ER activates this gene directly, as it is unlikely that enough time had elapsed to both transcribe and translate a downstream activator driven by KLF1. The up-regulation of *Zbtb7a* was found to be short-lived, and transcript levels began to decrease shortly after induction. This may have been due to the regulation of transcript levels through mechanisms of mRNA degradation, such as previously reported microRNAs [211], or potentially by KLF1 activating a repressor of *Zbtb7a* which out-competes KLF1 and down-regulates *Zbtb7a*. This repressor may even be ZBTB7A itself as we found ZBTB7A binds to its own promoter in ZBTB7A ChIP-Seq experiments (Figure 3.14). We also showed that *Zbtb7a* was up-regulated even in the presence of the translation inhibitor cycloheximide, which provides good evidence that it is a direct target of KLF1-ER in these cells. If it were an indirect target of KLF1-ER, *Zbtb7a* expression would not be up-regulated, as *de novo* protein synthesis would be required to translate the KLF1-ER target that activates *Zbtb7a*.

KLF1 was found to bind at the promoter of *ZBTB7A* in published ChIP-Seq experiments in both human and mouse cells, as well as in the first intron and upstream potential enhancer regions, which we validated by ChIP-qPCR experiments. We found enrichment of KLF1 at the proximal promoter in both humans and mice, as well as at the first intron. It has previously been reported that KLF1 is found at the first intron of many of its target genes [84], so the finding that it binds in the first intron of *ZBTB7A is* again consistent with the view that *ZBTB7A* is a direct target of KLF1. The binding at the distal upstream region of *ZBTB7A* is potentially indicative of an enhancer region. This could be further explored using various techniques such as Chromatin Conformation Capture (3C), which is used to detected looping of this region back to the promoter of *ZBTB7A*, or by Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET), which is a high-throughput sequencing technique to determine genome-wide *de novo* long-range chromatin interactions.

It is interesting to note that, in the experiments conducted here and previously, only a modest up-/down-regulation of *Zbtb7a* is seen in response to KLF1. As detectable levels of

Zbtb7a are still present in both *Klf1*$^{-/-}$ foetal livers and un-induced K1ER cells, this indicates other transcription factors may also be responsible for activating *Zbtb7a*. Previous research has shown that GATA-1, another master erythroid regulator, activates *Zbtb7a* [137]. Here a GATA-1 inducible cell line similar to K1ERs, called G1E cells, was used to show that *Zbtb7a* expression increases upon GATA-1 induction. However, *Zbtb7a* was found to increase at 12 hours post-induction, roughly the same time that *Klf1* was also found to be up-regulated. As GATA-1 is known to activate the expression of *Klf1* [212], and we know KLF1 acts quickly to up-regulate the expression of *Zbtb7a*, it may be that this gene is an indirect target of GATA-1, and under the direct regulation of KLF1. No cycloheximide experiments were performed in this paper, and hence it is not possible to definitively conclude whether or not KLF1 is needed for the expression of *Zbtb7a*. A ChIP-qPCR of GATA-1 was performed, and they find enrichment of GATA-1 at a region 400 bp upstream of the TSS of *Zbtb7a*. Given this, GATA-1 may be the primary activator of *Zbtb7a*, which would explain why *Zbtb7a* expression is high prior to activation by KLF1, and KLF1 may act to further up-regulate *Zbtb7a*. Or alternatively, GATA-1 and KLF1 work together in erythroid cells to activate *Zbtb7a,* as they do at many erythroid genes [84]. If this is the case, an additional non-erythroid regulator may be responsible for the ubiquitous expression of *Zbtb7a*. This is a likely scenario, given our data from Chapter 3, demonstrating *Zbtb7a* expression in non-erythroid tissues.

This opens up the interesting possibility that *Zbtb7a,* like *Klf3* and many other erythroid genes, has an erythroid-specific transcript. As our qPCR primers lie in exon 2 of this gene, we may well be missing the full scale of the up-regulation of *Zbtb7a* by KLF1 in erythroid cells. Chapter 5 discusses the prospect of an erythroid-specific transcript of *Zbtb7a.*

CHAPTER 5 – AN ERYTHROID-SPECIFIC TRANSCRIPT OF *ZBTB7A*

5.1 Chapter 5 Introduction

Many genes have various transcripts which are differentially expressed in tissues. This dynamic regulation adds an additional dimension to the control of gene expression, as these transcripts can be differentially regulated by tissue-specific transcription factors. Erythroid tissues are a good example of this phenomenon as there are a number of genes which have erythroid-specific transcripts, some of which have already been discussed in this thesis. The expression of erythroid-specific transcripts from a number of genes is regulated by the erythroid master regulators, KLF1 and GATA-1. Some transcripts give rise to novel proteins, as is the case with porphobilinogen deaminase (PBGD), mentioned in Chapter 1, which is regulated by a housekeeping promoter, active in many tissues, as well as an erythroid-specific promoter [8]. This erythroid-specific promoter induces expression of a different isoform of the PBGD protein, with unique functions in erythroid cells. There are also many cases where the tissue-specific transcripts will produce the same final protein but where the erythroid-specific promoter allows fine-tuning of expression level in erythroid cells only. One example, driven by GATA-1, is Uroporphyrinogen-III (URO) synthase, a haeme-biosynthetic enzyme. This gene has been found to possess two promoters, one a house-keeping promoter active in many tissues, and the other active only in erythroid cells and driven by GATA-1. Another example we have previously discussed in detail is *Klf3*. This gene has a ubiquitous promoter, *1a*, and an erythroid-specific promoter, *1b*, which is driven by KLF1 in erythroid cells, and again, both transcripts give rise to the same protein [94].

We hypothesised that *Zbtb7a* may also possess an erythroid-specific transcript. It has previously been reported that its expression is driven by GATA-1 in the inducible cell line, G1E [137], although transcript is present even in the absence of any GATA-1 in these cells. We have shown here that it is also driven by KLF1 in the inducible cell line, K1ER, although again, its expression was detected prior to induction. We have also detected expression of *Zbtb7a* in non-erythroid tissues of the mouse and in un-induced primary human CD34+ HSCs. From Western blots performed in the lab, we have seen no evidence of a novel protein isoform, so we hypothesised this potential erythroid-specific transcript will give rise to the same protein as the ubiquitous transcript. The ubiquitous transcript for *ZBTB7A* in both mice and humans is composed of three exons. The translation initiation codon (ATG) lies within the first few bases of the second exon for both species. We therefore hypothesised that any

differential erythroid-specific exons will lie upstream of exon 2, in what is currently annotated as the 5'UTR.

In this chapter, we investigated the potential of an erythroid-specific start point of transcription by analysing data from the Functional Annotation of the Mammalian Genome (FANTOM5) Semantic catalogue of Samples, Transcription initiation And Regulators (SSTAR) database. The FANTOM5 consortium used Cap Analysis of Gene Expression (CAGE) to map TSSs across a wide collection of samples from 975 human and 399 mouse tissues, primary cells and cell lines, and used this to create a comprehensive expression atlas of all the actively expressed genes [213, 214]. Their findings suggest that true "housekeeping" genes are not as common as is widely believed, but rather that there are a number of composite TSSs at the start of a gene, which cluster together and show independent cell-type-specific profiles [213]. This data also serves as a valuable bank of transcript-specific information across human and mouse tissues. CAGE works by capturing tags, which are nucleotide sequences roughly 30 bp in length, from the 5' end of all mRNA transcripts present. These tags are then sequenced and aligned to the genome to essentially form peaks over the areas with the greatest numbers of tags. This allows one to determine the various start points of transcription of all the transcribed genes in the cell. This method, however, does not give any indication of the length of the first exon, or alternative splicing that may be occurring in the 5'UTR. We therefore only used this data as an indication of the presence of an erythroid-specific TSS, and employed 5' Rapid Amplification of cDNA Ends (RACE) to determine the exact structure of these possible erythroid-specific transcript variants. For this analysis we used the K1ER cell line and erythroid cells derived from primary human HSCs from both foetal and adult sources.

5.2 A conserved erythroid-specific *ZBTB7a* promoter in human and mouse

The promoter region of the *Zbtb7a* gene has previously been investigated [137]. Here they found the presence of two CACCC- and two GATA- boxes, conserved in both humans and mice. These are consensus binding sites for the erythroid factors KLF1 and GATA-1, among others. The region described in this paper was found to lie directly adjacent to the start of exon 1 in the mouse, and aligns to a region which is roughly 270 bp upstream of the first exon of *ZBTB7A* in humans. We hypothesised that this is a potential erythroid-specific promoter, and activation by either KLF1 or GATA-1, or potentially both factors together,

results in the expression of an erythroid-specific transcript variant. We next sought further evidence for these transcripts in mouse and human cells.

5.3 A *Zbtb7a* transcript variant most highly expressed in foetal liver

We used the FANTOM5 SSTAR database to investigate whether *Zbtb7a* has a transcription start point which is enriched in mouse erythroid cells or tissues. In mice, seven different *Zbtb7a* TSSs are annotated, numbered in order of abundance of expression, shown across the tissues and cells investigated. The erythroid tissues included in the mouse were foetal liver at various developmental time points, which from around E14-E15 is a highly erythroid organ [46], as well as neonatal spleen at various time points. The magnitude of expression in particular tissues of the top five transcripts, labelled *Zbtb7a p1-p5*, have been arranged in order of expression according to *Zbtb7a p5,* which is most highly expressed in erythroid tissues (Figure 5.1). It is clear that the *Zbtb7a p5* transcript is neither very highly, nor widely expressed, compared to other *Zbtb7a* transcripts. But it is the most highly expressed transcript in the foetal liver at E12, E14 and E15 of development and, as such, is a good candidate for an erythroid-specific transcript.

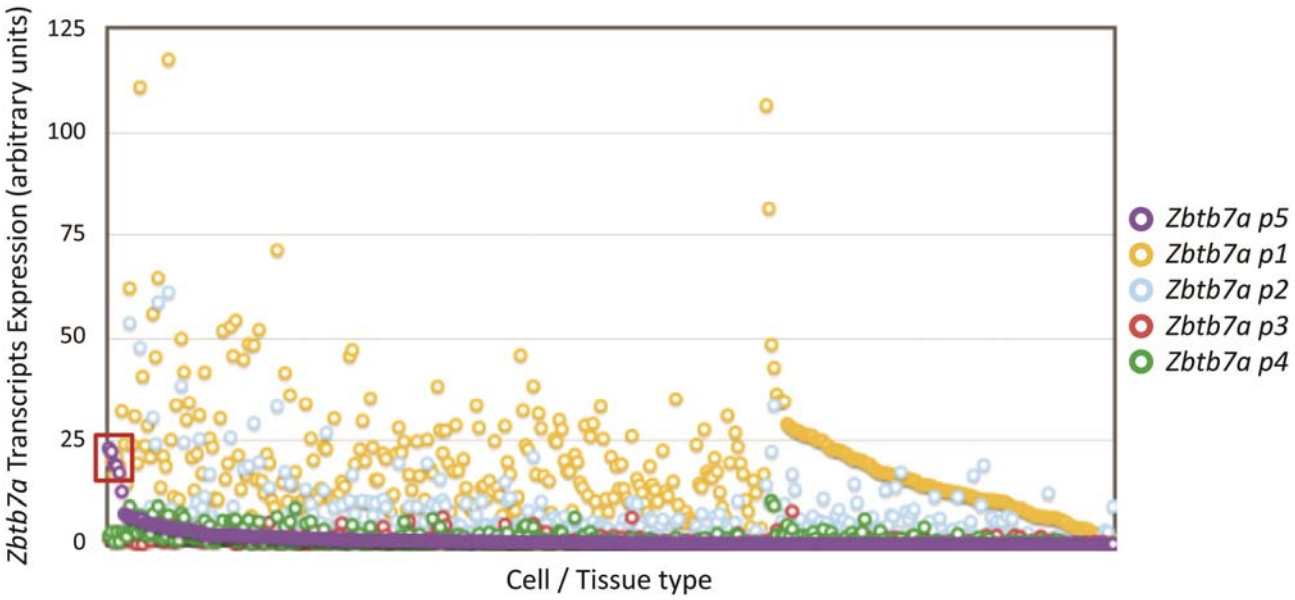


Figure 5.1. Mouse *Zbtb7a* transcript variants and their expression levels in various human tissues, discovered from the FANTOM5 (Functional Annotation of the Mammalian Genome) SSTAR (Semantic catalogue of Samples, Transcription initiation And Regulators) Project. These variants have been found using Cap Analysis of Gene Expression (CAGE). Tissues have been arranged based on the values for variant *Zbtb7a p5*, which is most highly expressed at E15 in the foetal liver. The y-axis here represents the relative expression levels, and along the x-axis are the various cell and tissue types. The red box indicates the erythroid tissues where *p5* transcript is the most highly expressed *Zbtb7a* transcript.

Table 5.1 lists the relative expression of each different isoform in the ten tissues with the greatest level of *Zbtb7a p5* expression. The highest expression of this transcript is in foetal liver at E15, which is when the *Zbtb7a*$^{-/-}$ is embryonically lethal due to severe anaemia. The second highest expression level is in the foetal liver at E14, which is the tissue where we saw the highest expression of *Zbtb7a* in those we investigated in Chapter 3 (Figure 3.2). Out of the top ten tissues in which the *p5* transcript is most highly expressed, seven of them are erythroid in nature, and one is megakaryocytic, which is closely related to the erythroid branch of haematopoiesis, and is also driven by GATA-1 [215]. The *Zbtb7a p1* transcript is also highly expressed in these tissues, but from Figure 5.1, we can see that it is also the most ubiquitously expressed transcript. By E16, the *p1* transcript has overtaken *p5* in expression. This may be due to the fact that at around E16 is when the foetal liver is beginning to become more hepatic in nature [216]. We hypothesise that the *p1* transcript is a ubiquitously expressed *Zbtb7a* transcript, whereas the *p5* transcript may be erythroid-specific.

Table 5.1. The various *Zbtb7a* transcripts expressed in mouse cell and tissue types, from the FANTOM5 (Functional Annotation of the Mammalian Genome) SSTAR (Semantic catalogue of Samples, Transcription initiation And Regulators) database. The tissues have been arranged according to expression levels of the *Zbtb7a p5* transcript, most highly expressed in erythroid-type tissues in the mouse. The numbers represent the expression levels of each isoform relative to a housekeeping gene.

Sample	*Zbtb7a p5*	*Zbtb7a p1*	*Zbtb7a p2*	*Zbtb7a p3*	*Zbtb7a p4*
Foetal liver, E15	23.13	23.12	1.83	1.62	2.03
Foetal liver, E14	22.26	19.79	2.52	1.80	2.92
Foetal liver, E12	19.35	16.68	0.67	1.17	0.50
Foetal liver, E16	18.81	20.98	2.89	2.89	2.89
Foetal liver, E13	17.10	18.87	1.97	1.57	0.59
CD41+ megakaryocyte	12.59	32.18	2.80	0.70	2.80
Spleen, neonate N10	7.35	24.11	1.32	1.88	2.07
Foetal liver, E17	6.87	14.51	3.56	2.29	1.78
CD326+ enterocyte	6.63	62.12	53.74	6.63	9.07
Intestine, neonate N6	6.53	23.88	8.29	0.50	3.52

The TSS for the *Zbtb7a p5* transcript identified in the FANTOM5 database lies 48 bp upstream of the first exon. This can be seen below in Figure 5.2 A.

5.3.1 5’ RACE design

In order to investigate the presence of an erythroid-specific transcript in more detail, we performed 5’RACE using un-induced and induced K1ER cells. We chose the time point of 2

hours post-induction as this is when we found the highest expression of the *Zbtb7a* transcript (as determined by qPCR with primers to exon 2) (Figure 4.10). As *Zbtb7a* has a GC rich first exon and GC rich regions are challenging to amplify by PCR, we designed 6 reverse primers in exon 2 to maximise our chances of success. Each successive primer was located further downstream of the annotated TSS and thus resulted in the amplification of a larger product. This sequential approach maximises the likelihood of the *Zbtb7a* being the only product to be amplified. If a non-specific product was being generated, it is unlikely that each primer would be found to amplify it. Following the manufacturer instructions for this 5'RACE kit, these primers were designed with melting temperatures above 70°C. During the synthesis of 5'RACE ready cDNA from the whole RNA extracts, a 5' cap is attached to each transcript, which provides a sequence for the binding of a generic forward primer provided by the kit. Below is a schematic of the genomic DNA of *Zbtb7a* (Figure 5.2 A), as well as the RefSeq cDNA transcript from NCBI (NM_010731.3) (Figure 5.2 B), each composed of three exons, and indicating the translation start site, and 5' and 3' UTRs. The figure also indicates the start point of transcription, identified in the FANTOM5 database, of the hypothesised erythroid-specific transcript (*p5*) of *Zbtb7a*. This start point is located 48 bp upstream of the annotated TSS. The asterisk denotes the site of the erythroid-specific promoter [137] (Figure 5.2 A), which in the case of the mouse is adjacent to the annotated TSS. The reverse primers for the 5'RACE are shown lying in exon 2 (Figure 5.2 B).

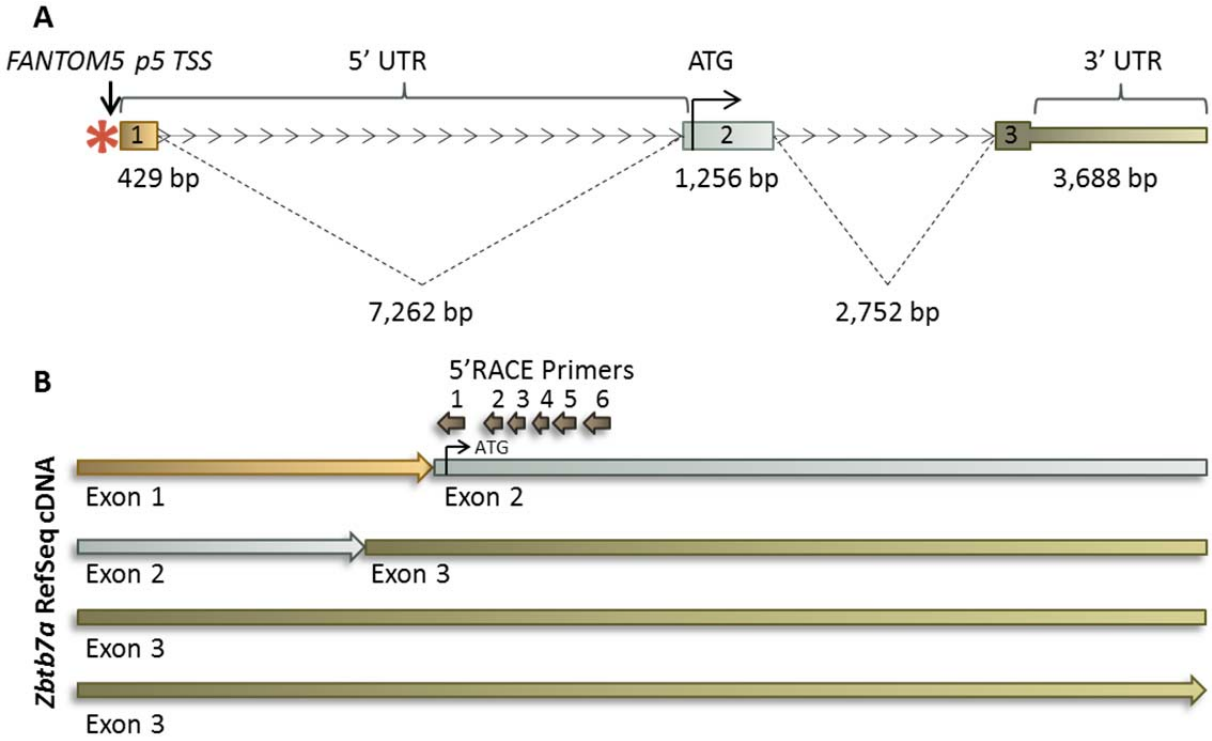
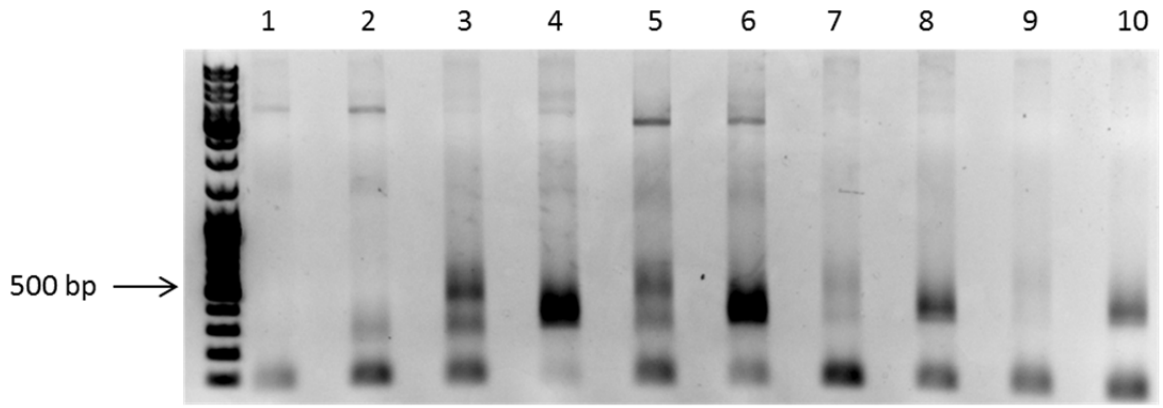
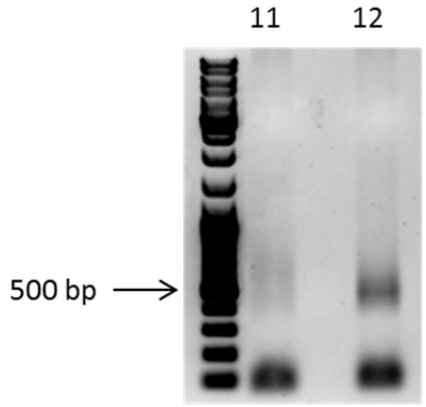


Figure 5.2. 5'RACE primer design for amplifying the 5' end of the mouse *Zbtb7a* transcript. 5'RACE was performed according to instructions in the kit. cDNA is produced from RNA with a 5' cap recognised by forward primers supplied by Clontech. Gene specific reverse primers were designed lying in Exon 2 to amplify potential transcript variants. **(A)** Genomic sequence of *Zbtb7a* showing exons 1-3 and intergenic regions, 5'- and 3' UTRs and the start point of translation, marked with ATG. Asterisk indicates the erythroid-specific promoter. **(B)** RefSeq cDNA sequence of *Zbtb7a* (NM_010731.3) showing exons and start point of translation, and each individual 5'RACE primer, labelled 1-6.

5.3.2 5'RACE Results

PCRs of un-induced and 2 hour induced K1ER cell cDNA samples were set up with the generic forward primer from the kit in each sample, and each individual reverse primer. The PCR products were run on an agarose gel (Figure 5.3). For each primer the induced samples displayed a unique band, but primers 2, 3 and 4 worked particularly well. The un-induced samples often showed two bands, indicative of two transcripts. The bands for the induced samples were more pronounced, suggesting it may be made up of a few different length fragments. It appeared that at least one of the fragments in the induced samples was of a similar size to the smaller band seen in the un-induced samples, and the larger band seen in the un-induced samples was not present in the induced samples. For the reactions where no bands were detected in the un-induced samples, there was still a band seen in the induced samples, indicating that there is a novel transcript being expressed in the presence of KLF1-ER, which is encouraging that this transcript may be erythroid-specific.





Lane	Sample	Primer
1	K1ER un-induced	1
2	K1ER induced 2 hours	
3	K1ER un-induced	2
4	K1ER induced 2 hours	
5	K1ER un-induced	3
6	K1ER induced 2 hours	
7	K1ER un-induced	4
8	K1ER induced 2 hours	
9	K1ER un-induced	5
10	K1ER induced 2 hours	
11	K1ER un-induced	6
12	K1ER induced 2 hours	

Figure 5.3. Induced K1ER cells express a different transcript of *Zbtb7a*. Whole mRNA was extracted from un-induced and induced K1ER cells and 5'RACE ready cDNA was made using the 5'RACE kit from Clontech. PCR reactions were set up with 5'RACE ready cDNA and gene specific reverse primers which lie in exon 2 of *Zbtb7a*. Products were run of a 2% agarose gel. A 100 bp DNA ladder was used to estimate the sizes of the fragments.

5.3.3 5'RACE Sequencing

These products were sequenced and mapped to the *Zbtb7a* locus using the BLAT search tool on the UCSC Genome Browser [217, 218]. Some of the sequences were found not to be *Zbtb7a*, or to lie in unexpected regions of the gene, such as exon 3, or on the complementary strand, and these were discarded from further analysis. Most of the sequences appeared to have start sites that started in the first exon of the gene. The sequences that mapped to the *Zbtb7a* transcript are tabulated below in Table 5.2, which shows the sample type in column one, un-induced in grey and induced in red, and the base number of exon 1 that the sequence was found to start at in the second column. The third column groups the samples together based on their TSSs.

Table 5.2. Collated sequencing data from un-induced and induced K1ER 5'RACE samples. The first column indicates the sample type, the second is the start point of transcription found in the first exon, this is the number of base pairs downstream of the defined TSS, and the third column groups the samples together by their TSS. Grey rows indicated un-induced samples and red rows indicate induced samples.

Sample	TSS	TSS Grouping
K1ER Un-induced	95	TSS 1
K1ER Un-induced	99	
K1ER Un-induced	102	
K1ER Un-induced	103	
K1ER Un-induced	103	
K1ER Un-induced	103	
K1ER Un-induced	108	
K1ER Un-induced	108	
K1ER Un-induced	108	
K1ER Un-induced	109	
K1ER Un-induced	116	
K1ER Un-induced	134	TSS 2
K1ER Un-induced	142	
K1ER Induced 2 hours	154	
K1ER Induced 2 hours	234	
K1ER Un-induced	238	TSS 3
K1ER Un-induced	240	
K1ER Un-induced	240	
K1ER Induced 2 hours	240	
K1ER Induced 2 hours	240	
K1ER Induced 2 hours	240	
K1ER Induced 2 hours	240	
K1ER Induced 2 hours	240	
K1ER Induced 2 hours	241	
K1ER Induced 2 hours	240	
K1ER Induced 2 hours	240	
K1ER Induced 2 hours	240	
K1ER Induced 2 hours	240	
K1ER Induced 2 hours	241	
K1ER Induced 2 hours	277	TSS 4
K1ER Un-induced	277	
K1ER Un-induced	277	
K1ER Induced 2 hours	284	
K1ER Un-induced	286	
K1ER Induced 2 hours	289	
K1ER Un-induced	306	TSS 5
K1ER Un-induced	314	
K1ER Un-induced	324	

There are five distinct groupings of start sites, but Groups 1 and 3 stand out, as the transcripts falling in these groups are largely divided by un-induced and induced samples respectively. Group 1 is found to start at around 103-108 bp downstream of the RefSeq TSS

in exon 1. All of the sequences with this start site were found to be from the un-induced samples. The start point of Group 3 was found to be further downstream in exon 1 at around 240 bp downstream of the TSS. There were a few transcripts from the un-induced samples that also started here, but the overwhelming majority were from induced samples, suggesting that it is an erythroid-specific start site. Figure 5.4 depicts the various TSSs that were found from compiling the 5'RACE data. The red arrows indicate the induced samples, and the grey arrows the un-induced. The clustering of the transcript start point of the induced samples at the +240 site is clear. Interestingly, this does not correspond to the *p5* TSS from FANTOM5 that was highly expressed in the foetal liver of the mouse (Figure 5.1), which was found to lie 48 bp upstream of exon 1. There are a number of reasons for why this might be the case, the primary one being that these are different cell types. While the foetal liver is predominantly erythroid at the time points investigated by FANTOM5, it still represents a mixed population of cells, and it is possible that the TSS identified is not erythroid-specific. However, the transcript we uncovered here by 5'RACE is clearly induced in K1ER cells, which indicates that it may be under the direct regulation of KLF1. As KLF1 is a master regulator of erythropoiesis, this is a good indication that this transcript is indeed erythroid-specific.

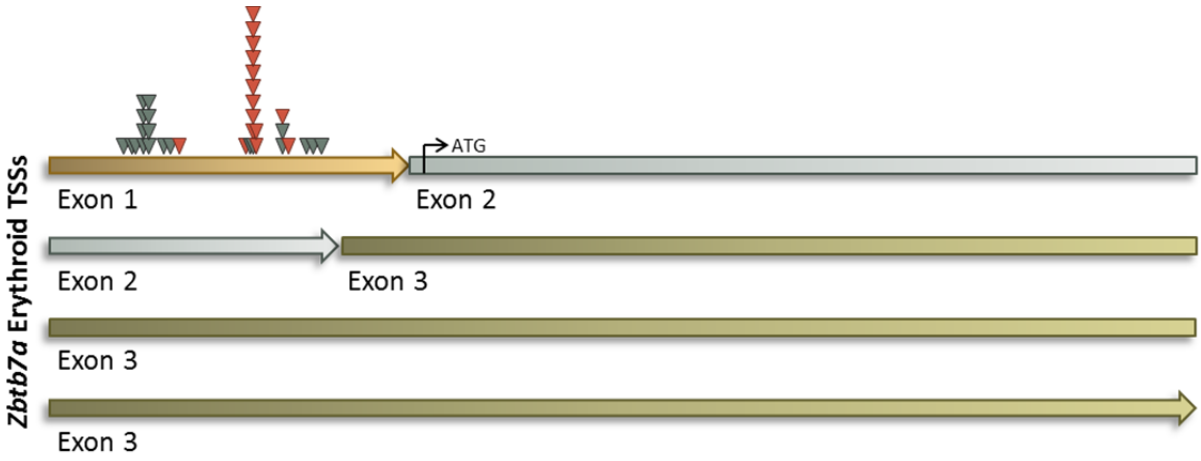


Figure 5.4. Transcriptional start sites of *Zbtb7a* transcripts in un-induced and induced K1ER cells found by 5'RACE. Grey arrow heads indicate un-induced samples and red arrow heads indicate induced samples. Induced cells appear to have a novel *Zbtb7a* transcript which begins at base pair +240 downstream of the annotated TSS of *Zbtb7a.*

5.4 *ZBTB7A* has a transcript variant highly expressed in reticulocytes in humans

We also used the FANTOM5 SSTAR database to investigate whether there were any transcripts that were highly expressed in human erythroid cells or tissues. The most relevant

erythroid human cell type that was investigated was the reticulocyte, derived from CD34+ HSCs. These are immature red blood cells just before they extrude their nuclei to form fully mature erythrocytes. These cells still contain RNA and hence could be used in CAGE experiments. In human cells, they identified eight different transcripts of *ZBTB7A*, numbered in order of relative expression. In this case, the *ZBTB7A p3* transcript was most highly expressed in reticulocytes compared to other transcripts, and the level of expression in these cells was second only to mast cells. Figure 5.5 shows the tissues of the top five most highly expressed transcripts, labelled *ZBTB7A p1-p5,* which have been arranged in order of expression according to the levels of transcript *p3*. This transcript is more abundant than the erythroid transcript that we examined above in mouse cells. Reticulocyte samples were differentiated from two different human donors, and both showed high expression of this particular transcript. Like in the mouse, the *ZBTB7A p1* transcript is also highly expressed in these cells, as well as the majority of other cell types and represents a ubiquitous transcript.

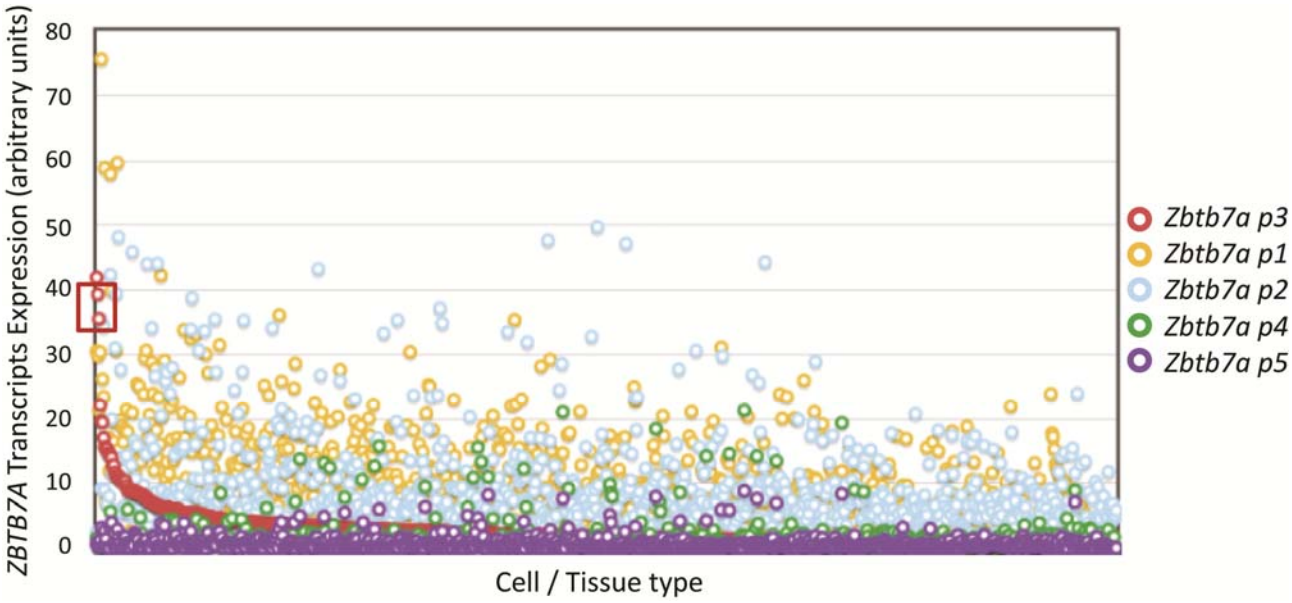


Figure 5.5. Human *ZBTB7A* transcript variants and their expression levels in various human tissues, uncovered from the FANTOM5 (Functional Annotation of the Mammalian Genome) SSTAR (Semantic catalogue of Samples, Transcription initiation And Regulators) Project. These variants have been found using Cap Analysis of Gene Expression (CAGE). Tissues have been arranged based on the values for variant *ZBTB7A p3*, which is most highly expressed in mast cells and reticulocytes. The y-axis here represents the relative expression levels, and along the x-axis are the various cell and tissue types. The red box indicates the erythroid tissues where *p3* transcript is the most highly expressed *Zbtb7a* transcript.

Table 5.3 lists the relative expression of each transcript in the top ten tissues of *ZBTB7A p3* expression. All of the top ten samples are cells of the haematopoietic system. As mentioned, the cell type which has the highest level of the *ZBTB7A p3* transcript is the mast

cell sample from one donor out of two. It is difficult to draw any conclusions from these various cell types, except to say that after reticulocytes, the expression level of this particular transcript drops off quite quickly and is overtaken by both *ZBTB7A p1* and *p2* in the CD14+ monocyte cells, which is the fourth entry on this list. By the tenth entry, the expression level of *ZBTB7A p3* is less than half of what it is in the reticulocytes, which may indicate that high levels of this transcript are largely restricted to erythroid-type cells. We hypothesised that *p3* is the erythroid transcript that we may identify experimentally by 5'RACE.

Table 5.3. Table of human cell and tissue types and the various ZBTB7A transcripts expressed in them, found from the FANTOM 5 (Functional Annotation of the Mammalian Genome) SSTAR (Semantic catalogue of Samples, Transcription initiation And Regulators) database. The tissues have been arranged according to expression levels of the *Zbtb7a p3* transcript, most highly expressed in mast cells, but also reticulocytes in humans. The numbers represent the expression levels of each isoform relative to a housekeeping gene.

Sample	*Zbtb7a p3*	*Zbtb7a p1*	*Zbtb7a p2*	*Zbtb7a p4*	*Zbtb7a p5*
Mast cell #1	41.82	30.55	2.55	1.58	0.62
Reticulocyte #2	39.22	29.80	1.57	0.26	1.05
Reticulocyte #1	35.52	21.27	0.00	0.22	0.22
CD14+ monocytes #1	22.10	30.39	35.48	1.02	1.16
Mast cell #4	19.57	75.62	9.09	1.14	3.16
CD14+ Monocytes #3	19.41	26.62	34.69	0.73	0.24
CD14+ Monocytes #2	17.00	23.43	34.39	1.89	0.38
Mast Cell #2	15.55	58.88	8.92	0.62	2.70
Acute myeloid leukaemia	15.51	7.05	2.82	0.00	1.41
Whole blood	15.29	16.82	6.88	0.00	0.00

The TSS for the *ZBTB7A p3* transcript identified in the FANTOM5 database lies 270 bp upstream of exon 1. This can be seen in Figure 5.6 A.

5.4.1 5' RACE design

We next set out to explore this transcript start point in more detail, and this was approached in the same way as for the mouse. In order to determine the presence of an erythroid-specific transcript in other human cells, we performed 5'RACE using un-differentiated and erythroid differentiated cells from CD34+ HSCs. We used CD34+ HSCs derived from both cord blood and bone marrow for this experiment to investigate any potential differences between foetal and adult tissues. These cells were induced as previously described in Chapter 2 (Figure 2.1). For the non-erythroid control samples, we used RNA extracted from the erythroid differentiation at week 1, when the cells are largely still CD34+. The RNA for the erythroid samples was extracted at week 3 of differentiation, when the cells

are predominantly erythroid and resemble reticulocytes, characterised by high levels of globins and erythroid-specific proteins such as KLF1. These cells were comparable to the reticulocyte cells investigated in the FANTOM5 analysis, although the exact method of differentiation may have varied. As in the mouse, the first exon of *ZBTB7A* is GC rich, so we again designed 6 reverse primers in exon 2 to maximise our chances of success. Each successive primer was again slightly further downstream of the TSS and resulted in the amplification of a larger product. Below is a diagram depicting the genomic DNA of the human *ZBTB7A* gene (Figure 5.6 A), as well as the RefSeq cDNA transcript from NCBI (NM_015898.2) (Figure 5.6 B). As with the mouse, *ZBTB7A* is composed of three exons, depicted here along with the translation start site, and 5' and 3' UTRs. This figure also depicts the TSS identified by FANTOM5 which lies approximately 270 bp upstream of the annotated TSS. The asterisk denotes the site of the erythroid-specific promoter [137], which in the case of the human lies approximate 270 bp upstream of the annotated TSS and adjacent to the TSS of the hypothesised erythroid-specific transcript discovered in FANTOM5 (Figure 5.6 A). The reverse primers for the 5'RACE are shown lying in exon 2 (Figure 5.6 B).

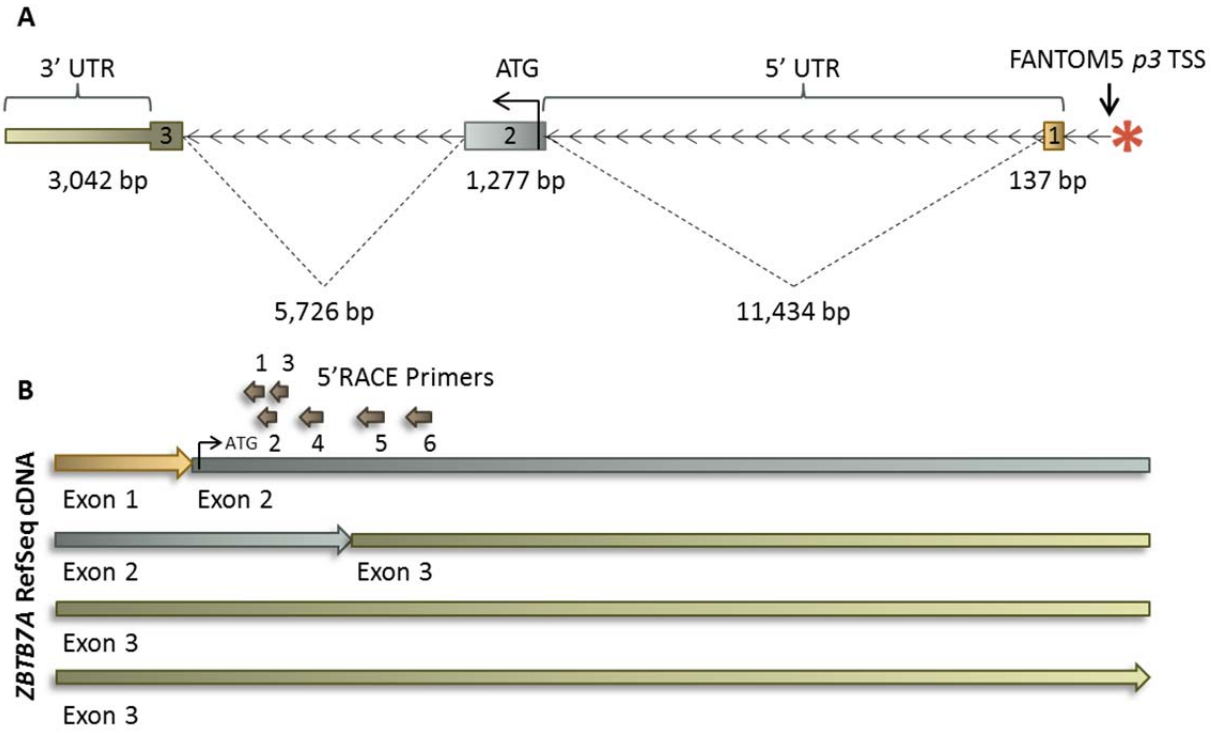
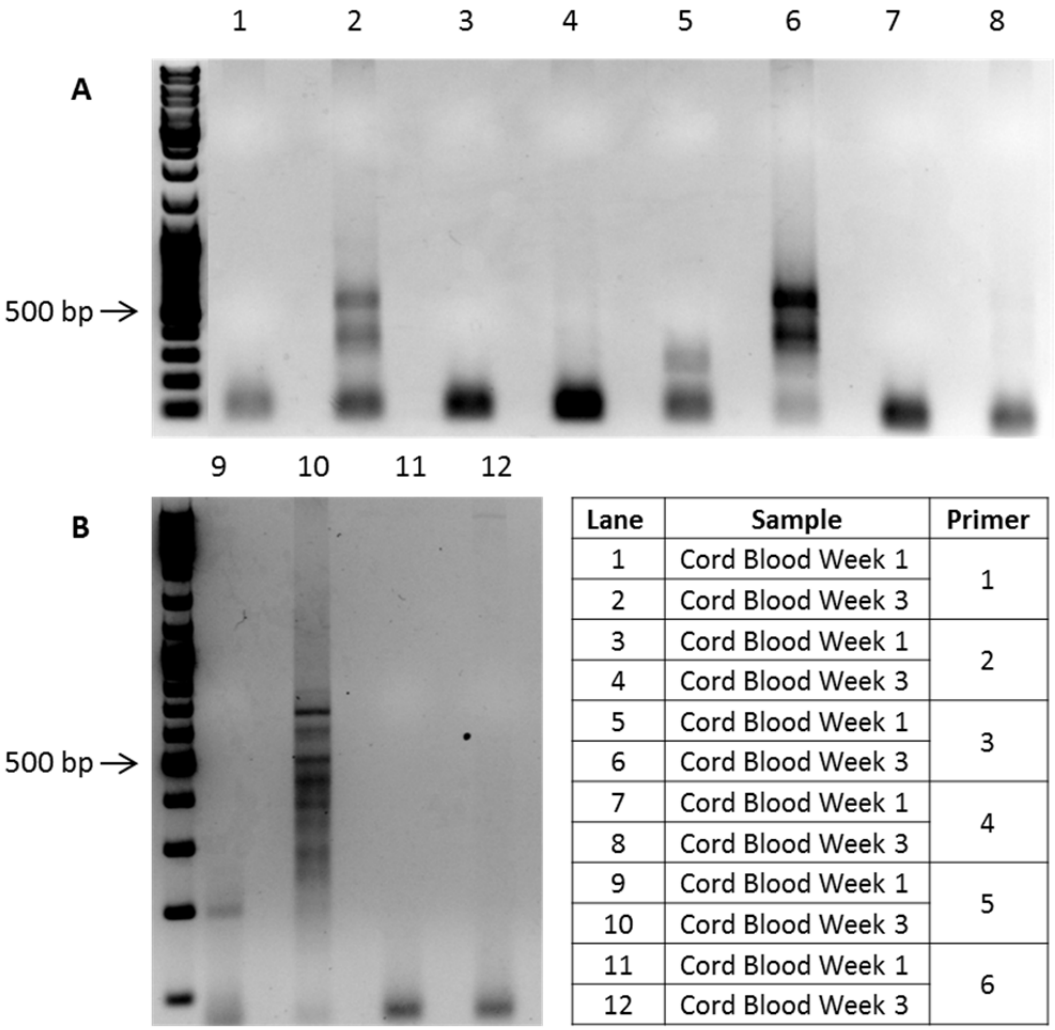


Figure 5.6. 5'RACE primer design for amplifying the 5' end of the human *ZBTB7A* transcript. 5'RACE was performed according to instructions in the kit. cDNA is produced from RNA with a 5' cap recognised by forward primers supplied by Clontech. Gene specific reverse primers were designed lying in Exon 2 to amplify potential transcript variants. **(A)** Genomic sequence of *ZBTB7A* showing exons 1-3 and intergenic regions, 5'- and 3' UTRs and the start point of translation, marked with ATG. Asterisk indicates the erythroid-specific promoter. **(B)** RefSeq cDNA sequence of *ZBTB7A* (NM_015898.2) showing exons and start point of translation, and each individual 5'RACE primer, labelled 1-6.

5.4.2 5'RACE results

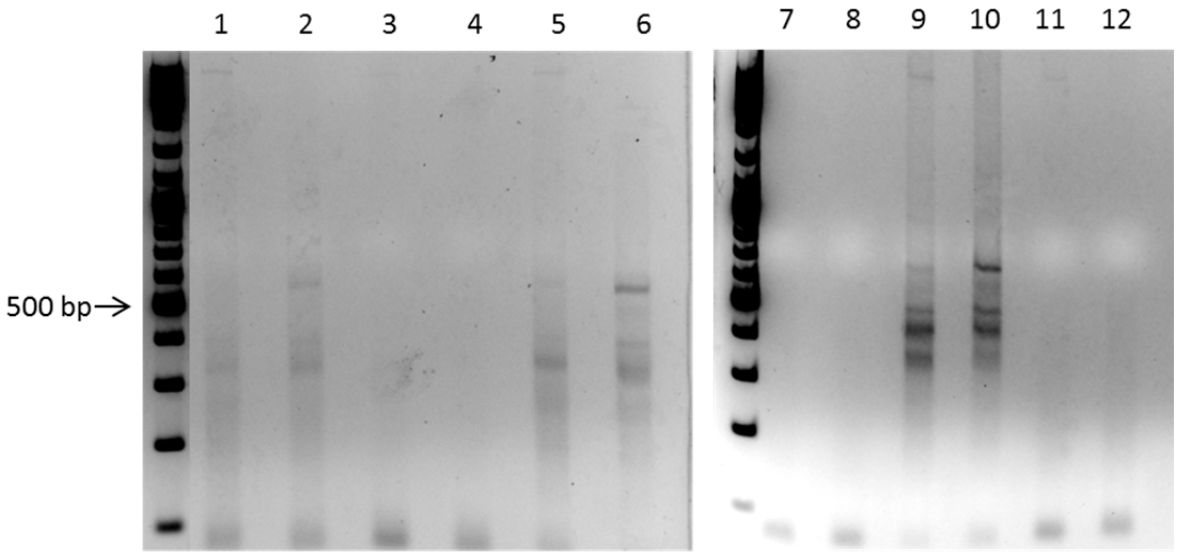
5'RACE experiments were conducted in human cells using undifferentiated and erythroid differentiated cord blood (Figure 5.7) and bone marrow (Figure 5.8) samples, using the generic forward primer and individual reverse primers targeted to exon 2 of *ZBTB7A*. From this we found that, for both cord blood and bone marrow, some of the primers were unsuccessful. Primer 2 (lanes 3 and 4), primer 4 (lanes 7 and 8) and primer 6 (lanes 11 and 12) showed no products for either week 1 or week 3 samples in cord blood and bone marrow. As was the case with some of the mouse assays, one primer (primer 1) showed no product in the undifferentiated week 1 samples for both cord blood and bone marrow, but produced two distinct bands in the differentiated week 3 samples which appeared to be the same as the bands seen for the other successful primers, 3 and 5. In the cord blood samples, primer 4 also seems to have produced two faint bands in the week 3 samples. In the top gel of the cord blood samples, showing lanes 1-8, two bands are visible in the week 3 samples (Figure 5.7 A). But when these bands were electrophoresed for longer, as was the case for the bottom gel

of the cord blood samples (Figure 5.7 B) and both gels for the bone marrow samples (Figure 5.8), it was found that these two bands resolved into several distinct bands. The week 0 samples for most primer pairs had much less product than the week 3 samples, regardless of band size. This may be due to PCR efficiency, or it may be indicative of the fact that there is less *ZBTB7A* transcript expressed in these cells, which would correlate with what was seen in Chapter 3 (Figure 3.5 and Figure 3.7).



Lane	Sample	Primer
1	Cord Blood Week 1	1
2	Cord Blood Week 3	
3	Cord Blood Week 1	2
4	Cord Blood Week 3	
5	Cord Blood Week 1	3
6	Cord Blood Week 3	
7	Cord Blood Week 1	4
8	Cord Blood Week 3	
9	Cord Blood Week 1	5
10	Cord Blood Week 3	
11	Cord Blood Week 1	6
12	Cord Blood Week 3	

Figure 5.7. CD34+ cord blood cells at week 1 and week 3 of erythroid differentiation express a different transcript of *ZBTB7A*. Whole mRNA was extracted from CD34+ cord blood cells at week 1 of differentiation, where they remain CD34+ and undifferentiated, and week 3 of differentiation, where they are erythroid cells, and 5'RACE ready cDNA was made using the 5'RACE kit from Clontech. PCR reactions were set up with 5'RACE ready cDNA and gene specific reverse primers which lie in exon 2 of *ZBTB7A*. Products were run of a 2% agarose gel. A 100 bp DNA ladder was used to estimate the sizes of the fragments.



Lane	Sample	Primer
1	Bone Marrow Week 1	1
2	Bone Marrow Week 3	
3	Bone Marrow Week 1	2
4	Bone Marrow Week 3	
5	Bone Marrow Week 1	3
6	Bone Marrow Week 3	
7	Bone Marrow Week 1	4
8	Bone Marrow Week 3	
9	Bone Marrow Week 1	5
10	Bone Marrow Week 3	
11	Bone Marrow Week 1	6
12	Bone Marrow Week 3	

Figure 5.8. CD34+ bone marrow cells at week 1 and week 3 of erythroid differentiation express a different transcript of *ZBTB7A*. Whole mRNA was extracted from CD34+ bone marrow cells at week 1 of differentiation, where they remain CD34+ and undifferentiated, and week 3 of differentiation, where they are erythroid cells, and 5'RACE ready cDNA was made using the 5'RACE kit from Clontech. PCR reactions were set up with 5'RACE ready cDNA and gene specific reverse primers which lie in exon 2 of *ZBTB7A*. Products were run of a 2% agarose gel. A 100 bp DNA ladder was used to estimate the sizes of the fragments.

5.4.3 5'RACE sequencing

Visible bands from the cord blood samples were sequenced. The sequences were then mapped to the *ZBTB7A* locus using the BLAT search tool on the UCSC Genome Browser [217, 218]. Some of the sequences were found not to be *ZBTB7A*, and these were discarded from the analysis. Most of the sequences from week 1 samples appeared to have start sites that were just upstream of exon 1, whereas week 3 samples started around 240 bp upstream. The sequence that mapped to the *ZBTB7A* locus are tabulated below in Table 5.4, which shows the sample type in column one, week 1 in grey and week 3 in red, and the number of

bases up- or downstream of the annotated RefSeq TSS where the sequence was found to start, is shown in the second column. The third column groups the samples together based on their TSSs.

Table 5.4. Collated sequencing data from CD34+ cord blood Week 1 and week 3 of erythroid differentiation 5'RACE samples. The first column indicates the sample type, the second is the start point of transcription found relative to the annotated TSS of the first exon, this is the number of base pairs up- or downstream of the defined TSS, and the third column groups the samples together by their TSS. Grey rows indicated Week 1 samples and red rows indicate week 3 samples.

Sample	TSS	TSS Grouping
Cord Blood Wk3	-276	TSS 1
Cord Blood Wk3	-241	
Cord Blood Wk3	-238	
Cord Blood Wk3	-238	
Cord Blood Wk3	-238	
Cord Blood Wk3	-233	
Cord Blood Wk3	-233	
Cord Blood Wk3	-229	
Cord Blood Wk3	-226	
Cord Blood Wk1	-81	TSS 2
Cord Blood Wk1	-28	
Cord Blood Wk1	-23	
Cord Blood Wk3	-22	
Cord Blood Wk1	23	
Cord Blood Wk1	Exon 2	
Cord Blood Wk1	Exon 3	

In the human samples, there were two groupings of start sites. In this case, distinct from our findings in the mouse, the first group was located at around 240 bp upstream of the TSS in exon 1 (-240 bp). All of these longer transcripts came from the erythroid week 3 samples, suggesting that it is an erythroid-specific start site. The second start site was further downstream. Some were still found upstream of the annotated TSS, and some internally in exon 1, and the majority of these, bar one exception, were found from the week 1 samples. Figure 5.9 depicts the various TSSs that were found in the sequencing analysis. The red arrows indicate the induced samples, and the grey arrows the un-induced. From this figure, the clustering of the induced samples at the -240 site is clear. The upstream region, which is transcribed in the erythroid cells, has been labelled Exon 0. The longest transcript found here at -276 bp, closely matches the TSS of *p3* discovered in the FANTOM5 project at 270 bp upstream (-270 bp) of exon 1. The bulk of the tags sequenced in the FANTOM5 project were found to lie around 240 bp upstream, which mirrors what we have found using 5'RACE. It is

interesting to note that the conserved region of the mouse and human promoter containing potential GATA-1 and KLF1 sites (indicated by an asterisk in Figure 5.6) as mentioned above, lies 270 bp upstream of the TSS of exon 1, which makes it directly adjacent to the TSS of this novel Exon 0 in humans. We therefore believe this to be to be a genuine erythroid-specific transcript.

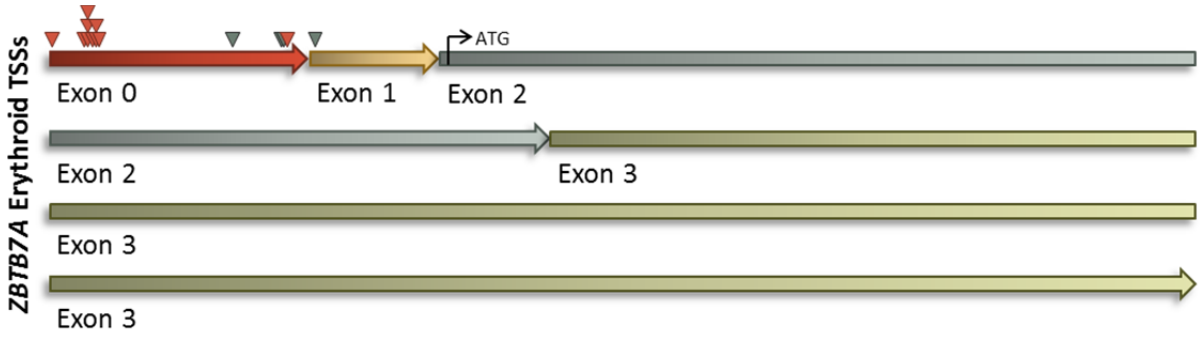


Figure 5.9. Transcriptional start sites of *ZBTB7A* transcripts in CD34+ cord blood cells at week 1 and week 3 of erythroid differentiation found by 5'RACE. Grey arrow heads indicate the week 1 samples and red arrow heads indicate the week 3 samples. Week 3 samples cells tend to cluster at around -238 bp upstream of the annotated TSS, which appears to be a novel erythroid *ZBTB7A* transcript. We have labelled this novel exon, exon 0.

5.4.4 Validation

As the 5'RACE results for the human samples indicated a novel TSS upstream of the first exon in erythroid cells, it was amenable to validation by qPCR. Even though this exon 0 is technically an extension of exon 1, and is not separated by an intron (Figure 5.10 A), we were still able to design primers that lay in exon 0 and exon 1 to test whether there was a greater up-regulation in this transcript in the erythroid differentiated CD34+ cord blood and bone marrow cells. Figure 5.10 shows the up-regulation of the entire *ZBTB7A* transcript, as the primers lie across the exon 2/3 boundary, in comparison to the new erythroid transcript, where the primers lie in exon 0 (Figure 5.10 A). This figure depicts the difference in up-regulation between the two transcripts when haematopoietic stem cells (week 1) are differentiated down the erythroid lineage (week 3). From this we can see that in both the cord blood (Figure 5.10 B) and the bone marrow (Figure 5.10 C) there is a markedly greater up-regulation of this erythroid-specific transcript, which was previously masked in the qPCR which did not cover this region (Figure 3.5-3.7). This is due to the fact that other transcript variants of ZBTB7A are expressed in haematopoietic stem cells at week 1, resulting in a higher expression level at base-line. However no detectable levels of the novel transcript from exon 0 is detectable in these cells, which results in a greater difference in this transcript, compared to the total levels of transcript during the differentiation process. Interestingly, the

bone marrow sample appears to have a greater up-regulation of this transcript than the cord blood.

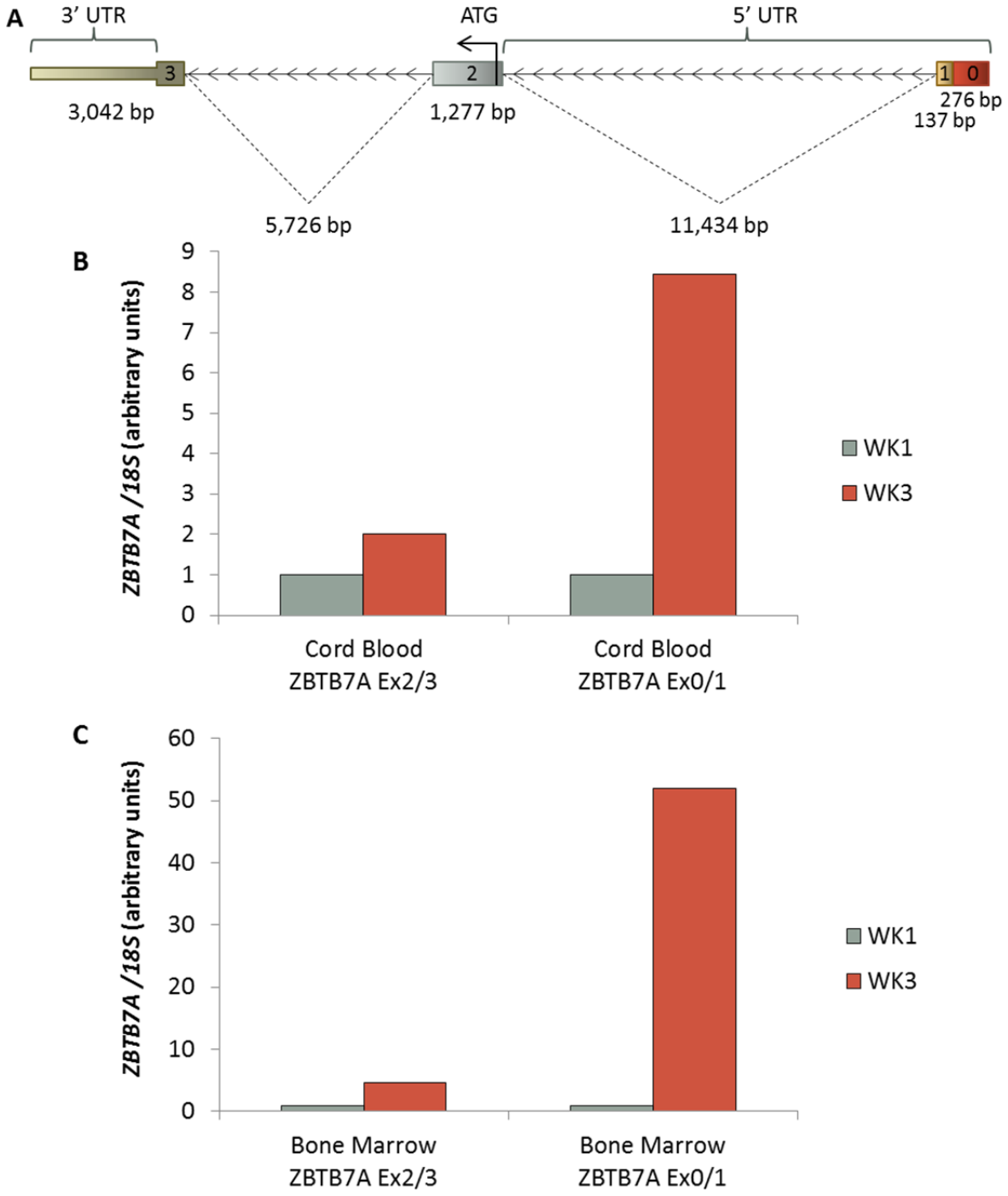


Figure 5.10. Validation of the erythroid-specific transcript of human *ZBTB7A* in cord blood and bone marrow erythroid differentiated CD34+ cells. (A) Genomic DNA of ZBTB7A indicating location of the putative exon 0. mRNA was extracted from cord blood **(B)** and bone marrow **(C)** CD34+ cells at week 1 of erythroid differentiation (indicated by grey bars), where the cells remain undifferentiated and CD34+, and week 3 of differentiation (indicated by red bars), where the cells are erythroid. Differential expression of *ZBTB7A* transcripts were examined with qPCR. The transcript emanating from the previously examined exon 2 and 3 boundary is shown first, followed by the extension of exon 1 labelled exon 0, discovered by the 5'RACE. Representative plots are shown for each.

5.5 Chapter 5 Discussion

In this chapter, we investigated the possibility of an erythroid-specific transcript of *ZBTB7A* in both human and mouse cells. There are a number of conserved CACCC- and GATA-motifs in the promoter of this gene which may be bound by the erythroid factors KLF1 and GATA-1, suggesting the possibility of an erythroid-specific start site. GATA-1 sites are often found slightly downstream of binding sites for SCL/TAL1 in erythroid cells [181]. These sites usually appear as an E-box/GATA composite motif with the sequence -CTG(N_9)GATA- [219], however no E-boxes, or these composite sites were found near the GATA sites in the conserved *ZBTB7A* promoter region. We hypothesised that we would see a difference in the TSS of this gene in erythroid cells, which would give rise to a novel transcript, but the same protein as the ubiquitously expressed gene. We initially hypothesised that there would be transcription of a novel 5' exon which would be alternatively spliced to exon 2 in erythroid cells, as is the case with many erythroid-specific transcripts, such as the one in *Klf3*. We began by examining data from the FANTOM5 project, which showed transcription start points enriched in erythroid cells of human and mouse. We also performed 5'RACE to fully ascertain the structure of 5'UTR of the erythroid transcripts, and then validated the novel transcript by qPCR in human samples.

The sequencing results of the 5'RACE suggested the situation is more complex than anticipated. From both the mouse and human analysis, we saw no evidence of alternative splicing of a novel erythroid-specific exon. Instead, we found a new start point of transcription, which, for the mouse, was internal in the first exon and slightly downstream of the annotated TSS, and for the human was upstream of the annotated TSS, but as an extension of the first exon rather than from a distinct upstream exon. For the mouse, this internal TSS did not match that found in the FANTOM5 database, which was found to be 48 bp upstream of the first exon. There are a number of reasons for why this might be the case, the primary one being that these are different cell types. While the foetal liver is predominantly erythroid at the time points investigated by FANTOM5, it still represents a mixed population of cells, and it is possible that the TSS identified is not erythroid-specific. However, the transcript we uncovered here by 5'RACE is clearly induced in K1ER cells, which indicates that it may be under the regulation of KLF1. As KLF1 is a master regulator of erythropoiesis, it is possible that this transcript is erythroid-specific; however experiments in non-erythroid cells types would be needed to confirm this. The transcript we identified by 5'RACE was difficult to validate by qPCR. Only a short upstream region was available to be targeted to test differential expression between the two transcripts, which was too GC rich to

achieve successful amplification by qPCR. This means that additional 5'RACE experiments would need to be utilised to investigate whether this particular TSS is found in other cell types, and to confirm that it is erythroid-specific.

The human erythroid-specific transcript we identified by 5'RACE did match the TSS discovered in the FANTOM5 database which we hypothesised was erythroid-specific. The validation of this transcript proved to be more straightforward than the mouse. As the new upstream TSS was sufficiently far enough away from the annotated TSS, there was a margin to successfully design probes targeted to this region. The major issue we faced with this validation was the lack of an intron separating the two exons, as any contaminating genomic DNA (gDNA) would influence results. During our standard RNA extraction protocol we include a DNase step prior to cDNA synthesis, which allows genomic DNA to be largely eliminated. However, for these primary cells, given the small number obtained from the differentiation, it was not possible to use our standard DNase method. Instead, cell lysates were passed through a column designed to eliminate gDNA, in the RNeasy Plus Mini kit (Qiagen, Clifton Hill, VIC, Australia). We therefore used ChIP-qPCR primers, which were specific for other promoters and intergenic regions, to determine the level of gDNA contamination. Amplification was detected at these ChIP primer sets in some samples, making assessment of the precise magnitude of increase in transcript levels difficult to define. Samples with the least amount of contaminating gDNA were selected, which is the reason only representative plots for each sample type were shown. These samples showed no detectable levels of amplification from the ChIP primers used.

It should be noted that, unlike the start point of translation, transcriptional start sites are not defined by a particular sequence, and there is some inherent fluctuation in where a transcript will begin. This is reflected in the sequences obtained from both mouse and human 5'RACE experiments, but is particularly apparent in the mouse, where the TSSs for the uninduced cells were found to range from +95 to +324 of the annotated *Zbtb7a* TSS. It is also interesting to note that no transcripts for either human or mouse were found that showed a TSS that exactly matched the annotated TSS.

If the protein arising from differentially expressed transcripts is the same, and the first exon of each of these transcripts is almost identical, then what is the reason for this erythroid-specific regulation? It could be argued that it is because *Zbtb7a* is critical for the maturation and survival of erythroid cells and therefore requires increased expression in these cells above basal levels. From knockout experiments, it is clear that *Zbtb7a* is important in an erythroid context [137, 138, 164]. It has been suggested to be activated by the master

erythroid regulator, GATA-1 [137], and we have shown in this thesis that it is regulated by another important erythroid regulator KLF1. Hence, GATA-1 and KLF1 may act together to bolster the expression of *ZBTB7A* to levels that are required for the correct maturation and survival of erythroid cells. Or perhaps, the reason for the erythroid-specific up-regulation may be, as we have suggested in Chapter 3, to correctly silence the expression of the embryonic *βh1-globin* in the mice, and foetal *γ-globin* in humans.

Ultimately, future research will focus on determining the functional relevance of this transcript *in vivo* in cell lines, as well as in a mouse model and primary human cells. There are various ways in which this can be achieved, but the recent technological advance of the CRISPR/Cas9 system which enables us to specifically knockout regions of the genome with ease, and with little to no off-target effects, in both cells and mouse models, will be employed here in future experiments. We plan to target the conserved GATA- and CACCC-motifs identified in the promoter regions of human and mouse *ZBTB7A* thought to be bound by the erythroid master transcription factors GATA-1 and KLF1 to determine if this impairs production of the erythroid specific start point, while leaving the ubiquitous start point intact.

CHAPTER 6 – GENERAL DISCUSSION & CONCLUSION

6.1 Summary

Diseases of the adult *β-globin* genes are some of the most prevalent monogenetic diseases worldwide. However, naturally occurring mutations, which lead to increased production of the foetal globins into adulthood - known as Hereditary Persistence of Foetal Haemoglobin (HPFH) - have been shown to result in less severe symptoms, and in many cases, to completely ameliorate these diseases. This has triggered intensive research into haemoglobin switching, as reactivating the foetal globin genes in patients with *β*-globinopathies is an attractive therapeutic target. A number of HPFH families present with mutations in the -200 region of the foetal *γ-globin* promoter, which have been shown to lead to an increase in foetal haemoglobin (HbF) ranging from 2.5% to 40% for particular mutations (Table 1.2). There are two mechanisms by which this can conceivably occur. Firstly, these mutations may create novel binding sites for activating transcription factors, and thereby activating *γ-globin*. Or they may result in the loss of repressor complex binding sites, resulting in the loss of *γ-globin* silencing. As there are a number of mutations in this region which result in the same phenotype, the latter scenario is more likely, however finding evidence of the responsible factors that bind to this region has proven elusive.

Since its discovery in 2008, *BCL11A* has been touted as the central factor in globin switching. Disruption of *BCL11A* expression has since been shown to be a good therapeutic target for treating adult *β*-globinopathies as it has been found to be required for *γ-globin* silencing, and yet dispensable for other normal erythrocyte function and production. In a proof of principle study, inactivation of *BCL11A* was used to correct the defects associated with SCD in a mouse model [111]. However, the mechanism of action of BCL11A has been difficult to elucidate, as there is currently no evidence that it binds directly to the promoter of the *γ-globin* genes. It has, however, been shown in various instances to be involved in promoting long-range looping interactions to the *β-globin* locus to silence the expression of the foetal globin genes [120, 220, 221].

BCL11A, however, is not the only important factor in foetal globin silencing, as the reactivation of *γ-globin* expression seen in the absence of BCL11A is not complete [110], and can be enhanced by the loss of other factors, for instance SOX6 [120], KLF1 [115], LSD1 and DNMT1 [222]. While many studies since the discovery of *BCL11A*'s role in the field of globin switching have been focused on determining its mechanism of action in silencing

γ-globin expression, uncovering novel repressors of *γ-globin* will add further understanding to this complex process.

Here we have provided evidence that ZBTB7A, a transcriptional repressor, plays an important role in erythropoiesis and binds to the -200 site in the *γ-globin* promoter. Our collaborators at Harvard University, Stuart Orkin and Takahiro Maeda, are currently investigating the effect of knocking out ZBTB7A on *γ-globin* expression. These experiments are akin to those which established the role of BCL11A as an important repressor of foetal globin expression [105, 110]. The predominant focus of this thesis has therefore been the regulation of *ZBTB7A* in humans and mice, as an essential complement to the work being conducted by these collaborators. It has previously been shown that *Zbtb7a* is activated by GATA-1 [137], and it has been hypothesised that *Zbtb7a* is activated by KLF1 on the basis of various microarray analyses [87], but regulation of *Zbtb7a* by KLF1 has never been explicitly examined in detail. Here we have provided evidence that KLF1 binds to the promoter of *Zbtb7a* and directly activates its expression in an erythroid context. We have also provided evidence of a potential erythroid-specific transcript, which appears to be under the regulation of KLF1, at least in the mouse cell line K1ER.

Importantly, mutations in KLF1 are associated with elevated levels of foetal globin [113]. KLF1 has been shown to directly activate *BCL11A* [112], which possesses an erythroid-specific enhancer [109]. Thus, individuals with reduced KLF1 activity have reduced levels of the repressor BCL11A, and thereby an increase in foetal globin. This has led to the idea of specifically knocking out BCL11A's erythroid enhancer [223]. It is hypothesized that this would reduce BCL11A in erythroid cells and increase foetal globin levels, but not impair levels of *BCL11A* in other cell types, where it is a vital gene.

We have approached ZBTB7A in a similar manner. We have also provided evidence that it is a foetal globin repressor that is activated by KLF1. We have investigated its regulation and discovered that it too has an erythroid-specific control element. *ZBTB7A* is an essential gene in many tissues, including erythroid cells, but again it may be possible to reduce its expression in erythroid cells without affecting expression elsewhere by the targeted removal of the erythroid control element. This reduction may serve as a way to boost foetal globin levels to treat haemoglobinopathies.

6.2 ZBTB7A and the field of globin switching

Globin switching is a complex process, involving many co-regulators and transcription factors shown to bind directly to the proximal promoters of the *β*-globin genes, or to mediate

looping interactions to various sites of the *β*-globin locus. Below is a simplified model for the network of key transcription factors involved in globin switching illustrating the insights into ZBTB7A discovered in through this body of work (Figure 6.1). The master erythroid transcription factor, GATA-1 activates *KLF1*, another important factor in erythropoiesis. One of the roles of KLF1 is to activate the adult *β-globin* gene at the time of haemoglobin switching [89, 90]. However, its regulation of this process is two-fold, as it is also involved in activating genes involved in the silencing foetal *γ-globin* expression, such as *BCL11A* [112], and as proposed previously [87] and validated here, *ZBTB7A*. GATA-1 has also been shown to activate *BCL11A* via an erythroid-specific enhancer located in the second intron of *BCL11A* [109]. BCL11A interacts with the *β*-LCR, and works together with SOX6 and GATA-1 to reconfigure the *β-globin* locus and silence the expression of *γ-globin* [120]. SOX6 has also been found to be involved in the direct silencing of the murine embryonic *εy-globin* [119], which is an orthologue of the human embryonic *ε-globin* gene [165], and may be involved in silencing this gene in human development (illustrated as a dotted line in Figure 6.1). We have provided evidence of ZBTB7A binding the -200 region of the promoters of both *γ-globin* genes, and hypothesise, from preliminary results from our collaborators, that it acts to silence the expression of these genes. We have also found evidence that ZBTB7A binds to the *β*-LCR, and may, like BCL11A, also play a role in facilitating long-range looping interactions to various regions of the *β*-globin locus. Our ChIP-Seq experiments have also shown that ZBTB7A binds to the promoter regions of *BCL11A, KLF1* and *GATA-1*, as well as its own promoter. Whether it is involved in the activation or silencing of these genes is unknown at this stage. Many transcription factors can both activate and repress their targets under different conditions and in conjunction with different regulators and co-factors. While ZBTB7A has predominantly been found to repress genes, it has also been found to be capable of activation [180]. In this diagram we illustrate the potential regulation of these genes as dotted lines showing both activation and repression at its targets.

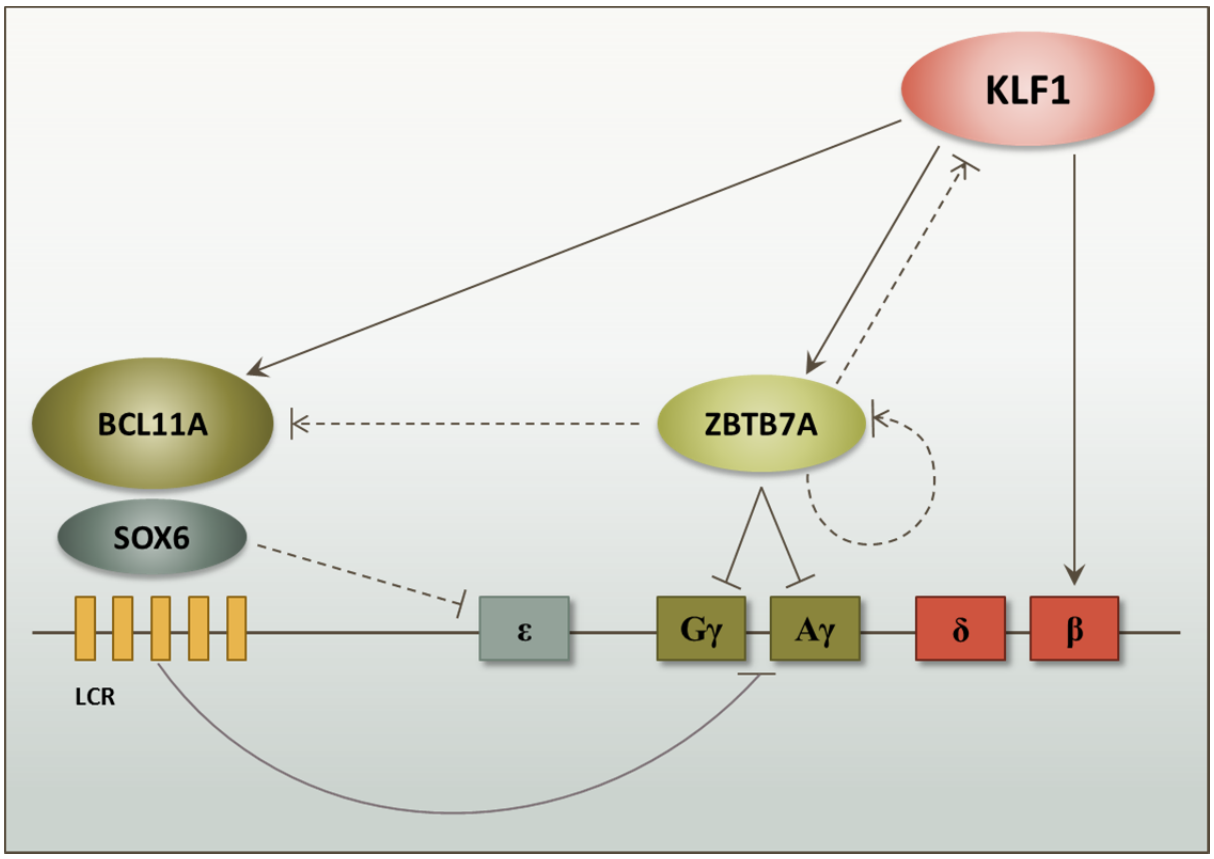


Figure 6.1. Model of haemoglobin switching at the *β*-globin locus. GATA-1 activates the erythroid transcription factor *KLF1*. KLF1, in turn, activates adult *β-globin* at the time of globin switching. KLF1 also drives the expression of transcription factors which are involved in the silencing of the *γ-globin* genes, *BCL11A* and *ZBTB7A*. BCL11A and SOX6 bind together at the locus control region (LCR) and facilitate looping interactions to various sites in the *β*-globin locus to silence the expression of foetal globin. SOX6 may also play a role in silencing the embryonic *ε-globin* earlier during development (dotted line). ZBTB7A binds to the -200 region of the *γ-globin* promoters and helps to silence their expression. This protein may also have roles in facilitating looping at the LCR. ZBTB7A also may act to either activate or silence its own expression and that of *KLF1* and *BCL11A* (dotted lines).

6.3 Future Directions

The findings presented in this thesis can be divided to support two main hypotheses. Firstly, that ZBTB7A is involved in the regulation of globin switching, and silencing of foetal globin is achieved through direct binding of ZBTB7A to the -200 region of the foetal globin promoter. This is evidenced predominantly by the demonstration that ZBTB7A is capable of binding the -200 region of the foetal globin promoter *in vitro* and *in vivo*, and that this binding is disrupted by mutations that result in HPFH [140]. Secondly, that *ZBTB7A* expression is specifically regulated by KLF1 and GATA-1 via an erythroid specific control element in the promoter of *ZBTB7A* in erythroid cells. This is evidenced by the direct regulation of *ZBTB7A* by KLF1, and the discovery of a potentially erythroid specific transcript variant. However, additional data is required to add further weight to these

hypotheses. The future directions for this project are, thus, also two-fold, firstly to determine whether ZBTB7A can indeed regulate the foetal globin gene through its binding to the -200 region, and secondly, to determine the function of the erythroid control element. Outlined below are a number of experiments which could potentially serve to provide greater evidence for these hypotheses.

The initial aim is to fully elucidate the mechanism by which ZBTB7A regulates foetal globin through the -200 region of the promoter. As discussed in Chapter 3, being able to capture a repressor binding to its target gene by ChIP is particularly challenging. Somewhat counterintuitively, ZBTB7A was found to bind to the -200 region in K562 cells, where foetal globin is highly expressed, and not adult bone marrow cells, where the gene is typically considered silenced. This is possibly due to the fact that repressors affect conformational changes in the chromatin of some target genes, and once they have inflicted these changes, they are no longer required to bind to these sites. In order to demonstrate that ZBTB7A regulates foetal globin through this promoter site, an elegant series of tests need to be undertaken in a suitable cell line, or at the appropriate time in primary cells. Chapter 3 provides evidence that ZBTB7A is capable of binding the -200 region of the foetal globin promoter in CD34+ cells derived from cord blood, which are harvested at the time that switching is actively occurring in humans, where some cells express foetal globin, and some adult globin. While technically challenging to work with, these cells present a perfect system in which to explore the role of ZBTB7A in globin switching, however, considering we have also observed binding of ZBTB7A to the -200 region in K562 cells, this cell line may present a more amenable alternative. Whichever system is chosen, the experiments remain the same. Using genome engineering techniques, the -200 mutations could be introduced into either of these cells. These mutations should result in the up-regulation of foetal globin in these cells, as they do in the individuals that carry them. Once these modified cells are established, ChIP experiments to determine ZBTB7A genomic occupancy can be carried out in parallel with WT cells to test the hypothesis that there would be abolished binding of this protein to the -200 site of the foetal globin promoter. A recent publication from this lab demonstrated that mutations resulting in HPFP can be introduced in cell lines using a TALEN mediated approach in order to up-regulate foetal globin. In this proof of principle study it verified that another HPFH mutation, single base pair T to C substitution at the -175 site of the foetal globin promoter, results in the creation of a novel binding site for the transcriptional activator TAL1 [224]. The same approach could be used for the -200 mutations, however as there are multiple mutations in this region, it may be prudent to begin with the mutation that shows the greatest increase in foetal globin in patients, namely the -196 C to T mutation in the *Aγ*

promoter [69, 70]. In order to demonstrate that it is indeed the loss of ZBTB7A binding to this site that results in the up-regulation of foetal globin, a "rescue" type approach using a modified ZBTB7A protein could be employed. In this case, the ZBTB7A protein could be modified to recognise the mutant -200 sequence and thereby once again bind this region and silence foetal globin.

The second aim arising from this thesis is to fully elucidate the role of the erythroid control element of *ZBTB7A*. As demonstrated in Chapter 5, the promoter of *ZBTB7A* contains a number of CACCC- and GATA-elements, some of which are conserved in the mouse. These are binding sites of the important erythroid transcription factors KLF1 and GATA-1, both of which have been shown to play important roles in regulating *ZBTB7A*. We have also provided evidence that a novel transcript variant is upregulated in erythroid cells. It is hypothesized that the function of this regulation element and novel transcript is to upregulate ZBTB7A in an erythroid specific manner to fully silence foetal globin in these cells. The first step in investigating this hypothesis will require confirming that this transcript is indeed erythroid specific. In order to achieve this, CD34+ haematopoietic stem cells can be differentiated down various haematopoietic lineages, and the expression of the novel *ZBTB7A* transcript variant can be compared to the expression of all *ZBTB7A* transcripts by qPCR in the manner that was utilized in Chapter 5. The analysis performed by the FANTOM5 consortium already suggests that this particular transcript will be highly expressed in both mast cells and the erythroid lineage. As GATA-1 also plays a role in megakaryocytes, it may also be plausible that this transcript is also upregulated in megakaryocytes, however, as the *Zbtb7a* KO mice have no megakaryocyte defects, it is unlikely that a potential knockdown in these cells will have adverse effects. While it is possible that this transcript may be expressed in other lineages, it is unlikely that it is under the regulation of both the erythroid transcription factors KLF1 and GATA-1 in these cells. Once the expression profile of this transcript has been investigated, it will be necessary to examine the role this particular transcript plays in erythroid cells. It is possible that the erythroid control element, regulated by KLF1 and GATA-1, can be used to fine-tune the expression of *ZBTB7A* in erythroid cells, without perturbing the levels in other cell types and at the same time leaving enough ZBTB7A in erythroid cells for their normal function. In order to investigate this in more detail, the binding sites of KLF1 and GATA-1 need to be strategically removed. Below is a diagram illustrating the various CACCC- and GATA-elements in the promoter of *ZBTB7A*. This diagram also illustrates proposed mutations to these sites which could be introduced to abolish the binding of both KLF1 and GATA-1.

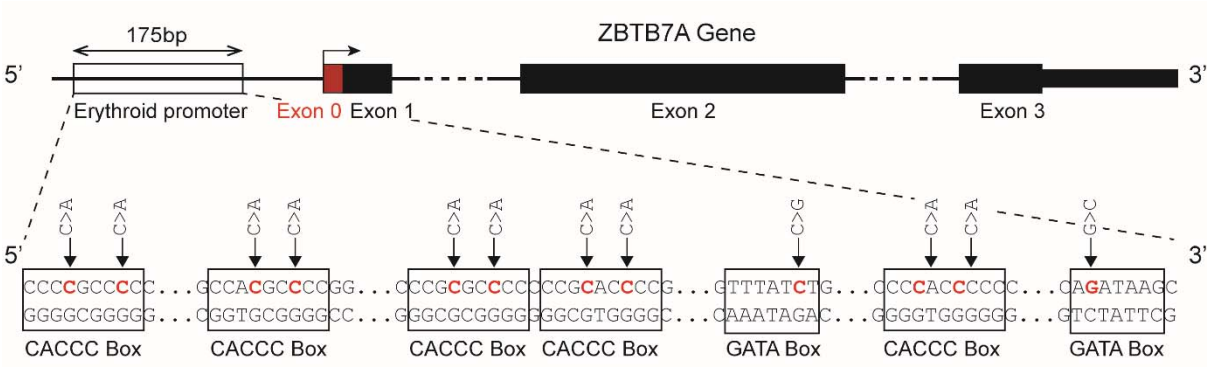


Figure 6. 2 The erythroid promoter of ZBTB7A and proposed mutations to binding sites for erythroid transcription factors. The ZBTB7A promoter contains a number of CACCC- and GATA-elements, binding sites for the erythroid transcription factors KLF1 and GATA-1, respectively. It is hypothesised that these factors upregulate a novel ZBTB7A transcript in erythroid cells. The proposed mutations would inhibit the binding of these factors and thereby silence the expression of this novel transcript variants in erythroid cells exclusively.

Determining the appropriate cell type in which to conduct these experiments is important. These cells must be shown to express the novel ZBTB7A transcript variant, as well as have the capacity to upregulate foetal globin. In this regard, the K562 cell line is a potential candidate. These cells have been found to express the transcript variant, and are able to be induced down the erythroid lineage. It has also recently been shown that, while these cells express high levels of foetal globin at base line, they do indeed have the capacity to upregulate this gene, as shown when the -175 HPFH mutation was introduced [224]. Once clonal cells containing these mutations have been established, the expression level of ZBTB7A will be examined by qPCR. It is expected that no expression of the erythroid specific transcript of ZBTB7A will be detected in these cells, which will result in an overall down-regulation of ZBTB7A at both the transcript and protein level. Following this it will be necessary to investigate whether there are any effects on cell survival and erythroid differentiation. Once it has been confirmed that there are no deleterious effects of this down-regulation, the expression of the globin genes can be investigated. It is hypothesised that an increase in foetal globin expression will be observed, which is likely to be accompanied by a comparable decrease in the levels of the adult globin [224]. This particular approach certainly has its pitfalls. Firstly, the exact location of the ubiquitous promoter of ZBTB7A is unknown, and may well lie in and around the erythroid promoter region. If the ubiquitous promoter is perturbed, then there is the potential to result in a knockout effect of ZBTB7A. As this gene is important in erythroid biology [137] and has been shown to play a role in various cancers [171-173, 176, 177], completely silencing its expression has foreseeable negative effects. Secondly, it is unknown if a down-regulation of this protein will

be sufficient to result in a large up-regulation of foetal globin. However, only a small up-regulation of foetal globin in patients with HPFH has been shown to be effective in alleviating the symptoms of haemoglobinapothies.

6.4 Conclusion

Here we have presented evidence that the transcriptional repressor ZBTB7A is capable of binding to the -200 site of the foetal *γ-globin* promoter *in vivo*. While it is not always found to occupy this region in all cell types, we believe that ZBTB7A acts via this site to repress foetal globin. We have previously provided evidence that mutations in the -200 region that cause HPFH, abolish binding of ZBTB7A, and our collaborators are currently investigating the repressive effects of ZBTB7A on *γ-globin.* We hypothesise that determining the occupancy of the repressor to its targets can be particularly challenging, and capturing the right time of silencing is important. We therefore plan to repeat the ChIP-Seq experiments in foetal-like cells which are undergoing haemoglobin switching. The predominant focus of this thesis has been investigating the regulation of ZBTB7A.

As ZBTB7A is an important transcription factor for normal erythroid development and survival [137], and considering its role in various cancers [171-173, 176, 177], complete silencing of this protein is likely to have multiple ill-effects. However, by targeting the erythroid-specific promoter through the use of genome editing, we may be able to reduce the expression in erythroid cells specifically, without completely silencing it. This reduction may indeed be sufficient to increase the levels of foetal globin by enough of a margin to reduce the severity of *β*-globin diseases.

References

1. **Levine M, Tjian R.** 2003. Transcription regulation and animal diversity. Nature **424:**147-151.
2. **Griffiths AJFM, J.H.; Suzuki, D.T.; Lewontin, R.C.; and Gelbart, W.M.** 2000. An Introduction to Genetic Analysis, 7th ed. New York : W.H. Freeman, c2000., United States.
3. **Patterson SD, Aebersold RH.** 2003. Proteomics: the first decade and beyond. Nat Genet **33:**311-323.
4. **Van Der Kelen K, Beyaert R, Inze D, De Veylder L.** 2009. Translational control of eukaryotic gene expression. Crit Rev Biochem Mol Biol **44:**143-168.
5. **Kutach AK, Kadonaga JT.** 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol **20:**4754-4764.
6. **Philipsen S, Talbot D, Fraser P, Grosveld F.** 1990. The β-globin dominant control region: hypersensitive site 2. EMBO J **9:**2159-2167.
7. **Ayoubi TA, Van De Ven WJ.** 1996. Regulation of gene expression by alternative promoters. FASEB J **10:**453-460.
8. **Chretien S, Dubart A, Beaupain D, Raich N, Grandchamp B, Rosa J, Goossens M, Romeo PH.** 1988. Alternative transcription and splicing of the human porphobilinogen deaminase gene result either in tissue-specific or in housekeeping expression. Proc Natl Acad Sci U S A **85:**6-10.
9. **Blackwood EM, Kadonaga JT.** 1998. Going the distance: a current view of enhancer action. Science **281:**60-63.
10. **Higgs DR, Wood WG.** 2008. Long-range regulation of α-globin gene expression during erythropoiesis. Curr Opin Hematol **15:**176-183.
11. **Vernimmen D, Marques-Kranc F, Sharpe JA, Sloane-Stanley JA, Wood WG, Wallace HA, Smith AJ, Higgs DR.** 2009. Chromosome looping at the human α-globin locus is mediated via the major upstream regulatory element (HS -40). Blood **114:**4253-4260.
12. **Guttman M, Rinn JL.** 2012. Modular regulatory principles of large non-coding RNAs. Nature **482:**339-346.
13. **Garrell J, Campuzano S.** 1991. The helix-loop-helix domain: a common motif for bristles, muscles and sex. Bioessays **13:**493-498.
14. **Roeder RG.** 2005. Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett **579:**909-915.

15. **Mitchell PJ, Tjian R.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245:**371-378.
16. **Hanna-Rose W, Hansen U.** 1996. Active repression mechanisms of eukaryotic transcription repressors. Trends Genet **12:**229-234.
17. **Su W, Jackson S, Tjian R, Echols H.** 1991. DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes Dev **5:**820-826.
18. **Su W, Middleton T, Sugden B, Echols H.** 1991. DNA looping between the origin of replication of Epstein-Barr virus and its enhancer site: stabilization of an origin complex with Epstein-Barr nuclear antigen 1. Proc Natl Acad Sci U S A **88:**10870-10874.
19. **Courey AJ, Jia S.** 2001. Transcriptional repression: the long and the short of it. Genes Dev **15:**2786-2796.
20. **Struhl K.** 1998. Histone acetylation and transcriptional regulatory mechanisms. Genes Dev **12:**599-606.
21. **Li B, Carey M, Workman JL.** 2007. The role of chromatin during transcription. Cell **128:**707-719.
22. **Pina C, Enver T.** 2007. Differential contributions of haematopoietic stem cells to foetal and adult haematopoiesis: insights from functional analysis of transcriptional regulators. Oncogene **26:**6750-6765.
23. **Robb L.** 2007. Cytokine receptors and hematopoietic differentiation. Oncogene **26:**6715-6723.
24. **Funnell A.** 2008. Identification of a regulatory network within the Krüppel-like factor (Klf) family PhD Thesis. University of Sydney, Sydney.
25. **Skalnik DG.** 2002. Transcriptional mechanisms regulating myeloid-specific genes. Gene **284:**1-21.
26. **Pallister CJ, Hancock JT.** 1995. Phagocytic NADPH oxidase and its role in chronic Granulomatous disease. British Journal of Biomedical Science **52:**149-156.
27. **Goldfarb AN.** 2009. Megakaryocytic programming by a transcriptional regulatory loop: A circle connecting RUNX1, GATA-1, and P-TEFb. J Cell Biochem **107:**377-382.
28. **Suzuki N, Suwabe N, Ohneda O, Obara N, Imagawa S, Pan X, Motohashi H, Yamamoto M.** 2003. Identification and characterization of 2 types of erythroid progenitors that express GATA-1 at distinct levels. Blood **102:**3575-3583.
29. **Ferreira R, Ohneda K, Yamamoto M, Philipsen S.** 2005. GATA-1 function, a paradigm for transcription factors in hematopoiesis. Mol Cell Biol **25:**1215-1227.

30. **Wong PM, Chung SW, Chui DH, Eaves CJ.** 1986. Properties of the earliest clonogenic hemopoietic precursors to appear in the developing murine yolk sac. Proc Natl Acad Sci U S A **83:**3851-3854.

31. **Keller G, Kennedy M, Papayannopoulou T, Wiles MV.** 1993. Hematopoietic commitment during embryonic stem cell differentiation in culture. Mol Cell Biol **13:**473-486.

32. **Haar JL, Ackerman GA.** 1971. A phase and electron microscopic study of vasculogenesis and erythropoiesis in the yolk sac of the mouse. Anat Rec **170:**199-223.

33. **Tavian M, Peault B.** 2005. Embryonic development of the human hematopoietic system. Int J Dev Biol **49:**243-250.

34. **McGrath KE, Koniski AD, Malik J, Palis J.** 2003. Circulation is established in a stepwise pattern in the mammalian embryo. Blood **101:**1669-1676.

35. **Ji RP, Phoon CK, Aristizabal O, McGrath KE, Palis J, Turnbull DH.** 2003. Onset of cardiac function during early mouse embryogenesis coincides with entry of primitive erythroblasts into the embryo proper. Circ Res **92:**133-135.

36. **De la Chapelle A, Fantoni A, Marks PA.** 1969. Differentiation of mammalian somatic cells: DNA and hemoglobin synthesis in fetal mouse yolk sac erythroid cells. Proc Natl Acad Sci U S A **63:**812-819.

37. **McGrath KE, Kingsley PD, Koniski AD, Porter RL, Bushnell TP, Palis J.** 2008. Enucleation of primitive erythroid cells generates a transient population of "pyrenocytes" in the mammalian fetus. Blood **111:**2409-2417.

38. **Heath DS, Axelrad AA, McLeod DL, Shreeve MM.** 1976. Separation of the erythropoietin-responsive progenitors BFU-E and CFU-E in mouse bone marrow by unit gravity sedimentation. Blood **47:**777-792.

39. **Stephenson JR, Axelrad AA, McLeod DL, Shreeve MM.** 1971. Induction of colonies of hemoglobin-synthesizing cells by erythropoietin in vitro. Proc Natl Acad Sci U S A **68:**1542-1546.

40. **Palis J, Robertson S, Kennedy M, Wall C, Keller G.** 1999. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. Development **126:**5073-5084.

41. **Zanarini MC, Skodol AE, Bender D, Dolan R, Sanislow C, Schaefer E, Morey LC, Grilo CM, Shea MT, McGlashan TH, Gunderson JG.** 2000. The Collaborative Longitudinal Personality Disorders Study: reliability of axis I and II diagnoses. J Pers Disord **14:**291-299.

42. **Rifkind RA, Chui D, Epler H.** 1969. An ultrastructural study of early morphogenetic events during the establishment of fetal hepatic erythropoiesis. J Cell Biol **40:**343-365.

43. **Russel ES, Thompson MW, McFarland E.** 1968. Analysis of effects of W and f genic substitutions on fetal mouse hematology. Genetics **58:**259-270.

44. **Steiner R.** 1973. On the kinetics of erythroid cell differentiation in fetal mice. II. DNA and hemoglobin measurements of individual erythroblasts during gestation. J Cell Physiol **82:**219-230.

45. **Liu XS, Li XH, Wang Y, Shu RZ, Wang L, Lu SY, Kong H, Jin YE, Zhang LJ, Fei J, Chen SJ, Chen Z, Gu MM, Lu ZY, Wang ZG.** 2007. Disruption of palladin leads to defects in definitive erythropoiesis by interfering with erythroblastic island formation in mouse fetal liver. Blood **110:**870-876.

46. **Baron MH, Isern J, Fraser ST.** 2012. The embryonic origins of erythropoiesis in mammals. Blood **119:**4828-4837.

47. **Lux CT, Yoshimoto M, McGrath K, Conway SJ, Palis J, Yoder MC.** 2008. All primitive and definitive hematopoietic progenitor cells emerging before E10 in the mouse embryo are products of the yolk sac. Blood **111:**3435-3438.

48. **Paulson RF, Shi L, Wu DC.** 2011. Stress erythropoiesis: new signals and new stress progenitor cells. Curr Opin Hematol **18:**139-145.

49. **Amato A, Cappabianca MP, Perri M, Zaghis I, Grisanti P, Ponzini D, Di Biagio P.** 2013. Interpreting elevated fetal hemoglobin in pathology and health at the basic laboratory level: new and known γ-gene mutations associated with hereditary persistence of fetal hemoglobin. Int J Lab Hematol.

50. **Sankaran VG, Xu J, Orkin SH.** 2010. Advances in the understanding of haemoglobin switching. Br J Haematol **149:**181-194.

51. **Noordermeer D, de Laat W.** 2008. Joining the loops: β-globin gene regulation. IUBMB Life **60:**824-833.

52. **Oneal PA, Gantt NM, Schwartz JD, Bhanu NV, Lee YT, Moroney JW, Reed CH, Schechter AN, Luban NL, Miller JL.** 2006. Fetal hemoglobin silencing in humans. Blood **108:**2081-2086.

53. **Liebhaber SA, Russell JE.** 1998. Expression and developmental control of the human α-globin gene cluster. Ann N Y Acad Sci **850:**54-63.

54. **Perkins AC, Peterson KR, Stamatoyannopoulos G, Witkowska HE, Orkin SH.** 2000. Fetal expression of a human Aγ globin transgene rescues globin chain imbalance but not hemolysis in EKLF null mouse embryos. Blood **95:**1827-1833.

55. **Baron MH.** 1997. Transcriptional control of globin gene switching during vertebrate development. Biochim Biophys Acta **1351:**51-72.

56. **Galkin O, Vekilov PG.** 2004. Mechanisms of homogeneous nucleation of polymers of sickle cell anemia hemoglobin in deoxy state. J Mol Biol **336:**43-59.

57. **Pauling L, Itano HA, et al.** 1949. Sickle cell anemia, a molecular disease. Science **109:**443.

58. **Frenette PS, Atweh GF.** 2007. Sickle cell disease: old discoveries, new concepts, and future promise. J Clin Invest **117:**850-858.

59. **Mabaera R, West RJ, Conine SJ, Macari ER, Boyd CD, Engman CA, Lowrey CH.** 2008. A cell stress signaling model of fetal hemoglobin induction: what doesn't kill red blood cells may make them stronger. Exp Hematol **36:**1057-1072.

60. **Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O, Steinberg MH, Klug PP.** 1994. Mortality in sickle cell disease. Life expectancy and risk factors for early death. N Engl J Med **330:**1639-1644.

61. **Garner C, Tatu T, Reittie JE, Littlewood T, Darley J, Cervino S, Farrall M, Kelly P, Spector TD, Thein SL.** 2000. Genetic influences on F cells and other hematologic variables: a twin heritability study. Blood **95:**342-346.

62. **Thein SL, Menzel S.** 2009. Discovering the genetics underlying foetal haemoglobin production in adults. Br J Haematol **145:**455-467.

63. **Watson J.** 1948. A study of sickling of young erythrocytes in sickle cell anemia. Blood **3:**465-469.

64. **Weatherall DJ, Clegg JB.** 2001. Inherited haemoglobin disorders: an increasing global health problem. Bull World Health Organ **79:**704-712.

65. **Bohmer RM.** 2003. Reactivation of fetal hemoglobin in adult stem cell erythropoiesis by transforming growth factor-β. J Hematother Stem Cell Res **12:**499-504.

66. **Friedman S, Schwartz E, Ahern E, Ahern V.** 1976. Variations in globin chain synthesis in hereditary persistence of fetal haemoglobin. Br J Haematol **32:**357-364.

67. **Collins FS, Stoeckert CJ, Jr., Serjeant GR, Forget BG, Weissman SM.** 1984. Gγ β+ hereditary persistence of fetal hemoglobin: cosmid cloning and identification of a specific mutation 5' to the Gγ gene. Proc Natl Acad Sci U S A **81:**4894-4898.

68. **Costa FF, Zago MA, Cheng G, Nechtman JF, Stoming TA, Huisman TH.** 1990. The Brazilian type of nondeletional Aγ-fetal hemoglobin has a C→G substitution at nucleotide -195 of the Aγ-globin gene. Blood **76:**1896-1897.

69. **Gelinas R, Bender M, Lotshaw C, Waber P, Kazazian H, Jr., Stamatoyannopoulos G.** 1986. Chinese Aγ fetal hemoglobin: C to T substitution at position-196 of the Aγ gene promoter. Blood **67:**1777-1779.

70. **Giglioni B, Casini C, Mantovani R, Merli S, Comi P, Ottolenghi S, Saglio G, Camaschella C, Mazza U.** 1984. A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and β-thalassemia. EMBO J **3:**2641-2645.

71. **Gilman JG, Mishima N, Wen XJ, Kutlar F, Huisman TH.** 1988. Upstream promoter mutation associated with a modest elevation of fetal hemoglobin expression in human adults. Blood **72:**78-81.

72. **Tasiopoulou M, Boussiou M, Sinopoulou K, Moraitis G, Loutradi-Anagnostou A, Karababa P.** 2008. Gγ -196 C→T, Aγ -201 C→T: two novel mutations in the promoter region of the γ-globin genes associated with nondeletional hereditary persistence of fetal hemoglobin in Greece. Blood Cells Mol Dis **40:**320-322.

73. **Pearson R, Fleetwood J, Eaton S, Crossley M, Bao S.** 2008. Kruppel-like transcription factors: a functional family. Int J Biochem Cell Biol **40:**1996-2001.

74. **McConnell BB, Yang VW.** 2010. Mammalian Kruppel-like factors in health and diseases. Physiol Rev **90:**1337-1381.

75. **Pei J, Grishin NV.** 2013. A new family of predicted Kruppel-like factor genes and pseudogenes in placental mammals. PLoS One **8:**e81109.

76. **Kaczynski J, Cook T, Urrutia R.** 2003. Sp1- and Kruppel-like transcription factors. Genome Biol **4:**206.

77. **van Vliet J, Crofts LA, Quinlan KG, Czolij R, Perkins AC, Crossley M.** 2006. Human KLF17 is a new member of the Sp/KLF family of transcription factors. Genomics **87:**474-482.

78. **Philipsen S, Suske G.** 1999. A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. Nucleic Acids Research **27:**2991-3000.

79. **Miller IJ, Bieker JJ.** 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. Mol Cell Biol **13:**2776-2786.

80. **Chen X, Bieker JJ.** 2001. Unanticipated repression function linked to erythroid Kruppel-like factor. Mol Cell Biol **21:**3118-3125.

81. **Luo Q, Ma X, Wahl SM, Bieker JJ, Crossley M, Montaner LJ.** 2004. Activation and repression of interleukin-12 p40 transcription by erythroid Kruppel-like factor in macrophages. J Biol Chem **279:**18451-18456.

82. **Feng WC, Southwood CM, Bieker JJ.** 1994. Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. J Biol Chem **269:**1493-1500.

83. **Raich N, Romeo PH.** 1993. Erythroid regulatory elements. Stem Cells **11:**95-104.

84. **Tallack MR, Whitington T, Yuen WS, Wainwright EN, Keys JR, Gardiner BB, Nourbakhsh E, Cloonan N, Grimmond SM, Bailey TL, Perkins AC.** 2010. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome Res **20:**1052-1063.

85. **Su MY, Steiner LA, Bogardus H, Mishra T, Schulz VP, Hardison RC, Gallagher PG.** 2013. Identification of biologically relevant enhancers in human erythroid cells. J Biol Chem **288:**8433-8444.

86. **Donze D, Townes TM, Bieker JJ.** 1995. Role of erythroid Kruppel-like factor in human γ- to β-globin gene switching. J Biol Chem **270:**1955-1959.

87. **Hodge D, Coghill E, Keys J, Maguire T, Hartmann B, McDowall A, Weiss M, Grimmond S, Perkins A.** 2006. A global role for EKLF in definitive and primitive erythropoiesis. Blood **107:**3359-3370.

88. **Shyu YC, Wen SC, Lee TL, Chen X, Hsu CT, Chen H, Chen RL, Hwang JL, Shen CK.** 2006. Chromatin-binding in vivo of the Erythroid Kruppel-like Factor, EKLF, in the murine globin loci. Cell Res **16:**347-355.

89. **Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F.** 1995. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature **375:**316-318.

90. **Perkins AC, Sharpe AH, Orkin SH.** 1995. Lethal β-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature **375:**318-322.

91. **Funnell AP, Norton LJ, Mak KS, Burdach J, Artuz CM, Twine NA, Wilkins MR, Power CA, Hung TT, Perdomo J, Koh P, Bell-Anderson KS, Orkin SH, Fraser ST, Perkins AC, Pearson RC, Crossley M.** 2012. The CACCC-binding protein KLF3/BKLF represses a subset of KLF1/EKLF target genes and is required for proper erythroid maturation in vivo. Mol Cell Biol **32:**3281-3292.

92. **Tallack MR, Perkins AC.** 2010. KLF1 directly coordinates almost all aspects of terminal erythroid differentiation. IUBMB Life **62:**886-890.

93. **Crossley M, Whitelaw E, Perkins A, Williams G, Fujiwara Y, Orkin SH.** 1996. Isolation and characterization of the cDNA encoding BKLF/TEF-2, a major CACCC-box-binding protein in erythroid cells and selected other cells. Mol Cell Biol **16:**1695-1705.

94. **Funnell APW, Maloney CA, Thompson LJ, Keys J, Tallack M, Perkins AC, Crossley M.** 2007. Erythroid Kruppel-like factor directly activates the basic Kruppel-like factor gene in erythroid cells. Mol Cell Biol **27:**2777-2790.

95. **Basu P, Morris PE, Haar JL, Wani MA, Lingrel JB, Gaensler KM, Lloyd JA.** 2005. KLF2 is essential for primitive erythropoiesis and regulates the human and murine embryonic β-like globin genes in vivo. Blood **106:**2566-2571.

96. **Anderson KP, Kern CB, Crable SC, Lingrel JB.** 1995. Isolation of a gene encoding a functional zinc finger protein homologous to erythroid Kruppel-like factor: identification of a new multigene family. Mol Cell Biol **15:**5957-5965.

97. **Kuo CT, Veselits ML, Barton KP, Lu MM, Clendenin C, Leiden JM.** 1997. The LKLF transcription factor is required for normal tunica media formation and blood vessel stabilization during murine embryogenesis. Genes Dev **11:**2996-3006.

98. **Wani MA, Means RT, Jr., Lingrel JB.** 1998. Loss of LKLF function results in embryonic lethality in mice. Transgenic Res **7:**229-238.

99. **Basu P, Lung TK, Lemsaddek W, Sargent TG, Williams DC, Jr., Basu M, Redmond LC, Lingrel JB, Haar JL, Lloyd JA.** 2007. EKLF and KLF2 have compensatory roles in embryonic β-globin gene expression and primitive erythropoiesis. Blood **110:**3417-3425.

100. **Alhashem YN, Vinjamur DS, Basu M, Klingmuller U, Gaensler KM, Lloyd JA.** 2011. Transcription factors KLF1 and KLF2 positively regulate embryonic and fetal β-globin genes through direct promoter binding. J Biol Chem **286:**24819-24827.

101. **Pang CJ, Lemsaddek W, Alhashem YN, Bondzi C, Redmond LC, Ah-Son N, Dumur CI, Archer KJ, Haar JL, Lloyd JA, Trudel M.** 2012. Kruppel-like factor 1 (KLF1), KLF2, and Myc control a regulatory network essential for embryonic erythropoiesis. Mol Cell Biol **32:**2628-2644.

102. **Turner J, Crossley M.** 1999. Basic Kruppel-like factor functions within a network of interacting haematopoietic transcription factors. Int J Biochem Cell Biol **31:**1169-1174.

103. **van Vliet J, Turner J, Crossley M.** 2000. Human Krüppel-like factor 8: a CACCC-box binding protein that associates with CtBP and represses transcription. Nucleic Acids Research **28:**1955-1962.

104. **Funnell AP, Mak KS, Twine NA, Pelka GJ, Norton LJ, Radziewic T, Power M, Wilkins MR, Bell-Anderson KS, Fraser ST, Perkins AC, Tam PP, Pearson RC, Crossley M.** 2013. Generation of mice deficient in both KLF3/BKLF and KLF8

reveals a genetic interaction and a role for these factors in embryonic globin gene silencing. Mol Cell Biol **33:**2976-2987.

105. **Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH.** 2008. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science **322:**1839-1842.

106. **Lettre G, Sankaran VG, Bezerra MA, Araujo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN, Orkin SH.** 2008. DNA polymorphisms at the BCL11A, HBS1L-MYB, and β-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc Natl Acad Sci U S A **105:**11869-11874.

107. **Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, Piras MG, Sestu N, Lai S, Dei M, Mulas A, Crisponi L, Naitza S, Asunis I, Deiana M, Nagaraja R, Perseu L, Satta S, Cipollina MD, Sollaino C, Moi P, Hirschhorn JN, Orkin SH, Abecasis GR, Schlessinger D, Cao A.** 2008. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of β-thalassemia. Proc Natl Acad Sci U S A **105:**1620-1625.

108. **Menzel S, Garner C, Gut I, Matsuda F, Yamaguchi M, Heath S, Foglio M, Zelenika D, Boland A, Rooks H, Best S, Spector TD, Farrall M, Lathrop M, Thein SL.** 2007. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat Genet **39:**1197-1199.

109. **Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L, Sabo PJ, Vierstra J, Voit RA, Yuan GC, Porteus MH, Stamatoyannopoulos JA, Lettre G, Orkin SH.** 2013. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science **342:**253-257.

110. **Sankaran VG, Xu J, Ragoczy T, Ippolito GC, Walkley CR, Maika SD, Fujiwara Y, Ito M, Groudine M, Bender MA, Tucker PW, Orkin SH.** 2009. Developmental and species-divergent globin switching are driven by BCL11A. Nature **460:**1093-1097.

111. **Xu J, Peng C, Sankaran VG, Shao Z, Esrick EB, Chong BG, Ippolito GC, Fujiwara Y, Ebert BL, Tucker PW, Orkin SH.** 2011. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science **334:**993-996.

112. **Zhou D, Liu K, Sun CW, Pawlik KM, Townes TM.** 2010. KLF1 regulates BCL11A expression and γ- to β-globin gene switching. Nat Genet **42:**742-744.

113. **Satta S, Perseu L, Moi P, Asunis I, Cabriolu A, Maccioni L, Demartis FR, Manunza L, Cao A, Galanello R.** 2011. Compound heterozygosity for KLF1 mutations associated with remarkable increase of fetal hemoglobin and red cell protoporphyrin. Haematologica **96:**767-770.

114. **Gallienne AE, Dreau HM, Schuh A, Old JM, Henderson S.** 2012. Ten novel mutations in the erythroid transcription factor KLF1 gene associated with increased fetal hemoglobin levels in adults. Haematologica **97:**340-343.

115. **Esteghamat F, Gillemans N, Bilic I, van den Akker E, Cantu I, van Gent T, Klingmuller U, van Lom K, von Lindern M, Grosveld F, Bryn van Dijk T, Busslinger M, Philipsen S.** 2013. Erythropoiesis and globin switching in compound Klf1::Bcl11a mutant mice. Blood **121:**2553-2562.

116. **Trakarnsanga K, Wilson MC, Lau W, Singleton BK, Parsons SF, Sakuntanaga P, Kurita R, Nakamura Y, Anstee DJ, Frayne J.** 2014. Induction of adult levels of β-globin in human erythroid cells that intrinsically express embryonic or fetal globin by transduction with KLF1 and BCL11A-XL. Haematologica **99:**1677-1685.

117. **Cantu C, Ierardi R, Alborelli I, Fugazza C, Cassinelli L, Piconese S, Bose F, Ottolenghi S, Ferrari G, Ronchi A.** 2011. Sox6 enhances erythroid differentiation in human erythroid progenitors. Blood **117:**3669-3679.

118. **Cohen-Barak O, Erickson DT, Badowski MS, Fuchs DA, Klassen CL, Harris DT, Brilliant MH.** 2007. Stem cell transplantation demonstrates that Sox6 represses εy globin expression in definitive erythropoiesis of adult mice. Exp Hematol **35:**358-367.

119. **Yi Z, Cohen-Barak O, Hagiwara N, Kingsley PD, Fuchs DA, Erickson DT, Epner EM, Palis J, Brilliant MH.** 2006. Sox6 directly silences ε-globin expression in definitive erythropoiesis. PLoS Genet **2:**e14.

120. **Xu J, Sankaran VG, Ni M, Menne TF, Puram RV, Kim W, Orkin SH.** 2010. Transcriptional silencing of γ-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev **24:**783-798.

121. **Ko LJ, Engel JD.** 1993. DNA-binding specificities of the GATA transcription factor family. Mol Cell Biol **13:**4011-4022.

122. **Leonard M, Brice M, Engel JD, Papayannopoulou T.** 1993. Dynamics of GATA transcription factor expression during erythroid differentiation. Blood **82:**1071-1079.

123. **Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH.** 1996. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Proc Natl Acad Sci U S A **93:**12355-12358.

124. **Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, Orkin SH, Costantini F.** 1991. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature **349:**257-260.

125. **Weiss MJ, Keller G, Orkin SH.** 1994. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. Genes Dev **8:**1184-1197.

126. **Pevny L, Lin CS, D'Agati V, Simon MC, Orkin SH, Costantini F.** 1995. Development of hematopoietic cells lacking transcription factor GATA-1. Development **121:**163-172.

127. **Evans T, Reitman M, Felsenfeld G.** 1988. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc Natl Acad Sci U S A **85:**5976-5980.

128. **Wall L, deBoer E, Grosveld F.** 1988. The human β-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. Genes Dev **2:**1089-1100.

129. **Martin DI, Orkin SH.** 1990. Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. Genes Dev **4:**1886-1898.

130. **Harju-Baker S, Costa FC, Fedosyuk H, Neades R, Peterson KR.** 2008. Silencing of Aγ-globin gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the -566 GATA site. Mol Cell Biol **28:**3101-3113.

131. **Bottardi S, Ross J, Bourgoin V, Fotouhi-Ardakani N, Affar el B, Trudel M, Milot E.** 2009. Ikaros and GATA-1 combinatorial effect is required for silencing of human γ-globin genes. Mol Cell Biol **29:**1526-1537.

132. **Woon Kim Y, Kim S, Geun Kim C, Kim A.** 2011. The distinctive roles of erythroid specific activator GATA-1 and NF-E2 in transcription of the human fetal γ-globin genes. Nucleic Acids Res **39:**6944-6955.

133. **Chlon TM, Dore LC, Crispino JD.** 2012. Cofactor-mediated restriction of GATA-1 chromatin occupancy coordinates lineage-specific gene expression. Mol Cell **47:**608-621.

134. **Keys JR, Tallack MR, Zhan Y, Papathanasiou P, Goodnow CC, Gaensler KM, Crossley M, Dekker J, Perkins AC.** 2008. A mechanism for Ikaros regulation of human globin gene switching. Br J Haematol **141:**398-406.

135. **Costa FC, Fedosyuk H, Chazelle AM, Neades RY, Peterson KR.** 2012. Mi2beta is required for γ-globin gene silencing: temporal assembly of a GATA-1-FOG-1-Mi2 repressor complex in β-YAC transgenic mice. PLoS Genet **8:**e1003155.

136. **Lee SU, Maeda T.** 2012. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. Immunol Rev **247:**107-119.

137. **Maeda T, Ito K, Merghoub T, Poliseno L, Hobbs RM, Wang G, Dong L, Maeda M, Dore LC, Zelent A, Luzzatto L, Teruya-Feldstein J, Weiss MJ, Pandolfi PP.** 2009. LRF is an essential downstream target of GATA-1 in erythroid development and regulates BIM-dependent apoptosis. Dev Cell **17:**527-540.

138. **Lee SU, Maeda M, Ishikawa Y, Li SM, Wilson A, Jubb AM, Sakurai N, Weng L, Fiorini E, Radtke F, Yan M, Macdonald HR, Chen CC, Maeda T.** 2013. LRF-mediated Dll4 repression in erythroblasts is necessary for hematopoietic stem cell maintenance. Blood **121:**918-929.

139. **Pessler F, Hernandez N.** 2003. Flexible DNA binding of the BTB/POZ-domain protein FBI-1. J Biol Chem **278:**29327-29335.

140. **Martyn G.** 2013. A potential role for ZBTB7A in the foetal to adult globin gene switch. University of New South Wales.

141. **Sambrook J, Fritsch EK, Maniatis T.** 1989. Molecular cloning. A laboratory manual., 2nd ed. Cold Spring Harbour Laboratory Press.

142. **Coghill E, Eccleston S, Fox V, Cerruti L, Brown C, Cunningham J, Jane S, Perkins A.** 2001. Erythroid Kruppel-like factor (EKLF) coordinates erythroid cell proliferation and hemoglobinization in cell lines derived from EKLF null mice. Blood **97:**1861-1868.

143. **Andrews NC, Faller DV.** 1991. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res **19:**2499.

144. **Schuller CE, Jankowski K, Mackenzie KL.** 2007. Telomere length of cord blood-derived CD34(+) progenitors predicts erythroid proliferative potential. Leukemia **21:**983-991.

145. **Giarratana MC, Rouard H, Dumont A, Kiger L, Safeukui I, Le Pennec PY, Francois S, Trugnan G, Peyrard T, Marie T, Jolly S, Hebert N, Mazurier C, Mario N, Harmand L, Lapillonne H, Devaux JY, Douay L.** 2011. Proof of principle for transfusion of in vitro-generated red blood cells. Blood **118:**5071-5079.

146. **Perdomo J, Verger A, Turner J, Crossley M.** 2005. Role for SUMO modification in facilitating transcriptional repression by BKLF. Mol Cell Biol **25:**1549-1559.

147. **Abramoff MD, Magalhaes PJ, Ram SJ.** 2004. Image processing with ImageJ. Biophoton Int **11:**36-42.

148. **Schneider CA, Rasband WS, Eliceiri KW.** 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods **9:**671-675.

149. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

150. **Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT.** 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods **48:**240-248.

151. **Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B.** 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res **40:**W622-627.

152. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357-359.

153. **Allison K.** 2014. Homer-idr first-pass. ZENODO.

154. **Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK.** 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell **38:**576-589.

155. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S.** 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:**2078-2079.

156. **Thorvaldsdottir H, Robinson JT, Mesirov JP.** 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform **14:**178-192.

157. **Machanick P, Bailey TL.** 2011. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics **27:**1696-1697.

158. **Bailey TL, Williams N, Misleh C, Li WW.** 2006. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res **34:**W369-373.

159. **Bailey TL.** 2011. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics **27:**1653-1659.

160. **Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A.** 2005. Galaxy: a platform for interactive large-scale genome analysis. Genome Res **15:**1451-1455.

161. **Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J.** 2010. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol **Chapter 19:**Unit 19 10 11-21.

162. **Goecks J, Nekrutenko A, Taylor J, Galaxy T.** 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol **11:**R86.

163. **Bauer DE, Orkin SH.** 2011. Update on fetal hemoglobin gene regulation in hemoglobinopathies. Curr Opin Pediatr **23:**1-8.

164. **Yu M, Riva L, Xie H, Schindler Y, Moran TB, Cheng Y, Yu D, Hardison R, Weiss MJ, Orkin SH, Bernstein BE, Fraenkel E, Cantor AB.** 2009. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. Mol Cell **36:**682-695.

165. **Newton DA, Rao KM, Dluhy RA, Baatz JE.** 2006. Hemoglobin is expressed by alveolar epithelial cells. J Biol Chem **281:**5668-5676.

166. **Eaton SA, Funnell AP, Sue N, Nicholas H, Pearson RC, Crossley M.** 2008. A network of Kruppel-like Factors (Klfs): Klf8 is repressed by Klf3 and activated by Klf1 in vivo. J Biol Chem **283:**26937-26947.

167. **Dean A, Ley TJ, Humphries RK, Fordis M, Schechter AN.** 1983. Inducible transcription of five globin genes in K562 human leukemia cells. Proc Natl Acad Sci U S A **80:**5515-5519.

168. **Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M.** 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res **22:**1813-1831.

169. **Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigo R,**

Hubbard TJ. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res **22:**1760-1774.

170. **Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght**

A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, International Human Genome Sequencing C. 2001. Initial sequencing and analysis of the human genome. Nature **409:**860-921.

171. **Apostolopoulou K, Pateras IS, Evangelou K, Tsantoulis PK, Liontos M, Kittas C, Tiniakos DG, Kotsinas A, Cordon-Cardo C, Gorgoulis VG.** 2007. Gene amplification is a relatively frequent event leading to ZBTB7A (Pokemon) overexpression in non-small cell lung cancer. J Pathol **213:**294-302.

172. **Choi WI, Kim Y, Kim Y, Yu MY, Park J, Lee CE, Jeon BN, Koh DI, Hur MW.** 2009. Eukaryotic translation initiator protein 1A isoform, CCS-3, enhances the transcriptional repression of p21CIP1 by proto-oncogene FBI-1 (Pokemon/ZBTB7A). Cell Physiol Biochem **23:**359-370.

173. **Jeon BN, Yoo JY, Choi WI, Lee CE, Yoon HG, Hur MW.** 2008. Proto-oncogene FBI-1 (Pokemon/ZBTB7A) represses transcription of the tumor suppressor Rb gene via binding competition with Sp1 and recruitment of co-repressors. J Biol Chem **283:**33199-33210.

174. **Aggarwal H, Aggarwal A, Agrawal DK.** 2011. Epidermal growth factor increases LRF/Pokemon expression in human prostate cancer cells. Exp Mol Pathol **91:**496-501.

175. **Aggarwal H, Aggarwal A, Hunter WJ, 3rd, Yohannes P, Khan AU, Agrawal DK.** 2011. Expression of leukemia/lymphoma related factor (LRF/Pokemon) in human benign prostate hyperplasia and prostate cancer. Exp Mol Pathol **90:**226-230.

176. **Liu XS, Haines JE, Mehanna EK, Genet MD, Ben-Sahra I, Asara JM, Manning BD, Yuan ZM.** 2014. ZBTB7A acts as a tumor suppressor through the transcriptional repression of glycolysis. Genes Dev **28:**1917-1928.

177. **Wang G, Lunardi A, Zhang J, Chen Z, Ala U, Webster KA, Tay Y, Gonzalez-Billalabeitia E, Egia A, Shaffer DR, Carver B, Liu XS, Taulli R, Kuo WP, Nardella C, Signoretti S, Cordon-Cardo C, Gerald WL, Pandolfi PP.** 2013. Zbtb7a suppresses prostate cancer through repression of a Sox9-dependent pathway for cellular senescence bypass and tumor invasion. Nat Genet **45:**739-746.

178. **Schaeffer EK, West RJ, Conine SJ, Lowrey CH.** 2014. Multiple physical stresses induce γ-globin gene expression and fetal hemoglobin production in erythroid cells. Blood Cells Mol Dis **52:**214-224.

179. **Chen JJ.** 2014. Stress-enhanced translation of γ-globin mRNA. Blood **124:**2622-2623.

180. **Choi WI, Jeon BN, Park H, Yoo JY, Kim YS, Koh DI, Kim MH, Kim YR, Lee CE, Kim KS, Osborne TF, Hur MW.** 2008. Proto-oncogene FBI-1 (Pokemon) and SREBP-1 synergistically activate transcription of fatty-acid synthase gene (FASN). J Biol Chem **283:**29341-29354.

181. **Xu Z, Huang S, Chang LS, Agulnick AD, Brandt SJ.** 2003. Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein LDB1 in erythroid gene expression and differentiation. Mol Cell Biol **23:**7585-7599.

182. **Yang B, Kirby S, Lewis J, Detloff PJ, Maeda N, Smithies O.** 1995. A mouse model for beta 0-thalassemia. Proc Natl Acad Sci U S A **92:**11608-11612.

183. **Lim SK, Bieker JJ, Lin CS, Costantini F.** 1997. A shortened life span of EKLF-/- adult erythrocytes, due to a deficiency of β-globin chains, is ameliorated by human γ-globin chains. Blood **90:**1291-1299.

184. **Nilson DG, Sabatino DE, Bodine DM, Gallagher PG.** 2006. Major erythrocyte membrane protein genes in EKLF-deficient mice. Exp Hematol **34:**705-712.

185. **Borg J, Papadopoulos P, Georgitsi M, Gutierrez L, Grech G, Fanis P, Phylactides M, Verkerk AJ, van der Spek PJ, Scerri CA, Cassar W, Galdies R, van Ijcken W, Ozgur Z, Gillemans N, Hou J, Bugeja M, Grosveld FG, von Lindern M, Felice AE, Patrinos GP, Philipsen S.** 2010. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. Nat Genet **42:**801-805.

186. **Huang da W, Sherman BT, Lempicki RA.** 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc **4:**44-57.

187. **Huang da W, Sherman BT, Lempicki RA.** 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res **37:**1-13.

188. **Kihm AJ, Kong Y, Hong W, Russell JE, Rouda S, Adachi K, Simon MC, Blobel GA, Weiss MJ.** 2002. An abundant erythroid protein that stabilizes free α-haemoglobin. Nature **417:**758-763.

189. **Keys JR, Tallack MR, Hodge DJ, Cridland SO, David R, Perkins AC.** 2007. Genomic organisation and regulation of murine α-haemoglobin stabilising protein by erythroid Kruppel-like factor. Br J Haematol **136:**150-157.

190. **Dumitriu B, Patrick MR, Petschek JP, Cherukuri S, Klingmuller U, Fox PL, Lefebvre V.** 2006. Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. Blood **108:**1198-1207.

191. **Vu TT, Gatto D, Turner V, Funnell AP, Mak KS, Norton LJ, Kaplan W, Cowley MJ, Agenes F, Kirberg J, Brink R, Pearson RC, Crossley M.** 2011. Impaired B cell development in the absence of Kruppel-like factor 3. J Immunol **187:**5032-5042.

192. **Piazza F, Costoya JA, Merghoub T, Hobbs RM, Pandolfi PP.** 2004. Disruption of PLZP in mice leads to increased T-lymphocyte proliferation, cytokine production, and altered hematopoietic stem cell homeostasis. Mol Cell Biol **24:**10456-10469.

193. **Kang BY, Miaw SC, Ho IC.** 2005. ROG negatively regulates T-cell activation but is dispensable for Th-cell differentiation. Mol Cell Biol **25:**554-562.

194. **Beaulieu AM, Zawislak CL, Nakayama T, Sun JC.** 2014. The transcription factor Zbtb32 controls the proliferative burst of virus-specific natural killer cells responding to infection. Nat Immunol **15:**546-553.

195. **Gilligan DM, Lozovatsky L, Gwynn B, Brugnara C, Mohandas N, Peters LL.** 1999. Targeted disruption of the beta adducin gene (Add2) causes red blood cell spherocytosis in mice. Proc Natl Acad Sci U S A **96:**10717-10722.

196. **Yenerel MN, Sundell IB, Weese J, Bulger M, Gilligan DM.** 2005. Expression of adducin genes during erythropoiesis: a novel erythroid promoter for ADD2. Exp Hematol **33:**758-766.

197. **Rybicki AC, Schwartz RS, Qiu JJ, Gilman JG.** 1994. Molecular cloning of mouse erythrocyte protein 4.2: a membrane protein with strong homology with the transglutaminase supergene family. Mamm Genome **5:**438-445.

198. **Dooner GJ, Barker JE, Gallagher PG, Debatis ME, Brown AH, Forget BG, Becker PS.** 2000. Gene transfer to ankyrin-deficient bone marrow corrects spherocytosis in vitro. Exp Hematol **28:**765-774.

199. **Guernsey DL, Jiang H, Campagna DR, Evans SC, Ferguson M, Kellogg MD, Lachance M, Matsuoka M, Nightingale M, Rideout A, Saint-Amant L, Schmidt PJ, Orr A, Bottomley SS, Fleming MD, Ludman M, Dyack S, Fernandez CV, Samuels ME.** 2009. Mutations in mitochondrial carrier family gene SLC25A38 cause

nonsyndromic autosomal recessive congenital sideroblastic anemia. Nat Genet **41:**651-653.

200. **Song CZ, Gavriilidis G, Asano H, Stamatoyannopoulos G.** 2005. Functional study of transcription factor KLF11 by targeted gene inactivation. Blood Cells Mol Dis **34:**53-59.

201. **Warren AJ, Colledge WH, Carlton MB, Evans MJ, Smith AJ, Rabbitts TH.** 1994. The oncogenic cysteine-rich LIM domain protein RBTN2 is essential for erythroid development. Cell **78:**45-57.

202. **Nam CH, Lobato MN, Appert A, Drynan LF, Tanaka T, Rabbitts TH.** 2008. An antibody inhibitor of the LMO2 protein complex blocks its normal and tumorigenic functions. Oncogene **27:**4962-4968.

203. **Brandt SJ, Koury MJ.** 2009. Regulation of LMO2 mRNA and protein expression in erythroid differentiation. Haematologica **94:**447-448.

204. **Inoue A, Fujiwara T, Okitsu Y, Katsuoka Y, Fukuhara N, Onishi Y, Ishizawa K, Harigae H.** 2013. Elucidation of the role of LMO2 in human erythroid cells. Exp Hematol **41:**1062-1076 e1061.

205. **Harada H, Harada Y, O'Brien DP, Rice DS, Naeve CW, Downing JR.** 1999. HERF1, a novel hematopoiesis-specific RING finger protein, is required for terminal differentiation of erythroid cells. Mol Cell Biol **19:**3808-3815.

206. **Keel SB, Doty RT, Yang Z, Quigley JG, Chen J, Knoblaugh S, Kingsley PD, De Domenico I, Vaughn MB, Kaplan J, Palis J, Abkowitz JL.** 2008. A heme export protein is required for red blood cell differentiation and iron homeostasis. Science **319:**825-828.

207. **Littlewood TD, Hancock DC, Danielian PS, Parker MG, Evan GI.** 1995. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. Nucleic Acids Res **23:**1686-1690.

208. **Chen X, Bieker JJ.** 2004. Stage-specific repression by the EKLF transcriptional activator. Mol Cell Biol **24:**10416-10424.

209. **Siatecka M, Xue L, Bieker JJ.** 2007. Sumoylation of EKLF promotes transcriptional repression and is involved in inhibition of megakaryopoiesis. Mol Cell Biol **27:**8547-8560.

210. **Frontelo P, Manwani D, Galdass M, Karsunky H, Lohmann F, Gallagher PG, Bieker JJ.** 2007. Novel role for EKLF in megakaryocyte lineage commitment. Blood **110:**3871-3880.

211. **Poliseno L, Pitto L, Simili M, Mariani L, Riccardi L, Ciucci A, Rizzo M, Evangelista M, Mercatanti A, Pandolfi PP, Rainaldi G.** 2008. The proto-oncogene LRF is under post-transcriptional control of MiR-20a: implications for senescence. PLoS One **3:**e2542.

212. **Crossley M, Tsang AP, Bieker JJ, Orkin SH.** 1994. Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. J Biol Chem **269:**15440-15444.

213. **Consortium F, the RP, Clst, Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassman T, Kulakovskiy IV, Lizio M, Itoh M, Andersson R, Mungall CJ, Meehan TF, Schmeier S, Bertin N, Jorgensen M, Dimont E, Arner E, Schmidl C, Schaefer U, Medvedeva YA, Plessy C, Vitezic M, Severin J, Semple C, Ishizu Y, Young RS, Francescatto M, Alam I, Albanese D, Altschuler GM, Arakawa T, Archer JA, Arner P, Babina M, Rennie S, Balwierz PJ, Beckhouse AG, Pradhan-Bhatt S, Blake JA, Blumenthal A, Bodega B, Bonetti A, Briggs J, Brombacher F, Burroughs AM, Califano A, Cannistraci CV, Carbajo D, Chen Y, Chierici M, Ciani Y, Clevers HC, Dalla E, Davis CA, Detmar M, Diehl AD, Dohi T, Drablos F, Edge AS, Edinger M, Ekwall K, Endoh M, Enomoto H, Fagiolini M, Fairbairn L, Fang H, Farach-Carson MC, Faulkner GJ, Favorov AV, Fisher ME, Frith MC, Fujita R, Fukuda S, Furlanello C, Furino M, Furusawa J, Geijtenbeek TB, Gibson AP, Gingeras T, Goldowitz D, Gough J, Guhl S, Guler R, Gustincich S, Ha TJ, Hamaguchi M, Hara M, Harbers M, Harshbarger J, Hasegawa A, Hasegawa Y, Hashimoto T, Herlyn M, Hitchens KJ, Ho Sui SJ, Hofmann OM, Hoof I, Hori F, Huminiecki L, Iida K, Ikawa T, Jankovic BR, Jia H, Joshi A, Jurman G, Kaczkowski B, Kai C, Kaida K, Kaiho A, Kajiyama K, Kanamori-Katayama M, Kasianov AS, Kasukawa T, Katayama S, Kato S, Kawaguchi S, Kawamoto H, Kawamura YI, Kawashima T, Kempfle JS, Kenna TJ, Kere J, Khachigian LM, Kitamura T, Klinken SP, Knox AJ, Kojima M, Kojima S, Kondo N, Koseki H, Koyasu S, Krampitz S, Kubosaki A, Kwon AT, Laros JF, Lee W, Lennartsson A, Li K, Lilje B, Lipovich L, Mackay-Sim A, Manabe R, Mar JC, Marchand B, Mathelier A, Mejhert N, Meynert A, Mizuno Y, de Lima Morais DA, Morikawa H, Morimoto M, Moro K, Motakis E, Motohashi H, Mummery CL, Murata M, Nagao-Sato S, Nakachi Y, Nakahara F, Nakamura T, Nakamura Y, Nakazato K, van Nimwegen E, Ninomiya N, Nishiyori H, Noma S, Noma S, Noazaki T, Ogishima S, Ohkura N, Ohimiya H, Ohno H, Ohshima M, Okada-Hatakeyama M, Okazaki Y, Orlando**

V, Ovchinnikov DA, Pain A, Passier R, Patrikakis M, Persson H, Piazza S, Prendergast JG, Rackham OJ, Ramilowski JA, Rashid M, Ravasi T, Rizzu P, Roncador M, Roy S, Rye MB, Saijyo E, Sajantila A, Saka A, Sakaguchi S, Sakai M, Sato H, Savvi S, Saxena A, Schneider C, Schultes EA, Schulze-Tanzil GG, Schwegmann A, Sengstag T, Sheng G, Shimoji H, Shimoni Y, Shin JW, Simon C, Sugiyama D, Sugiyama T, Suzuki M, Suzuki N, Swoboda RK, t Hoen PA, Tagami M, Takahashi N, Takai J, Tanaka H, Tatsukawa H, Tatum Z, Thompson M, Toyodo H, Toyoda T, Valen E, van de Wetering M, van den Berg LM, Verado R, Vijayan D, Vorontsov IE, Wasserman WW, Watanabe S, Wells CA, Winteringham LN, Wolvetang E, Wood EJ, Yamaguchi Y, Yamamoto M, Yoneda M, Yonekura Y, Yoshida S, Zabierowski SE, Zhang PG, Zhao X, Zucchelli S, Summers KM, Suzuki H, Daub CO, Kawai J, Heutink P, Hide W, Freeman TC, Lenhard B, Bajic VB, Taylor MS, Makeev VJ, Sandelin A, Hume DA, Carninci P, Hayashizaki Y. 2014. A promoter-level mammalian expression atlas. Nature **507:**462-470.

214. **Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, Abugessaisa I, Fukuda S, Hori F, Ishikawa-Kato S, Mungall CJ, Arner E, Baillie JK, Bertin N, Bono H, de Hoon M, Diehl AD, Dimont E, Freeman TC, Fujieda K, Hide W, Kaliyaperumal R, Katayama T, Lassmann T, Meehan TF, Nishikata K, Ono H, Rehli M, Sandelin A, Schultes EA, t Hoen PA, Tatum Z, Thompson M, Toyoda T, Wright DW, Daub CO, Itoh M, Carninci P, Hayashizaki Y, Forrest AR, Kawaji H, consortium F.** 2015. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol **16:**22.

215. **Shivdasani RA, Fujiwara Y, McDevitt MA, Orkin SH.** 1997. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. Embo Journal **16:**3965-3973.

216. **Khurana S, Mukhopadhyay A.** 2008. Hematopoietic progenitors from early murine fetal liver possess hepatic differentiation potential. Am J Pathol **173:**1818-1827.

217. **Kent WJ.** 2002. BLAT--the BLAST-like alignment tool. Genome Res **12:**656-664.

218. **Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D.** 2002. The human genome browser at UCSC. Genome Res **12:**996-1006.

219. **Kassouf MT, Hughes JR, Taylor S, McGowan SJ, Soneji S, Green AL, Vyas P, Porcher C.** 2010. Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. Genome Res **20:**1064-1083.

220. **Jawaid K, Wahlberg K, Thein SL, Best S.** 2010. Binding patterns of BCL11A in the globin and GATA-1 loci and characterization of the BCL11A fetal hemoglobin locus. Blood Cells Mol Dis **45:**140-146.

221. **Sankaran VG, Xu J, Byron R, Greisman HA, Fisher C, Weatherall DJ, Sabath DE, Groudine M, Orkin SH, Premawardhena A, Bender MA.** 2011. A functional element necessary for fetal hemoglobin silencing. N Engl J Med **365:**807-814.

222. **Xu J, Bauer DE, Kerenyi MA, Vo TD, Hou S, Hsu YJ, Yao H, Trowbridge JJ, Mandel G, Orkin SH.** 2013. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proc Natl Acad Sci U S A **110:**6518-6523.

223. **Hardison RC, Blobel GA.** 2013. Genetics. GWAS to therapy by genome edits? Science **342:**206-207.

224. **Wienert B, Funnell AP, Norton LJ, Pearson RC, Wilkinson-White LE, Lester K, Vadolas J, Porteus MH, Matthews JM, Quinlan KG, Crossley M.** 2015. Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. Nat Commun **6:**7085.

Appendix I

Table A1. Murine Primers

	Gene Name	Database numbers	Forward Sequence	Reverse Sequence
Genotyping Primers	*Klf1* WT	3510/3511	AAGGCCACTTCCAGCTCTTTCGCG	TTGGAGTAGCTCTTCCCGCAGCCT
	Neomycin	3512/3591	AGGGTTGGTGACTGGGCCTTTGGG	TAAGGGCCAGCTCATTCCTCCCACT
Real-Time RT-PCR Primers	*Bcl11a*	3431/3432	CCAGAGGATGACGATTGTTTATCA	GAAGTTTATCTGCTATGTGTTCCTGTTT
	Hbb-B1	2744/2745	CACTGTGACAAGCTGCATGT	TAGTGGTACTTGTGAGCCAG
	Hbb-B2	3395/3396	GCACCTGACTGATGCTGAGA	CTGGGTCCAAGGGTAGACAA
	Hbb-Bh1	2581/2741	CTCAAGGAGACCTTTGCTCA	AATCACCAGCTTCTGCCAGGC
	Hbb-y	2732/2733	CAAAGAGAGTTTTTGTTGAAGGAGGAG	AAAGTTCACCATGATGGCAAGTCTGG
	Klf1	1917/1918	AGACTGTCTTACCCTCCATCAGTACA	CCGCCACCACTTGAGGAA
	Klf3 1a	2094/2095	GTCATGTGACTGCCCAGAGTT	CTGCCAAGCGCGGTC
	Klf3 1b	2098/2099	GGTGGAATTCTGTTCAGGTCAAC	CCACGCCTTCTAGGGTGTTCT
	Klf3 Ex4/5	2096/2097	GAAATGTCACCCCCTTTAATGAAC	CACGATGACGGAAGGATGGT
	Pu.1	3547/3548	AGCGATGGAGAAAGCCATAG	TCTGCAGCTCTGTGAAGTGG
	Runx1	4307/4308	TGGCTTCAGACAGCATTTTT	CAGCGCCTCGCTCATCTT
	Zbtb7a	5039/5040	GCGACGTGGTGATTCTTGTG	CGACGTGAACAGCTTCTTGAAGT
ChIP Primers	*Zbtb7a* Promoter	5634/5635	TGTAGCTCTAACGGCCCTAGAAA	AGGAGGCAGGTGGATCTCTGT
	Hbb-B1 Promoter	4441/4442	GCTTCTGACATAGTTGTGTTGACTCA	CAGCAGCCTTCTCAGCATCA
	Klf3 +10 kb	2177/2178	GCCTGCGGGAGGTGATTAC	TTCTTGAAGCAAAGCCAAGAATATC
	Klf3 1b Promoter	2175/2176	CTGGGTGTGGGCAGAATCTT	GCCAGGGCGAGTCCAACT
	Zbtb7a +5.3 kb	5763/5764	GGAAGGAACAGCAAAAAGTCAACT	CAGGATACACAGTGCAGAGAAGCT
	Zbtb7a +6.9 kb	5757/5758	CTCAGTCAAGGAGCCTGAAAGAC	CACACCTCCCTTGAAGTTCCAA
	Zbtb7a 1st Intron	5660/5661	CCACAAAGTCTTGGCAATAAAAAG	GGAGGCGTGTACCGTGTGT
	Zbtb7a -7.7 kb	5652/5653	AGCAGAACGTGTACGAGATCGA	TGGCGGTGTAGGCGAAGT
5'RACE Primers	*mZbtb7a 5'RACE I*	5825	Provided in 5'RACE Kit (Clontech Laboratories Inc. Mountain View, CA, USA)	GCCAGCCATCTTCCGCGACA
	mZbtb7a 5'RACE II	5826		CTCAGGATGTCGCTGCTGTGGTCC
	mZbtb7a 5'RACE III	5827		CCCTGAGTCCGCTGCTCGTTCAG
	mZbtb7a 5'RACE IV	5828		TCCACAAGAATCACCACGTCGCA
	mZbtb7a 5'RACE V	5829		AGCGGTGCGTGGGGAACTCACGTC
	mZbtb7a 5'RACE VI	5830		GAACAGCTTCTTGAAGTACTGGCTGCAG

Table A2. Human Primers

	Gene Name	Database numbers	Forward Sequence	Reverse Sequence
Real-Time RT-PCR Primers	*BCL11A*	4111/4112	CGAGCACAAACGGAAACAATG	GATTAGAGCTCCATGTGCAGAACG
	HbB	4069/4070	TGTCCACTCCTGATGCTGTTATG	GGCACCGAGCACTTTCTTG
	HbD	3994/3995	AACCTCAAGGGCACTTTTTCT	GGAAACAGTCCAGGATCTCAA
	HbE	4141/4142	TGCTGAGGAGAAGGCTGCCG	TGGGTCCAGGGGTAAACAACGAGG
	HbG	2524/2525	CCTGTCCTCTGCCTCTGCC	GGATTGCCAAAACGGTCAC
	KLF1	4105/4106	CCACAGCCGAGACCGCCTTGACC	CTCTCATCGTCCTCTTCCTCCC
	KLF3	3950/3951	ACCCAGTTCCTGTCAAGCAA	TCAGGCAATGGTGTGGAGTA
	ZBTB7A	5041/5042	AAGCCCTACGAGTGCAACATCT	CAGGTCGTAGTTGTGGGCAAA
	ZBTB7A Exon 0	5894/5895	GCTGCGGCACCT TTAAGACA	GCGTCACTGCCCCTACA
ChIP Primers	*BIM* Promoter	4376/4377	CGGGTTGGGGTAGGTGAG	GGCGTGTTTACCGGAGTAAC
	HbB Promoter	3825/3826	GGAGGGCTGAGGGTTTGAAGTCC	TGTCCTTGGCTCTTCTGGCACTG
	HbE Promoter	4445/4446	CACAAACTTAGTGTCCATCCATCAC	CCCTGTTCTCCATGGTACTTAAAAG
	HbG Promoter	1925/1926	TCAATGCAAATATCTGTCTGAAACG	CAAGGCTATTGGTCAAGGCAA
	HbG Promoter -1.1kb	1965/1966	GAGATCATGGATCACTTTCAGAG	AAGTATTTATGGTGGTTTTTTGG
	HbG Promoter -2.1kb	1967/1968	CCTGACCAGGAACCAGCAGAAAAG	AAGGTGCTATAACAAAATAGCATAG
	HbG Promoter -3.1kb	1969/1970	ATGTGGGTTTTGATGAGCAAAT	ACCTTTTACTCCCACTTGCAGAAC
	KLF1 Promoter	4374/4375	TCAAATTAGCCTGGCGTTCAA	AATGGTGGGCCAGTTGTCA
	ZBTB7A +15.4 kb	5672/5673	CACAGAGCTCAAACTGGCTTGA	ACTCGCACTCCCAGGAACAC
	ZBTB7A +8.2 kb	5670/5671	GCCACCGCGTGGATGA	TGTGTGGCTTTTTGTGTCTGACT
	ZBTB7A Intron Peak 1	5660/5661	CCACAAAGTCTTGGCAATAAAAAG	GGAGGCGTGTACCGTGTGT
	ZBTB7A Intron Peak 2	5664/5665	CAAAAGCAATCAGACGTCATGTC	CTGGCCGCCCTTGTGTAC
	ZBTB7A Promoter	5656/5657	CAGGACTCAGTTTCCCCTTCCT	CGCTCACTAAACGGCAGCTAA
5'RACE Primers	*hZBTB7A 5'RACE I*	5819	Provided in 5'RACE Kit (Clontech Laboratories Inc. Mountain View, CA, USA)	TCAGGATGTCGCTGCTGTGGTCG
	hZBTB7A 5'RACE II	5820		CTCGTTCAGCCCACTCAGGATGT
	hZBTB7A 5'RACE III	5821		TGCGTCCGCTGCTCGTTCAGC
	hZBTB7A 5'RACE IV	5822		TCCACCAGGATCACCACGTCGCA
	hZBTB7A 5'RACE V	5823		GCTTCTTGAAGTACTGGCTGCAGGC
	hZBTB7A 5'RACE VI	5824		AAGTCGATCTCGTACACGTTCTGCTG GT

Appendix II

The following figures are representative primary amplification plots from real-time qPCR data in order to display the relative expression levels of important genes in various tissues examined in this thesis.

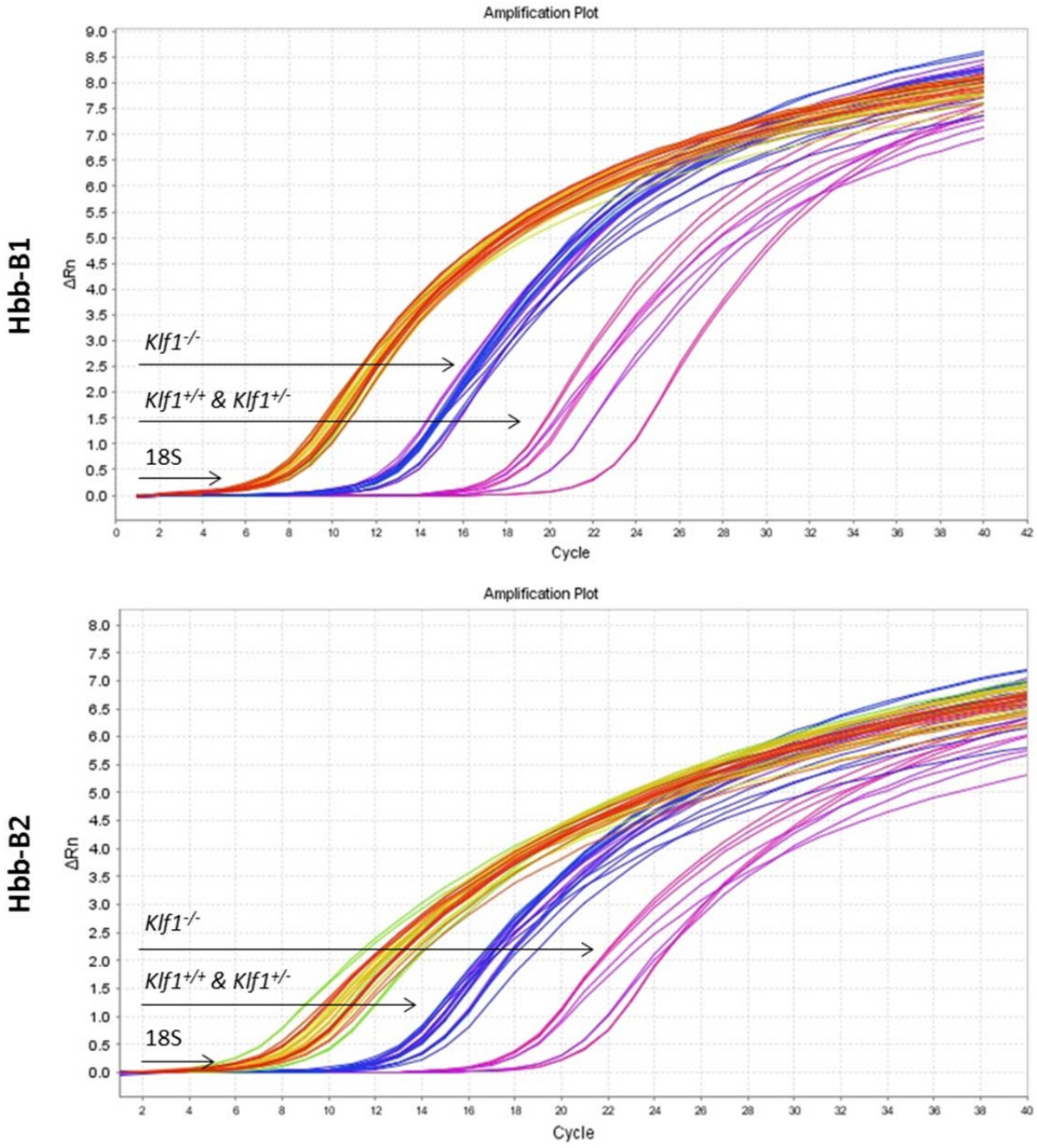


Figure II.1. Real-time quantitative PCR amplification plots of murine adult *β-globin* genes in *Klf1*$^{+/+}$, *Klf1*$^{+/-}$ and *Klf1*$^{-/-}$ E13.5 foetal livers. Amplification plots display the threshold cycle (C_T) of each gene. Whole RNA was extracted from E14.5 litter-matched *Klf1*$^{+/+}$, *Klf1*$^{+/-}$ and *Klf1*$^{-/-}$ foetal livers. Red, yellow and green curves represent 18S expression levels, purple curves represent expression levels of the adult *β-globin* genes, *Hbb-B1* and *Hbb-B2* in *Klf1*$^{+/+}$ and *Klf1*$^{+/-}$ foetal livers, and pink curves represent expression levels of *Hbb-B1* and *Hbb-B2* in *Klf1*$^{-/-}$ cells.

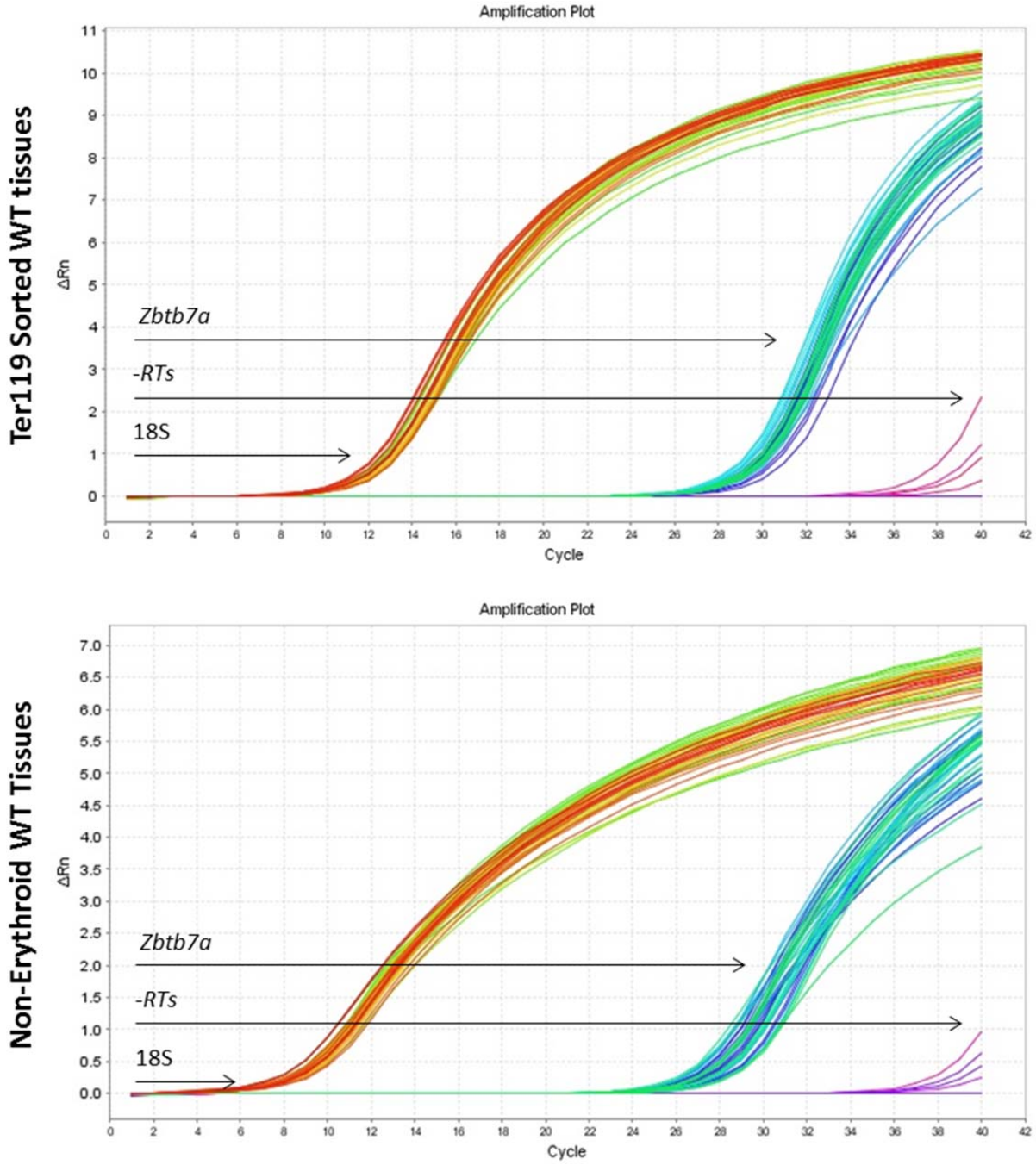


Figure II.2. Real-time quantitative PCR amplification plots of murine *Zbtb7a* in various wild type tissues. Amplification plots display the threshold cycle (C_T) of each gene. Yolk sac was harvested at E10.5, and foetal livers at E14.5. All adult tissues were harvested from mice aged between 10-12 weeks. Erythroid tissues were sorted with anti-TER119 Magnetic Microbeads. All other tissues were homogenized using a TissueLyserII. Whole RNA was extracted from the TER119+ erythroid cells, as well as the homogenized non-erythroid tissues. Red, yellow and green curves represent 18S expression levels, light and dark blue curves represent expression levels of the *Zbtb7a* in the various tissues, while the pink curves represent expression levels of *Zbtb7a* in negative RT controls.

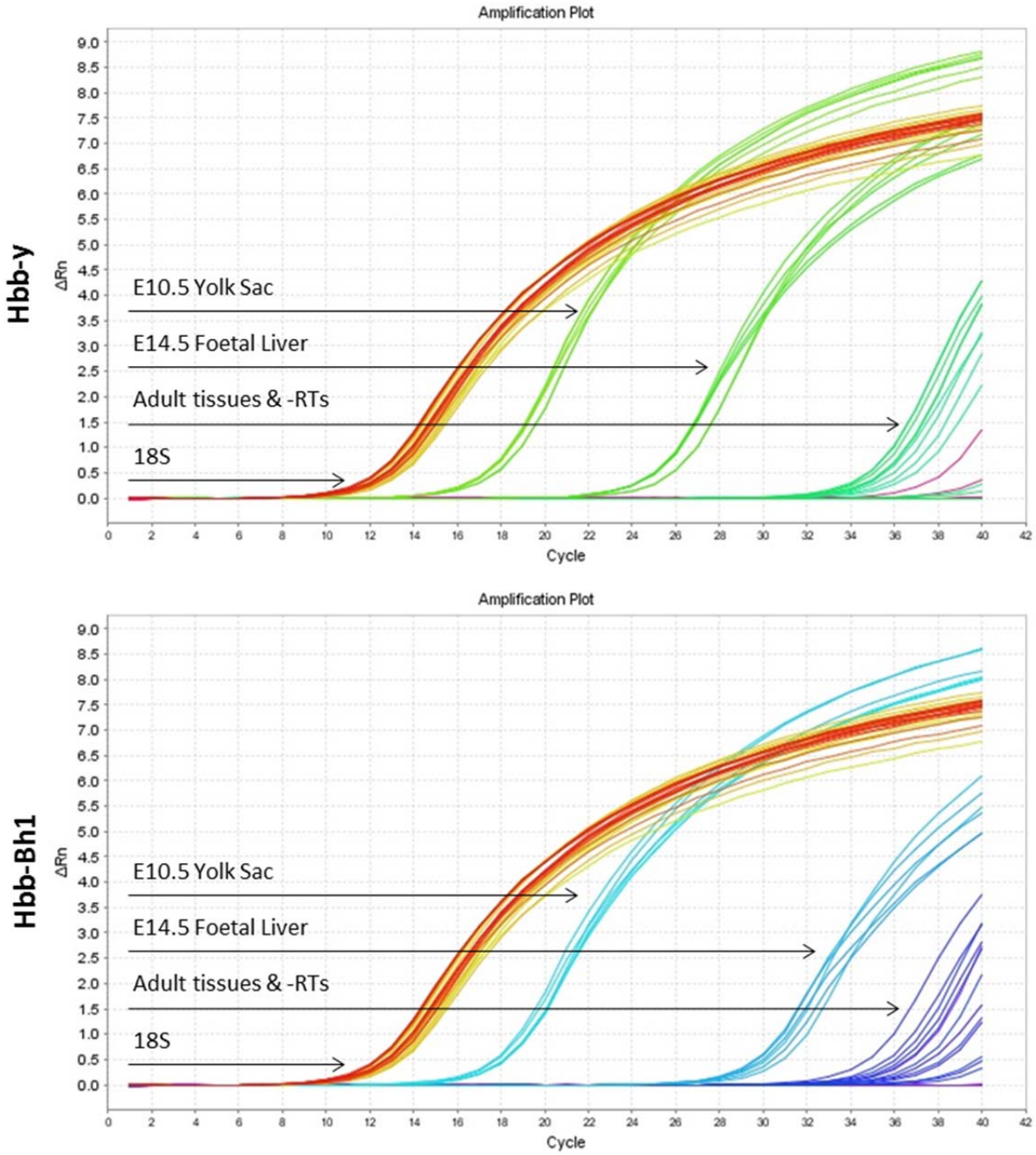


Figure II.3. Real-time quantitative PCR amplification plots of murine embryonic *β-globin* genes in various Ter119+ wild type tissues. Amplification plots display the threshold cycle (C_T) of each gene. Yolk sac was harvested at E10.5, and foetal livers at E14.5. All adult tissues were harvested from mice aged between 10-12 weeks. Erythroid tissues were sorted with anti-TER119 Magnetic Microbeads. All other tissues were homogenized using a TissueLyserII. Whole RNA was extracted from the TER119+ erythroid cells, as well as the homogenized non-erythroid tissues. Red and yellow curves represent 18S expression levels, light blue curves represent expression levels of the embryonic genes *Hbb-y* and *Hbb-Bh1* in yolk sac and foetal liver samples, while the dark blue and purple curves represent expression levels of *Hbb-y* and *Hbb-Bh1* in adult tissues and negative RT controls.

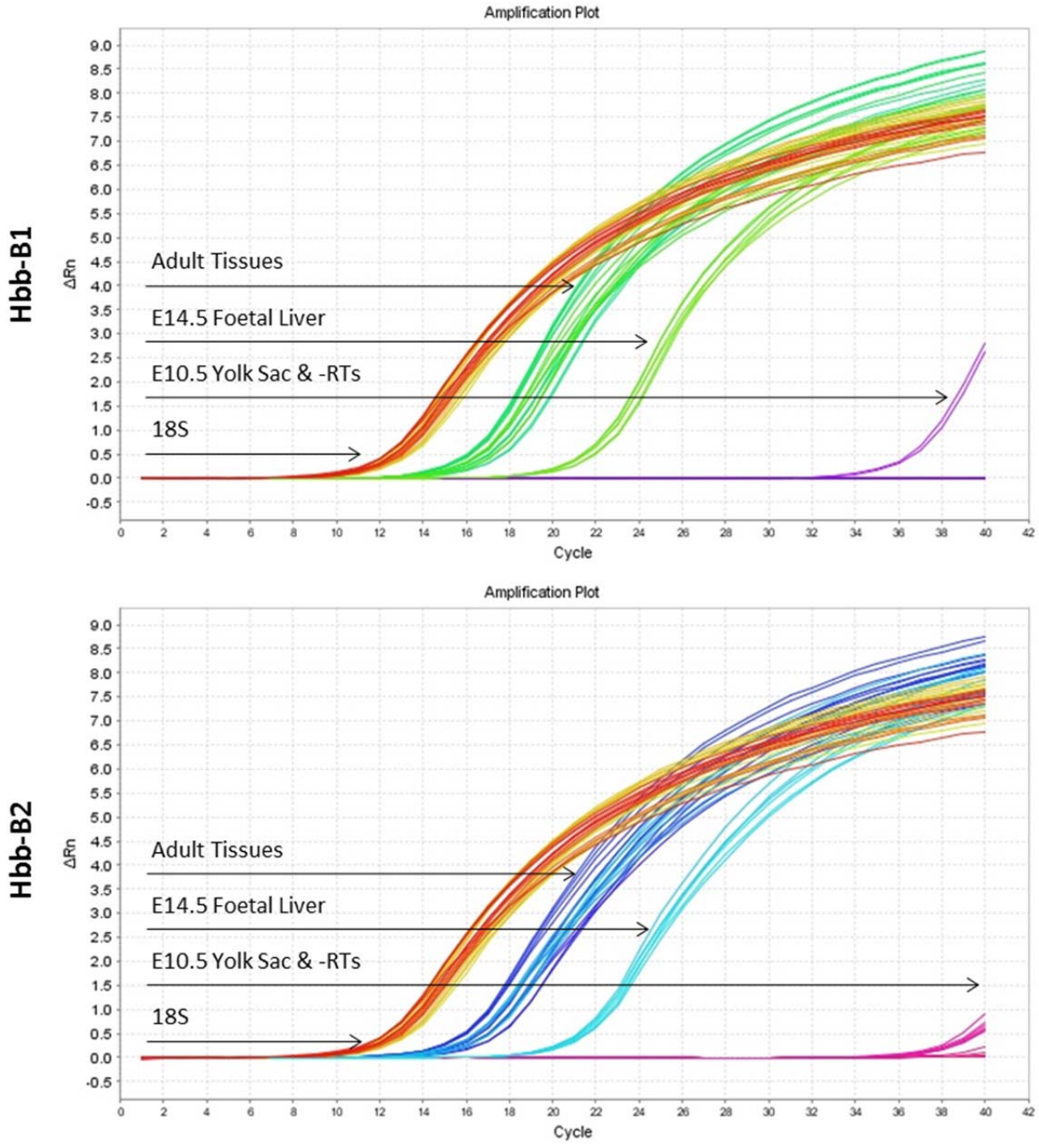


Figure II.4. Real-time quantitative PCR amplification plots of murine adult *β-globin* genes in various Ter119+ wild type tissues. Amplification plots display the threshold cycle (C_T) of each gene. Yolk sac was harvested at E10.5, and foetal livers at E14.5. All adult tissues were harvested from mice aged between 10-12 weeks. Erythroid tissues were sorted with anti-TER119 Magnetic Microbeads. All other tissues were homogenized using a TissueLyserII. Whole RNA was extracted from the TER119+ erythroid cells, as well as the homogenized non-erythroid tissues. Red and yellow curves represent 18S expression levels, dark blue curves represent expression levels of the adult globin genes *Hbb-B1* and *Hbb-B2* in adult samples, while the light blue represent expression levels in E14.5 foetal liver, and pink curves represent expression levels in E10.5 yolk sac and negative RT controls.

Appendix III

The following figures depict ZBTB7A binding to various loci in K562 and CD34+ bone marrow cells differentiated toward the erythroid lineage at two separate time points, compared to the input for each sample. The inputs for each time point of the bone marrow samples have been combined into a single track, while for K562s, each input is displayed under the corresponding sample.

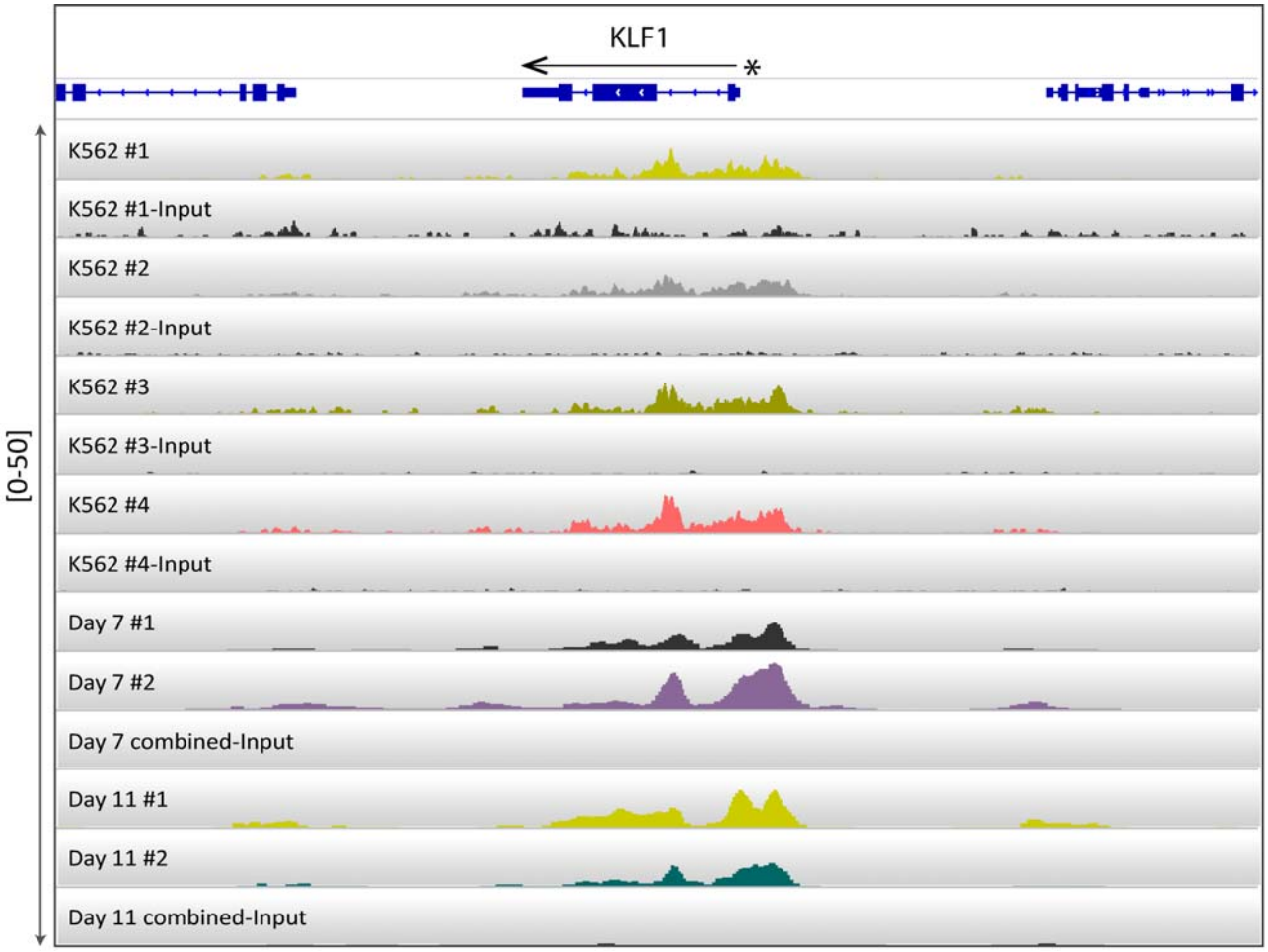


Figure III.1. ZBTB7A binding is enriched at the promoters of *KLF1* above background pulldown of input samples in human cells. ChIP-Seq peaks were viewed using IGV v2.3 and indicate that ZBTB7A binds to the promoters of both *KLF1*. This binding is seen across all K562 and bone marrow samples. The asterisks indicate the positon of the promoters and the arrows indicate the direction of translation. The enrichment of binding at this promoter above input can be seen across all samples.

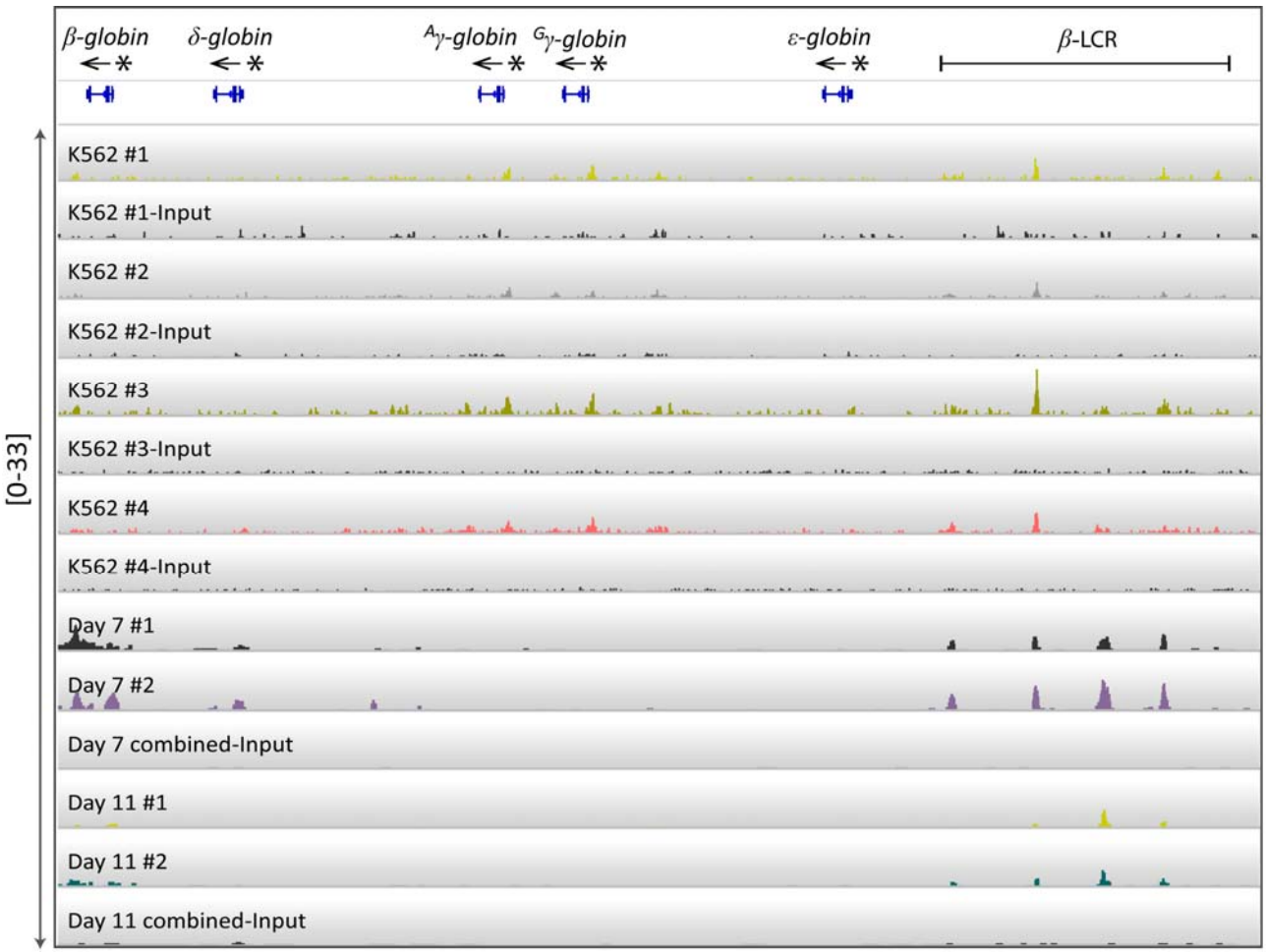


Figure III.2. ZBTB7A binding is enriched at regulatory regions of the *β*-globin locus above background pulldown of input samples. ChIP-Seq data revealed that ZBTB7A binds to the DNase Hypersensitive sites of the β-Locus Control Region (LCR) across both K562 and bone marrow samples, as well as the -200 site of the both the *γ-globin* promoters in all K562 replicates. The asterisks indicate the positon of the promoters and the arrows indicate the direction of translation. The enrichment of binding at this promoter above input can be seen across all samples.