The regulation of alpha-like globin gene expression throughout development and differentiation

A thesis submitted for the degree of
Doctor of Philosophy in Medical Sciences

Susannah R. Holliman

University College
University of Oxford

MRC Doctoral Training Program
RDM Scholars Program

Weatherall Institute of Molecular Medicine

Michaelmas Term 2024

Abstract

A key question in molecular biology is how genes are turned on and off throughout development and differentiation? Precise spatio-temporal gene regulation involves complex interplay between cis- and trans-acting regulatory factors. Cis-acting elements include enhancers, promoters, and insulators. Trans-acting factors include transcription factors (TFs), co-factors, chromatin remodelling factors, and ncRNAs. In this thesis, I use the well characterised α-globin gene locus to study the roles of the CTCF-binding site (CBS) insulator elements at the 3' of the α-globin sub-TAD in normal erythropoiesis and the transcription factor KLF1 in developmental gene switching of the α-like globin genes.

First, I characterise a mouse model and show that deletion of the 3' CBSs (θ1, θ2, HS+44, and HS+48) has a surprisingly minimal role in formation or maintenance of the α-globin sub-TAD. I find no significant changes to the enhancer interaction, chromatin accessibility, or gene expression profiles of the α-globin locus. I discuss, instead, the possibility that the α-globin genes themselves may delimit the α-globin sub-TAD.

Next, I characterise a novel KLF1$^{StrepII\text{-}FKBP}$ mouse model. A major difficulty when studying KLF1, particularly for low cell-input assays, is limited availability of suitable antibodies for chromatin immunoprecipitation (ChIP). I surprisingly find that the Strep-II tag is insufficient for robust chipmentation (low-input ChIP) of KLF1$^{StrepII\text{-}FKBP}$. I discuss the factors that might make KLF1 difficult to immunoprecipitate and suggest future optimisations to the chipmentation protocol.

Finally, I investigate the role of KLF1 in α-like globin gene switching. Some humans with KLF1 mutations upregulate embryonic ζ-globin in definitive erythroblasts. Understanding mechanisms of ζ-globin (de)repression has important therapeutic implications for treatment of severe α-thalassemia. I use

the FKBP-V degron tag to quickly and effectively degrade KLF1$^{StrepII\text{-}FKBP}$ protein in primary definitive erythroblasts to test whether KLF1 depletion recapitulates ζ-globin de-repression in mouse. I observe many of the expected cellular and gene expression effects, including modest upregulation of ζ-globin at the RNA but not protein level. I discuss some discrepancies between the human patients and mouse model that may explain this. Overall, I present the KLF1$^{StrepII\text{-}FKBP}$ model as a novel and powerful tool for future use beyond globin switching to investigate the numerous erythroid functions of KLF1.

Acknowledgements

I would like to start at the beginning and thank my grandparents, Mary and Derek, for conveniently having both a pond and a microscope. I would learn many years later that they were, in fact, both science teachers and that these “games” were science lessons in disguise. Touché.

Thank you to my past supervisors – Dr Magomet Aushev, Prof. Dawn Coverley, and Dr Angelika Feldmann – for giving a younger, shyer version of myself the confidence that perhaps I could do a PhD, too. Without your kindness, patience, and belief I would never have taken such a bet on myself.

Then come the biggies. A massive thank you to my supervisors, Doug Higgs and Mira Kassouf, for your guidance and support. Doug, your kindness, gentle leadership, and love for science are everything that could ask for in a supervisor. I feel extremely lucky to have studied under such an excellent scientist as yourself. Mira, thank you for your constant support and guidance. Particular thanks go to my mentor, Chris Babbs. Thank you, Chris, for sharing your encyclopaedic knowledge of zeta-globin with me and for answering every inane and repetitive question I had with a saintly-level of patience. Thank you, to you and to Rosie, for even putting a roof over my head when needed. I could not have completed this thesis without your support, kindness, and good humour.

I can certainly say my PhD experience would have been nowhere near as fun or bearable without the past and present members of the Higgs lab. I feel extremely lucky to have been part of a lab as supportive, friendly, and caring as ours – so thank you to every Higgs lab member. Special thanks go to those whose work laid the foundations for this project: Rob Beagrie, Caz Harrold, and Lucy Cornell. Thank you to my fellow members of the “zeta group” – Ayesha and Siyu – for your zeta-themed enthusiasm, patience, and always remembering the details that manage to escape me time and again. Particular thanks go to Ed, Seren, and Kinam for never failing to make me laugh (or to listen while I cried). Speaking of

crying, thank you to Nicole and Ali for always being there to mop me up or take us out for a little quodtail when needed.

To my housemates. Nikki, we've been together from day one and I find it hard to put into words how much you mean to me now, but I think you get it. Bei Bei, you might be the most wonderful person on planet Earth and I'm so grateful that I get to live and laugh with you every day. Emily, you are such a gentle soul with the most wicked sense of humour. I love all three of you very deeply and our little "family" means the world to me.

I would of course like to thank my family, for always looking after and supporting me. I must include in this my other set of grandparents, Carol and Derek (a different Derek, thank god), for taking us out hiking every year since I can remember in the most beautiful place on Earth. These hikes brought me some of the most peace and relief from the more stressful moments of the last four years. I hope that I've done you all proud.

My apologies for the length and gushiness of this section. It is Christmas Eve and I've had a couple of glasses of wine.

A final thank you for the support from the core facilities at the WIMM, particularly Kevin Clark at the Flow Cytometry Facility, Jackie Sloane-Stanley at the Transgenics Facility, and the Sequencing Facility. Thank you also to my Transfer and Confirmation team – Irene Thompson, Tom Milne, and Rob Beagrie – for your time, scientific guidance, and perspective.

This work was generously funded by the MRC Doctoral Training Program, the Radcliffe Department of Medicine Scholars Program, and the University College War Memorial Fund.

List of Publications

Cornell, L., Harrold, C., **Holliman, S.**, *et al*. (in review) “A functional overlap between actively transcribed genes and chromatin boundary elements”, *EMBO*

Holliman, S., *et al.* (2024) “A functional overlap between actively transcribed genes and chromatin boundary elements”. *Haemoglobin Switching Meeting*. Rio Grande, Puerto Rico 10-14th November.

Table of Contents

List of Abbreviations

3C	Chromosome conformation capture
3D	Three dimensional
Aa	Amino acids
AGM	Aorta-gonad-mesonephros
APH	Acetylphenylhydrazine
ATAC-seq	Assay for transposase-accessible chromatin with next generation sequencing
BFU-E	Burst forming unit erythroid
BHFS	Hb Bart's Hydrops Fetalis Syndrome
BSA	Bovine Serum Albumin
CBS	CTCF binding site
CDA	Congenital dyserythropoietic anaemia
CFU-E	Colony forming unit erythroid
ChIP-seq	Chromatin immunoprecipitation with next generation sequencing
CMEI	Central macrophages of erythroblastic islands
CMP	Common myeloid progenitor
Ct	cycle threshold
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dsDNA	double stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene-bis(oxyethylenenitrilo)tetraacetic acid
EMP	erythro-myeloid progenitor
eRNA	enhancer RNA
ESC	Embryonic stem cell
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
gDNA	genomic DNA
Hb	Haemoglobin
Het	Heterozygous
HGB	Haemoglobin (level, %)
Hom	Homozygous
HPEH	Hereditary persistance of embryonic haemoglobin
HPFH	Hereditary persistance of fetal haemoglobin
HSC	Haematopoietic stem cell

HSPC	Haematopoietic stem and progenitor cell
IEF	Isoelectric focussing
LYM	Lymphocyte
MCC	Micro-Capture-C
MCH	Mean corpuscular haemoglobin
MCV	Mean corpuscular volume
MEL	Mouse erythroleukaemia (cell line)
MEP	Megakaryocyte and erythrocyte progenitor
mESC	mouse embryonic stem cell
mFL	mouse Fetal Liver
MON	Monocyte
MPP	Multipotent progenitors
Nan	Neonatal anaemia (mouse model)
ncRNA	non-coding RNA
NGS	Next generation sequencing
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PIC	Protease inhibitor cocktail
POI	Protein of interest
PTM	Post-translational modification
qPCR	quantitative PCR
qRT-PCR	quantitative real time PCR
RBC	Red blood cell
Retic	Reticulocyte
REX	Range extender elements
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing
SDS	Sodium dodecyl sulfate
sgRNA	single guide RNA
shRNA	short hairpin RNA
SNR	Signal:noise ratio
STED	Stimulated emission depletion (microscopy)
TAD	Topologically associating domains
TAE	Tris-acetate-EDTA
TCA	Trichloroacetic acid
TF	Transcription factor
TSS	Transcription start site
UTR	Untranslated region
WBC	White blood cell
WIMM	Weatherall Institute of Molecular Medicine
WT	Wildtype
YFP	Yellow fluorescent protein

ZnF	Zinc finger
HbH	Haemoglobin H disease
HDAC	Histone deacetylase
mRNA	messenger RNA
IP	Immunoprecipitation
ACH	Active chromatin hub
LCR	Locus control region
ZPE	ζ-promoter edited
ddPCR	digital droplet PCR
FISH	Fluorescent *in situ* hybridisation assay
CpG	5'-Cytosine-Guanine-3' dinucleotide

Chapter 1 – Introduction

How are mammalian genes switched on and off during development and differentiation? This is the question at the heart of modern molecular genetics. Tackling this is central to understanding how these processes are perturbed in genetic disease and to develop new therapeutic strategies for such disorders.

There are many levels at which gene expression is regulated. This includes the dynamics of RNA transcription, RNA processing and stability, the rate of translation to protein, and the post-translational modification of a protein to alter its function or half-life. The most fundamental of these control points, however, is the decision whether to transcribe a gene or not.

Broadly speaking, gene transcription is mediated by cis- and trans-acting regulatory factors. Cis-acting factors are DNA elements that are involved in initiating or modifying transcription. They include gene promoters, enhancers, and insulator elements. Trans-acting factors refer to gene products – RNA or proteins – which bind to cis-regulatory elements to alter gene expression. They predominantly include transcription factors (TF) and co-factors (that alter the activity of TFs and give rise to epigenetic modification of TFs and chromatin).

In the Higgs Laboratory, we study the function and interplay of these regulatory factors in red blood cells, specifically in the regulation of α-like globin gene expression throughout erythroid differentiation and development. This is a powerful model with which to study principles of gene regulation. First, haematopoiesis and erythropoiesis are among the most well characterised cell systems, due in large part to their relative abundance and ease of access. Second, erythropoiesis is a terminal differentiation pathway and altering globin expression within this does not affect cell fate. Lastly, the α-globin locus is extremely well characterised. We have identified all cis-acting regulatory elements within this locus and a comprehensive suite of general and erythroid-specific trans-acting factors that act here. We have demonstrated the contribution

of these cis- and trans-acting elements to human disease, including α-thalassemia, ATRX syndrome, and Hb Bart's hydrops fetalis syndrome (BHFS). Our current work aims to build a comprehensive view of how these regulatory factors interact to fine-tune gene expression. In this thesis, I investigate the roles of CTCF insulator elements and the trans-acting TF, KLF1, in erythroid differentiation and development.

1.1 – Haematopoiesis

Haematopoiesis involves the production, specification, and differentiation of all blood cell lineages. This process is broadly conserved across vertebrates (**Orkin and Zon, 2008**) and occurs in three major developmental waves.

In mouse, the first wave of haematopoiesis arises around E7.5 from primitive erythroid progenitors (EryP-CFC) in the yolk sac blood islands. This gives rise to primitive erythroid cells which circulate through the embryo as a nucleated, semi-synchronous cohort of differentiating erythroblasts (**Fig. 1.1; Palis *et al.*, 1999**). Primitive erythroblasts predominate between E9.5-E12.5, although they are detectable in the peripheral blood until shortly after birth (**Palis *et al.*, 1999; Kingsley *et al.*, 2004**).

At E8.25, the second transient haematopoietic wave emerges from a separate population of erythro-myeloid progenitors (EMPs) in the yolk sac (**Palis *et al.*, 1999**). The multi-potential EMPs produce definitive erythroid cells, which are smaller than primitive cells and enucleate before entering the circulation. By E10.5, EMPs have colonised the fetal liver, from which they contribute erythrocytes until at least E16.5 (**Fig. 1.1**; **Palis *et al.*, 1999**).

The third wave of haematopoiesis is derived from haematopoietic stem cells (HSCs). These first emerge from the aorta-gonad-mesonephros (AGM) at ~E10.5 (**de Bruijn *et al.*, 2000**) and subsequently migrate to the fetal liver from ~E11.5 and then to the bone marrow around birth (**Fig. 1.1**; **Palis, 2014**). The mFL is the predominant source of erythropoiesis from ~E12.5-birth. The bone marrow is the

primary site of haematopoiesis in post-natal life, except in times of erythropoietic stress when extramedullary haematopoiesis in the spleen occurs.

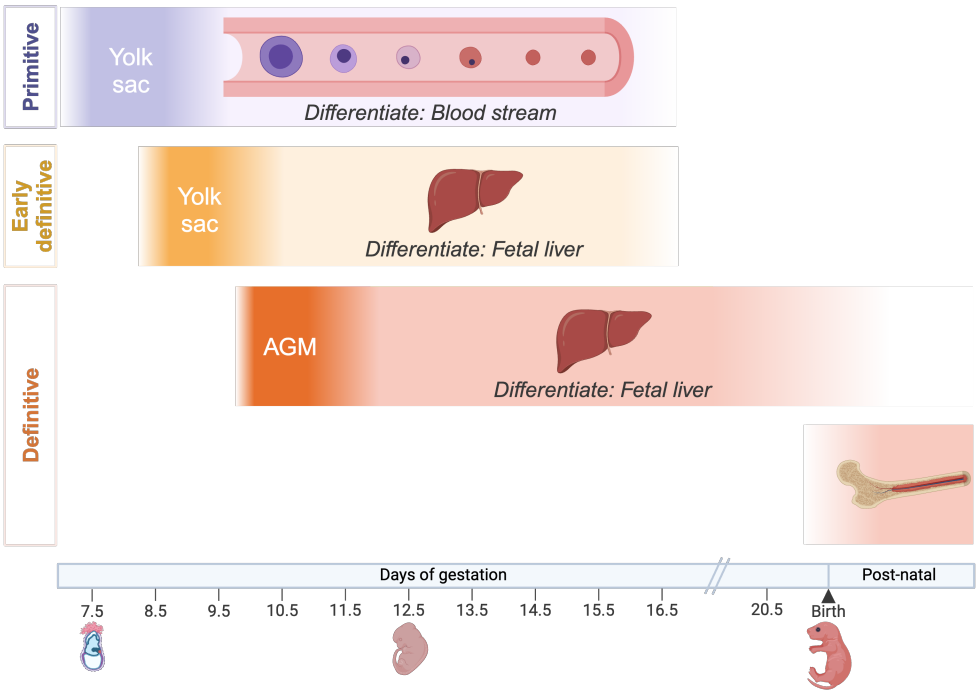

Fig. 1.1 – *Haematopoiesis in mouse development*. Primitive haematopoiesis arises from EryP-CFC cells in the embryonic yolk sac and differentiates in the blood stream. Early definitive haematopoiesis arises from EMP cells in the yolk sac and differentiates in the fetal liver. Definitive erythropoiesis arises from HSCs in the AGM, which migrate to the fetal liver and later to the bone marrow around the time of birth. Accordingly, definitive erythroid cells differentiate in the fetal liver and then in the post-natal bone marrow.

1.2 – Erythropoiesis

Erythropoiesis is the differentiation of mature erythrocytes from haematopoietic stem and progenitor cells (HSPCs). Primitive erythroid cells are larger and differentiate and enucleate as a semi-synchronous cohort in the embryonic blood stream. Definitive erythroid cells, however, differentiate in the fetal liver and adult bone marrow, and are ordinarily only released into the blood stream after enucleation. Both share the same broad stages of erythropoiesis (**Fig. 1.2; Palis *et al.*, 2014**). This thesis will predominantly focus on definitive erythropoiesis.

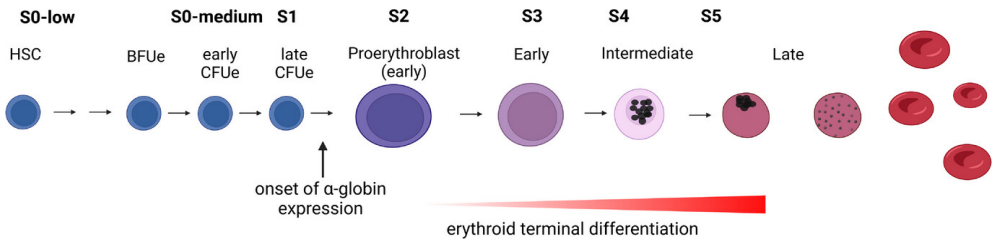

Fig. 1.2 – *Erythroid differentiation.* Reproduced with permission from **Kassouf *et al.*, 2023**. Original legend: "Schematic representation of erythropoiesis. S0-low to S5 mark seven defined cellular stages of erythropoiesis representing an immunophenotyping-based purification strategy [(**Pop *et al.*, 2010; Chapter 4, Fig 4.3**)] that allows the isolation of the desired population from mouse fetal livers. The stages of erythroid differentiation are shown starting with Hematopoietic Stem Cells (HSC) committing, amongst other lineages, to the erythroid progenitors (burst colony forming unit and colony forming unit erythroid, BFUe and CFUe respectively). α-globin expression is first detected and progressively increases as cells differentiate from proerythroblasts to mature red blood cells, as indicated by the gradually increasing red colour in the red triangle. The schematic representation of the progressively maturing erythroid cells highlight the morphologically distinguishable stages both in the varying size and hemoglobinisation states of the cells."

1.3 – Globin gene expression and haemoglobin switching

Functional haemoglobin protein is a heterotetramer, composed of two α-like and two β-like globin subunits. For both α-like and β-like globins, different paralogous globin genes are expressed at different developmental stages. The α-like and β-like globin gene loci are encoded on separate chromosomes (**Fig. 1.3**), but within each locus the genes are positioned in the same order as their order of developmental expression, with the embryonic globins closest to the enhancer elements.

In humans, there are five β-like globin genes. An embryonic ε-globin, two fetal γ-globins, a minor adult δ-globin, and a major adult β-globin (**Fig. 1.3A**). Mouse β-like globins are similar, with two embryonic genes, ε_Y-globin and β_{h1}-globin, and two adult β-globins. β_{h1}-globin may be considered the closer orthologue to human γ-globin, due to their similar expression profiles across the three developmental waves of erythropoiesis (primitive, EMP-derived early definitive, and HSC-derived late definitive) when compared within humanised mouse models (**McGrath *et al.*, 2011**), but ultimately there is no exact γ-globin orthologue in

mouse. Consequently, most studies of γ-globin have used humanised mouse models and transgenes. At the α-globin locus, in human and mouse, there is one embryonic ζ-globin gene and two duplicated adult α-globin genes. For the α- and β-globin loci, globin gene expression is regulated by a cluster of erythroid-specific enhancer-like elements lying upstream of the genes.

The switching of globin gene expression from embryonic to fetal/adult and fetal to adult is referred to as haemoglobin switching. The α- and β-globin switching events occur at different developmental stages (**Fig. 1.3**) and are mostly independently controlled.

For the α-like globins, there are two ζ- to α-globin switching events. One is considered the intra-lineage switch (occurring within the primitive erythroid lineage) and the other the inter-lineage switch (the switch from ζ-expressing primitive to ζ-repressing definitive erythroid lineages). The intra-lineage switch occurs within terminally differentiating primitive erythroid cells. In these primitive erythroblasts, ζ-globin is initially expressed, but that expression gradually decreases as α-globin transcription increases. Conversely, in definitive erythroid cells, ζ-globin is always silenced and only α-globin is expressed. Therefore, the transition from primitive to definitive erythropoiesis is considered to be the α-like inter-lineage switch. The net effect of this in normal development is the expression of ζ-globin until ~E8.5-10.5 in mouse and ~7 weeks in human embryos (**Russel *et al.*, 1998**). Understanding the mechanism of ζ-globin repression in definitive erythroblasts has the most clinical relevancy for potential therapeutic de-repression in patients with severe α-thalassemia (**Section 1.5**).

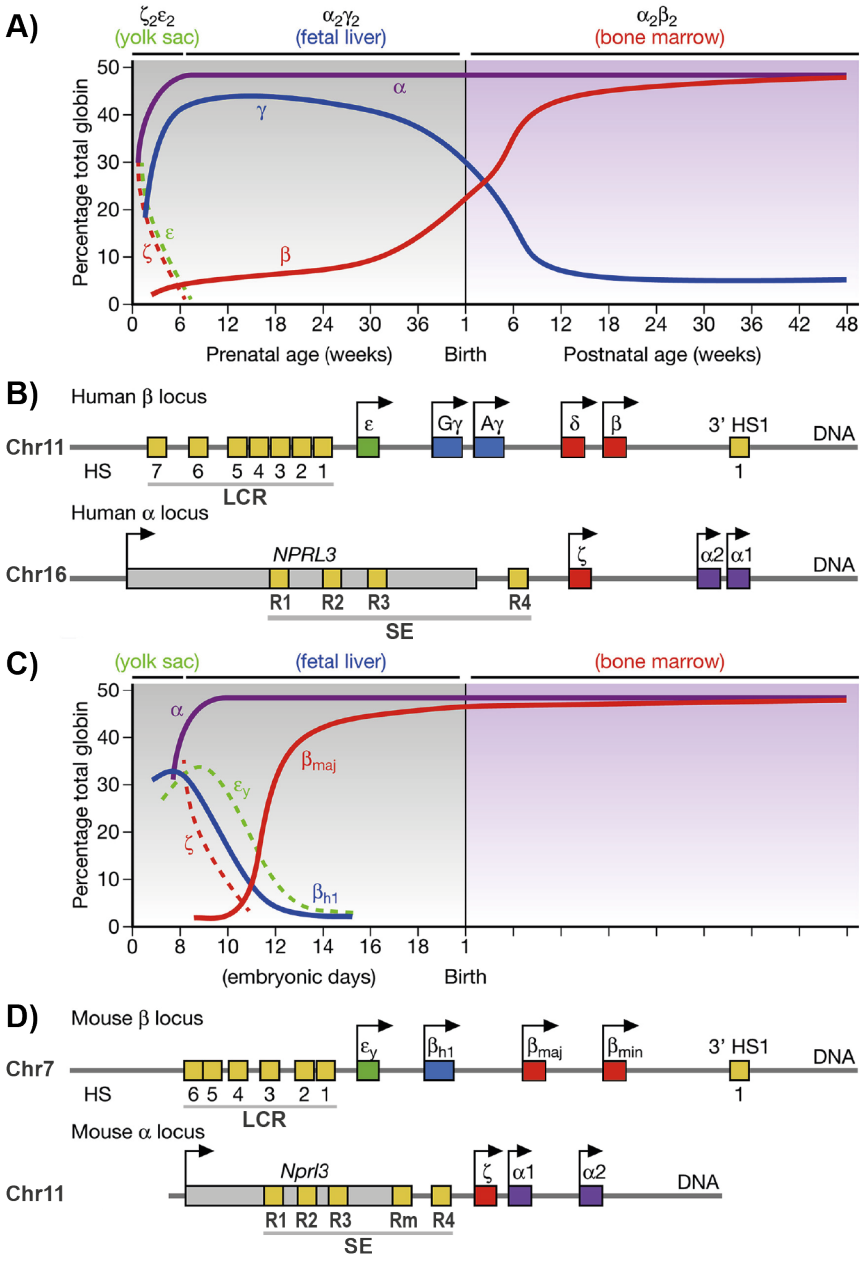

Fig. 1.3 – *Haemoglobin switching in development.* Adapted from **Blobel *et al.*, 2015**. Relative expression of each globin gene during in human **(A)** and mouse **(C)** development, with the corresponding α-globin and β-globin gene loci illustrated below (**B + D**, not to scale). Enhancer-like elements annotated in yellow: "LCR" = locus control region, "SE" = superenhancer.

1.4 – The α-globin gene locus

The mouse α-globin gene locus contains an embryonic ζ-globin gene (*Hba-x*), two duplicated adult α-globin genes (*Hba-a1/2*), and two θ-globin pseudogenes (*Hbq1a/b* (aka *Hbq-1/2*); lowly transcribed, non-coding, α-like homologues; **Fig. 1.4A**). Upstream of this are five enhancer-like elements – R1, R2, R3, Rm, and R4 – which fulfil the definition of a superenhancer and contain binding sites for erythroid TFs, including KLF1, GATA1, NFE2, and TAL1 (**Whyte *et al.*, 2013; Hay *et al.*, 2016**). R1 and R2 act as strong, canonical enhancers elements, whereas R3, Rm, and R4 act as facilitators. These facilitators lack significant enhancer activity in isolation but potentiate the activity of R1 and R2 in a position-dependent manner (**Blayney *et al.*, 2023**).

Together, these elements are located within a ~70 kb, erythroid-specific sub-TAD within a larger, cell type-invariant 165 kb TAD (**Oudelaar *et al.*, 2020**), which are flanked by broadly convergently orientated CTCF bound sites (CBSs) (**Hanssen *et al.*, 2017**). TADs are domains of preferentially self-interacting chromatin that typically emerge from CTCF-delimited cohesin loop extrusion (**Fundenberg *et al.*, 2016; explored in Chapter 3**). With the exception of Rm, the above cis-acting elements are largely conserved between mouse and human (**Hughes *et al.*, 2005**).

In definitive erythroblasts, ζ-globin is hypoacetylated, not expressed, and doesn't contact the rest of the sub-TAD (**Fig. 1.4A; Hua *et al.*, 2021; King *et al.*, 2021**). Even in primitive erythropoiesis, ζ-globin is less reliant on the enhancers than the α-genes are. This independence might be due to differences in the α- and ζ-globin promoters. The ζ-globin promoter is longer, more highly evolutionarily conserved, and contains more TF binding motifs than the α-globin promoters (**Hughes *et al.*, 2005; King *et al.*, 2021**).

It has been shown by Capture-C (**Davies *et al.*, 2016, Oudelaar *et al.*, 2020; Fig. 1.4A**), MCC (**Hua *et al.*, 2021**), and super-resolution imaging (**Brown *et al.*, 2018; Fig. 1.4B**) that the α-globin enhancers and promoters come into close proximity within the sub-TAD domain and that these interactions correlate with the onset of α-globin transcription (**Oudelaar *et al.*, 2020**). It is hypothesised that

these interactions manifest as a highly dynamic, 3D transcription hub that could recruit high concentrations of activating trans-acting factors to facilitate *Hba-a1/2* gene transcription (**Fig. 1.4C**).

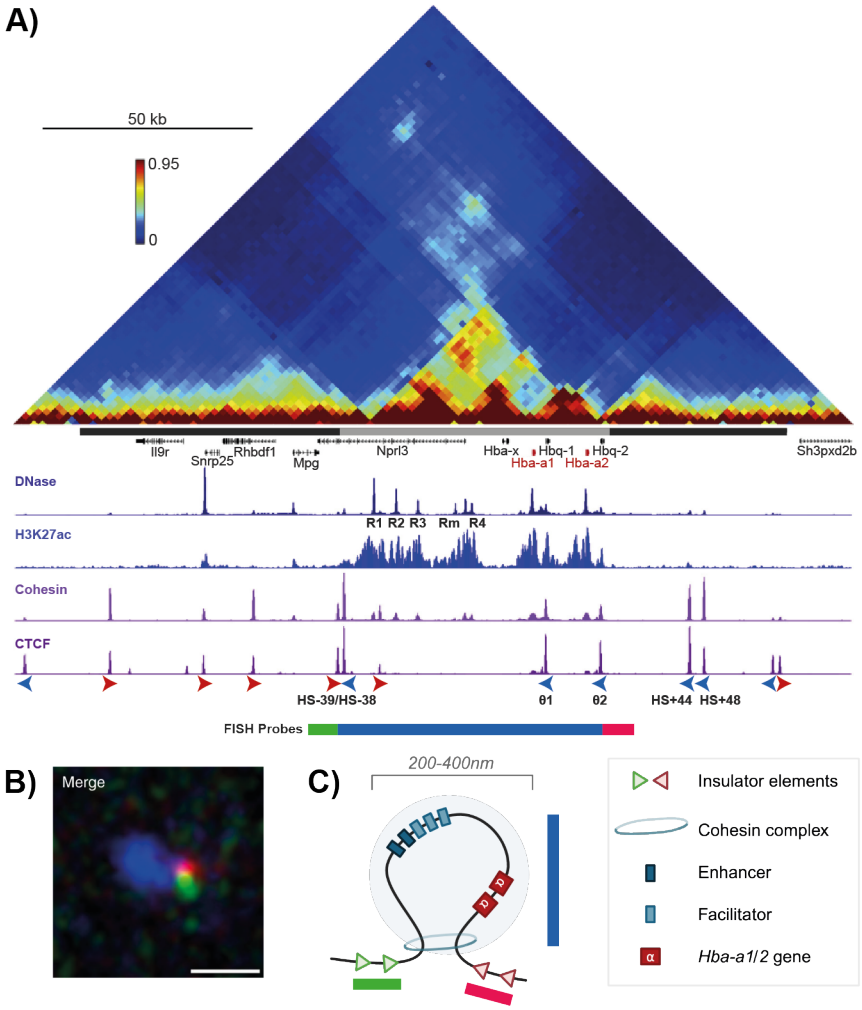

Fig 1.4 – *The α-globin gene locus in definitive erythroblasts*. A) Adapted from **Oudelaar *et al.*, 2021**. Tiled-C map of interaction probabilities across the α-globin TAD (black bar) and nested sub-TAD (grey bar). Below are the gene annotation, DNase hypersensitivity assay (chromatin accessibility), H3K27ac ChIP (active chromatin mark), cohesin ChIP, and CTCF ChIP tracks. The orientation of the CTCF binding sites (CBSs) are annotated with red (forward) and blue (reverse) arrows. **B)** Super-resolution (STED) imaging of the α-globin locus using DNA fluorescent *in situ* hybridisation (DNA FISH) shows the flanks of the sub-TAD come into close physical proximity. Scale bar = 0.5 μm. (**Brown *et al.*, 2018**). **C)** A simple proposed model of the α-globin transcription hub. Cohesin loop extrusion may bring the CBS that flank the sub-TAD into close physical proximity (green and red FISH probes). The α-globin enhancers and promoters may interact within the resulting transcription hub domain (blue FISH probes).

1.5 – α-thalassemia

α-thalassemia is one of the most common monogenic disorders worldwide. It is a hereditary anaemia caused by loss of α-globin expression from between one to all of the α-globin genes. Disease severity increases with the number of mutant α-globin alleles inherited, from asymptomatic thalassemia trait to lethal BHFS (**Table 1.1; Harteveld and Higgs, 2010; Piel and Weatherall, 2014**).

Table 1.1 – ***The spectrum of α-thalassemia disease genotypes and associated phenotypes*****.** Disease severity increases with loss of functional α-globin gene expression.

α locus genotype	Phenotype	Predominant adult Hb	Symptoms
(αα/αα)	Normal	$\alpha_2\beta_2$	Asymptomatic
(αα/α-)	Silent carrier	$\alpha_2\beta_2$	Asymptomatic
(αα/--) (α-/α-)	Thalassemia trait	$\alpha_2\beta_2$	Asymptomatic
(α-/--)	HbH disease	$\beta_2\beta_2$ $\alpha_2\beta_2$	Variable anaemia. May include splenomegaly, jaundice, and acute episodes of haemolysis
(--/--)	Bart's Hydrops Fetalis Syndrome (BHFS)	$\beta_2\beta_2$ $\zeta_2\beta_2$	Severe fetal hypoxia; high risk neonatal death and maternal complications

α-thalassemia is most commonly caused by deletion of one or both α-globin genes from the chromosome. The most common of these deletion alleles is the $--^{SEA}$ allele, which leaves the ζ-globin gene intact (**Fig. 1.5; Amid *et al.*, 2024**). In these patients, de-repression of endogenous ζ-globin could be an effective therapeutic strategy to ameliorate symptoms of severe α-thalassemia.

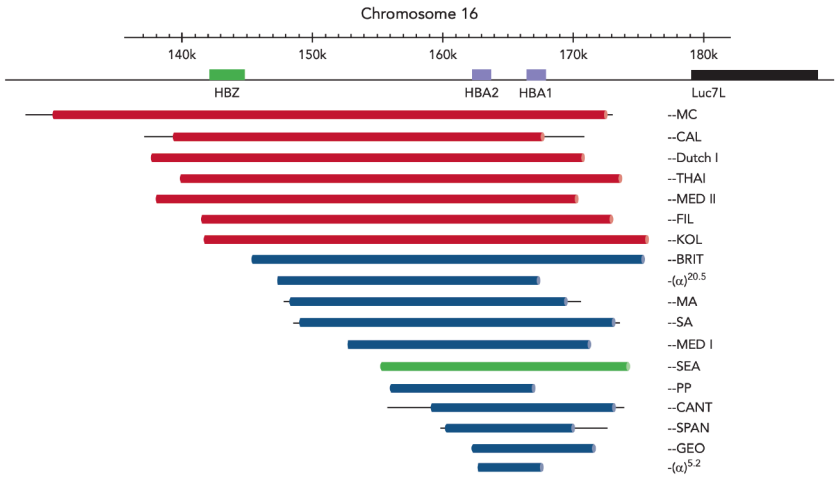

Fig. 1.5 – *The --^{SEA} deletion leaves the ζ-globin gene intact.* Reproduced with permission from **Amid *et al.*, 2024**. Original legend: "Deletions of 2 α-genes giving rise to α^0-thalassemia. Deletions that involve the ζ-globin gene (HBZ) are depicted in red. By far the most common α^0-thalassemia deletion involved in Hb Bart's hydrops fetalis ($--^{SEA}$ deletion) is depicted in green. Bold lines indicate sequences known to be deleted. Thin lines indicate regions of uncertainty around the breakpoints. Reproduced with permission, with minor modifications, from Harteveld and Higgs [2010].".

1.6 – Cis-acting regulatory elements

Cis-acting factors are DNA elements that are involved in initiating or modifying transcription. There are three main types of these.

Promoters

"Promoter" typically refers to the transcription start site (TSS) of a gene (where Pol II and general TFs bind to form the pre-initiation complex (PIC)) and the proximal upstream sequence to this, which may contain binding sites for other TFs or chromatin modifying complexes that influence PIC assembly, transcription initiation, and transcription elongation (**Andersson and Sandelin, 2020**). Most importantly, the promoter is the site of PIC assembly and transcription initiation. For widely expressed genes, such as housekeeping genes, the promoter is usually sufficient to regulate normal gene expression. For tissue-specific or developmental-stage specific genes, however, long-range enhancer elements

are often required to ensure correct spatio-temporal expression (**Bower and Kvon, 2025**).

Enhancers and facilitators

Enhancers are not defined by their sequence but rather their function: enhancers act at a distance to upregulate transcription from a gene promoter, irrespective of the distance or orientation from the transcription start site. There are an estimated ~900,000 enhancers in the mammalian genome (**ENCODE *et al.*, 2020**). Enhancers might be marked by common signatures, such as the presence of certain TF motifs, histone marks (e.g. H3K27ac, H3K4me1), or enhancer RNA (eRNA) transcription, but none of these have been found to be universal (**Halfon, 2019**).

Facilitator and range extender (REX) elements are new sub-classes of element that lack intrinsic enhancer activity but act to support enhancer-promoter interaction and/or potentiate transcription (**Blayney *et al.*, 2023; Bower *et al.*, 2024**).

Insulators

An insulator refers to an element which can reduce gene expression when placed between an enhancer and its cognate gene promoter. Typically, this is thought to involve blocking cohesin loop extrusion. CBSs are the most abundant and canonical insulator element, reviewed further in **Chapter 3**. However, CBSs are not the only possible type of insulator. Translocation of cohesin may be stalled by other protein complexes such as transcription and replication complexes (**Bonev *et al.*, 2017; Chahar *et al.*, 2022; Jeppsson *et al.*, 2022; Banigan *et al.*, 2023; H. Zhang *et al.*, 2023; S. Zhang *et al.*, 2023**).

There is emerging evidence for functional overlap between the cis-acting elements. There are several examples of enhancers acting as promoters, promoters acting as enhancers, and transcribing promoters acting as insulators (**Andersson and Sandelin, 2020; S. Zhang *et al.*, 2023; Cornell *et al.*, manuscript in preparation**). Nonetheless these classifications remain useful;

having an agreed “syntax” of the genome aids collective research into the functions of the non-coding genome.

1.7 – Trans-acting factors

Trans-acting factors refer to gene products which bind to cis-regulatory elements to alter gene expression. These include TFs, co-factors, chromatin remodelling factors, and ncRNAs. Five key TF examples relevant to this thesis are outlined below.

KLF1

KLF1 (EKLF) is an erythroid-specific, master TF. In both mouse and human, the KLF1 gene spans ~3 kb, encoding three exons and two introns. This translates to a 358 aa (37.8 kDa) protein in mouse or 362 aa (38.22 kDa) protein in humans, with distinct functional domains (**Fig. 1.6**). The N-terminal transactivation domains (TADs) are largely, but not exclusively, the site of interactions with select transcription initiation factors and chromatin modifiers. Meanwhile, the C-terminal zinc finger (ZnF) domains mediate KLF1 binding to its CCM-CRC-CCN motif, supposedly regulating interactions at >1000 gene promoters and enhancers (**Miller and Bieker, 1993; Feng, Southwood and Bieker, 1994; Tallack *et al.*, 2010; Huang *et al.*, 2024**). Both termini are subject to post-translational modifications that modulate KLF1 function. This includes phosphorylation, ubiquitination, acetylation, and sumoylation (**Ouyang, Chen and Bieker, 1998; Quadrini and Bieker, 2006; Zhang *et al.*, 2001; Siatecka, Xue and Bieker, 2007;** reviewed in **Yien and Bieker, 2013)**. Mutations in both regions have been implicated in various forms of hereditary anaemia.

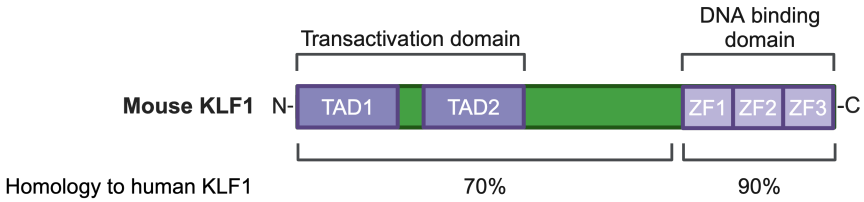

Fig. 1.6 – *The structure of KLF1*. Schematic of the mouse KLF1 protein structure, not to scale. KLF1 protein contains: two N-terminal transactivation domains (TAD1-2; dark

purple); three C-terminal zinc finger DNA-binding domains (ZnF1-3; purple); a disordered intervening sequence (dark green). The % homology to the human KLF1 amino acid sequence is annotated below.

KLF1 has many critical erythroid TF functions. KLF1 is reported to bind hundreds of erythroid genes and can recruit transcription initiation factors, including TFIIH and TAF9 (**Mas *et al.*, 2011; Sengupta *et al.*, 2009**), and promote transcription through interaction and recruitment of chromatin modifying complexes, P300/CBP and SWI/SNF (**Armstrong *et al.*, 1998; Zhang *et al.*, 2001**). These transcription activating functions are reflected in gene expression assays in KLF1$^{-/-}$ mice, in which 690 genes were downregulated in the absence of KLF1 (**Tallack *et al.*, 2012**). The same study also found 118 genes to be upregulated, reflecting the possible transcription repressor functions of KLF1 via interactions with NuRD (a histone deacetylase) complex (**Siatecka *et al.*, 2007**) and PIAS family proteins (**Siatecka *et al.*, 2015**).

GATA1

Like KLF1, GATA1 is a master erythroid ZnF TF with transcription activator and repressor functions. GATA1 is essential for erythropoiesis but, unlike KLF1, GATA1 also has roles in non-erythroid haematopoietic lineages (**Barbarini *et al.*, 2019**). In early erythropoiesis GATA1 is upregulated and in late erythropoiesis is silenced (**Gutiérrez *et al.*, 2019**; **Appendix Fig. A4.3**).

Notably, there is a close association between KLF1 and GATA1. Not only does GATA1 bind the *Klf1* promoter to upregulate *Klf1* transcription (**Crossley *et al.*, 1994**), but closely spaced GATA/CACC motifs can be found as an erythroid module in many erythroid promoters and distal enhancers, such as the β-globin LCR (**Philipsen, Pruzina and Grosveld, 1993; Tallack *et al.*, 2010**). As such, the two TFs can act cooperatively to regulate many of their target genes (**Gregory *et al.*, 1996; Tallack and Perkins, 2010**). For example, KLF1 can interact with the GATA1-containing erythroid pentameric complex (GATA1/TAL1/LDB1/LMO2/CBFA2T3) (**Bieker and Phillipsen 2024**).

BCL11A & LRF

KLF1 is also known to directly upregulate two transcriptional repressors, BCL11A and LRF (**Zhou *et al.*, 2010; Norton *et al.*, 2017**).

Bcl11a encodes a Krüppel-like C2H2 ZnF TF. As well as erythroid roles in fetal globin regulation, BCL11A is also highly expressed in the developing mammalian brain and is implicated in several neurodevelopmental disorders (**Peron *et al.*, 2024**). Crucially, the *Bcl11a* gene has an erythroid-specific enhancer which may allow specific targeting for therapeutic reactivation of fetal haemoglobin (**Bauer *et al.*, 2013**).

LRF is a ZnF TF encoded by the *Zbtb7a* (*Lrf*) gene and expressed in several other tissues including the brain, thymus, and adipocytes. Both KLF1 and GATA1 can bind and regulate expression of *Lrf*. LRF in turn can antagonistically bind near to GATA1 motif sites and recruit the chromatin repressive complex Polycomb Repressive Complex 2 (PRC2) to mediate gene silencing (**Constantinou *et al.*, 2019**).

BCL11A and LRF are both known to directly repress embryonic and fetal globin genes (**Liu *et al.*, 2018; Martyn *et al.*, 2018; King *et al.*, 2021; S. Liu, unpublished**). This is thought to be mediated at least in part by recruitment of the repressive, histone deacetylase NuRD complex (**Sankaran *et al.*, 2008; Masuda *et al.*, 2016**).

Trans-acting regulatory factors clearly form complex gene regulatory networks, often with multiple feedback loops, functional redundancies, and large numbers of target genes. Chapter 4 of this thesis describes some of the challenges of studying KLF1 – principally poor antibody availability for ChIP-seq – and describes a novel mouse model to overcome these. Chapter 5 of this thesis outlines evidence for the action of KLF1 in the KLF1-BCL11A/LRF-ζ-globin pathway and seeks to test this further with acute depletion of KLF1 in definitive erythroblasts.

1.8 – Thesis aims

In this thesis, I specifically investigate two of the above regulatory factors – CBSs and KLF1 – at the α-globin locus, to answer the following three questions:

1. To what extent does the cis-regulatory factor, CTCF-binding sites, delimit the 3' boundary of the mouse α-globin sub-TAD?

2. How does the DNA binding profile of the transcription factor, KLF1, change throughout terminal definitive erythropoiesis?

3. Does acute depletion of KLF1 in a mouse model derepress ζ-globin expression in definitive erythroblasts?

Chapter 2 – Materials and Methods

2.1 – List of common reagents

BSA (Sigma, A3059-50G)

cOmplete protease inhibitor cocktail (PIC) (Roche, 11873580001)

DMSO (Sigma, D2650-5X5ML)

EDTA (Invitrogen, AM9260G)

EGTA (Alpha Aesar, J60767)

Ethanol (Supelco, 1009832511)

FBS (Sigma, F7524-500ML)

Formaldehyde (Sigma, 47608-250ML)

Glycine (Sigma, G1726)

Hoechst 33258 (Invitrogen, H3569)

NaCl (Invitrogen, AM9760G)

NaOAc (Invitrogen, AM9740)

N,N-Dimethylformamide (Sigma, 227056-100ML)

Nuclease-free H_2O (Sigma, W4502-1L)

Methanol (Sigma, 32213)

$MgCl_2$ (Invitrogen, AM9530G)

PBS (Gibco, 10010-015)

Proteinase K (ThermoFisher, EO0491)

RNAse (Roche, 1119915)

SDS (Invitrogen, AM9820)

Triton X (Sigma, T8787)

Tween-20 (Sigma, P9416-50ML)

2.2 – Mouse model generation

All mouse work was performed in accordance with UK Home office regulations, under the appropriate animal licenses. Animal husbandry was conducted by the Weatherall Institute for Molecular Medicine (WIMM) Mouse Transgenics Core Facility.

2.2.1 – Design and generation of the Δθ1-θ2-44-48 CTCF mouse

The Δθ1-θ2-44-48 mouse model (**Fig. 2.3.1**) was designed by C. Harrold and B. Davies and engineered by B. Davies. In brief, oocytes/blastocysts from Δθ1-θ2 mice (**Cornell *et al.*, manuscript in preparation**) were microinjected with CRISPR/Cas9 expression constructs and sgRNA targeting directly upstream of HS+44 and downstream of HS+48, resulting in the deletion of both CBS and the ~4 kb intervening region (**Fig. 2.3.1**). Founder mice harbouring the desired mutation were crossed with C57BL/6 mice to obtain heterozygotes before being bred to homozygosity by B. Davies and J. Sharpe. This mouse model is further described in **Cornell *et al.*, manuscript in preparation**.

2.2.2 – Design and generation of the KLF1$^{StrepII\text{-}FKBP}$ mouse

The KLF1$^{StrepII\text{-}FKBP}$ mouse model (**Fig. 3.2.2**) was conceptualised and designed by Robert Beagrie (WIMM) in conjunction with the Wellcome Centre for Human Genetics. The KLF1$^{StrepII\text{-}FKBP}$ mouse was generated by Daniel Biggs (Wellcome Centre for Human Genetics). In brief, linear donor template DNA was microinjected into the pronuclei of C57BL6/J zygotes for integration by homologous recombination. After re-implantation, resulting progenies were genotyped, a founder identified, and successful germline transmission confirmed by PCR and Sanger Sequencing.

2.3 – Genotyping the mice

The genotyping strategies for Δθ1-θ2-44-48 and KLF1$^{StrepII\text{-}FKBP}$ are illustrated in **Fig. 2.3.1** and **Fig. 2.3.2**, respectively, and described below. Images of full genotyping gels are in **Appendix A3.1-3, A4.1, A4.2**. For both models, all samples appeared correctly edited and no samples were found to have an incomplete mutation series.

2.3.1 – gDNA isolation & genotyping PCR

Genomic DNA (gDNA) was isolated from Δθ1-θ2-44-48 mouse ear notches by incubation with DirectPCR Lysis Reagent (Viagen, 301-C) and 5 µL/mL Proteinase K overnight at 56ºC. Reactions were incubated at 95ºC, 10 minutes, to inactivate Proteinase K before taking 5 µL lysate forward for PCR.

gDNA was isolated from KLF1$^{StrepII\text{-}FKBP}$ mouse ear notch or fetal tissue using the DNeasy Blood & Tissue kit (Qiagen, 69506) according to the manufacturer's protocol with the following amendment: samples were spun down at 2000g for 2 minutes before applying supernatant to columns. All gDNA were quantified by Nanodrop spectrophotometry before PCR reactions were set up according to the manufacturer's protocol using the primers, polymerase, and PCR conditions in **Table 2.3.1.**

Table 2.3.1 – ***Primers and PCR cycling conditions for genotyping*** Δθ1-θ2-44-48 (red) and KLF1$^{StrepII\text{-}FKBP}$ (blue). Using IMMOLASE DNA Polymerase (Bioline, BIO-21047) or Platinum™ PCR SuperMix High Fidelity [1.1x] (Invitrogen, 12532-016). The products of Δθ1-θ2-44-48 PCR #3 were digested with BamHI-HF (NEB; R3136S) in CutSmart Buffer (NEB; B7204S).

#	Primer name	Sequence	Target	WT product (bp)	Mutant product (bp)	Enzyme	Annealing temp (°C)	Extension
1	007 008	CCCTGGCGGCCTCTTG GTCCAGGACAACAATGCAGC	**θ1**	802	N/A (Δθ1θ2)	Immolase	59	30 secs
2	009 010	TGTGGAAGACAGAAGAGGGC CTGAAACACAAGAGGCGCTG	**θ1**	N/A (Δθ1θ2)	506 (Δθ1θ2)	Immolase	65	30 secs
3	075 076	TTCCGAAGGACTCGGGAAGC ATGCACAGAGGCAATGCAGC	**θ2**	1008	Post-PCR: 1008; Post-BamHI digestion: 206 & 778 (Δθ2 & Δθ1θ2)	Immolase	62	30 secs
4	067 068	TGTCTGGAACTATCTGGAAACTGA GGGAAGAATAGACTAGACACATCA	**HS+44 / +48**	4520	967 (Δ3'4CTCF)	Immolase	57	1 min (only amplifies deletion allele)
5	073 077	TGCTTGGCCTTCAGGTTTCA AAGTTCCGAGTCCCGCCAC	**HS+48**	534	N/A (Δ44-48)	Immolase	60	30 secs
1	044 045	TCTCCCCTTATACAACAGGGACTGCA TCATCA AGCCACCATGTGGTTGCTTGTAATTG AACTCAG	**Upstream KLF1**	292	N/A	Platinum	55	69 sec
2	046 047	AGGTTCGCTCGCTCAGACGAACTGAC GGC ACA GTG TGA AGC TCC CTT CTA GAT C	**KLF1$^{StrepII-FKBP}$ Construct (internal)**	465	956	Platinum	55	69 sec
3	048	CATCTGGGATCAGGAGGTTGAGACAC TCA CTGGGGAGATGGTTTCCACCTGCACT	**KLF1$^{StrepII-FKBP}$ Construct (internal - external)**	N/A	1145	Platinum	55	69 sec

2.3.2 – Gel electrophoresis

PCR products were run with Gel Loading Dye (New England Biolabs, B7024S) on 1% agarose gel (Sigma, A9539-500G) in TAE buffer, using O'GeneRuler DNA Ladder mix (ThermoFisher, 11803983) ladder, unless specified otherwise. Gels were imaged on a Biorad Gel Doc XR+ with Image lab software. Gel images were edited in FIJI (rotated, colour inverted, brightness/contrast adjusted) and annotated in Microsoft Powerpoint.

2.3.4 – Sanger sequencing designs for Δθ1-θ2-44-48 and KLF1$^{StrepII\text{-}FKBP}$ mice

To verify the correct edits of both Δθ1-θ2-44-48 and KLF1$^{StrepII\text{-}FKBP}$ mouse models, I took gDNA from at least one breeding homozygous adult and performed PCRs using Q5 High-Fidelity 2X Master Mix (New England Biolabs; M0492S) according to the manufacturer's protocol, with the PCR cycling conditions and primers in **Tables 2.3.2, 2.3.3**. PCR reactions were run on 0.8% TAE agarose gels (as in **Section 2.3.2**), target bands excised, and PCR products purified with the QIAquick gel extraction kit (Qiagen, 28506), according to the manufacturer's protocol. PCR products were sent for in-house Sanger sequencing using the primers in **Table 2.3.3**, which verified the expected edited genotypes.

Table 2.3.2 – *PCR cycling conditions to generate templates for Sanger sequencing* of the Δθ1-θ2-44-48 (red) and KLF1$^{StrepII\text{-}FKBP}$ (blue). Using Q5 High-Fidelity 2X Master Mix (New England Biolabs, M0492S).

Step	Temp (°C)	Time	Cycles
Initial denaturation	98	30 sec	
Denaturation	98	10 sec	x40
Annealing	**66/68**	10 sec	
Extension	72	**60/204 sec**	
Final Extension	72	2 min	
Hold	10	∞	

Table 2.3.3 – *Primers for PCR and Sanger sequencing* of Δθ1-θ2-44-48 (red) and KLF1$^{StrepII\text{-}FKBP}$ (blue).

Model	PCR #	Primer name	Sequence	Target	Product (bp)	Function
Δθ1-θ2-44-48	1	008 009	GTCCAGGACAACAATGCAGC TGTGGAAGACAGAAGAGGGC	θ1	1284	PCR
	2	009 065	TGTGGAAGACAGAAGAGGGC ATGCATCTTTTTCGCCGCTGAAGTCTC	θ2	1201	PCR
	3	067 068	TGTCTGGAACTATCTGGAAACTGA GGGAAGAATAGACTAGACACATCA	HS+44 / +48	967	PCR
	-	005	GGTACGGGCTCCTGACACTC	**PCR 1**	-	Seq
	-	006	TGGAGGGGAAAGCCACGAAG	**PCR 1+2**	-	Seq
	-	066	CCCACATCTCCAGCAACCCG	**PCR 2**	-	Seq
	-	069	AAGCAAGCACTTCCCATCCA	**PCR 3**	-	Seq
	-	070	TGTGTCATAAGGAAGCCAGGG	**PCR 3**	-	Seq
KLF1$^{StrepII\text{-}FKBP}$	1	047 046	GGCACAGTGTGAAGCTCCCTTCTAGATC AGGTTCGCTCGCTCAGACGAACTGAC	Insert sequence	956	PCR
	2	047 061	GGCACAGTGTGAAGCTCCCTTCTAGATC AACTAGAGGGGTGAACCCGA	3' homology arm junction	1879	PCR
	-	046	AGGTTCGCTCGCTCAGACGAACTGAC	**PCR 1**	-	Seq
	-	063	GAGGTTTGTCCTGAAGGCTCC	**PCR 1**	-	Seq
	-	048	CATCTGGGATCAGGAGGTTGAGACACTCA	**PCR 2**	-	Seq
	-	059	ACTAGAGGGGTGAACCCGAA	**PCR 2**	-	Seq

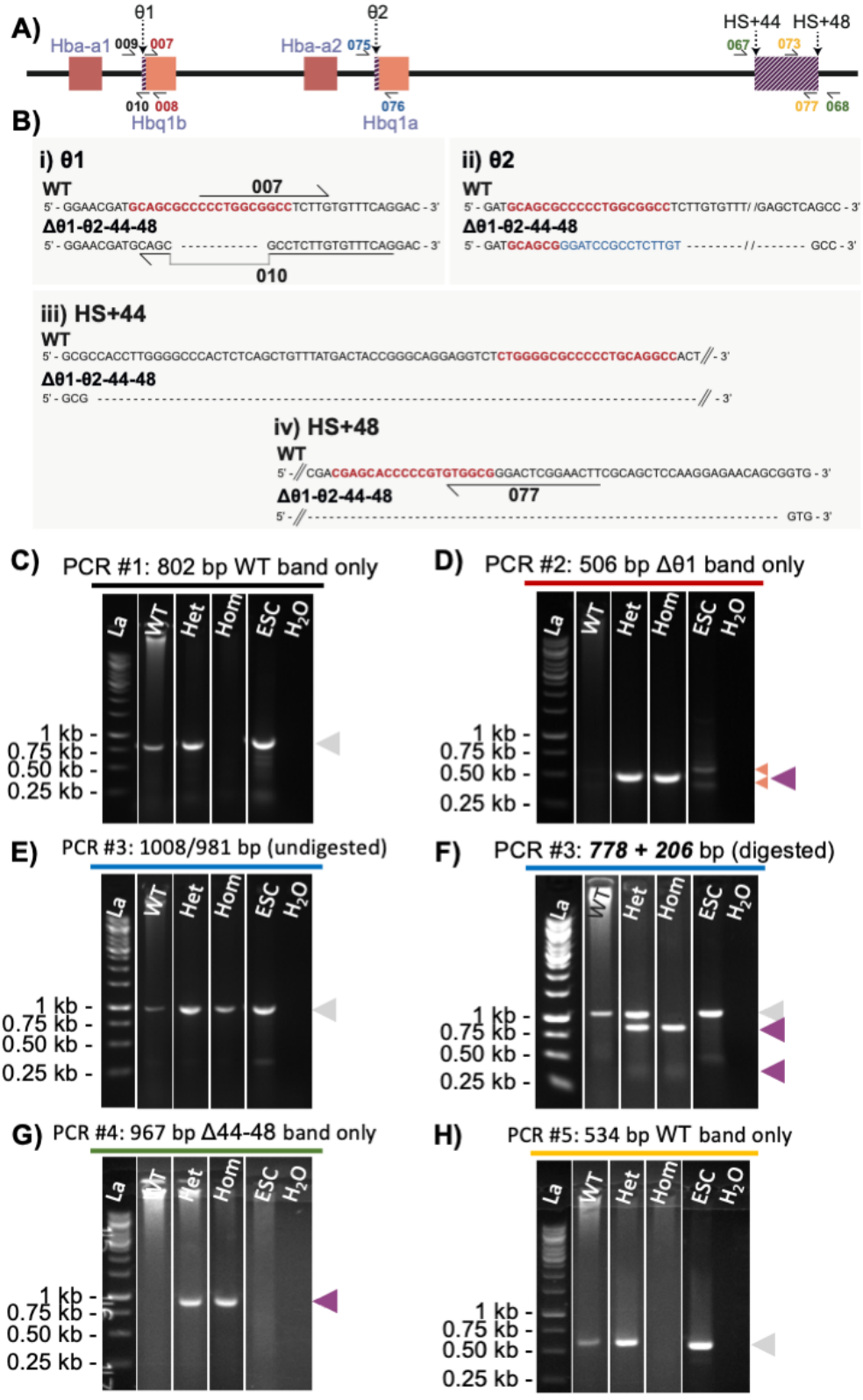

Figure 2.3.1 – *Δθ1-θ2-44-48 genotyping strategy identifies WT, het, and hom mice.* **A)** Schematic of the Δθ1-θ2-44-48 α-globin locus with approximate positions of genotyping primers for genotyping PCR #1 (black), PCR #2 (red), PCR #3 (blue), PCR #4 (green), and PCR #5 (yellow). Dark and light orange bars represent the *Hba-a1/2* and *Hbq1b/a* genes, respectively. Purple and white striped bars indicate DNA sequences deleted in Δθ1-θ2-44-48 mice (11 bp Δθ1; 10 bp Δθ2; 3553 bp ΔHS+44/48). The original sites of the θ1, θ2, HS+44, and HS+48 CTCF binding sites (CBS) are annotated. Not to

scale. **B)** Schematics of the θ1, θ2, HS+44, and HS+48 CBS sequences in WT (upper sequence; core CBS motif in red text) and deleted sequences in Δθ1-θ2-44-48 (lower sequence; dashed lines indicate deleted bases). Black arrows indicate positions of relevant genotyping primers. Slashes indicate DNA sequence excluded from schematic for brevity. **C-H)** Agarose gel electrophoresis of genotyping PCRs #1-#5, using PCR primer pairs 007/008, 009/010, 075/076, 067/068, and 073/077 respectively. For 4 representative samples: WT, het, and hom littermates, an unrelated mESC control "ESC", and a no gDNA control "H_2O". "La" = Fermentas O'GeneRuler 1kb DNA Ladder. All complete gels in **Appendix A3.1-3**. Titles above each gel indicate expected product sizes. **C)** PCR #1 uses primers (009 and 010) that bind the intact θ1 CBS (**B-i**) to produce an 802 bp band in samples containing at least one WT θ1 allele (light grey arrow). **D)** PCR #2 uses primers (007 and 008) that selectively bind over the junction of the expected 11 bp θ1 deletion (Δθ1; **B-i**) to produce a 506 bp amplicon in samples containing at least one Δθ1 allele (purple arrow). Small orange arrows indicate off target amplifications present in the unrelated ESC WT control, only. **E)** PCR #3 (075 and 076) amplifies 1008 or 981 bp specifically over θ2 (grey arrow). The expected 42 bp θ2 CBS deletion (Δθ2) was designed to insert a novel BamHI restriction site (**B-ii**, 15 bp sequence in blue). **F)** Amplicon digestion with BamHI results in 778 and 206 bp digestion fragments (purple arrows) in samples containing at least one Δθ2 allele. Samples with at least one WT θ2 allele retain the 1008 bp band (grey arrow). **G)** PCR #4 primers (067 and 068) bind unedited sequence flanking the Δ44-48 deletion. By using an extension time optimised for amplification of the shorter Δ44-48 amplicon, a single 967 bp band is expected from the successfully edited Δ44-48 alleles (purple arrow). **H)** PCR #5 uses primers (073 and 077) located within the deleted Δ44-48 region, to produce a 534 bp amplicon in WT alleles only (grey arrow).

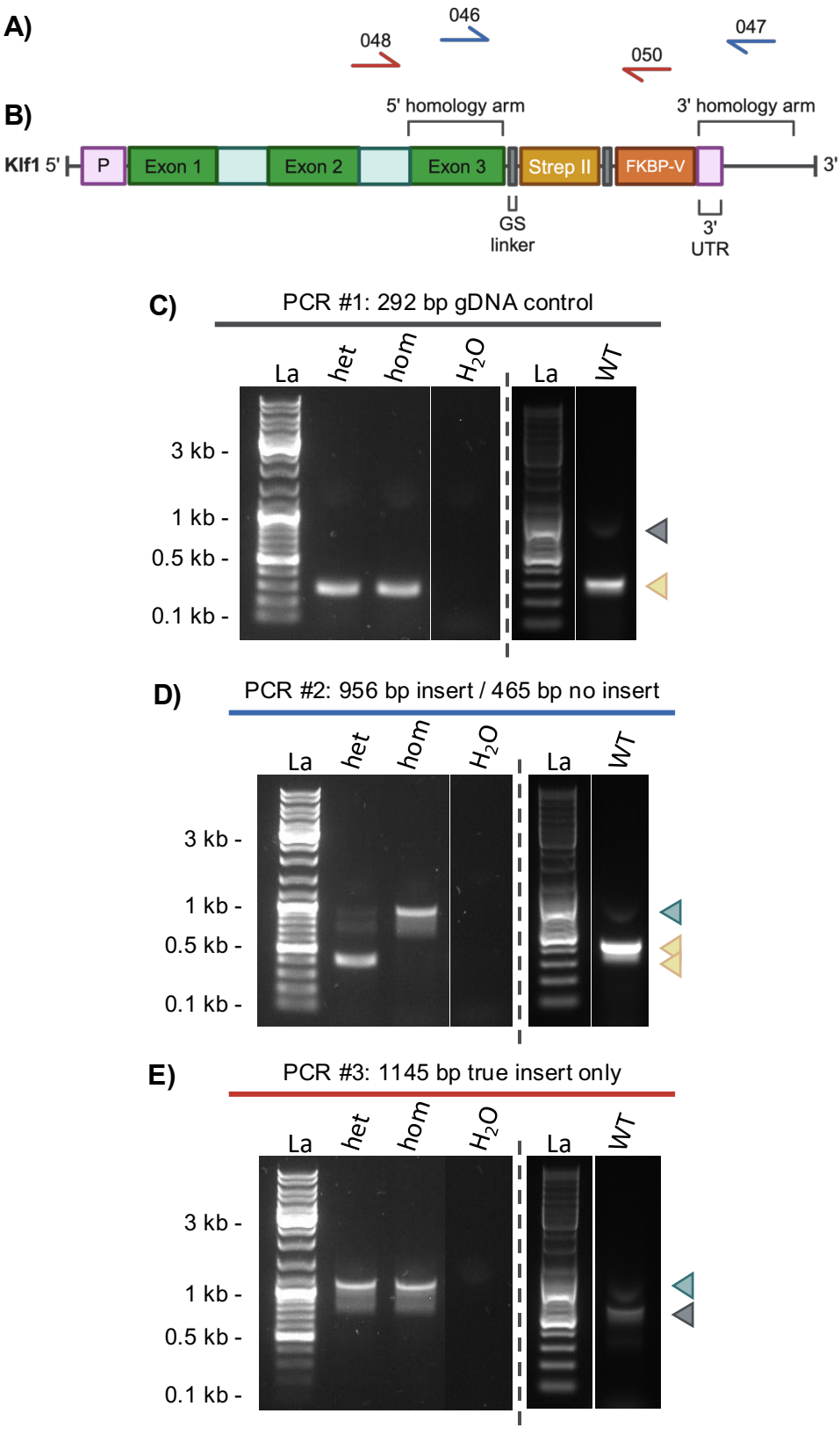

Fig 2.3.2 – *Genotyping strategy for KLF1$^{StrepII\text{-}FKBP}$ identifies WT, het, and hom mice.* **A)** Relative position of genotyping primers for genotyping PCR #2 (blue) and PCR #3 (red), with respect to B). Not to scale. **B)** Diagram of the recombinant mouse *Klf1* locus. "P" = *Klf1* promoter. Light green boxes = *Klf1* introns. Orange and grey boxes = inserted sequence encoding StrepII and FKBP-V tags (orange), flanked by GS linker sequences

(grey). Not to scale. **C-E)** Agarose gel electrophoresis of genotyping PCRs #1-#3, using PCR primer pairs 044/045, 046/047, and 048/050 respectively, for 3 representative samples: het, hom, and an unrelated WT control. "H_2O" = no gDNA. "La" = Thermo Scientific GeneRuler DNA Ladder Mix. For each PCR, WT was run on a different gel to het, hom, and H_2O. All complete gels in **Appendix Fig. A4.1, A4.2**. Green arrows indicate amplicons from KLF1$^{StrepII\text{-}FKBP}$ alleles. Yellow arrows indicated amplicons from WT alleles. Grey arrows indicate off-target/contaminant bands. **C)** PCR #1 is a control PCR that confirms presence of template gDNA. The primers (044 and 045) bind to a region downstream of the endogenous *Klf1* gene, on chromosome 8. PCR amplification produces a single 292 bp band in all samples containing intact mouse gDNA (yellow arrow). **D)** PCR #2 amplifies a region over the insert sequence, internal to the donor construct's homology arms. The forward primer, 046, binds in *Klf1* exon 3 (5' homology arm) and the reverse primer, 047, binds downstream of the *Klf1* coding sequence (within the 3' homology arm). This produces a larger 956 bp band (green arrow) if the StrepII-FKBP DNA sequence is present, or a shorter 465 bp band if amplifying over the WT *Klf1* sequence (yellow arrows). In het samples, PCR falsely biased amplification of the shorter WT band. Additionally, as both primers lie within the homology arms of the donor template DNA, this PCR could also produce a 956 bp band in samples with off-target insertions of the donor sequence. To overcome these limitations, I designed PCR #3 to identify true insertions into *Klf1*. **E)** For PCR #3, the forward primer, 048, binds within the endogenous *Klf1* gene sequence, upstream and external to the 5' homology arm. The downstream primer, 050, binds to *FKBP* sequence, internal to the insert. No band should be amplified in WT samples (differentiating them from het samples when analysed alongside PCR #2). A 1145 bp band should be amplified in het or hom sample with at least one true *KLF1*$^{StrepII\text{-}FKBP}$ insertion (green arrow). In all WT, het, and hom samples for PCR #3, a faint off-target band at ~800 bp is visible (grey arrow). In most samples from all PCRs and the no gDNA "H2O" control, there is an unexpected, "smiling" band at ~1000 bp (grey arrow in C)), suggestive of contamination of the PCR reactants or loading dye. Neither of these erroneous bands confound interpretation of the genotyping results.

2.4 – Mouse haematology

All haematologic analysis were performed on adult (>7 weeks) male and female Δθ1-θ2-44-48 or KLF1$^{StrepII\text{-}FKBP}$ mice and age-matched C57BL/6J WT controls, N > 3 (See **Method 2.13.1** for minimum sample size calculation). >80 µL peripheral blood was collected in EDTA before immediate use in the following assays. Red and white blood cell haematologic parameters were measured on a Horiba Medical Scil Vet abc Plus+ instrument. To prepare blood smears, 2 µL of peripheral blood was spread on glass slides, dried, stained with the Hematek Stain Pak's modified Wright-Giemsa stain (Siemens Healthcare, 23-044626), and mounted. For reticulocyte counts, >10 µL peripheral blood was incubated in an equal volume of BCB stain (1% w/v Briliant Cresyl Blue Powder + 0.9% (w/v) NaCl) for >4 hours at room temperature. 2 µL blood/BCB was then smeared as described above.

All blood films were imaged on an Olympus BX60 light microscope with Infinity Analyze (V6.5) software. Reticulocyte counts and blood smear assessments were verified by two independent researchers, not blinded to genotype unless specified. Reticulocytes were classified and scored according to **Fig. 2.4.1**.

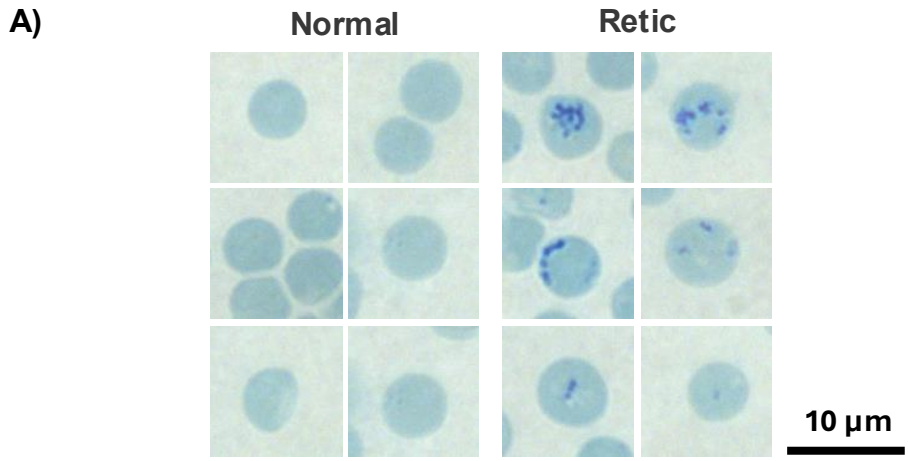

Fig 2.4.1 – *Example reticulocyte classifications*. N = 3, n = 265-776. Reticulocytes are the final precursor cell to the mature erythrocyte. A normal reticulocyte count range in mouse peripheral blood is ~1-6% of total RBC count (**O'Connell *et al*., 2015**) and can be a proxy for rate of RBC production in the bone marrow. Reticulocyte count may be decreased in cases of anaemia with decreased haemoglobin or RBC production. Likewise, circulating reticulocyte count can be increased when RBC production in the bone marrow is elevated, such as in response to haemolytic anaemia (**Rai, Wilson and Moosavi, 2023).**

2.5 – Cell Culture

2.5.1 – Spleen harvest

Adult Δθ1-θ2-44-48 mice were treated with acetylphenylhydrazine (APH) to induce extramedullary haematopoiesis. Spleens were dissected and dissociated in FACS buffer (PBS + 10% FBS) by repeated syringe aspiration. Spleen homogenates were passed through a 70 µm cell strainer (Falcon, 352350) before LS column (Miltenyi Biotec, 130-042-401) selection for $Ter119^{+}$ according to the manufacturer's protocol. In brief, aliquots of <200 x 10^6 cells were incubated with 30 µg/mL PE-Ter119 antibody (Miltenyi Biotec, 130-102-336) and anti-PE beads (Miltenyi Biotec, 130-048-801). Cells were passed through a 30 µm cell strainer into LS columns for magnetic bead selection. $Ter119^{-ve}$ and $Ter119^{+}$ fractions were verified by flow cytometry (**Method 2.6, Fig. 2.5.1**) and $Ter119^{+}$ erythroblasts immediately taken for downstream applications.

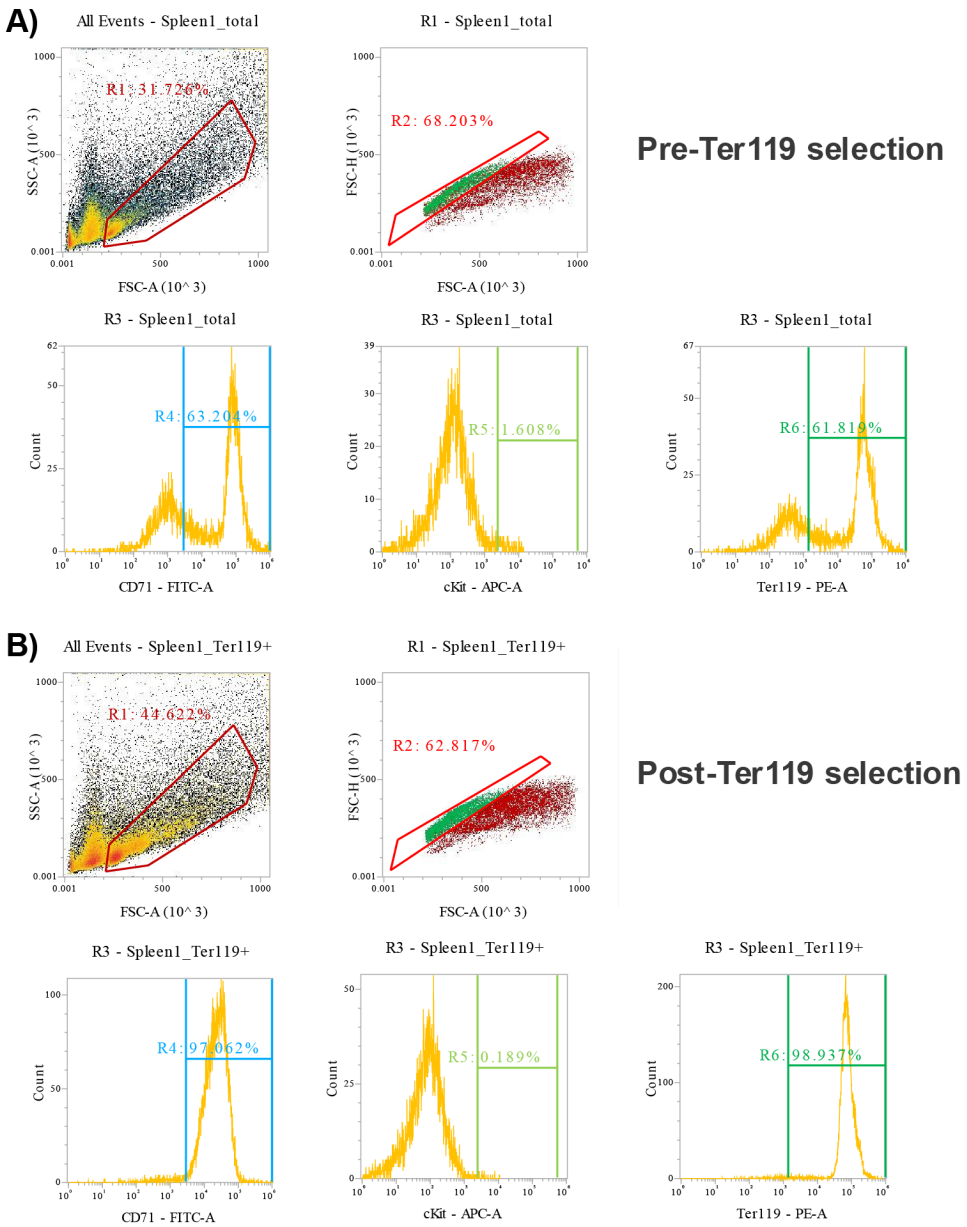

Fig 2.5.1 – *Verifying Ter119+ selection of splenic erythroid cells.* Flow cytometry on Δθ1-θ2-44-48 APH-treated splenic erythroid cells, before (**A**) and after (**B**) selection for Ter119.

2.5.2 – Embryonic blood harvest

E10.5-E12.5 embryos were bled from severed umbilical cords into PBS + 10 mM EDTA. The collected embryonic blood was passed through a MACS Pre-Separation Filter (Miltenyi Biotec, 130-041-407) and spun at 300 g, 5 minutes, RT. The PBS supernatant was removed and blood pellets frozen at -80^{o}C for long-term storage.

2.5.3 – Fetal liver harvest & culture

Mouse fetal livers (mFL) were isolated from E12.5 embryos and each dissociated in 1 mL expansion media (**Table 2.5.1, 2.5.2**) by passing 10 times through a 23G needle (Agani, AN-2332R). Isolated mFL cells were then cultured at 37°C, with 5% CO_2, and passaged according to **Table 2.5.3** for up to 7 days.

Cells were frozen on day 2 of culture by resuspending aliquots of 3-5 million live cells in 1 mL expansion media with 10% DMSO. Cells were gradually frozen to -80^{o}C and stored long term in liquid nitrogen (-160^{o}C). 24 hours after thawing, cells were centrifuged (300 g, 5 minutes, RT) and resuspended to 1 million/mL in fresh expansion media. Otherwise, cells were passaged normally and recovered for a minimum of 48 hours before further experimentation.

2.5.4 – Fetal liver dTAG treatment

mFL cells were treated on day 2 (fresh) or day 4 (thawed) of culture with 0.2 μM dTAG-13 (Bio-Techne, 6605, lot: 20221212) or an equivalent volume of DMSO. For each day of dTAG-13/DMSO treatment, cells were centrifuged (300 g, 5 minutes, RT) and resuspended to 1 million/mL in fresh expansion media with dTAG-13 or DMSO.

Table 2.5.1 – *mFL base media.* All media and reagents prepared under sterile conditions.

Reagent	Supplier, cat #	[Final]
StemPro-34 SFM	Gibco, 10639011	-
Nutrient Supplement	Gibco, 10639011	1 X
Penicillin/Streptomycin	Gibco, 15070-063	1 X
L-glutamine	Gibco, 25030-024	2 mM

Table 2.5.2 – *mFL expansion media.* All media and reagents prepared under sterile conditions.

Reagent	Supplier, cat #	[Final]
Base Media	-	-
mSCF (resuspended in PBS + 0.5% BSA)	PeproTech, 250-03-100UG	50 ng/uL
Eprex Epoetin alfa	Janssen-Cilag, GTIN: 15012674903206	1 U/mL
Dexamethasone (stock diluted 1:10 in PBS)	Hamelm, DEXA3.3, BN: 209508	0.4 µg/mL

Table 2.5.3 – *Passaging E12.5 mFL erythroid cells.* On day 1 of culture, add 2 mL fresh expansion media (Σ = 3 mL) and move to a 6-well plate, without counting or centrifuging.

Culture day	Description	Cell number	Post-passage concentration	Culture vessel
Day 0	Harvest	1-5 million	-	12-well plate
Day 1	Recovery		-	6-well plate
Day 2-7	Expansion/ Treatment	<10 million	~1 million/mL	T-25
		>10 million	~1 million/mL	T-75

2.6 – Flow cytometry & FACS

2.6.1 – Δθ1-θ2-44-48 spleen cell flow cytometry

In brief, cells were stained with the antibody panel in **Table 2.6.1** and 1 µg/mL Hoechst 33258 before running on a Attune NXT flow cytometer.

Table 2.6.1 – ***Antibody panel for flow cytometry of Δθ1-θ2-44-48 spleen cells.***

Epitope	Conjugate	Supplier	Cat #	Lot #	Dilution
CD71	**FITC**	eBioscience	11-0711-85	-	1:200
Ter119	**PE**	BD Pharmingen	553673	-	1:100
c-kit	**APC**	Biolegend	105812	-	1:100

2.6.2 – KLF1$^{StrepII\text{-}FKBP}$ mFL cell FACS

In brief, cells were blocked with 200 µg/mL anti-Rabbit IgG (ChromePure, 015-000-003) then stained with 0.67 µg/mL Hoechst 33258 and the antibody panel in **Table 2.6.2**. Cells were sorted with the Sony MA900 into expansion media and recovered at 37°C, 5% CO_2, for ~6 hours before use in subsequent protocols. FACS and flow cytometry data was analysed with FlowJo (v.10.9.0).

Table 2.6.2 – ***Antibodies used for flow cytometry and FACS of KLF1$^{StrepII\text{-}FKBP}$ mFL cells.***

Epitope	Conjugate	Supplier	Cat #	Lot #	Dilution
Ter119	**Biotin**	BD Biosciences	553672	6113560	1/100
Streptavidin	**APC**	Biolegend	405207	B223856	1/300
CD71	**PE-Cy7**	Biolegend	113811	B221915	1/1500

2.7 – Protein methods

2.7.1 – IEF

To collect primitive blood for IEF, E12.5 embryo yolk sacs were individually bled into 1 mL PBS. Samples were spun (300 g, 5 min), the PBS supernatant removed, and cell pellets stored at -80°C until use. To collect definitive samples, cultured E12.5 mFL cells were washed once in PBS and cell pellets frozen at -80°C until use.

IEF was performed with the RESOLVE Haemoglobin kit (Revvity, FR-9120) and protocol from the Special Haematology & HMDL Department, Leicester Royal Infirmary, as adapted by Ayesha Ejaz (DPhil Student, Higgs laboratory). In brief, 5 uL of whole blood or cell pellets of >12.5 million cells, were resuspended in 30-50 uL Hb Elution Solution (RESOLVE) and incubated (10 min, RT) to lyse. Cell culture lysates were filtered with 0.45 µm centrifugal filters (Merk Millipore, UFC30HVNB, following manufacturer's protocol) to remove nuclear DNA. 5 uL of each lysate was loaded onto the IEF gel and run for 1.5 hours, 1500 V. The pI-separated haemoglobin proteins were then transferred onto a nitrocellulose membrane (Amersham, 10600096) by compression in transfer buffer (50% ethanol, 50% PBS) for 15 minutes at room temperature. The nitrocellulose membrane was then washed in deionised H_2O before proceeding with my western blotting protocol, as outlined in **Methods 2.7.2**, from blocking onwards and using the antibodies outlined in **Table 2.7.1**. After transfer, the IEF gel was fixed in 10% TCA for 10 min, shaking. The gel was then washed with DI H_2O before incubation with gel stain (Revvity, FR-9367) for 10 minutes. The gel was then washed in DI H_2O for >1 hour and air-dried overnight.

Table 2.7.1 – ***Anti-ζ-globin antibodies tested for IEF/Western blot.*** The full IEF and Western blot membranes are available in **Appendix Fig. A5.4**. For details of the in-house anti-ζ-globin antibody, see **Luo *et al.*, 1988**. Ab228651 was found to cross react with mouse ζ-globin and chosen for the final IEF figure.

Cat #	Lot #	Epitope	Supplier	Host species	Clonality	[Stock]	Working dilution
-	-	ζ-globin (mouse, human)	In-house	Mouse	Mono	n/a	1:10
101272, Clone C01B3	B1909	HBZ (human)	Santa Cruz	Mouse	Mono	100 µg/mL	1:100
Ab228651	1008261-6	HBZ (human)	Abcam	Rabbit	Poly	1 mg/mL	1:1000

2.7.2 – Western blot

Whole cell lysis

2 x 10^6 mFL cells were washed once in PBS + 10 µL/mL protease inhibitor cocktail (PIC; Sigma-Aldrich, P8340) then resuspended in 100 µL RIPA buffer (Sigma-Aldrich, R0278) +10 µL/mL PIC by triturating each pellet five times. Lysates were incubated for 30 minutes, 4°C, rotating, before centrifuging for 30 minutes, >21,000 g, 4°C. 90 µL supernatant + 90 µL loading dye (125mM Tris-HCl, 4% SDS, 50% glycerol, 0.2 g OrangeG (Sigma: O7252-25G), and 1M DTT (ThermoFisher: R0861)) + 20 µL 1M DTT (Sigma-Aldrich, 43816) was then boiled at 95°C for 10 minutes. Lysates were then immediately placed on ice for storage at -20°C or immediate use in Western blotting. All samples and reagents are kept on ice unless otherwise specified.

Subcellular fractionation

Protocol for extraction of sub-cellular protein fractions was modified from **Baghirova *et al.*, 2015**. By lysing cells with increasing detergent stringencies, the cytoplasmic membrane, organelle membranes, and nuclear membrane are sequentially lysed to release their compartment proteins. In brief, 3.5 x 10^6 mFL cells were washed in PBS + PIC. Cells were resuspended in 200 µL Lysis buffer A (**Table 2.7.2, Table 2.7.3**) and incubated for 30 minutes, 4°C, rotating. Samples were centrifuged 2000 g, 10 minutes, 4°C, and the supernatant cleared as the "cytoplasmic" fraction. The pellet was resuspended in 200 µL Lysis buffer B (**Table 2.7.4**) and incubated 30 minutes, 4°C. Samples were centrifuged 7000 g,

10 minutes, 4°C, and the supernatant cleared as the “organelle” fraction. The pellet was resuspended in 200 µL Lysis buffer C (**Table 2.7.5**) and incubated 30 minutes, 4°C, rotating. Samples were centrifuged 7800 g, 10 minutes, 4°C, and the supernatant cleared as the “nuclear” fraction. Lysates were boiled with loading dye and stored, as in preparation of whole cell lysates.

Table 2.7.2 – *Lysis Base Buffer.*

Reagent	Supplier, cat. #	[Final]
NaCl	Invitrogen, AM9760G	150 mM
HEPES (pH 7.4)	Gibco, 15630-080	50 mM
Hexylene Glycol	Sigma, 112100-500G	1 M
Nuclease free H_2O	Sigma, W4502-1L	-

Table 2.7.3 – *Lysis buffer A.* For lysis of cytoplasmic membranes

Reagent	Supplier, cat. #	[Final]
Lysis Base Buffer	-	-
Digitonin	Sigma, D141-500MG	1% v:v
PIC	Sigma, P8340	10 µL/mL

Table 2.7.4 – *Lysis buffer B.* For lysis of organelle membranes.

Reagent	Supplier, cat. #	[Final]
Lysis Base Buffer	-	-
IGEPAL	Sigma, I8896	1% v:v
PIC	Sigma, P8340	10 µL/mL

Table 2.7.5 – *Lysis buffer C.* For lysis of nuclear membranes. Can use 0.02 g SDS powder per 20 mL, or 20% SDS solution as below. For the sodium deoxycholate, first dissolve 0.5 g in 2.5 mL nuclease-free H_2O for a 20% stock solution.

Reagent	Supplier, cat. #	[Final]
Lysis Base Buffer	-	-
SDS	Invitrogen, AM9820	0.1% w:v
Sodium deoxycholate	Sigma, D6750-100G	0.5% w:v
Benzonase	Sigma, E1014-25KU	17.5 µL/mL
PIC	Sigma, P8340	10 µL/mL

Fluorescent Western blot

30 µL freshly boiled whole cell lysates were loaded onto NuPAGE 4-12% Bis-Tris Protein Gels (ThermoFisher, NP0335BOX) alongside 3 µL Chameleon Duo Pre-Stained Protein Ladder (Li-Cor, 928-60000). For SDS-PAGE, gels were run with NuPAGE MOPS SDS Running Buffer (ThermoFisher, NP0001) for 3 minutes, 170 V, then 90 minutes, 120 V. Gels were transferred onto a 0.45 µm nitrocellulose membrane (Amersham,10600096) in 1X NuPAGE Transfer buffer (ThermoFisher, NP0006) and 25% methanol for 90 minutes, 30 V. After transfer, the nitrocellulose membrane was rinsed in de-ionised H_2O, optionally stained with Ponceau S solution (Sigma-Aldrich, P7170), then blocked for 1 hour, room temperature, in 5% milk in PBS, with agitation. Membranes were then incubated overnight, 4^{o}C, with primary antibody (see **Table 2.7.6**) in 5% milk in PBS + 0.2% Tween-20. Membranes were then washed three times in PBS + 0.1% Tween-20 before incubation with species-matched IRDye-conjugated secondary antibody (see **Table 2.7.7**) in 5% milk + 0.2% Tween-20 for 1 hour, room temperature, in the dark. Membranes were then washed three times in PBS + 0.1% Tween-20 for 5 minutes each, then once in PBS for 10 minutes.

Membranes were imaged with the Li-Cor Odessy M. Images were edited in FIJI (rotated, colour inverted, brightness/contrast adjusted, and exported to .tiff) and annotated in Microsoft Powerpoint.

Quantifying Western blots

Brightness/contrast adjusted Western blots images were quantified using FIJI. In brief, images were blurred with a median filter, radius = 2. Lanes were then isolated, quantified, and signal intensity (area under curve; AUC) of the bands above background were measured. Each sample band AUC is divided by the control band AUC (from within the same lane) and multiplied by 100 to give normalised peak size (arbitrary units). This should be considered a semi-quantitative measure.

Table 2.7.6 – *Primary antibodies for fluorescent Western blotting*. "H" = human, "Ms" = mouse, "Rb" = rabbit, "O" = others.

Cat #	Supplier	Epitope	Host species	Clonality	Target species	Expected band (kDa)	Dilution
2118	Cell Signalling	**GAPDH**	Rb	mono	H, Ms, O	37	1:1000
ab1791	Abcam	**Histone H3**	Rb	poly	H, Ms, O	~17	1:5000
sc-55492	Santa Cruz	**Pol II**	Ms	mono	H, Ms, O	~200	1:1000
34850	Qiagen	**Strep-II**	Ms	mono	-	-	1:1000

Table 2.7.7 – *Secondary antibodies used for fluorescent Western blotting*. "Ms" = mouse, "Rb" = rabbit.

Cat #	Supplier	Host species	Clonality	Target species	IRDye	Dilution
ab216772	Abcam	Goat	poly	Ms	800CW	1:1000
ab216776	Abcam	Goat	poly	Ms	680RD	1:1000
ab216773	Abcam	Goat	poly	Rb	800CW	1:1000
ab216777	Abcam	Goat	poly	Rb	680RD	1:1000

2.8 – Gene expression assays

2.8.1 – qPCR

Aliquots of 5-10 x 10^5 cells were snap frozen in Trizol reagent (Sigma, T9424) and stored at -80°C. RNA was purified with the Direct-zol RNA Mini Prep kit (Zymo, #2050), quantified with HS RNA Qubit (Invitrogen, Q32852), and quality checked with HS RNA Tapestation (Agilent). An approximately equal amount (0.3-1 µg) of RNA was taken for cDNA synthesis with SuperScript III first-strand synthesis SuperMix for qRT-PCR (Invitrogen, 11752-250). RT-qPCR was performed using TaqMan Fast Advanced Master Mix (Applied Biosystems, 4444557) and exon-spanning TaqMan probes (**Table 2.8.1**) to detect mRNA only. Plates were set up with a sample-centric layout (all probes per sample on the same plate) and all plates run on the same QuantStudio 3 RT-PCR machine (Applied Biosystems) to

reduce technical variation. No template and no reverse-transcriptase controls were run with each experiment.

Technical triplicate Ct values falling within 1 Ct value were mean averaged. The ΔΔCt method was used to normalise to Rps18 and to the experimental control as specified. For alpha:beta ratios, the second ΔΔCt normalisation was performed against β-globin (alpha dCT / beta dCT). All qPCR experiments were performed in biological triplicate.

Table 2.8.1 – *Taqman probes for RT-qPCR*.

Taqman probe	Catalogue ID	Target
Rps18 [60X]	Mm02601777_g1	Exons 3-4
Hba-a2 [60X]	Mm02580841_g1	Exons 1-2
Hba-x [60X]	Mm00439255_m1	Exons 1-2
Hbb-b1 [20X]	Mm01611268_g1	Exons 1-2
Hbb-y [20X]	Mm00433936_g1	Exons 1-2
Hbb-bh1 [20X]	Mm00433932_g1	Exons 1-2
Klf1 [20X]	Mm00516096_m1	Exons 1-2
Bcl11a [20X]	Mm00479358_m1	Exons 3-4
Zbtb7a [20X]	Mm00657132_m1	Exons 2-3
Sh3xpd2b [20X]	Mm00616672_m1	Exons 9-10
Ubtd2 [20X]	Mm00612868_m1	Exons 1-2

2.9 – NGS assays

2.9.1 – ATAC-seq

Assay for transposase-accessible chromatin (ATAC)-seq was performed using the Illumina Tagment DNA and buffer kit (Illumina, 20034197), as previously described (**Buenrostro *et al.*, 2015**; **Hay *et al.*, 2016**). In brief, ~7.5 x 10^4 cells were washed in PBS, resuspended in 50 µL ATAC lysis buffer (**Table 2.9.1**), and spun 10 minutes, 500 g, 4°C. Pellets were gently resuspended in 50 uL transposition reaction mixture (**Table 2.9.2**) and incubated 30 minutes, 37°C. Immediately following transposition, tagmented DNA was purified with MinElute PCR Purification Kit (Qiagen, 28004). ATAC libraries were indexed with the NEBnext High-Fidelity 2X PCR Master Mix, as in **Tables 2.9.3**, **2.9.5**, using PCR cycling conditions in **Table 2.9.4**. The PCR amplified libraries were then purified

with the Qiaquick PCR cleanup kit (Qiagen, 28106). Successful library transposition and indexing was verified by HS D1000 Tapestation (Agilent, 5067-5584) (**Fig. 2.9.1**) before library quantification by HS dsDNA Qubit (Invitrogen, Q33230).

Table 2.9.1 – [1X] *ATAC lysis buffer.*

Reagent	Supplier, cat. #	[Final]
Tris HCL (pH 7)	Invitrogen, AM9850G	10 mM
NaCl	Invitrogen, AM9760G	10 mM
$MgCl_2$	Invitrogen, AM9530G	3 mM
Nuclease free H_2O	Sigma, W4502-1L	-
IGEPAL CA-630	Sigma, I8896	0.1%

Table 2.9.2 – *Transposition reaction mixture.* Using the Illumina Tagment Enzyme and Buffer kit (20034210).

Reagent	Supplier, cat. #	Volume (µL)
[2X] Tagment DNA Buffer	Illumina, 15027866	25
Tn5 Transposase Enzyme	Illumina, 15027865	2.5
Nuclease free H_2O	Sigma, W4502-1L	22.5

Table 2.9.3 – *NEBnext PCR reactants.* See **Table 2.9.5** for Nextera primer sequences.

Reagent	Supplier, cat. #	Volume (µL)
NEBNext High-Fidelity [2x] PCR Master Mix	New England Biolabs, M0541	25
Nuclease free H_2O	Sigma, W4502-1L	10
[25 µM] Nextera Primer 1.1 (universal)	-	2.5
[25 µM] Nextera Primer 2.x (barcode)	-	2.5
Transposed DNA	-	10

Table 2.9.4 – *NEBnext High-Fidelity PCR cycling conditions*.

Step	Temp (°C)	Time	Cycles
Adapter end repair	72	5 min	
Initial denaturation	98	30 sec	
Denaturation	98	10 sec	x 9-12
Annealing	63	30 sec	
Extension	72	60 sec	
Final Extension	72	1 min	
Hold	4	∞	

Table 2.9.5 – ***Nextera indexing primer sequences*.**

Function	Primer ID	Sequence
Universal	1.1	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Indexing	2.1	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
	2.2	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
	2.3	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
	2.4	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
	2.5	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
	2.6	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
	2.7	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
	2.8	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
	2.9	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
	2.10	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
	2.11	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
	2.12	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
	2.13	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
	2.14	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
	2.15	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
	2.16	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT

	2.17	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGG GCTCGGAGATGT
	2.18	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGG CTCGGAGATGT
	2.19	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGG CTCGGAGATGT
	2.20	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGG CTCGGAGATGT
	2.21	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGG CTCGGAGATGT
	2.22	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGG CTCGGAGATGT
	2.23	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGG GCTCGGAGATGT
	2.24	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGG GCTCGGAGATGT

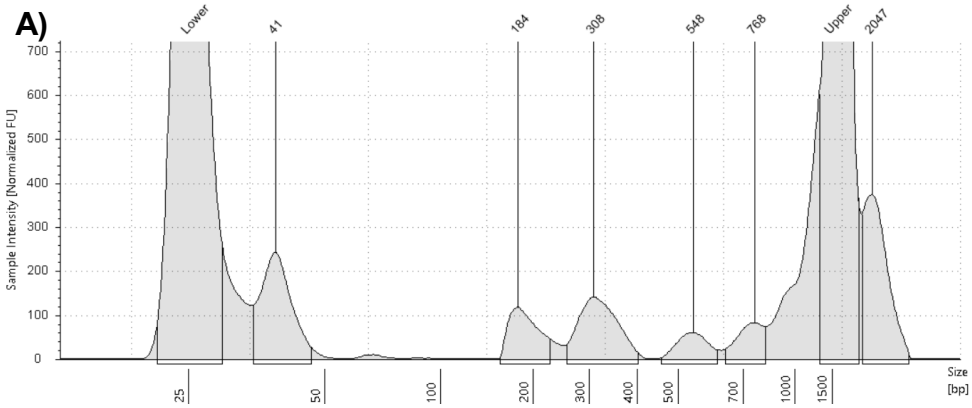

Fig. 2.9.1 – *An example ATAC library D1000 HS Tapestation quality trace*, post-tagmentation and indexing.

2.9.2 – KLF1 Chipmentation

Chipmentation was performed as previously described, with the following amendments. 2.5 x 10^5 cells were fixed in 1 mL PBS + 0.4% formaldehyde for 10 minutes, before being quenched with 125 mM glycine. Pellets were snap frozen and stored at -80°C.

For chipmentation, 5 µL Protein A + 5 µL Protein G Dynabeads (Invitrogen, 10002D, 10003D) were blocked in 150 µL blocking solution (**Table 2.9.6**) + 16 µL anti-StrepII monoclonal antibody (Qiagen, 34850), 2-4 hours, 4°C, then washed twice in blocking solution. Antibody concentration was empirically determined.

Fixed cell pellets were resuspended in 120 µL lysis buffer (**Table 2.9.7**) and incubated on ice, >30 minutes. Lysates were sonicated with the Covaris ME220, conditions in **Table 2.9.8**. Sonicates were incubated with 1% Triton X, 10 minutes room temperature, to neutralise SDS. 5 µL each sample was taken for purification (RNase A and proteinase K treated and Qiaquick column purified (Qiagen, 28106) and library preparation as input control. Sonicates were analysed by HS D1000 Tapestation (Agilent) before incubating overnight, 4°C, with the antibody-bound Protein A + G beads. The bead-bound samples were then sequentially washed once each with chipmentation wash buffers A, C, D, TE, and twice with Tris (**Tables 2.9.9-11**), then resuspended in Tagmentation buffer (**Table 2.9.12**). The transposition reactions were incubated for 10 minutes precisely, 37°C, before Tn5 inactivation by addition of wash buffer A (without PIC). Beads were resuspended in nuclease free H_2O and immediately proceeded to library preparation with NEBnext High-Fidelity 2X PCR Master Mix, as in **Table 2.9.13**, with primers **Table 2.9.5**, using PCR cycling conditions in **Table 2.9.14**. Indexed libraries were purified with AmpureXP SPRI beads (Beckman Coulter, A63880) and eluted in 0.1X TE. Libraries were checked with HS dsDNA Tapestation and HS dsDNA Qubit (**Fig. 2.9.2**). Purified inputs were indexed and quality controlled with the same protocol.

Table 2.9.6 – *Chipmentation blocking solution*.

Reagent	Supplier, cat. #	[Final]
PBS	Gibco, 10010-015	-
BSA	Sigma, A3059-50G	0.5%
cOmplete PIC	Sigma, 11873580001	1 X

Table 2.9.7 – *Chipmentation lysis buffer*.

Reagent	Supplier, cat. #	[Final]
Tris pH 8	Invitrogen, AM9850G	50 mM
EDTA	Invitrogen, AM9260G	10 mM
20% SDS	Invitrogen, AM9820	0.5 %
Nuclease free H_2O	Sigma, W4502-1L	-
cOmplete PIC	Sigma, 11873580001	1 X

Table 2.9.8 – *Covaris sonication conditions* for single fixed mFL.

Cell type	Fixation	PIP	DF	CPB	Time (s)	Average power
mFL	Single	75	9	1000	500	6.75

Table 2.9.9 – *Chipmentation wash buffer A.* Low salt wash buffer.

Reagent	Supplier, cat. #	Final concentration
Tris-HCl pH 7	Invitrogen, AM9850G	10 mM
EDTA pH 8.0	Invitrogen, AM9260G	1 mM
EGTA pH 8.0	Alpha Aesar, J60767	0.5 mM
Triton X-100	Sigma, T8787	1%
SDS (20%)	Invitrogen, AM9820	0.1%
Na Deoxycholate (5%)	Invitrogen, AM9740	0.1%
NaCl	Invitrogen, AM9760G	140 mM
Nuclease free H_2O	Sigma, W4502-1L	-
cOmplete PIC	Sigma, 11873580001	1 X

Table 2.9.10 – *Chipmentation wash buffer B.* High salt wash buffer.

Reagent	Supplier, cat. #	Final concentration
Tris-HCl pH 7	Invitrogen, AM9850G	10 mM
EDTA pH 8.0	Invitrogen, AM9260G	1 mM
EGTA pH 8.0	Alpha Aesar, J60767	0.5 mM
Triton X-100	Sigma, T8787	1%
SDS (20%)	Invitrogen, AM9820	0.1%
Na Deoxycholate (5%)	Invitrogen, AM9740	0.1%
NaCl	Invitrogen, AM9760G	500 mM
Nuclease free H_2O	Sigma, W4502-1L	-
cOmplete PIC	Sigma, 11873580001	1 X

Table 2.9.11 – *Chipmentation wash buffer C.* LiCl salt wash buffer.

Reagent	Supplier, cat. #	Final concentration
Tris-HCl pH 7	Invitrogen, AM9850G	10 mM
EDTA pH 8.0	Invitrogen, AM9260G	1 mM
EGTA pH 8.0	Alpha Aesar, J60767	0.5 mM

Triton X-100	Sigma, T8787	1%
SDS (20%)	Invitrogen, AM9820	0.1%
Na Deoxycholate (5%)	Invitrogen, AM9740	0.1%
LiCl	Sigma, 62476-100G	250 mM
Nuclease free H_2O	Sigma, W4502-1L	-
cOmplete PIC	Sigma, 11873580001	1 X

Table 2.9.12 – ***[1X] tagmentation buffer.***

Reagent	Supplier, cat. #	[Final]
Tris pH7	Invitrogen, AM9850G	10 mM
$MgCl_2$	Invitrogen, AM9530G	5 mM
Dimethylformamide	Sigma, 227056-100ML	10%
Nuclease free H_2O	Sigma, W4502-1L	-

Table 2.9.13 – ***Chipmentation indexing PCR reactants.***

Reagent	Supplier, cat. #	Volume (µL)
NEBNext High-Fidelity [2x] PCR Master Mix	New England Biolabs, M0541	25
[25 µM] Nextera Primer 1.1 (universal)	-	0.25
[25 µM] Nextera Primer 2.x (barcode)	-	0.25
Transposed DNA	-	24.5

Table 2.9.14 – ***Chipmentation NEBnext High-Fidelity PCR cycling conditions.***

Step	Temp (°C)	Time	Cycles
Adapter end repair	72	5 min	
De-crosslink chromatin	95	5 min	
Denaturation	98	10 sec	x 12
Annealing	63	30 sec	
Extension	72	60 sec	
Final Extension	72	1 min	
Hold	4	∞	

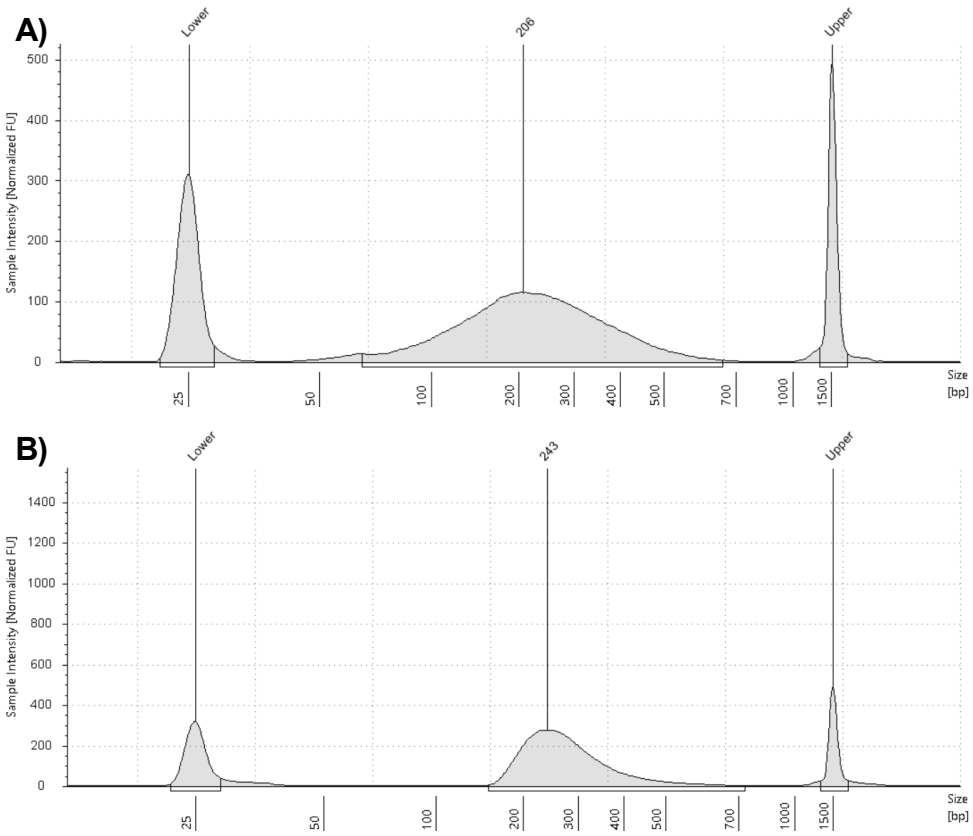

Fig. 2.9.2 – *Chipmentation Quality Control*. Tapestation (Agilent) traces of chipmentation samples after sonication (**A**) and after indexing (**B**).

2.9.3 – CTCF ChIP-seq

CTCF ChIP-seq was performed with the Merk ChIP Assay Kit (Merk Millipore, 17-295), with the following amendments. 10 x 10^6 cells were fixed in 1% formaldehyde for 10 minutes, quenched with 125 mM glycine, then snap frozen and stored at -80°C. For ChIP, pellets were resuspended in ChIP lysis buffer (Merk Millipore, 20-163), incubated for 10 minutes, room temperature, then sonicated with the Bioruptor Pico (Diagenode) for 4 minutes (8 cycles). Sonicates were centrifuged at 13,000 g, 10 minutes, 8°C, to pellet cell debris. Supernatants, containing the chromatin, were diluted in 2 mL ChIP dilution buffer (**Table 2.9.15**). 100 µL (5%) was taken for ChIP input. Remaining sample was immunoprecipitated overnight with 10 µL anti-CTCF Rabbit Polyclonal antibody (Sigma Aldrich, 07-729), 4°C, rotating.

Samples were incubated with 60 µL Protein A agarose slurry (Merk Millipore, 16-157c), 1 hour, 4°C, then washed with low salt, high salt, LiCl, and TE wash buffers, before eluting in 250 µL elution buffer (**Table 2.9.16**). At this point, inputs were thawed, diluted in elution buffer, and taken forward as samples. Samples were de-crosslinked at 65°C, 4 hours, treated with RNase (Roche: 12849100), 37°C, 60 minutes, and proteinase K (BioLabs: P81075), 45°C, 60 minutes. Immunoprecipitated (IP) DNA was purified by phenochloroform extraction, checked with D1000 Tapestation (Agilent), and quantified with dsDNA Qubit. Immunoprecipitation efficiency was verified by serial dilution qPCR with Fast SYBR Green (Applied Biosystems, 4385612) and the primers in **Table 2.9.17.**

An approximately equal mass of input and IP DNA was taken for indexing (<1 µg) with NEBNext Ultra II DNA Library Prep Kit (New England Biolabs: NEB: 7645S/L), according to the manufacturer's protocol, with the following notes. Indexing PCR was performed for 7-11 cycles (depending on input DNA concentration) using NEBNext Mulitplex Oligos (New England Biolabs: E7335S/L and E7500S/L). Indexed libraries were analysed with D1000 Tapestation (Agilent) and quantified using the KAPA Library Quantification Complete Kit (Roche: KK4824).

Table 2.9.15 – *ChIP dilution buffer*.

Reagent	Supplier, cat. #	Final concentration
Tris-HCl pH 7	Invitrogen, AM9850G	16.7 mM
EDTA pH 8.0	Invitrogen, AM9260G	1.2 mM
Triton X-100	Sigma, T8787	1.1%
SDS (20%)	Invitrogen, AM9820	0.01%
NaCl	Invitrogen, AM9760G	167 mM
Nuclease free H_2O	Sigma, W4502-1L	-
cOmplete PIC	Sigma, 11873580001	1 X

Table 2.9.16 – ***ChIP elution buffer.***

Reagent	Supplier, cat. #	[Final]
SDS	Invitrogen, AM9820	1%
$NaHCO_3$	ThermoFisher, 123360010	0.1 M
NF-H2O	Sigma, W4502-1L	-

Table 2.9.17 – ***Primers to quantify ChIP efficiency***, using Fast SYBR Green qPCR.

Primer	Control	Primer direction	Primer sequence (5'-3')
NPRL3	Negative	Fwd	CAGAGCTCCCAGACAACCAG
		Rev	AGCACACCCAGGGTTCTCTA
Hs-38	Positive	Fwd	TGAGAAGGCTGGCCTTTGAG
		Rev	CCCAGGGAATGAATGCCAGT

2.9.4 – NG Capture-C

Next-generation Capture-C was performed as previously described (**Davies *et al.*, 2016; Cornell *et al.*, in revision**). 10 x 10^6 $Ter119^+$ APH-spleen cells were used per biological replicate. 3C libraries were prepared with DpnII restriction enzyme (New England Biolabs, R0543M). Illumina TruSeq adaptors were added using the NEBNext Ultra II DNA Library Prep Kit for Illumina (Illumina, E7645) according to the manufacturer's instructions. Capture enrichment was performed using NimbleGen SeqCap EZ Hybridization and Wash Kit (Roche, 05634261001), NimbleGen SeqCap EZ Accessory Kit v2 (Roche, 07145594001), and previously published custom biotinylated DNA oligonucleotides (R1 and HS-38 viewpoints (**Hanssen *et al.,* 2017**); α-globin promoters viewpoints (**Davies *et al.,* 2016**); **Table 2.9.18**). Indexed libraries were quantified using the KAPA Library Quantification Complete Kit (Roche: KK4824).

Table 2.9.18 – ***Genomic coordinates of Capture C probes.***

Probe	mm9 coordinates
HS38	Chr11:32138079-32139000
HS39	Chr11:32137176-32137426
R1	Chr11:32145277-32146568
R2	Chr11:32151060-32151883
Hba-a1	Chr11:32182969-32183821
Hba-a2	Chr11:32195804-32196638
Hbb-b1	Chr7:110961967-110962817
Hbb-b2	Chr7:110975976-110976456
Scl25A37	Chr14:69902454-69903469

2.10 – NGS data analysis

For all NGS experiments, indexed samples and inputs were pooled as a 4 nM library and sequenced with a high-output v2 75-cycle (or 300-cycle for Capture-C) kits on the Illumina NextSeq platform. For ATAC, chipmentation, and ChIP I aimed for a sequencing depth of ~20-40 million reads per sample. For NG Capture-C, 250k reads per viewpoint per sample. Actual sequencing depth was variable. Analyses were performed by the author unless otherwise specified.

2.10.1 – ATAC, ChIP, and chipmentation analysis

Fastq file quality was assessed using FASTQC (Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data, n.d.), and the reads aligned to mm9 (Δθ1-θ2-44-48) or mm10 (KLF1$^{StrepII\text{-}FKBP}$) using bowtie (**Langmead *et al.*, 2009**). Non-aligning reads were trimmed using Cutadapt trimgalore (**Martin, 2011**) and then realigned to the mm9/10 genome using bowtie. All reads which still failed to align were extracted, and flashed using FLASH (**Magoč and Salzberg, 2011**), before realignment to the mm9/10 genome using bowtie. All files containing successfully aligning reads were concatenated and aligned to the mm9/10 genome together using bowtie. Alignment SAM files were then filtered, sorted, and PCR duplicates removed, using SAMtools (samtools view, sort, and rmdup) (**H. Li *et al.*, 2009**). The resultant BAM file was indexed using SAMtools index and converted to a bigwig file using deepTools

bamCoverage (**Table 2.10.1**; **Ramírez *et al.*, 2016**). Biological and/or technical replicate bigwig files were merged using wiggletools mean (**Zerbino *et al.*, 2014**). All bigwigs in this thesis were visualised with the University of California Santa Cruz (UCSC) genome browser (**Kent *et al.*, 2002**). PDFs of tracks were downloaded from UCSC and annotated in Adobe Illustrator.

Table 2.10.1 – *bamCoverage settings to create bigwig files* for UCSC visualisation or Lanceotron peak calling.

Application	Normalisation	smoothLength	extendReads
UCSC	RPKM	300	No
Lanceotron	RPKM	-	Yes

2.10.2 – NG Capture-C analysis

Capture-C data were analysed using an in-house pipeline developed by Jelena Telenius (WIMM) described previously (**Davies *et al.*, 2016**). In brief, reads were quality controlled, mapped to the mm9 genome, capture-containing reads identified, interaction fragments reported, and visualisation files generated. A second in-house pipeline, CaptureCompare developed by Damien Downes (WIMM), was used to generate the Capture-C interaction comparison profiles (*https://github.com/Hughes-Genome-Group/CaptureCompare*). In brief, unique interactions were normalised to cis interactions, averaged (n = 3), and the difference between the means calculated. For visualisation, means and standard deviations were binned (150 bp) and smoothed with a sliding window (3 kb).

2.10.3 – KLF1 ChIP comparative analyses

Comparative ChIP analyses and visualisations were generated with assistance from Lucy Cornell (WIMM). In brief, mean-averaged bigwigs were peak called using Lanceotron (**Hentges *et al.*, 2022**) and individually thresholded by total peak score (**Table 2.10.2**). Called peaks were proximity merged using bedtools merge (**Quinlan and Hall, 2010**). For motif analysis, all peaks within 50 bp were merged. For Venn diagrams, heatmaps, and UCSC visualisation, all peaks within 1 kb were merged to avoid repeated calling of broad peaks.

To generate Venn diagrams, bedtools intersect was used to make pair-wise comparisons and plotted with RStudios eulerr package. To generate PWM motif graphics, FASTA sequences under called peaks were retrieved with bedtools getfasta and motif analysis performed using the MEME suite (**Bailey *et al.*, 2015**). Heatmaps were generated using deepTools computeMatrix and plotHeatmap (**Ramírez *et al.*, 2016**).

Table 2.10.2 – ***Total peak score thresholds for Lanceotron-called peaks.***

Dataset	seq	All peaks	Peak threshold	Merge all peaks within:	Peaks after threshold/merging
Holliman	paired	**1110**	0.1		**174**
Mukherjee	single	**22764**	0.01	1 kb	**574**
Huang	paired	**3978**	0.5		**1210**

2.11 – AlphaFold

Protein structure predictions were generated with AlphaFold 3 (**Abramson *et al.*, 2024**) and visualised with PyMol (ver. 2.4.1). Protein and DNA sequence details in **Table 2.11.1** and **Table 2.11.2**.

Table 2.11.1 – *Amino acid sequences (single letter code) used as input for AlphaFold 3.* The recombinant tag is highlighted in bold.

Protein Name	Sequence (N-C)	Reference
KLF1	MASAETVLPSISTLTTLGQFLDTQEDFLKWWRSEETQD LGPGPPNPTGPSLHVSLKSEDPSGEDDERDVTCAWD PDLFLTNFPGSESPGTSRTCALAPSVGPVAQFEPPESL GAYAGGPGLVTGPLGSEEHTSWAHPTPRPPAPEPFVA PALAPGLAPKAQPSYSDSRAGSVGGFFPRAGLAVPAA PGAPYGLLSGYPALYPAPQYQGHFQLFRGLAAPSAGG TAPPSFLNCLGPGTVATELGATAIAGDAGLSPGTAPPK RSRRTLAPKRQAAHTCGHEGCGKSYSKSSHLKAHLRT HTGEKPYACSWDGCDWRFARSDELTRHYRKHTGHRP FCCGLCPRAFSRSDHLALHMKRHL	UniProt (ID: P46099)
KLF1$^{\text{StrepII-FKBP}}$	MASAETVLPSISTLTTLGQFLDTQEDFLKWWRSEETQD LGPGPPNPTGPSLHVSLKSEDPSGEDDERDVTCAWD PDLFLTNFPGSESPGTSRTCALAPSVGPVAQFEPPESL GAYAGGPGLVTGPLGSEEHTSWAHPTPRPPAPEPFVA PALAPGLAPKAQPSYSDSRAGSVGGFFPRAGLAVPAA PGAPYGLLSGYPALYPAPQYQGHFQLFRGLAAPSAGG TAPPSFLNCLGPGTVATELGATAIAGDAGLSPGTAPPK RSRRTLAPKRQAAHTCGHEGCGKSYSKSSHLKAHLRT HTGEKPYACSWDGCDWRFARSDELTRHYRKHTGHRP FCCGLCPRAFSRSDHLALHMKRHL**PGGSGGGGSGSA** **WSHPQFEKGGGSGGGSGGSAWSHPQFEKGSGGGG** **SGGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGK** **KVDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSVG** **QRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE**	Custom

Table 2.11.2 – *Hbb-b1 promoter DNA sequence used as input for AlphaFold 3*. The canonical KLF1 binding motif is highlighted in bold. Equivalent to the 21-nucleotide oligo used for gel shift (EMSA) assays in the discovery of KLF1.

DNA oligo	Sequence (top strand)	Reference
Hbb-b1 promoter	5'-AGCTAG**CCACACCCT**GAAGCT-3'	Miller and Bieker, 1993

2.12 – Data visualisation

Unless otherwise stated, all graphs in this thesis were made with RStudio (ver. 2023.12.1) and the ggplot2 package. Original schematics were made with BioRender.com and Microsoft PowerPoint. Figures were assembled in Adobe Illustrator (2022) or Microsoft PowerPoint.

2.13 – Statistical analyses

All statistical analyses were performed in RStudio, unless otherwise specified.

2.13.1 – Power calculations

First, a power calculation was performed to determine the minimum sample size required to detect an anaemic phenotype in the mouse haematology data. As the Δθ1-θ2-44-48 and KLF1$^{StrepII\text{-}FKBP}$ models could be expected to cause anaemia by impairing adult globin expression, haemoglobin concentration (HGB) was chosen to calculate sample size. Based on previously reported mean female HGB of 14.9 ± 1.0 g/dL (**Raabe *et al*., 2011**) and with a view to detect anaemia defined as HGB 25% below WT mean HGB (**Hay *et al*., 2016**), a minimum power (likelihood of true positive results) of 0.8 and a P = 0.05, I calculated a required sample size of n = 3 per treatment group. Calculations were performed with G*Power 3.1 software. Haematological analysis was performed on more mice if available, to improve power above 0.8.

2.13.2 – Two-way ANOVA

Two-way ANOVA with replications was performed where appropriate to detect significant differences between mean values. For all analyses, homoscedasticity of residuals (assumption of normal variance) was verified by Shapiro-Wilk normality test. Where explanatory variables were found to have a significant (P-threshold ≧ 0.05) effect on the dependent variable, Tukey's Honest Significant Difference (Tukey HSD) post-hoc test was performed. Tukey HSD P-values are annotated on figures where appropriate.

2.13.3 – Chi-Squared

Chi-squared Goodness of Fit tests were used to calculate whether the ratio of genotypes of Δθ1-θ2-44-48 and KLF1$^{StrepII\text{-}FKBP}$ pups at weaning fit a Mendelian ratio of inheritance.

Chapter 3 – Defining the 3' boundary of the α-globin sub-TAD

3.1 – Introduction

3.1.1 – TADs and their regulation of gene expression

Mammalian genomes are organised into hierarchical structures, from chromosome territories to specific enhancer-promoter interactions. A recurring question concerns the relationship between genome structure and function. From 2002 onwards, the development of assays to measure chromosome conformation (**Dekker *et al*., 2002**) has enabled researchers to investigate genome structure at ever-increasing resolutions (reviewed in **Goel and Hansen, 2020).** A major breakthrough in understanding higher order chromatin structure emerged from the development of Hi-C which led to the definition of Topologically Associating Domains (TADs; **Lieberman-Aiden *et al*., 2009; Dixon *et al*., 2012; Nora *et al*., 2012; Sexton *et al*., 2012**). These are ~100-1000s kb sections of DNA in which sequences within the TAD interact with ~2-fold higher frequency than with sequences beyond the TAD. TADs are visible on Hi-C maps as triangular domains of increased interaction frequency, and they represent a statistical enrichment of interactions between a defined segment of contiguous DNA (**Figure 3.1A**).

TADs are thought to form due to cohesin loading across the genome and translocating across chromatin – a process known as chromatin loop extrusion (**Fig. 3.1B**). Once loaded, cohesin bi-directionally extrudes the DNA fibre, bringing the two extruded strands into close physical proximity. When cohesin reaches an insulator element, the cohesin stalls and eventually disengages from the DNA. This extrusion and stalling is evidenced by the characteristic “dots and

stripes" seen emanating from TAD boundaries in Hi-C maps (**Figure 3.1A**). These loop extrusion events are thought to be required for TAD formation.

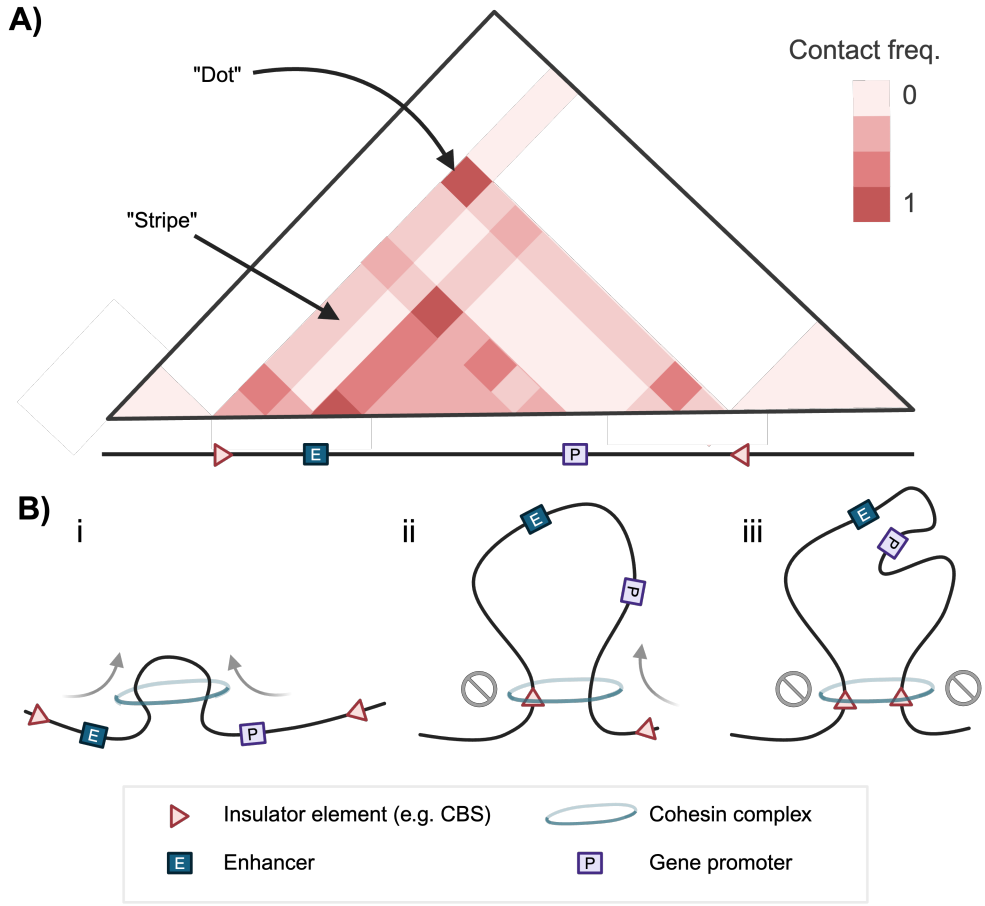

Fig. 3.1 – *Topologically Associating Domains (TADs) are domains of increased interaction frequencies formed by cohesin loop extrusion.* A) Schematic of a TAD in a hypothetical heatmap. TADs can be detected by proximity ligation assays, such as Capture-C or other 3C assays. The enriched DNA-DNA interactions can be visualised in contact heatmaps. "Dots" are where two regions are likely to spend more time in closer proximity, such as between enhancers and their cognate promoters or between two TAD-flanking insulator elements. "Stripes" are artefacts of cohesin tracking when cohesin stalls at one boundary element but continues to track the other DNA strand, such as in B-ii. **B)** Schematic representation of cohesin-mediated chromatin loop extrusion. **B-i)** CTCF loads onto DNA and bi-directionally extrudes the chromatin fibre. **B-ii)** Extrusion can be stalled by blocking factors, such as CTCF. **B-iii)** Together this brings the contiguous, boundary-delineated DNA fibre into closer physical proximity. This process is thought to facilitate the enhancer-promoter contacts implicated in gene transcription.

3.1.2 – The insulator element: CTCF

The best characterised insulator elements in mammals are sequences that bind CCCTC-binding factor (CTCF). CTCF is an 82 kDa transcription factor with known roles in gene activation and repression, as well as in TAD insulation. Structural and functional studies have shown that translocation of cohesin is blocked when it interacts with the N-terminus of the bound CTCF protein. Thus, CTCF acts as an orientation dependent block to the translocation of cohesin (**Fig. 3.2**) and plays an important role in delimiting TADs (reviewed in **Hansen *et al.*, 2020**). Notably, TAD boundaries are often enriched for clusters of CTCF bound sites (CBSs) (**Dixon *et al.*, 2012; Kentepozidou *et al.*, 2020**) and studies have shown that, at some loci, multiple CBSs are required for effective TAD boundary insulation (**Despang *et al.*, 2019; Jia *et al.*, 2020; Huang *et al.*, 2021**).

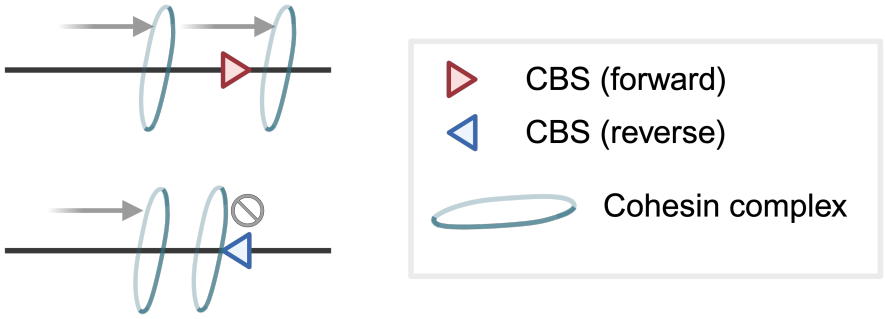

Fig. 3.2 – *CTCF blocks cohesin loop extrusion in an orientation-dependent manner.*

3.1.3 – Sub-TADs

Initial Hi-C studies suggested TADs to be largely invariant across different cell types and species (**Dixon *et al.*, 2012; Nora *et al.*, 2012**). Further improvement of the resolution of Hi-C assays to the ~1 kb scale revealed the existence of sub-TADs: smaller (10s-1000s kb), less-well insulated contact domains nested within TADs and also largely delimited by CTCF (**Phillips-Cremins *et al.*, 2013; Rao *et al.*, 2014**). In contrast to TADs, sub-TADs show considerable structural variation during differentiation and development. There is increasing evidence that cell-specific sub-TADs are formed via increased recruitment of cohesin at poised and active enhancers (**Kagey *et al.*, 2010; Faure *et al.*, 2012; Hua *et al.*, 2021;**

Rinzema *et al.*, 2022; Geogiades *et al.*, 2023) and that the loops resulting from cohesin chromatin extrusion are delimited by oppositely-orientated CTCF insulators flanking the sub-TAD (**Rao *et al.*, 2014; de Wit *et al.*, 2015; Guo *et al.,* 2015; Vietri Rudan *et al.*, 2015**). Importantly, as well as cell-type specific structures, sub-TADs often contain lineage-specific enhancers and their cognate promoters.

3.1.4 – Investigating the boundaries of the α-globin sub-TAD

The α-globin gene locus serves as a paradigm for the study of the relationship between 3D genome structure and function. In mouse, the α-globin genes and their cognate super-enhancer elements lie within a ~165 kb TAD together with flanking, widely expressed genes (**Introduction Fig. 1.4**). During erythroid differentiation, an erythroid-specific, ~70 kb sub-TAD emerges that encompasses the α-globin genes and super-enhancer, excluding the flanking genes. However, the formation and function of this domain are yet to be fully understood.

Recent work from our lab suggests that cohesin-mediated loop extrusion plays an important role in normal α-globin gene regulation within the sub-TAD. Firstly, we have demonstrated that cohesin and its loading protein, NIPBL, are enriched at the active enhancers in erythroid cells (**Hua *et al.*, 2021**). Likewise, acute degradation of cohesin causes a reduction in α-globin expression, especially in earlier stages of erythropoiesis (**Stolper *et al.*, 2023**). We have also shown that this cohesin loop extrusion and super-enhancer activity can be delimited by convergently orientated CBSs (**Hanssen *et al.*, 2017; Kassouf *et al.*, 2022**). Finally, we have demonstrated that blocking cohesin translocation with a CBS insertion between the α-globin promoters and enhancers reduces enhancer-promoter contact probability and α-globin transcription (**Stolper *et al.*, 2023; Tsang *et al.*, 2024**). Together, these findings suggest that loop extrusion within the sub-TAD plays an important role in mediating enhancer-promoter interaction and normal α-globin expression.

The question then arises as to the role of specific CBS insulators in this process? We have previously demonstrated that the 5' boundary of the α-globin sub-TAD is delineated by the CBSs, HS-38, and HS-39 (**Table 3.1**). In **Hanssen *et al.*,**

2017, we demonstrated that specific deletion of HS-38 and HS-39 in a mouse model was sufficient to cause 5' extension of the sub-TAD interaction domain and ectopic transcription of the upstream genes, *Mpg*, *Rhbdf1,* and *Snrnp25*. This suggests that HS-38 and HS-39 CBSs demarcate the 5' sub-TAD boundary.

Table 3.1 – *Summary of the α-locus CTCF binding site (CBS) deletion models to date.* Schematics are of the α-globin sub-TAD (dark blue rectangle), flanking CBS (light blue lines), and deleted CBS (red lines).

Model	Deleted CBS	α:β	Sub-TAD	Flanking gene expression
WT		-	-	-
Δ38-39		No change	5’ extension	5’ ↑
Δθ1		(no data)	(no data)	No change
Δθ2		(no data)	No change	No change
Δθ1-θ2		No change	No change	No change
Δ44-48		No change	Minimal 3’ extension	No change
Δθ1-θ2-44-48		?	?	?

In light of this, I sought to determine whether CBSs are also responsible for delimiting the 3' boundary of the sub-TAD. Experiments by C. Harrold, (**Cornell *et al.*, manuscript in preparation**), generated mouse models with pairwise, CRISPR-Cas9-mediated, specific ablation of the 3' CBSs: θ1/θ2 and HS+44/+48. Interestingly, neither pair of CBS disruptions impacted the α-globin sub-TAD topology or expression of downstream genes, *Sh3pxd2b* and *Ubtd2* (**Table 3.1**). Indeed, there is precedent for CBSs at TAD boundaries to be dispensable for TAD insulation (**Despang *et al.*, 2019**) or to act to strengthen, rather than block, enhancer-promoter interactions (**Guo *et al.*, 2015**). However, there is also

evidence that multiple CTCF sites can act together to insulate TAD boundaries. For example, TAD boundaries are often found to be enriched for clusters of CTCF binding sites (**Dixon *et al.*, 2012; Kentepozidou *et al.*, 2020**). In transgenic mouse experiments, *de novo* TAD insulation strength can be increased by, in some cases even contingent upon, the insertion of multiple ectopic CBSs (**Huang *et al.*, 2021**). Likewise, mouse models with sequential deletion or interference with dCas9 (catalytically “dead” Cas9) of endogenous tandem arrays of CTCF binding sites revealed synergistic/additive topological effects (**Jia *et al.*, 2020**). This led us to hypothesise that the θ1/θ2 and HS+44/+48 CBSs might act synergistically to define the α-globin sub-TAD 3' boundary.

Therefore, we designed a mouse model with a quadruple deletion of the 3' CBSs to test whether θ1, θ2, HS+44, and HS+48 act together to insulate the α-globin sub-TAD domain (**Table 3.1**).

3.1.5 – Aims of Chapter 3

In this chapter, I investigate the extent to which the 3' boundary of the α-globin sub-TAD is delimited by the θ1/θ2 and HS+44/+48 CBSs by characterising the Δθ1-θ2-44-48 mouse model. I first address whether removing these four CBSs alters the normal interaction profile and limits of the sub-TAD. Second, I look at whether removal of these CBSs leads to altered interactions between the super-enhancer elements and neighbouring promoters, and if there is accompanying dysregulation of *Hba-a1/2* and other downstream genes.

Aims:

1. Characterise the erythroid phenotype of the Δθ1-θ2-44-48 mouse model.

2. Establish whether the 3' structural boundary of the α-globin sub-TAD is delimited by the θ1/2 and HS+44/+48 CTCF binding sites.

3. Establish whether the 3' functional boundary of the α-globin sub-TAD is delimited by the θ1/2 and HS+44/+48 CTCF binding sites.

3.2 – Results

3.2.1 – Survival of the Δθ1-θ2-44-48 mouse model

First, the Δθ1-θ2-44-48 mouse model was assessed to determine whether perturbing the four 3' CTCF sites had an adverse effect on mouse survival or phenotype, for example by causing anaemia.

All hom Δθ1-θ2-44-48 adult mice were morphologically normal and indistinguishable from WT by visual inspection. The genotypes of progeny from heterozygous crosses were counted to determine whether the Δθ1-θ2-44-48 allele was segregating within the population in the expected Mendelian ratios. If the Δθ1-θ2-44-48 allele were to have no effect on survival to weaning, then we can expect the ratio of adult WT:het:hom to be 1:2:1. All available (N = 22) pups were genotyped after weaning. This gave 1 WT, 14 het, and 7 hom pups, which did not significantly differ from the expected Mendelian ratio by Chi-squared Goodness of Fit test (*P* = 0.0859) (**Fig. 3.3**). I concluded the Δθ1-θ2-44-48 genotype has no significant effect on mouse survival to adulthood.

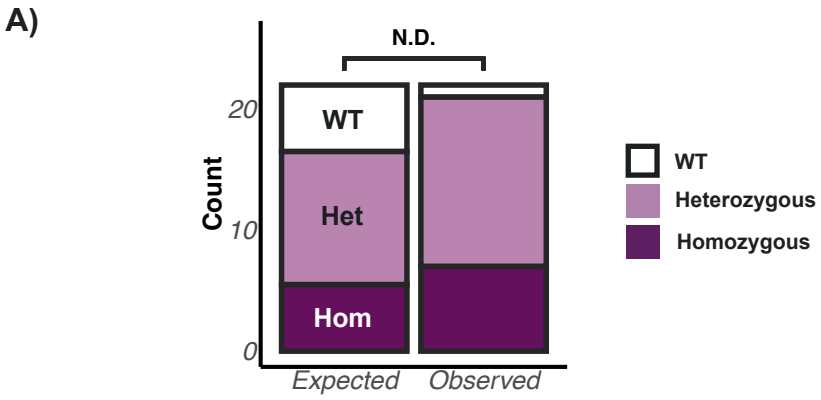

Figure 3.3 – *Δθ1-θ2-44-48 mice are born in Mendelian ratios*. A) Observed counts at weaning age of Δθ1-θ2-44-48 homozygous (hom), heterozygous (het), and wildtype (WT) mice, from heterozygous crosses. Counts do not significantly differ from an expected Mendelian ratio, as calculated by Chi-squared Goodness of Fit test (N = 22, X^2 = 4.9091, *d.f.* = 2, *P-value* = 0.0859).

3.2.2 – Δθ1-θ2-44-48 mouse red cell haematology

Our laboratory has previously observed that mice with moderate anaemia may not display an obvious anaemic phenotype (**Hay *et al*., 2016**). To ensure that Δθ1-θ2-44-48 adult mice were haematologically normal, red blood cell haematology, reticulocyte counts, and blood films were measured, as outlined in **Methods 2.4**.

First, I performed power calculation to verify the minimum sample size required to detect an anaemic phenotype in the mice, which came to n ≧ 3 per treatment group (**Methods 2.13.1**). Peripheral blood was then obtained from all available adult mice (N = 19; WT: n = 3, Het: n = 12, Hom: n = 4) and measured for haemoglobin level (HGB; g/dL), red blood cell count (RBC; %), mean corpuscular volume (MCV; μm^3), and mean corpuscular haemoglobin (MCH; pg). All measured parameters were within a normal range (**Fig. 3.4A**). Likewise, mean reticulocyte counts were within a normal range, with no significant effect of genotype (**Fig. 3.4B**). Peripheral blood films were normal (**Fig. 3.4C**). This suggests that Δθ1-θ2-44-48 mice are not under erythropoietic or anaemic stress. I conclude that there is no effect of the Δθ1-θ2-44-48 genotype on mouse haematology beyond normal background variation.

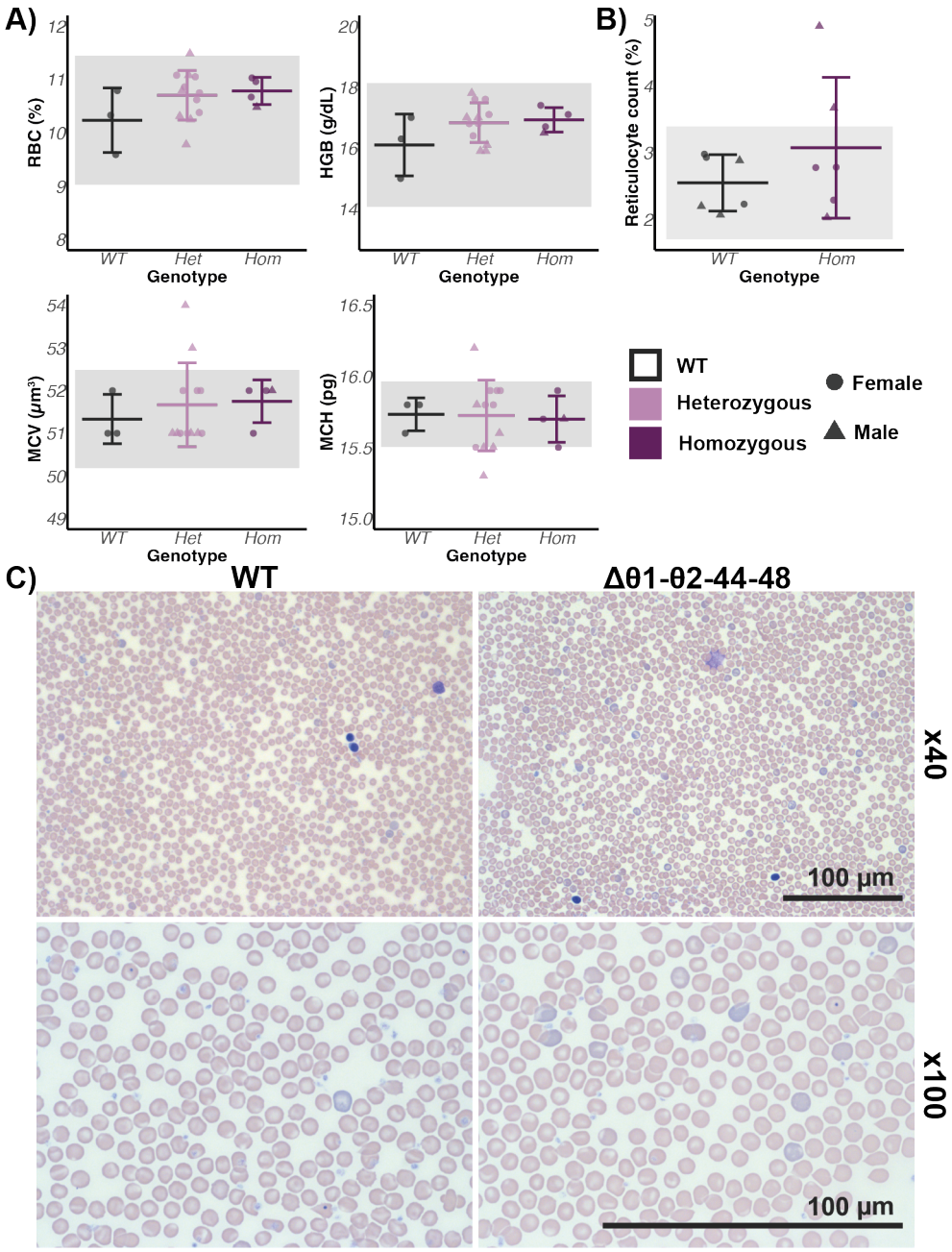
A)
RBC (%)
HGB (g/dL)
MCV (μm³)
MCH (pg)
WT
Het
Hom
Genotype
B)
Reticulocyte count (%)
WT
Hom
Genotype
WT
Heterozygous
Homozygous
Female
Male
C)
WT
Δθ1-θ2-44-48
x40
x100
100 μm
100 μm

Figure 3.4 – *Δθ1-θ2-44-48 mice are haematologically normal*. A) Mixed-sex population mean values for: red blood cell count (RBC; %), mean haemoglobin (HGB; g/dL), mean corpuscular volume (MCV; μm^3), and mean corpuscular haemoglobin (MCH; pg) in WT, het, and hom adult peripheral blood. N = 19; WT: n = 3, Het: n = 12, Hom: n = 4. Extended data in **Appendix Table A3.1**. **B)** Mean percentage reticulocyte counts for WT and hom, mixed-sex adult peripheral blood. N = 6. **A + B)** No significant effect of genotype was detected by two-way ANOVA ($P > 0.05$). Error bars = ± 1 standard deviation. Grey shading indicates the normal range per parameter, calculated as ± two standard deviations from the WT mean (**Raabe *et al.*, 2011**). **C)** Representative peripheral blood films from adult, male, WT, and hom Δθ1-θ2-44-48 mice. Wright-Giemsa staining dyes basic structures (e.g. haemoglobin) red/pink and acidic structures (e.g. nuclei and RNA) blue/purple. Imaged at X40 and X100 magnification. Scale bars = 100 μm.

3.2.3 – The effect of Δθ1-θ2-44-48 on α-globin sub-TAD organisation

Next, I investigated the molecular phenotype of the Δθ1-θ2-44-48 model to determine how deletion of the 3' CBSs affects the α-globin sub-TAD boundaries.

Adult mice can be treated with acetylphenylhydrazine (APH) to induce haemolytic anaemia and extramedullary haematopoiesis. This results in massive enrichment of splenic erythropoiesis, such that one spleen may yield ~0.3-1 x10^9 erythroblasts. I harvested erythroblasts from the spleens of APH-treated, adult Δθ1-θ2-44-48 mice and processed for: ATAC-seq assay for average chromatin accessibility; ChIP-seq for average CTCF chromatin binding; Capture-C to detect the relative probability of interaction between a specific DNA locus – the "capture viewpoint" – and any other genomic position, averaged over several million nuclei.

CTCF ChIP-seq confirmed the specific and complete abrogation of CTCF binding from each of the deleted CBSs in the Δθ1-θ2-44-48 model (**Fig. 3.5A**). The CTCF binding profile is unchanged at all other loci across the α-globin TAD. Similarly, the chromatin accessibility profile of the locus is unchanged in Δθ1-θ2-44-48 mice, barring the expected loss of the small and discrete peaks of chromatin accessibility that ordinarily lie over the θ1, θ2, HS+44, and HS+48 CBSs in WT (**Fig. 3.5A**).

I then analysed the effect of Δθ1-θ2-44-48 on sub-TAD structure by Capture-C from viewpoints taken at or close to the boundaries of the α-globin sub-TAD, HS-38, and *Hba-a1/2* (**Fig. 3.5B, C**). From this we see that both interaction profiles are largely unchanged in Δθ1-θ2-44-48 compared to WT. From any viewpoint within the TAD, in WT (black), contact probability across the 3' boundary decreases in a stepwise fashion. The sharpest decrease in interaction frequency is immediately downstream of the *Hba-a2* gene. There is then a sharp but smaller decrease in intra-sub-TAD interaction frequency following the HS+44/HS+48 pair of CBSs. Finally, there is a minor decrease in interaction frequency at the extreme 3' HS+65/HS+66 CBSs.

In the Δθ1-θ2-44-48 model (purple) compared to WT, the interaction frequency of the HS-38 site decreases marginally at the Δθ1 site. There is a small loss of

interaction, compared to WT, surrounding the ΔHS+44/HS+48 sites (note the sites themselves and intervening sequences are deleted and so not plotted). Meanwhile, there is a gain of interaction frequency at the intact, downstream HS+65/HS+66 CBSs. There is a very slight extension of interaction frequency of HS-38 with the DNA sequence extending 3' of the α-globin TAD domain. Due to the signal exclusion zones imposed during data analysis, the Capture-C signal at the θ2 position, which lies close to the *Hba-a2* gene/viewpoint, is excluded from the plotting view. Overall, the interaction profile of HS-38, the known boundary element at the 5' of the α-globin sub-TAD, is largely unchanged by Δθ1-θ2-44-48, save for a slight “spillover” of interaction frequency in the 3' direction.

Likewise, there is a similar marginal change in interaction profile from the *Hba-a1/2* combined Capture-C viewpoint. The 5' boundary of the interaction frequency of *Hba-a1/2* remains intact at the HS-38/HS-39 loci and the 3' interaction boundary is still sharply defined as just 3' of the *Hba-a2* gene. Similar to the HS-38 Capture-C viewpoint, there is a moderate gain of interaction frequency with the HS+65/HS+66 CBSs. There is minimal interaction spillover to the sequence 3' of the TAD.

Taken together, these reciprocal Capture-C viewpoints indicate that the α-globin sub-TAD and TAD domains are largely unchanged by the Δθ1-θ2-44-48 model.

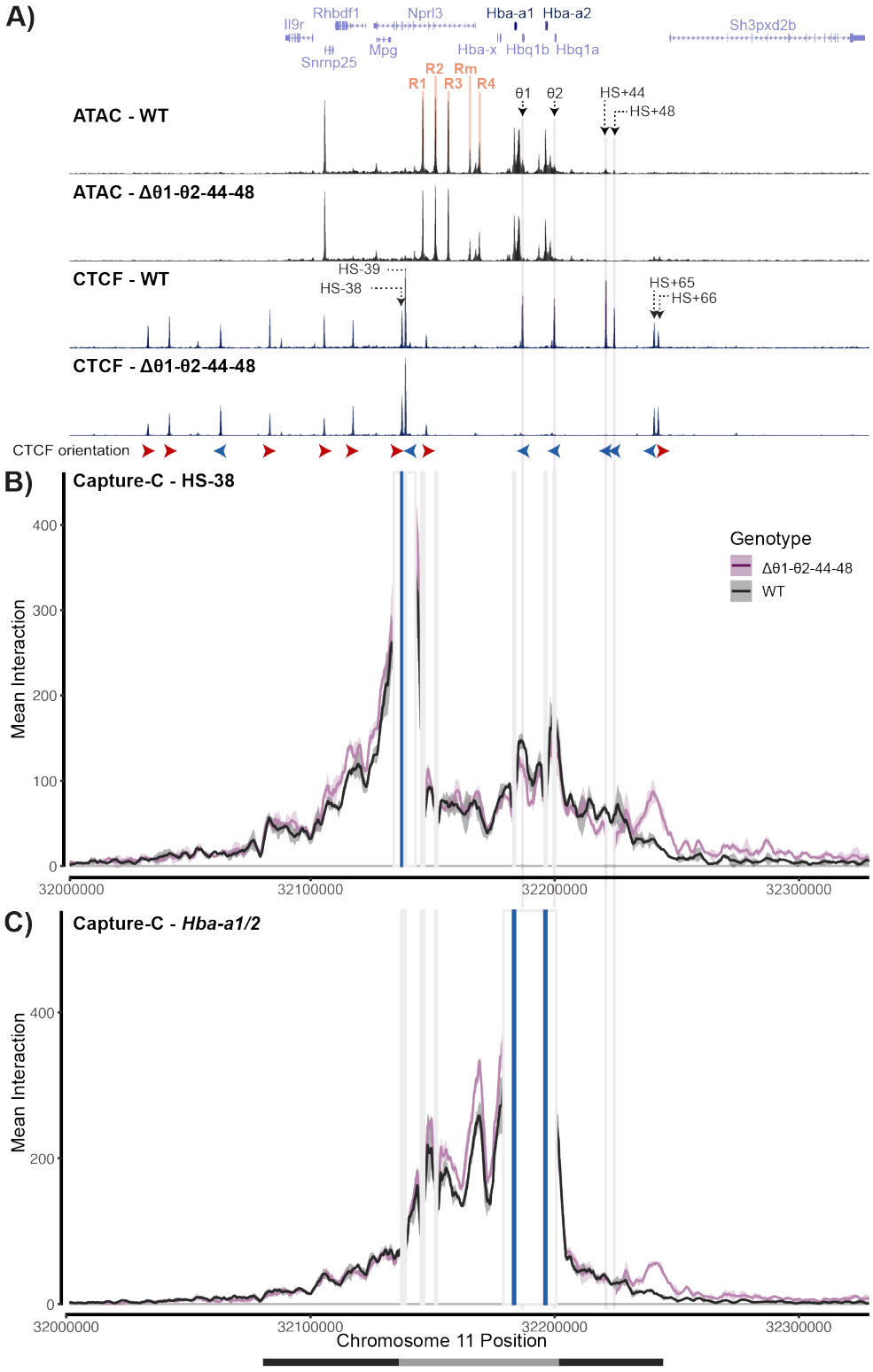

Figure 3.5 – *Δθ1-θ2-44-48 causes minor extension of the α-globin sub-TAD boundary.* **A)** Average ATAC-seq (dark grey; N = 3) and CTCF ChIP-seq (navy; N = 2) traces from WT and homozygous Δ44-48-θ1-θ2, APH-treated adult mouse spleen. Normalised by RPKM. Pertinent CTCF sites annotated in black. α-globin super enhancer elements (R1-4 and Rm) annotated in orange. UCSC gene annotation in blue, above. CTCF binding site orientation annotated with red (forward-oriented) and blue (reverse-oriented) arrows, below. **B)** Capture-C profile from the HS-38 viewpoint. **C)** Capture-C

profile from the combined *Hba-a1* and *Hba-a2* viewpoints. All Capture-C was performed in biological triplicate from APH-treated, adult mouse, Ter119^{+} spleen. Blue vertical bar indicates Capture-C viewpoint(s). White vertical box is the plotting exclusion zone, ± 1 kb from the Capture-C viewpoint(s). Light grey vertical bars indicate co-captured probe loci and corresponding plotting exclusion zones. Deleted CBSs are highlighted in pale pink. Below Capture-C profiles, thick black bar indicates the position of the TAD domain, present in erythroid and erythroid-precursor cells (chr11:32,080,00032,245,000). The thick grey bar represents the position erythroid-specific α-globin sub-TAD domain (chr11:32,136,000-32,202,000).

3.2.4 – The effect of Δθ1-θ2-44-48 on enhancer-promoter interactions and 3' gene expression

Having established that the structure of the α-globin sub-TAD was largely intact in Δθ1-θ2-44-48 erythroblasts, I next investigated whether the function (i.e. enhancer-promoter insulation effect) of the sub-TAD was perturbed.

I investigated this with Capture-C from the viewpoints of the strongest α-globin enhancer elements, R1 and R2 (**Fig. 3.6**). The interaction profiles of both enhancers were broadly unchanged across the α-globin TAD. The enhancers showed normal interactions with their cognate *Hba-a1/2* gene targets, despite deletion of the θ1 and θ2 CBSs which lie proximal to the *Hba-a1/2* genes. Similar to the interaction profiles from the HS-38 and *Hba-a1/2* viewpoints, the interaction frequency of R1/R2 retains the normal "stepwise" pattern of interaction decrease over the 3' sub-TAD boundary and gains modest increases over the downstream HS+65/HS+66 elements. Crucially, there was no increase in R1/R2 interactions with the downstream DNA sequence or the *Sh3pxd2b* or *Ubtd2* gene promoters.

To be certain whether Δθ1-θ2-44-48 could cause upregulation of the downstream genes, *Sh3pxd2b* and *Ubtd2*, I performed RT-qPCR on spleen erythroblasts from WT and Δθ1-θ2-44-48 mice. *Sh3pxd2b* and *Ubtd2* are ordinarily transcriptionally silent in WT erythroid cells and remained undetectable by RT-qPCR in Δθ1-θ2-44-48 (**Fig. 3.6D**). I also detected no significant change in the relative expression of α:β globin gene transcripts (**Fig. 3.6D**), as could be expected from the normal interaction profile between the enhancers and the *Hba-a1/2* genes.

From this, I conclude that Δθ1-θ2-44-48 has not significantly affected the structural or the functional boundary of the α-globin sub-TAD. This suggests that the θ1, θ2, HS+44, and HS+48 CBSs are not required for the insulation of the active α-globin enhancer elements from the downstream genes in mature erythroblasts.

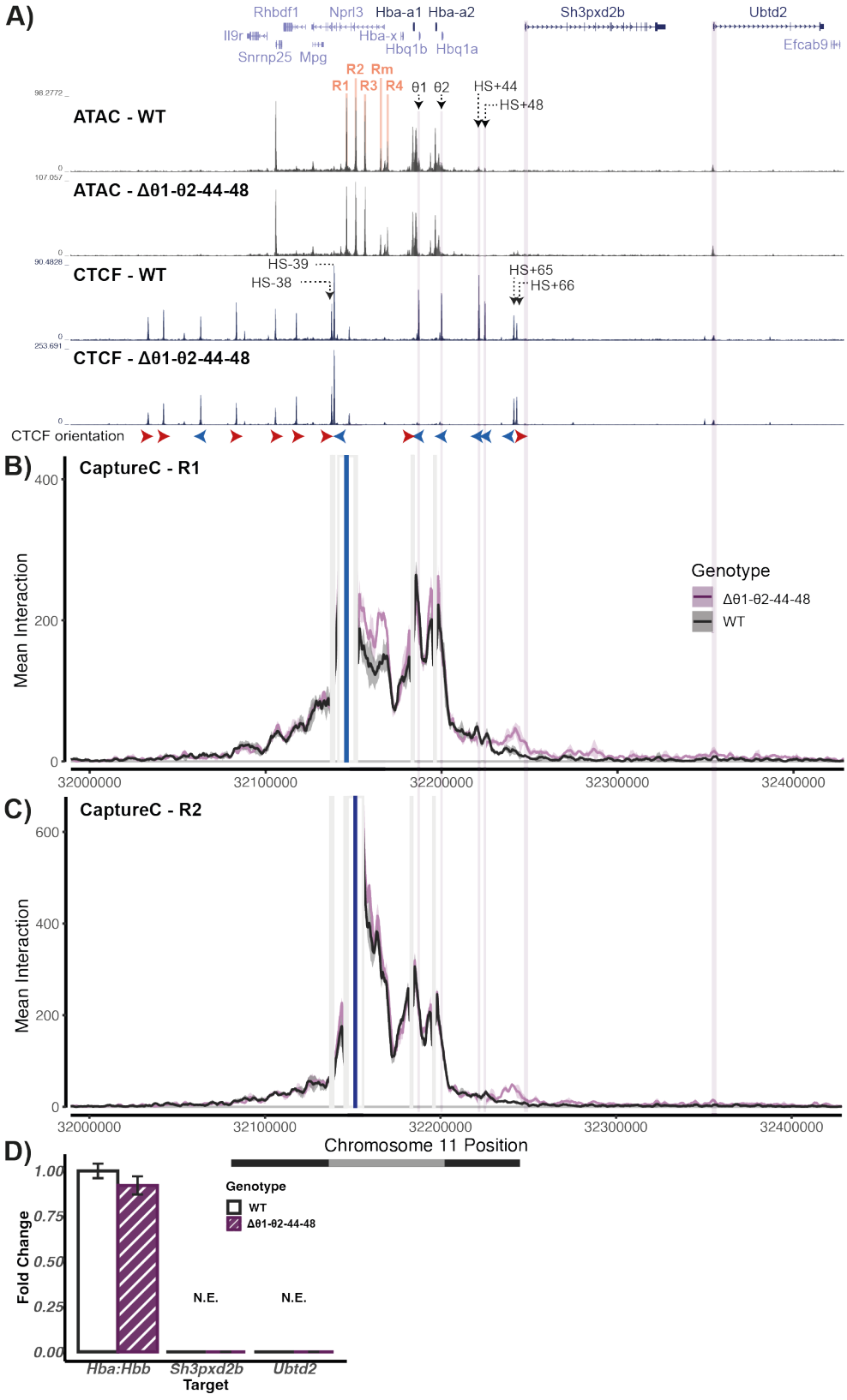

Figure 3.6 – *Δθ1-θ2-44-48 does not affect the enhancer-promoter insulation function of the α-globin 3' sub-TAD boundary.* A) Average ATAC-seq (dark grey; N = 3) and CTCF ChIP-seq (navy; N = 2) traces from wild type (WT) and homozygous Δ44-48-θ1-θ2, APH-treated adult mouse spleen. Normalised by RPKM. Pertinent CTCF sites annotated in black. α-globin super enhancer elements (R1-4 and Rm) annotated in

orange. UCSC gene annotation in blue, above. CTCF binding site orientation annotated with red (forward-oriented) and blue (reverse-oriented) arrows, below. **B+C)** Capture-C interaction frequency profiles from the viewpoint of **B)** the R1 and **C)** the R2 enhancers. In biological triplicate from APH-treated, adult mouse, $Ter119^{+}$ spleen. Chr11:31,988,000-32,430,000. Blue vertical bar indicates Capture-C viewpoint(s). White vertical box is the plotting exclusion zone, ± 1 kb from the Capture-C viewpoint(s). Light grey vertical bars indicate co-captured probe loci and corresponding plotting exclusion zones. Deleted CTCF sites are highlighted in pale pink. Below Capture-C profiles, horizontal black bar indicates the position of the TAD domain, present in erythroid and erythroid-precursor cells (chr11:32,080,00032,245,000). The horizontal grey bar indicates the position of erythroid-specific, α-globin sub-TAD domain (chr11:32,136,000-32,202,000). **D)** RT-qPCR in $Ter119^{+}$ spleen erythroblasts, showing dCT values normalised to *Rps18* housekeeper. Hba:Hbb is the ratio of *Hba-a1/2* to *Hbb-b1* dCT values. “N.E.” indicates no expression detected.

3.3 – Discussion

3.3.1 – θ1/θ2 and HS+44/+48 CTCF binding sites do not delimit the 3' sub-TAD

In this chapter, I have directly addressed my aims to characterise the Δθ1-θ2-44-48 mouse model to determine whether the θ1/θ2 and HS+44/+48 CBSs might act synergistically to define the structural and/or functional 3' boundary of the α-globin sub-TAD.

By Capture-C from strategic viewpoints across the sub-TAD, I determine that Δθ1-θ2-44-48 does not significantly perturb the 3' sub-TAD structural boundary or the functional insulation of the α-globin super-enhancer from downstream genes, *Sh3pxd2b* and *Ubtd2*, in erythroblasts. Rather, the θ1/θ2 and HS+44/+48 CBSs appear to confer a minor boundary effect, with the majority of the sub-TAD insulation remaining proximal to the *Hba-a1/2* genes.

It must be considered that functional redundancy of the downstream HS+65/+66 CBSs could be conserving the sub-TAD 3' boundary, and that a Δθ1-θ2-44-48-65-66 knockout model would reveal a redundant/synergistic boundary effect of all six CBSs. However, as there is a relatively minor decrease in interaction frequency at HS+65/+66 and the largest and steepest decrease in interaction frequency remains local to *Hba-a1/2*, even in the Δθ1-θ2-44-48 model, it seems unlikely that major changes to the 3' sub-TAD boundary would occur in a Δθ1-θ2-44-48-65-66 model.

Further evidence against a major role of HS+65/66 comes from Micro-Capture-C (MCC) assays from the viewpoint of the 5' boundary element, HS-38, in WT mouse erythroid and ES cells by **Hua *et al.*, 2021**. MCC is a high-resolution Capture-C-style assay, providing interaction profiles at base-pair resolution. Hua *et al.* observed, upon erythroid differentiation, strong and punctate increases in interaction frequencies of HS-38 with *Hba-a1/2*, θ1/*Hbq1b*, θ2/*Hbq1a*, HS+44, and HS+48. Meanwhile, HS-38 interaction with HS+65/66 was equally extremely low in both ESCs and erythroid cells, further suggesting that HS+65/66

contributes minimally to erythroid-specific α-globin sub-TAD insulation in the WT context.

Finally, the mouse HS+65/+66 CBSs sequences are not conserved from mouse to human (**Appendix Fig. S3.4**), despite high conservation of the sub-TAD domain structure between human and mouse. Even the core CBS sequence of HS+65 diverges by two nucleotides from the orthologous rat sequence (**Appendix Fig. S3.4**). In conclusion, HS+65/66 appears to lie outside of the sub-TAD interaction domain and contributes minimally to boundary insulation, if at all.

3.3.2 – Remaining putative 3' boundary elements

Rather, it would appear that in Δθ1-θ2-44-48 the major sub-TAD boundary remains local to the α-globin genes themselves. This raises the interesting possibility that some other cis-acting element(s), besides θ1/θ2, is acting as the major 3' sub-TAD boundary. There are three candidate elements: two “θ-minor” CTCF binding sites, the *Hbq1b/a* genes, and the *Hba-a1/2* genes.

Close analysis of CTCF ChIP-seq at the *Hba-a1/2* sites reveals two very small, previously unidentified, CTCF peaks that lie immediately upstream of each of the θ1/θ2 sites and likely correspond to underlying 5'-CCCCCAC-3' motifs (**Fig. 3.7).** These sites were discounted during initial designs as they were not called as significant DNAse hypersensitive sites or CTCF ChIP-seq peaks (**Hughes *et al.*, 2014**; **Hanssen *et al.*, 2017**). In order to fully dismiss CBSs as major boundary elements at the mouse α-globin sub-TAD, these “θ1/θ2 minor” CTCF sites would have to be knocked out, in combination with Δθ1-θ2-44-48, and the α-globin sub-TAD interaction profiles re-analysed. However, it should be considered that no such “theta minor” CTCF sites are present at the corresponding loci in humans, rat, or any other species queried, despite high conservation of the domain structure between human and mouse (**Fig. 3.7B**). These CBSs are also in the “forward” orientation, so not convergently orientated relative to the sub-TAD. The orientation and lack of evolutionary conservation of these “theta minor” sites suggests that a vital role in mouse α-globin sub-TAD delineation is unlikely.

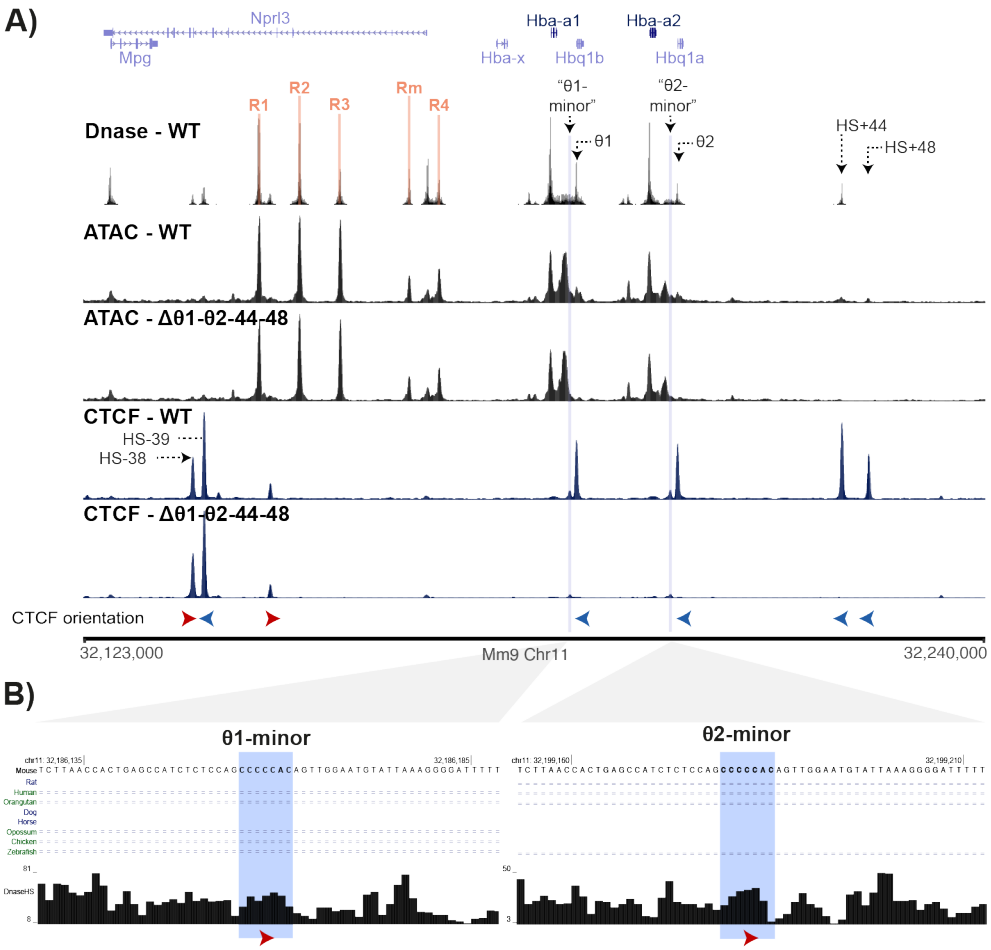

Fig. 3.7 – *"θ1/2 minor" CTCF binding motifs have weak CTCF binding and are not conserved in human*. A) DNAse hypersensitivity ("Dnase"/"DnaseHS"; **Hanssen *et al.*, 2017**), ATAC-seq, and CTCF ChIP-seq tracks in WT and Δθ1-θ2-44-48 adult mouse spleen erythroblasts. "θ1/2 minor" CTCF binding motifs are highlighted in blue vertical lines and coincident with the *Hbq1b* and *Hbq1a* promoters, respectively. Pertinent CTCF sites annotated in black. α-globin super enhancer elements (R1-4 and Rm) annotated in orange. UCSC gene annotation in blue, above. CTCF binding site orientation annotated with red (forward-oriented) and blue (reverse-oriented) arrows, below. **B)** UCSC Multiz Vertebrate Multi-species Alignment over the "θ1/2 minor" CBS. Top line is mm9 reference sequence. Core CBS of the putative motifs are highlighted in blue. Below, no line indicates no bases in the aligning species. Double line indicates that aligning species has one or more unalignable bases in the gap region. The "θ1/2 minor" CBS are not conserved in any of the other species and do not correspond to DNAse hypersensitivity footprints, suggesting that they are unlikely to be significant insulators in adult erythroblasts.

A more convincing argument, therefore, lies with the growing evidence that transcribing elements can also function as TAD boundaries (**Dixon *et al.*, 2012; Bonev *et al.*, 2017; Chahar *et al.*, 2022; Banigan *et al.*, 2023; H. Zhang *et al.*,**

2023; S. Zhang *et al.*, 2023). Moreover, at the time of writing, follow up work to the Δθ1-θ2-44-48 model has since been performed that provides strong evidence that the transcribing *Hba-a1/2* genes themselves may be the primary boundary of the 3' sub-TAD **(Cornell *et al.*, manuscript in preparation**). First, RNA-seq in a Δθ1-θ2 mouse showed that the normal *Hba-a1*:*Hba-a2* transcript ratio of 70:30 is preserved in erythroblasts. This suggests that θ1, which lies between *Hba-a1* and *Hba-a2,* does not act to block the *Hba-a2* promoter from the enhancers. In further experiments by L. Cornell, a boundary activity reporter mESC cell line was used to insert single test sequences between the α-globin enhancers and the YFP-tagged *Hba-a1* gene. After induction of erythroid differentiation, the insulation strength of each insert was measured by Hba-a1-YFP reporter readout by flow cytometry. This demonstrated that insertion of an α-globin promoter, gene-body, and promoter with gene-body could repress *Hba-a1-YFP* expression to varying degrees. RNA-sequencing verified that insulation strength strongly positively correlated with the transcriptional output of the inserted sequence. Further work to observe domain structure while specifically abrogating transcription at the *Hba-a1* and *Hba-a2* genes or depleting cohesin would solidify these conclusions.

The *Hbq1b/a* genes are transcribed in mouse erythroblasts but have unknown function (**Hsu *et al.*, 1988**). The θ1/θ2 CBSs coincide with the *Hbq1b/a* promoters. It is plausible that *Hbq1b/a* transcription might contribute to the 3' boundary similarly to the *Hba-a1/2* genes. This may even explain the evolutionary conservation of *Hbq1b/a*, despite the fact that they do not contribute functional haemoglobin protein. Moreover, MCC from the viewpoint of HS-38 (**Hua *et al.*, 2021**) revealed that, in erythroid cells, HS-38 has an approximately two-fold higher interaction probability with the regions of the θ1 CBS/*Hbq1b* promoter and θ2 CBS/*Hbq1a* promoter than with either of *Hba-a1/2* genes. This suggests that *Hbq1b/a* may even form the strongest boundary element at this locus. It is conceivable that the combination of *Hbq1b/a* transcription and CTCF binding amounts to greater insulation strength than *Hba-a1/2* transcription alone. Then again, the *Hbq1b/a* genes are constitutively accessible in DNase hypersensitivity assays, unlike the erythroid-specific α-globin genes, suggesting their conservation may result from a non-erythroid function (**Hughes *et al.*, 2005**).

MCC assays in mouse models with Δθ1-θ2 and inclusion of *Hbq1b/a* in knockout or transcriptional abrogation studies would shed further light on this and provide interesting insight into the function of these poorly understood genes.

3.3.3 – The potential insulator role of θ1/θ2 and HS+44/+48 before or beyond terminal erythropoiesis

Despite the minimal role in α-globin sub-TAD insulation in late erythroblasts, θ1/θ2 and HS+48 are highly evolutionarily conserved (**Hughes *et al.*, 2005; Appendix Fig. A3.4**). This raises the question of whether the main function θ1/θ2 and HS+44/+48 lies beyond a mature erythroid context.

Sh3pxd2b and *Ubtd2* are not expressed in erythroid cells, but both are expressed in heart, brain, lung, spleen, and adipocyte tissues, among others. This raises the possibility that the θ1/θ2 and HS+44/+48 CBSs may play a greater boundary role in these non-erythroid cell types. Published Hi-C data in non-erythroid tissues should be reanalysed to determine whether a cell type-invariant TAD boundary coincides with the 3' CBS. Similarly, Capture-C experiments and gene expression analyses in non-erythroid tissues could reveal a more disruptive phenotype of the Δθ1-θ2-44-48 model. If so, these would suggest the 3' CBSs may function to separate the inactive α-globin enhancers and promoters from the active *Sh3pxd2b* and *Ubtd2* genes.

Likewise, it is possible that the θ1/θ2 and HS+44/+48 CBSs play a greater insulation role when establishing the α-globin sub-TAD in early erythroid progenitors. Within current temporal resolution limits, the α-globin sub-TAD appears to form concomitant with *Hba-a1/2* transcription initiation (**Oudelaar and Beagrie, 2020**), but it is plausible that early cohesin loop extrusion events require the θ1/θ2 and HS+44/+48 CBSs to halt loop extrusion, and thereby bring the *Hba-a1/2* genes close to the enhancer elements, before transcription of the *Hba-a1/2* genes is established. Detecting these extrusion events, which may be rare, in early erythroid progenitors is beyond the sensitivity of contemporary capture-style experiments. Super-resolution imaging may be better able to detect the spatial dynamics of enhancer-promoter proximity events, cohesin location and their relation to transcriptional outcome in rare cell populations (**Chen *et al.*, 2018;**

Alexander *et al.*, 2019; Gabriele *et al.*, 2022; Mach *et al.*, 2022; Platania *et al.*, 2023). Although the association between 3C-style interaction frequencies and spatial proximity as determined by imaging is debated (**Giorgetti and Heard, 2016; Huang *et al.*, 2021; Gomez-Acuna *et al.*, 2024**). Further technical advancements in the sensitivity, precision, and throughput of all these techniques are required for this hypothesis to be directly tested.

3.3.4 – A revised model for α-globin sub-TAD insulation

Taken together, I propose a revised model in which, in the context of mature erythroblasts, the transcribing *Hba-a1/2* genes themselves delineate the α-globin sub-TAD by physically blocking cohesin-mediated loop extrusion in the majority of loop-extrusion events. Meanwhile the θ1/θ2 and HS+44/+48 CTCF binding sites may serve as minor, or "insurance", boundary elements, in the minority of cases where loop extrusion surpasses *Hba-a1/2* (**Fig. 3.8**). In early erythroid progenitors where *Hba-a1/2* expression is low or nil, these CBSs may play a more vital role in establishing α-globin enhancer-promoter contacts or insulating promiscuous activity of the strong α-globin enhancers.

Functional overlap of enhancer and promoter elements is well documented, but the ability of promoter elements to act as insulators is much less well understood. This work is an important step in the growing literature determining the boundary elements at the α-globin locus and the general principles of 3D genome regulation. Further work to demonstrate the cell-type or differentiation-stage specificity of the 3' CTCF binding sites' function will help to unravel why and how a cis-regulatory element's function may differ so markedly between cell types.

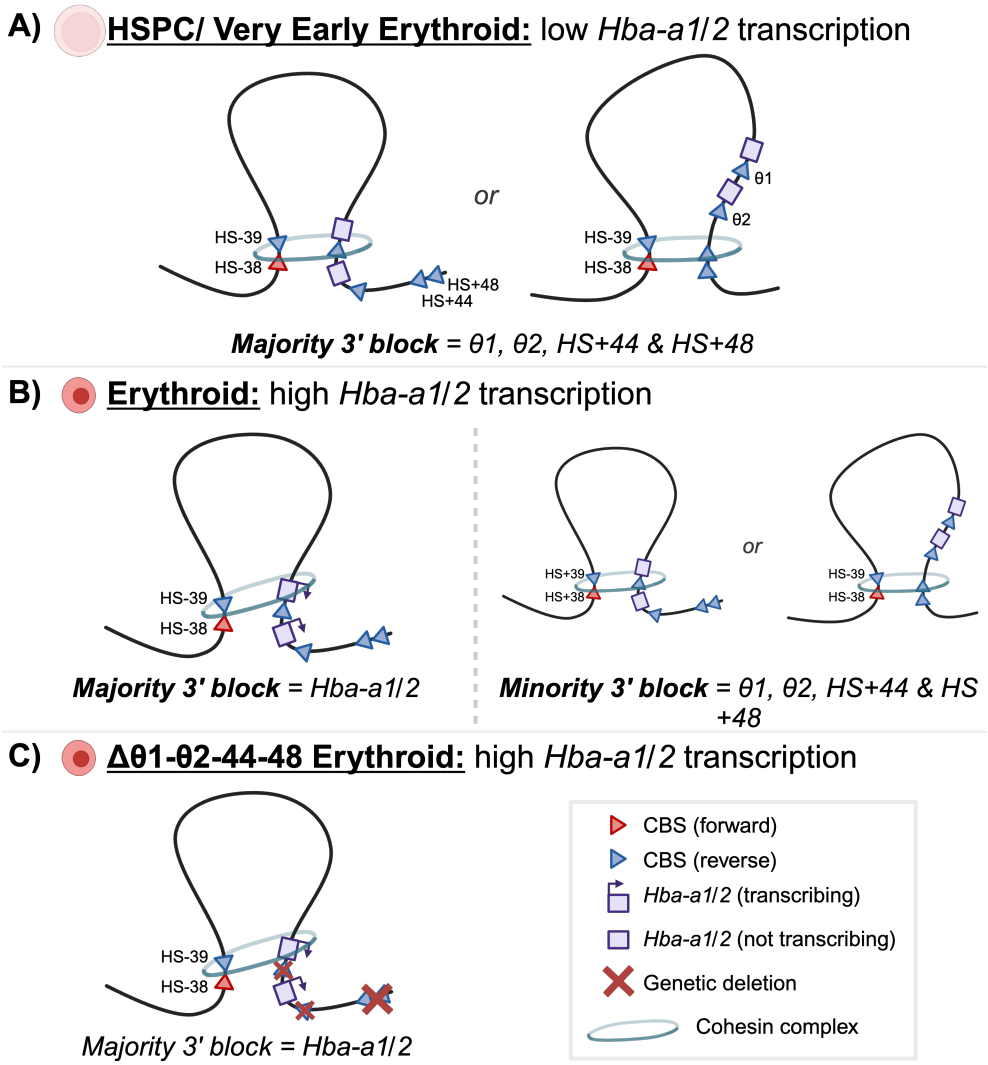

Figure 3.8 – *A revised model of the boundary elements to cohesin-mediated chromatin loop extrusion at the α-globin sub-TAD locus*. A) In very early erythroid progenitor/precursor cells, when *Hba-a1/2* transcription is low or nil, the CTCF binding sites (CBS) θ1, θ2, HS+44, and HS+48 may confer the major block against loop extrusion at the 3' end. This might help to establish the sub-TAD domain and facilitate enhancer-promoter interactions before transcription complex assembly is frequent/stable enough to insulate cohesin itself. **B)** In erythroid cells, assembly of the "bulky" transcription machinery at α-globin genes (including pol II, transcription factors (TFs) and large TF complexes, such as Mediator or chromatin remodellers) is a more effective block to loop extrusion and forms the major 3' boundary to the now well-established sub-TAD. **C)** In Δθ1-θ2-44-48 erythroid cells, the major blocking elements - the α-globin genes - remain intact and the sub-TAD structure is largely unaffected.

3.4 – Chapter 3 conclusions

In this chapter, I set out to characterise the Δθ1-θ2-44-48 mouse model and establish whether deletion of the θ1/2 and HS+44/+48 CTCF binding sites perturbed the α-globin sub-TAD's 3' boundary. I conclude that the Δθ1-θ2-44-48 mouse is phenotypically and haematologically normal and that Δθ1-θ2-44-48 does not significantly perturb the structure or function of the 3' boundary. This is in surprising contrast to the known boundary function of the HS-38/-39 CBSs at the 5' of the sub-TAD. From my results, reinforced by subsequent experiments from our laboratory, I propose a revised model of the α-globin sub-TAD boundaries, in which the transcribing *Hba-a1/2* genes are the major boundary elements while θ1/2 and HS+44/+48 CSBs contribute only a minor boundary effect in erythroid cells. I have provided a novel example of evolutionarily conserved CBSs that are redundant to TAD boundary formation. Moreover, in context with recent evidence for transcription of *Hba-a1/2* acting as a boundary element, this work contributes to novel evidence for actively transcribing promoters having dual function as promoters and boundary elements.

3.5 – Chapter 3 Acknowledgements

A massive thank you goes to Caroline Harrold who, in collaboration with Lars Hanssen and Ben Davies, designed and generated the Δθ1-θ2-44-48 mouse model and laid much of the technical and theoretical groundwork for this project. Thank you to Jackie Sharpe and Jackie Sloane-Stanley for performing the animal husbandry work. Thank you to Mira Kassouf and Joseph Blayney, for introducing me to the lab and teaching me the NGS techniques that were instrumental for this chapter. Thank you to Jelena Telenius and Damien Downes, for providing the bioinformatic pipeline to analyse the Capture-C data. Thank you, also, to Lucy Cornell for the fantastic scientific discussions, colour-scheme obsessions and her follow-up work on the boundary effect of the α-globin genes that has furthered the narrative of the CTCF story – and for the many laughs and coffee breaks along the way.

Chapter 4 – Cellular and molecular characterisation of erythropoiesis in the KLF1$^{StrepII\text{-}FKBP}$ mouse model

4.1 – Introduction

In this introduction, I give a brief overview of the functions of KLF1, some of the technical limitations faced by those studying KLF1, and how I seek to overcome these limitations with a StrepII-FKBP tagged KLF1 mouse model.

4.1.1 – The role of KLF1 in haematopoietic lineage fate decisions

KLF1 is known to play a role in cell fate choice at several stages of haematopoiesis (**Fig. 4.1**). KLF1 expression starts at very low levels in the multipotent progenitors and increases as cells differentiate to common myeloid progenitors (CMP), favouring myeloid over lymphoid lineages. KLF1 expression then increases again as megakaryocyte and erythrocyte progenitors (MEPs) emerge (**Palii *et al.*, 2019**). From MEP cells, mutual antagonism between KLF1 and FLI1 promotes either the erythroid or megakaryocyte lineages, respectively (**Frontelo *et al.*, 2007;** reviewed in **Siatecka and Bieker, 2011**). The most marked increase in KLF1 expression occurs in terminal erythroid differentiation. There is also evidence suggesting that KLF1 may play a role in specifying macrophages, particularly the central macrophages of erythroblastic islands (CMEIs) (**Luo *et al.*, 2004**; **Porcu *et al.*, 2011; Xue *et al.*, 2014**).

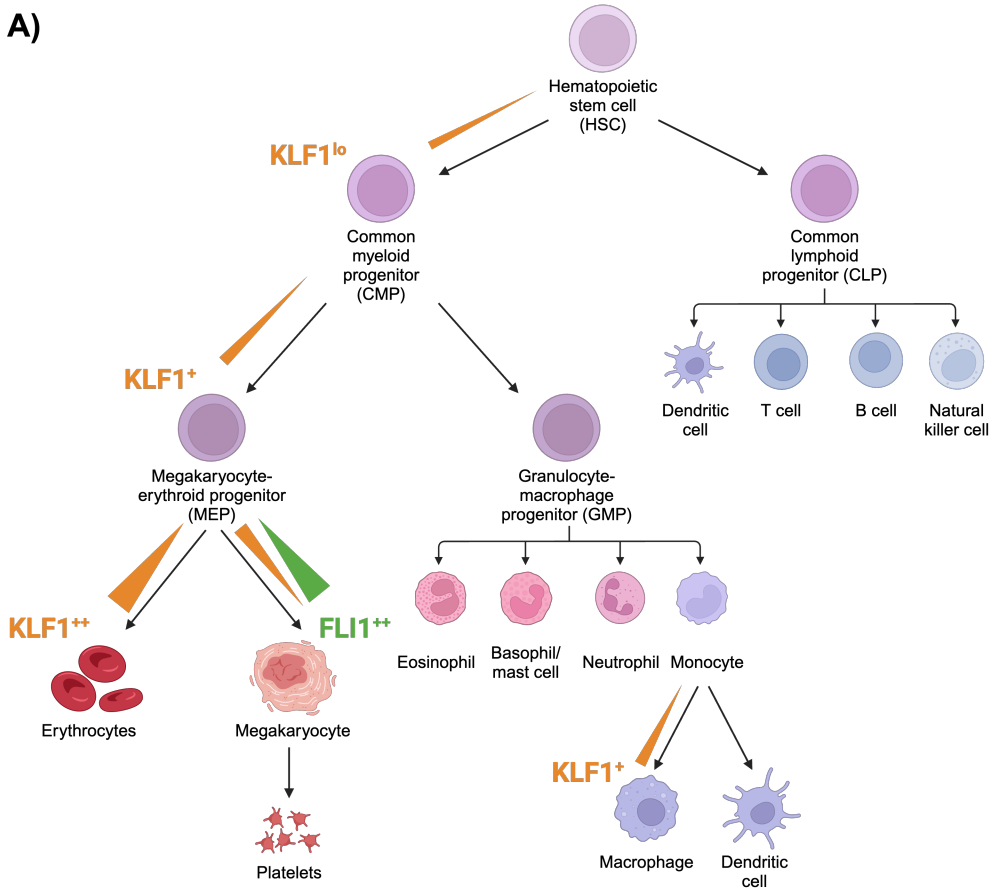

Fig. 4.1 – *Schematic summary of KLF1 expression in haematopoiesis*. Thin oranges wedges indicate changes in KLF1 expression. Thick orange or green wedges indicate induction of high expression of KLF1 or FLI1, respectively.

4.1.2 – KLF1 expression profile in erythropoiesis

While *Klf1* expression is modest in MEP cells, it markedly increases as cells differentiate through the erythroid lineage, in both human and mouse (**Corces *et al.*, 2016; Ludwig *et al.*, 2019; Oudelaar *et al.*, 2020**). **Fig. 4.2A** shows the expression profile of *Klf1* RNA in E13.5 mouse fetal liver (mFL) cells, as measured by single-cell RNA-sequencing (**Oudelaar *et al.*, 2020**). These samples were sorted by a FACS gating strategy based on the expression of the CD71 and Ter119 cell surface markers (**Fig. 4.2B; Zhang *et al.*, 2003; Pop *et al.*, 2010**). This well-established method gates erythroid cells into "S0-S5" populations, which approximately follows the trajectory of erythroid differentiation as indicated in **Fig. 4.2C**. *Klf1* peaks at early terminal erythroid differentiation.

To the best of my knowledge, there is no published data with an erythroid timecourse of mouse KLF1 protein expression. This is likely due, in part, to the lack of reliable antibodies against KLF1, discussed further in **section 4.1.5**. However, a single-cell CyTOF timecourse experiment in differentiating erythroid cells from human umbilical cord assayed KLF1 expression (and 15 other TF proteins and 11 cell surface markers; **Palii *et al.*, 2019**). These data span from multipotent progenitors (MPP) to basophilic erythroblasts and show a large and steady increase in KLF1 throughout, at a similar rate as their co-measured increase in *KLF1* RNA. Given the strong functional conservation of KLF1 between human and mouse, it is likely that the *Klf1* RNA expression profile measured by **Oudelaar *et al.*, 2020**, is reflective of the mouse KLF1 protein expression profile.

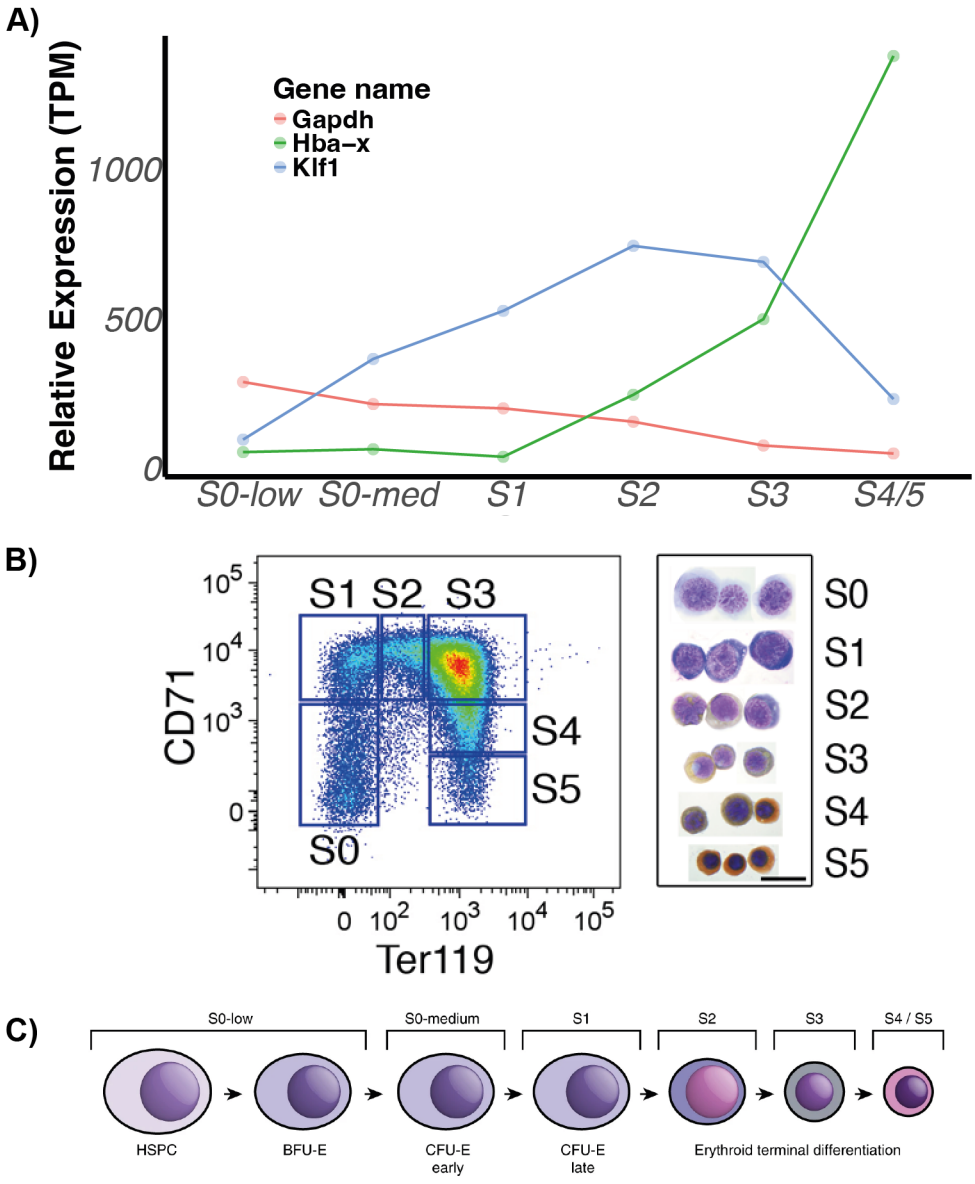

Fig. 4.2 – *Klf1 expression increases through mouse erythropoiesis.* A) Relative RNA expression (transcripts per million; TPM) from single-cell RNA-sequencing data in CD71/Ter119 FACS-sorted E13.5 mFL erythroblast populations. Replotted with permission from **Oudelaar *et al.*, 2020**. **B)** Example FACS plot and cytospin preparations demonstrating the CD71/Ter119 FACS gating strategy used to isolate erythroid progenitors from an E14.5 mFL. Scale bar = 20 μm. Figure reproduced from **Pop *et al.*, 2010**. **C)** Schematic indicates the approximate erythroid differentiation stages enriched in the S0-S4/5 FACS sorted groups. Reproduced from Oudelaar *et al.*, 2020.

4.1.3 – The role of KLF1 in erythropoiesis

KLF1 is required for both primitive and definitive erythropoiesis. KLF1 can therefore be detected in all sites of haematopoiesis, from the E7.5 yolk sac and circulating primitive blood, to the fetal liver and adult bone marrow and spleen (**Southwood *et al.*, 1996**). However, KLF1 knockout mice display a far less severe phenotype in the primitive erythroid compartment than definitive. This is thought to be due to functional redundancy of KLF1 with other TFs, particularly KLF2, in primitive erythropoiesis (**Basu *et al.*, 2007; Alhashem *et al.*, 2011**). Meanwhile, KLF1$^{-/-}$ causes embryo death between E14.5-E16.5, following the primitive to definitive switch (**Nuez *et al.*, 1995; Perkins *et al.*, 1995**). The majority of KLF1 functional studies therefore concern definitive erythropoiesis, which has the most mechanistic and clinical relevance.

KLF1 expression is essential for erythropoiesis. As well as erythroid/megakaryocyte lineage fate decisions, KLF1 has integral functions across cell cycle, cell membrane formation, iron metabolism, globin chaperones, haemoglobin synthesis, red cell metabolism, and enucleation (summarised in **Table 4.1**). Many of these functions have been determined through a combination of KLF1 loss/gain of function and knockout studies, and by studying patients with inherited variants of KLF1. Most of these key erythroid functions are attributable to KLF1's role as a transcription factor, for which WT KLF1 has an estimated >1000 direct gene targets (**Tallack *et al.*, 2010; 2012; Huang *et al.*, 2024**). Therefore, validating the DNA binding and TF activity of KLF1 is central to understanding its roles in erythropoiesis and, by extension, in human disease.

Table 4.1 – *Example genes verified as direct targets of KLF1 and their known erythroid functions*.

Erythrocyte function	Example KLF1 gene targets	Gene function	Reference
Erythroid differentiation	*Tal1*	Erythroid TF; complex with GATA1; binds *Klf1* promoter	Kassouf *et al.*, 2010 Hung *et al.*, 2021
Cell cycle	*E2f2*	S-phase entry and DNA synthesis	Pilon *et al.*, 2008 Tallack *et al.*, 2009
Cell membrane	*Dmtn*	Erythroid cytoskeleton and membrane integrity	Siatecka *et al.*, 2010 Mukherjee *et al.*, 2022
Iron metabolism	*Slc25a37*	Mitochondrial iron transport	Tallack *et al.*, 2010
Red cell metabolism	*Pklr*	Glycolysis/ATP production	Viprakasit *et al.*, 2014
Globin chaperones	*Ahsp*	Binds free α-globin chains	Pilon *et al.*, 2006
Haemoglobin synthesis	*Hbb-bt/s*	Adult β-globin chains	Miller and Bieker, 1993 Tallack *et al.*, 2010
	Hba-a1/2	Adult α-globin chains	Tallack *et al.*, 2010

4.1.4 – Key experiments in the history of KLF1

KLF1 (EKLF) was discovered in 1993 by subtractive hybridisation studies in the mouse erythroleukaemia (MEL) cell line (**Miller and Bieker *et al.*, 1993**). It was the first of a family of 18 Krüppel-like factors to be identified, named after their structural similarity to the Krüppel protein in *Drosophila*. Miller *et al.* also determined the expression profile of KLF1 and its canonical 5'-CCA-CAC-CCT-3' binding motif at the mouse β-globin promoter, which has since been refined to 5'-CCM-CRC-CCN-3' (where M = A/C and R = A/G; **Tallack *et al.*, 2010; Siatecka *et al.*, 2010**). In 1995, the first KLF1 knockout mice were produced and characterised (**Nuez *et al.*, 1995; Perkins *et al.*, 1995**). Both studies recorded the embryonic lethality of KLF1$^{-/-}$ between E14.5 and E16.5, the presence of microcytic red cells with membrane defects, and severe β-thalassemia. While the cause of embryonic death was first presumed to be β-thalassemia, subsequent experiments supplementing the KLF1$^{-/-}$ mouse with human transgenic γ-globin

did not rescue the KLF1$^{-/-}$ phenotype (**Perkins *et al.*, 2000**). From this, it was clear that KLF1 had additional essential functions beyond globin regulation.

Since then, the role of KLF1 in health and disease has been an active area of research and is extensively reviewed elsewhere (**Siatecka and Bieker, 2011; Tallack and Perkins, 2013; Viprakasit *et al.*, 2014; Perkins *et al.*, 2016; Caria *et al.*, 2022; Bieker and Philipsen, 2024**). Notably, two pivotal advances in our understanding of KLF1 came with the advent of high-throughput next generation sequencing.

To study the role of a TF, we need to understand its genome-wide DNA binding profile, identifying gene targets in relevant tissues, and the impact of the TF's knockout or knockdown on the expression of those genes and resulting phenotype. The first KLF1 ChIP-seq and RNA-seq experiments in WT vs KLF1$^{-/-}$ mice shed such a light on the global binding and functions of KLF1 (**Tallack *et al.*, 2010; 2012**). This identified 945 putative binding sites for KLF1 and 808 dysregulated genes in the KLF1$^{-/-}$ mouse. Ontology analyses of these genes affirmed that KLF1 is involved in the regulation of tens, if not hundreds, of key erythroid pathways. Similar functional studies in mouse models of human KLF1 disorders, such as the neonatal anaemia (Nan) mouse model of CDA IV, have revealed even more nuance behind the complex effects of KLF1 perturbation (**Mukherkjee and Bieker, 2022**)**.** More recent iterations of KLF1 ChIP-seq have increased the estimated number of KLF1 binding sites to 1144 (**Huang *et al.*, 2024**). However, there are two major limitations to these studies.

First, studying gene expression changes in total knockout models is limited by the lengthy delay between KLF1 perturbation and experimental readout. This makes it difficult to discern primary effects of KLF1 knockout from secondary and tertiary effects; a particular problem when studying a highly pleiotropic TF. shRNA knockdown studies for KLF1 have been used to uncover the role of KLF1 in patient samples with HPFH (**Borg *et al.*, 2010; Zhou *et al.*, 2010**). However, TF-target gene transcription can respond to TF changes in the order of minutes while transduction with an shRNA lentivirus typically has a delay of several days before gene expression readout. Additionally, these experiments are confounded by

unknown off-target effects of shRNAs and residual protein persistence. Indeed, without rapid depletion of KLF1 at the protein level, it remains difficult to untangle primary from secondary gene target effects. Second, there are concerns in the field over the accuracy and reliability of the existing KLF1 ChIP-seq data.

4.1.5 – Review of pre-existing KLF1 ChIP-seq experiments

Historically, KLF1 has been an extremely difficult protein to ChIP. Notable attempts are outlined in **Table 4.2**. This difficulty has been speculated to be for many reasons. **Tallack *et al.*, 2010** and **Eaton *et al.*, 2008**, suggested binding competition with KLF3 and thus produced their – the first ever – KLF1 ChIP-seq in KLF3$^{-/-}$ mice (**Tallack *et al.*, 2010**). Another reason is that sometimes a protein is simply difficult to ChIP. This might be, for example, if it has a low residence time when binding chromatin, although the true reason is rarely determined. In the case of KLF1, however, an outstanding unresolved concern in the field that could explain poor ChIP performance is the quality of the available anti-KLF1 antibodies. ChIP-seq is crucially dependent on strong and specific binding of the antibody to the protein of interest. There have been two main antibodies against mouse KLF1 (**Perkins *et al.*, 1995; Siatecka *et al.*, 2010**), but there is unresolved dispute as to which, if either, accurately and reliably detects endogenous KLF1.

To compensate for this, previous KLF1 ChIPs have all employed sub-optimal adjustments to the standard ChIP protocol. As well as the “WT” KLF1 ChIP in the KLF3$^{-/-}$ model, other adjustments, outlined in **Table 4.2**, include pooling tens of millions of cells, averaging across many technical replicates, and sequencing to extremely high depths. These adjustments are not only time-consuming and expensive, but also preclude the use of rare cell populations, such as early erythroid progenitors in which KLF1 is present and may be functioning. To my knowledge, there has been one previous attempt to tag endogenous KLF1 in mouse, but the anti-HA tag ChIP-seq for KLF1 in this model still required several tens of millions of cells and yielded poor signal:noise ratio (SNR) (**Pilon *et al.*, 2011**). While these approaches have elucidated many roles of KLF1 in erythropoiesis over the past 3 decades, there remains an unmet need for a reliable, low cell input system in which to do so. Such a ChIP for endogenous

KLF1 is necessary to corroborate these findings and further explore the complex role of KLF1 in erythropoiesis and haemoglobin switching.

To overcome the above limitations of previous KLF1 experiments, I herein describe the design and characterisation of a KLF1$^{StrepII\text{-}FKBP}$ mouse model and demonstrate its potential use in a low cell input KLF1 ChIP.

Table 4.2 – *Summary of previous KLF1 ChIP-seq experiments in mouse and human erythroblasts*. "n/a" = data not available.

Reference	Genotype	Cell type	Replicates	Antibody	Cell # per IP	Total reads	GEO accession:
Tallack *et al.*, 2010	WT & KLF3$^{-/-}$, combined	E14.5 mFL	2 biological replicates, each of 4 pooled littermates.	Rabbit polyclonal anti-KLF1 (Perkins et al., 1995; Keys et al., 2007)	8 x 10^7	17,438,921	GSE20478
Pilon *et al.*, 2011	HA-KLF1-TAP	E13.5 mFL, CD71/Ter119 **sorted** early vs late erythroblasts	>3 independent cell sorts	Mouse monoclonal anti-HA (F-7, sc-7392X)	n/a	35 million per replicate	n/a
Ilsley *et al.*, 2017	KLF1-ER	K1-ER MEL cell line	2 biological replicates, 3 pooled technical reps each	Anti-KLF1 antibody (Perkins et al., 1995)	n/a	n/a	GSE92620 GSM995443
Mukherjee & Bieker, 2022	WT (vs nan/+)	E13.5 mFL	3 biological replicates, where possible biological replicates were littermates.	Mouse monoclonal antibody anti-KLF1 "7B2a"	7.5 x 10^6	n/a	GSE210779
Huang *et al.*, 2024	WT (vs KLF1$^{D45\text{-}mutant}$)	E14.5 mFL	1 (sequencing library generated from 2 pooled replicates)	Anti-KLF1 antibody (Perkins et al., 1995)	n/a	n/a	GSE94351
Su *et al.*, 2013	WT	Late erythroid (S3/4) sorted, CD71$^+$	n/a	Goat polyclonal, anti-human KLF1, ab2483, Abcam	1 × 10^7	>18 million after mapping	GSE43626
Liu *et al.*, 2024	WT (vs GATA2AS KO)	HUDEP2	2	Goat polyclonal, anti-human KLF1, ab2483, Abcam	5 x 10^6	n/a	GSE213779

4.1.6 – The KLF1$^{StrepII\text{-}FKBP}$ mouse model

The KLF1$^{StrepII\text{-}FKBP}$ mouse is a recombinant model harbouring an insertion of a Twin-StrepII and an FKBP-V tag at the endogenous *Klf1* gene. Model design is described further in **Results 4.2.1**.

Streptavidin is a bacterial protein with extremely high binding affinity for biotin molecules (vitamin B7). StrepII is an engineered biotin analogue widely used in molecular biology applications, including protein purifications and ELISA. The Twin-StrepII tag, with two StrepII peptides joined by a short linker sequence, decreases the K_{off} rate of StrepII-streptavidin associations, thereby improving StrepII binding stability (**Schmidt *et al.*, 2013**). Herein, the Twin-StrepII tag used in this mouse model is referred to as StrepII. By adding a StrepII epitope to endogenous KLF1, I aimed to resolve debates surrounding anti-KLF1 antibody specificity and, with its extraordinary strength and affinity of binding, reduce the high cell numbers required of previous KLF1 ChIP experiments.

FKBP/dTAG-13 is a proteolysis targeting chimera (PROTAC) system for acute and targeted protein degradation that utilises the highly conserved, endogenous eukaryotic E3 ubiquitin ligase protein degradation pathway. FKBP12^{F36V}, herein referred to as FKBP-V or FKBP, is an engineered peptide tag that can be recombinantly expressed in-frame with a protein of interest. Addition of dTAG-13, a ligand for both FKVP-V and cereblon (CRBN), recruits the cell's endogenous CRBN (E3 ubiquitin ligase) complex for inducible, rapid (within 30 minutes), near-complete, and reversible target protein degradation (**Nabet *et al.*, 2018**). FKBP tags are powerful tools for studying protein function but have been known to affect protein stability and/or function (**Layden *et al.*, 2021; Mehta *et al.*, 2022**), and so thorough characterisation of any recombinant FBKP model is essential. The FKBP-V tag in my KLF1$^{StrepII\text{-}FKBP}$ model was generated to facilitate several novel investigations of KLF1 function. Specifically, I aimed to study the role of KLF1 in embryonic to adult α-like globin switching, explored in **Chapter 5**.

4.1.7 – Aims of Chapter 4

The KLF1$^{StrepII\text{-}FKBP}$ mouse therefore stands to address several outstanding and novel questions of KLF1 function in erythropoiesis. However, it was important to fully characterise this mouse model to ensure the normal function of the tagged KLF1 protein. In this chapter, I characterise the KLF1$^{StrepII\text{-}FKBP}$ mouse and seek to establish a low cell input ChIP method – chipmentation – for KLF1$^{StrepII\text{-}FKBP}$ to overcome existing barriers in the study of KLF1.

Aims:

1. To phenotype cellular and molecular erythropoiesis in the KLF1$^{StrepII\text{-}FKBP}$ mouse model.

2. Establish chipmentation in the KLF1$^{StrepII\text{-}FKBP}$ model to further characterise the genome-wide binding profile of endogenous KLF1 in primary erythroid cells.

3. Use chipmentation in the KLF1$^{StrepII\text{-}FKBP}$ model to reveal the differential genome-wide binding profile of endogenous KLF1 across terminal erythroid differentiation.

4.2 – Results

4.2.1 – Design of the KLF1$^{StrepII\text{-}FKBP}$ mouse model

As illustrated in **Fig. 4.3A**, donor template DNA sequence encoding a twin StrepII tag and an FKBP-V tag was previously inserted by homologous recombination into the endogenous *Klf1* gene of C57BL6/J zygotes. The StrepII-FKBP sequence is inserted directly downstream of the last exon and upstream of the stop codon and 3' UTR. The StrepII-FKBP peptide tag is therefore translated to the C-terminus of KLF1 (**Fig. 4.3B**). Disordered GS linker sequences between the KLF1 coding sequence, the StrepII tag, and the FKBP-V tag encourage proper folding of all three peptide sequences after translation. **Fig. 4.3C-E** are AlphaFold predictions of WT KLF1 and KLF1$^{StrepII\text{-}FKBP}$ protein folding conformations, bound to an oligo from the *Hbb-b1* promoter containing the canonical KLF1 binding motif (**Miller and Bieker, 1993**). There is extremely high similarity between the predicted conformations of the three zinc finger domains (ZnFs) between KLF1 and KLF1$^{StrepII\text{-}FKBP}$. Notably, there is very high confidence in the predicted structures of the ZnFs and StrepII-FKBP domains and the StrepII-FKBP is not predicted to interact with, or affect the DNA binding of, the ZnFs. As expected, the unstructured N-terminus has low confidence and low similarity between KLF1 and KLF1$^{StrepII\text{-}FKBP}$.

Of course, generative AI prediction models, such as AlphaFold, can only produce results based on an averaged synthesis of pre-existing training data. These predictions are not based on any experimental data from KLF1 or KLF1$^{StrepII\text{-}FKBP}$ specifically. However, it is reassuring that the positioning of the StrepII-FKBP tag at the C-terminus is not expected to interfere with the proximal ZnFs' structure and presumably with its function.

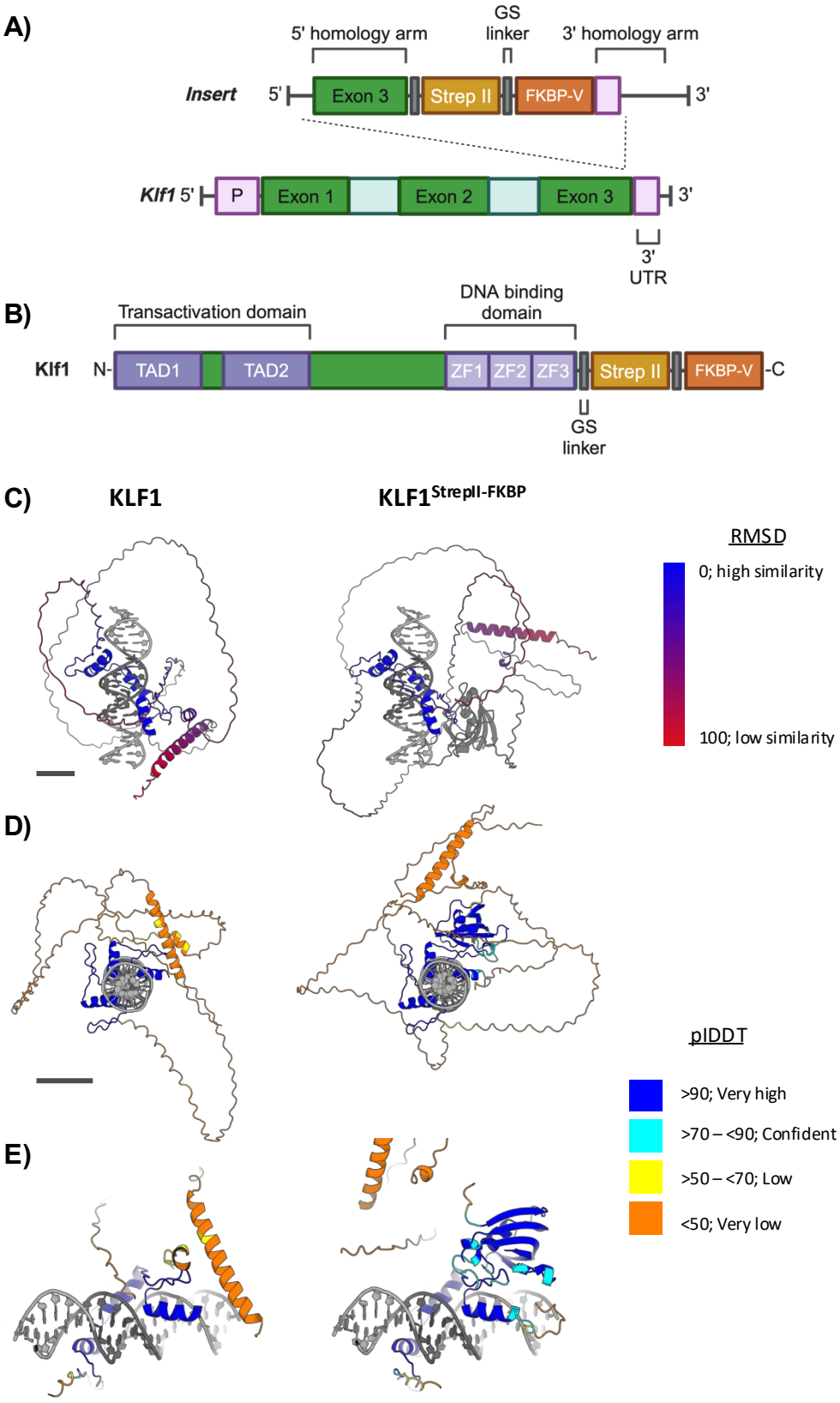

Fig 4.3 – *Schematics of the KLF1$^{StrepII\text{-}FKBP}$ engineering strategy and recombinant protein.* A) Schematic of the genome engineering strategy for KLF1$^{StrepII\text{-}FKBP}$. Template dsDNA encoding the StrepII and FKBP-V tag flanked by homology arms for recombination into endogenous *Klf1* after the third exon. Not to scale. "P" = promoter. "3'

UTR" = 3' untranslated region. Light green boxes = introns. **B)** Schematic of the KLF1$^{StrepII\text{-}FKBP}$ protein sequence. Not to scale. Dark green box = KLF1 coding sequence. "TAD", light purple = transactivation domain. "ZF", dark purple = C2H2 zinc finger domain. **C-E)** AlphaFold predictions of KLF1 and KLF1$^{StrepII\text{-}FKBP}$ protein folding with the canonical *Hbb-b1* KLF1 binding motif. **C)** Protein is coloured by RMSD (root mean square deviation; a measure of similarity between superimposed protein structures) between KLF1 and KLF1$^{StrepII\text{-}FKBP}$. The three zinc finger domains have near identical conformations (blue). **D + E)** Protein is coloured by pIDDT (a measure of local folding confidence). The zinc finger domains and StrepII-FKBP tag have very high predicted folding confidence. **E)** A close up of the zinc finger domains interacting with the KLF1 binding motif. The 21-nucleotide dsDNA oligo from the *Hbb-b1* promoter is shaded in light grey, with the canonical KLF1 binding motif (5'-CCACACCCT-3') in dark grey (**Miller and Bieker, 1993**). Scale bars = 20Å.

4.2.2 – Survival and phenotype of the KLF1$^{StrepII\text{-}FKBP}$ mouse

Next, I assessed the KLF1$^{StrepII\text{-}FBKP}$ mouse model to determine whether the StrepII and FKBP-V protein tags on endogenous KLF1 (in the *absence* of dTAG-13 ligand) had an adverse effect on mouse survival or phenotype.

First, the genotypes of progeny from heterozygous crosses were counted to determine whether the KLF1$^{StrepII\text{-}FBKP}$allele was segregating within the population in the expected Mendelian ratios (genotyping strategy in **Methods 2.3**). If the KLF1$^{StrepII\text{-}FBKP}$ allele were to have no effect on survival to weaning, then we can expect the ratio of adult WT:het:hom to be 1:2:1. All available (N = 40) pups were genotyped after weaning**.** This gave 18 WT, 16 het, and 6 hom pups, which significantly differs from the expected Mendelian ratio, as calculated by Chi-squared Goodness of Fit test (P = 0.01228, N = 40) (**Fig. 4.4**; **Methods 2.13.3**). This implies that, surprisingly, the KLF1$^{StrepII\text{-}FBKP}$ genotype may have a negative effect on mouse survival to adulthood.

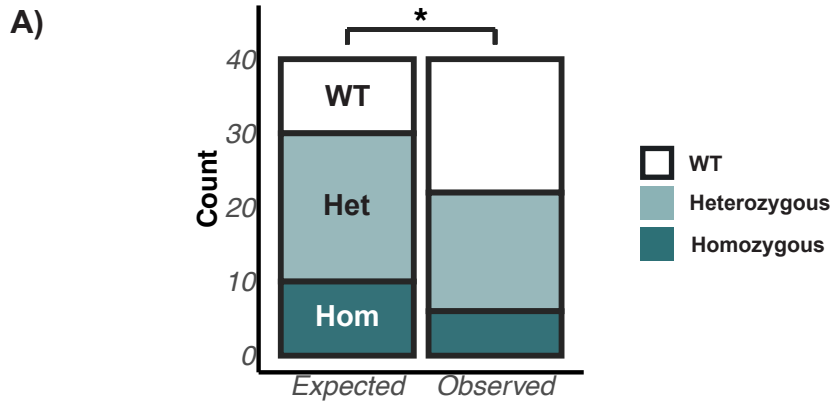

Fig. 4.4 – *The KLF1$^{StrepII\text{-}FBKP}$ allele does not segregate in a Mendelian ratio among progeny of heterozygous crosses.* A) Observed counts of KLF1$^{StrepII\text{-}FBKP}$ homozygous (hom), heterozygous (het), and wildtype (WT) mice, at weaning age, significantly differ from an expected Mendelian ratio, as calculated by Chi-squared Goodness of Fit test (N = 40, X^2 = 8.8, *d.f.* = 2, *P-value* = 0.01228).

To further investigate this, I compared the number of KLF1$^{StrepII\text{-}FBKP}$ pups from hom x hom crosses alive at birth to the number alive at weaning. All of the 41 eligible pups survived to weaning, suggesting that the skew in het x het progeny is not caused by a loss of hom pups between birth and weaning. Furthermore, no WT, het, or hom progeny died prematurely post-weaning and all adult het and hom KLF1$^{StrepII\text{-}FBKP}$ mice were phenotypically normal and indistinguishable from WT by visual inspection (data not shown).

It is possible that hom pups are more likely to perish before birth. To investigate this, I counted the number of live, runted, and partially reabsorbed pups from homozygous crosses harvested at embryonic timepoint E12.5. If KLF1$^{StrepII\text{-}FBKP}$ is negatively affecting the endogenous function of KLF1 then there may be an abnormal rate of runting or reabsorption in the E12.5 KLF1$^{StrepII\text{-}FBKP}$ pups. The results are summarised in **Table 4.3**. The high rate of runting and reabsorption among the "27SEP23" litter is likely attributable to the age of the sow at conception and I have therefore discounted these progenies from this analysis as anomalies. Of the remaining 36 total pups, we see a comparable rate of losses to a typical WT litter (**email correspondence with Jackie Sloane-Stanley**). I conclude the skewed allelic ratio of the heterozygous crosses is unlikely to be caused by loss of homozygous embryos prior to E12.5.

Table 4.3 – *Embryo counts from KLF1$^{StrepII\text{-}FBKP}$ homozygous crosses at E12.5*. Embryos were classified as "reabsorbed" if they were very small/morphologically atypical, i.e. in the process of being reabsorbed by the mother. Embryos were classified as "runted" if they were noticeably smaller/less developed than their littermates but still viable. Both were excluded from further culture and further analyses. *The sow of this litter was older than that of the rest of the crosses and the live embryos visually appeared less developed, closer to E11.5 in morphology.

Harvest date	Sow age at harvest (weeks)	# Live	# Reabsorbed	# Runted	# Total
27SEP23*	31.4	0	4	6	10
24OCT23	21.3	8	0	0	8
04DEC23	27.1	8	0	1	9
24JUN24	11	9	1	0	10

By process of elimination, I conclude that the allelic skew among KLF1$^{StrepII\text{-}FBKP}$ progeny from heterozygous crosses is likely caused either by random chance or by loss of the KLF1$^{StrepII\text{-}FBKP}$ pups between E12.5 and birth (~E20.5).

To determine any further effects of KLF1$^{StrepII\text{-}FBKP}$ on mouse health and phenotype post-weaning, I compared total body mass and relative spleen mass in adult, mixed-sex KLF1$^{StrepII\text{-}FBKP}$ mice. Spleen masses are given as a percentage of total body mass, to normalise for any confounding effect of body size on spleen mass. I find that there is no significant effect of genotype on total body mass (**Fig. 4.5A**; ANOVA: $P = 0.684$) or relative spleen mass (**Fig. 4.5B**; ANOVA: $P = 0.12$). There was also no significant effect of sex, or interaction between sex and genotype, on body mass or relative spleen mass. This suggests that there is no significant effect of KLF1$^{StrepII\text{-}FBKP}$ on adult growth. The lack of splenomegaly suggests that KLF1$^{StrepII\text{-}FBKP}$ is unlikely to be inducing high levels of stress erythropoiesis.

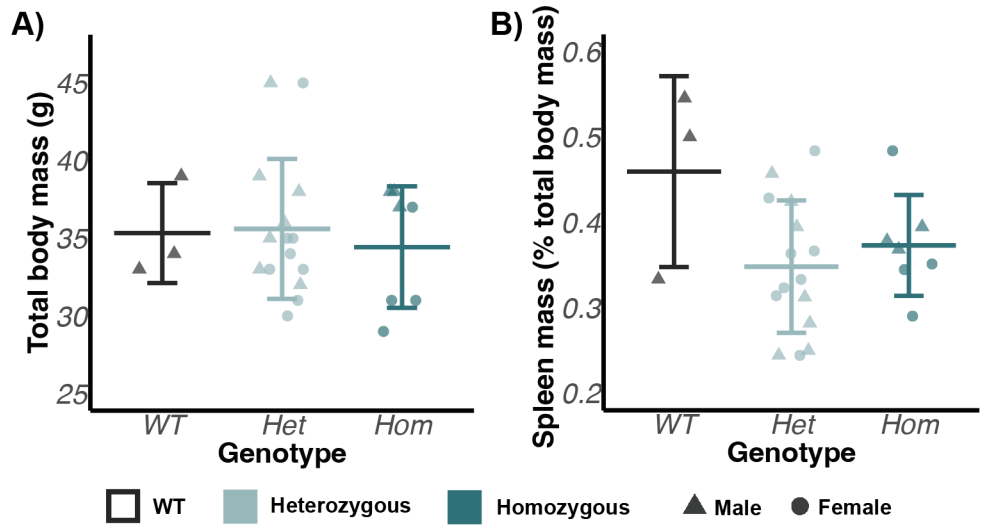

Fig. 4.5 – *KLF1$^{StrepII\text{-}FBKP}$ mice have normal total body and relative spleen masses.* **A)** Mean total body mass (grams) in KLF1$^{StrepII\text{-}FBKP}$ mice. There is no statistically significant effect of genotype or sex on total body mass, as calculated by two-way ANOVA ($P \geqq 0.05$). **B)** Mean spleen mass normalised as a percentage of total body mass. There is no statistically significant effect of genotype or sex on relative spleen mass, as calculated by two-way ANOVA with reps ($P \geqq 0.05$). N = 22. Error bars = ± one standard deviation. All WT and hom mice are age matched.

4.2.3 – Haematology of KLF1$^{StrepII\text{-}FBKP}$ mice

As spleen mass is unaffected by the KLF1$^{StrepII\text{-}FBKP}$ allele, it is unlikely that there are any severe erythropoietic abnormalities that would induce extramedullary erythropoiesis in the KLF1$^{StrepII\text{-}FBKP}$ mouse. However, given the wide variety of erythroid differentiation and haematopoietic cell fate functions governed by KLF1, I wanted to further investigate the potential haematological effects of KLF1$^{StrepII\text{-}FBKP}$.

Peripheral blood was obtained from 3 male and 3 female WT, het, and hom KLF1$^{StrepII\text{-}FBKP}$ mice and measured for red blood cell (RBC) parameters (**Fig. 4.6**; **Methods 2.4**; extended data in **Appendix Table A4.1**). All mean RBC parameters for het and hom mice were within, or close to, a normal range (**Fig. 4.6A, B**). At the sample size available, genotype does not significantly explain any of the variation in these data. Particularly, the normal reticulocyte counts suggest that the KLF1$^{StrepII\text{-}FBKP}$ mice are not under erythropoietic or anaemic stress and have a normal rate of RBC production in the bone marrow (**Fig. 4.6B**).

Likewise, blood films of hom KLF1$^{StrepII\text{-}FBKP}$ adult mouse peripheral blood appear indistinguishable from aged-matched WT controls by visual inspection (**Fig. 4.6C**). I therefore conclude that KLF1$^{StrepII\text{-}FBKP}$ has no effect on the tested red cell parameters.

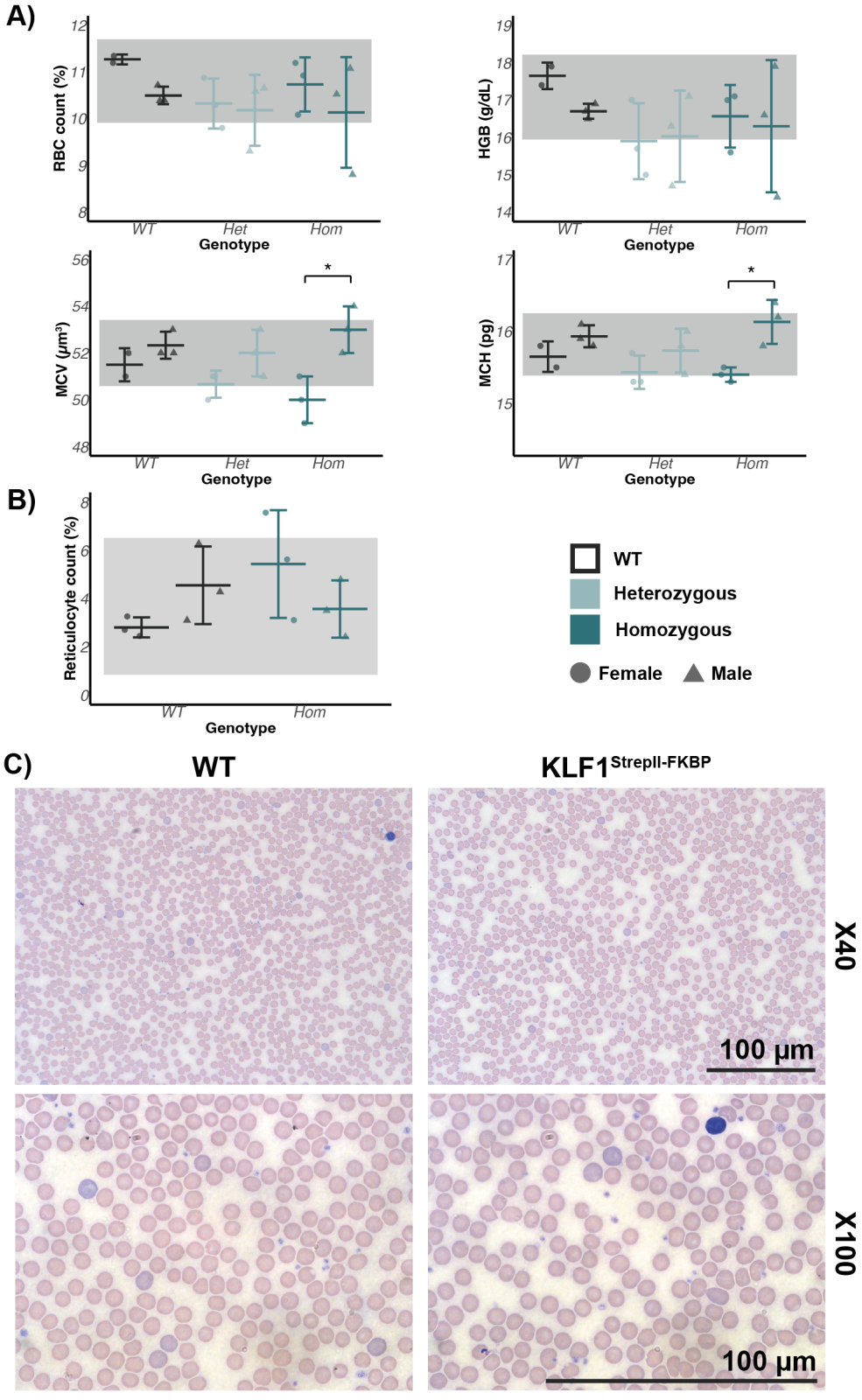

Fig. 4.6 – *KLF1$^{StrepII\text{-}FKBP}$ mice are haematologically normal*. A) Mean values for: red blood cell count (RBC; %), haemoglobin level (HGB; g/dL), mean corpuscular volume (MCV; μm^3), and mean corpuscular haemoglobin (MCH; pg) in age-matched WT, heterozygous (het), and homozygous (hom) adult peripheral blood. N = 3. Significant pairwise differences, determined by two-way ANOVA with Tukey's HSD post-hoc testing, are annotated as follows: "*" = 0.01 < *P* < 0.05. Extended data in **Appendix Table A4.1**. **B)** Mean reticulocyte counts (as % of total RBCs) from age-matched WT and hom adult peripheral blood. There is no significant effect of genotype on mean reticulocyte count (ANOVA: *P* = 0.374; N = 3, n = 460-772). Grey shading indicates the normal range per parameter, calculated as ± two standard deviations from the WT mean (**Raabe *et al.*, 2011**). Error bars = ± one standard deviation. **C)** Representative peripheral blood films from female, adult, WT, and hom KLF1$^{StrepII\text{-}FKBP}$ mice. Wright-Giemsa staining dyes basic structures (e.g. haemoglobin) red/pink and acidic structures (e.g. nuclei and RNA) blue/purple. Imaged at X40 and X100 magnification. Scale bars = 100 μm.

Similarly, there was minimal effect of KLF1$^{StrepII\text{-}FBKP}$ on white blood cell (WBC) haematology (**Fig 4.7**). In males, I observed no difference between WT and hom for all parameters measured. However, there were statistically significant differences between female WT and hom mice for total WBC, LYM, and, to a lesser extent, MON counts. In each case, the WT female data are consistently lower than all other measurements recorded for WT males, hom males, and hom females, which all lie within a close range. This raises the question of whether there normally is a sexual dimorphism in WBC counts, thus we are observing a sexually dimorphic effect of KLF1$^{StrepII\text{-}FBKP}$, or if the WT female result is anomalous itself.

Published results on sexual dimorphism in WBC populations, however, do not find a significant effect of sex on normal WBC or LYM counts (**Doeing *et al.*, 2003**). Indeed, in this publication, mean WBC and LYM counts from WT C57BL/6 mice are in a more similar range to the WT male, hom male, and hom female LYM counts than the WT female counts recorded here. This suggests that this difference in WT vs hom WBC/LYM/MON counts are unlikely to be caused by an increase in the hom counts, but rather in normal biological variation causing lower WT female counts.

Variation in all haematological counts can arise for many reasons, including factors such as the age, genetic background, nutrition, and stress (**O'Connell *et***

***al.*, 2015**). The WT controls used in this experiment were not littermate controls, so this observation may be reflective of such biological or handling variations (although all mice measured were age-matched and of C57BL/6 background). As such, it may be the WT female mice that significantly differ from the rest and there is potentially no biologically significant effect of KLF1$^{StrepII\text{-}FBKP}$ on WBC/LYM.

Overall, I conclude that, within the sample size available, KLF1$^{StrepII\text{-}FBKP}$ mice appear to have a normal WBC phenotype but additional biological replicates, ideally with WT littermate controls, are required to confirm this.

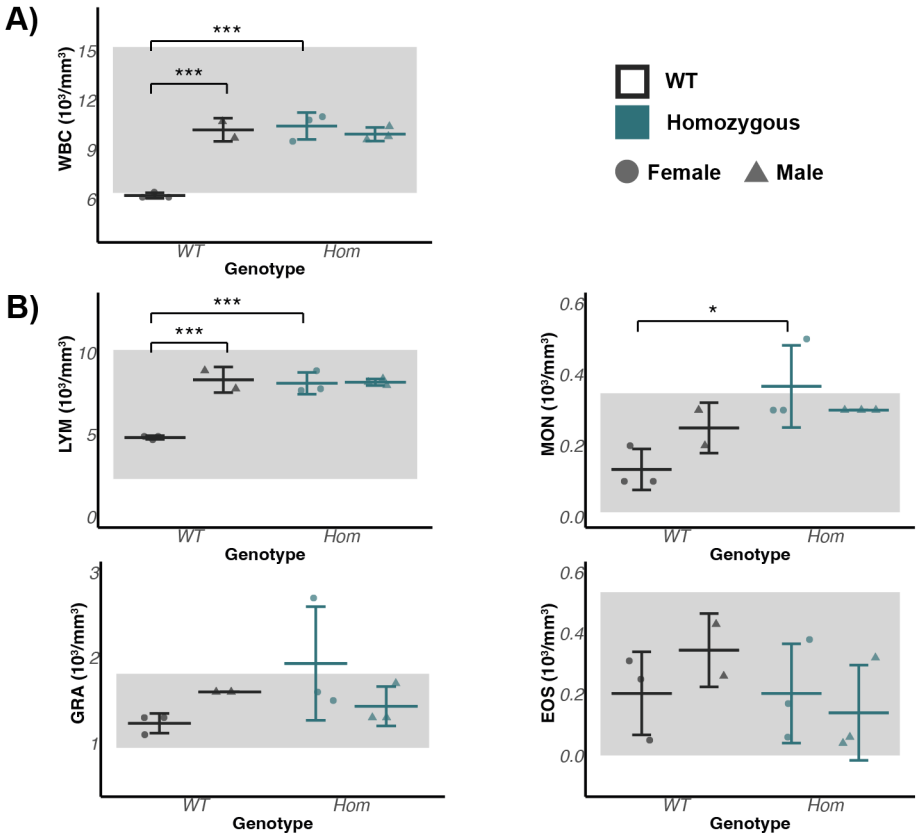

Figure 4.7 – *KLF1$^{StrepII\text{-}FKBP}$ mice have normal WBC counts.* A) Mean values for: total white blood cell count (WBC; 10^3/mm^3). **B)** Mean values for lymphocyte count (LYM; 10^3/mm^3), monocyte count (MON; 10^3/mm^3), granulocyte count (GRA; 10^3/mm^3), and eosinophil count (EOS; 10^3/mm^3) in age-matched WT and homozygous (hom) adult peripheral blood. N = 3. Significant pairwise differences, determined by two-way ANOVA with Tukey's HSD post-hoc testing, are annotated as follows: "*" = 0.01 ≦ P < 0.05; "**" = 0.001 ≦ P < 0.01; "***" = P < 0.001. Grey shading indicates the normal range per parameter, calculated as ± two standard deviations from the WT mean (**Raabe *et al.*, 2011**). Error bars = ± one standard deviation.

4.2.4 – Assessing the erythroid phenotype in cultured KLF1$^{StrepII\text{-}FBKP}$ mFL cells

Next, I wanted to affirm that there was a normal RBC phenotype in the primary cell culture system to be used in subsequent experiments.

The mouse fetal liver (mFL) is a major site of definitive haematopoiesis in embryos from ~E12.5 to birth. As such, a mFL is typically comprised near entirely of developing erythroblasts, from HSPCs to enucleated reticulocytes. A typical E12.5 mFL contains >95% CD71^{+}/Ter119^{+} erythroblasts (S1-S3; **Introduction Fig. 4.2B, C**). Easy to harvest, well characterised in the literature, and devoid of the genetic and phenotypic inconsistencies of cultured cell lines, mFLs are a prime source of definitive erythroid cells. Moreover, while one E12.5 mFL can yield ~1-5 million cells, mFL erythroblasts grow well in culture and can be easily expanded, typically doubling with each day of culture for up to 7 days. This has several benefits. mFL culture abrogates the need for column-based selection to remove the few contaminating cell lineages, as only erythroid cells persist in the culture system. Additionally, *ex vivo* culture allows researchers to reduce overall mouse use, by expanding, freezing, and recovering mFL cells from a single litter for multiple experiments. Lastly and most importantly for this work, when held in so-called expansion media (**Methods 2.5.3**) there is proliferation of all erythroid cells, including some early erythroblasts (from HSPC, BFU-E, and CFU-E cells), allowing the researcher to maintain and expand a complete complement of erythroid cells, with only a slight drift toward later stage (S3-S5) cells. This allows the study of experimental manipulation effects at all stages of erythropoiesis. It is widely accepted that the mFL definitive cells are comparable to the definitive RBCs of the adult bone marrow and peripheral blood.

Therefore, I wanted to affirm that there was normal proliferation and phenotype of the cultured E12.5 KLF1$^{StrepII\text{-}FBKP}$ mFL cells by qualitative and quantitative observations of the culture system. All biological replicates of hom mFL cell cultures consistently had expected, good rates of proliferation (~doubling every 24hrs) and cell viability (>90%, data not shown). Cytospins from bulk, cultured

KLF1$^{StrepII\text{-}FBKP}$ mFL cells show that, throughout cell culture, there is a consistent normal complement of early to late mFL cells with normal morphologies (**Fig. 4.8**).

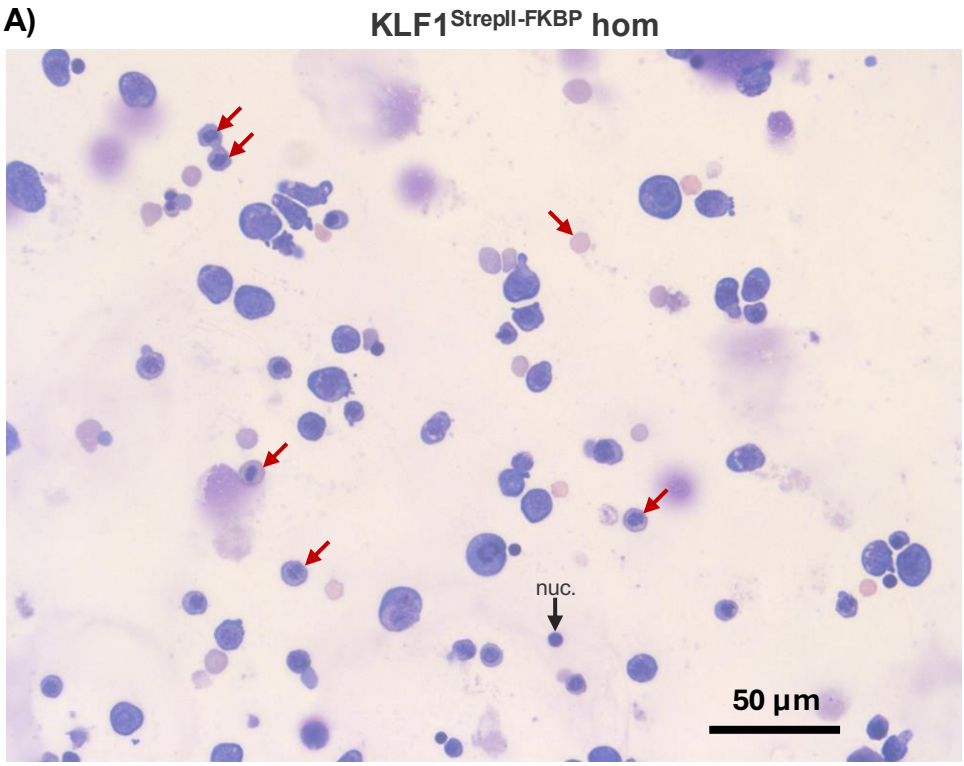

Fig. 4.8 – ***There is a full range of normal, terminally differentiating, erythroid cells in KLF1$^{StrepII\text{-}FKBP}$ mFL cell cultures*****. A)** E12.5 KLF1$^{StrepII\text{-}FKBP}$ mFL cells, cultured for 5 days in expansion media. Red arrows point to examples of late-stage erythroblasts present in the culture, indicating normal erythroid differentiation progression and enucleation. "nuc." = an enucleated erythroid cell nucleus. Scale bar = 50 µm. Imaged at X40 magnification.

I next wanted to quantitively confirm if the distribution of differentiating erythroid populations within cultured KLF1$^{StrepII\text{-}FBKP}$ mFL cells was normal. For this, I did flow cytometry on primary cultures of WT, KLF1$^{StrepII\text{-}FBKP}$ het, and KLF1$^{StrepII\text{-}FBKP}$ hom mFL cells (**Fig. 4.9A**). The WT, het, and hom cultures all show the expected "horseshoe" distribution when immunophenotyped using the erythroid cell surface markers, CD71 and Ter119 (see **Introduction Fig. 4.2**). As is typical, WT and hom each display a slight enrichment for the late S3 population. Overall, the percentages of each sub-population (S0-S4/5) are within typical ranges (**Fig. 4.9B**).

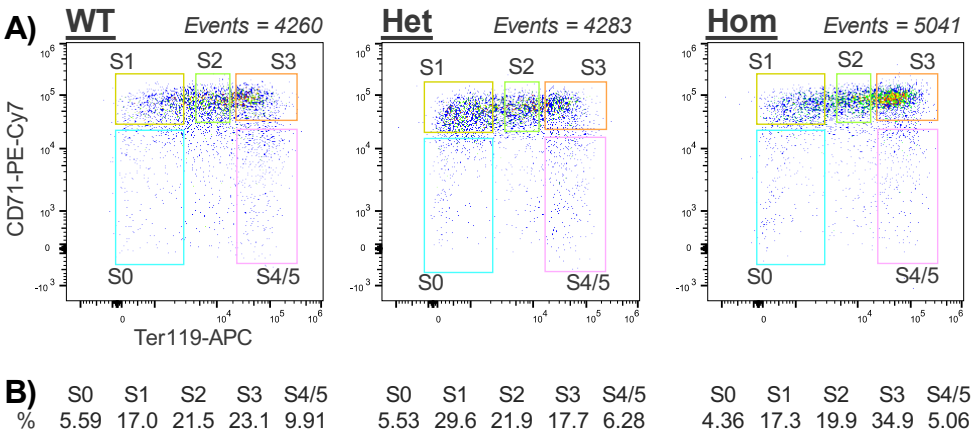

Fig 4.9 – *The distribution of erythroid populations is equal across genotypes.* A) Flow cytometry on E12.5 mFL cells cultured for 5 days in expansion media and sorted for CD71 and Ter119 expression. "S0" = CD71$^{low/mid}$/Ter119^{low}. "S1" = CD71^{high}/Ter119^{low}. "S2" = CD71^{high}/Ter119^{mid}. "S3" = CD71^{high}/Ter119^{high}. "S4/5" = CD71$^{low/mid}$/Ter119^{high} **B)** Percentage of cells per gate for each sample. N = 1.

It is known that KLF1$^{-/-}$ cells do not express Ter119 (**Siatecka and Bieker, 2011**). I therefore wanted to verify that the CD71/Ter119 FACS sorted S0-S4/5 populations in my primary KLF1$^{StrepII\text{-}FBKP}$ hom cultures, which should still have functional KLF1 protein, correspond to the expected, Ter119-expression-stratified cell types. **Fig. 4.10** shows images of cells from cultured and sorted KLF1$^{StrepII\text{-}FBKP}$ hom primary mFL samples. This confirms that cells from each FACS gate morphologically conform to the expected erythroid differentiation stages (see **Introduction Fig. 4.2B, C**).

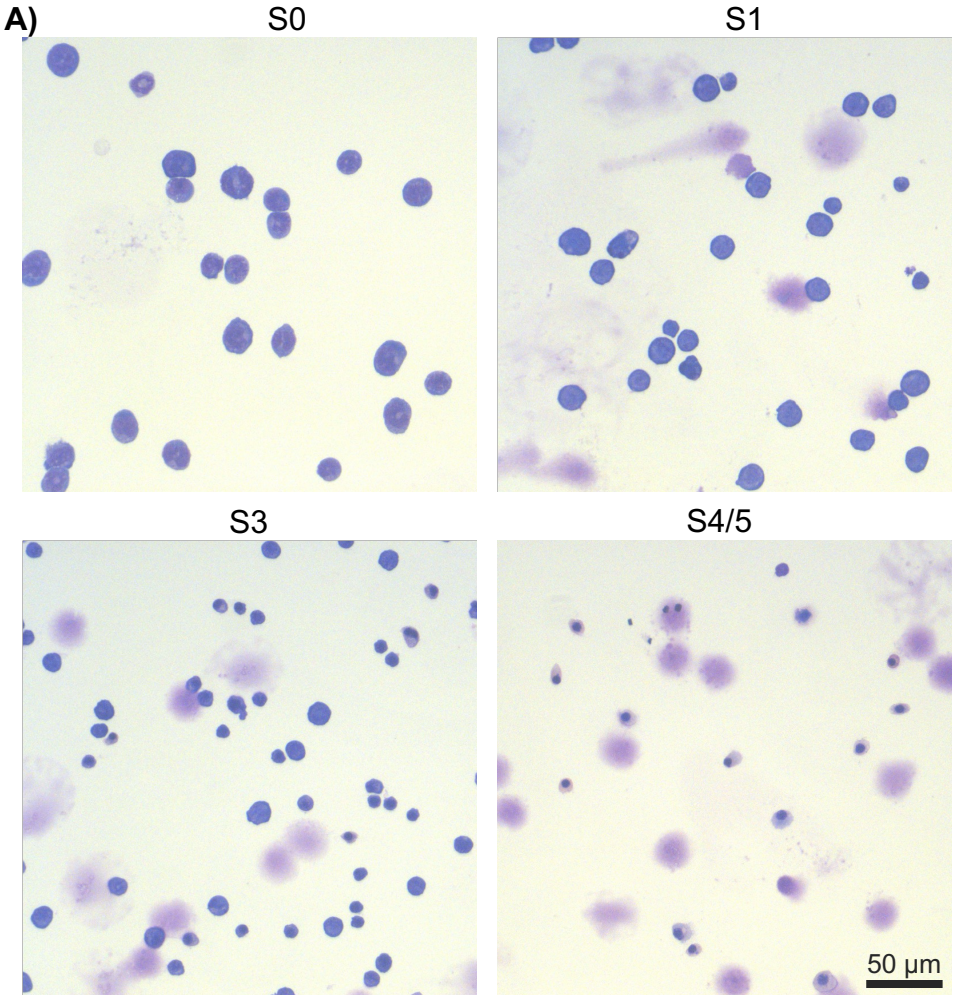

Fig. 4.10 – *mFL cells are sorted into S0-S4/5 erythroid populations after FACs.* A) E12.5 KLF1$^{StrepII\text{-}FKBP}$ mFL cultured cells, FACS sorted by CD71/Ter119 expression. "S0": CD71$^{low/mid}$/Ter119^{low}. "S1": CD71^{high}/Ter119^{low}. "S2": CD71^{high}/Ter119^{mid}. "S3": CD71^{high}/Ter119^{high}. "S4/5": CD71$^{low/mid}$/Ter119^{high}. Scale bar = 50 µm. Imaged at X40 magnification.

Overall, I conclude that the KLF1$^{StrepII\text{-}FBKP}$ mFL cells have normal growth, are morphologically normal, and follow a normal pattern of CD71 and Ter11 expression which can be used to stratify the mFL cells by an established method of terminal erythroid differentiation staging.

4.2.5 – Verifying the StrepII tag with Western blot

Having established the KLF1$^{StrepII\text{-}FKBP}$ mouse was phenotypically normal, I next wanted to determine whether the StrepII-tag was working and could be used to ChIP for KLF1.

First, I performed a Western blot on whole cell lysates from WT, het, and hom KLF1$^{StrepII\text{-}FKBP}$ E12.5 mFL cells. As expected, there was no KLF1$^{StrepII\text{-}FKBP}$ in WT and a strong KLF1$^{StrepII\text{-}FKBP}$ doublet signal in the hom mFL lane at ~70kDa (**Fig. 4.11A**). The expected size of KLF1$^{StrepII\text{-}FKBP}$ is 53.7 kDa, however several factors, such as the denatured conformation and PTMs of a protein, can affect migration through the polyacrylamide gel, so it is not a major concern that KLF1$^{StrepII\text{-}FKBP}$ runs higher than expected. Indeed, WT KLF1 is 37.7 kDa in size, but has been documented to run closer to ~45 kDa in published Western blots (**Southwood *et al.,* 1996**). It is also well recorded for KLF1 to run as a doublet in SDS-PAGE gels, due to the differential binding of interacting proteins to KLF1 in the cytoplasm vs nucleus (**Southwood *et al.*, 1996; Quadrini *et al.*, 2008; Yien and Bieker, 2013**).

The signal intensity of the KLF1$^{StrepII\text{-}FKBP}$ doublet in the het lane is ~33% of the hom doublet (**Fig. 4.11B**). Western blot is semi-quantitative and, despite best efforts, is invariably subject to variations in protein concentration during lysate preparation and gel loading. This quantification should therefore not be interpreted as an absolute value. However, it could be speculated that, if the tagged protein does have higher turnover or lower translation rates than the endogenous protein, then there may be a true ~70:30 KLF1:KLF1$^{StrepII\text{-}FKBP}$ ratio in het samples. Overall, this Western blot confirms that the StrepII tag works for specific and semi-quantitative detection of KLF1$^{StrepII\text{-}FKBP}$.

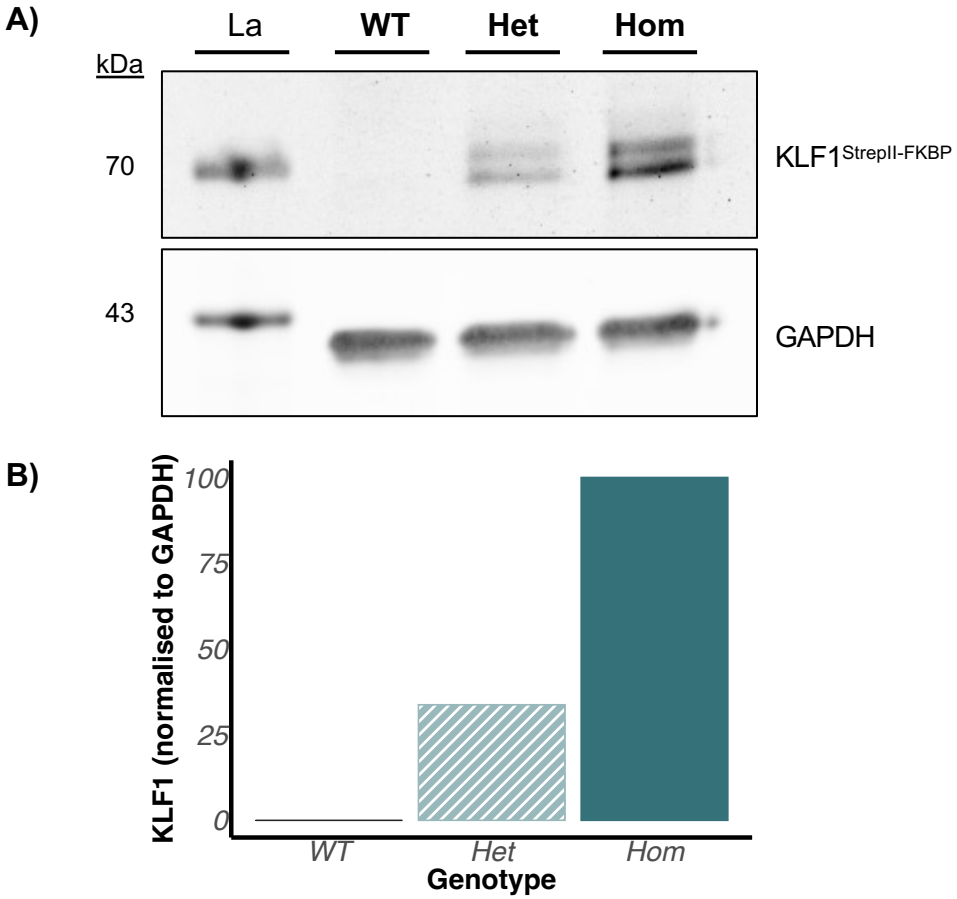

Fig 4.11 –*KLF1$^{StrepII\text{-}FKBP}$ protein is expressed and detected in het and hom, but not WT mFL cells.* **A)** Fluorescent Western blot of whole cell lysates from WT, het, and hom, cultured E12.5 mFL cells. KLF1$^{StrepII\text{-}FKBP}$ was detected with an anti-StrepII antibody (Qiagen, 34850). "La" = protein ladder lane. **B)** Quantification of Western blot data. KLF1$^{StrepII\text{-}FKBP}$ signal intensity was normalised to GAPDH signal intensity per lane and plotted as a percentage of the signal intensity of the hom KLF1$^{StrepII\text{-}FKBP}$ band. No KLF1$^{StrepII\text{-}FKBP}$ signal was detected in the WT lane. *N* = 1. Full details of antibodies and analysis are available in **Methods 2.7.2.**

4.2.6 – KLF1$^{StrepII\text{-}FKBP}$ bulk chipmentation vs published KLF1 ChIP

One of the major hurdles to studying KLF1 is dispute over anti-KLF1 antibody specificity. I sought to overcome this by using an anti-StrepII antibody for ChIP pulldown of KLF1$^{StrepII\text{-}FKBP}$. Given the high specificity and strength of binding between StrepII and anti-StrepII antibodies – engineered analogues of biotin and streptavidin – I also aimed to reduce the cell input requirements by using chipmentation, a low cell-input ChIP-seq method.

To my surprise, even after several rounds of optimisations, the KLF1$^{StrepII\text{-}FKBP}$ chipmentation yielded consistently low signal to noise ratio (SNR) (**Fig. 4.12**). Nevertheless, I proceeded to comparative analysis with existing published KLF1 ChIP-seq datasets, to assess whether the chipmentation was detecting true KLF1 peaks and worth further optimisation. For this comparison, I reanalysed the two most recently published ChIP-seq data sets in primary WT mFL. These were generated using the two most widely approved anti-KLF1 antibodies, one from the Bieker laboratory (**Mukherjee and Bieker, 2022**) and one from the Perkins laboratory (**Huang *et al.*, 2024**). The three datasets are henceforth referred to as Holliman, Mukherjee, and Huang. For comparison of ChIP conditions and published analysis results, see **Table 4.4**.

Table 4.4 – *Summary of previous KLF1 ChIP-seq experiments in mouse erythroblasts*. * Huang sequencing library generated from 2 pooled replicates. “n/a” = data not available.

Reference	Genotype	Cell type	Replicates	Antibody	Cell # per IP	GEO accession:	# Peaks published	# Peaks re-analysed
Mukherjee & Bieker, 2022	WT (vs nan/+)	E13.5 mFL	3 biological replicates	Ms monoclonal anti-KLF1 “7B2a”	7.5 x 10^6	GSE210779	9150	574
Huang *et al.*, 2024	WT (vs KLF1$^{D45\text{-}mutant}$)	E14.5 mFL	1*	Rb polyclonal Anti-KLF1 (**Perkins *et al.*, 1995**)	n/a	GSE94351	1144	1210
Holliman	KLF1$^{StrepII\text{-}FKBP}$	E12.5 mFL	3 technical replicates	Anti-StrepII	2.5 x 10^5	-	-	174

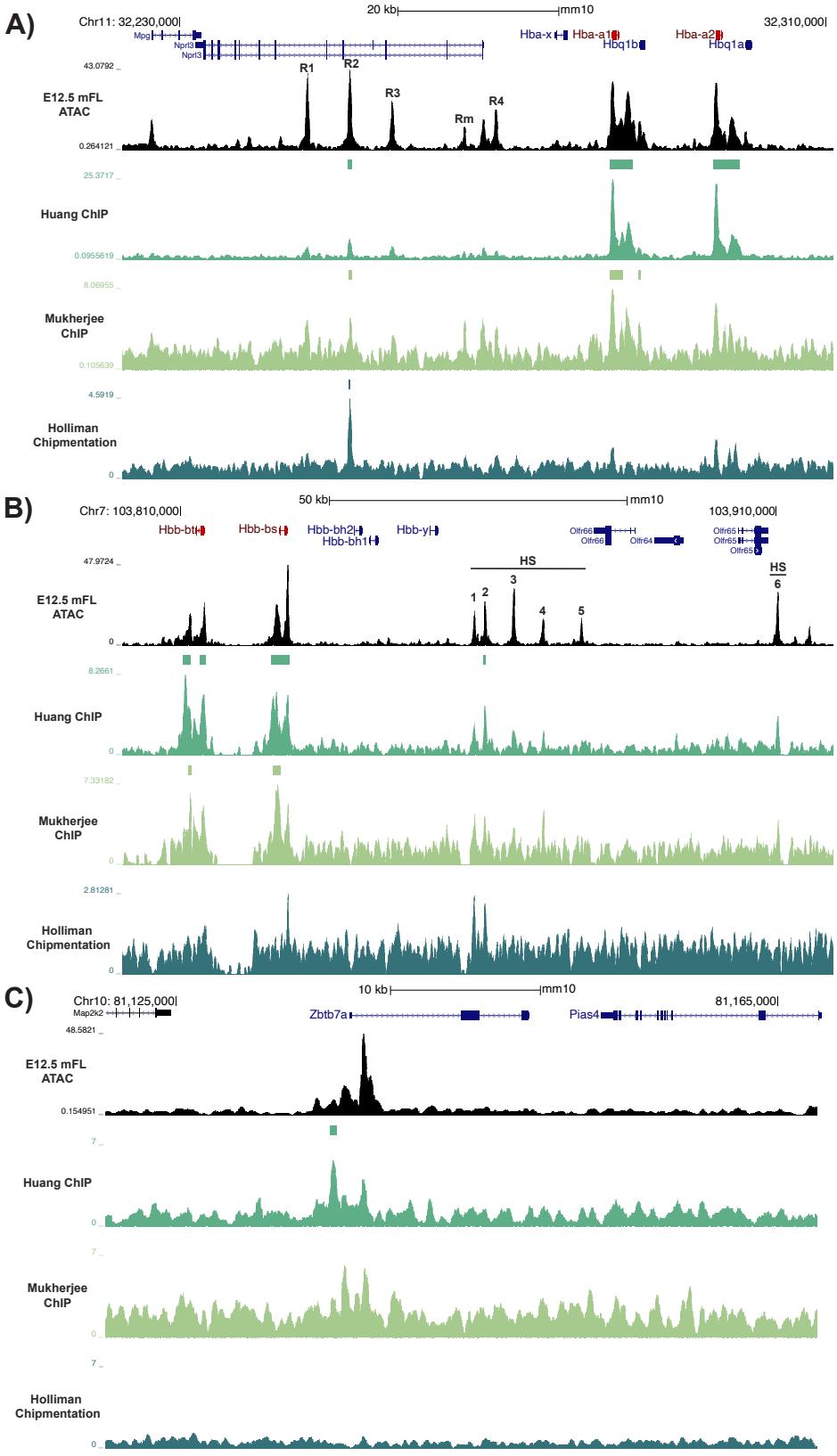

Fig. 4.12 – *KLF1 peaks in Holliman and Mukherjee are sometimes below the threshold of peak detection*. UCSC tracks comparing three KLF1 ChIP experiments: Huang *et al*., 2024 (GSE94351); Mukerjee and Bieker, 2022 (GSE210779); this thesis (**Methods 2.9.2**). At **A)** α-globin locus, **B)** β-globin locus, and **C)** *Lrf* (*Zbtb7a*). UCSC

gene annotations and scale bars are at the top of each panel, with adult α- and β-globin genes highlighted in red. ATAC tracks in E12.5 mFL (**Blayney *et al.*, 2023**; GSE220463) are below in black. The α-globin enhancers are annotated "R1-3, Rm, R4" and the β-globin locus control region "HS1-6". Peaks, called with Lanceotron, are annotated as coloured bars directly above each ChIP track.

There were 174 peaks called in Holliman, 574 in Mukherjee, and 1210 in Huang. Approximately one quarter of all peaks in Holliman and Mukherjee are present in Huang (**Fig. 4.13A**). However, there are only eight common peaks between Holliman and Mukherjee. Despite this, *de novo* motif discovery revealed enrichment of the core KLF1 motif (C[A/G]CCC) at peak summits in all three datasets, suggesting that each has successfully pulled down and called true KLF1 peaks (**Fig. 4.13B**).

As well as calling the greatest number of peaks, the Huang dataset has very strong ChIP signal intensity (**Fig. 4.12**; **Fig. 4.13C-i**). Meanwhile, the range of signal intensities at the Mukherjee peaks are much lower than in Huang, indicating a less sensitive detection of KLF1 binding events. Conversely, the Holliman peaks have relatively strong signal at all called sites, despite having the fewest total peaks (**Fig. 4.13C-iii**). As the Holliman peaks are enriched for the KLF1 motif and have similar signal intensities in the Huang dataset (**Fig. 4.13C-iv**), this suggests that, despite not all being called as peaks in the Huang dataset, they are real KLF1 peaks. With the same true for Mukherjee peaks (**Fig. 4.13B, Fig. 4.13C-vi**) this would mean that there could be closer to >1500 true KLF1 bound sites in the E12-14 mFL genome. Finally, visual inspection of UCSC tracks, e.g. at the *Hba-a1/2* and *Hbb-bt/s* and their enhancers, show that there may be many more KLF1 peaks in Holliman and Mukherjee that are below the threshold for peak detection and could be improved upon with optimisations to SNR (**Fig. 4.12**).

Interestingly, in line with **Fig. 4.13A**, there is almost no signal in Mukherjee at the sites of Holliman peaks and vice versa (**Fig. 4.13C-v**), despite both having greater similarity to Huang. This suggests that the Mukherjee and Holliman ChIPs might be capturing a different subset of true KLF1 peaks. This exemplifies the need for

verification of antibody specificity in methods such as ChIP. For instance, perhaps the Mukherjee anti-KLF1 and the Holliman anti-StrepII antibodies are biased towards detecting KLF1 in different chromatin contexts (e.g. in an activating complex with SWI/SNF vs a repressive complex with NuRD). Whereas the anti-KLF1 antibody used by Huang might be able to detect KLF1 in both contexts. Further bioinformatic or experimental analysis would be required to determine this. Such comparisons to reveal differences in the chromatin/sequence context of the differentially called peaks could include analysis of: chromatin markers (e.g. H3K27ac, H3K4me1, H3K4me3) at called peaks; the identity of peaks (e.g. at gene promoters, introns, or enhancers); gene ontology of peaks over gene promoters. Alternatively, another explanation could simply be a lack of sensitivity of the Holliman and Mukherjee datasets, which could be improved upon with further optimisations. Certainly, there is room for further optimisation of the anti-KLF1$^{StrepII\text{-}FKBP}$ chipmentation protocol, discussed further in **Discussion 4.3**.

Notably, my chipmentation protocol used in total ~750k cells, compared to 22.5 million cells used in Mukherjee. Despite my chipmentation having called ~1/3 of the total number of peaks as the Mukherjee ChIP, I maintained a similar accuracy (25 % peaks also detected in Huang) while using ~30-fold fewer cells. This is very promising as a basis for developing the KLF1$^{StrepII\text{-}FKBP}$ chipmentation.

Regarding the Huang ChIP, as well as having the strongest overall signal strength, my analysis identified the total number of Huang KLF1 peaks to be in a similar range to the ~1000 KLF1 dependent genes predicted from RNA-seq experiments in KLF1$^{-/-}$ mice (**Tallack *et al.*, 2012**). One conclusion from this comparative analysis is to consider the Huang dataset as a tentative “gold standard” KLF1 ChIP. Although this will require proven repeatability of the Huang ChIP results, particularly while questions of antibody specificity remain, this could provide a benchmark with which to compare future KLF1$^{StrepII\text{-}FKBP}$ chipmentation optimisations.

In conclusion, I believe it will be very valuable to continue optimising the KLF1$^{StrepII\text{-}FKBP}$ chipmentation protocol, to not only provide independent verification of the peaks identified by Huang *et al.*, but also to enable researchers

to examine KLF1 binding in more experimental contexts where cell number is limiting.

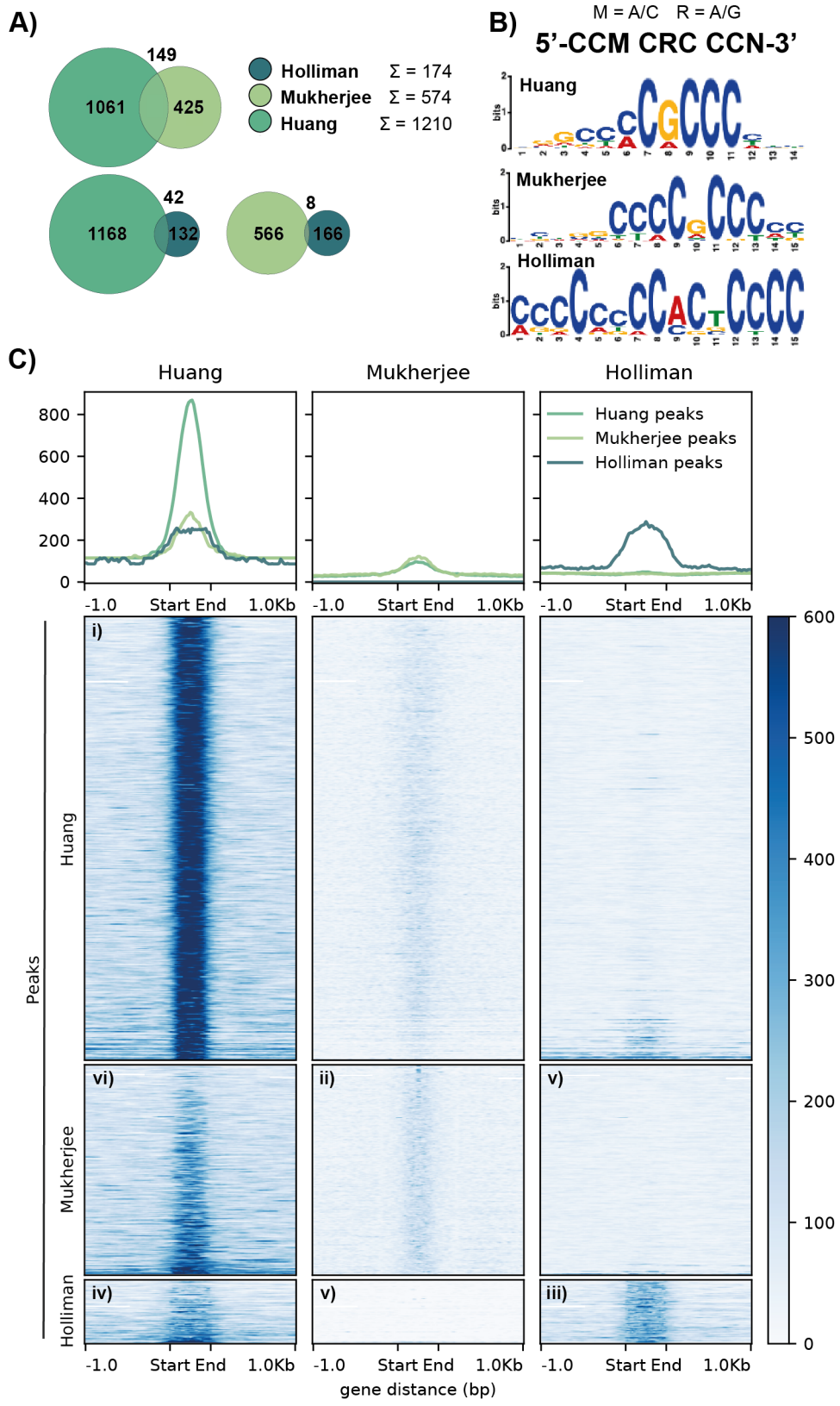

Fig. 4.13 – *Meta-analysis of Huang, Mukherjee, and Holliman KLF1 ChIP-seq.* A) Venn diagrams of total peaks called in Huang ChIP (**Huang *et al.*, 2024**; GSE94351),

Mukherjee ChIP (**Mukherjee *et al.*, 2022**; GSE210779) and Holliman chipmentation (this thesis; **Methods 2.9.2**) datasets. **B)** *De novo* motif discovery reveals enrichment of the core KLF1 motif (C[A/G]CCC) at peak summits in all three datasets. **C)** ChIP/chipmentation signal intensity (RPKM) averaged line plots (upper panel) and heatmaps (lower panels) at peak summits ± 1 kb in Huang, Mukherjee, and Holliman datasets.

4.2.7 – KLF1 S0-S4/5 sorted chipmentation

A major benefit of chipmentation is that it has low enough cell input requirements to facilitate use with FACS-sorted cell populations. This enables precise studies of TF recruitment and binding dynamics in rare populations that until recently have been below the limit of detection. In our laboratory, we have previously used the CD71/Ter119 FACS protocol (**Introduction Fig. 4.2**; **Method 2.6**) to perform CUT&RUN – a similar, low cell input ChIP method – for key erythroid TFs (GATA1, GATA2, NF-E2) and histone marks (H3K27ac/me3, H3K4me1/3; **R. Beagrie, unpublished data**). Not only would sorted-cell chipmentation for KLF1$^{StrepII\text{-}FKBP}$ add another key erythroid TF to our existing knowledge of chromatin dynamics in mFL erythropoiesis, but it would also add an unprecedented temporal resolution to existing KLF1 ChIP-seq data.

To this end, I sorted E12.5 cultured mFL cells into the S1, S2, S3, and S4/5 populations by CD71/Ter119 FACS before performing anti-StrepII chipmentation for KLF1$^{StrepII\text{-}FKBP}$. UCSC tracks at the α- and β-globin loci, from two averaged biological replicates, are displayed in **Fig. 4.14**. This revealed two very interesting results.

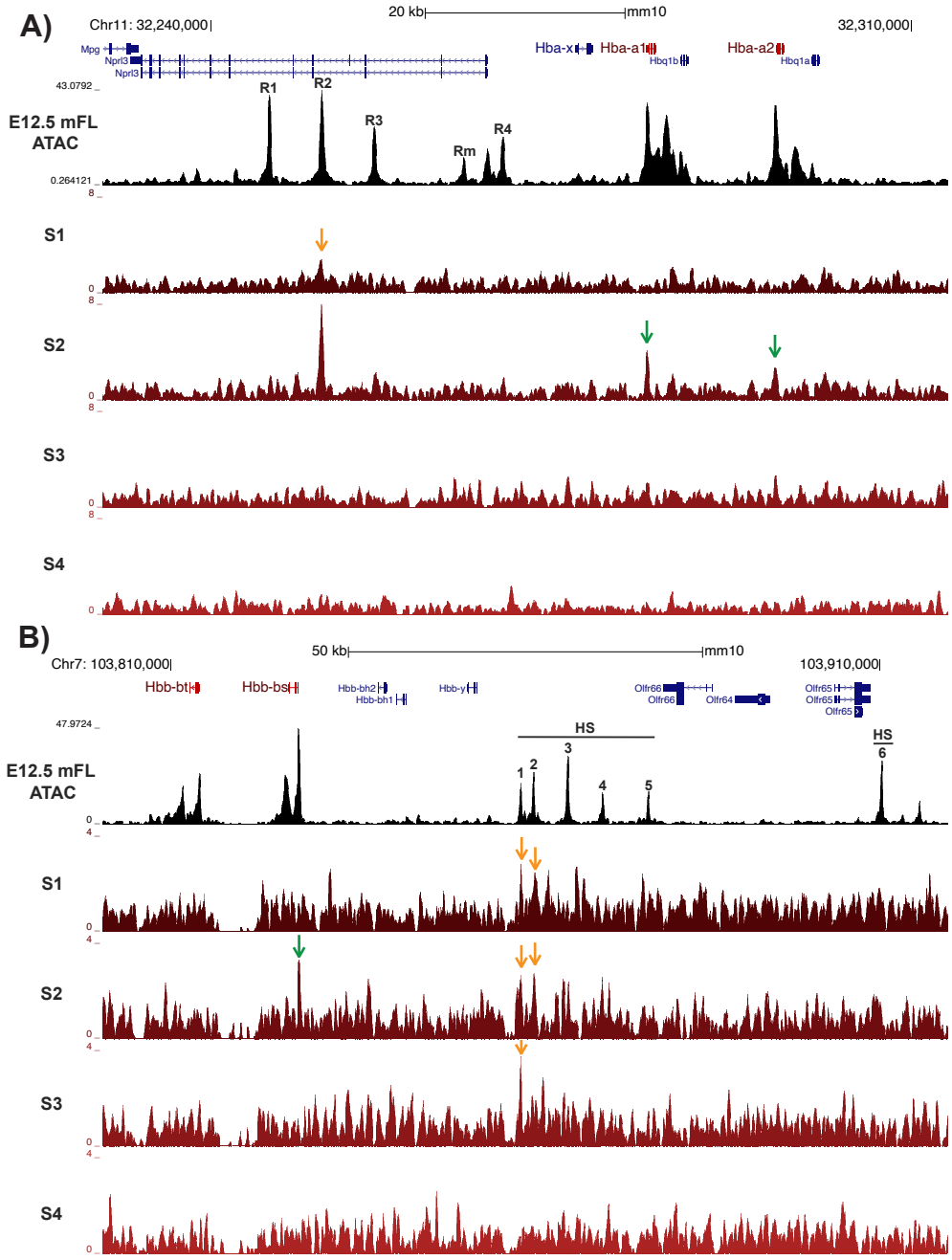

Fig. 4.14 – *KLF1$^{StrepII\text{-}FKBP}$ is detected at the strongest superenhancer elements before the adult globin gene promoters.* UCSC tracks of KLF1$^{StrepII\text{-}FKBP}$ chipmentation, at **A)** α-globin locus and **B)** β-globin locus. From E12.5 mFL cells, FACS sorted by CD71/Ter119 expression. "S1": CD71^{high}/Ter119^{low}. "S2": CD71^{high}/Ter119^{mid}. "S3": CD71^{high}/Ter119^{high}. "S4/5": CD71$^{low/mid}$/Ter119^{high}. UCSC gene annotations and scale bars are at the top of each panel, with adult α- and β-globin genes highlighted in red. ATAC tracks in E12.5 mFL (**Blayney *et al.*, 2023**; GSE220463) are below in black. The α-globin enhancers are annotated "R1-3, Rm, R4" and the β-globin locus control region "HS1-6". Orange arrows indicate putative KLF1$^{StrepII\text{-}FKBP}$ peaks at globin enhancers. Green arrows indicate KLF1$^{StrepII\text{-}FKBP}$ peaks at adult globin gene promoters. Tracks are averages from 2 independent biological replicates.

First, at all inspected loci and in both replicates (data not shown), the KLF1$^{StrepII\text{-}FKBP}$ signal is far superior in the S2 cells than any other fraction (**Fig. 4.14**). This observation will require further verification, but there is additional, circumstantial, supporting evidence for this. First, KLF1 mRNA expression is known to be highest in S2 cells (**Introduction Fig. 4.2, Oudelaar and Beagrie 2020**). Second, KLF1 is known to interact and cooperatively bind DNA with GATA1. We have previously observed that GATA1 CUT&RUN signal strength is low in S0, high in S1 and S2, and negligible in S3 cells (**Appendix Fig. A4.3; Beagrie, unpublished**). As S2 cells consistently make up ~20% of a mFL culture **(Fig. 4.9**), this raises the question as to what extent the high SNR observed in previously published KLF1 ChIP data has been due to the use of bulk mFL samples in which ~80% of the material would not yield a KLF1 signal, perhaps in addition to variable antibody quality.

Secondly, at both adult globin loci KLF1 appears to bind the enhancer elements in S1 cells (orange arrows) before binding globin gene promoters in S2 cells (green arrows; **Fig. 4.14**). Perhaps KLF1 is first recruited to adult globin enhancers and then chromatin folding brings KLF1 into contact with the globin genes. This would fit with previously published data on chromatin folding dynamics at the α-globin locus where enhancer-promoter contacts are of low frequency in S1 cells but greatly increase as cell move into S2 and S3 (**Oudelaar *et al.*, 2020**). An alternative explanation could be technical limitations masking a true but perhaps weaker KLF1 ChIP signal in S1 cells, meaning that only signal from the strongest peaks, e.g. R2 enhancer, is detected here. If KLF1 were present but less well detected in S1 cells, this would also be more in line with our known GATA1 binding profile (**Appendix Fig. A4.3; R. Beagrie, unpublished**). Further optimisation of this chipmentation will be essential to distinguish between these possibilities.

4.3 – Discussion

In this chapter, I have found that adult KLF1$^{StrepII\text{-}FKBP}$ mice and their cultured mFL cells are haematologically normal and morphologically indistinguishable from WT. Despite this, I did observe a significant reduction in the number of KLF1$^{StrepII\text{-}FKBP}$ hom mice present at weaning age. By process of elimination, I determined that any loss of hom KLF1$^{StrepII\text{-}FKBP}$ pups is likely occurring in late gestation, between E12.5 and birth. Notably, KLF1$^{-/-}$ embryos are known to perish between ~E14.5-E16.5 (**Nuez *et al.*, 1995; Perkins *et al.*, 1995**), thought to be due to the timing of the primitive-to-definitive erythropoietic switch. If there is a slight effect of the StrepII-FKBP tags on KLF1 function or turnover, it would appear that the KLF1$^{StrepII\text{-}FKBP}$ embryos are sensitive to this at the same developmental window.

Notably, it has been observed in other mouse models with FKBP or similar degron tags on endogenous proteins of interest (POI), that the presence of the degron tag can interfere with POI function or turnover (**Morawska and Ulrich, 2013; Nora *et al.*, 2017; Layden *et al.*, 2021**). For example, **Mehta *et al.*, 2022**, describe their BCL11A^{FKBP} mouse model, harbouring a C-terminal FKBP12^{F36V} tag at endogenous *Bcl11a*, as having such a dramatic effect that homozygous BCL11A^{FKBP} mice, without any dTAG-13 treatment, phenocopy *Bcl11a* knockout mice. BCL11A^{FKBP} mice are born in non-mendelian ratios, with live hom pups dying shortly after birth. By cycloheximide pulse-chase and Western blot experiments, they determined that the FKBP tag had reduced the half-life of BCL11A from ~24.5 hours to ~7 hours, resulting in significantly reduced BCL11A protein abundance. While I do not observe such a dramatic effect in my KLF1$^{StrepII\text{-}FKBP}$ model, it is entirely possible that there is a small effect of the FKBP peptide on KLF1 function or turnover in my model, which could account for the non-mendelian inheritance of KLF1$^{StrepII\text{-}FKBP}$ and the very modest trends observed in RBC and WBC haematology data.

It is possible, therefore, that the allelic skew I observe is caused by reduced function or stability of KLF1, resulting in loss of pups later in gestation when definitive erythropoiesis dominates. An ideal experiment to investigate this further would have been a series of het x het crosses with an embryonic time course of

morphological assessments and genotyping of the resulting embryos (e.g. daily from E10.5-20.5) to determine the timing and cause of death. However, I decided the cost of mouse life for this was too great to be justified, particularly considering how the allelic skew is relatively mild, could still be due to random chance, and there is no discernible phenotype in the adult hom mice or FL cultures that prevents us from using this model, regardless.

If dysfunction of KLF1$^{StrepII\text{-}FKBP}$ is suspected, it seems most likely for that to also be due to increased protein turnover, as in the findings of **Mehta *et al.*, 2022**. To test this further, I would like to perform a cycloheximide pulse-chase experiment to determine the half-life of KLF1$^{StrepII\text{-}FKBP}$, compared to endogenous KLF1's known half-life of 6-9hrs (**Quadrini and Bieker, 2006**). However, as the StrepII-FKBP tags are proximal to the ZnF domains of KLF1, they could also be affecting its DNA binding function. An EMSA assay could make a semi-quantitative assessment of this. However, with the limited time available, I chose not to prioritise these experiments, particularly as KLF1 function appears to be normal.

Nevertheless, the potential for reduced stability of KLF1$^{StrepII\text{-}FKBP}$ is curious and it is valuable to bear in mind that it is near impossible to modify a protein sequence and not affect its function and/or stability to some degree. But, as long as the appropriate internal controls are used, I conclude that this model is suitable to take forward for subsequent functional studies of KLF1.

My next aim for this chapter was to establish a chipmentation for KLF1$^{StrepII\text{-}FKBP}$ to verify the genome-wide binding profile of KLF1, free from the bias of questionable anti-KLF1 antibodies. I show that KLF1$^{StrepII\text{-}FKBP}$ can be detected using an anti-StrepII antibody in both Western blot and chipmentation assays, although further optimisation of the chipmentation is required. Nonetheless, by comparing my chipmentation data to two recently published KLF1 ChIP datasets (**Huang *et al.*, 2024 and Mukherjee and Bieker, 2022**), I saw that there was relatively little concordance in the results from the two most mainstream anti-KLF1 antibodies. This exemplifies doubts regarding existing anti-KLF1 antibody sensitivity and specificity, highlighting the need for independent verification of ChIP-style data acquired with any anti-KLF1 antibody. I also found that my

KLF1$^{StrepII\text{-}FKBP}$ chipmentation peaks overlapped more with the Huang ChIP, suggesting that the Perkins antibody use by Huang *et al*. might be better suited to detecting KLF1 binding. However, this anti-KLF1 antibody is not commercially available, will be subject to batch-variations, and still requires independent verification. I conclude that continued chipmentation optimisations in the KLF1$^{StrepII\text{-}FKBP}$ model will be very valuable to the field.

A significant benefit of chipmentation over ChIP-seq is the low cell-input requirement. My final aim for this chapter was to exploit this to uncover, for the first time, the genome-wide binding profile of KLF1 in erythroblast populations stratified by differentiation. From my results, it appears that KLF1 may only – or most strongly – be binding chromatin in the S2 (proerythroblast) fraction of cells. It is very intriguing to speculate whether part of the difficulties faced by those trying to ChIP for KLF1 over the last decade has been due to the majority of cells in a bulk mFL population having intrinsically low or no KLF1 binding. This result exemplifies the need to have established, and next to optimise, a low cell-input chipmentation method for KLF1.

For future work, there are several aspects of the KLF1$^{StrepII\text{-}FKBP}$ chipmentation that I would optimise. First would be to attempt chipmentation in S1/S2 (CD71^{+}/Ter119$^{-ve/low}$) -enriched erythroblasts. There are also additional, universal tricks to optimise a chipmentation, such as pooling multiple IPs before library preparation to increase library complexity and averaging sequencing results from additional technical and/or biological replicates. However, these will need to be balanced against total cell number requirements. As such, steps like titrating the volume of protein A/G beads and optimising chipmentation wash buffer stringencies might be preferable starting points. Another optimisation to consider is the issue of chipmentation input tracks.

When there is low SNR in a ChIP-seq-style experiment, subtracting the signal of an input from the sample signal can reduce background and improve SNR. A ChIP-seq input typically involves taking 2-10% of sonicated chromatin from the sample, incubating it with beads without any antibody, washing the beads, and thus continuing to library preparation with only chromatin that was non-

specifically bound to the beads. This is currently not feasible in chipmentation, as the extremely low cell number per sample leaves you with essentially no chromatin after "immunoprecipitating" ~5% without antibody. Instead, it is common for researchers to do library preparation directly on 5% of the sample's sonicated chromatin. The problem with this, however, is that the resulting input track represents signal from total chromatin, including everything that would normally be washed away. For example, I sequenced inputs for all my chipmentation reactions but, as there were at least as many strong signals in the inputs that weren't present in the IPs, with just the occasional common background peak, input-subtracted sample tracks were far noisier than the sample tracks alone (results not shown). There is an unmet need in the field to develop a suitable chipmentation input protocol, which would be of particular use in chipmentations with low SNR or un-verified genome binding profiles, as are both the case with the KLF1 chipmentation.

4.4 – Chapter 4 conclusions

I herein present a viable mouse model that has potential to overcome the existing limitations of anti-KLF1 antibodies, a pressing unmet need in the KLF1 field (**Bieker and Philipsen, 2024**). In conclusion, I find the molecular and cellular phenotype of the KLF1$^{StrepII\text{-}FKBP}$ mouse appears almost indistinguishable from WT. This suggests that the StrepII-FKBP tags do not significantly interfere with KLF1 function. It remains possible that the FKBP tag has a slight negative effect on KLF1 turnover, given the previously recorded effects of FKBP tags on protein stability in other FKBP-tagged models. This might explain the potential vulnerability of the mice to KLF1$^{StrepII\text{-}FKBP}$ between E12.5-birth. As such, an additional cycloheximide pulse-chase style experiment would help to discern this. Importantly, I have shown that the StrepII tag works for Western blot and, tentatively, for chipmentation of KLF1. Further optimisations of the KLF1$^{StrepII\text{-}FKBP}$ chipmentation protocol will unlock the full potential of this model to study KLF1 genome-wide binding patterns. The low cell-input requirement of chipmentation will make this a powerful system in which to study KLF1 binding dynamics in rare

biological or experimental samples, such as stratified erythroblasts. Overall, I have met my first aim to characterise the KLF1$^{StrepII\text{-}FKBP}$ mouse and have laid important technical groundwork towards my next two aims for understanding the genome-wide binding of KLF1 in bulk and rare mFL cell populations. I find the KLF1$^{StrepII\text{-}FKBP}$ mouse to be a suitable model to take forward to address several outstanding and novel questions regarding KLF1 gene targets and functions.

4.5 – Chapter 4 acknowledgements

For this chapter, big thanks go to Rob Beagrie who, with Daniel Biggs and Ben Davies, designed and generated the KLF1$^{StrepII\text{-}FKBP}$ mouse model and assisted with the design and review of this project. Thank you to Jackie Sloane-Stanley for taking care of the animal husbandry work. Thank you to Kevin Clark at the FACS facility for helping me to set up the FACS sorts and for being constantly, infectiously, cheerful and patient (no matter how many times I broke/forgot how to use the FACS machine). Thank you to Gabby Sherwood, for her work on KLF1$^{StrepII\text{-}FKBP}$ characterisation in her Master's thesis. Massive credit and thanks go to Lucy Cornell, for helping me to write, use, and interpret the code and results for the comparative ChIP analyses. Thank you to Kinam Gupta for helping me to brainstorm the many rounds of chipmentation optimisations, for generously helping me out whenever I took on too many experiments at once, and for every "quick chat" on our way out the building that would always brighten my day. A final thank you to my live-in structural biology consultant, Emily Rowland, for the one-to-one crash course in AlphaFold.

Chapter 5 – The effect of acute KLF1 degradation on ζ-globin expression

5.1 – Introduction

KLF1 has several well-characterised roles in regulating expression of the globin genes. Yet the specific role of KLF1 in the regulation of ζ-globin, the embryonic α-like globin gene, is relatively understudied. In this chapter, I use the KLF1$^{StrepII\text{-}FKBP}$ mouse model to interrogate the role of KLF1 in ζ-globin gene regulation in definitive erythropoiesis.

5.1.1 – ζ-globin

ζ-globin is the embryonic α-like globin gene (*Hba-x* in mouse, *HBZ* in human). It is ordinarily expressed in only primitive erythroid cells, where it is partly activated by the α-globin enhancer-like elements and marked by H3K4me3, H3K4me1, H3K27ac, and ATAC-accessibility (**Fig. 5.1**; **King *et al.*, 2021**). Conversely, in definitive erythropoiesis, the ~10 kb sequence encompassing *Hba-x* does not contact the active enhancer elements, is devoid of these active histone marks, is considered inaccessible by ATAC-seq, and is transcriptionally silenced (**King *et al.*, 2021**). How exactly ζ-globin is so differentially regulated between primitive and definitive erythropoiesis is poorly understood.

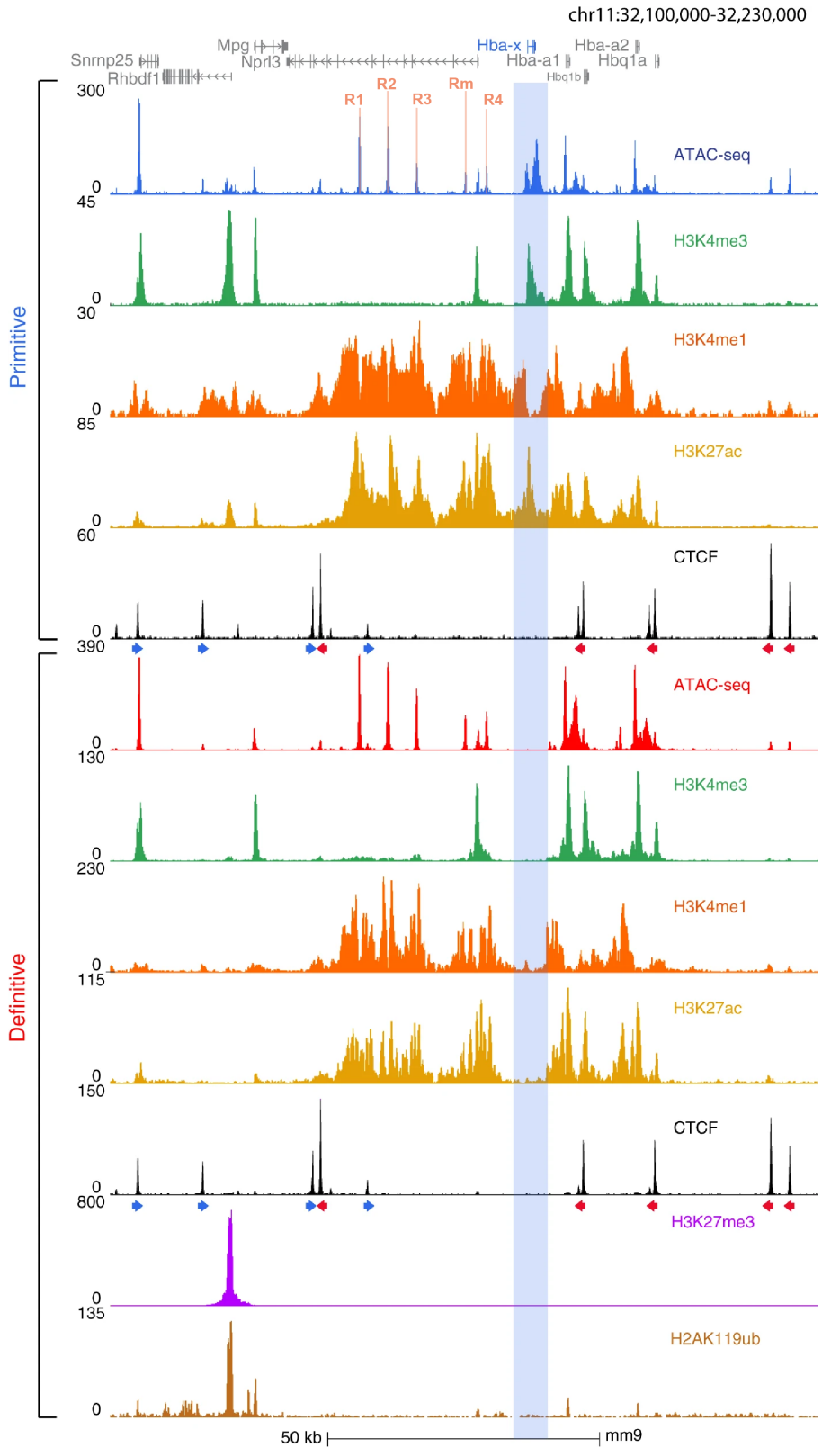

Fig. 5.1 – Hba-x *is devoid of all measured chromatin marks of accessibility in definitive erythroid cells.* Figure adapted with permission from **King *et al.*, 2021**. ATAC-seq and ChIP-seq tracks for H3K4me3, H3K4me1, H3K27ac, and CTCF in primitive (upper) and definitive (lower) mouse erythroid cells. The *Hba*-x gene is highlighted in blue. The α-globin enhancers (superenhancer) are annotated in orange. The orientation

of CTCF sites is show as blue (forward) or red (reverse) arrows below CTCF tracks. H3K27me3 and H2AK119ub are histone markers of polycomb repression.

An unmet need in the field is the thorough characterisation of the factors involved, and the extent of their involvement, in the regulation of ζ-globin gene expression. Developing a better understanding of ζ-globin silencing in definitive erythropoiesis could not only reveal new insights into the general mechanisms of developmental gene switching events, but could identify new therapeutic targets for the reactivation of ζ-globin in α-thalassemia and Hb Bart's hydrops fetalis syndrome (BHFS).

5.1.2 – The therapeutic potential of ζ-globin

There is a clinical precedent for the upregulation of a patient's own developmentally silenced genes in adult tissues to rescue a disease phenotype. Most relevant are the on-going phase III clinical trials to upregulate fetal γ-globin for the treatment of β-thalassemia and sickle cell disease. Both pharmacologic (ClinicalTrials.gov Identifier: NCT04586985) and autologous stem cell gene editing strategies (NCT03655678; NCT03745287; NCT03282656) show promising results (**Esrick *et al.*, 2021; Locatelli *et al.* 2024; Frangoul *et al.*, 2024**). Similarly, there is therapeutic potential to de-repress ζ-globin in patients with α-thalassemia.

In pre-clinical mouse models, ectopic expression of human ζ-globin is able to rescue lethal α-thalassemia phenotypes, suggesting a beneficial therapeutic function of Hb Portland II ($\zeta_2\beta_2$) in adult blood (**Russell and Liebhaber, 1998; He and Russell, 2004).** One concern when translating this to patients is the higher O_2 binding affinity of $\zeta_2\beta_2$ than adult haemoglobin ($\alpha_2\beta_2$) (**He and Russell, 2004**). However, there are reports of individuals with naturally occurring globin chain variants that are associated with even higher O_2 binding affinities than $\zeta_2\beta_2$ and who are asymptomatic (**Wajcman and Galactéros, 1996**). Moreover, ζ-globin is persistently expressed in some healthy carriers or HbH patients harbouring large α-globin deletion alleles, such as the $--^{SEA}$ allele (**Chui *et al.*, 1986; Tang *et al.*, 1992; Lafferty *et al.*, 2000**). Indeed, unusually high ζ-globin expression is

thought to be a key contributor to the rare survival of some patients with BHFS (**Songdej, Babbs and Higgs, 2017**). It is hypothesised that these survivors co-inherit positive modifiers of *HBZ*. In one such remarkable case, a BHFS patient survived their first year of life before diagnosis (**D. Songdej, DPhil thesis, 2015**). With no functional α-globin genes, they survived entirely on their upregulation of ζ-globin and occasional blood transfusion for jaundice. Together, these observations suggest that de-repression of ζ-globin could be an effective therapeutic strategy to ameliorate symptoms in patients with severe α-thalassemia. Further investigation of endogenous ζ-globin regulation and feasible mechanisms of de-repression will be paramount to exploit its full therapeutic potential.

5.1.3 – Known ζ-globin gene regulation factors

In addition to potential therapeutic benefits, a better understanding of ζ-globin regulation could improve our understanding of the general principles of gene silencing and developmental gene switching events. Indeed, our extensive characterisation of the α-globin locus has previously revealed several gene regulatory mechanisms that are applicable genome-wide. In comparison to the β-like and adult α-globin genes, however, ζ-globin gene regulation is far less well understood.

Some clues as to the cis-regulatory elements governing ζ-globin expression come from transgenic assays. Full-length *Hba-x* gene constructs, driven by the *Hba-x* promoter and β-globin LCR, undergo normal ζ-globin switching (**Spangler, Andrews and Rubin, 1990**). Likewise, placement of an α-globin promoter and gene sequence into the position of ζ-globin has a normal α-globin expression profile (**S. Liu., unpublished**). This suggests that the key cis-regulatory sequences lie within the full-length ζ-globin gene sequence. This was further refined by a series of 5' truncations of the *Hba-x* promoter which concluded that the proximal 128 bp of the promoter sequence is sufficient for normal developmental regulation (**Sabath *et al.*, 1993**). Although there may be additional essential sequences, as a parallel series of deletion constructs implicated the

promoter, intron 1, and 3' UTR for normal ζ-globin regulation (**Vyas *et al.*, 1992; Liebhaber *et al.*, 1996**).

Likewise, some trans-acting regulatory factors have been identified as negative regulators of ζ-globin: namely KLF1, BCL11A, and LRF. No positive regulatory factors *specific* to ζ-globin expression have yet been identified, beyond the normal complement of core transcription machinery and erythroid TFs (**D. Songdej, DPhil thesis, 2015; Goode *et al.*, 2016**).

5.1.4 – The roles of KLF1, BCL11A, and LRF in ζ-globin regulation

The primary way KLF1 is known to regulate ζ-globin is by direct upregulation of BCL11A and LRF (summarised in **Fig. 5.2**).

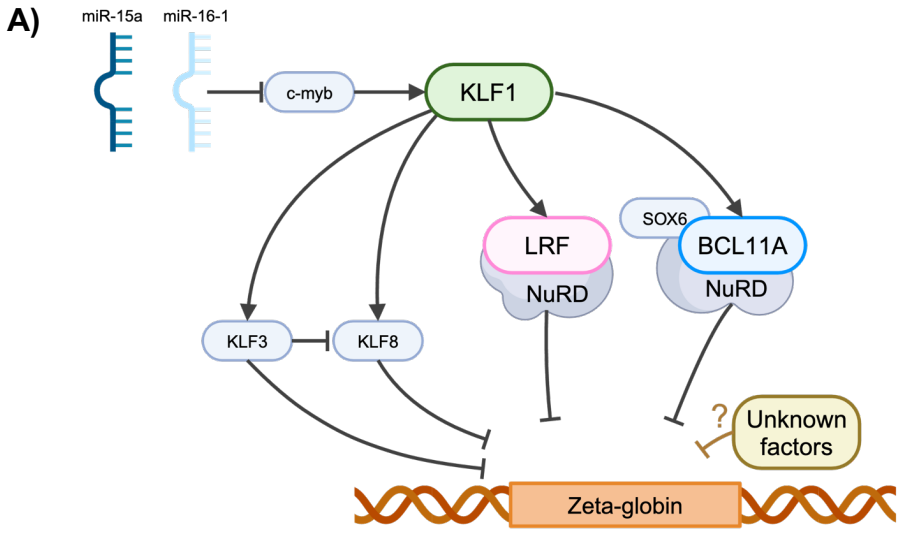

Fig. 5.2 – *A summary of the known ζ-globin regulatory pathways*. There are several known trans-acting regulatory factors involved in ζ-globin gene regulation. KLF1 indirectly silences ζ-globin by promoting *Lrf* (*Zbtb7a*) and *Bcl11A* expression. LRF and BCL11A are the two most well-characterised direct ζ-globin regulators. They bind the ζ-globin promoter in human and mouse to recruit NuRD histone deacetylase complex and confer transcriptional silencing. Together they account for ~50% ζ-globin silencing. All the factors presented here have been experimentally or clinically linked to ζ-globin upregulation. This is non-exhaustive; there are more known regulators of *Klf1*, *Bcl11a,* and *Lrf*. Pointed arrows indicate positive regulation. Flat-headed arrows indicate negative regulation. Yellow “unknown factors” indicates unknown regulatory pathways that might confer ~50% ζ-globin silencing in definitive erythroid cells. These may or may not be linked to KLF1.

BCL11A and LRF are negative transcriptional regulators of ζ-globin. Their binding is detected at the *HBZ* promoter by ChIP-seq in definitive-like erythroid cell lines (**Martyn *et al.*, 2018**). Accordingly, there are BCL11A and LRF binding motifs in the ζ-globin promoter and double knockout of these motifs, or the TFs themselves, partially derepresses ζ-globin expression in human and mouse cells (**King *et al.*, 2021; Liu *et al.*, manuscript in preparation**). ζ-globin silencing is thought to be mediated at least in part by histone deacetylation. BCL11A and LRF can recruit the NuRD histone deacetylase (HDAC) complex and treatment of definitive erythroblasts with HDAC inhibitors derepressed ζ-globin to a minor extent (**King *et al.*, 2021; Bradner *et al.*, 2010**). Double knockout of LRF and BCL11A or their ζ-globin promoter binding sites only reactivates ζ-globin to ~20% of total α-like mRNA (**King *et al.*, 2021**; **Liu *et al.*, manuscript in preparation**). However, in E10.5 primitive erythroblasts, ~40% of all α-like transcripts originate from *Hba-x* (**Kingsley *et al.*, 2006; King *et al.*, 2021**). As such, there appears to be additional, yet-unknown mechanisms governing >50% of ζ-globin repression.

Whether KLF1 mediates its control over ζ-globin exclusively via LRF and BCL11A is unknown, but there are several clinical and experimental examples of KLF1 dysfunction leading to upregulated ζ-globin.

5.1.5 – KLF1 mutations in HPEH

Thanks to the rising accessibility of genome sequencing, a growing number of KLF1 disorders are being characterised and diagnosed – several of which present with elevated fetal and embryonic globins (**Perkins *et al.*, 2016**). Key clinical cases and mouse models with evidence of ζ-globin de-repression are summarised in **Table 5.1**. Unsurprisingly, knockout of human KLF1 is ordinarily lethal and there is only one recorded KLF1-null survivor (**Magor *et al.*, 2015**). This patient and others with heterozygous, or compound heterozygous, mutations in KLF1 all express elevated ζ-globin (hereditary persistence of embryonic haemoglobin; HPEH). Similarly, mouse models of these human disorders can derepress ζ-globin (**Table 5.1**). A major difficulty when evaluating this, however, is that ζ-globin expression is typically far less often or less thoroughly reported than its β-like and adult α-globin orthologues. The extent and

frequency of ζ-globin de-repression may therefore be underestimated. For example, despite at least three independent mouse lines with extensive β-like globin analyses, there is extremely little published on whether ζ-globin is dysregulated in the KLF1$^{-/-}$ mice.

The lack of *Hba-x* expression data is not explained by the demise of KLF1$^{-/-}$ embryos around the time of the primitive-definitive switch, as these mice die from severe β-thalassemia and deranged erythropoiesis, rather than a total lack of definitive erythropoiesis. Indeed, E12-E15 KLF1$^{-/-}$ definitive erythroid mFL material can be obtained and such expression analyses have been performed (**Tallack *et al.*, 2012; Hung *et al.*, 2021**). Rather, whether ζ-globin is ever included or analysed in the first place, let alone dysregulated, often goes unreported (**Table 5.1**). Acute KLF1 knockdown, e.g. by shRNA, has been attempted before, but failed to control for differentiation effects or report ζ-globin expression (**Borg *et al.*, 2010**). Likewise, all these experiments use either bulk mFL cells, which fail to control for potential differentiation differences, or use wrongly Ter119-stratified populations (**Table 5.1**). For example, **Hodge *et al.*, 2006**, claimed KLF1 *induction* upregulated *Hba-x* by ~6-fold. But their experiments appear to have been performed in bulk cell line cultures. If KLF1 is important for differentiation progression, and *Hba-x* is lowly but increasingly expressed in differentiating erythroid cells (**Appendix Fig. A5.1**), then differentiation advancement in the KLF1-induced cultures may well account for this observation. In line with this, they also observed upregulation of *Hbb-bh1*, *Hbb-y* – two well validated, KLF1-*repressed* genes that also are increasingly expressed throughout erythroid differentiation (**Appendix Fig. A5.1**).

KLF1 is clearly implicated in the definitive ζ-globin repression pathway. However, the exact extent and mechanism of this, particularly in the mouse, is unknown. The KLF1$^{StrepII\text{-}FKBP}$ mouse presents a valuable new model in which this might be addressed.

Table 5.1 – *Examples of known ζ-globin upregulation in human and mouse, in primary definitive erythroid samples*. Hb Portland I = $\zeta_2\gamma_2$. **Perkins *et al.*, 2016**, review contains further examples of KLF1 patients with upregulated embryonic and fetal globins. All examples are in primary cells unless otherwise stated. * From supplementary data: https://genome.cshlp.org/content/22/12/2385/suppl/DC1 and https://www.ebi.ac.uk/gxa/experiments/E-GEOD-33979. ** α and beta globin depletion using Mouse/Rat GLOBINclear Kit (Ambion). Unclear if *Hba-x* depleted. *Hbb-bh1* and *Hbb-y* were sequenced and upregulated.

Mutation (disease)	Species	↑ ζ-globin	Reference	Notes
KLF1$^{-/-}$	Human	**4.5-fold elevated *HBZ* by RNA-seq**. *Suspected* 24% Hb Portland at 6-months by HPLC.	**Magor *et al.*, 2015**	Only recorded human survivor of KLF1$^{-/-}$. RNA-seq data in Supp. Table 4.
Compound heterozygous KLF1 mutations (NSHA)	Human	**4 – 17.6% $\zeta_2\gamma_2$**	**Viprakasit *et al.*, 2014**	Patients with highest (>16%) Hb Portland were also SEA carriers.
KLF1^{W313L}	Human	Significant **ζ-globin protein.**	**Ireland and Rees, 2021**	ζ-globin protein upregulation confirmed in further work (A. Ejaz; unpublished)
KLF1^{E325K} (CDA-VI)	Human	**Very faint ζ-globin band** in haemolysate. 0.2% haemolysate ζ-globin. Heterologous ζ-globin expression by immunofluorescence.	**Tang *et al.*, 1993**	
		2.9% $\zeta_2\gamma_2$ by IEF.	**Arnaud *et al.*, 2010**	
KLF1^{K288X}	Human	*(*ζ-globin not reported)	**Borg *et al.*, 2010**	RNA microarray in bulk cultured primary HEPs.
Hba$^{SEA/+}$	Human	**0.13 – 0.42% ζ-globin** chains by radioimmuno assay.	**Chui *et al.*, 1986**	
		ζ-globin detected by ELISA.	**Lafferty *et al.*, 2000**	
Hba$^{SEA/SEA}$ (BHFS)	Human	Median **17% $\zeta_2\gamma_2$** at birth.	**Songdej *et al.*, 2017**	
Trisomy 13 (Patau Syndrome)	Human	Increased embryonic haemoglobins	**Huehns *et al.*, 1964**	Suspected to act via miR-15a, miR-16-1, MYB, and KLF1.
BCL11A$^{-/-}$	Human	**1% *HBZ*** of total α-like mRNA.	**King *et al.*, 2021**	HUDEP-2

		56% of cells express HBZ protein, by flow cytometry		
LRF$^{-/-}$	Human	**3% *HBZ*** of total α-like mRNA. 87% of cells express HBZ protein, by flow cytometry.	**King *et al.*, 2021**	HUDEP-2
BCL11A$^{-/-}$ LRF$^{-/-}$ *DKO*	Human	**~15% *HBZ*** of total α-like mRNA. 87% of cells express HBZ protein, by flow cytometry.	**King *et al.*, 2021**	HUDEP-2
Klf1$^{-/-}$	Mouse	*(Hba-x* not reported)	**Perkins *et al.*, 1995**	Normal ζ-globin RNA in +/- and -/- E11 yolk sac (primitive).
		(Hba-x not reported)	**Nuez *et al.*, 1995**	Not reported at all.
		No ectopic *Hba-x* RNA detected.	**Perkins *et al.*, 1996**	Rnase protection assay. E15 mFL: 1 hom, 4 hets.
		Hba-x upregulated 5.9-fold when KLF1-ER induced.	**Hodge *et al.*, 2006**	RNA microarray in KLF1$^{-/-}$, KLF1-ER inducible definitive cell line.
		Reduced to ~30% *Hba-x* mRNA in bulk KLF1$^{-/-}$ vs WT in primitive blood.	**Basu *et al.*, 2007**	E10.5 yolk sac (primitive).
		(Hba-x not reported; confounded)	**Pilon *et al.*, 2008**	RNA microarray in CD71/Ter119-sorted KLF1$^{-/-}$ mice = results confounded.
		(Hba-x not reported)	**Pilon *et al.*, 2011**	G1E vs G1E-ER4 (GATA1-inducible) RNA-seq.
		(Hba-x not reported)	**Siatecka *et al.*, 2011**	Review.
		Hba-x not significantly changed.	**Tallack *et al.*, 2012***	First RNA-seq in KLF1$^{-/-}$ mice. Depleted** for α and β RNA.
		No change *Hba-x* RNA in KLF1$^{+/-}$ by qPCR.	**Esteghamat *et al.*, 2013**	Note: heterozygote.
		(Hba-x not reported)	**Hung *et al.*, 2021**	KLF1$^{-/-}$ vs WT, RNA microarray. (neither *Hba-x* nor any β-like globins in supp.

				tables of up/down-regulated genes, only *Hba-a1/2*)
KLF1$^{H350R/-}$ (NSHA)	Mouse	"No elevated embryonic globin".	**Huang *et al.*, 2024**	AKA KLF1$^{mommeD45/-}$
Klf1$^{E339D/+}$	Mouse	**>16-fold elevated *Hba-x*** by RNA-seq	**Gillinder *et al.*, 2017**	AKA Klf1^{nan}. E14.5 Fetal liver, KLF1$^{Nan/+}$ vs WT. From reanalysis in **Nébor *et al.*, 2018**, supplementary data.
		Hba-x $\log_2$ fold change by RNA-seq: 4.9 ProE (**28.9x**) 7.9 Baso (**239x**) 9.07 Poly (**537x**) 4.7 Ortho (**26x**)	**Nébor *et al.*, 2018**	Spleen, KLF1$^{Nan/+}$ vs anaemic WT. Supplementary data.
Bcl11a$^{-/-}$	Mouse	**Significant *Hba-x* de-repression** by RNA microarray (>14-fold).	**Xu *et al.*, 2011**	CD71^{+}/Ter119^{+} bone marrow. Erythroid-specific BCL11A^{CKO}.
		Several-thousand-fold elevated *Hba-x*, single cell RT-qPCR.	**King *et al.*, 2021**	
		5-fold to >100-fold increase *Hba-x* RNA.	**Esteghamat *et al.*, 2013**	*Bcl11a* CKO
Lrf$^{-/-}$	Mouse	**~2000-fold to 4000-fold elevated *Hba-x*** by RNA-seq. **Significant ζ-globin protein** by IEF.	**Masuda *et al.*, 2016**	Supp. Fig S2
		Several-thousand-fold elevated *Hba-x*, single cell RT-qPCR.	**King *et al.*, 2021**	
Hba-x promoter BCL11A motif mutation	Mouse	**3% *Hba-x*** of total α-like mRNA.	(Liu *et al.*, manuscript in preparation)	G1E-ER4 cells.
Hba-x promoter LRF motif mutation	Mouse	**15% *Hba-x*** of total α-like mRNA.	(Liu *et al.*, manuscript in preparation)	G1E-ER4 cells.
Hba-x promoter BCL11A & LRF motif mutation	Mouse	**22% *Hba-x*** of total α-like mRNA.	(Liu *et al.*, manuscript in preparation)	G1E-ER4 cells.

5.1.6 – Expected effects of KLF1 perturbation on α-like globins

KLF1 is a well-characterised regulator of the β-like globin genes. In mouse, KLF1 directly promotes adult β-globin expression and indirectly represses the embryonic β-like globins, *Hbb-bh1* and *Hbb-y*, via BCL11A and LRF. This is extensively reviewed elsewhere (**Tallack and Perkins, 2013; Santos *et al.*, 2024; Bieker and Philips, 2024**).

In contrast to the β-like globin genes, there is relatively little published research on the role of KLF1 in α-like globin gene expression. In the literature, KLF1 knockout mice show profound β-thalassaemia while *Hba-a1/2* RNA is relatively unperturbed, as measured by semi-quantitative RNAse protection assays (**Perkins *et al.,* 1995; Nuez *et al.*, 1995; Coghill *et al.*, 2001**). This is despite presence of KLF1 motifs at the *Hba-a1/2* promoters, where KLF1 binding is detected by ChIP (**Chapter 4; Vernimmen *et al.*, 2007; Tallack *et al.*, 2010; Huang *et al.*, 2024**). To the best of my knowledge, since those first experiments in KLF1$^{-/-}$ mouse, no one has reported a more accurate quantification of *Hba-a1/2* RNA specifically (**Drissen *et al.*, 2005; Hodge *et al.*, 2006; Nilson *et al.*, 2006; Pilon *et al.*, 2008; Tallack *et al.*, 2009, 2010, 2012; Hung *et al.*, 2021; Huang *et al.*, 2024**).

Meanwhile, KLF1 is thought to indirectly repress *Hba-x* by directly promoting transcription of the repressive transcription factors, *Bcl11a* and *Lrf*. To determine the expected magnitude of *Hba-x* de-repression after KLF1$^{StrepII\text{-}FKBP}$ depletion, we ideally would look to *Hba-x* expression in a KLF1$^{-/-}$ mouse model. Unfortunately, there also appears to be no published quantification of *Hba-x* RNA expression in definitive KLF1$^{-/-}$ erythroblasts (**Table 5.1**). Instead, we might look to KLF1 mutant mouse models and human KLF1 patients. Klf1 compound heterozygous mice express variable *Hba-x*, ranging 0 to 500-fold higher transcript abundance than controls, with ζ-globin protein levels typically not reported (**Table 5.1**). The only known KLF1$^{-/-}$ human patient was recorded to have more a modest 4.5-fold increase in *HBZ* mRNA expression and suspected <24% Hb Portland (**Magor *et al.,* 2015**). Human compound heterozygous KLF1 patients typically have of ~0.2 – 17% ζ-globin protein (**Table 5.1**). Interestingly, mice with direct

Bcl11a or *Lrf* perturbations have more evidence for higher levels of *Hba-x* de-repression than KLF1$^{-/-}$ alone (**Table 5.1**). This may be because KLF1$^{-/-}$ mice still express some LRF (**Norton *et al.*, 2017**) and BCL11A (**Zhou *et al.*, 2010**).

The primary aim of this chapter is to determine whether acute KLF1$^{StrepII\text{-}FKBP}$ depletion phenocopies the ζ-globin de-repression phenotype observed in KLF1-perturbed patient and mouse models. This serves the ultimate purpose of determining whether this mouse model is suitable to use to investigate the mechanism and extent of KLF1 regulation of ζ-globin. From the above literature review, we might reasonably hypothesise KLF1 knockdown to have no effect on *Hba-a1/2* expression but to at least modestly derepress *Hba-x*.

5.1.7 – Aims for Chapter 5

To address this ζ-globin knowledge gap and investigate whether *Hba-x* is upregulated in KLF1$^{-/-}$ mouse, as it is in human, I use the KLF1$^{StrepII\text{-}FKBP}$ mouse to acutely deplete KLF1 and assay ζ-globin expression in differentiation-matched definitive erythroid cells.

Aims:

1. To characterise the KLF1$^{StrepII\text{-}FKBP}$ mouse model for effective KLF1 protein depletion and its effects.
2. To determine whether acute KLF1$^{StrepII\text{-}FKBP}$ depletion derepresses *Hba-x* RNA and protein expression in definitive mouse fetal liver erythroblasts.

5.2 – Results

5.2.1 – Acute depletion of KLF1

First, I checked that KLF1$^{StrepII\text{-}FKBP}$ could be quickly and completely degraded in mFL cells with a time-course of dTAG-13 treatment and anti-StrepII Western blotting (**Fig. 5.3, Appendix Fig. A5.2,** N = 2). This revealed that ~ 94% of KLF1 was depleted within 30 minutes of dTAG-13 treatment, rising to >99% depletion from 24 hours onwards. Subcellular fractionation of mFL cells treated for 24 hours with dTAG-13 or DMSO confirmed depletion to sub-detectable levels of KLF1 from both the nuclear and cytoplasmic fractions within 24 hours (**Fig. 5.4**).

A concern with any acute degron depletion of a DNA or chromatin-binding protein is persistence of the POI on the chromatin at sub-detectable levels by Western blot. As this could have important functional consequences when interpreting subsequent gene expression analyses, I attempted a chipmentation for KLF1$^{StrepII\text{-}FKBP}$ after 24-hours treatment with dTAG-13 or DMSO, to confirm KLF1 knockdown at the chromatin level. Surprisingly, I was unable to repeat my anti-StrepII chipmentation from Chapter 4 in the DMSO treatment sample (**Appendix Fig. A5.3**, N = 3), reflecting the continued difficulty of KLF1 IP. A lack of reproducibility of the anti-StrepII KLF1 chipmentation method is concerning. Perhaps enriching for the S2 cell fraction before chipmentation would improve SNR and thus the reproducibility of this ChIP.

Nonetheless, I have demonstrated near complete depletion of KLF1 in the nuclear fraction by Western blot. If low levels persist in the chromatin fractions, it remains highly likely that the dramatically depleted KLF1 nuclear concentration will affect the overall gene expression output of KLF1-dependent genes. This system is therefore likely suitable for studying expression effects in primary KLF1 target genes.

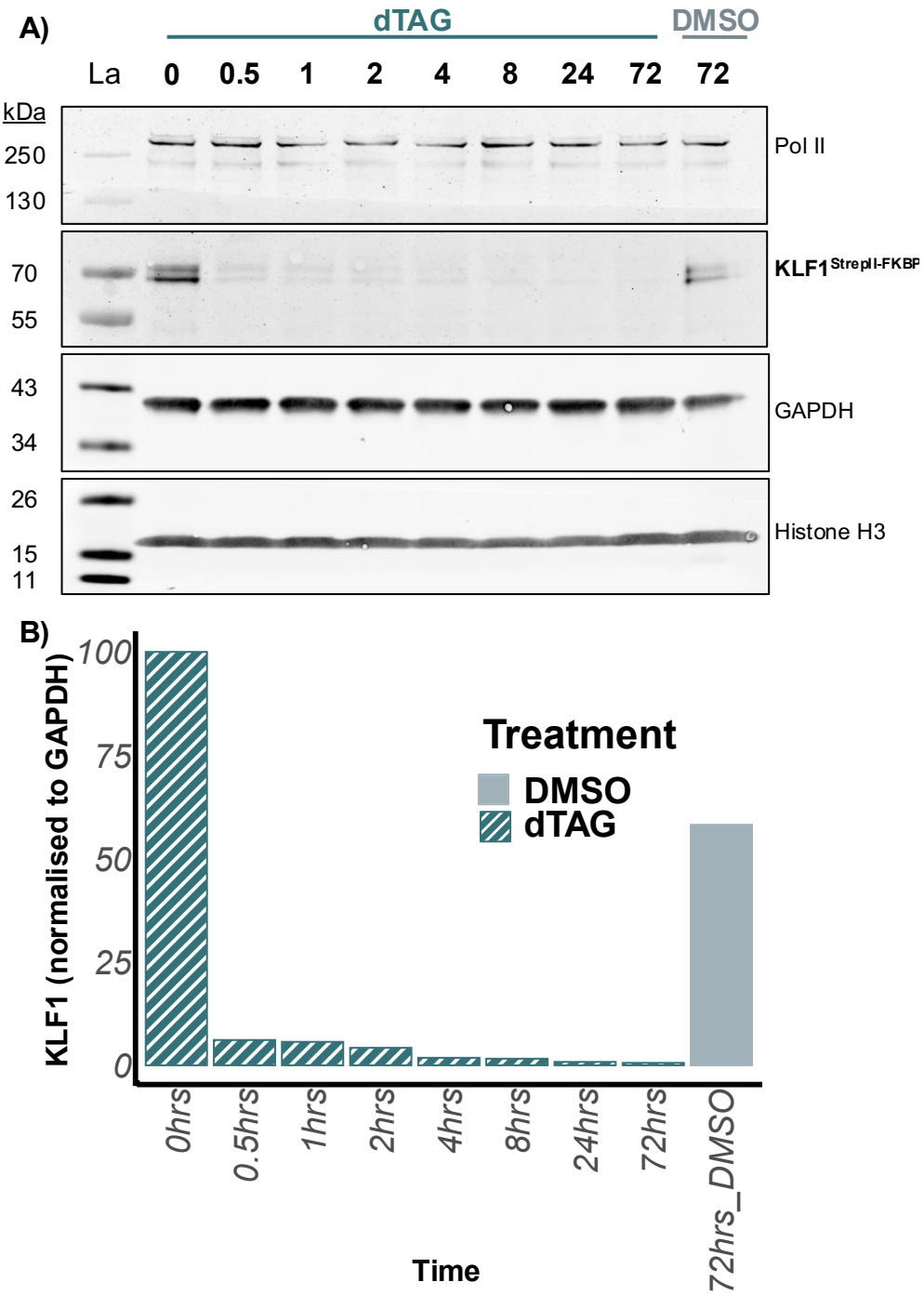

Fig 5.3 – *KLF1 is quickly, acutely, and persistently depleted by dTAG-13 treatment*. **A)** Western blot timecourse of KLF1$^{StrepII\text{-}FKBP}$ mFL cells cultured for 0-72 hours in expansion media with 0.2 μM dTAG-13 or an equivalent volume DMSO. 3 loading controls were tested: Pol II polymerase, GAPDH, and histone H3. KLF1$^{StrepII\text{-}FKBP}$ was detected with an anti-StrepII antibody (Qiagen, 34850). **B)** Quantification of KLF1$^{StrepII\text{-}FKBP}$ signal, normalised to GAPDH and to 0hrs. Fluorescent western blot is a semi-quantitative technique and should be interpreted with caution. A second biological replicate is in **Appendix A5.2**.

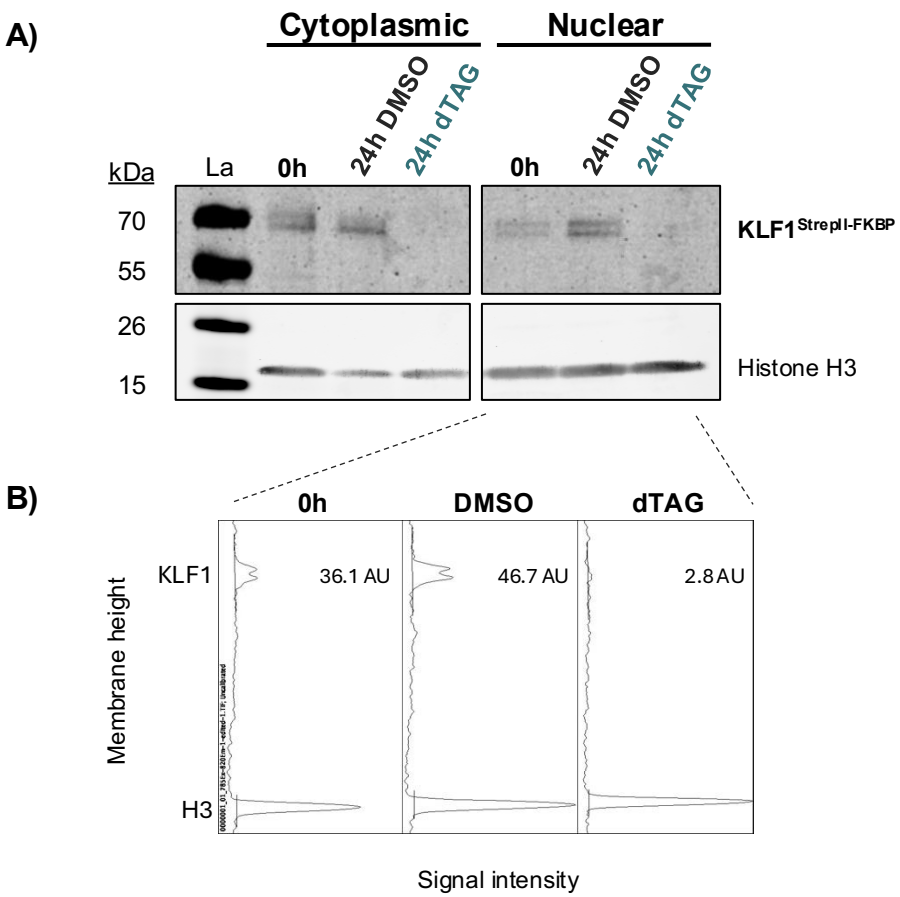

Fig. 5.4 – *KLF1 is degraded from nuclear and cytoplasmic fractions to undetectable levels by 24 hours dTAG-13 treatment.* A) Subcellular fractionation Western blot of KLF1$^{StrepII\text{-}FKBP}$ mFL cells cultured for 0 or 24 hours with 0.2 μM dTAG-13 or an equivalent volume DMSO. KLF1$^{StrepII\text{-}FKBP}$ was detected with an anti-StrepII antibody (Qiagen, 34850). Non-equivalent volumes of protein lysates were loaded per lane. **B)** Signal intensity histogram of the nuclear fraction lanes from Western blot A). KLF1$^{StrepII\text{-}FKBP}$ signal (area under curve, AUC) was normalised to histone H3 loading control AUC to give relative signal intensity of the KLF1 bands (in arbitrary units; "AU"). N = 1.

5.2.2 – The effect of acute KLF1 knockdown on erythroid differentiation

Having established the potential for thorough knockdown of KLF1$^{StrepII\text{-}FKBP}$, I next assessed the effects of KLF1 depletion on mFL differentiation, to ensure that I compare like-for-like cell populations between the dTAG-13-treated knockdown and DMSO control samples in subsequent gene expression assays.

KLF1 is known to play complex roles in regulating erythroid differentiation, but there are two main stages at which these manifest. Briefly, first in early terminal erythropoiesis KLF1 regulates cell cycle genes, such as *E2f2*, which control the rapid cell divisions that govern the S0-S1 transition from erythroblast proliferation to terminal erythroid differentiation (**Pilon *et al.*, 2008; Tallack *et al.*, 2009**). Second, in late-stage erythroblasts, KLF1 promotes expression of p18 (*Cdkn2c*) and p27 (*Cdkn1b*) that mediate the cell cycle exit required for enucleation and reticulocyte formation. As such, KLF1$^{-/-}$ erythroblasts fail to progress past the orthochromatic stage (~S4/5) to reticulocytes (**Gnanapragasam *et al.*, 2016; Gnanapragasam and Bieker, 2017**). Following KLF1 knockdown we therefore might expect an increase in the proliferating early S0 erythroblasts (~BFU-E/CFU-E) and fewer terminally differentiating, more highly haemoglobinised erythroblasts.

I periodically measured mFL cultures for growth rate and overall haemoglobinisation. dTAG-13 treatment resulted in increased cell growth rate and the resulting cell pellets were paler than DMSO controls (**Fig. 5.5**). This suggests a possibly larger population of early-stage proliferative erythroblasts with KLF1 depletion, although paler cell pellets could also be due to defective haemoglobinisation, another KLF1 regulated process.

An obvious next step would be to perform CD71/Ter119 flow cytometry to assess erythroid differentiation. However, due to a quirk of KLF1 biology, this is not so straightforward.

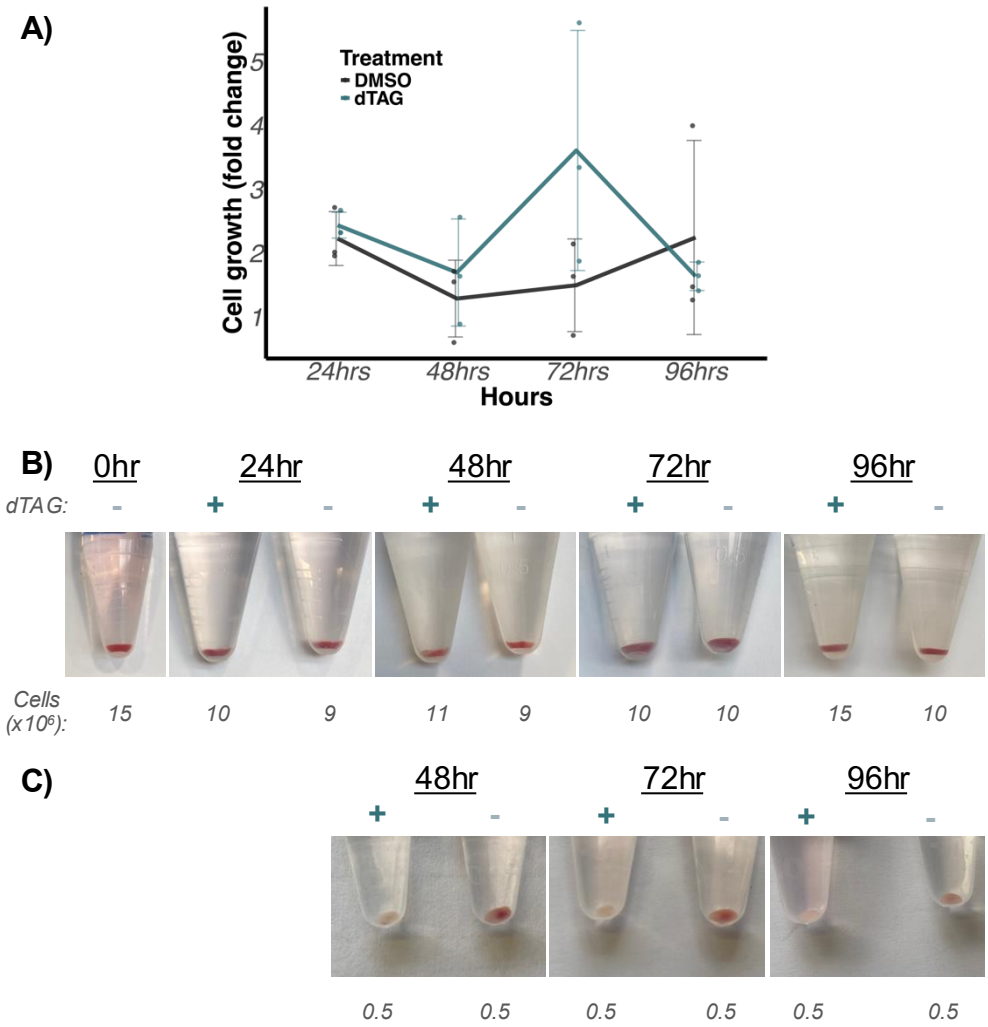

Fig. 5.5 – ***Cell pellets are paler and growth rate is faster in mFL cells with acute KLF1 depletion.*** KLF1$^{StrepII\text{-}FKBP}$ mFL cells were cultured up to 96 hours with 0.2 μM dTAG-13 or an equivalent volume DMSO. **A)** Line plot of mean cell growth rate, expressed as the fold change in total cell count from the previous timepoint. N = 3. Error bars = ± one standard deviation. **B)** Photos of representative example cell pellets from a typical differentiation. Cells treated with dTAG-13 were slightly paler than DMSO controls for most biological replicates. **C)** In 2/9 differentiations (from 7 biological replicates), cells pellet colour differences were more obvious.

5.2.3 – The effect of acute KLF1 knockdown on Ter119 and erythroid differentiation

A particular difficulty when studying KLF1 in erythropoiesis is the loss of Ter119 expression in KLF1 knockout cells, with obvious consequence on CD71/Ter119-based immunophenotyping (**Siatecka and Bieker, 2011**). I hypothesised however that, given that untreated KLF1$^{StrepII\text{-}FKBP}$ cells express Ter119 normally (**Chapter 4**, **Fig. 4.9**), there would be a period of time between acute KLF1 knockdown and Ter119 turnover in which the cell surface marker would still be present and sufficient to stratify the differentiating erythroid populations. To assess this, I performed a 72-hour treatment with dTAG-13 or DMSO control and periodically assessed Ter119 expression and cell differentiation.

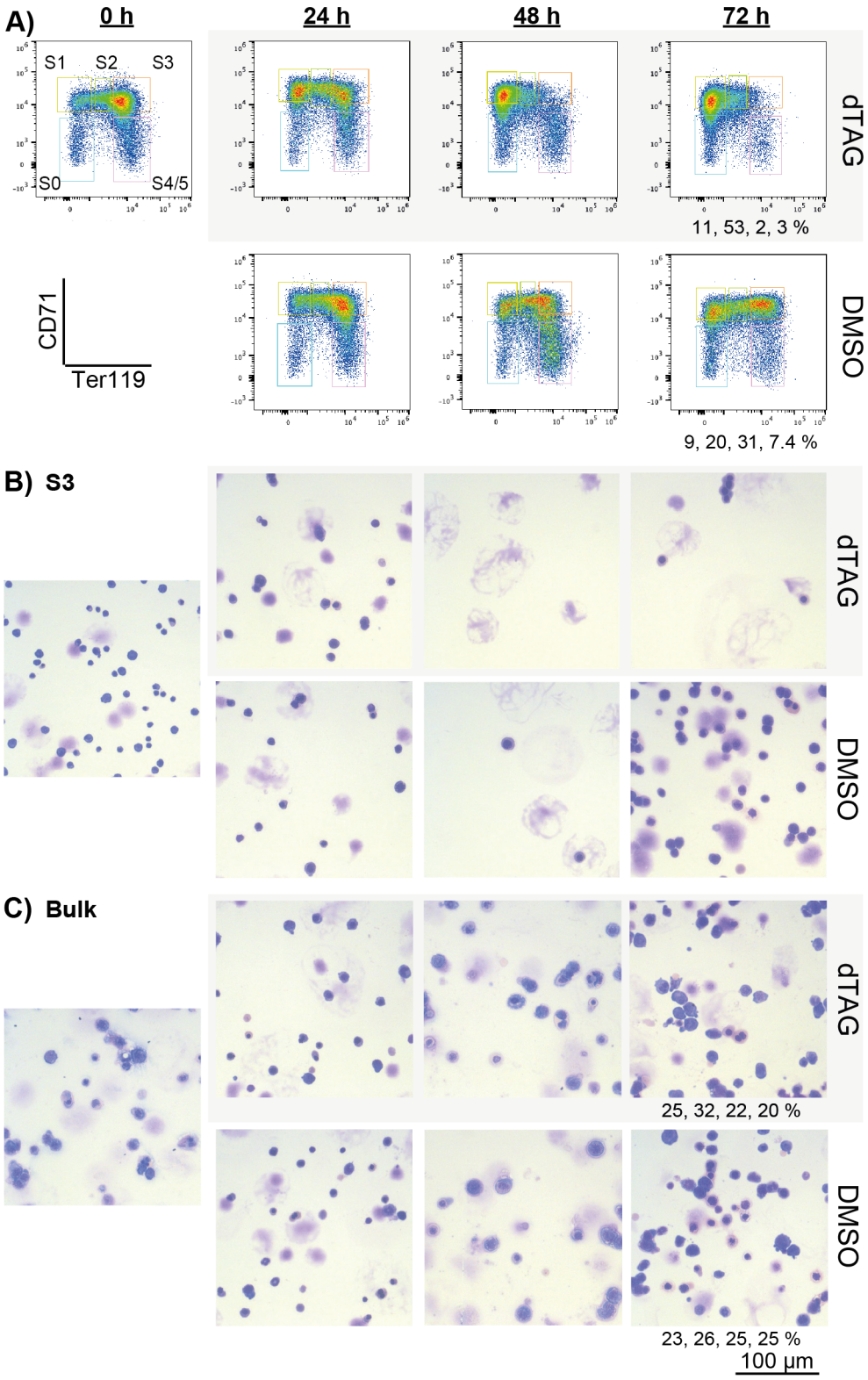

Fig. 5.6 – *Expression of the Ter119 cell surface marker is preserved for at least 24 hours post-KLF1 knockdown.* **A)** FACS plots of KLF1$^{StrepII\text{-}FKBP}$ mFL cells 0-72 hours after treatment with dTAG-13 to deplete KLF1. "0 h" is untreated with dTAG-13 or DMSO. At 48 and 72 hours there is a marked depletion of Ter119^{+} cells (S3-5 fractions). The percentage of cells treated for 72 hours in the S0, S1, S3, and S4/5 FACS gates are annotated below their respective plots. **Table 5.2** lists all FACS gate population

percentages. 500,000 cells were sorted per sample. **B)** Representative cytospin images of KLF1$^{StrepII\text{-}FKBP}$ mFL cells (N = 4) enriched for the S3 fraction 0-72 hours after treatment with dTAG-13 or DMSO. **C)** Representative cytospin images of bulk mFL cultures (N = 4) 0-72 hours after treatment with dTAG-13 or DMSO. Late erythroblast cells can be seen at all timepoints. The percentage of cells treated for 72 hours in the S0, S1, S3, and S4/5 FACS gates are annotated below their respective plots; n > 100. Scale bar = 100 μm for all images.

After 24 hours with dTAG-13, the proportion of cells in the S3 (CD71^{high}/Ter119^{high}; mid-late terminal erythroid) fraction was approximately half that compared to DMSO control but S3 cells remained separable by FACS (**Fig. 5.6A, B; Table 5.2; Table 5.3**). By 48 and 72 hours, the dTAG-13-treated S3 population was significantly diminished and very few cells could be isolated (**Fig. 5.6A, B; Table 5.2**). Notably, of those few separable S3 cells at 72 hours, all appeared to be late terminal erythroblasts by visual inspection (**Fig. 5.6B**). This suggests that some residual Ter119 expression persists that can be used to isolate a limited number of S3 cells.

Table 5.2 – ***Expression of the Ter119 cell surface marker is preserved for at least 24 hours post-KLF1 knockdown.*** The percentages of live/total cells and sorted-fractions/live cells from KLF1$^{StrepII\text{-}FKBP}$ mFL cells treated for 0-72 hours with dTAG-13 or DMSO. "S0" = CD71$^{low/mid}$/Ter119^{low}. "S1" = CD71^{high}/Ter119^{low}. "S2" = CD71^{high}/Ter119^{mid}. "S3" = CD71^{high}/Ter119^{high}. "S4/5" = CD71$^{low/mid}$/Ter119^{high}. The percentage of cells in the S3-S5 fractions is similar between 24 hours treatment with dTAG-13 or DMSO. 500,000 cells were sorted per sample.

		(%)					
Time Point	**Treatment**	**Live cells**	**S0**	**S1**	**S2**	**S3**	**S4/5**
0 hrs	none	97.1	3.98	8.42	12.1	47.7	11.6
24 hrs	DMSO	98.1	3.73	13.4	13.8	**37.2**	13.2
24 hrs	dTAG	97.7	4.61	24.4	18	**23.7**	12
48 hrs	DMSO	98.5	6.6	13.8	11.7	18.4	25.6
48 hrs	dTAG	96.2	8.72	50.1	6.65	2.6	8.39
72 hrs	DMSO	98	9.21	20.5	14.1	31.2	7.36
72 hrs	dTAG	95.4	11.5	52.9	10.3	1.9	3.08

Table 5.3 – *Biological replicates of the % S3 cells* in mFL populations cultured for 24 hours with dTAG-13 or DMSO. "S3" = $CD71^{high}/Ter119^{high}$.

Experiment ID	% S3 cells		% difference
	24 hrs DMSO	24 hrs dTAG	
08AUG23	33.6	18.9	56.4
27SEP23	33.0	11.5	34.9
24OCT23	38.2	24.4	63.7
24JUN24	37.2	13.2	35.4
Average:	35.5	17.0	**47.6**

Cytospin slides prepared from bulk (unsorted) mFL cultures showed erythroblasts of all terminal differentiation stages present at 48 and 72 hours with dTAG-13 or DMSO (**Fig. 5.6C**). Therefore, the lack of $Ter119^{+}$ S3 cells is likely not from a differentiation defect, but rather a lack of Ter119 expression, consistent with previous reports (**Siatecka and Bieker, 2011; Gnanapragasam *et al.*, 2016**).

Together, these findings suggest that CD71/Ter119 FACS can be used to separate terminally differentiating erythroblasts up to 24 hours after KLF1 knockdown. S3 cells from later timepoints may be successfully isolated, albeit at much lower numbers and should be interpreted with caution. S1 and S2 fractions are not reliably separable at any timepoint following KLF1 knockdown, as contamination with $Ter119^{-ve}$, later-stage erythroblasts is inevitable.

5.2.4 – The gene expression effects of acute KLF1 depletion in a late erythroid subset

Human patients and mouse models with mutant or deleted KLF1 express elevated ζ-globin. To better understand the mechanism(s) behind this, we must first establish whether acute depletion of KLF1$^{StrepII\text{-}FKBP}$ can recapitulate expected gene expression changes, in particular the *Hba-x* de-repression phenotype.

Given that Ter119 persists at 24 hours after KLF1 knockdown, I used CD71/Ter119 FACS to isolate a comparable mid/late terminal erythroid (S3) population from mFL cells that had been cultured for 24 hours with dTAG-13 or DMSO control. RT-qPCR then revealed gene expression changes in dTAG-13 vs DMSO treated cultures. For comparison, the baseline mRNA expression for each of these genes in S0-S5 mFL cells are in **Appendix A.5.1**.

Klf1 expression was not significantly changed at 24 hours, indicating that KLF1 protein knockdown does not affect *Klf1* mRNA production/turnover (**Fig. 5.7A**). *Lrf*, *Bcl11a,* and *Hbb-b1* are well-validated direct gene targets of KLF1. Of these, *Bcl11a* mRNA was significantly decreased after KLF1 knockdown. This suggests that KLF1 has been depleted to functionally significant levels, but that perhaps *Lrf* and *Hbb-b1* are not as easily downregulated by acute KLF1 loss alone.

Meanwhile, there was modest but not statistically significant upregulation of *Hbb-y* and *Hbb-bh1* (**Fig. 5.7B**). Contrastingly, in the KLF1$^{-/-}$ mouse model *Hbb-bh1* and *Hbb-y* are each upregulated ~11-fold (**Tallack *et al.*, 2012**). *Hbb-bh1* and *Hbb-y* are known direct targets of BCL11A and LRF; BCL11A knockout upregulates *Hbb-bh1* and *Hbb-y* (**Xu *et al.*, 2011; Mehta *et al.*, 2022**) while LRF knockout upregulates *Hbb-bh1* only (**Masuda *et al.* 2016; King *et al.*, 2021**). Residual BCL11A and particularly LRF protein likely explain the modest effects observed for these embryonic β-like genes.

From the literature, we might expect KLF1 knockdown to have no effect on *Hba-a1/2* expression but to at least modestly derepress *Hba-x*. At the resolution of this experiment, I observe a slight but not statistically significant increase in *Hba-a1/2*

expression and a significant 3.3-fold upregulation of *Hba-x*. In line with these literature and experimental observations, after 24 hours' treatment with dTAG-13, there is an increased α:β mRNA ratio, which is indicative of β-thalassemia, (**Fig. 5.7D**) and increased ζ-globin as a percentage of total α-like globin mRNA (**Fig. 5.7E**). Although it should be noted that these trends are subtle and do not reach statistical significance.

Overall, I conclude that degradation of KLF1 modestly derepresses ζ-globin in this model.

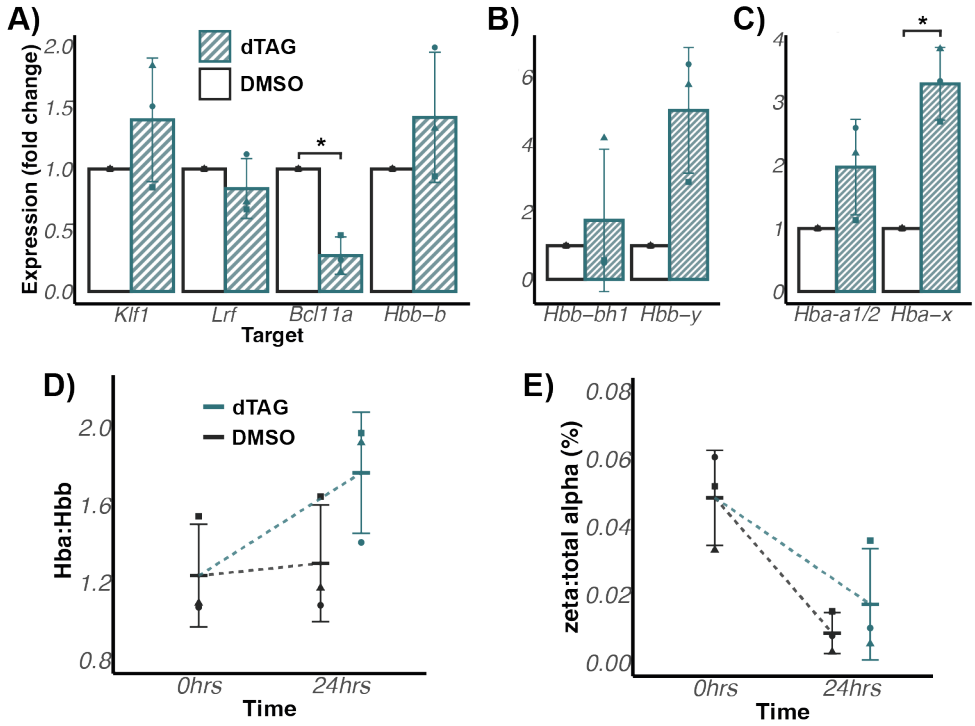

Fig 5.7 – Hba-x *is derepressed in late terminal erythroid (S3) cells after acute KLF1 knockdown.* RT-qPCR in CD71/Ter119 FACS sorted, E12.5, KLF1$^{StrepII\text{-}FKBP}$ mFL cells cultured for 24 hours with dTAG-13 or DMSO control. **A-C)** Relative gene expression fold change compared to DMSO control. Samples are normalised with the ΔΔCT method: first to *Rps18* housekeeping control, then to DMSO for each biological replicate. Statistical significance was calculated by one-sample Student's t-test, using the normalised DMSO control value of 1 as the reference theoretical value. "*" = $P \leqq 0.05$. **D)** Ratio of *Hba-a1/2* ΔCT to *Hbb-b1* ΔCT, where ΔCT is normalised to *Rps18*. **E)** *Hba-x* expression as a percentage of total α-like globin expression. (*Hba-x* ΔCT) / (*Hba-x* ΔCT + *Hba-a1/2* ΔCT) x 100, where ΔCT is normalised to *Rps18*. Error bars = ± one standard deviation. N = 3.

5.2.5 – The effect of KLF1 knockdown on Hba-x protein expression

As *Hba-x* mRNA was increased in the dTAG-treated S3 population, I next sought to verify if Hba-x was also derepressed at the protein-level. To allow maximum time for ζ-globin protein translation, I collected bulk mFL samples for isoelectric focusing (IEF) and IEF-Western blot after 96 hours of dTAG or DMSO treatment. ζ-globin protein was not detected by IEF or anti-ζ-globin IEF-Western blot after 96-hour KLF1$^{StrepII\text{-}FKBP}$ knockdown.

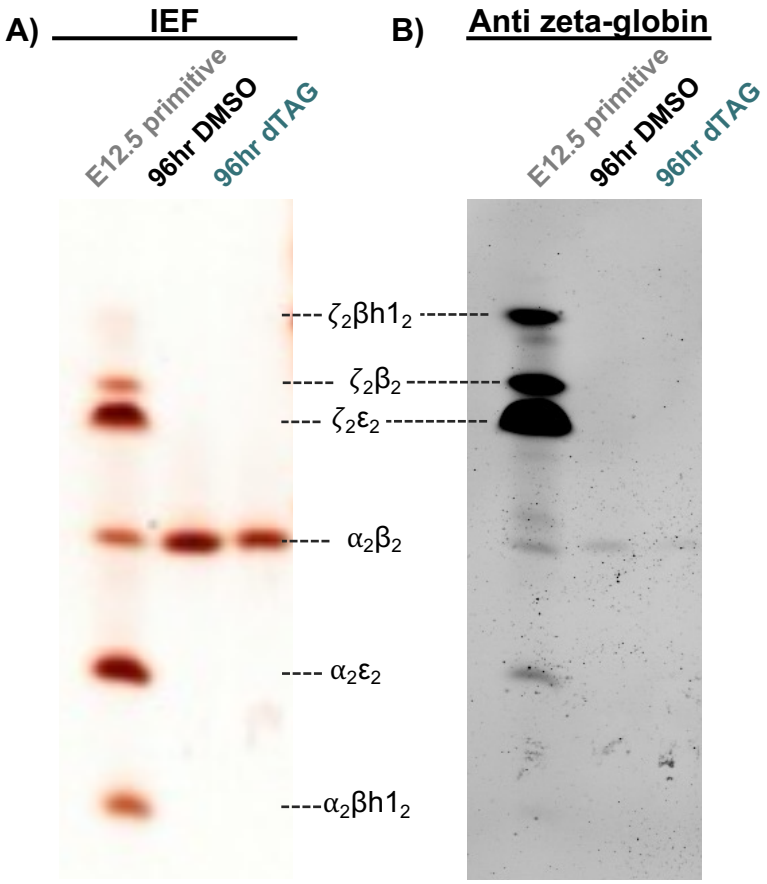

Fig 5.8 – *ζ-globin protein is not upregulated to detectable levels by IEF-Western blot.* A) Isoelectric focussing (IEF) gel separates haemoglobin tetramers by isoelectric point (pH at which the protein has no net electrical charge). Samples, left to right: E12.5 mouse primitive embryonic blood (ζ-globin positive control); definitive E12.5 mFL cells cultured for 96 hours with DMSO; definitive E12.5 mFL cells cultured for 96 hours with dTAG-13. **B)** Western blot, transferred from the IEF gel and probed with an anti-HBZ antibody (Abcam, Ab228651). Bands have been identified and annotated based on unpublished mouse IEF-Western blot for ζ-globin (S. Liu; data not shown). "ε" = Hbb-y. Full gel and blot in **Appendix Fig. A5.4**.

5.3 – Discussion

In this chapter, I have demonstrated that KLF1$^{StrepII\text{-}FKBP}$ protein can be rapidly and effectively depleted with dTAG-13 treatment. This acute depletion phenocopies at least some of the expected effects of KLF1 knockout, including: increased cell growth rate; impaired haemoglobinisation (both of which might result from impaired differentiation); loss of Ter119 cell surface marker expression; reduced *Bcl11a* expression; increased *Hbb-y* and *Hba-x* RNA expression; and an increased α:β transcript ratio. Together this confirms KLF1 was depleted to a functionally significant level and establishes the KLF1$^{StrepII\text{-}FKBP}$ mouse as a viable model to further study the effects of acute KLF1 depletion.

However, caution should be taken as some of the expected effects of KLF1 depletion were not observed in this system. Principally, there was no reduction of *Hbb-b1* or *Lrf*. The acuteness of KLF1 depletion and the timing of mRNA measurement could explain the lack of *Hbb-b1* repression. KLF1 is thought to be essential for the formation of the active chromatin hub (ACH) at the β-globin locus by binding to the β-globin gene promoters and the LCR enhancers, facilitating their interaction (**Drissen *et al.*, 2004**). The KLF1 binding motifs within the adult β-globin gene promoters have stronger binding affinity for KLF1 than those of the fetal/embryonic promoters (**Donze *et al.*, 1995**; **Tallack *et al.*, 2010**). Thus, KLF1-driven promoter competition plays a key role in β-globin switching (reviewed in **Tallack and Perkins, 2013**). With total KLF1 knockout, this aspect of promoter competition is lost and haemoglobin switching is impaired. With acute KLF1$^{StrepII\text{-}FKBP}$ depletion in definitive erythroblasts, however, the adult β-globin promoters would already be established within the ACH and be highly transcribed. Already in this transcriptionally competent and promoter-competition "dominant" state, the adult β-globin genes might be impervious to acute loss of KLF1. Meanwhile, *Lrf* is known to be directly activated by KLF1 and *Lrf* transcription is rapidly upregulated (within 1 hour) upon KLF1 induction in a K1ER (KLF1-inducible) cell line (**Norton *et al.*, 2017**). With that, it is surprising not to have observed any acute *Lrf* downregulation upon KLF1 depletion. But it is also thought that other activating TFs, such as GATA1, are involved in LRF regulation and KLF1$^{-/-}$ does not entirely repress *Lrf* mRNA or protein expression (**Norton *et al.*, 2017**).

Therefore, functionally redundant TF(s) may still promote normal *Lrf* expression after acute KLF1 depletion. Clearly, the complexities and number of parallel and functionally redundant mechanisms at play need to be accounted and controlled for when planning future experiments in the KLF1$^{StrepII\text{-}FKBP}$ model.

5.3.1 – De-repression of *Hba-x* at the transcript level

Additionally, the extent of *Hba-x* transcriptional de-repression observed was lower than expected. I observed *Hba-x* mRNA upregulated by 3.3-fold after loss of KLF1. This is a similar magnitude of de-repression to the reported KLF1$^{-/-}$ human patient with a 4.5-fold increase in *HBZ* by RNA-seq. However, this is lower upregulation than other KLF1-mutant mouse models where *Hba-x* is derepressed up to several hundred-fold. These KLF1 mutants are thought to cause transcriptional dysregulation by dominant *changes* to KLF1's TF function, rather than abrogation of KLF1, which could explain the different effects. However, there are also several differences between these long-term KLF1 perturbation models and acute KLF1$^{StrepII\text{-}FKBP}$ depletion that might explain the lower magnitude of ζ-globin de-repression in KLF1$^{StrepII\text{-}FKBP}$ depletion.

Residual KLF1 protein

First, in a KLF1 genetic knockout model there is no residual KLF1 protein. In the KLF1$^{StrepII\text{-}FKBP}$ depletion model, there may be low levels (<1%) of residual KLF1 protein at the timepoints studied. It cannot be excluded, without a reliable KLF1 ChIP or more sensitive protein quantitation method, that some residual KLF1 protein remains bound to chromatin.

Residual transcription factors

Second and most likely to mitigate *Hba-x* de-repression is the persistence of the repressive transcription factors, BCL11A and LRF, after 24-hours KLF1$^{StrepII\text{-}FKBP}$ depletion. Although *Bcl11a* mRNA was significantly reduced by 24-hours KLF1$^{StrepII\text{-}FKBP}$ depletion, the protein half-life of BCL11A is ~24 hours (**Mehta *et al.*, 2022**). We therefore might expect ~50% residual BCL11A and normal levels of LRF at this timepoint, which could carryover a significant repressive effect. Although notably, in BCl11A$^{-/+}$ mice *Hba-x* is not significantly upregulated (**King**

***et al.*, 2021**), suggesting that the KLF1 depletion is having a greater effect beyond BCL11A reduction alone. Nevertheless, the persistence of negative repressors is likely limiting the de-repression effect on *Hbb-bh1*, *Hbb-y, and Hba-x* compared to total KLF1 knockout or perturbation models.

To overcome the confounding effects of residual BCL11A/LRF, it would be interesting to cross the KLF1$^{\text{StrepII-FKBP}}$ mouse with our ζ-promoter edited (ZPE) mouse model that harbours specific deletions of the BCL11A and LRF binding motifs within the *Hba-x* promoter, which itself shows ~15-20% *Hba-x* transcripts of total α-like globin mRNA (**Liu *et al.*, manuscript in preparation**). It would be extremely interesting to test whether acute KLF1 depletion in this context could further modify *Hba-x* expression, thus definitively determining whether there are any additional *Hba-x* regulatory pathways involving KLF1.

Stress erythropoiesis

Third, a major difference between systemic KLF1 perturbance and acute depletion is stress erythropoiesis. Stress erythropoiesis is a response to anaemic stress and can involve several different mechanisms and signalling pathways, resulting in increased medullary haematopoiesis or induction of extra-medullary haematopoiesis in the spleen. The erythroid niche in which an erythroblast develops is intricately linked to its development. In stress erythropoiesis, human γ-globin but not ε-globin is upregulated. Whether ζ-globin is also upregulated as a stress response is unknown. However, it is plausible that some aspect of stress erythropoiesis compounds KLF1 perturbation to further alter globin gene expression. Alternatively, but not necessarily mutually exclusively, chronic stress *in vivo* might select for survival of erythroblasts with the highest expression of embryonic/fetal globins, leading to an enrichment of these globins compared to *in vitro* cultured cells. One way to control for this might be to treat the mice with acetylphenylhydrazine before mFL harvest, with or without KLF1 depletion. Assessing if there is a stepwise increase in *Hba-x* de-repression, compared to WT, from stress alone and from stress with KLF1$^{\text{StrepII-FKBP}}$ depletion could address this potential confounder.

A "hit and run" mechanism of KLF1 action

Finally, an early "hit and run" mechanism of KLF1 action could produce differential effects in total KLF1 knockout vs acute knockdown models. Indeed, KLF1 has been proposed to function as pioneer factor to establish open, poised, and active chromatin via its recruitment of CBP/P300 acetylase and SWI/SNF chromatin remodeller complexes (**Gillinder *et al.*, 2018**; **Mukherjee *et al.*, 2022**). KLF1 is known to have transcription activating and repressing functions, that these are regulated by differential PTMs, and that KLF1 carries different PTMs and functions at different stages of erythropoiesis (**Bieker and Philipsen, 2024**). Furthermore, transcriptional repressors are thought to act via these transient "hit and run" mechanisms while activators bind target genes more frequently (**Shah *et al.*, 2019**) We cannot rule out that KLF1 might act directly and distinctly at *Hba-x* in early erythropoiesis to establish *Hba-x* repression and its effect on *Hba-x* is missed by knocking out KLF1 in this relatively late stage of erythropoiesis.

Differentiation differences

A final consideration is, despite best efforts, a potential differentiation discrepancy. From the FACS analysis (**Section 5.2.3**), I have shown successful isolation of S3 cells from bulk populations cultured for 24 hours with dTAG-13 or DMSO and that these S3 populations appear comparable by visual inspection. However, it cannot be ruled out, without further molecular characterisation, that the two S3 populations are not identical. Perhaps only the more mature erythroblasts retain their Ter119 expression after KLF1 depletion. This could explain the slight increases in pan-globin expression, although this could also be technical variation. Although, even if this is the case, the fold change in ζ-globin upregulation exceeds that of the adult globins, suggesting a true de-repression effect above a difference in erythroblast maturity.

To improve upon this and enable more refined analysis of the effects of KLF1 depletion on precise erythroblast populations in future, we will need to employ a Ter119-independent FACS strategy. As Ter119 and CD44 expression is dramatically reduced with KLF1 knockout (**Siatecka and Bieker, 2011**), imaging-based flow cytometry methods hold the most promise. This has been used to

separate terminally differentiating KLF1$^{-/-}$ erythroblasts based on cell size, morphology, and CD71/c-kit expression (**Gnanapragasam *et al.*, 2016**), effectively separating cells from early precursors through to enucleating erythroblasts. Combining image flow data with acute KLF1 depletion would allow for several more nuanced questions to be asked. First, longer timecourses could be performed, which would allow us to differentiate primary, secondary, and tertiary transcriptional effects of acute KLF1$^{StrepII\text{-}FKBP}$ depletion. I also suspect that this would allow us to see more exaggerated transcriptional changes after KLF1$^{StrepII\text{-}FKBP}$ depletion, especially at genes such as the fetal and embryonic globins that rely on intermediate repressors. This will be an essential method to maximise the potential of the KLF1$^{StrepII\text{-}FKBP}$ model.

5.3.2 – De-repression of *Hba-x* at the protein level

Notably, in this model system, no ζ-globin protein was detected by IEF-Western blot after 96-hours KLF1$^{StrepII\text{-}FKBP}$ depletion. This contrasts the KLF1$^{-/-}$ patient who was suspected to have <24% Hb Portland (**Magor *et al.*, 2015**). None of the examples of KLF1$^{-/-}$ mice outlined in **Table 5.1** report on ζ-globin protein, but *Bcl11a* and *Lrf* knockout studies observe significant ζ-globin protein expression. In the acute depletion of KLF1$^{StrepII\text{-}FKBP}$, it is possible that ζ-globin protein was present below the detection threshold of this IEF-Western blot. However, it is perhaps more likely that ζ-globin protein had not been produced at all. There are a few possible explanations for this.

One explanation could be that ζ-globin is never derepressed at the protein level in any KLF1$^{-/-}$ model. The only recorded KLF1$^{-/-}$ patient had <24% *suspected* but not confirmed ζ-globin by HPLC and there is no record of ζ-globin protein upregulation in KLF1$^{-/-}$ mice. Alternatively, the ζ-globin protein detected in the KLF1$^{-/-}$ patient was true, and this could be a discrepancy in human vs mouse ζ-globin regulation. Clearly, the first step to discriminate between these would be to specifically assay for ζ-globin protein in definitive blood from the KLF1$^{-/-}$ patient and in the KLF1$^{-/-}$ mouse. As KLF1$^{-/-}$ is ordinarily embryonic lethal, it is difficult to measure ζ-globin protein in mouse definitive peripheral blood but it could instead be measured in cultured E14.5 KLF1$^{-/-}$ fetal livers. A benefit of the KLF1$^{StrepII\text{-}FKBP}$

model, that would complement this investigation and overcome potential issues of primitive blood contamination, is the ability to deplete KLF1 in adult mice and measure for Hba-x de-repression in their peripheral blood. This would allow us to study KLF1 loss in the most clinically relevant tissue – the adult peripheral blood – which is unavailable in total KLF1 knockout models.

Alternatively, it is possible that acute KLF1$^{StrepII\text{-}FKBP}$ depletion derepresses *Hba-x* at the RNA but not the protein level. ζ-globin is widely considered to be transcriptionally "silent" in definitive erythroid cells. While it is true that ζ-globin protein is ordinarily undetectable in mouse and human peripheral blood, it is in fact common to see low levels of accessibility and transcription of the *Hba-x* gene (**Fig. 5.9**). Interpretation of the ζ-globin gene as "silent" might just be consequence of constant comparison to the extremely highly expressed, acetylated and accessible *Hba-a1/2* genes. We must therefore consider that *Hba-x* silencing may occur at the post-transcription as well as transcription level.

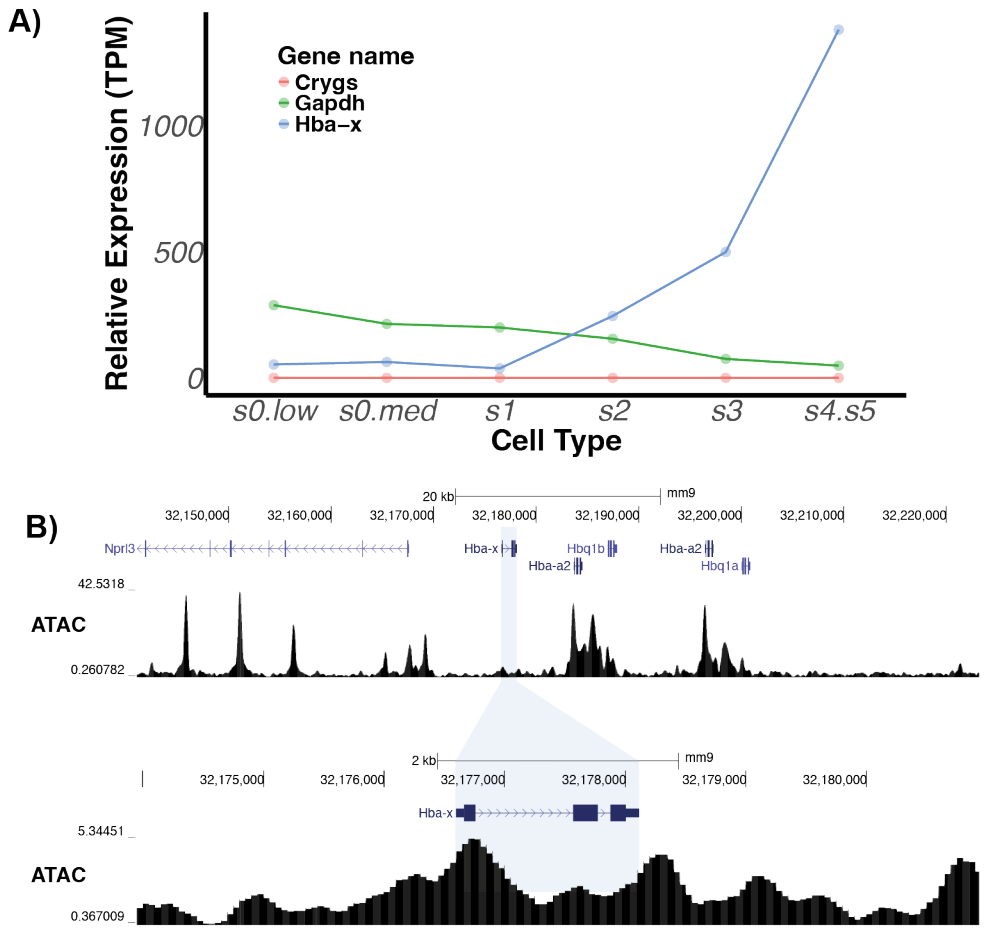

Fig. 5.9 – *ζ-globin mRNA is present in definitive erythroid cells.* A) ζ-globin mRNA is detected at a level more comparable to the housekeeping control, *Gapdh*, than to the lineage-restricted *Crygs* gene. *Crygs* encodes the most significant of the γ-crystallin proteins of the adult eye lens tissue. Like ζ-globin, *Crygs* is tissue-specifically and developmentally regulated. Relative RNA expression (transcripts per million; TPM) from single-cell mRNA-sequencing data in CD71/Ter119 FACS-sorted E13.5 mFL erythroblast populations. Replotted with permission from **Oudelaar *et al.*, 2020**. **B)** UCSC ATAC-seq tracks in definitive erythroid cells. Low levels of chromatin accessibility (ATAC-seq) are found over the *Hba-x* gene and proximal promoter in E12.5 mFL cells.

WT humans and mice consistently show low level ζ-globin gene activity in definitive erythroblasts but with no detectable ζ-globin protein (**Chui *et al.*, 1989; Tang *et al.*, 1992; King *et al.*, 2021**). It is an intriguing possibility that there is a post-transcriptional level of ζ-globin gene silencing in definitive erythroblasts that is unaffected by acute KLF1 knockdown. Evidence for a mechanism of post-transcriptional silencing of ζ-globin comes primarily from a series of transgenic studies in the 1990s. Here, ζ-globin mRNA was shown to have approximately half the stability in adult erythroblasts of α-globin mRNA, attributed to differences in their 3' UTRs (**Russell, Lee and Liebhaber, 1998a; Russell *et al.,* 1998b**). First, is ζ-globin's shorter polyA tail, which might affect mRNA stability and translation efficiency (**Russell, Lee and Liebhaber, 1998a; Russell *et al.,* 1998b; Thompson *et al.*, 2017**). Second, the ζ-globin 3' UTR was shown to harbour a single C>G substitution (3' UTR position 42), relative to α-globin, that reduced assembly of the α-complex, a mRNA-stabilising protein complex (**Kildejian, Wang and Liebhaber, 1995; Russell *et al.*, 1998b**). ζ-globin mRNA stability has been also shown to rely on additional factors, such as binding of the RNA stabilising protein, AUF1, to an AGUG (3' UTR position 57-60) motif in the 3' UTR (**He *et al.*, 2014**).

Together, these experiments show multiple potential points of post-transcriptional control of ζ-globin mRNA stability. Whether or how there is differential control of ζ-globin mRNA stability between primitive and definitive erythroid cells, in addition to transcriptional regulation, is unknown. My observations in the KLF1$^{\text{StrepII-FKBP}}$ model mirror other observations in our laboratory that ζ-globin is more easily derepressed at the transcription than protein level. Arguably, my results add

circumstantial evidence in favour of a tentative, post-transcriptional silencing mechanism of *Hba-x* that would be extremely interesting to investigate further. As a first step, digital droplet PCR (ddPCR) or fluorescent in-situ hybridisation (FISH) for sensitive quantitation of *nascent Hba-x* transcripts could help to distinguish whether there is greater fold change increase in nascent *Hba-x* transcription than at the mRNA level, which would indicate mRNA degradation.

Together, this paints a picture of there perhaps being two levels of ζ-globin repression. A level of transcriptional repression, which is partly overcome by acute KLF1$^{StrepII\text{-}FKBP}$. And second, a level of post-transcriptional repression, which does not seem to be overcome by acute KLF1$^{StrepII\text{-}FKBP}$ depletion but is derepressed in systemic KLF1 perturbation models (**Table 5.1**).

5.4 – Chapter 5 Conclusions

In this chapter, I have demonstrated that KLF1$^{StrepII\text{-}FKBP}$ protein can be rapidly and acutely depleted with dTAG-13 treatment. This acute depletion phenocopies several of the expected effects of KLF1 knockout and confirms KLF1 was depleted to a functionally significant level. ζ-globin appears to be derepressed at the mRNA but not protein level. However, this mRNA de-repression effect is relatively modest and would benefit from follow up work with more refined methods – particularly image-based flow-cytometry – to increase the certainty of this conclusion. Some expected KLF1$^{-/-}$ effects, such as reduced *Lrf* and *Hbb-b1* or Hba-x protein expression were not observed at 24-hours post-knockdown. But the discrepancies between the acute depletion and total knockout models themselves highlight particularly interesting areas for future research, such as TF coordination and functional redundancy, *Hba-x* RNA stability, and the potential confounding effects of stress erythropoiesis between *in vivo* vs *in vitro* models. Nonetheless, this establishes the KLF1$^{StrepII\text{-}FKBP}$ mouse as a viable and valuable model to further study the effects of acute KLF1 depletion.

5.5 – Chapter 5 Acknowledgements

Thank you to Chris Babbs, for taking me under his wing and introducing me to the world of ζ-globin with endless enthusiasm. Thank you to Siyu Liu and Ayesha Ejaz for all their shared help, advice, and knowledge. Thank you to Aude-Anaïs Olijnik, for teaching me how to do her (incredibly neat) Western blot protocol. As always, thank you to Jackie Sloane-Stanley for taking care of the animal work. Thank you to Rob Beagrie for advising on the experimental design. Finally, thank you very much to Doug and Mira for their grounding advice and support through the rollercoaster of experiments (and emotions) that went into this chapter.

Chapter 6 – Discussion

A key question in molecular biology is how are genes turned on and off throughout development and differentiation? Precise spatio-temporal gene regulation involves complex interplay between cis- and trans-acting regulatory factors. Cis-acting elements include enhancers, promoters, and insulators. Trans-acting factors include transcription factors (TFs), co-factors, chromatin remodelling factors, and ncRNAs. Understanding how these factors are individually regulated, and together integrate, to fine-tune gene expression is essential to answering this question. The α-globin locus is an extremely well characterised gene locus which integrates all key types of cis- and trans-acting factors to confer precise developmental control over the α-like globin genes. I have used this model to study two such factors – CTCF binding site (CBS) insulators and the erythroid-specific TF KLF1 – in α-like globin gene regulation throughout erythroid differentiation and development.

6.1 – The relative contribution of CTCF to gene regulation

It is widely accepted that enhancers and promoters interact before or around the time of transcription initiation. Controlling this interaction event is one level of gene regulation. CTCF-insulated, cohesin-mediated loop extrusion is just one – and by no means the only – mechanism that can influence enhancer-promoter interactions. Moreover, CBSs are not the only type of insulator element and loop extrusion may also be limited by transcription and replication complexes **(Bonev *et al.*, 2017; Chahar *et al.*, 2023; Jeppsson *et al.*, 2022; Banigan *et al.*, 2023; H. Zhang *et al.*, 2023; S. Zhang *et al.*, 2023).** I outline some of the theories of how enhancer-promoter contact may be established and act to regulate transcription and the extent to which CTCF-delimited loop extrusion contributes to this.

6.2 – The nature of enhancer-promoter interactions

The *formation* of an enhancer-promoter contact

There are three broad and widely accepted theoretical mechanisms by which an enhancer-promoter interaction may be established. One is cohesin loop extrusion, in which cohesin translocating throughout the locus brings the enhancer and promoter into close physical proximity and/or tethers the enhancer and promoter into a TAD/sub-TAD of increased interaction probability (**Fig. 6.1A-i**). Second is a diffusion-based model, whereby the enhancer, promoter, and TF proteins "freely" diffuse and coalesce into a transcriptionally permissive hub (**Fig. 6.1A-ii**). Third is a protein bridging model, where specific protein-protein interactions, for example dimerisation of LDB1, YY1, or mediator, tethers the enhancer to the promoter (**Fig. 6.1A-iii**). Notably, in each, homotypic interactions result in a locally high concentration of proteins to form the transcriptionally permissive environment.

The *nature* of an enhancer-promoter contact

Once assembled, by whichever mechanism, there is then the question of how an enhancer-promoter interaction relates to transcription initiation? Again, there are three broadly accepted, overarching mechanisms for this. First is a "transcription hub" model. In this, the enhancer-promoter interaction (established by whichever mechanism) might seed the accumulation of a locally high concentration of all the necessary TFs, co-factors, chromatin modifiers, and Pol II proteins required for transcription initiation. The enhancer and promoter may both lie within this hub when transcription occurs (**Fig. 6.1B-i**). Second is an enhancer-promoter "kissing" model. In this, the enhancer-promoter contact is transient, perhaps conferring a signal (such as transcription activator proteins or PTMs) between enhancer and promoter (**Fig. 6.1B-ii**). Transcription might occur after the contact event, when enhancer and promoter are not necessarily in proximity. Third is a "kissing hub" model. In this, the enhancer(s) may form a transcriptionally permissive hub (by any of the mechanisms above) and the promoter can "dip" in and out of this (e.g. by diffusion or loop extrusion), without necessarily making close physical contact with the enhancers (**Fig. 6.1B-iii**).

These proposed mechanisms are not exhaustive and are extensively reviewed elsewhere (**Yang and Hansen, 2024**). One great difficulty when trying to resolve these models is that they are by no means mutually exclusive. Loop extrusion may help to form protein bridges, or protein bridges may help to seed a wider transcription hub, for example. Each of these mechanisms may contribute to different degrees at different loci – or at the same locus under different conditions, e.g. *de novo* versus steady-state expression. To further complicate matters, what even constitutes a functional enhancer-promoter "contact" is highly debatable. Is general proximity (~200-400 nm) sufficient to bring transcription activators together to promote transcription? Is closer contact (~20-40 nm) actually required to confer specific proteins or PTMs from enhancer to promoter? These are some of the core questions in the current field of molecular gene regulation.

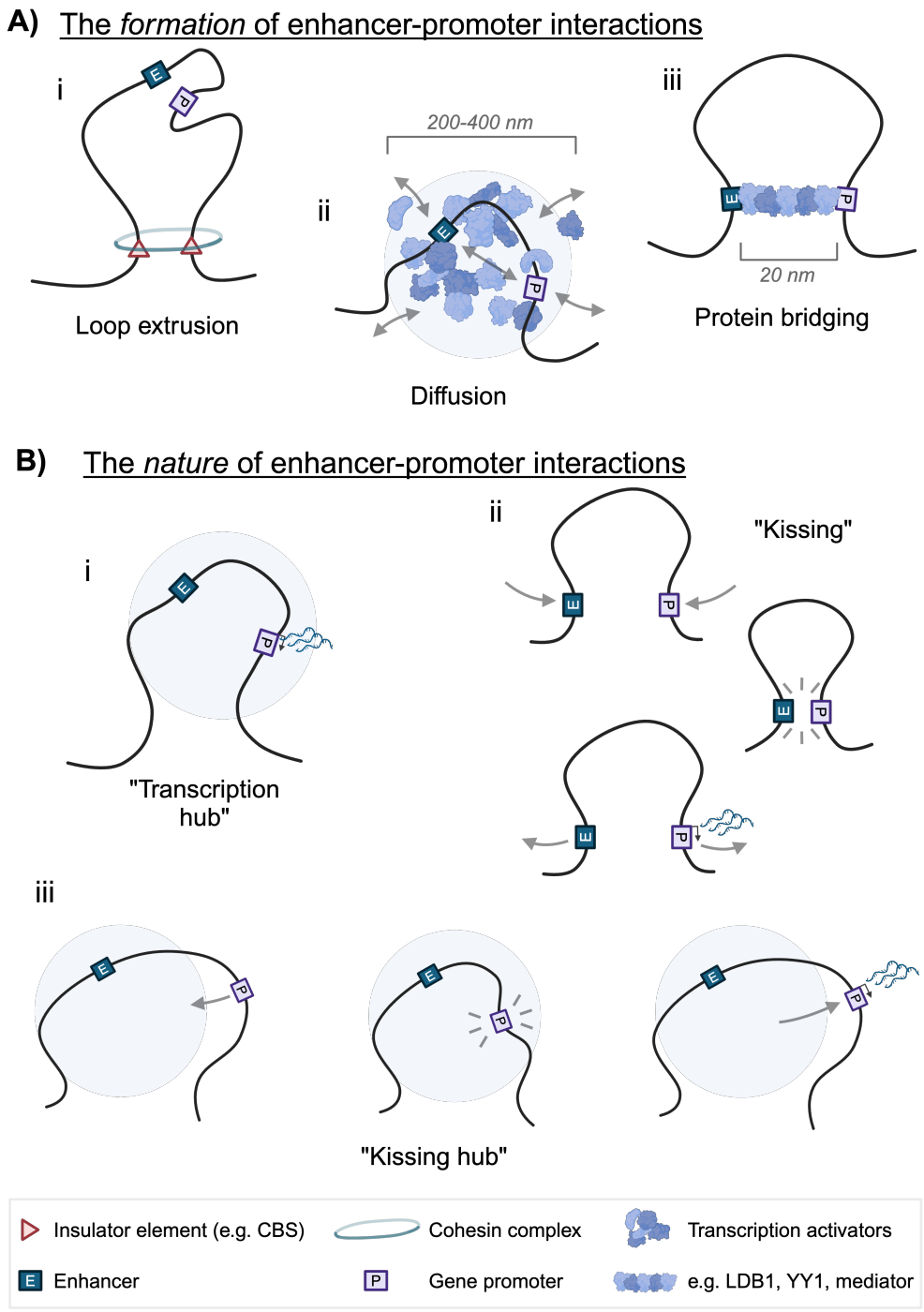

Fig. 6.1 – *Broad theoretical mechanisms for the formation and nature of enhancer-promoter interactions.* A) *The formation of enhancer-promoter interactions*. **A-i)** Loop extrusion: the cohesin complex loads onto chromatin and bi-directionally extrudes the DNA until it reaches an insulator element. This is a highly dynamic process that increases the likelihood of interaction within the looped domain (TAD or sub-TAD) that is formed. **A-ii)** Diffusion: the enhancer and promoter elements, transcription factors, co-factors, chromatin modifiers, and Pol II might "freely" diffuse towards one another. **A-iii)** Protein bridging: dimerisation, oligomerisation, or simultaneous enhancer-promoter binding of proteins or protein complexes, such as LDB1, YY1, or mediator, might facilitate enhancer-promoter interaction formation. Both B-ii and B-iii are likely to be facilitated by

homotypic protein-protein interactions, for example between unstructured protein domains or specific dimerisation/interaction domains. **B)** *The nature of enhancer-promoter interactions.* I.e. how do enhancer-promoter interactions relate to transcription initiation? **B-i)** Transcription hub model: the enhancer-promoter interaction establishes a locally high concentration of transcription activator proteins. This forms a transcription-activating microenvironment or "hub". Enhancer, promoter, and activator proteins are within close physical proximity for maintenance of the hub and for transcription initiation. **B-ii)** Enhancer-promoter kissing model: single or multiple, discrete, transient, enhancer-promoter proximity event(s) facilitate transcription initiation. This may be to confer a signal, such as activator proteins or PTMs, between the enhancer and promoter. Transcription initiation may occur upon or after the proximity event(s). **B-iii)** Kissing hub model: the enhancer(s) may form a transcription-activating hub domain with locally high concentrations of activator proteins and independently of the gene promoter. The promoter may then "dip" into the enhancer hub to acquire activating proteins or PTMs. Transcription may occur upon or after the "dipping" event. Close physical proximity between enhancer and promoter might not be required. None of these mechanisms are necessarily mutually exclusive. Any combination of formation mechanisms may conceivably contribute to any of the contact nature models. Different mechanisms may apply at different gene loci.

With good reason, there has been scepticism over the functional importance of loop extruded TADs/sub-TADs for gene regulation. Several independent experiments that acutely knocked down CTCF or cohesin came to similar conclusions: A/B compartments were largely unaffected, TADs/sub-TADs were dissolved, and transcription was broadly unaffected (**Nora *et al.*, 2017; Schwarzer *et al.*, 2017; Rao *et al.*, 2017; Hyle *et al.*, 2019**). However, in all of these, significant transcriptional effects were observed at a subset of loci. Notably, **Rao *et al.*, 2017**, suggest that super-enhancers are particularly reliant on cohesin for bringing them into contact with their cognate promoters, and cohesin loss was more likely to reduce super-enhancer-mediated gene expression. Notably, none of the CTCF/cohesin knockdown experiments specifically probed transcription effects across cell differentiation. What if TAD/sub-TAD formation is most critical in establishing *de novo* gene expression, e.g. at tissue-specific loci, but is dispensable for expression maintenance after an active chromatin hub has been formed?

6.3 – The extent of the role of CBSs at the α-globin locus

We have repeatedly observed a role for cohesin loop extrusion at the α-globin locus. Acute knockdown of cohesin reduced α-globin transcription by ~30% in early erythroid differentiation (**Stolper *et al.*, 2023**). Knockout of the 5' CBS did not affect *Hba-a1/2* transcription, but did cause ectopic enhancer interactions and 5' gene expression (**Hanssen *et al.*, 2017**). Insertion of reverse-orientated CBS or a transcribing *Hba-a1* gene between the enhancer elements and *Hba-a1* strongly reduced endogenous *Hba-a1* expression, accumulated cohesin by RAD21 ChIP-seq, and formed a boundary to enhancer interactions by Capture-C (**Stolper *et al.*, 2023; Tsang *et al.*, 2023; and Cornell *et al.*, manuscript in preparation**). This suggests cohesin loop extrusion is important for establishing α-globin transcription in early erythroid cells, perhaps by helping to establish a transcriptionally active hub *de novo*. These data point to a model whereby cohesin may load at the α-globin enhancers and bidirectionally extrude chromatin. CBSs do play a role in this, as the major 5' sub-TAD boundary is the HS-38/HS-39 CBSs (**Hanssen *et al.*, 2017**). However, CBSs are not the only type of boundary element, as I have shown that CBSs appear to be dispensable for sub-TAD integrity at the 3' boundary (**Chapter 3**; **Cornell *et al.*, manuscript in preparation**). A total CTCF knockout model, with deletion of all 5' and 3' CBS, would help to confirm the absolute contribution of CBS to α-globin regulation. If similar effects to the HS-38/HS-39 deletion model are observed, this would confirm discrete function at the 5' boundary only. If greater boundary dysfunction is observed, however, that might suggest cooperative function between flanking CBS. As neither CBS deletion from the 5' or 3' boundary affected α-globin expression or intra-sub-TAD interactions (**Hanssen *et al.*, 2017; Chapter 3**), a total CTCF knockout model would also help answer whether CBS are required to establish or maintain α-globin expression at all.

Importantly, cohesin depletion reduced, rather than abolished, α-globin expression. While loop extrusion almost certainly plays a role, it is not the only mechanism capable of establishing α-globin transcription. Further investigation of potential protein bridging factors, such as YY1, LDB1, and mediator, will help to determine their contribution to α-globin regulation. Likewise, efforts are being

made to establish super-resolution live cell imaging of the α-globin enhancers, promoters, and nascent RNA. This, combined with the characterisation of the type and abundance of protein factors present before, during, and after a transcriptional burst, will help to resolve the mechanism(s) linking enhancer-promoter contact to α-globin expression.

6.4 – KLF1 and ζ-globin regulation: a species difference?

Regardless of how sub-TADs and enhancer-promoter contacts form, it is universally acknowledged that tissue-specific transcription factors play a key role in the expression of developmentally regulated genes.

In humans, patients with heterozygous or compound heterozygous mutations that compromise or ablate *KLF1* express embryonic ζ-globin protein in adult life. There is a conserved regulatory pathway – the KLF1-BCL11A/LRF-ζ-globin axis – connecting KLF1 to ζ-globin silencing in human and mouse definitive erythropoiesis alike. However, there is no strong evidence that KLF1 knockout is sufficient to derepress ζ-globin in mouse as it appears to in human. This could be due to the lack of published work reporting ζ-globin expression. Alternatively, it may be that KLF1 knockout is insufficient to derepress ζ-globin in the mouse. In the third results chapter, I used the KLF1$^{StrepII\text{-}FKBP}$ mouse to investigate this. I find that acute degradation of KLF1$^{StrepII\text{-}FKBP}$ modestly derepressed *Hba-x* RNA to a level similar to the KLF1$^{-/-}$ human case. However, unlike the KLF1$^{-/-}$ patient but in keeping with KLF1$^{-/-}$ mice, I observe no upregulation of ζ-globin protein. Overall, I cannot conclude that KLF1 depletion significantly derepresses *Hba-x* in mouse.

This raises the question of how the KLF1-BCL11A/LRF-ζ-globin regulatory pathway differs between mouse and human. KLF1, BCL11A, and LRF structure, function, and regulation are generally well conserved between human and mouse (**Caria *et al.*, 2022; Beaker and Philipsen, 2024; Satterwhite *et al.*, 2001; Chondrou *et al.*, 2020**). In line with this, there is a similar level of low-to-moderate residual BCL11A and LRF expression in KLF1$^{-/-}$ human and mouse (**Zhou *et al.*, 2010; Magor *et al.*, 2015; Norton *et al.*, 2017**). It seems unlikely therefore that differential abrogation or action of intermediate repressors is the cause. Instead, perhaps we are observing a difference in ζ-globin-specific regulation.

The ζ-globin promoter and coding sequence are highly conserved between human and mouse (**Hughes *et al.*, 2005**). However, other non-coding regions are implicated in ζ-globin regulation and are less well conserved. For example, human *HBZ* contains an intronic CpG island that is not present in mouse *Hba-x*. There is also low conservation of the 3' UTR, despite being implicated in post-transcriptional ζ-globin regulation, as discussed in **Chapter 5**. These regions could account for the discrepancy in ζ-globin regulation between human and mouse, but very little is known about the regulatory contribution of each.

Clearly, there is a need to further investigate the roles of the intronic and 3' UTR sequences in ζ-globin regulation. First, verification of the exact contribution of each non-coding sequence to ζ-globin regulation should be made. These results must then be carefully compared between human and mouse to establish any species differences. This will be instrumental for the long-term aim to therapeutically reactivate ζ-globin. Overall, the work of this thesis highlights the need to verify ζ-globin regulation experiments in human and mouse and to not assume that effects will translate across species.

6.5 – KLF1$^{\text{StrepII-FKBP}}$: a powerful tool for studying KLF1

In future, the inducible KLF1$^{\text{StrepII-FKBP}}$ knockdown model will be a powerful tool to address outstanding questions of KLF1's roles and requirements in its many erythroid functions.

For example, this model could address the outstanding question of whether KLF1's role in enucleation is restricted to orthochromatic erythroblasts, or whether KLF1 also plays an earlier role in the cytoskeleton or cell membrane to establish an enucleation-permissive state (**Bieker and Philipsen, 2024**)? This could be achieved with relatively simple assays in the KLF1$^{\text{StrepII-FKBP}}$ model. For example, CD71/Ter119 FACS sorting of mFL cells into S0-S5 enriched populations; culture in differentiation media with dTAG-13; comparison of reticulocyte formation with flow cytometry for thiazole orange (an RNA stain) and CD71.

Similar mFL stratification protocols could be applied to measure the window of sensitivity to, and effects of, KLF1 on a range of other erythroid functions, from cell cycle to iron transport. However, as noted in **Chapter 5's** preliminary exploration of the gene expression effects after KLF1 depletion, the dependency of erythroid differentiation on KLF1 will confound most results unless accounted for. It will be imperative therefore to use a Ter119-independent, erythroid stratification protocol – such as image flow-cytometry – for such future work.

A proven repeatable ChIP method for KLF1 remains a critical but unmet need for the study of KLF1. Despite limited success so far, the StrepII tag feature of KLF1$^{StrepII\text{-}FKBP}$ may yet yield such results with continued optimisation. However, if reproducible ChIP for KLF1 cannot be achieved even with the StrepII tag, the direct binding targets of KLF1 might still be verified by utilising the degron feature of the model. For example, acute KLF1 depletion followed closely by nascent RNA-seq could identify some of the KLF1-dependent genes. This of course might produce false positives (e.g. fast-acting stress response genes) and false negatives (e.g. genes which may be primary targets of KLF1 but are bound by functionally redundant activating TFs), but this could at least provide a list of KLF1-sensitive genes with which to functionally cross-reference the conflicting lists of KLF1 binding sites previously identified by ChIP (**Mukherjee *et al.*, 2022; Huang *et al.*, 2024**). This would have the advantage over previous RNA-profiling experiments comparing KLF1$^{-/-}$ and WT mice (**Hodge *et al.*, 2006; Tallack *et al.*, 2012**) in that acute KLF1 depletion could exclusively identify primary KLF1 target genes, without secondary or tertiary effects of KLF1 knockout.

6.6 – Conclusions

In this thesis, I investigated the action of cis- and trans-acting regulatory factors at the α-globin locus, specifically the 3' CBS in normal erythropoiesis and the transcription factor KLF1 in developmental gene switching of the α-like globin genes. I found that the 3' CBS play a surprisingly minimal role in the formation/maintenance of the α-globin sub-TAD; deletion of the θ1, θ2, HS+44, and HS+48 CBSs led to no significant changes to the enhancer interaction,

chromatin accessibility, or gene expression profiles of the α-globin locus. I suggest, instead, that the α-globin genes themselves may delimit the α-globin sub-TAD. Then, for my investigation of KLF1 in α-like globin switching, I thoroughly characterised a KLF1$^{StrepII\text{-}FKBP}$ mouse model and demonstrated its potential as a powerful tool for studying KLF1 biology. A major difficulty in the study of KLF1 is limited availability of suitable antibodies for chromatin immunoprecipitation (ChIP). In my second chapter, I surprisingly found that the Strep-II tag – an engineered, high-affinity binding partner for streptavidin analogues – was insufficient for robust chipmentation (low-input ChIP) of KLF1$^{StrepII\text{-}FKBP}$. I discuss further modifications to the chipmentation protocol that should be attempted and the intriguing consideration that other feature(s) of KLF1 make it difficult to immunoprecipitate, regardless of antibody or epitope. This could be, for example, if KLF1 is only bound to chromatin – or somehow only "ChIPable" – in a specific subset of terminally differentiating erythroblasts. Finally, I demonstrated the use of the FKBP-V degron tag to quickly and effectively degrade KLF1$^{StrepII\text{-}FKBP}$ protein in primary cultured erythroblasts. This recapitulated many of the expected cellular and gene expression effects that could be expected from acute KLF1 degradation, including modest upregulation of the embryonic α-like globin gene, ζ-globin. The degron feature of the KLF1$^{StrepII\text{-}FKBP}$ model is a particularly novel and powerful tool with potential uses beyond globin switching to investigate the numerous erythroid functions of KLF1.

7 – Appendix

7.1 – Appendix for Chapter 3

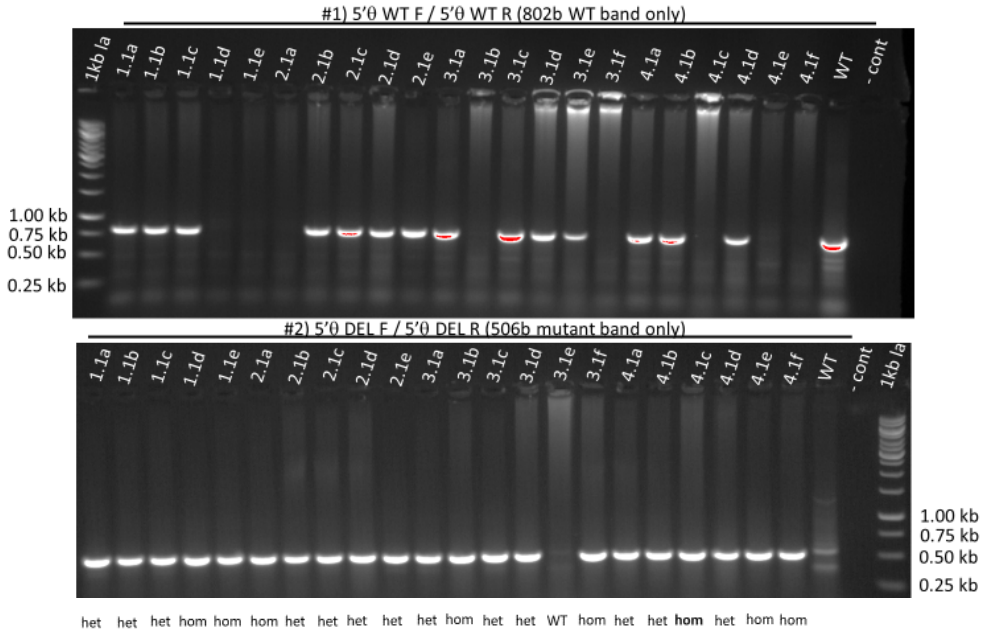

Fig. A3.1 – ***Complete Δθ1-θ2-44-48 genotyping gels: PCR 1 + 2.***

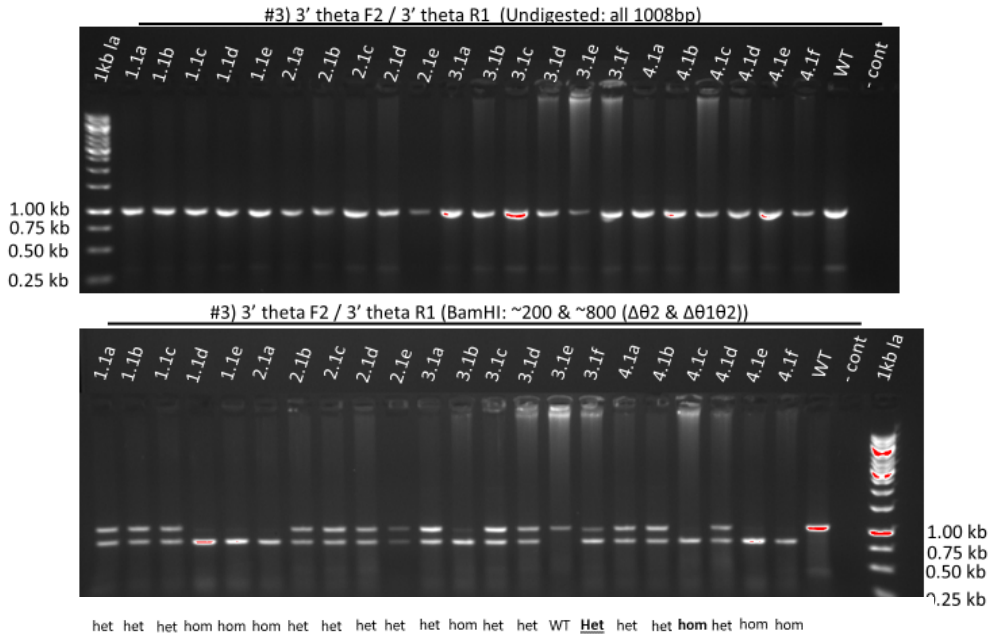

Fig. A3.2 – ***Complete Δθ1-θ2-44-48 genotyping gels: PCR 3.***

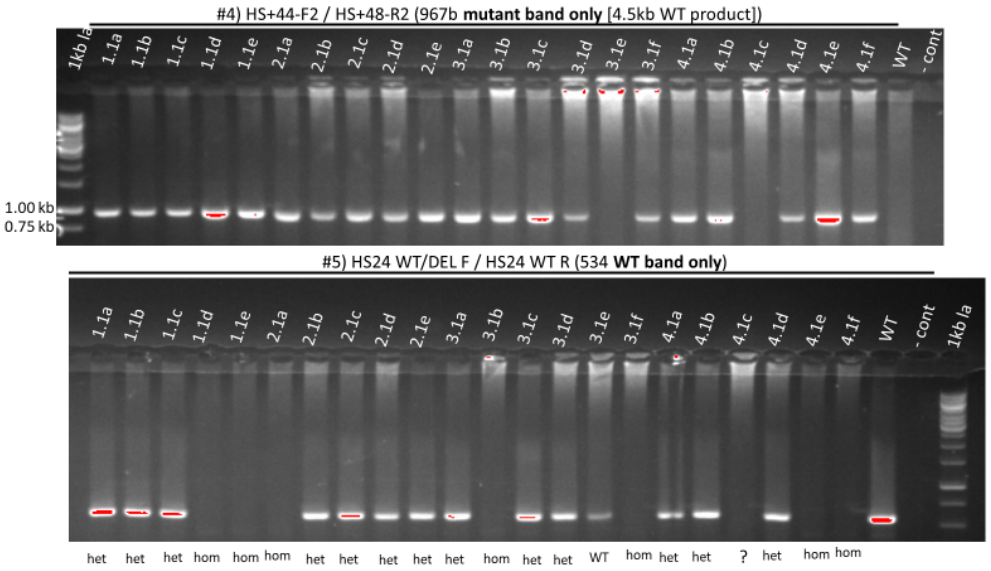

Fig. A3.3 – ***Complete Δθ1-θ2-44-48 genotyping gels: PCR 4 + 5.***

Table A3.1 – *Complete haematology data for Δθ1-θ2-44-48 mouse model*. Anomalies (where blood samples coagulated during collection) are highlighted in red and were excluded from further data analysis.

Animal Name	Zygosity	Sex	Date measured	RBC (%)	HGB (g/dl)	HCT (%)	MCV (μm^3)	MCH (pg)	MCHC (g/dl)	RDW (%)	PLT ($10^3/mm^3$)	MPV (μm^3)
WT1-M	WT	M	2024	**10.41**	**16.1**	48.9	**47**	**15.5**	32.9	16.5	809	6.1
WT2-M	WT	M	2024	**10.8**	**16.6**	50.6	**47**	**15.4**	32.8	16.4	605	6.4
WT3-M	WT	M	2024	**10.46**	**16.1**	48.6	**46**	**15.4**	33.1	16.4	1037	5.7
CTCF16.1a	Hom	M	2024	**11.59**	**19**	55.7	**48**	**16.4**	34.1	15.7	608	5.8
CTCF16.1b	Hom	M	2024	**10.34**	**16.7**	50.7	**49**	**16.1**	32.9	15.7	747	5.5
CTCF16.1c	Hom	M	2024	**10.83**	**17.4**	52.9	**49**	**16.1**	32.9	15.9	668	5.4
WT1-F	WT	F	2024	**10.86**	**17.6**	52	**48**	**16.2**	33.8	15	769	5.6
WT2-F	WT	F	2024	**11.3**	**19.1**	56.8	**50**	**16.9**	33.7	14.3	987	5.3
WT3-F	WT	F	2024	**11.27**	**18.6**	56.1	**50**	**16.5**	33.2	14.8	747	5.5
CTCF16.1e	Hom	F	2024	**10.43**	**16.8**	49.1	**47**	**16.1**	34.2	15.2	773	5.7
CTCF16.1f	Hom	F	2024	**11.27**	**18.2**	53.3	**47**	**16.1**	34.1	15.2	918	5.5
CTCF16.1g	Hom	F	2024	**11.15**	**18**	52.9	**47**	**16.1**	33.9	14.9	790	5.4
CTCF1.1a	Het	M	2020	10.79	17	56.9	53	15.8	30	11.7	645	5.5
CTCF1.1b	Het	M	2020	11.49	17.8	58.6	51	15.5	30.3	11.4	852	5.9
CTCF1.1c	Het	M	2020	11.08	17	56	51	15.3	30.3	11.6	947	5.6
CTCF2.1b	Het	F	2020	10.75	17.1	55.4	51	15.9	30.8	11.7	722	5.1
CTCF2.1c	Het	F	2020	10.62	16.8	54.6	51	15.8	30.7	11.3	700	5.2
CTCF2.1d	Het	F	2020	10.38	16.4	53.6	52	15.8	30.7	11.3	787	5.2
CTCF2.1e	Het	F	2020	10.86	16.8	55.5	51	15.5	30.3	10.9	723	5.5
CTCF3.1c	Het	F	2020	11.05	17.6	57.6	52	15.9	30.5	11.1	682	5.4
CTCF3.1d	Het	F	2020	11.08	17.6	57.3	52	15.9	30.8	10.5	681	5.6

CTCF4.1a	Het	M	2020	10.25	15.9	52.4	51	15.5	30.3	10.6	840	5.4
CTCF4.1b	Het	M	2020	10.31	16.1	52.6	51	15.6	30.6	10.8	671	5.4
CTCF4.1d	Het	M	2020	9.78	15.9	52.7	54	16.2	30.1	12	586	5.4
CTCF1.1d	Hom	F	2020	10.96	17.4	56.5	52	15.9	30.8	11.7	610	5.9
CTCF1.1e	Hom	F	2020	11.03	17.1	56.3	51	15.5	30.3	10.8	751	5.9
CTCF3.1b	Hom	F	2020	10.67	16.7	55.4	52	15.7	30.2	11.1	728	5.2
CTCF3.1f	Hom	F	2020	1.91	2.8	9.4	50	14.9	30.1	11.5	169	5.4
CTCF4.1c	Hom	M	2020	10.48	16.5	54.2	52	15.7	30.4	11.4	756	5.4
CTCF4.1e	Hom	M	2020	1.48	2.2	7.2	48	14.8	30.6	11	147	5.5
CTCF4.1f	Hom	F	2020	6.4	9.8	32.2	50	15.3	30.4	11.2	502	5.4
WT1	WT	F	2020	9.59	15	48.8	51	15.6	30.6	10.4	630	5.1
WT2	WT	F	2020	10.33	16.3	52.9	51	15.8	30.9	10.8	758	5.1
CTCF3.1e	WT	F	2020	10.79	17	56.3	52	15.8	30.2	11	806	5.5

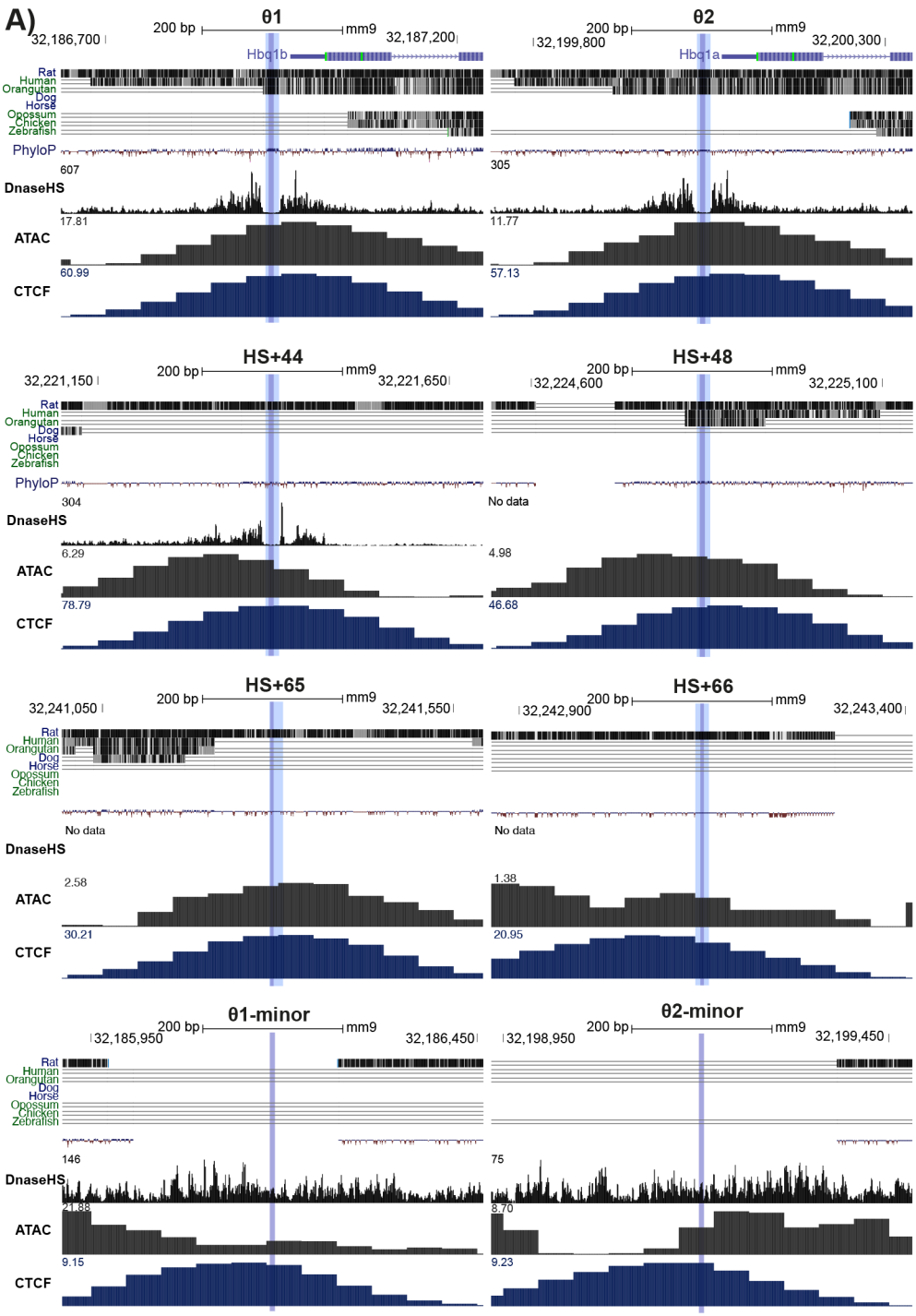

Figure A3.4 – ***UCSC genome browser 600 bp view over the CTCF binding sites referred to in Chapter 3***. Top track is UCSC gene annotation in blue. CTCF binding sites (CBSs) are highlighted vertically in light blue, with the core CBS motif in dark blue. Below this are UCSC Multiz Vertebrate Multi-species Alignments to the reference mm9 mouse genome: no line = no bases in the aligning species; double line = aligning species has one or more unalignable bases in the gap region. UCSC PhyloP conservation scores indicate relative speed of genomic sequence evolution compared to that expected under natural drift, where >0 indicates conservation (slower than expected evolution; blue) and <0 indicates negative selection (faster than expected evolution; red). The final three tracks are DNAse hypersensitivity (**Hanssen *et al.*, 2017**), ATAC-sequencing, and CTCF ChIP-seq in primary adult mouse erythroblasts.

7.2 – Appendix for Chapter 4

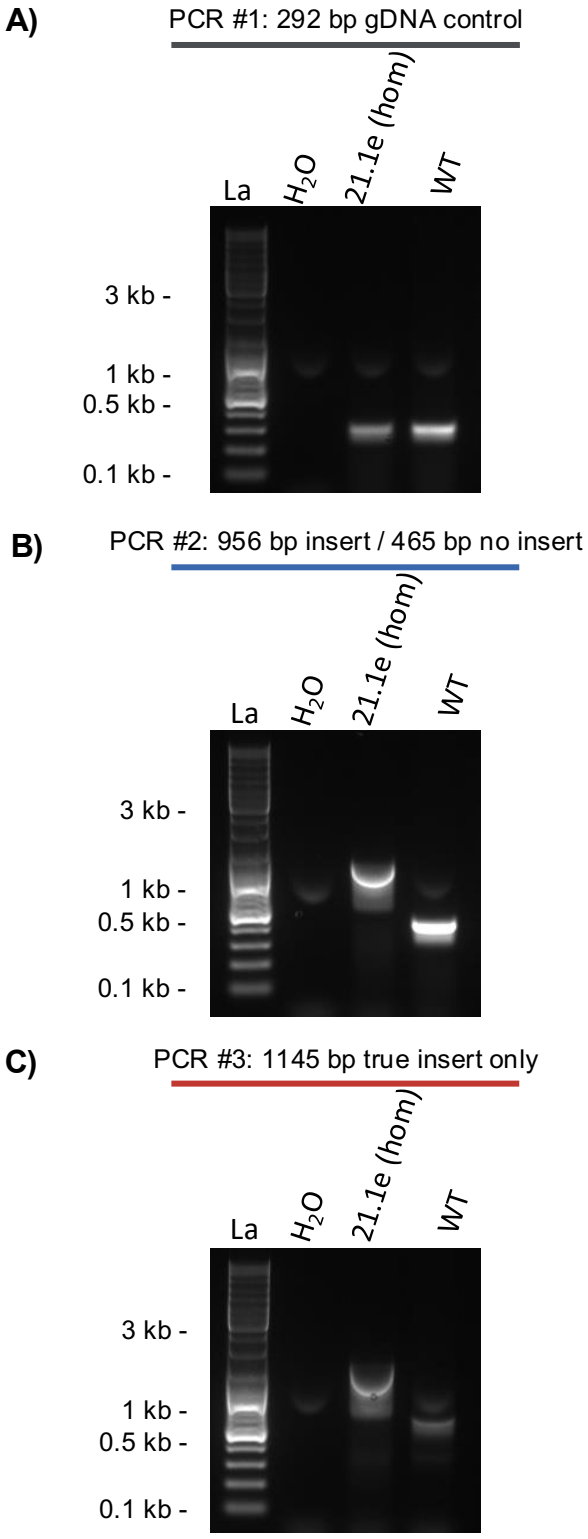

Fig. A4.1 – *Complete KLF1$^{StrepII\text{-}FKBP}$ genotyping gels for WT sample.*

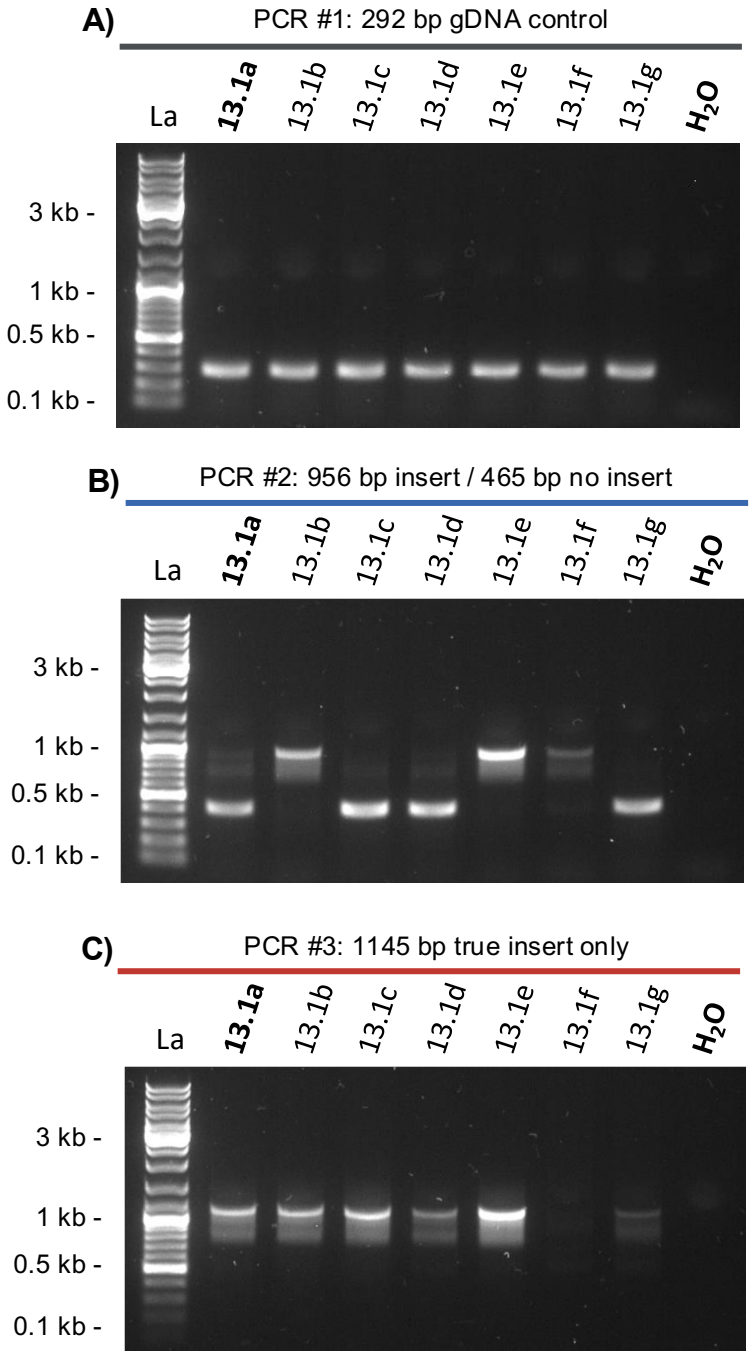

Fig. A4.2 – *Complete KLF1$^{StrepII\text{-}FKBP}$ genotyping gels for het and hom samples.* The samples used in **Fig. 2.3.2** are highlighted in bold.

Table A4.1 – *Complete haematology data for KLF1$^{StrepII\text{-}FKBP}$ mouse model*.

Animal Name	Zygosity	Sex	Date measured	RBC (%)	HGB (g/dl)	HCT (%)	MCV (μm^3)	MCH (pg)	MCHC (g/dl)	RDW (%)	PLT (10^3/mm^3)	MPV (μm^3)
KLF6.3a	WT	M	2023	10.39	16.5	54.7	53	15.9	30.2	11.2	783	5.2
KLF6.3c	WT	M	2023	10.71	16.9	55.7	52	15.8	30.5	11.4	851	5.2
KLF6.3f	WT	M	2023	10.38	16.7	54.3	52	16.1	30.8	11.4	744	5.1
KLF6.3e	Het	M	2023	9.3	14.7	48.2	52	15.8	30.4	11.6	431	5.4
KLF13.1a	Het	M	2023	10.65	17.1	56.4	53	16	30.3	10.5	399	5.2
KLF13.1d	Het	M	2023	10.58	16.3	53.6	51	15.4	30.5	11.6	699	5.5
KLF6.3b	Hom	M	2023	10.52	16.6	54.6	52	15.8	30.5	11.9	630	5.2
KLF11.1b	Hom	M	2023	8.8	14.4	47.4	54	16.4	30.4	12.9	571	5.5
KLF13.1b	Hom	M	2023	11.07	17.9	58.7	53	16.2	30.6	11	732	5.4
WT1-F	WT	F	2023	11.34	17.9	58.8	52	15.8	30.5	10.6	659	5.5
WT2-F	WT	F	2023	11.19	17.4	57.3	51	15.5	30.3	10.9	720	5.4
KLF13.1c	Het	F	2023	9.8	15	50	51	15.3	30.1	11.9	776	5.1
KLF13.1g	Het	F	2023	10.87	17	55.9	51	15.7	30.5	11.8	713	5.5
KLF6.3g	Het	F	2023	10.29	15.7	51.5	50	15.3	30.5	11.9	638	5.2
KLF11.1g	Hom	F	2023	11.19	17.1	56	50	15.3	30.5	12.2	588	5.5
KLF11.1h	Hom	F	2023	10.08	15.6	49.8	49	15.4	31.3	11.9	601	5.4
KLF13.1e	Hom	F	2023	10.92	17	56	51	15.5	30.3	11.8	778	5.2

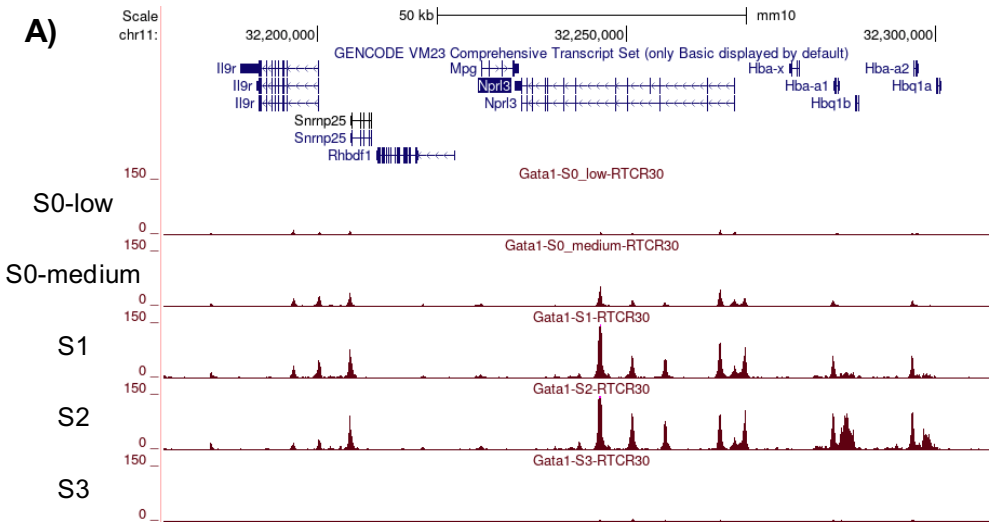

Fig. A4.3 – ***GATA1 binding profile in S0-S3 erythroblasts.*** **A)** Unpublished UCSC tracks from R. Beagrie. CUT&RUN for GATA1 in CD71/Ter119 FACS separated mFL cells.

7.3 – Appendix for Chapter 5

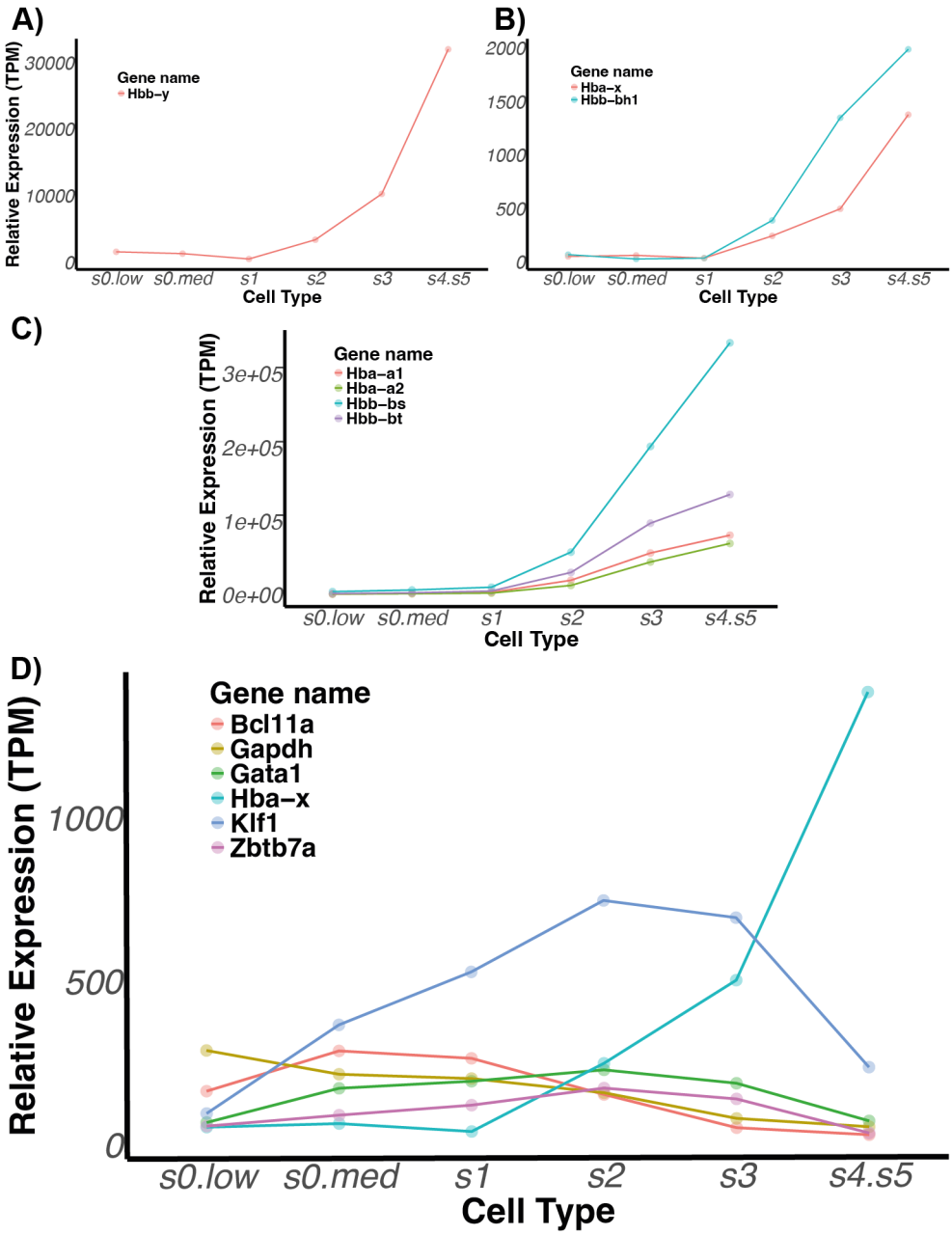

Fig. A5.1 – *Normal expression profile of key erythropoietic genes.* A) Embryonic β-like *Hbb-y*. **B)** Embryonic β-like *Hbb-bh1* and α-like *Hba-x*. **C)** Adult α- and β-like globin genes. **D)** Key erythroid genes discussed in this thesis. *Zbtb7a* = *Lrf*. Relative RNA expression (transcripts per million; TPM) from single-cell mRNA-sequencing data in CD71/Ter119 FACS-sorted E13.5 mFL erythroblast populations. Replotted with permission from **Oudelaar *et al.*, 2020**.

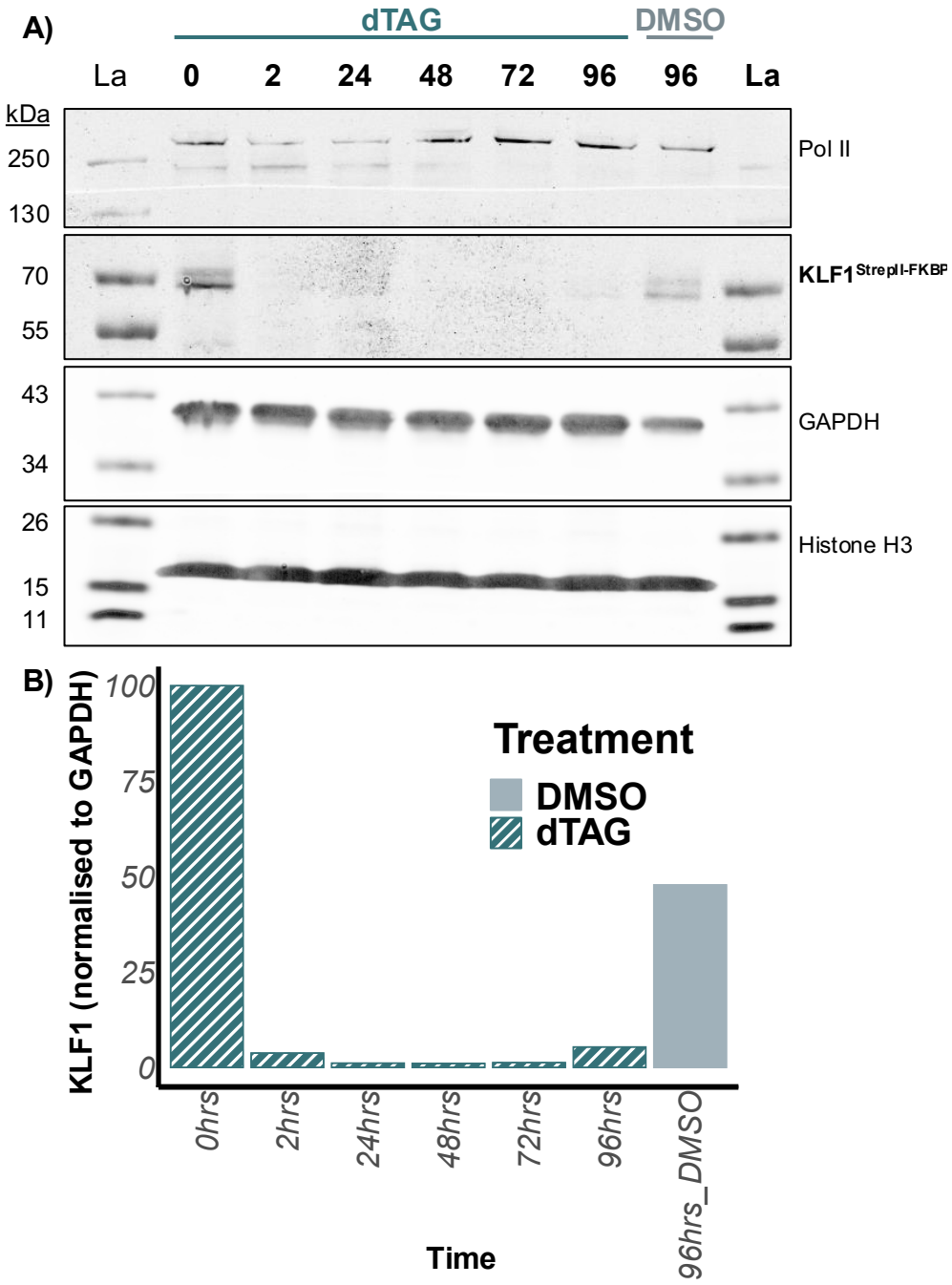

Fig A5.2 – ***Second biological replicate showing KLF1 is quickly, acutely, and persistently depleted by dTAG-13 treatment*****. A)** Western blot timecourse of KLF1$^{StrepII\text{-}FKBP}$ mFL cells cultured for 0-96 hours in expansion media with 0.2 μM dTAG-13 or an equivalent volume DMSO. 3 loading controls were tested: Pol II polymerase, GAPDH, and histone H3. KLF1$^{StrepII\text{-}FKBP}$ was detected with an anti-StrepII antibody (Qiagen, 34850). **B)** Quantification of KLF1$^{StrepII\text{-}FKBP}$ signal, normalised to GAPDH and to 0hrs. Fluorescent western blot is a semi-quantitative technique and should be interpreted with caution.

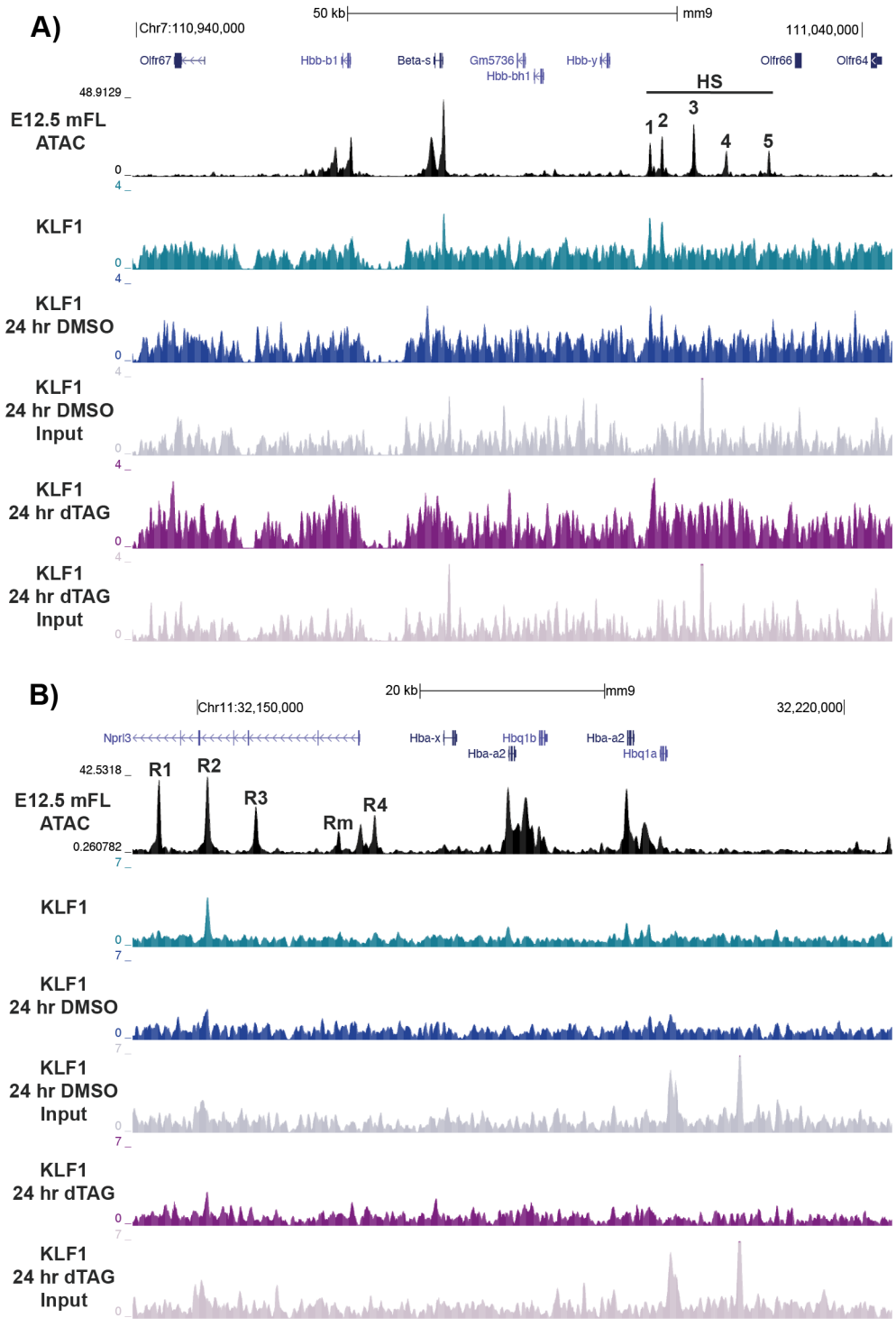

Fig. A5.3 – *Chipmentation for KLF1$^{StrepII\text{-}FKBP}$ after dTAG and DMSO treatment.* Continued on next page.

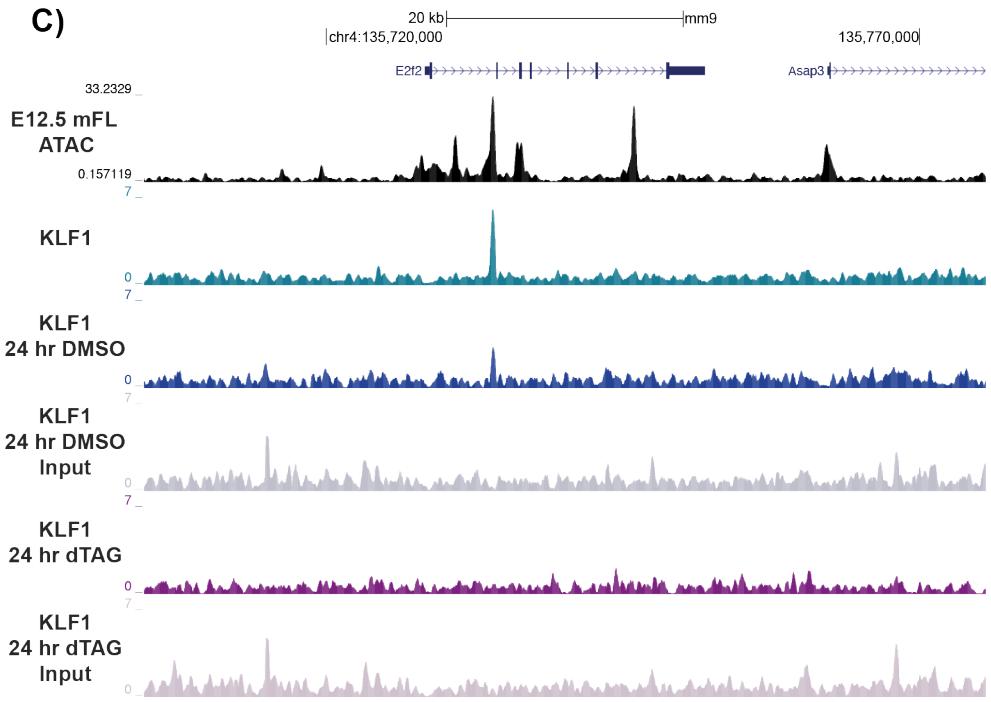

Fig. A5.3 continued – *Chipmentation for KLF1$^{StrepII\text{-}FKBP}$ after dTAG and DMSO treatment.* UCSC tracks for anti-strepII chipmentation in KLF1$^{StrepII\text{-}FKBP}$ E12.5 mFL cells cultured for 24 hours with 0.2 μM dTAG-13 or an equivalent volume DMSO. At the **A)** β-globin locus, **B)** α-globin locus, and **C)** *E2f2* locus. "KLF1" track corresponds to the "Holliman" chipmentation from Chapter 4. N = 3. Input tracks, generated from 5% volume of input sonicated chromatin, are displayed below each treatment track. UCSC gene annotations and scale bars are at the top of each panel. ATAC tracks in E12.5 mFL (**Blayney *et al.*, 2023**; GSE220463) are below in black. The α-globin enhancers are annotated "R1-3, Rm, R4" and the β-globin locus control region "HS1-6".

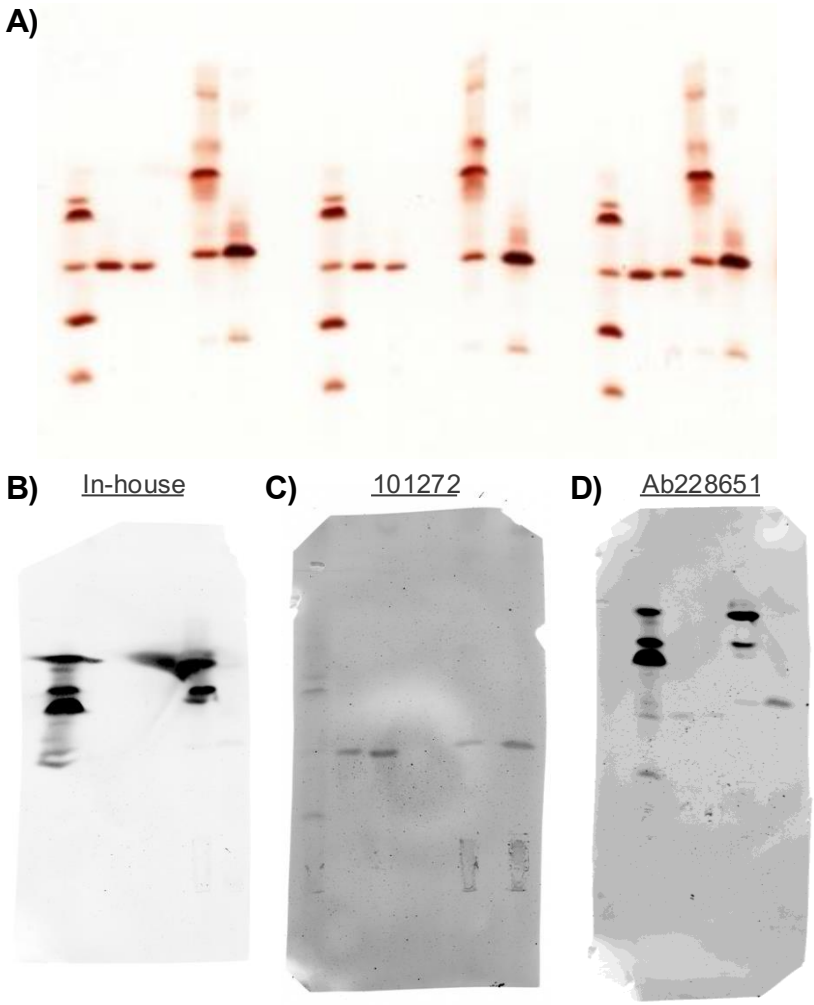

Fig. A5.4 – ***Full images of IEF and western blot for ζ-globin*****.** Full IEF (**A**) and subsequent anti-ζ-globin Western blot (**B-D**) membranes for **Chapter 5 Results Fig. 5.8**. Samples, left to right: E12.5 mouse primitive embryonic blood (ζ-globin positive control); definitive E12.5 mFL cells cultured for 96 hours with DMSO; definitive E12.5 mFL cells cultured for 96 hours with dTAG-13; human BHFS patient "KD" peripheral blood (ζ-globin positive control); human normal control peripheral blood (ζ-globin negative control).

8 – References

Abramson, J. *et al.* (2024) ‘Accurate structure prediction of biomolecular interactions with AlphaFold 3’, *Nature*, 630(8016), pp. 493–500. Available at: https://doi.org/10.1038/s41586-024-07487-w.

Alhashem, Y.N. *et al.* (2011) ‘Transcription Factors KLF1 and KLF2 Positively Regulate Embryonic and Fetal β-Globin Genes through Direct Promoter Binding’, *The Journal of Biological Chemistry*, 286(28), pp. 24819–24827. Available at: https://doi.org/10.1074/jbc.M111.247536.

Amid, A. *et al.* (2024) ‘Hemoglobin Bart’s hydrops fetalis: charting the past and envisioning the future’, *Blood*, 144(8), pp. 822–833. Available at: https://doi.org/10.1182/blood.2023023692.

Andersson, R. and Sandelin, A. (2020) ‘Determinants of enhancer and promoter activities of regulatory elements’, *Nature Reviews Genetics*, 21(2), pp. 71–87. Available at: https://doi.org/10.1038/s41576-019-0173-8.

Armstrong, J.A., Bieker, J.J. and Emerson, B.M. (1998) ‘A SWI/SNF–Related Chromatin Remodeling Complex, E-RC1, Is Required for Tissue-Specific Transcriptional Regulation by EKLF In Vitro’, *Cell*, 95(1), pp. 93–104. Available at: https://doi.org/10.1016/S0092-8674(00)81785-7.

Arnaud, L. *et al.* (2010) ‘A Dominant Mutation in the Gene Encoding the Erythroid Transcription Factor KLF1 Causes a Congenital Dyserythropoietic Anemia’, *The American Journal of Human Genetics*, 87(5), pp. 721–727. Available at: https://doi.org/10.1016/j.ajhg.2010.10.010.

Baghirova, S. *et al.* (2015) ‘Sequential fractionation and isolation of subcellular proteins from tissue or cultured cells’, *MethodsX*, 2, pp. 440–445. Available at: https://doi.org/10.1016/j.mex.2015.11.001.

Bailey, T.L. *et al.* (2015) ‘The MEME Suite’, *Nucleic Acids Research*, 43(W1), pp. W39-49. Available at: https://doi.org/10.1093/nar/gkv416.

Banigan, E.J. *et al.* (2023) ‘Transcription shapes 3D chromatin organization by interacting with loop extrusion’, *Proceedings of the National Academy of Sciences*, 120(11), p. e2210480120. Available at: https://doi.org/10.1073/pnas.2210480120.

Barbarani, G. *et al.* (2019) ‘The Pleiotropic Effects of GATA1 and KLF1 in Physiological Erythropoiesis and in Dyserythropoietic Disorders’, *Frontiers in Physiology*, 10. Available at: https://doi.org/10.3389/fphys.2019.00091.

Basu, P. *et al.* (2007) 'EKLF and KLF2 have compensatory roles in embryonic β-globin gene expression and primitive erythropoiesis', *Blood*, 110(9), pp. 3417–3425. Available at: https://doi.org/10.1182/blood-2006-11-057307.

Bauer, D.E. *et al.* (2013) 'An Erythroid Enhancer of BCL11A Subject to Genetic Variation Determines Fetal Hemoglobin Level', *Science*, 342(6155), pp. 253–257. Available at: https://doi.org/10.1126/science.1242088.

Bieker, J.J. and Philipsen, S. (2024) 'Erythroid Krüppel-Like Factor (KLF1): A Surprisingly Versatile Regulator of Erythroid Differentiation', in T. Borggrefe and B.D. Giaimo (eds) *Transcription factors in blood cell development*. Cham: Springer Nature Switzerland, pp. 217–242. Available at: https://doi.org/10.1007/978-3-031-62731-6_10.

Blayney, J.W. *et al.* (2023) 'Super-enhancers include classical enhancers and facilitators to fully activate gene expression', *Cell*, 186(26), pp. 5826-5839.e18. Available at: https://doi.org/10.1016/j.cell.2023.11.030.

Bonev, B. *et al.* (2017) 'Multiscale 3D Genome Rewiring during Mouse Neural Development', *Cell*, 171(3), pp. 557-572.e24. Available at: https://doi.org/10.1016/j.cell.2017.09.043.

Borg, J. *et al.* (2010) 'Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin', *Nature Genetics*, 42(9), pp. 801–805. Available at: https://doi.org/10.1038/ng.630.

Bower, G. *et al.* (2024) 'Conserved Cis-Acting Range Extender Element Mediates Extreme Long-Range Enhancer Activity in Mammals'. bioRxiv, p. 2024.05.26.595809. Available at: https://doi.org/10.1101/2024.05.26.595809.

Bower, G. and Kvon, E.Z. (2025) 'Genetic factors mediating long-range enhancer–promoter communication in mammalian development', *Current Opinion in Genetics & Development*, 90, p. 102282. Available at: https://doi.org/10.1016/j.gde.2024.102282.

Bradner, J.E. *et al.* (2010) 'Chemical genetic strategy identifies histone deacetylase 1 (HDAC1) and HDAC2 as therapeutic targets in sickle cell disease', *Proceedings of the National Academy of Sciences*, 107(28), pp. 12617–12622. Available at: https://doi.org/10.1073/pnas.1006774107.

Brown, J.M. *et al.* (2018) 'A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions', *Nature Communications*, 9(1), p. 3849. Available at: https://doi.org/10.1038/s41467-018-06248-4.

de Bruijn, M.F.T.R. *et al.* (2000) 'Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo', *The EMBO Journal*, 19(11), pp. 2465–2474. Available at: https://doi.org/10.1093/emboj/19.11.2465.

Buenrostro, J. *et al.* (2015) 'ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide', *Current protocols in molecular biology / edited by*

Frederick M. Ausubel ... [et al.], 109, p. 21.29.1-21.29.9. Available at: https://doi.org/10.1002/0471142727.mb2129s109.

Caria, C.A., Faà, V. and Ristaldi, M.S. (2022) 'Krüppel-Like Factor 1: A Pivotal Gene Regulator in Erythropoiesis', *Cells*, 11(19), p. 3069. Available at: https://doi.org/10.3390/cells11193069.

Chahar, S. *et al.* (2023) 'Transcription induces context-dependent remodeling of chromatin architecture during differentiation', *PLOS Biology*, 21(12), p. e3002424. Available at: https://doi.org/10.1371/journal.pbio.3002424.

Chondrou, V. *et al.* (2020) 'LRF/*ZBTB7A* conservation accentuates its potential as a therapeutic target for the hematopoietic disorders', *Gene*, 760, p. 145020. Available at: https://doi.org/10.1016/j.gene.2020.145020.

Chui, D.H.K. *et al.* (1986) 'Embryonic ζ-Globin Chains in Adults: a Marker for α-Thalassemia-1 Haplotype Due to a >17.5-kb Deletion', *New England Journal of Medicine*, 314(2), pp. 76–79. Available at: https://doi.org/10.1056/NEJM198601093140203.

Chui, D.H.K. *et al.* (1989) 'Human Embryonic ζ-Globin Chains in Fetal and Newborn Blood', *Blood*, 74(4), pp. 1409–1414. Available at: https://doi.org/10.1182/blood.V74.4.1409.1409.

Coghill, E. *et al.* (2001) 'Erythroid Kruppel-like factor (EKLF) coordinates erythroid cell proliferation and hemoglobinization in cell lines derived from EKLF null mice', *Blood*, 97(6), pp. 1861–1868. Available at: https://doi.org/10.1182/blood.V97.6.1861.

Constantinou, C. *et al.* (2019) 'The multi-faceted functioning portrait of LRF/ZBTB7A', *Human Genomics*, 13(1), p. 66. Available at: https://doi.org/10.1186/s40246-019-0252-0.

Corces, M.R. *et al.* (2016) 'Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution', *Nature Genetics*, 48(10), pp. 1193–1203. Available at: https://doi.org/10.1038/ng.3646.

Crossley, M. *et al.* (1994) 'Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1.', *Journal of Biological Chemistry*, 269(22), pp. 15440–15444. Available at: https://doi.org/10.1016/S0021-9258(17)40698-3.

Davies, J.O.J. *et al.* (2016) 'Multiplexed analysis of chromosome conformation at vastly improved sensitivity', *Nature Methods*, 13(1), pp. 74–80. Available at: https://doi.org/10.1038/nmeth.3664.

Dekker, J. *et al.* (2002) 'Capturing Chromosome Conformation', *Science*, 295(5558), pp. 1306–1311. Available at: https://doi.org/10.1126/science.1067799.

Despang, A. *et al.* (2019) 'Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture', *Nature Genetics*, 51(8), pp. 1263–1271. Available at: https://doi.org/10.1038/s41588-019-0466-z.

Dixon, J.R. *et al.* (2012) 'Topological domains in mammalian genomes identified by analysis of chromatin interactions', *Nature*, 485(7398), pp. 376–380. Available at: https://doi.org/10.1038/nature11082.

Donze, D., Townes, T.M. and Bieker, J.J. (1995) 'Role of Erythroid Kruppel-like Factor in Human γ- to β-Globin Gene Switching (∗)', *Journal of Biological Chemistry*, 270(4), pp. 1955–1959. Available at: https://doi.org/10.1074/jbc.270.4.1955.

Drissen, R. *et al.* (2004) 'The active spatial organization of the β-globin locus requires the transcription factor EKLF', *Genes & Development*, 18(20), pp. 2485–2490. Available at: https://doi.org/10.1101/gad.317004.

Drissen, R. *et al.* (2005) 'The Erythroid Phenotype of EKLF-Null Mice: Defects in Hemoglobin Metabolism and Membrane Stability', *Molecular and Cellular Biology*, 25(12), pp. 5205–5214. Available at: https://doi.org/10.1128/MCB.25.12.5205-5214.2005.

Eaton, S.A. *et al.* (2008) 'A network of Krüppel-like Factors (Klfs). Klf8 is repressed by Klf3 and activated by Klf1 in vivo', *The Journal of Biological Chemistry*, 283(40), pp. 26937–26947. Available at: https://doi.org/10.1074/jbc.M804831200.

Esrick, E.B. *et al.* (2021) 'Post-Transcriptional Genetic Silencing of *BCL11A* to Treat Sickle Cell Disease', *New England Journal of Medicine*, 384(3), pp. 205–215. Available at: https://doi.org/10.1056/NEJMoa2029392.

Esteghamat, F. *et al.* (2013) 'Erythropoiesis and globin switching in compound Klf1::Bcl11a mutant mice', *Blood*, 121(13), pp. 2553–2562. Available at: https://doi.org/10.1182/blood-2012-06-434530.

Faure, A.J. *et al.* (2012) 'Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules', *Genome Research*, 22(11), pp. 2163–2175. Available at: https://doi.org/10.1101/gr.136507.111.

Feng, W.C., Southwood, C.M. and Bieker, J.J. (1994) 'Analyses of beta-thalassemia mutant DNA interactions with erythroid Krüppel-like factor (EKLF), an erythroid cell-specific transcription factor.', *Journal of Biological Chemistry*, 269(2), pp. 1493–1500. Available at: https://doi.org/10.1016/S0021-9258(17)42283-6.

Frangoul, H. *et al.* (2024) 'Exagamglogene Autotemcel for Severe Sickle Cell Disease', *The New England Journal of Medicine*, 390(18), pp. 1649–1662. Available at: https://doi.org/10.1056/NEJMoa2309676.

Frontelo, P. *et al.* (2007) 'Novel role for EKLF in megakaryocyte lineage commitment', *Blood*, 110(12), pp. 3871–3880. Available at: https://doi.org/10.1182/blood-2007-03-082065.

Fudenberg, G. *et al.* (2016) 'Formation of Chromosomal Domains by Loop Extrusion', *Cell Reports*, 15(9), pp. 2038–2049. Available at: https://doi.org/10.1016/j.celrep.2016.04.085.

Georgiades, E. *et al.* (2023) 'Active regulatory elements recruit cohesin to establish cell-specific chromatin domains'. bioRxiv, p. 2023.10.13.562171. Available at: https://doi.org/10.1101/2023.10.13.562171.

Gillinder, K.R. *et al.* (2017) 'Promiscuous DNA-binding of a mutant zinc finger protein corrupts the transcriptome and diminishes cell viability', *Nucleic Acids Research*, 45(3), pp. 1130–1143. Available at: https://doi.org/10.1093/nar/gkw1014.

Gillinder, K.R. *et al.* (2018) 'KLF1 Acts As a Pioneer Transcription Factor to Open Chromatin and Facilitate Recruitment of GATA1', *Blood*, 132(Supplement 1), pp. 501–501. Available at: https://doi.org/10.1182/blood-2018-99-119608.

Gnanapragasam, M.N. *et al.* (2016) 'EKLF/KLF1-regulated cell cycle exit is essential for erythroblast enucleation', *Blood*, 128(12), pp. 1631–1641. Available at: https://doi.org/10.1182/blood-2016-03-706671.

Gnanapragasam, M.N. and Bieker, J.J. (2017) 'Orchestration of late events in erythropoiesis by KLF1/EKLF', *Current Opinion in Hematology*, 24(3), p. 183. Available at: https://doi.org/10.1097/MOH.0000000000000327.

Goel, V.Y. and Hansen, A.S. (2021) 'The macro and micro of chromosome conformation capture', *Wiley Interdisciplinary Reviews. Developmental Biology*, 10(6), p. e395. Available at: https://doi.org/10.1002/wdev.395.

Goode, D.K. *et al.* (2016) 'Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation', *Developmental Cell*, 36(5), pp. 572–587. Available at: https://doi.org/10.1016/j.devcel.2016.01.024.

Gregory, R.C. *et al.* (1996) 'Functional Interaction of GATA1 With Erythroid Krüppel-Like Factor and SP1 at Defined Erythroid Promoters', *Blood*, 87(5), pp. 1793–1801. Available at: https://doi.org/10.1182/blood.V87.5.1793.1793.

Guo, Y. *et al.* (2015) 'CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function', *Cell*, 162(4), pp. 900–910. Available at: https://doi.org/10.1016/j.cell.2015.07.038.

Gutiérrez, L. *et al.* (2020) 'Regulation of GATA1 levels in erythropoiesis', *IUBMB Life*, 72(1), pp. 89–105. Available at: https://doi.org/10.1002/iub.2192.

Halfon, M.S. (2019) 'Studying transcriptional enhancers: the founder fallacy, validation creep, and other biases', *Trends in genetics : TIG*, 35(2), pp. 93–103. Available at: https://doi.org/10.1016/j.tig.2018.11.004.

Hansen, A.S. (2020) ‘CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism’, *Nucleus*, 11(1), pp. 132–148. Available at: https://doi.org/10.1080/19491034.2020.1782024.

Hanssen, L.L.P. *et al.* (2017) ‘Tissue-specific CTCF–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo’, *Nature Cell Biology*, 19(8), pp. 952–961. Available at: https://doi.org/10.1038/ncb3573.

Harteveld, C.L. and Higgs, D.R. (2010) ‘α-thalassaemia’, *Orphanet Journal of Rare Diseases*, 5, p. 13. Available at: https://doi.org/10.1186/1750-1172-5-13.

Hay, D. *et al.* (2016) ‘Genetic dissection of the α-globin super-enhancer in vivo’, *Nature Genetics*, 48(8), pp. 895–903. Available at: https://doi.org/10.1038/ng.3605.

He, Z. *et al.* (2014) ‘Structural determinants of human ζ-globin mRNA stability’, *Journal of Hematology & Oncology*, 7, p. 35. Available at: https://doi.org/10.1186/1756-8722-7-35.

He, Z. and Russell, J.E. (2004) ‘Effect of ζ-globin substitution on the O2-transport properties of Hb S in vitro and in vivo’, *Biochemical and Biophysical Research Communications*, 325(4), pp. 1376–1382. Available at: https://doi.org/10.1016/j.bbrc.2004.10.180.

Hentges, L.D. *et al.* (2022) ‘LanceOtron: a deep learning peak caller for genome sequencing experiments’, *Bioinformatics*, 38(18), pp. 4255–4263. Available at: https://doi.org/10.1093/bioinformatics/btac525.

Hodge, D. *et al.* (2006) ‘A global role for EKLF in definitive and primitive erythropoiesis’, *Blood*, 107(8), pp. 3359–3370. Available at: https://doi.org/10.1182/blood-2005-07-2888.

Hua, P. *et al.* (2021) ‘Defining genome architecture at base-pair resolution’, *Nature*, 595(7865), pp. 125–129. Available at: https://doi.org/10.1038/s41586-021-03639-4.

Huang, H. *et al.* (2021) ‘CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains’, *Nature Genetics*, 53(7), pp. 1064–1074. Available at: https://doi.org/10.1038/s41588-021-00863-6.

Huang, S. *et al.* (2024) ‘Mutations in linker-2 of KLF1 impair expression of membrane transporters and cytoskeletal proteins causing hemolysis’, *Nature Communications*, 15(1), p. 7019. Available at: https://doi.org/10.1038/s41467-024-50579-4.

Huehns, E.R. *et al.* (1964) ‘DEVELOPMENTAL HEMOGLOBIN ANOMALIES IN A CHROMOSOMAL TRIPLICATION: D1 TRISOMY SYNDROME*’, *Proceedings of the National Academy of Sciences of the United States of America*, 51(1), pp. 89–97.

Hughes, J.R. *et al.* (2005) 'Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences', *Proceedings of the National Academy of Sciences of the United States of America*, 102(28), pp. 9830–9835. Available at: https://doi.org/10.1073/pnas.0503401102.

Hung, C.-H. *et al.* (2021) 'A Positive Regulatory Feedback Loop between EKLF/KLF1 and TAL1/SCL Sustaining the Erythropoiesis', *International Journal of Molecular Sciences*, 22(15), p. 8024. Available at: https://doi.org/10.3390/ijms22158024.

Hyle, J. *et al.* (2019) 'Acute depletion of CTCF directly affects MYC regulation through loss of enhancer–promoter looping', *Nucleic Acids Research*, 47(13), pp. 6699–6713. Available at: https://doi.org/10.1093/nar/gkz462.

Ilsley, M.D. *et al.* (2017) 'Krüppel-like factors compete for promoters and enhancers to fine-tune transcription', *Nucleic Acids Research*, 45(11), pp. 6572–6588. Available at: https://doi.org/10.1093/nar/gkx441.

Jeppsson, K. *et al.* (2022) 'Cohesin-dependent chromosome loop extrusion is limited by transcription and stalled replication forks', *Science Advances*, 8(23), p. eabn7063. Available at: https://doi.org/10.1126/sciadv.abn7063.

Jia, Z. *et al.* (2020) 'Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection', *Genome Biology*, 21, p. 75. Available at: https://doi.org/10.1186/s13059-020-01984-7.

Kagey, M.H. *et al.* (2010) 'Mediator and cohesin connect gene expression and chromatin architecture', *Nature*, 467(7314), pp. 430–435. Available at: https://doi.org/10.1038/nature09380.

Kassouf, M.T. *et al.* (2022) *Multipartite super-enhancers function in an orientation-dependent manner*. preprint. Molecular Biology. Available at: https://doi.org/10.1101/2022.07.14.499999.

Kent, W.J. *et al.* (2002) 'The human genome browser at UCSC', *Genome Research*, 12(6), pp. 996–1006. Available at: https://doi.org/10.1101/gr.229102.

Kentepozidou, E. *et al.* (2020) 'Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains', *Genome Biology*, 21(1), p. 5. Available at: https://doi.org/10.1186/s13059-019-1894-x.

Kiledjian, M., Wang, X. and Liebhaber, S.A. (1995) 'Identification of two KH domain proteins in the alpha-globin mRNP stability complex.', *The EMBO Journal*, 14(17), pp. 4357–4364.

King, A.J. *et al.* (2021) 'Reactivation of a developmentally silenced embryonic globin gene', *Nature Communications*, 12(1), p. 4439. Available at: https://doi.org/10.1038/s41467-021-24402-3.

Kingsley, P.D. *et al.* (2004) 'Yolk sac–derived primitive erythroblasts enucleate during mammalian embryogenesis', *Blood*, 104(1), pp. 19–25. Available at: https://doi.org/10.1182/blood-2003-12-4162.

Kingsley, P.D. *et al.* (2006) '"Maturational" globin switching in primary primitive erythroid cells', *Blood*, 107(4), pp. 1665–1672. Available at: https://doi.org/10.1182/blood-2005-08-3097.

Lafferty, J.D. *et al.* (2000) 'A Reliable Screening Test to Identify Adult Carriers of the (--SEA) alpha0-Thalassemia Deletion', *American Journal of Clinical Pathology*, 114(6), pp. 927–931. Available at: https://doi.org/10.1309/26G7-BQH4-93BV-UR0Q.

Langmead, B. *et al.* (2009) 'Ultrafast and memory-efficient alignment of short DNA sequences to the human genome', *Genome Biology*, 10(3), p. R25. Available at: https://doi.org/10.1186/gb-2009-10-3-r25.

Layden, H.M. *et al.* (2021) 'A protocol for rapid degradation of endogenous transcription factors in mammalian cells and identification of direct regulatory targets', *STAR Protocols*, 2(2), p. 100530. Available at: https://doi.org/10.1016/j.xpro.2021.100530.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics (Oxford, England)*, 25(16), pp. 2078–2079. Available at: https://doi.org/10.1093/bioinformatics/btp352.

Lieberman-Aiden, E. *et al.* (2009) 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science (New York, N.Y.)*, 326(5950), pp. 289–293. Available at: https://doi.org/10.1126/science.1181369.

Liebhaber, S.A. *et al.* (1996) 'Developmental Silencing of the Embryonic ζ-Globin Gene: Concerted Action of the Promoter and the 3′-Flanking Region Combined with Stage-Specific Silencing by the Transcribed Segment', *Molecular and Cellular Biology*, 16(6), pp. 2637–2646. Available at: https://doi.org/10.1128/MCB.16.6.2637.

Liu, G. *et al.* (2024) 'Long noncoding RNA *GATA2AS* influences human erythropoiesis by transcription factor and chromatin landscape modulation', *Blood*, 143(22), pp. 2300–2313. Available at: https://doi.org/10.1182/blood.2023021287.

Liu, N. *et al.* (2018) 'Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch', *Cell*, 173(2), pp. 430-442.e17. Available at: https://doi.org/10.1016/j.cell.2018.03.016.

Locatelli, F. *et al.* (2024) 'Exagamglogene Autotemcel for Transfusion-Dependent β-Thalassemia', *New England Journal of Medicine*, 390(18), pp. 1663–1676. Available at: https://doi.org/10.1056/NEJMoa2309673.

Ludwig, L.S. *et al.* (2019) 'Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis', *Cell Reports*, 27(11), pp. 3228-3240.e7. Available at: https://doi.org/10.1016/j.celrep.2019.05.046.

Luo, H. *et al.* (1988) 'A Novel Monoclonal Antibody Based Diagnostic Test for α-Thalassemia-1 Carriers Due to the (—SEA/) Deletion', *Blood*, 72(5), pp. 1589–1594. Available at: https://doi.org/10.1182/blood.V72.5.1589.1589.

Luo, Q. *et al.* (2004) 'Activation and Repression of Interleukin-12 p40 Transcription by Erythroid Kruppel-like Factor in Macrophages *', *Journal of Biological Chemistry*, 279(18), pp. 18451–18456. Available at: https://doi.org/10.1074/jbc.M400320200.

Magoč, T. and Salzberg, S.L. (2011) 'FLASH: fast length adjustment of short reads to improve genome assemblies', *Bioinformatics*, 27(21), pp. 2957–2963. Available at: https://doi.org/10.1093/bioinformatics/btr507.

Magor, G.W. *et al.* (2015) 'KLF1-null neonates display hydrops fetalis and a deranged erythroid transcriptome', *Blood*, 125(15), pp. 2405–2417. Available at: https://doi.org/10.1182/blood-2014-08-590968.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), pp. 10–12. Available at: https://doi.org/10.14806/ej.17.1.200.

Martyn, G.E. *et al.* (2018) 'Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding', *Nature Genetics*, 50(4), pp. 498–503. Available at: https://doi.org/10.1038/s41588-018-0085-0.

Mas, C. *et al.* (2011) 'Structural and functional characterization of an atypical activation domain in erythroid Kruppel-like factor (EKLF)', *Proceedings of the National Academy of Sciences of the United States of America*, 108(26), pp. 10484–10489. Available at: https://doi.org/10.1073/pnas.1017029108.

Masuda, T. *et al.* (2016) 'Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin', *Science*, 351(6270), pp. 285–289. Available at: https://doi.org/10.1126/science.aad3312.

Mehta, S. *et al.* (2022) 'Temporal resolution of gene de-repression and proteome changes upon PROTAC-mediated degradation of BCL11A protein in erythroid cells', *Cell Chemical Biology*, 29(8), pp. 1273-1287.e8. Available at: https://doi.org/10.1016/j.chembiol.2022.06.007.

Miller, I.J. and Bieker, J.J. (1993) 'A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins.', *Molecular and Cellular Biology*, 13(5), pp. 2776–2786.

Moore, J.E. *et al.* (2020) 'Expanded encyclopaedias of DNA elements in the human and mouse genomes', *Nature*, 583(7818), pp. 699–710. Available at: https://doi.org/10.1038/s41586-020-2493-4.

Mukherjee, K. and Bieker, J.J. (2022) 'EKLF/Klf1 regulates erythroid transcription by its pioneering activity and selective control of RNA Pol II pause-release', *Cell Reports*, 41(12), p. 111830. Available at: https://doi.org/10.1016/j.celrep.2022.111830.

Nabet, B. *et al.* (2018) 'The dTAG system for immediate and target-specific protein degradation', *Nature Chemical Biology*, 14(5), pp. 431–441. Available at: https://doi.org/10.1038/s41589-018-0021-8.

Nébor, D. *et al.* (2018) 'Mutant KLF1 in Adult Anemic Nan Mice Leads to Profound Transcriptome Changes and Disordered Erythropoiesis', *Scientific Reports*, 8(1), p. 12793. Available at: https://doi.org/10.1038/s41598-018-30839-2.

Nilson, D.G. *et al.* (2006) 'Major erythrocyte membrane protein genes in EKLF-deficient mice', *Experimental Hematology*, 34(6), pp. 705–712. Available at: https://doi.org/10.1016/j.exphem.2006.02.018.

Nora, E.P. *et al.* (2012) 'Spatial partitioning of the regulatory landscape of the X-inactivation centre', *Nature*, 485(7398), pp. 381–385. Available at: https://doi.org/10.1038/nature11049.

Norton, L.J. *et al.* (2017) 'KLF1 directly activates expression of the novel fetal globin repressor ZBTB7A/LRF in erythroid cells', *Blood Advances*, 1(11), pp. 685–692. Available at: https://doi.org/10.1182/bloodadvances.2016002303.

Nuez, B. *et al.* (1995) 'Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene', *Nature*, 375(6529), pp. 316–318. Available at: https://doi.org/10.1038/375316a0.

O'Connell, K.E. *et al.* (2015) 'Practical Murine Hematopathology: A Comparative Review and Implications for Research', *Comparative Medicine*, 65(2), pp. 96–113.

Orkin, S.H. and Zon, L.I. (2008) 'Hematopoiesis: An Evolving Paradigm for Stem Cell Biology', *Cell*, 132(4), pp. 631–644. Available at: https://doi.org/10.1016/j.cell.2008.01.025.

Oudelaar, A.M. *et al.* (2020) 'Dynamics of the 4D genome during in vivo lineage specification and differentiation', *Nature Communications*, 11(1), p. 2722. Available at: https://doi.org/10.1038/s41467-020-16598-7.

Ouyang, L., Chen, X. and Bieker, J.J. (1998) 'Regulation of Erythroid Krüppel-like Factor (EKLF) Transcriptional Activity by Phosphorylation of a Protein Kinase Casein Kinase II Site within Its Interaction Domain *', *Journal of Biological Chemistry*, 273(36), pp. 23019–23025. Available at: https://doi.org/10.1074/jbc.273.36.23019.

Palii, C.G. *et al.* (2019) 'Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate', *Cell Stem Cell*, 24(5), pp. 812-820.e5. Available at: https://doi.org/10.1016/j.stem.2019.02.006.

Palis, J. *et al.* (1999) 'Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse', *Development*, 126(22), pp. 5073–5084. Available at: https://doi.org/10.1242/dev.126.22.5073.

Palis, J. (2014) 'Primitive and definitive erythropoiesis in mammals', *Frontiers in Physiology*, 5. Available at: https://doi.org/10.3389/fphys.2014.00003.

Parkins, A.C., Sharpe, A.H. and Orkin, S.H. (1995) 'Lethal β-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF', *Nature*, 375(6529), pp. 318–322. Available at: https://doi.org/10.1038/375318a0.

Perkins, A. *et al.* (2016) 'Krüppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants', *Blood*, 127(15), pp. 1856–1862. Available at: https://doi.org/10.1182/blood-2016-01-694331.

Perkins, A.C. *et al.* (2000) 'Fetal expression of a human Aγ globin transgene rescues globin chain imbalance but not hemolysis in EKLF null mouse embryos', *Blood*, 95(5), pp. 1827–1833. Available at: https://doi.org/10.1182/blood.V95.5.1827.004k10_1827_1833.

Perkins, A.C., Gaensler, K.M. and Orkin, S.H. (1996) 'Silencing of human fetal globin expression is impaired in the absence of the adult beta-globin gene activator protein EKLF', *Proceedings of the National Academy of Sciences of the United States of America*, 93(22), pp. 12267–12271. Available at: https://doi.org/10.1073/pnas.93.22.12267.

Peron, A. *et al.* (2024) 'BCL11A intellectual developmental disorder: defining the clinical spectrum and genotype-phenotype correlations', *European Journal of Human Genetics*, pp. 1–13. Available at: https://doi.org/10.1038/s41431-024-01701-z.

Philipsen, S., Pruzina, S. and Grosveld, F. (no date) 'The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the a globin locus control region'.

Phillips-Cremins, J.E. *et al.* (2013) 'Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment', *Cell*, 153(6), pp. 1281–1295. Available at: https://doi.org/10.1016/j.cell.2013.04.053.

Piel, F.B. and Weatherall, D.J. (2014) 'The α-Thalassemias', *New England Journal of Medicine*. Edited by D.L. Longo, 371(20), pp. 1908–1916. Available at: https://doi.org/10.1056/NEJMra1404415.

Pilon, A.M. *et al.* (2008) 'Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2', *Molecular and Cellular Biology*, 28(24), pp. 7394–7401. Available at: https://doi.org/10.1128/MCB.01087-08.

Pilon, A.M. *et al.* (2011) 'Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte

differentiation', *Blood*, 118(17), pp. e139-148. Available at: https://doi.org/10.1182/blood-2011-05-355107.

Pop, R. *et al.* (2010) 'A Key Commitment Step in Erythropoiesis Is Synchronized with the Cell Cycle Clock through Mutual Inhibition between PU.1 and S-Phase Progression', *PLoS Biology*, 8(9), p. e1000484. Available at: https://doi.org/10.1371/journal.pbio.1000484.

Porcu, S. *et al.* (2011) 'Klf1 Affects DNase II-Alpha Expression in the Central Macrophage of a Fetal Liver Erythroblastic Island: a Non-Cell-Autonomous Role in Definitive Erythropoiesis ▿ ', *Molecular and Cellular Biology*, 31(19), pp. 4144–4154. Available at: https://doi.org/10.1128/MCB.05532-11.

Quadrini, K.J. and Bieker, J.J. (2006) 'EKLF/KLF1 is ubiquitinated in vivo and its stability is regulated by activation domain sequences through the 26S proteasome', *FEBS Letters*, 580(9), pp. 2285–2293. Available at: https://doi.org/10.1016/j.febslet.2006.03.039.

Quadrini, K.J., Gruzglin, E. and Bieker, J.J. (2008) 'Non-random subcellular distribution of variant EKLF in erythroid cells', *Experimental cell research*, 314(7), pp. 1595–1604. Available at: https://doi.org/10.1016/j.yexcr.2008.01.033.

Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841–842. Available at: https://doi.org/10.1093/bioinformatics/btq033.

Raabe, B.M. *et al.* (2011) 'Effects of Weekly Blood Collection in C57BL/6 Mice', *Journal of the American Association for Laboratory Animal Science : JAALAS*, 50(5), pp. 680–685.

Rai, D., Wilson, A.M. and Moosavi, L. (2024) 'Histology, Reticulocytes', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK542172/ (Accessed: 26 August 2024).

Ramírez, F. *et al.* (2016) 'deepTools2: a next generation web server for deep-sequencing data analysis', *Nucleic Acids Research*, 44(W1), pp. W160-165. Available at: https://doi.org/10.1093/nar/gkw257.

Rao, S.S.P. *et al.* (2014) 'A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping', *Cell*, 159(7), pp. 1665–1680. Available at: https://doi.org/10.1016/j.cell.2014.11.021.

Rao, S.S.P. *et al.* (2017) 'Cohesin Loss Eliminates All Loop Domains', *Cell*, 171(2), pp. 305-320.e24. Available at: https://doi.org/10.1016/j.cell.2017.09.026.

Rinzema, N.J. *et al.* (2022) 'Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes', *Nature Structural & Molecular Biology*, 29(6), pp. 563–574. Available at: https://doi.org/10.1038/s41594-022-00787-7.

Russell, J.E. *et al.* (1998) 'Sequence Divergence in the 3′ Untranslated Regions of Human ζ- and α-Globin mRNAs Mediates a Difference in Their Stabilities and Contributes to Efficient α-to-ζ Gene Developmental Switching', *Molecular and Cellular Biology*, 18(4), pp. 2173–2183.

Russell, J.E., Lee, A.E. and Liebhaber, S.A. (1998) 'Full Developmental Silencing of the Embryonic ζ-Globin Gene Reflects Instability of its mRNA', *Annals of the New York Academy of Sciences*, 850(1), pp. 386–390. Available at: https://doi.org/10.1111/j.1749-6632.1998.tb10499.x.

Russell, J.E. and Liebhaber, S.A. (1998) 'Reversal of lethal alpha- and beta-thalassemias in mice by expression of human embryonic globins', *Blood*, 92(9), pp. 3057–3063.

Sabath, D.E. *et al.* (1993) 'Analysis of the Human *ζ*-Globin Gene Promoter in Transgenic Mice', *Blood*, 82(9), pp. 2899–2905. Available at: https://doi.org/10.1182/blood.V82.9.2899.2899.

Sankaran, V.G. *et al.* (2008) 'Human Fetal Hemoglobin Expression Is Regulated by the Developmental Stage-Specific Repressor *BCL11A*', *Science*, 322(5909), pp. 1839–1842. Available at: https://doi.org/10.1126/science.1165409.

Santos, G.P. dos *et al.* (2024) 'Transcriptional regulators of fetal hemoglobin', *Hematology, Transfusion and Cell Therapy*, 46, pp. S258–S268. Available at: https://doi.org/10.1016/j.htct.2024.06.001.

Satterwhite, E. *et al.* (2001) 'The BCL11 gene family: involvement of BCL11A in lymphoid malignancies', *Blood*, 98(12), pp. 3413–3420. Available at: https://doi.org/10.1182/blood.V98.12.3413.

Schmidt, T.G.M. *et al.* (2013) 'Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants', *Protein Expression and Purification*, 92(1), pp. 54–61. Available at: https://doi.org/10.1016/j.pep.2013.08.021.

Schwarzer, W. *et al.* (2017) 'Two independent modes of chromatin organization revealed by cohesin removal', *Nature*, 551(7678), pp. 51–56. Available at: https://doi.org/10.1038/nature24281.

Sengupta, T. *et al.* (2009) 'Distinct modes of gene regulation by a cell-specific transcriptional activator', *Proceedings of the National Academy of Sciences of the United States of America*, 106(11), pp. 4213–4218. Available at: https://doi.org/10.1073/pnas.0808347106.

Sexton, T. *et al.* (2012) 'Three-dimensional folding and functional organization principles of the Drosophila genome', *Cell*, 148(3), pp. 458–472. Available at: https://doi.org/10.1016/j.cell.2012.01.010.

Shah, M. *et al.* (2019) 'Hit and Run Transcriptional Repressors Are Difficult to Catch in the Act', *BioEssays*, 41(8), p. 1900041. Available at: https://doi.org/10.1002/bies.201900041.

Siatecka, M., Lohmann, F., *et al.* (2010) 'EKLF Directly Activates the p21WAF1/CIP1 Gene by Proximal Promoter and Novel Intronic Regulatory Regions during Erythroid Differentiation', *Molecular and Cellular Biology*, 30(11), pp. 2811–2822. Available at: https://doi.org/10.1128/MCB.01016-09.

Siatecka, M., Sahr, K.E., *et al.* (2010) 'Severe anemia in the Nan mutant mouse caused by sequence-selective disruption of erythroid Kruppel-like factor', *Proceedings of the National Academy of Sciences of the United States of America*, 107(34), pp. 15151–15156. Available at: https://doi.org/10.1073/pnas.1004996107.

Siatecka, M. *et al.* (2015) 'Transcriptional Activity of Erythroid Kruppel-like Factor (EKLF/KLF1) Modulated by PIAS3 (Protein Inhibitor of Activated STAT3)', *The Journal of Biological Chemistry*, 290(15), pp. 9929–9940. Available at: https://doi.org/10.1074/jbc.M114.610246.

Siatecka, M. and Bieker, J.J. (2011) 'The multifunctional role of EKLF/KLF1 during erythropoiesis', *Blood*, 118(8), pp. 2044–2054. Available at: https://doi.org/10.1182/blood-2011-03-331371.

Siatecka, M., Xue, L. and Bieker, J.J. (2007) 'Sumoylation of EKLF Promotes Transcriptional Repression and Is Involved in Inhibition of Megakaryopoiesis', *Molecular and Cellular Biology*, 27(24), pp. 8547–8560. Available at: https://doi.org/10.1128/MCB.00589-07.

Songdej, D. *et al.* (2017) 'An international registry of survivors with Hb Bart's hydrops fetalis syndrome', *Blood*, 129(10), pp. 1251–1259. Available at: https://doi.org/10.1182/blood-2016-08-697110.

Southwood, C.M., Downs, K.M. and Bieker, J.J. (1996) 'Erythroid Krüppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny', *Developmental Dynamics*, 206(3), pp. 248–259. Available at: https://doi.org/10.1002/(SICI)1097-0177(199607)206:3<248::AID-AJA3>3.0.CO;2-I.

Spangler, E.A., Andrews, K.A. and Rubin, E.M. (1990) 'Developmental regulation of the human zeta globin gene in transgenic mice', *Nucleic Acids Research*, 18(23), pp. 7093–7097. Available at: https://doi.org/10.1093/nar/18.23.7093.

Stolper, R.J. *et al.* (2023) *Loop extrusion by cohesin plays a key role in enhancer-activated gene expression during differentiation*. preprint. Molecular Biology. Available at: https://doi.org/10.1101/2023.09.07.556660.

Su, M.Y. *et al.* (2013) 'Identification of Biologically Relevant Enhancers in Human Erythroid Cells', *The Journal of Biological Chemistry*, 288(12), pp. 8433–8444. Available at: https://doi.org/10.1074/jbc.M112.413260.

Tallack, M.R. *et al.* (2009) 'EKLF/KLF1 Controls Cell Cycle Entry via Direct Regulation of E2f2', *The Journal of Biological Chemistry*, 284(31), p. 20966. Available at: https://doi.org/10.1074/jbc.M109.006346.

Tallack, M.R. *et al.* (2010) 'A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells', *Genome Research*, 20(8), pp. 1052–1063. Available at: https://doi.org/10.1101/gr.106575.110.

Tallack, M.R. *et al.* (2012) 'Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq', *Genome Research*, 22(12), pp. 2385–2398. Available at: https://doi.org/10.1101/gr.135707.111.

Tallack, M.R. and Perkins, A.C. (2013) 'Three fingers on the switch: Krüppel-like factor 1 regulation of: γ-globin: to: β-globin: gene switching', *Current Opinion in Hematology*, 20(3), p. 193. Available at: https://doi.org/10.1097/MOH.0b013e32835f59ba.

Tang, W. *et al.* (1992) 'Human embryonic zeta-globin chain expression in deletional alpha-thalassemias', *Blood*, 80(2), pp. 517–522.

Tang, W. *et al.* (1993) 'Expression of Embryonic *ζ*-Globin and *ϵ*-Globin Chains in a 10-Year-Old Girl With Congenital Anemia', *Blood*, 81(6), pp. 1636–1640. Available at: https://doi.org/10.1182/blood.V81.6.1636.1636.

Thompson B., Wickens M., Kimble J. (2007) in Translational Control in Biology and Medicine (Mathews M. B., Sonenberg N., Hershey J. W. B., eds) pp. 507–544, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York [Google Scholar]

Tsang, F.H. *et al.* (2024) 'The characteristics of CTCF binding sequences contribute to enhancer blocking activity', *Nucleic Acids Research*, 52(17), pp. 10180–10193. Available at: https://doi.org/10.1093/nar/gkae666.

Two brothers with a novel KLF1 mutation | Blood | American Society of Hematology (no date). Available at: https://ashpublications.org/blood/article/138/12/1087/476923/Two-brothers-with-a-novel-KLF1-mutation (Accessed: 14 January 2025).

Vernimmen, D. *et al.* (2007) 'Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression', *The EMBO Journal*, 26(8), pp. 2041–2051. Available at: https://doi.org/10.1038/sj.emboj.7601654.

Vietri Rudan, M. *et al.* (2015) 'Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture', *Cell Reports*, 10(8), pp. 1297–1309. Available at: https://doi.org/10.1016/j.celrep.2015.02.004.

Viprakasit, V. *et al.* (2014) 'Mutations in Krüppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression', *Blood*, 123(10), pp. 1586–1595. Available at: https://doi.org/10.1182/blood-2013-09-526087.

Vyas, P. *et al.* (1992) 'Regulation of human embryonic globin genes zeta 2 and epsilon in stably transformed mouse erythroleukemia cells', *Blood*, 80(7), pp. 1832–1837.

Wajcman, H. and Galactéros, F. (1996) 'Abnormal hemoglobins with high oxygen affinity and erythrocytosis', *Hematology and Cell Therapy*, 38(4), pp. 305–312. Available at: https://doi.org/10.1007/s00282-996-0305-4.

Whyte, W.A. *et al.* (2013) 'Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes', *Cell*, 153(2), pp. 307–319. Available at: https://doi.org/10.1016/j.cell.2013.03.035.

Xu, J. *et al.* (2011) 'Correction of Sickle Cell Disease in Adult Mice by Interference with Fetal Hemoglobin Silencing', *Science*, 334(6058), pp. 993–996. Available at: https://doi.org/10.1126/science.1211053.

Xue, L. *et al.* (2014) 'Extrinsic and intrinsic control by EKLF (KLF1) within a specialized erythroid niche', *Development (Cambridge, England)*, 141(11), pp. 2245–2254. Available at: https://doi.org/10.1242/dev.103960.

Yien, Y.Y. and Bieker, J.J. (2013) 'EKLF/KLF1, a Tissue-Restricted Integrator of Transcriptional Control, Chromatin Remodeling, and Lineage Determination', *Molecular and Cellular Biology*, 33(1), pp. 4–13. Available at: https://doi.org/10.1128/MCB.01058-12.

Zerbino, D.R. *et al.* (2014) 'WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis', *Bioinformatics*, 30(7), pp. 1008–1009. Available at: https://doi.org/10.1093/bioinformatics/btt737.

Zhang, H. *et al.* (2023) 'CTCF and R-loops are boundaries of cohesin-mediated DNA looping', *Molecular Cell*, 83(16), pp. 2856-2871.e8. Available at: https://doi.org/10.1016/j.molcel.2023.07.006.

Zhang, J. *et al.* (2003) 'Role of Ras signaling in erythroid differentiation of mouse fetal liver cells: functional analysis by a flow cytometry-based novel culture system', *Blood*, 102(12), pp. 3938–3946. Available at: https://doi.org/10.1182/blood-2003-05-1479.

Zhang, S. *et al.* (2023) 'Enhancer–promoter contact formation requires RNAPII and antagonizes loop extrusion', *Nature Genetics*, pp. 1–9. Available at: https://doi.org/10.1038/s41588-023-01364-4.

Zhang, W. *et al.* (2001) 'Site-Specific Acetylation by p300 or CREB Binding Protein Regulates Erythroid Krüppel-Like Factor Transcriptional Activity via Its Interaction with the SWI-SNF Complex', *Molecular and Cellular Biology*, 21(7), pp. 2413–2422. Available at: https://doi.org/10.1128/MCB.21.7.2413-2422.2001.

Zhou, D. *et al.* (2010) 'KLF1 regulates BCL11A expression and γ- to β-globin gene switching', *Nature Genetics*, 42(9), pp. 742–744. Available at: https://doi.org/10.1038/ng.637.