The regulation of ζ-globin expression

Andrew John King

Exeter College



A thesis submitted for the degree of

Doctor of Philosophy

University of Oxford

Weatherall Institute of Molecular Medicine

Medical Sciences Division

Trinity Term 2019

Abstract

Expression of the embryonic α-like globin gene, ζ-globin, is limited to the early stages of primitive erythropoiesis. Under normal circumstances it is completely silenced in definitive erythroid cells, despite lying adjacent to a “super-enhancer” composed of five separate enhancer elements.

Here, the chromatin environment has been characterized within the mouse α-globin cluster between primitive erythroblasts when ζ-globin expression is maximal, and definitive erythroblasts, when ζ-globin is silenced. In primitive erythropoiesis, the α globin locus is contained within an identical ~80kb erythroid-specific self-interacting chromosomal domain as found in definitive erythroid cells, delimited by the same CTCF boundary elements and with the same five enhancer elements. Changes in chromatin accessibility across the locus between primitive and definitive erythropoiesis are limited to the ζ-globin promoter which is markedly accessible in primitive erythropoiesis, and inaccessible in definitive erythroblasts. NG-Capture C shows that ζ-globin remains within the same self-interacting domain as in definitive erythroid cells although the contact between the enhancers and the ζ-globin promoter is significantly reduced.

Recently, a novel immortalised human erythroid cell line, the HUDEP-2 line, has been described. Here, it is shown that this serves as an appropriate model for studying ζ-globin silencing. Using the HUDEP-2 cell line, primary human erythroid cells, primitive and definitive erythroid cells in mice and previously published data sets, it is shown that histone deacetylation and chromatin remodelling are responsible for ζ-globin silencing: the PRC2 complex, DNA methylation and H3K9me2/me3 do not play a role.

BCL11A and LRF together are responsible for silencing of the fetal stage β-like γ-globin gene and its murine orthologue. Using mouse models where these transcription factors have been perturbed, it is apparent that both play a crucial role in the silencing of ζ-globin with their removal leading to increased ζ-globin expression and changes in chromatin accessibility over the ζ-globin gene. Knockout of *BCL11A* and *LRF* in HUDEP-2 cells has similar effects, with knockout of both together leading to an increase in expression of ζ-globin to ~15% of the total α-like globins. Analysis of published CUT&RUN and ChIP-seq data sets show that both BCL11A and LRF bind directly to the ζ-globin promoter.

Knockout of *LRF* and *BCL11A* together in definitive erythroid cells does not lead to restoration of the level of ζ-globin expression to that seen in primitive erythropoiesis. This means that other transcription factors must be responsible. To facilitate screening of activating and repressive transcription factors and complexes, a new mouse model, where a yellow fluorescent protein variant replaces exon 3 of the endogenous ζ-globin gene, has been generated and characterised. This can also be used for screening drugs capable of de-repressing ζ-globin. Because exon 3 of the ζ-globin gene is removed in this mouse model, this knocks out ζ-globin when present in homozygosity. Live births of ζ-null mice are significantly decreased, but viable phenotypically normal mice can be born, indicating that ζ-globin is advantageous, but not essential, for normal development.

Acknowledgements

I must start by thanking my two supervisors, Professor Douglas Higgs and Dr Christian Babbs. Doug has provided fantastic support, direction and mentorship over the past four years. He has never missed an email, or a meeting, and is everything you could ask for in a supervisor. Moreover, his unfailing enthusiasm for science is extraordinary and inspirational. Thanks for letting me be a small part of it. Chris agreed to supervise me and let me take on the ζ-globin project; he unfailingly and patiently has dealt with all my (often daily) queries and provided supervision, help and direction. It wouldn't have been possible to complete the thesis without him.

I owe a debt of gratitude to all members of the Higgs', Gibbons', Buckle and Hughes' laboratories for their help. In particular, I would like to thank: Danuta Jeziorska for looking after me so well when I first arrived, for helpful discussions regarding genome engineering, and for providing significant help with generating the DNA methylation data; Damien Downes, for help with Capture-C, ChIP-seq, ATAC-seq and late-night computational queries and Rob Beagrie for his help with the single cell data generation, analysis and computational support. All are good friends, fantastic scientists and will have stellar careers. I look forward to one day saying I knew them when they were starting out.

Jacqueline Sharpe, Jackie Sloane-Stanley, Sue Butler and Hannah Robinson have provided invaluable assistance with mouse work. Philip Hublitz has provided on-going cheerful support with genome engineering queries. Jelena Telenius has written two of the pipelines used here and provided additional help with computational issues as has Ron Schwessinger. Hele Francis helped generate some of the data here and will be a great loss to science, and Yavor Bozhilov has been a companion in tissue culture, and also tolerated my teasing with good humour. Lars Hanssen has been a good friend over the years as well as a scientific sounding board. James Davies has provided excellent advice and support, both scientific and clinical as well as a memorable time in Barcelona.

Finally, I must thank Duantida (Bee) Songdej for laying the groundwork, and to Siyu Liu for carrying on the project after I returned to clinical practice.

Outside of the lab, I would like to thank Debbie Hay, who went beyond the call of duty in preparing for my PYA and additionally has provided on-going support and encouragement. I would also like to thank the teaching team at Exeter, in particular Nick Talbot and Hussein Al-Mossawi, the students for being so fun to teach and Caitlin Rigler, for joining the lab for her rotational project.

I will end by thanking my friends and family. Rotherham is a long way from Oxford, not just geographically, and my parents have always supported me, even if they have no idea what I'm doing. Special mention goes to Paolo Polzella who has been a good friend, as well as helping with flow cytometry late into the evenings. Finally, Caroline Watson has provided help with python and Illustrator, as well as being much kinder to me than I deserve. Hopefully now I won't bring the laptop on holiday.

This work was funded by the Wellcome Trust, and the Anya Sturdy Fund and I am very grateful to both for their financial assistance.

List of Abbreviations

^{o}C	Degrees Celsius
3C	Chromosome confirmation capture
aCGH	array Comparative genomic hybridization
ANOVA	Analysis of Variance
APC	Allophycocyanin
APC-Cy7	Allophycocyanin-cyanine7
ATAC-seq	Assay for Transposase Accessible Chromatin using sequencing
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BB700	Brilliant blue 700
BCB	Brilliant cresyl blue
BCL11A	B-cell lymphoma/leukemia 11A
Bcl11a$^{\Delta 10kb}$	*Bcl11a* with 10kb enhancer block deleted
BELA	Bristol Erythroid Line - Adult
BHFS	Haemoglobin Bart's Hydrops Fetalis Syndrome
bp	Base pair
BSA	Bovine serum albumin
Cas	CRISPR associated
CD	Cluster of differentiation
cDNA	complimentary DNA
CGI	CpG island
ChIP-seq	Chromatin Immunoprecipitation followed by high-throughput sequencing
CN	Copy Number
Cre	Cyclization recombination recombinase
CRISPR	Clustered regularly interspaced short palindromic repeat
CtBP	C terminal Binding Protein
CTCF	CCCTC-binding factor
DE	Differentially expressed
DHS	DNase I hypersensitive site
DKO	Knockout of BCL11A and LRF

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleotide Acid
DSB	Double stranded break
EB	Embryoid Body
EDTA	Ethylenediaminetetraacetic acid
EHMT	Euchromatic histone-lysine N-methyltransferase
EMP	Erythro-Myeloid Progenitor
Epo	Erythropoietin
ErD	Definitive erythroid cells
ErP	Primitive erythroid cells
ESC	Embryonic Stem Cell
FACS	Fluorescent activated cell sorting
FCS	Fetal Calf Serum
FITC	Fluorescein isothiocyanate
FKF	Frt-flanked kanamycin selection gene
Flox	Flanked by *loxP*
FMO	Fluorescent minus one
FOG1	Friend of GATA-1
Frt	Flippase recognition target
g	gram
G418	Geneticin
GATA1	GATA-binding factor 1
GFP	Green fluorescent protein
GPA	Glycophorin A
gRNA	guide RNA
H3K4me1	Histone H3 lysine 4 monomethylation
H3K4me3	Histone H3 lysine 4 trimethylation
H3K9me1	Histone H3 lysine 9 monomethylation
H3K9me2	Histone H3 lysine 9 dimethylation
H3K9me3	Histone H3 lysine 9 trimethylation
H3K27ac	Histone H3 lysine 27 acetylation
H3K27me3	Histone H3 lysine 27 trimethylation
Hb	Haemoglobin
HbA	Haemoglobin A

HbF	Haemoglobin F
HbH	Haemoglobin H
HDAC	Histone deacetylase
HDR	Homology Directed Repair
Het	Heterozygous
HiDEP	Human iPS derived erythroid progenitors
Hom	Homozygous
HPFH	Hereditary persistence of fetal haemoglobin
HPV16	Human Papilloma Virus 16
HS	Hypersensitive
HSC	Haematopoietic stem cells
HUDEP	Human umbilical cord blood derived erythroid progenitor
IEF	Isoelectric focusing
IMDM	Iscove's Modified Dulbecco's Medium
iPSC	induced pluripotent stem cell
kb	kilobase
KLF1	Krueppel-like factor 1
KLF3	Krueppel-like factor 3
KLF8	Krueppel-like factor 8
L	Litre
LCR	Locus control region
LD	Lineage deplete
LDB1	LIM domain binding protein 1
LIF	Leukaemia inhibitory factor
LIN28B	Lin-28 Homolog B
LMO2	Lim-only 2
LOH	Loss of heterozygosity
loxP	locus of crossover in P1
LRF	Leukaemia/Lymphoma Related Factor
LS	Lineage select
M	Molar
m	metre
MACS	Magnetic activated cell sorting
Mb	Megabase

MCH	Mean corpuscular haemoglobin
MCS	Multi-conserved species
MCV	Mean corpuscular volume
MEL	Mouse erythroleukemia
$MgCl_2$	Magnesium chloride
min	minute
MMEJ	Microhomology mediated end joining
mRNA	messenger RNA
MTG	Monothioglycerol
mTOR	mammalian target of Rapamycin
n.s.	not significant
NEB	New England Biolabs
NF-E2	Nuclear Factor, Erythroid 2
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
Nprl3	Nitrogen Permease Regulator-Like 3
NuRD	Nucleosome remodeling and deacetylase Complex
OMIM	Online Mendelian Inheritance in Man
PAM	Protospacer adjacent motif
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PE	Phycoerythrin
PE-Cy7	Phycoerythrin -cyanine7
PerCP/Cy5.5	Peridinin-chlorophyll-protein complex: cyanine5.5
PGK	Phosphoglycerate kinase
POGZ	Pogo Transposable Element Derived with ZNF Domain
Pol II	RNA polymerase II
PRC2	Polycomb Repressive Complex 2
P/S	Penicillin/Streptomycin
RBC	Red Blood Cell
RFP	Red fluorescent protein
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
ROSE	Rank ordering of super-enhancers

RPKM	Reads per kilobase per million mapped
RPL13A	Ribosomal Protein L13a
RPMI	Roswell Park Memorial Institute media
RT-qPCR	Real-time quantitative polymerase chain reaction
RT	Reverse transcriptase
SAP	Sin-associated protein
SCA	Sickle Cell Anaemia
SCF	Stem Cell Factor
SCL	Stem-cell leukaemia
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEA	Southeast Asia
SNP	Single nucleotide polymorphism
SOX6	Sry-Box 6
SSC	Saline-sodium-citrate
TAD	Topologically associated domain
TALENs	Transcription activator like effector nucleases
TSS	Transcriptional start site
UCSC	University of California Santa Cruz
UTR	Untranslated region
WIMM	Weatherall Institute of Molecular Medicine
WT	Wild-type
YAC	Yeast artificial chromosome
YFP	Yellow fluorescent protein
ZFN	Zinc Finger Nucleases

Table of Contents

1. Introduction 1

1.1 Erythropoiesis during mammalian development 1

1.2 Globin gene expression during human development 4

1.3 Globin gene expression during mouse development 9

1.4 Molecular mechanisms leading to α-thalassaemia 11

1.5 Clinical characteristics of α-thalassaemia and the Haemoglobin Bart's Hydrops fetalis syndrome (BHFS) 15

1.6 Prevention and treatment of the BHFS and outcomes in long-term survivors 17

1.7 Reactivation of ζ-globin as a potential therapy 20

1.7.1 Bart's Hydrops Fetalis Syndrome 20

1.7.2 Sickle cell anaemia 21

1.8 Chromatin structure 22

1.9 Gene silencing 25

1.10 Regulation of α-globin expression in definitive erythropoiesis in humans 30

1.10.1 Structure of the α-globin locus in humans 30

1.10.2 Transcription factors 32

1.10.3 Epigenetic regulation and histone modifications 33

1.11 Regulation of α-globin gene expression in mice 34

1.11.1 Differences between the murine and human α-globin loci 34

1.11.2 The α-globin enhancer cluster classifies as a super-enhancer in murine erythroid cells with the individual enhancers acting independently 36

1.11.3 The α-globin locus forms a self-interacting domain in erythroid cells flanked by CTCF sites 37

1.12 The regulation of ζ-globin in primitive and definitive erythropoiesis 41

1.12.1 Structure of the ζ-globin gene 41

1.12.2 The influence of cis-acting regulatory elements on ζ-globin expression 43

1.12.3 Trans-acting factors responsible for ζ-globin repression 47

1.13 Transcriptional differences between primitive and definitive erythropoiesis 55

1.14 Erythroid cellular models 56

1.15 The genome editing revolution 58

1.16 Summary and outline of the thesis 61

2 Materials and Methods 63

2.1 Cell Culture and related techniques 63

2.1.1 Culture of mouse embryonic stem cells 63

2.1.2 Differentiation of mouse embryonic stem cells 63

2.1.3 Primary Primitive Erythroblasts 63

2.1.4 Definitive erythroid cells derived from fetal liver 64

2.1.5 Isolation and selection of definitive erythroid cells from spleens 65

2.1.6 HUDEP-2 cell culture 65

2.1.7 Primary erythroid culture from $CD34^+$ cells 66

2.1.8 B16f10 cell culture 67

2.2 Mouse methods and husbandry 68
2.2.1 Mouse maintenance and breeding 68
2.2.2 Phenylhydrazine treatment of mice 68
2.2.3 Genotyping 69
2.2.4 Haematological analysis 72

2.3 Genome engineering 72
2.3.1 CRISPR/Cas9 construct cloning 72
2.3.2 Surveyor Assay (IDT) 73
2.3.3 Cloning of the Hba-x-mVenus donor construct 74
2.3.4 Transfection of mouse embryonic stem cells 75

2.4 Cell biology techniques 77
2.4.1 Flow cytometry – human cells 77
2.4.2 Flow cytometry – murine cells 80
2.4.3 FACS sorting – primary fetal livers 81
2.4.4 FACS sorting – primitive erythroid cells 82
2.4.5 Wright stain of cytospins 83

2.5 Routine molecular biology/protein techniques 83
2.5.1 DNA extraction 83
2.5.2 Sanger sequencing 84
2.5.3 RNA extraction and cDNA synthesis 84
2.5.4 RT-qPCR analysis 85
2.5.5 Multiplex single cell RT-qPCR 86
2.5.6 RNA standard curve preparation (performed by Rob Beagrie) 89
2.5.7 Southern blotting 90
2.5.8 Analysis of haemoglobin 91

2.6 Validation of the ζ-Venus model 91
2.6.1 Correlation between *Hba-x* and *Venus* expression and fluorescent signal in embryoid bodies (EBs) 91
2.6.2 Fixing and dissection of embryos 92
2.6.3 Drug screening optimisation 92

2.7 Next generation sequencing techniques 93
2.7.1 The Assay for transposase-accessible chromatin using sequencing (ATAC-seq) 93
2.7.2 Chromatin immunoprecipitation analysis followed by high throughput sequencing (ChIP-seq) 94
2.7.3 Next Generation (NG) Capture C 97
2.7.4 Bisulfite sequencing 101
2.7.5 Array comparative genomic hybridization 101

2.8 Computational Analysis 103
2.8.1 Analysis of ATAC-seq and ChIP-seq data 104
2.8.2 Identifying regions of differential ChIP-seq enrichment 105
2.8.3 Classification of Enhancer Elements and Super-enhancer analysis 105
2.8.4 Analysis of Capture-C data 106
2.8.5 Mapping and visualisation of RNA-seq data 106
2.8.6 Analysis of aCGH data 107
2.8.7 Analysis of Bisulfite sequencing data 107
2.8.8 Analysis of single cell RT-qPCR data 108
2.8.9 Data visualisation and graphs 108

3 The epigenetic landscape and self-interacting domain in primitive erythroblasts at the murine α-globin gene locus 109

3.1 Introduction 109

3.2 Results 112

3.2.1 Source of primary primitive and definitive erythroid cells 112
3.2.2 In primitive erythroid cells, the *Hba-x* gene is accessible and is marked by histone acetylation 114
3.2.3 CTCF binding throughout the α-globin locus is broadly unchanged between primitive and definitive erythroid cells 117
3.2.4 The α-globin enhancer cluster classifies as a super-enhancer (SE) in primitive erythroid cells 119
3.2.5 The chromatin compartment formed in primitive erythropoiesis is almost identical to that found in definitive erythroid cells 121
3.2.6 The *Hba-x* gene promoter contacts the enhancers at a lower frequency in definitive erythroid cells although it remains within the self-interacting domain 125
3.2.7 The embryonic globin genes at the murine β-globin locus together with an intergenic region form a small self-interacting region in contrast to the changes seen at the α-globin locus 126
3.2.8 A novel hypersensitive element is seen within the α-globin locus at E11.5 130

3.3 Discussion 132

4 The effect of Lrf and Bcl11a on the endogenous globin loci in mouse 142

4.1 Introduction 142

4.2 Mouse models: choice and validation 143
4.2.1 Lrf knockout 143
4.2.2 *Bcl11a* enhancer knockout 145

4.3 Results 152
4.3.1 *Lrf* knockout leads to increased apoptosis and a differentiation block 152
4.3.2 Knockout of *Lrf* leads to *Hba-x* and *Hbb-bh1* globin gene derepression and profound changes in chromatin accessibility in early fetal life 154
4.3.3 Fetal definitive *Lrf*$^{-/-}$ erythroid cells undergo further embryonic globin gene silencing when cultured *in vitro* for 8 days 161
4.3.4 Knockout of the *Bcl11a* composite enhancer leads to embryonic globin derepression and mild chromatin accessibility changes in early fetal erythroid cells 165
4.3.5 Knockout of the *Bcl11a* erythroid specific enhancers causes embryonic globin gene de-repression in adult erythroid cells with minimal changes in chromatin accessibility 171

4.4 Discussion 175

5 The chromatin landscape in human erythroid cells at the α-globin locus and the effect of LRF and BCL11A knockout on ζ-globin expression 181

5.1 Introduction 181

5.2 Results 185
5.2.1 Characterisation of the HUDEP-2 cell line 185
5.2.2 Knockout of *LRF* and *BCL11A* separately leads to embryonic and fetal globin de-repression in the HUDEP-2 cell line 198
5.2.3 Knockout of *LRF* and *BCL11A* separately leads to marked changes in chromatin accessibility and histone acetylation at the β-globin locus with minimal changes seen at the α-globin locus 202
5.2.4 LRF and BCL11A bind directly to the ζ-globin gene 205
5.2.5 Knockout of *BCL11A* and *LRF* together leads to greater embryonic and fetal globin de-repression than knockout of each individual transcription factor in HUDEP-2 cells 207
5.2.6 Knockout of BCL11A and LRF together leads to an increase in chromatin accessibility over the ζ-globin gene 210
5.2.7 DNA methylation is not necessary for repression of ζ-globin and methylation of the CpG island located at the 3' end becomes methylated as transcription increases 211
5.2.8 The ζ-globin gene is not marked by H3K9me3 or H3K27me3 in primary definitive erythroid cells and inhibition of EHMT1/EHMT2 does not lead to ζ-globin de-repression. 214

5.3 Discussion 216

6 Generation and characterisation of a ζ-globin reporter mouse 222

6.1 Introduction 222

6.2 Results 227

6.2.1 ζ-globin reporter design 227
6.2.2 Targeting Strategy 228
6.2.3 Screening of targeted ESCs 232
6.2.4 Validation of the ζ-Venus reporter in the EB culture system 235
6.2.5 Mouse generation 237
6.2.6 Validation of ζ-Venus *in vivo* 239
6.2.7 Validation of ζ-Venus erythroid cells for drug screening 243
6.2.8 The effect of ζ-globin knockout on murine survival 245

6.3 Discussion 247

7 General discussion and conclusions 253

7.1 ζ-globin is advantageous but not essential for normal development 253

7.2 There are decreased contacts between the enhancers and the ζ-globin promoter when definitive and primitive erythroid cells are compared 254

7.3 Silencing of ζ-globin is mediated by chromatin remodelling and histone deacetylation 255

7.4 LRF and BCL11A play an important role in ζ-globin repression 258

7.5 Potential therapeutic strategies to de-repress ζ-globin 262

7.6 Summary 264

8 References 265

List of Figures

Figure 1.1 A simplified diagram of the α- and β- multigene clusters in human. 5
Figure 1.2 Haemoglobin switching at the human α- and β-globin loci. 7
Figure 1.3 A simplified diagram of the murine α and β globin loci. 9
Figure 1.4 Haemoglobin switching at the murine α- and β-globin loci. 10
Figure 1.5 The common deletions leading to α-thalassaemia 12
Figure 1.6 The prevalence and distribution of mutations leading to deletional α-thalassaemia. 13
Figure 1.7 Histone modifications at active gene promoters, enhancers and repressed genes. 24
Figure 1.8 Gene silencing mechanisms. 29
Figure 1.9 An erythroid specific self-interacting domain forms in erythroid cells 40
Figure 1.10 Chromatin accessibility at the α-globin locus in definitive and primitive erythropoiesis. 44
Figure 1.11 Nanostring RNA quantification of the ratio of *Hba-a1/2* and *Hba-x* to β-like globin transcripts in each of the enhancer knockout models. 45
Figure 2.1 *Bcl11a*$^{\Delta 10kb}$ genotyping strategy 70
Figure 2.2 β-YAC genotyping strategy 70
Figure 2.3 *LRF* flox and knockout genotyping strategy. 71
Figure 2.4 Surveyor assay. 74
Figure 2.5 Flow cytometry on adult blood and blood from a terminated fetus with Hb Bart's hydrops fetalis. 79
Figure 2.6 RNA expression from different sorted fractions of HUDEP-2 cells after antibody staining with ζ-globin and γ-globin. 80
Figure 2.7 Sorting Protocol for primary fetal livers 82
Figure 3.1 Primitive and definitive erythroid cells used for experiments. 113
Figure 3.2 ChIP-seq for H3K4me3, H3K4me1, H3K27ac and CTCF in primitive erythroid cells at the α-globin locus 116
Figure 3.3 ChIP-seq for H3K4me3, H3K4me1, H3K27ac, H3K27me3 and CTCF in definitive erythroid cells at the α-globin locus. 116
Figure 3.4 CTCF binding throughout the locus in embryonic stem (ES) cells, primitive, erythroid cells and definitive erythroid cells. 117
Figure 3.5 CTCF occupancy overlap in primitive and definitive erythropoiesis. 119
Figure 3.6 Chromatin states in primitive cells determined using GenoSTAN 120
Figure 3.7 Super-enhancers in primitive erythroid cells 121
Figure 3.8 NG Capture-C profiles from the *Rhbdf1* promoter, the HS-38 CTCF element and the R1 erythroid enhancer 124
Figure 3.9 NG Capture-C profiles from the *Hba-x* promoter and the *Hba-a1/2* promoters. 127
Figure 3.10 Overlaid, normalized NG Capture-C data at the β-globin locus. 128
Figure 3.11 Chromatin accessibility at the α-globin locus during development. 131
Figure 3.12 Proposed configuration of the α-globin locus in primitive and definitive erythropoiesis based upon the NG Capture-C data. 137
Figure 4.1 *LRF* floxed mouse model. 144
Figure 4.2 Chromatin accessibility, histone marks and interaction profile at the *Bcl11a* gene locus in murine primitive and definitive erythroid cells. 148

Figure 4.3 Representative blood smears and reticulocyte staining for WT, heterozygous and homozygous mice for the *Bcl11a*$^{\Delta 10kb}$ deletion. 150
Figure 4.4 Chromatin accessibility and *Bcl11a* expression in WT definitive, primitive (E10.5) and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ definitive erythroid cells.. 151
Figure 4.5 Erythroid differentiation in freshly isolated fetal liver (E12.5), defined by the CD71/Ter119 flow cytometric profile in *LRF*$^{-/-}$ mice 152
Figure 4.6 Representative plots of erythroid differentiation in *Lrf*$^{Flox/Flox}$, *Lrf*$^{Flox/-}$ and *Lrf*$^{-/-}$ from E12.5 fetal livers. 153
Figure 4.7 PCA Analysis for single erythroid cells from *Lrf*$^{Flox/Flox}$, *Lrf*$^{Flox/-}$ and *Lrf*$^{-/-}$ erythroid cells from E12.5 fetal livers. 156
Figure 4.8 Chromatin accessibility and single cell expression at the α-globin locus in *Lrf*$^{Flox/Flox}$, *Lrf*$^{Flox/-}$ and *Lrf*$^{-/-}$ erythroid cells from the S3 population. 159
Figure 4.9 Chromatin accessibility and single cell expression at the β-globin locus in *Lrf*$^{Flox/Flox}$, *Lrf*$^{Flox/-}$ and *Lrf*$^{-/-}$ erythroid cells from the S3 population 160
Figure 4.10 Differentiation of *Lrf*$^{Flox/Flox}$ and *Lrf*$^{-/-}$ littermates in an erythroid fetal liver differentiation culture system. 162
Figure 4.11 Chromatin accessibility and expression of the α-like globins at the α-globin locus in *Lrf*$^{Flox/Flox}$ and *Lrf*$^{-/-}$ cells from the *in vitro* fetal liver culture system 163
Figure 4.12 Chromatin accessibility and expression of the β-like globins at the β-globin locus in *Lrf*$^{Flox/Flox}$ and *Lrf*$^{-/-}$ cells from the *in vitro* fetal liver culture 164
Figure 4.13 PCA Analysis for single erythroid cells for all stages (S0-S3) and S3 alone, from WT, heterozygote and homozygote *Bcl11a*$^{\Delta 10kb}$ knockout cells. 168
Figure 4.14 Chromatin accessibility and single cell expression at the α-globin locus in *Bcl11a*$^{WT/WT}$, *Bcl11a*$^{\Delta 10kb/WT}$ and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ erythroid cells from the S3 population 169
Figure 4.15 Chromatin accessibility and single cell expression at the β-globin locus in *Bcl11a*$^{WT/WT}$, *Bcl11a*$^{\Delta 10kb/WT}$ and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ erythroid cells from the S3 population. 170
Figure 4.16 Chromatin accessibility and expression of the α-like globins from the α-globin locus in WT and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ splenic Ter119+ erythroblasts.. 173
Figure 4.17 Chromatin accessibility and expression of the β-like globins from the β-globin locus in WT and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ splenic Ter119+ erythroblasts..... 174
Figure 5.1 Differentiation protocol for HUDEP-2 cells.. 186
Figure 5.2 Flow cytometric analysis of expression of cell surface membrane proteins during HUDEP-2 differentiation. 187
Figure 5.3 Sorting protocol which can be used to separate HUDEP-2 cells at Day 10 of differentiation 188
Figure 5.4 Expression analysis for *HBA*, *HBB*, *HBD*, *HBZ*, *HBG* and *HBE* from HUDEP-2 cells in expansion phase and at Day 10 of differentiation 190
Figure 5.5 Chromatin accessibility at the α-globin locus in HUDEP-2 cells 192
Figure 5.6 Chromatin accessibility at the β-globin locus in HUDEP-2 cells 193
Figure 5.7 Histone marks and CTCF binding for primary erythroid cells and HUDEP-2 cells at the α-globin locus 195
Figure 5.8 Histone marks and CTCF binding for primary erythroid cells and HUDEP-2 cells at the β-globin locus..... 196
Figure 5.9 Karyotype of a WT HUDEP-2 cell line. 197
Figure 5.10 Differentiation of the WT, *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ clone 199
Figure 5.11 Expression and protein analysis globins in WT, *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ HUDEP-2 clones 201

Figure 5.12 Chromatin accessibility and histone marks for WT *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ HUDEP-2 cells at the α-globin locus. 203
Figure 5.13 Chromatin accessibility and histone marks for WT *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ HUDEP-2 cells at the β-globin locus. 206
Figure 5.14 ChIP-seq and CUT&RUN data for LRF and BCL11A binding across the α-globin locus. 206
Figure 5.15 Day 10 differentiation of a WT and double L*RF/BCL11A* knockout clone 208
Figure 5.16 Expression and haemoglobin production from the *LRF/BCL11A* double knockout clone. 209
Figure 5.17 Chromatin accessibility, assessed by ATAC-seq at the a α-globin locus and β-globin locus in WT and DKO cells. 210
Figure 5.18 DNA methylation across the *HBZ* gene in primary erythroid cells and the HUDEP-2 clones. 213
Figure 5.19 Normalised ChIP-seq data for H3K9me3 and H3K27me3 in fetal and adult erythroblasts. 214
Figure 5.20 RNA-seq data from primary erythroid cells derived from a CD34+ culture system following addition of a vehicle control (DMSO; top) and a specific inhibitor of EHMT1/2 (UNC0638). 215
Figure 6.1 Absorption spectra of haemoglobin; excitation and emission spectra of different fluorescent proteins 225
Figure 6.2 Design of the ζ-globin reporter. 227
Figure 6.3 Constructs used for initial ES cell targeting 230
Figure 6.4 Targeting of mouse ESCs. 231
Figure 6.5 PCR screening of targeted clones. 232
Figure 6.6 Southern blotting strategy and results 234
Figure 6.7 Correlation between fluorescent signal and expression. 237
Figure 6.8 Mouse generation. 238
Figure 6.9 Fluorescence at E10.5 240
Figure 6.10 Fluorescent signal in definitive erythroid cells 241
Figure 6.11 Sections of a. WT/WT and b. ζ-Venus/WT embryos. 242
Figure 6.12 The effect of adding an HDAC inhibitor to erythroid cells. 243
Figure 6.13 High throughput drug screening protocol 244
Figure 6.14 A homozygous ζ-Venus mouse with its WT littermate. 245
Figure 6.15 Workflow of CRISPR/Cas9 lentiviral screen. 249
Figure 7.1 Proposed model explaining silencing of ζ-globin 257
Figure 7.2 Transcriptional network responsible for repressing ζ-globin. 259

List of Tables

Table 2.1 Oligonucleotide primers used for CRISPR/Cas9 construct cloning. 73
Table 2.2 Primers used for generation of ζ-Venus Mouse 76
Table 2.3 TaqMan probes used for RT-qPCR analysis.. 86
Table 2.4 Custom primers used for RT-qPCR analysis. 86
Table 2.5 Taqman assays used in single cell experiments 88
Table 2.6 Custom Taqman probes used in single cell experiments 89
Table 2.7 Antibodies used in ChIP experiments 95
Table 2.8 Primers used to check enrichment following ChIP 96
Table 2.9 Oligonucleotide probes used for Capture-C experiments. All had a 5' biotinylated modification. 100
Table 2.10 Nested primers used for bisulfite sequencing. 102
Table 2.11 Publicly available data-sets used for ChIP-seq. 104
Table 4.1 Hematologic parameters for adult mice for WT, heterozygous and homozygous *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ models.. 149
Table 4.2 Genes used in the single cell analysis for the *Lrf*$^{-/-}$ mice. 155
Table 4.3 Genes used for analysis of *BCL11A*$^{\Delta 10kb}$ mice 167
Table 5.1 aCGH on the WT HUDEP-2 cell line. 198
Table 5.2 Consensus sites for BCL11A and LRF in the human ζ-globin promoter 207
Table 6.1 Number of each genotype of adult mice born for WT, heterozygous (Het) and homozygous (Hom) ζ-globin knockout. 246
Table 6.2 Number of each genotype for WT, Het and Hom ζ-globin knockout at E12.5. 247

1. Introduction

Two of the best characterized mammalian multi-gene clusters, the α and β globin gene loci, which together form haemoglobin in red blood cells, have been studied for over sixty years. Careful scientific work examining these genes and their regulatory elements have elucidated many of the general principles by which mammalian genes are normally switched on and off during cell-fate decisions, and how this may be perturbed in human genetic disease. In this thesis, I examine how the expression of a single gene (ζ-globin) which lies within the α-globin cluster, can be regulated so precisely throughout development and differentiation. Understanding the regulation of this single gene at a model locus may provide insights into how other genes' expression is controlled. Separately, it may offer potentially new therapeutic avenues for a common lethal genetic condition: severe α-thalassaemia.

1.1 Erythropoiesis during mammalian development

Red blood cells (RBCs) carry oxygen and carbon dioxide between the lungs and tissues. Given their distinctive morphology, easy availability and abundance, they have been studied for centuries, and are amongst the most well understood of any cell type within the body (Hajdu, 2003).

The mouse is the classical model used to study mammalian physiology and consequently, erythropoiesis during murine development is the best characterised

(Baron et al., 2013; Palis, 2014). In the mouse (and almost certainly in other mammals), erythropoiesis occurs in three distinct waves (Baron et al., 2013; Palis, 2014).

The initial wave of haematopoietic activity arises outside the embryo within the yolk sac from “blood islands”; these cells differentiate from the mesoderm shortly after gastrulation and give rise to a semi-synchronous wave of nucleated erythroid precursors (primitive erythroid cells; ErP) within the embryo (Palis et al., 1999; Kingsley et al., 2004; Kingsley et al., 2006). In the mouse, which has a typical gestation of 20 days, these embryonically derived cells are predominantly seen between E9.5-E12.5 with their progenitors arising at E7.25 (Palis et al., 1999). Although they are only produced for a short period, they circulate until (and shortly after) birth, maturing and eventually enucleating (Kingsley et al., 2004).

ErP play a vital role in the transition between the embryo and fetus in the mammal. In addition to their role in oxygen delivery, they may also scavenge reactive oxygen species (Kingsley et al., 2013) and generate shear forces necessary for vascular remodelling (Blatnik et al., 2005).

A second wave of transient erythropoiesis arises from within the yolk sac at E8.25 in mice just prior to the onset of circulation (Palis et al., 1999). The erythroid progenitors produced are associated with the emergence of multipotential hematopoietic progenitor cells termed erythro-myeloid (EMP) progenitors (Palis et al., 2001). Definitive erythroid progenitors expand rapidly in the yolk sac, and are subsequently found in the bloodstream and the fetal liver (Palis et al., 1999) being

detectable between E11.5 and E12.5 (McGrath et al., 2011). These are considered the first definitive erythroid cells (ErD). ErPs are normally easier to distinguish morphologically from ErD by their larger size.

In adults, all circulating blood cells including erythroid cells are derived from hematopoietic stem cells, defined by their ability to repopulate a lethally irradiated mouse with the entire blood system. Hematopoietic stem cells are initially formed in the aorta-gonad-mesonephros (AGM) region and emerge from major arterial vessels in the developing embryo at E10.5 in the mouse (Kumaravelu et al., 2002): this is the third wave of erythropoiesis. They subsequently expand in the fetal liver, and then transition to their permanent location in the bone marrow; this transition begins at E17.5 and is completed after birth (Blazsek et al., 2000).

In humans, erythropoiesis during development is less well defined, due to the inaccessibility of material for experiments and the ethical constraints involved in making such observations. However, it seems probable that the same three waves of erythropoiesis occur. A cohort of cells originating in the yolk sac (ErPs) between days 17-21 are released into the circulation as the heart begins to beat corresponding to the first wave; these cells then mature and enucleate (Van Handel et al., 2010). It is likely, though not yet proven, that a second transient wave of erythropoiesis originating in the yolk sac around day 23 migrate to the fetal liver as EMPs; these are the first definitive erythroid cells (Migliaccio et al., 1986). Finally, at around day 28-35, cells arise from the floor of the dorsal aorta where they undergo a transition to Hematopoietic stem cells; these subsequently contribute to

blood formation throughout life (Ivanovs et al., 2011). By approximately 70 days, all new red cells are derived from these hemopoietic stem cells.

1.2 Globin gene expression during human development

The haemoglobin molecule, which carries oxygen in the blood, is a tetramer composed of two α-like and two β-like globin peptides along with heme moieties. Its production is encoded by the α and β gene clusters including the genes and their enhancers, which are located on chromosome 16 and 11 respectively in humans (Figure 1.1).

ErPs uniquely express embryonic globins, termed ζ-globin (the embryonic form of α-globin) and ε-globin (the embryonic form of β-globin) (Peschle et al., 1985). Together, these combine to form haemoglobin Gower I (ζ_2 ε_2). This is the major haemoglobin present in the earliest ErPs studied, with the embryonic globins making up approximately 80% of the total globins from ErPs examined at 5 weeks (Peschle et al., 1985). The remaining 20% comprise α-globin and γ-globin (from the α-globin and β-globin locus respectively). Over the first 5-7 weeks of development, there is a maturational switch at the α-globin locus, during which ζ-globin is silenced, with α-globin then comprising the entirety of the α-like globins from that point onwards; this corresponds to the switch to definitive erythropoiesis (Peschle et al., 1985). Studies in mice suggests that ζ-globin expression is also silenced within the primitive erythroid lineage during maturation (Kingsley et al., 2006), and there is some evidence to suggest this also occurs in humans (Peschle et al., 1985).

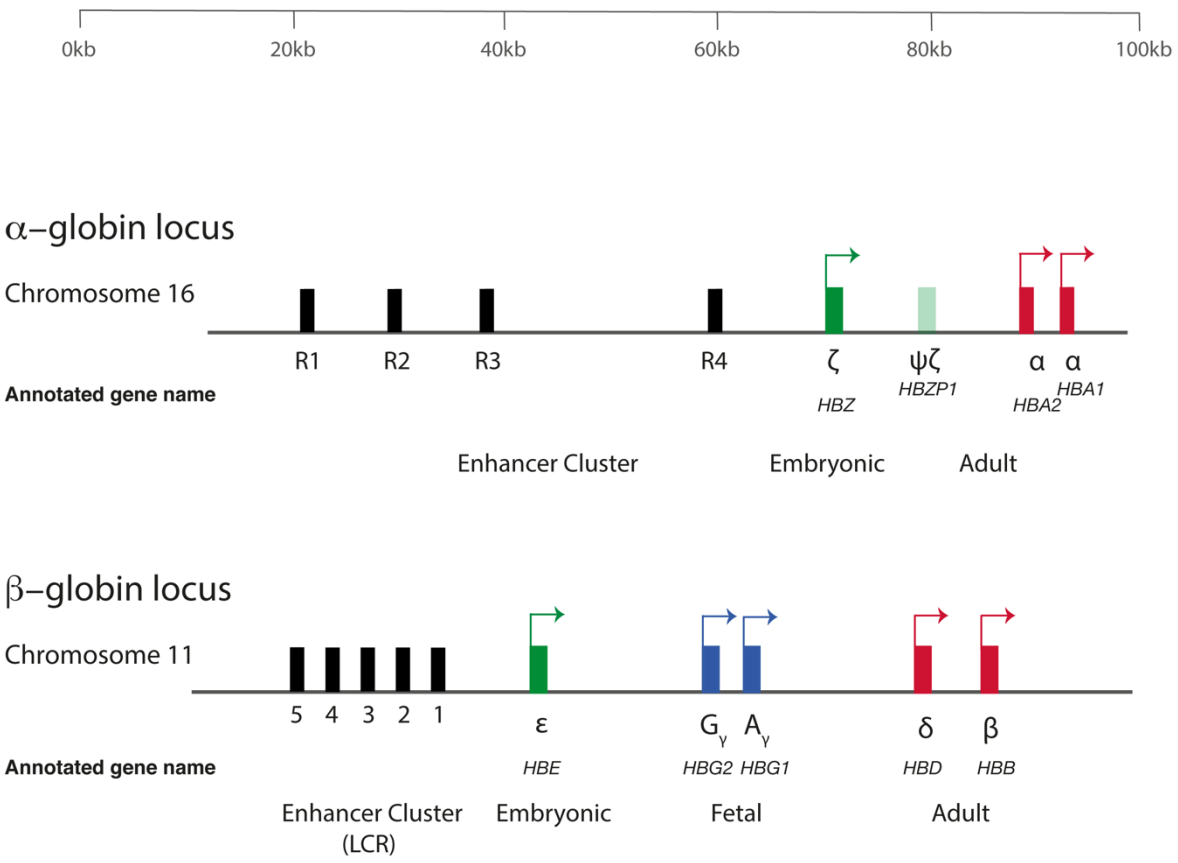


Figure 1.1 A simplified diagram of the α- and β- multigene clusters in human. The coding genes and enhancer elements are indicated, along with the annotated gene names. The embryonic globin genes are shown in green, the fetal globin genes in blue, and the adult globin genes in red. There are paralogous α-globin genes and γ-globin genes, as shown. The ψζ gene (explained in more detail on page 41) is also shown for reference.

Because α-globin is always expressed, haemoglobin Gower II ($\alpha_2\varepsilon_2$) is also present in the earliest ErPs; similarly, because γ-globin is expressed there are small amounts of Haemoglobin Portland I ($\zeta_2\gamma_2$) and HbF ($\alpha_2\gamma_2$).

At the β-globin locus, there is a similar temporal switch to that occurring at the α-globin locus, although a fetally-expressed globin, γ-globin replaces ε-globin. This switch appears to occur slightly later than the ζ- to α-switch, with ε-globin being silenced at approximately 8 weeks (Peschle et al., 1985).

A further switch occurs at the β-globin locus approximately 3 months after birth, such that γ-globin is silenced, and the adult globins (β- and δ-globin) are expressed. This means that during fetal life, the predominant haemoglobin is HbF ($\alpha_2\gamma_2$); this then changes to HbA ($\alpha_2\beta_2$) after birth. HbA_2 ($\alpha_2\delta_2$) normally accounts for 2-3% of the total haemoglobin and serves no clear physiological function (Steinberg and Rodgers, 2015).

Haemoglobin switching during development, alongside the predominant haemoglobin for each stage is shown in Figure 1.2.

The embryonic globin genes are strongly conserved throughout evolution and their expression timing is remarkably consistent. In all species examined, including birds, the orthologs to the ζ globin gene are expressed in embryonic erythroid cells, and the orthologs of the α-globin genes are expressed in fetal and adult erythroid cells (Hardison, 2012). Similarly, embryonic β-like globin genes are always expressed from the β-globin locus, although the γ-globin and β-globin orthologs have not retained the same timing of expression in all mammalian orders. In most mammals, the γ-globin and β-globin orthologs are expressed in embryonic and fetal life respectively, whereas in simian primates, including humans, they have a specific fetal and adult timing of expression (Hardison, 2012).

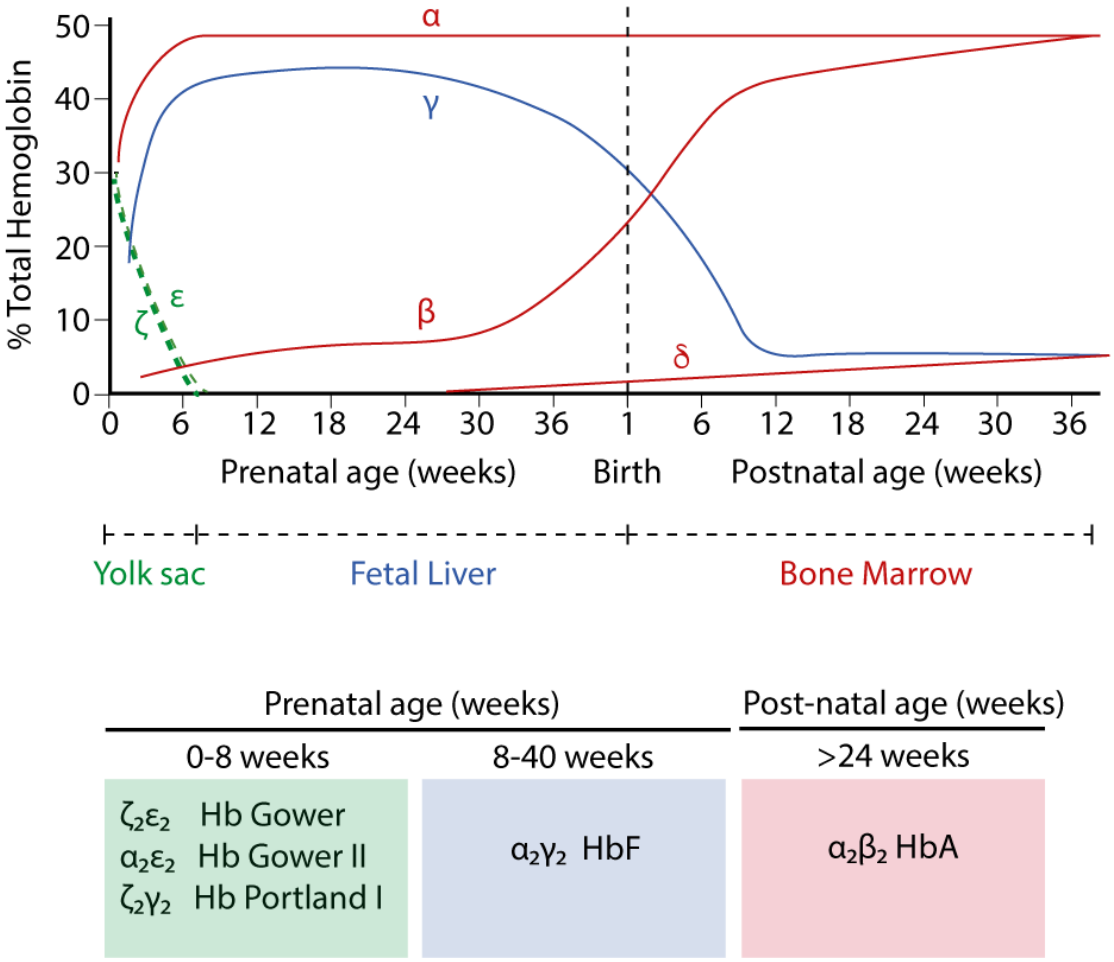


Figure 1.2 Haemoglobin switching at the human α- and β-globin loci. The embryonic globins are shown in green, the fetal globins are shown in blue and the adult globins are shown in red. Primitive erythropoiesis, derived from the yolk sac, is characterized by expression of ζ-globin (from the α-globin locus) and ε-globin (from the β-globin locus). These are silenced at approximately 7 and 8 weeks' gestation respectively. α-globin then accounts for the entirety of the transcriptional output from the α-globin locus. At the β-globin locus, there is a switch to a fetal globin (γ-globin) during fetal life, and then a second switch to the adult β-globin. The predominant type of haemoglobin corresponding to each developmental stage is shown, alongside the main site of haematopoiesis, corresponding to each stage (adapted from Blobel et al., 2015).

The haemoglobin molecules expressed during development have an increased oxygen affinity during both embryogenesis and fetal life (Maurer et al., 1970; Tuchinda et al., 1975; He and Russell, 2001). A widely held, yet largely unsubstantiated explanation for the evolutionary pressure on embryonic and fetal globins, is that they enhance *trans*-placental oxygen delivery to the developing offspring from the mother. This hypothesis, however, ignores the observation that embryonic globins are also found in birds (where gas exchange is via diffusion rather than trans-placental); women with mutations leading to haemoglobins which have a higher affinity than embryonic and fetal haemoglobins are also perfectly capable of delivering normal infants (Charache et al., 1985). An alternative explanation is that the properties of the fetal and embryonic haemoglobins confer an as of yet unexplained evolutionary benefit related to their protein structure, perhaps independent of their gas exchanging function.

It remains unclear whether ζ-globin is essential for normal development. There are currently no reports of individuals with deletion of both functional ζ-globin genes (Fei et al., 1989; Felice et al., 1986; Winichagoon et al., 1982). There are many reports where one copy of ζ-globin has been removed; such deletions are accompanied by deletion of the linked α-globin genes. Heterozygous deletion of ζ-globin can be as high as 0.4-2.8% in some populations; duplications of ζ-globin have also been reported. Individuals with heterozygous deletions do not have a phenotype suggestive of early developmental abnormalities associated with removal of one copy of ζ-globin. In mouse models where the murine ζ-globin gene is deleted, some mice are viable dependent on strain background (Leder et al., 1997); however these models were made using homologous recombination with neomycin selection,

meaning a PGK-neomycin cassette was left in place of the ζ-globin gene. Viable mice had α-thalassaemia, probably because the ectopic promoter competed with α-globin for the α-globin enhancers (Leder et al., 1997). This almost certainly confounded interpretation of the results.

1.3 Globin gene expression during mouse development

The mouse is an important tractable model for studying haemoglobin switching, and there is significant conservation of structure and function. As in humans, there are embryonic and adult specific globins, although there is no equivalent of fetal haemoglobin in the mouse (Blobel et al., 2015). A schematic of the mouse α- and β-globin loci is shown in figure 1.3 and the haemoglobin switching during development is shown in figure 1.4.

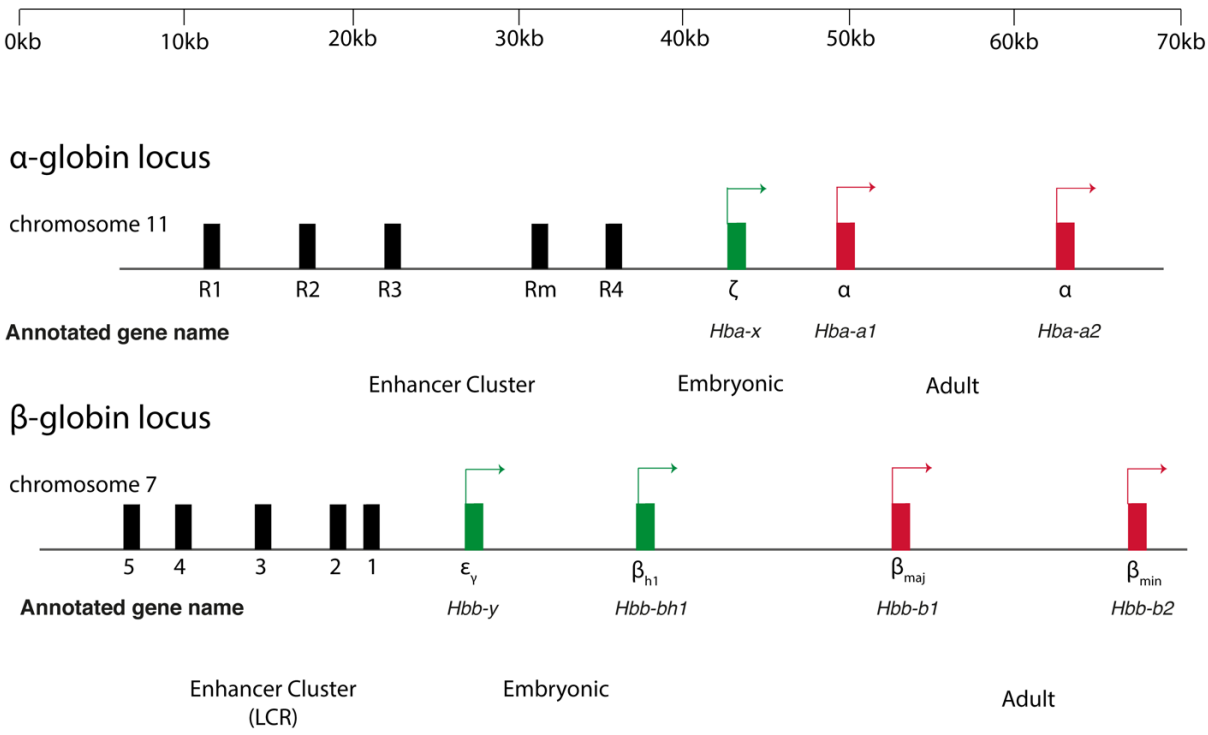


Figure 1.3 (previous page) A simplified diagram of the murine α and β globin loci. The coding genes and enhancer elements are indicated. The α-globin locus has an additional mouse specific enhancer element (Rm). The β-globin locus has two embryonic genes (εγ and βh1) and two adult globin genes (βmajor and βminor); these are expressed in an 80/20 ratio in adult life (Hutton et al., 1962).

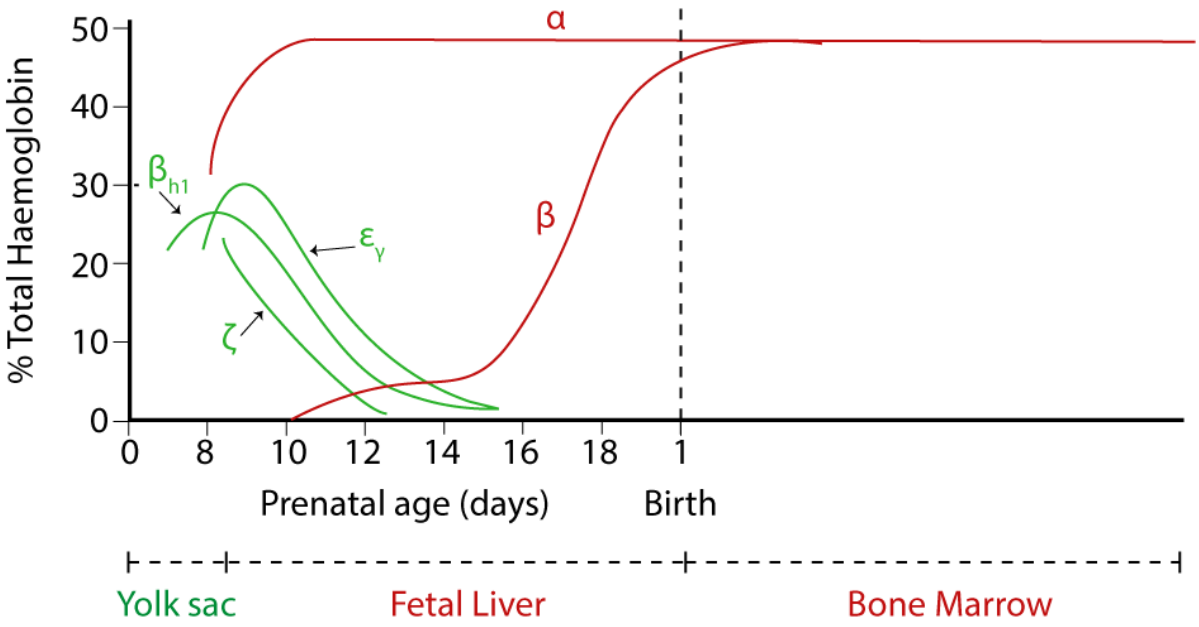


Figure 1.4 Haemoglobin switching at the murine α- and β-globin loci. Primitive erythropoiesis, derived from the yolk sac, is characterized by expression of embryonic globins (ζ, εγ and βh1); definitive erythroid cells express the adult globins. Primitive erythroid cells also express α-globin. There is no "fetal" stage equivalent to that seen in humans (adapted from Blobel et al., 2015).

Murine primitive erythroid cells undergo a maturational switch in their globin expression, with the silencing of ζ-globin at the α-globin locus, and a switch between βh1 and εγ globin at the β-globin locus between E10.5 and E12.5 (Kingsley et al., 2006). Definitive erythroid cells derived from EMPs express small amounts of βh1,

but otherwise express adult globins only (McGrath et al., 2011). Cells derived from the third wave of erythropoiesis only express adult globins.

1.4 Molecular mechanisms leading to α-thalassaemia

Production of globin peptides is tightly balanced under normal circumstances so that during both development and adulthood, the α and β globin-like peptides are synthesized in a 1:1 ratio (Weatherall et al., 1965). Decreased synthesis of either chain leads to a condition called thalassaemia (denoted α-thalassaemia if there are decreased α chains) or β-thalassaemia (if there are decreased β chains), with the severity of the thalassaemia depending upon the degree of imbalance of globin synthesis (Higgs et al., 2012).

Naturally occurring mutations in the α-globin locus leading to α-thalassaemia are found at high frequencies throughout all tropical and sub-tropical regions of the world; it appears that such variants have been selected because carriers have a selective advantage where *falciparum* malaria is endemic (Flint et al., 1986). α-thalassaemia is most commonly caused by one of over fifty deletions removing either one (ζ -α, referred to as α^{+}-thalassaemia) or both (ζ --, referred to as α^{0}-thalassaemia) α-globin genes (Harteveld and Higgs, 2010). Some deletions remove all the α-like genes (-/--) including ζ-globin. The α-globin genes are embedded within two highly homologous 4kb duplication units. The most common deletions causing deletion of one α-globin gene thus occur as a result of reciprocal recombination events with the other chromosome ($\alpha\text{-}^{3.7}$ and $-\alpha^{4.2}$) (Harteveld and Higgs, 2010). Aside from deletions, α-thalassaemia may be caused by single point

mutations, the most common of which is the constant spring mutation, which mutates the stop codon and leads to an extended mRNA and an unstable protein (Harteveld and Higgs, 2010). Occasionally, α-thalassaemia may be caused by deletions affecting the main *cis*-acting enhancer elements, leaving the α-globin genes intact (Hatton et al., 1990). The distribution of the deletional α-thalassaemias world-wide is shown in Figure 1.5 and the common mutations leading to the deletional α-thalassaemias are shown in Figure 1.6.

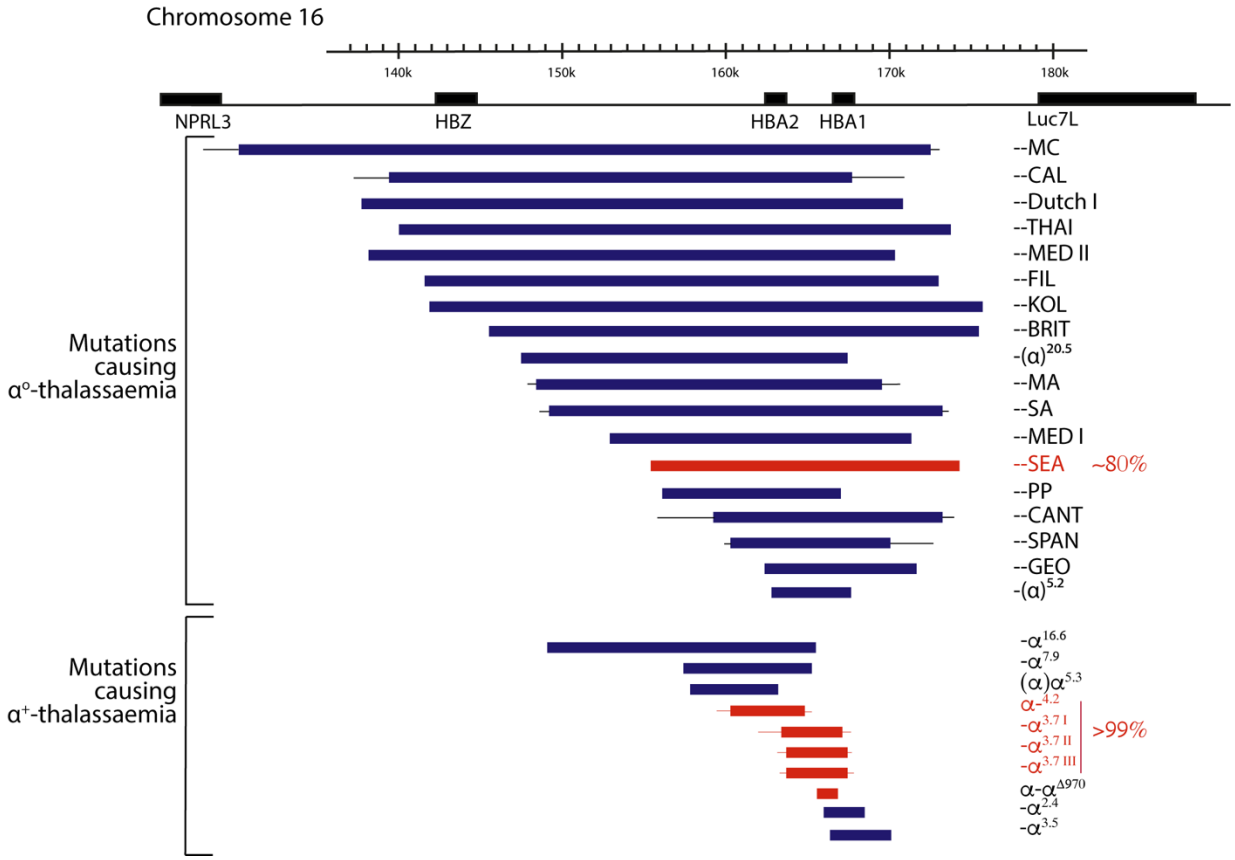


Figure 1.5 The common deletions leading to α-thalassaemia. The most frequent mutations and their incidence are highlighted in red. Adapted from (Harteveld and Higgs, 2010). Thin lines indicate uncertainty in the breakpoints.

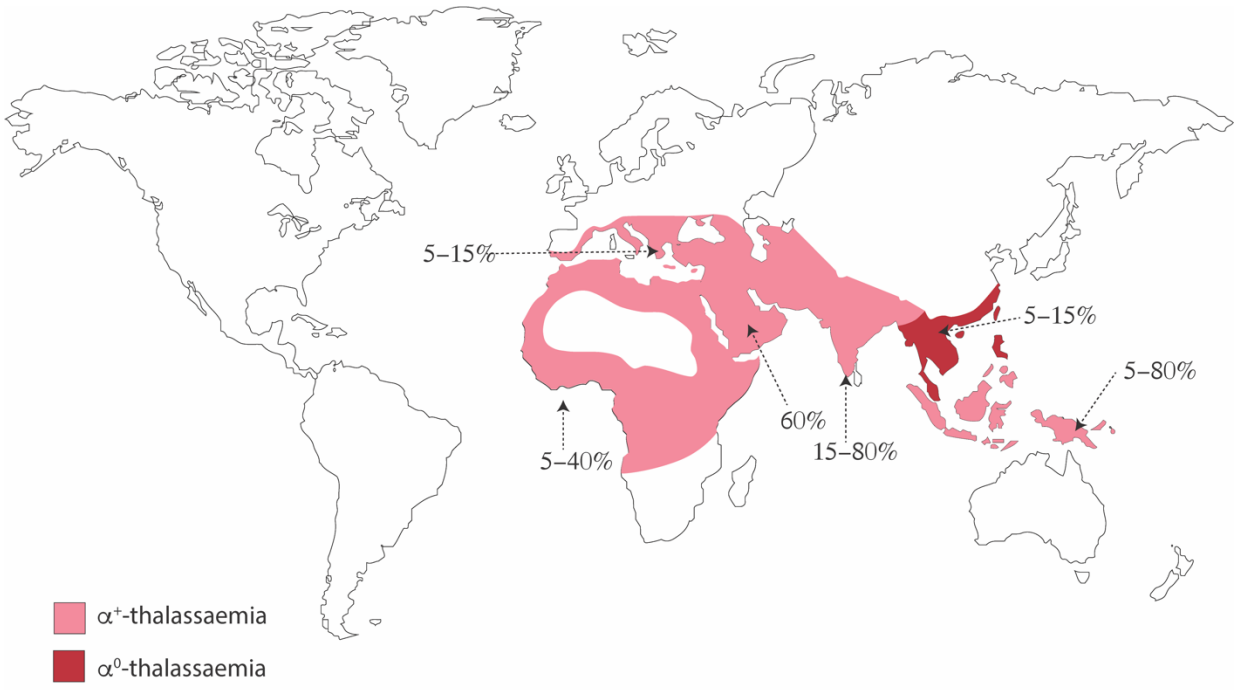


Figure 1.6 The prevalence and distribution of mutations leading to common deletional α-thalassaemia. The frequency varies amongst populations; hence the range in values. Adapted from (Piel and Weatherall, 2014). α^0-thalassaemia is normally found in certain areas of south-east Asia, although may also rarely occur in the Mediterranean regions. α^+-thalassaemia is also common in south-east Asia.

Rarely, large deletions (encompassing 1-2Mb) are seen involving the short arm of chromosome 16, which not only remove the α-globin genes and their regulatory elements, but also a number of surrounding genes (Wilkie et al., 1990). Affected individuals also have a variety of facial and skeletal abnormalities, associated with developmental delay. This is referred to as the ATR-16 Syndrome (OMIM #141750).

Mutations in a chromatin associated protein, ATRX, located on the X-chromosome, leads to a rare inherited form of mental retardation (ATR-X Syndrome OMIM #301040) associated with α-thalassaemia (Gibbons, 2012). α-thalassaemia may also occur as an acquired disorder in the setting of myelodysplastic syndrome associated with mutations in the chromatin remodeller ATRX (OMIM #300448); the haematological phenotype associated with this is often severe (Gibbons, 2012). Exactly how mutations in ATRX impact upon α-globin expression remains unclear and is an active area of research.

There are also more unusual mechanisms by which α-globin transcription may be reduced, despite all the *cis*-acting enhancer elements and α-globin genes being intact. One such example is where a single nucleotide variant leads to a new promoter-like element, thereby leading to reduced interactions between the enhancer elements and the α-globin genes (De Gobbi et al., 2006). A separate example is where a deletion leads to the juxtaposition of a widely expressed gene (LUC7L) with a normal α-globin gene; anti-sense transcription via the LUC7L gene then leads to aberrant methylation of the CpG island associated with the promoter of the α-globin gene with associated silencing (Tufarelli et al., 2003). Thorough

characterisation of such rare, typically non-coding mutations have provided significant insights into mammalian gene regulation more generally. This has helped cement the position of α-globin as a model locus for investigating the normal principles of transcription and gene regulation and how this is perturbed in human genetic disease. By far the most common mutations clinically however, are deletions.

1.5 Clinical characteristics of α-thalassaemia and the Haemoglobin Bart's Hydrops fetalis syndrome (BHFS)

The majority of patients harbouring α-globin gene deletions are asymptomatic. Deletion or inactivation of one or two (either in *cis* or *trans*) α-globin genes, commonly referred to as α-thalassaemia trait, leads to a mild microcytic anaemia (Piel and Weatherall, 2014). No treatment is required in such cases, apart from recognition of the cause, and avoidance of unnecessary investigations and treatment (for example iron supplementation). Deletion or inactivation of three genes (--/α-) leads to HbH disease, named for the abnormal β_4 tetramer, visible as sub-cellular inclusions. Patients with HbH disease usually express <30% of the normal α-globin levels and display a variable phenotype with the anaemia ranging from transfusion dependent to asymptomatic (Piel and Weatherall, 2014). Severely affected patients may require regular transfusions to avoid extramedullary haematopoiesis, splenomegaly and growth retardation. Should a regular transfusion program be initiated, it typically leads to iron overload with its associated endocrine, cardiac and hepatic complications and iron chelation is therefore recommended (Piel and Weatherall, 2014).

One particular deletion leading to α^0-thalassaemia, ζ --^{SEA} occurs at a high frequency throughout Southeast Asia (~3-14% depending upon the region). Homozygotes for this mutation have the haemoglobin Bart's Hydrops Fetalis Syndrome (BHFS), a lethal form of anaemia (Chui and Waye, 1998; Songdej et al., 2017). In some regions of Southeast Asia, as many as 1:200 conceived infants have this condition (Lemmens-Zygulska et al., 1996). It is thought that this balanced polymorphism results from a strong selective advantage for the ζ --^{SEA} allele in such areas.

Individuals with BHFS have no α-globin. Non-functional homotetramers of fetal gamma globin chains (γ_4; Hb Bart's) make up most of the Hb in the erythrocytes of affected infants. They are profoundly anaemic with an average Hb of 64 g/l; embryonic Hb Portland ($\zeta_2\gamma_2$), which is the only functional oxygen carrying haemoglobin in these infants, accounts for approximately 20% of this (Li et al., 2010).

The cardinal clinical features of BHFS are those of a pale oedematous infant with signs of cardiac failure and prolonged intra-uterine anaemia. Affected fetuses have pronounced hepatosplenomegaly, retardation in brain growth, a wide range of developmental abnormalities and there is gross enlargement of the placenta. Untreated children with the BHFS almost always either die, either *in utero* (typically between 23-38 weeks) or shortly after birth. This condition also carries a serious risk, including mortality, to the mother of the affected fetus, with up to 50% developing severe pre-eclampsia, at least in part due to placentomegaly (Jatavan et al., 2018).

1.6 Prevention and treatment of the BHFS and outcomes in long-term survivors

Given the high frequency of BHFS in people of Southeast Asian origin, the lethal degree of anaemia, the associated developmental abnormalities and the risks to the life of pregnant mothers, prevention of this condition is of paramount importance. The most important diagnostic criteria to detect carriers of all types of thalassaemia are microcytosis (MCV <80fL) and/or hypochromia (mean corpuscular haemoglobin [MCH] <27pg) (Chui and Waye, 1998). When such indices are found in a pregnant female this should prompt a full diagnostic algorithm to determine if she carries thalassaemia and if so which type (Ryan et al., 2010).

The partner of any carrier with α-thalassaemia should be tested to determine if they carry α-thalassaemia and the risk of the fetus having BHFS established. Subsequent fetal diagnosis includes ultrasonography, chorionic villus sampling or fetal blood sampling and termination if appropriate (Chui and Waye, 1998; Ryan et al., 2010). Successful screening programs are region and hospital dependent, with some centres reporting excellent compliance and almost complete prevention of BHFS (Lai et al., 2018).

Although many affected fetuses have the expected features of the BHFS, there can be considerable variability in the intrauterine clinical course. Some pregnancies, particularly those resulting from large α-thalassaemia deletions including the embryonic ζ-gene (-/--), terminate unnoticed or early in gestation. Others continue almost to term and are stillborn or critically ill at birth. Some children do not become

hydropic and are born spontaneously (Chan et al., 2018; Songdej et al., 2017). The environmental and genetic factors responsible for the diverse clinical courses of BHFS are currently unknown.

With improving methods of intra-uterine diagnosis and neonatal care of critically ill new-born infants there are an increasing number of reports of affected fetuses being treated with intra-uterine transfusion and/or neonates being treated by intensive care and transfusion. This first such case occurred in 1986 (Beaudry et al., 1986) and a recent review identified and reported the outcome in 69 affected individuals (Songdej et al., 2017). Such cases normally arise because of the parents' wish to continue with the pregnancy despite the diagnosis, or because the condition is not diagnosed until birth. To date, 39 of the 69 patients have survived beyond the age of 5 years; the oldest surviving patient is in her mid-30s. Typically, affected pregnancies require intrauterine transfusions to remain viable (although this isn't always required) and once born, affected individuals require life-long intensive blood transfusion programs and appropriate iron chelation (Amid et al., 2016). The only curative option remains allogeneic bone marrow transplantation and this has been successfully performed in at least 16 patients (Songdej et al., 2017).

Developmental abnormalities are commonly seen in BHFS, likely due to hypoxia during organogenesis and fetal development. Many abnormalities have been reported including microcephaly, hydrocephaly, hypoplasia of the lungs, hypospadias and cardiac defects (Songdej et al., 2017). Since anaemia develops as soon as the ζ-α switch occurs, developmental abnormalities may occur prior to diagnosis and implementation of intrauterine transfusion. Even with optimal post-

natal treatment, growth retardation and a degree of neurodevelopmental delay are often observed (Songdej et al., 2017).

Advances in bone marrow transplantation and fetal medicine have led to the possibility of intrauterine bone marrow transplants. There is currently a Phase I clinical trial using maternal (haploidentical) donors in fetuses affected by severe α-thalassaemia, with the donor cells administered between 18 and 25 weeks, in combination with a blood transfusion (https://clinicaltrials.gov/ct2/show/NCT02986698). Whether the lack of conditioning and immunosuppression will permit adequate engraftment and allow recipients to become transfusion independent is unknown. There are three previous reports using intrauterine bone marrow transplants for severe α-thalassaemia; none demonstrated successful engraftment or had positive outcomes (Derderian et al., 2014). Considerable efforts are being made to improve outcomes in fetal transplant, given this is a potential treatment for numerous inherited disorders besides haemoglobinopathies. One interesting strategy involves depletion of endogenous haematopoietic stem cells using an antibody directed against c-kit (Witt et al., 2018) and this shows promise in pre-clinical animal models. It remains to be seen however, whether this will translate successfully into clinical practice.

1.7 Reactivation of ζ-globin as a potential therapy

1.7.1 Bart's Hydrops Fetalis Syndrome

Given the majority of mutations leading to BHFS leave the ζ-globin gene intact, it may be possible to activate this gene in the definitive erythroid cells of affected patients. Production of ζ-globin in sufficient quantities would then take the place of the absent α-globin peptides, and form functional, oxygen carrying haemoglobin.

Evidence from pre-clinical models suggests this is a valid hypothesis. Transgene experiments in a mouse model of severe α-thalassaemia indicate that continued expression of ζ-globin is sufficient to rescue a lethal phenotype (Russell and Liebhaber, 1998). A somewhat artificial transgene, consisting of the human α-globin promoter, the human ζ-globin coding sequence and four elements of the human β-globin locus control region was inserted into mice heterozygous for deletion of their endogenous α-globin genes; they were then crossed to homozygosity. Mice with the transgene were identical to wild-type controls, with a normal life-span and normal litter size, despite having no functional α-globin.

Expression of ζ-globin in adult definitive cells would mean the main functional haemoglobin was Portland II ($\zeta_2\beta_2$) assuming the embryonic and fetal β-like globins were unperturbed. The latter would not necessarily hold true depending upon the approach used to increase ζ-globin however, given there are common pathways responsible for repression of both embryonic and fetal globins at both loci.

The functional characteristics of Portland II have been examined, again using mouse models (He and Russell, 2001). Complex transgenic mouse models were created which expressed both mouse and human haemoglobins (including Portland II), and the human haemoglobins were then purified and subjected to detailed *in vitro* physiological and biochemical studies. Portland II has a higher oxygen affinity and a lower subunit co-operativity than HbA; it also appears somewhat more unstable than HbA using thermal, chemical and mechanical perturbations. However, it still appears to be functional. Haemoglobin Portland I ($\zeta_2\gamma_2$) is more stable than Portland II, although it still has a higher oxygen affinity than HbA. Importantly, other naturally occurring mutations in the globin chains, associated with a higher oxygen affinity than Portland II are asymptomatic (Wajcman and Galacteros, 1996) suggesting that the higher oxygen affinity would be unlikely to have significant clinical sequalae.

1.7.2 Sickle cell anaemia

Sickle cell anaemia (SCA), occurring due to a homozygous single missense point mutation [Glu6Val, rs334] in the β-globin gene [HBB]) is a serious world-wide health problem with up to 300,000 affected infants born annually (Piel et al., 2017; Ware et al., 2017), leading to a global population of tens of millions of affected individuals. The deoxygenated haemoglobin molecule in affected individuals self-assembles into polymeric fibres that distort erythrocyte morphology into abnormally shaped sickle-cells (a process termed sickling). These fibres cause haemolytic anaemia and vasocclusion, with both acute and chronic consequences (Piel et al., 2017; Ware et al., 2017). Despite significant improvements in supportive care, life

expectancy for patients with SCA is still reduced by approximately 30 years, and there are considerable efforts to develop new therapeutic strategies.

In vitro, ζ-globin appears to have anti-sickling properties when as few as 10% of the α-globin subunits binding to the abnormal sickle β-globin molecule are replaced by ζ-globin chains (He and Russell, 2004). In a mouse model of SCA, ζ-globin normalises erythrocyte morphology, reduces haemolysis, and restores normal end-organ (renal) function when co-expressed as a transgene without significant negative effects upon oxygen transportation (He and Russell, 2004). The mechanism underlying this has not been fully elucidated but may involve destabilisation of the HbS polymers formed in sickle cell disease (Safo et al., 2015). These findings, which have not been validated in human studies, point towards a potential novel combined therapeutic strategy for sickle cell disease: delivery of a pharmacological agent which targets a common pathway responsible for repression of both ζ-globin and the normally silenced β-like γ-globin, thereby leading to a greater clinical benefit than increasing either alone.

1.8 Chromatin structure

In eukaryotes, DNA is packaged into chromatin. The base unit of chromatin is the nucleosome, which consists of 147bp of DNA wrapped around a histone octamer composed of two pairs of core histones: H2A, H2B, H3 and H4 (Zhang et al., 2015). Nucleosomes are further compacted into higher order chromatin structures by histone chaperones and nucleosome remodelling factors (Zhang et al., 2015). Positioning of nucleosomes throughout the genome modifies the availability of

binding sites to transcription factors, and ultimately therefore, to transcriptional machinery (Radman-Livaja and Rando, 2010); by contrast, nucleosome depleted regions permit access. “Open” chromatin assays such as DNase I followed by high-throughput sequencing (DNase-seq) (Boyle et al., 2008) and the more straightforward assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Buenrostro et al., 2013) allows genome-wide mapping of nucleosome depleted regions, and thus identification of putative regulatory regions of the genome including boundary elements, promoters and enhancers.

Nucleosome packing is controlled by modifications of amino-acids within particular histone proteins by enzymes and ATP-dependent chromatin remodelling complexes which directly move nucleosomes. ATP dependent chromatin remodelling machines typically occur as multiprotein complexes: for example the Nucleosome remodeling and deacetylase (NuRD) complex (Varga-Weisz, 2001). These redistribute nucleosomes along the DNA template using energy derived from the hydrolysis of ATP (Saha et al., 2006). A multitude of enzymes may modify histones with at least 130 different post-translational histone modifications reported (Tan et al., 2011). Transcriptionally active and silenced genes carry different histone modifications, which are usually found on the N-terminal tails of core histones although modifications have also been reported within the globular structure of histones (Zhang et al., 2015).

Active gene promoters normally carry high levels of lysine acetylation on the H3 and H4 tails and trimethylation of the fourth lysine residue on H3 (H3K4me3). Marks associated with repressed genes include trimethylation of lysine 27 on H3

(H3K27me3) and mono, di and trimethylation of lysine 9 on H3 (Zhang et al., 2015). In the context of gene regulation, acetylated lysine residues lose their positive charge, reducing interactions between the negatively charged DNA and the positively charged histones, and therefore allowing a more accessible environment (Zhang et al., 2015). Other marks associate with enhancers including H3K27ac (acetylation of lysine 27 in H3), and H3K4me1 (monomethylation of lysine 4 in H3). Histone modifications genome wide for any given cell type can be mapped throughout the genome using chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) (Barski et al., 2007). Although the exact function of many histone modifications is not completely understood, there is ample evidence that their deposition is required for normal gene expression. Common modifications associated with enhancers, promoters and repressed genes are shown in Figure 1.7.

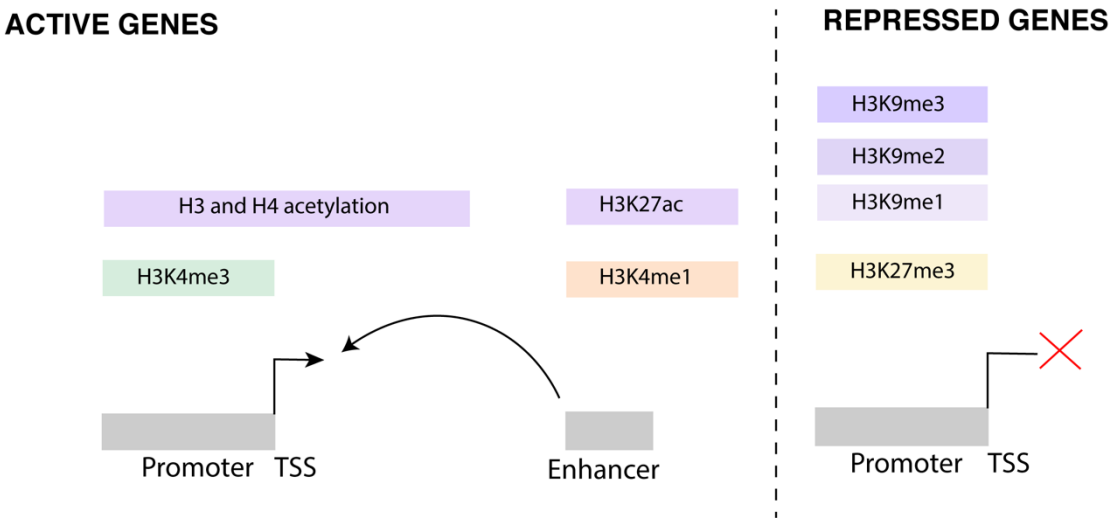


Figure 1.7 Histone modifications at active gene promoters, enhancers and repressed genes. TSS; transcriptional start site.

1.9 Gene silencing

Understanding how genes may be silenced is a key question in biology. Whilst there is much still to learn, a number of key mechanisms have been delineated over the last four decades. Broadly, there may be modifications to histones, which lead to chromatin compaction via recruitment of protein complexes; separately, the DNA itself may be modified through methylation.

Histone acetylation, which facilitates gene activation, is highly reversible, and deacetylation is therefore a mechanism by which genes may be silenced. Lysine residues are acetylated by histone acetyltransferases and are removed by histone deacetylases (HDACs). To date, 18 different mammalian HDACs have been identified, and have been divided into four sub-classes based upon their sequence similarity to their yeast counterparts (Kelly and Cowley, 2013). The sub-family with the most definitive role in regulating mammalian gene is Class I; this contains HDACs 1, 2, 3 and 8 (Kelly and Cowley, 2013).

HDAC1/2 are present in the mammalian nucleus (although there is some evidence that they may be found in the cytoplasm), and are always found in multiprotein complexes; they are incapable of binding DNA directly, and therefore their function is defined by the complex they are incorporated into. If they are not contained within a complex, their activity is minimal. The best characterised HDAC1/2 complexes are the NuRD, Sin3A and CoREST complexes (Kelly and Cowley, 2013; Seto and Yoshida, 2014).

Mammalian cells encode two homologues of SIN3 (SIN3A/B) which associate individually with HDAC1 and HDAC2, RBBP4 and RBBP7, SDS3 and SAP18 and SAP30; SAP stands for SIN-associated protein. Other studies have identified additional binding partners (Lai and Wade, 2011; Laugesen and Helin, 2014).

The NuRD complex contains the ATP-dependent CHD3/4 helicase subunit, HDAC1 and HDAC2, and additional component including the scaffolding proteins GATAD2AB, MBD2/3 (which mediates binding to DNA), and MTA1/2/3 (which mediates binding to transcription factors).

The CoREST complex incorporates HDAC1 or HDAC2, the scaffolding protein CoREST and lysine specific demethylase 1 (LSD1), alongside additional subunits including Sox-like-protein, ZNF217 and BHC80 (Laugesen and Helin, 2014).

The NCoR/SMRT complex associates with HDAC3, GPS2 (a G-protein pathway suppressor) and the transducing β-like factors TBL1 and TBLR1(Stanya and Kao, 2009).

The key feature of these complexes is that they serve as scaffolds for assembling catalytic activities at relevant genomic loci. Traditionally they are described as co-repressors, leading to gene inactivation via their histone deacetylation activity and chromatin modelling. All have been implicated in repression of γ-globin at the β-globin locus(Xu et al., 2013) via interactions with the transcription factor BCL11A.

A separate well characterised repressive complex is the polycomb repressive complex which is found throughout metazoan species; in mammals the multiprotein complex is referred to as PRC2. PRC2 contains the core components EZH2 (or its closely related homolog EZH1), EED and SUZ12; association with the histone chaperone RBBP4/7 is also necessary for activity in vivo (Margueron and Reinberg, 2011). PRC2 catalyses trimethylation of H3K27; although the precise function of this histone modification remains unclear, it correlates with gene silencing and associates with other gene silencing pathways including H3K9me3 (Margueron and Reinberg, 2011). PRC2 and DNA methylation however, are mutually antagonistic (Bartke et al., 2010).

DNA methylation, whereby methylation is added to the fifth position of cytosine via a covalent bond, is one of the most studied and best characterised epigenetic modifications. It is found throughout many diverse eukaryotic lineages including all land plants, certain fungi and mammals (Zemach et al., 2010). Despite its ancient origins, however, it is not universally present; indeed, it is absent in some model organisms including *Drosophila melanogaster*, *Caenorhabditis elegans*, fission yeasts and bakers' yeasts (Zemach et al., 2010; Feng et al. 2010). In mammals, methylation is primarily restricted to the CpG context (whereby C lies adjacent to G). Globally, mammalian genomes are CpG deleted; however around 10% of CpGs lie in CG dense regions termed CpG islands (CGIs). CpG islands are defined as stretches of DNA 500–1500 bp long, with a CG: GC ratio of more than 0.6. They tend to be associated with TSS with about half annotated as promoters (Feng et al., 2010). CGIs at TSS are typically unmethylated; however, methylation is associated with silencing, either through direct inhibition of transcription factor binding or by

recruitment of methyl binding domain proteins that recruit chromatin modifying activities to DNA (Klose and Bird, 2006). Methylation is controlled by enzymes responsible for de novo methylation (DNMT3A and DNMT3B), enzymes responsible for removal of methylation (TET1, TET2, and TET3), and a set of factors which copy methylation during DNA replication (DNMT1 and UHRF1) (Klose and Bird, 2006).

In organisms ranging from the model fission yeast *Schizosaccharomyces pombe* to human, constitutive heterochromatin which is compacted and transcriptionally repressed, is marked by H3K9me1/2 and H3K9me3 (Becker et al., 2016). In mammals, these modifications are catalysed by five members of the SET-domain containing family of methyltransferases. SETDB1 and the related enzymes SUV39H1 and SUV39H2 contribute to both H3K9me2 and H3K9me3 modifications (Rea et al., 2000) while GLP and G9a (also known as EHMT1 and EHMT2, respectively) catalyse H3K9me1 and H3K9me2 (Tachibana et al., 2005). These marks are also associated with gene repression in facultative heterochromatin. H3K9me2/me3 are bound by Heterochromatin protein 1 (HP1) which can self-oligomerize and recruit repressive histone modifiers, contributing to heterochromatin spread and compaction (Bannister et al., 2001).

These different silencing mechanisms are summarised in Figure 1.8. Although presented separately, there may be interplay between them, for example DNA methylation can be targeted by H3K9me3 (Rose and Klose, 2014).

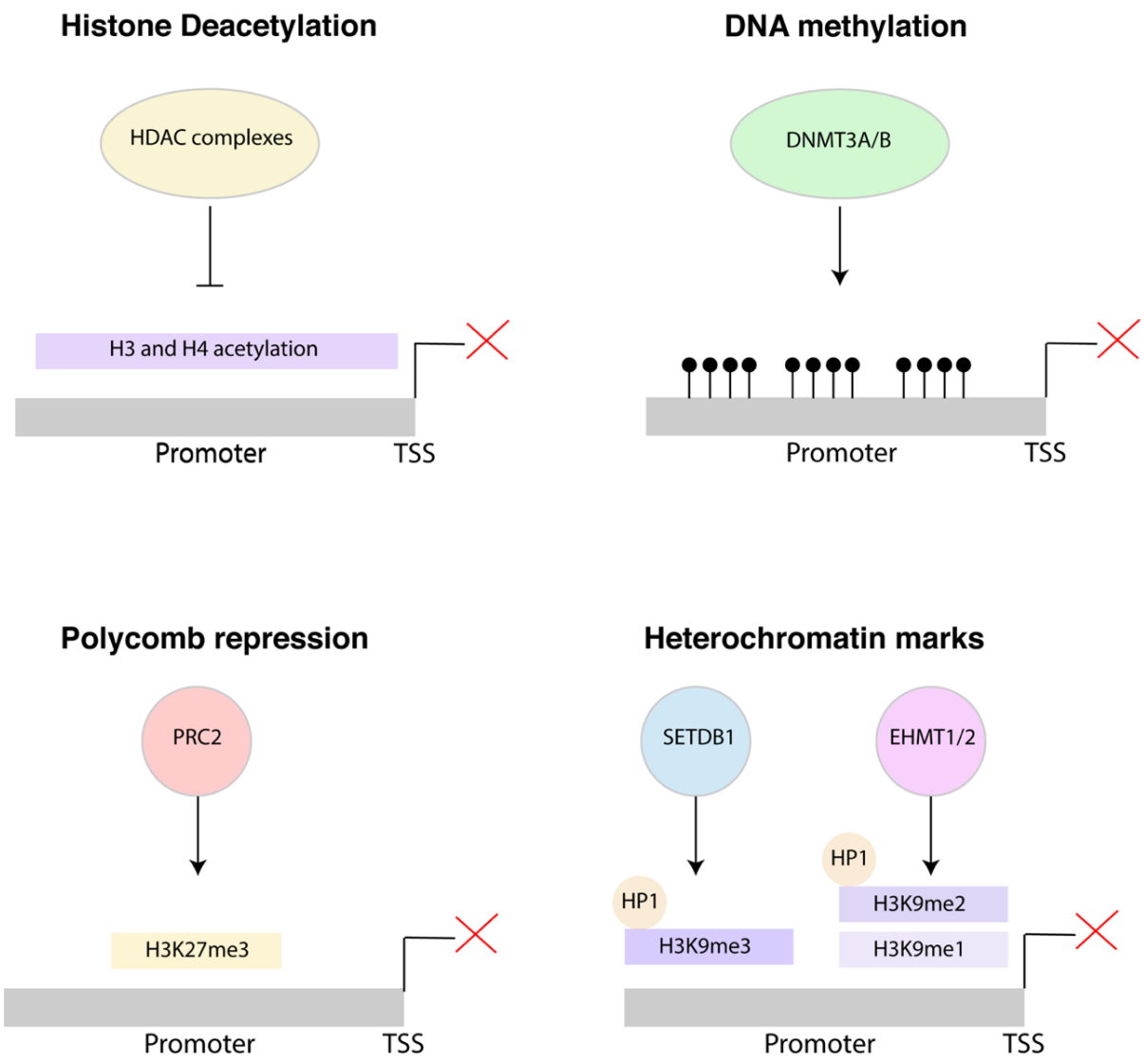


Figure 1.8 Gene silencing mechanisms. The associated histone marks and enzymes or complexes responsible for mediating silencing are shown.

1.10 Regulation of α-globin expression in definitive erythropoiesis in humans

1.10.1 Structure of the α-globin locus in humans

In humans, the α-globin gene cluster lies close to the telomere of chromosome 16 (16p13.3) in a conserved region of constitutively open chromatin surrounded by house-keeping genes. Expression of α-globin is encoded by the paralogous *HBA2* and *HBA1* genes. *HBZ*, which encodes ζ-globin expression, lies 5' of the two α-globin genes. Two other coding genes lie within the cluster; these produce a detectable transcript but no functional protein (Goh et al., 2005; Leung et al., 1987). *HBM* (μ) lies between *HBZ* and *HBA2*; *HBQ1* (θ) lies 3' of *HBA1*. Interestingly, both *HBM* and *HBQ1* are expressed at a higher level in fetal erythroid tissue in comparison to adult (Goh et al., 2005; Leung et al., 1987). *HBQ1* therefore appears to disobey the general rule of co-linearity whereby genes are arranged along the chromosome in the order in which they are expressed during development. Two pseudogenes are also present, termed ψζ (encoded by *HBZP1*) and ψα (encoded by *HBAP1*); these are homologues of the ζ-globin and α-globin gene respectively and contain mutations that render them incapable of producing a functional protein (Harteveld and Higgs, 2010). The complete order of the globin genes, including the pseudogenes is 5′-ζ-ψζ-μ-ψα-α2-α1-θ-3′.

Expression of α-globin in adult tissue is governed by four highly conserved *cis*-regulatory elements (or enhancers) (Hughes et al., 2005) termed multi-species conserved sequences (MCS) R1-R4. A comparative sequence analysis of 22

species has identified a conserved syntenic region of ~135kb which contains the α-globin gene, the α-like genes, and conserved regulatory elements. Three of these elements lie within introns of the upstream *NPRL3* gene (R1-R3). Presumably, the embedded regulatory elements have ensured the conserved synteny between *NPRL3* and the α-globin genes which is observed throughout vertebrate evolution (Hardison, 2012; Hughes et al., 2005). The precise function of NPRL3 remains uncertain, although evidence points towards a role in the mTOR pathway (Kowalczyk et al., 2012a). Knockout of *NRPL3* in mice leads to cardiac abnormalities and embryonic lethality late in gestation.

The MCS elements correspond to DNase I hypersensitive sites (DHSs) which in human are termed HS-48 (R1), HS-40 (R2), HS-33 (R3) and HS-10 (R4). The numbers indicate the distance in kilobases 5' of the *HBZ* mRNA cap.

Mutations in humans combined with laboratory transgene experiments indicate that R2 is the most important regulatory element for α-globin transcription in adult erythroid cells. All naturally occurring mutations which cause α-thalassaemia by deleting the up-stream regulatory elements include R2 (Harteveld and Higgs, 2010). R2 is also the most highly conserved of the enhancer elements throughout evolution (Hughes et al., 2005). In a humanized mouse model, where the entire mouse α-globin locus was replaced with the human locus and regulatory elements, deletion of the 1.1kb region containing R2 resulted in a reduction of human α-globin expression to ~2% of the wild-type chromosome (Wallace et al., 2007). R2 is also capable of driving expression of α-globin alone in transgenic mouse models and cell lines (Higgs et al., 1990; Sharpe et al., 1992).

Whilst R2 is critical for normal α-globin expression, in its natural chromosomal environment, it is not essential. The report of a patient homozygous for a 3.3kb deletion of R2 with insertion of a small 39nt orphan sequence (Coelho et al., 2010) demonstrated that α-globin expression was maintained at approximately 30% of normal (Mettananda et al., 2017a). Importantly, the 39nt sequence had no binding motifs for general or erythroid specific transcription factors. Examination of another patient demonstrated that deletion of R1 and R2 together leads to expression of α-globin at a level <1% of normal on the affected chromosome (Viprakasit et al., 2006). It therefore appears that R3 and R4 have a minimal positive effect on α-globin expression (assuming the enhancer elements act independently). The location of a patient where R1 alone is deleted would confirm or refute its presumed importance in α-globin expression in humans; such a mutation however, has not yet been described.

1.10.2 Transcription factors

Transcription factors bind to DNA sequences, thereby controlling tissue-specific expression of genes. *Cis*-acting regulatory regions and promoters contain multiple binding motifs whose order, orientation and spacing play a critical role in regulating gene expression (Spitz and Furlong, 2012). In erythroid cells, key transcription factors including GATA-binding factor 1 (GATA1), stem-cell leukaemia (SCL), transcription factor E2-alpha (E2A), Lim-only 2 (LMO2) and LIM domain binding protein 1 (Ldb1) are recruited to all four up-stream enhancer elements together with RNA polymerase II (Pol II) (De Gobbi et al., 2007). However, only R2, and to a

lesser extent, R1, bind the key erythroid transcription factor NFE2 in primary human erythroblasts (De Gobbi et al., 2007). Similarly, Krueppel-like factor 1 (KLF1) only appears to bind to R2 based upon ChIP-seq data (Norton et al., 2017). NFE2 is necessary for Pol II recruitment to the promoter at the β-globin locus (Johnson et al., 2001). It is likely that a similar mechanism exists at the α-globin locus, whereby all enhancers serve to recruit Pol II and other general transcription co-factors and multiprotein complexes (such as p300 and Mediator) and then NFE2 (via MCS-R2) is responsible for its transfer to the α-globin gene promoters and subsequent transcription (Vernimmen et al., 2009). During this process, there are physical interactions between enhancers and promoters.

1.10.3 Epigenetic regulation and histone modifications

The human α-globin gene cluster's chromatin environment has been well characterised. The locus is highly GC rich, with multiple CpG islands associated with gene promoters and gene bodies; these CpG islands are unmethylated in all tissues examined (Bird et al., 1987). In non-erythroid tissues, the α-globin genes are silenced through PRC2 (Garrick et al., 2008). H3K27me3 is noted at the α-globin gene promoters in non-erythroid tissues; SUZ12 is also seen to bind (Garrick et al., 2008).

In erythroid cells, when the α-globin gene is activated, PRC2 is detached and the H3K27me3 chromatin mark erased. Demethylation of H3K27me3 appears to be mediated via the Jumonji domain-containing protein 3 (Vernimmen et al., 2011). The promoters of the α-globin genes then become marked by histone H3 lysine

trimethylation (H3K4me3), which is associated with active gene promoters (De Gobbi et al., 2007) through the action of the trithorax group of proteins (in particular Set/MLL histone methyltransferases) (Schuettengruber et al., 2011). Together with the appropriate tissue specific transcription factors, this then drives α-globin transcription.

In erythroid cells, each of the enhancer elements are marked by histone H3 lysine monomethylation (H3K4me1) which is associated with active enhancers (De Gobbi et al., 2007). There is also a broad domain of histone acetylation extending across the locus (De Gobbi et al., 2007; Anguita et al., 2001) encompassing the α-globin genes and their enhancer elements. Both of these histone modifications are absent in non-erythroid tissues.

1.11 Regulation of α-globin gene expression in mice

Given the ease with which the mouse may be genetically manipulated, significant insights have been made regarding the regulation of α-globin expression by studying the endogenous mouse α-globin locus; in particular with regards to the higher order chromatin structure which forms when the α-globin genes are active, and the contribution of each enhancer element to α-globin transcription.

1.11.1 Differences between the murine and human α-globin loci

In mice, the α-globin locus is located on chromosome 11, and is surrounded by the same house-keeping genes as found in the human locus (Vernimmen, 2014).

Whilst the basic structure of the α-globin locus is conserved, there are a number of differences in comparison to that found in the human.

As well as the duplicated, coding α-globin genes (encoded by the genes *Hba-a1* and *Hba-a2*) and the embryonic ζ-globin gene (*Hba-x*) the mouse α-globin locus has two α-like genes (θ-globin) which produce a low level of transcript in adult erythroid cells (but no functional protein), and three pseudogenes, which produce no transcript. The θ-globin genes are encoded by *Hbq-1b* and *Hbq-1a*. The coding genes are arranged in the order: 5'-ζ-α1-5'θ-α2-3'θ-3' with the two θ-genes designated as 5' or 3' depending upon their location (Vernimmen, 2014).

In addition to the four conserved MCS-R elements, the mouse has an additional, fifth element, located at HS-12 outside the *Nprl3* gene; this is denoted as Rm (the mouse specific regulatory element) (Hughes et al., 2005). The MCS-R elements are located at HS-31 (R1), HS-26 (R2), HS-21 (R3) and HS-8 (R4) (Vernimmen, 2014).

The most prominent difference between the mouse and human α-globin loci is the absence of the large CpG islands present at the human α-globin gene promoters (Hughes et al., 2005); it is widely recognised that genome wide, rodents have fewer CpG islands than humans (Antequera and Bird, 1993) due to greater CpG island loss through evolution. In mice, polycomb repression appears to play no role in the silencing of the α-globin genes in non-erythroid cells (Garrick et al., 2008) in contrast to the human locus.

1.11.2 The α-globin enhancer cluster classifies as a super-enhancer in murine erythroid cells with the individual enhancers acting independently

Super-enhancers are groups of putative enhancers (as defined by associated histone marks such as H3K4me1 or H3K27ac) lying in close genomic proximity with unusually high levels of Mediator binding measured by ChIP-seq (Whyte et al., 2013). Their definition appears to be somewhat arbitrary: individual enhancers (typically identified by histone signatures with or without binding of tissue specific transcription factors) are "stitched" together if they lie within 12.5kb of each other, and are then ranked by their degree of Mediator binding or H3K27ac (Whyte et al., 2013). Those with the highest Mediator binding or H3K27ac (determined by an arbitrary, mathematically derived cut-off) are designated as super-enhancers. The original description, and numerous papers since, have proposed that super-enhancers act as switches to determine cell fate, and may form a functional unit greater than the sum of their parts (Whyte et al., 2013).

The α-globin enhancer cluster meets the definition of a super-enhancer in murine erythroid cells following classification of enhancers by histone signatures and erythroid transcription factor binding; indeed it is the second highest ranked enhancer cluster based on Mediator binding (Hay et al., 2016). Using a series of homozygous knockout mouse models, where each α-globin enhancer element was deleted individually and then in two combinations (R1/R2 and R2/R3) it was determined that the enhancers act additively. Removal of R1 and R2 caused α-globin transcription to fall by 40% and 50% respectively with their combinatorial knockout leading to a drop of 90%. Removal of R4 caused α-globin transcription to

fall by 10%; removal of Rm and R3 had no effect, with removal of R2 and R3 together having the same effect as R2 alone (Hay et al., 2016). Mathematical modelling confirmed no clear "higher-order" effects above those seen for individual enhancers (Dukler et al., 2016). Careful *in vivo* exploration of other multi-enhancer clusters in a similar fashion will help delineate whether "super-enhancers" are a discrete class of element with co-operative higher order properties.

1.11.3 The α-globin locus forms a self-interacting domain in erythroid cells flanked by CTCF sites

Chromosome Confirmation Capture (3C) and associated technologies have allowed visualisation of physical contacts between genomic elements in nuclear space, averaged across populations of cells (Dekker et al., 2002). All 3C-based methods operate on the principle that regions of interacting chromatin can be cross-linked, cut and then re-ligated; this then means that DNA regions lying adjacent to each other in the nucleus are joined together. High throughput sequencing followed by appropriate bioinformatic analysis allows the interacting regions to be determined (Davies et al., 2017). Hi-C, which generates contact maps between all parts of the genome (all vs all), demonstrates that mammalian chromatin is organized into active and inactive domains (A and B compartments). The A compartment contains regions of chromatin which are gene-dense, highly expressed and DNase I accessible; the B compartment is relatively inaccessible and gene poor (Lieberman-Aiden et al., 2009).

Within each compartment, chromosomes are organized into structural domains, termed topologically associated domains (TADs) (Dixon et al., 2012) of approximately 1Mb. These are highly self-interacting regions of chromatin delimited by domain boundaries. Within TADs, there are further structures termed "sub-TADs" or "contact domains" (Rao et al., 2014); these vary in size between 40kb and 3Mb, with a median size of 185kb. Thus, chromatin architecture is fractal, consisting of progressively larger interacting domains. TADs are usually considered to be cell type invariant (Dixon et al., 2012) whilst contact domains may vary between different types (Rao et al., 2014).

The highly conserved zinc-finger protein CCCTC-binding factor (CTCF) is a key factor in establishing mammalian architecture, acting as a boundary element (Dixon et al., 2012; Oudelaar et al., 2017). It has 11 zinc fingers, and binds a non-palindromic motif, allowing its orientation of binding to be determined. CTCF is enriched at the boundaries of self-interacting domains (including TADs and sub-TADs), typically in convergent orientations. It is now clear that the orientation of CTCF sites plays an important role in forming functional boundaries (Gómez-Marín et al., 2015; Vietri Rudan et al., 2015; Guo et al., 2015).

Self-interacting domains appear to form via loop extrusion, whereby a loop extrusion factor, containing two DNA binding units, associates with chromatin and travels in opposite directions, forming a progressively larger loop until the factor is stalled at the CTCF boundary elements (Sanborn et al., 2015). Cohesin, a ring shaped protein complex with ATPase activity, is currently the lead candidate for the loop extrusion factor (Wendt et al., 2008).

More recently, an additional model of transcription regulation has been described termed phase separation (Hnisz et al., 2017). This proposes that transcriptional control is driven by the formation of phase separated condensates, whereby enhancers recruit and concentrate large numbers of transcription factors and co-activators leading to separate liquid phases for activation domains (Hnisz et al., 2017). This compartmentalisation increases the specificity of interactions whilst still allowing a dynamic contacting process between promoters and enhancers. Evidence from live cell imaging experiments supports this model (Boija et al., 2018; Cho et al., 2018) in both mammalian cells and yeast. Heterochromatin has also been reported to form discrete phase-separated droplets, mediated by HP1 (Larson et al., 2017). Polymer modelling supports the combination of both phase separation and loop extrusion as an explanation for how chromatin is organised (Nuebler et al., 2018).

At the α-globin locus, there are a series of convergent CTCF sites focused upon the α-globin enhancers in definitive erythroid cells (Hanssen et al., 2017); (shown in this thesis in Figure 3.2). Using NG-Capture-C (Davies et al., 2016), which provides high resolution chromatin confirmation capture data in a one vs all approach, an erythroid-specific self-interacting domain and its boundary elements have been observed. Two CTCF sites, located 38 and 39kb upstream of the *Hba-x* mRNA cap (called HS-38 and HS-39) form the upstream boundary, with two other CTCF sites (located 44 and 48kb downstream), called HS44 and HS48 potentially forming a downstream boundary. The self-interacting domain (Figure 1.9) isn't seen in embryonic stem cells (ESCs), when the α-globin genes are transcriptionally silent. Removal of the upstream CTCF elements allows the α-globin enhancers to act promiscuously upon the normally silenced surrounding genes, including *Rhbdf1*

which is repressed by PRC2 in definitive erythroid cells (Hanssen et al., 2017), highlighting the importance of CTCF in constraining enhancer activity.

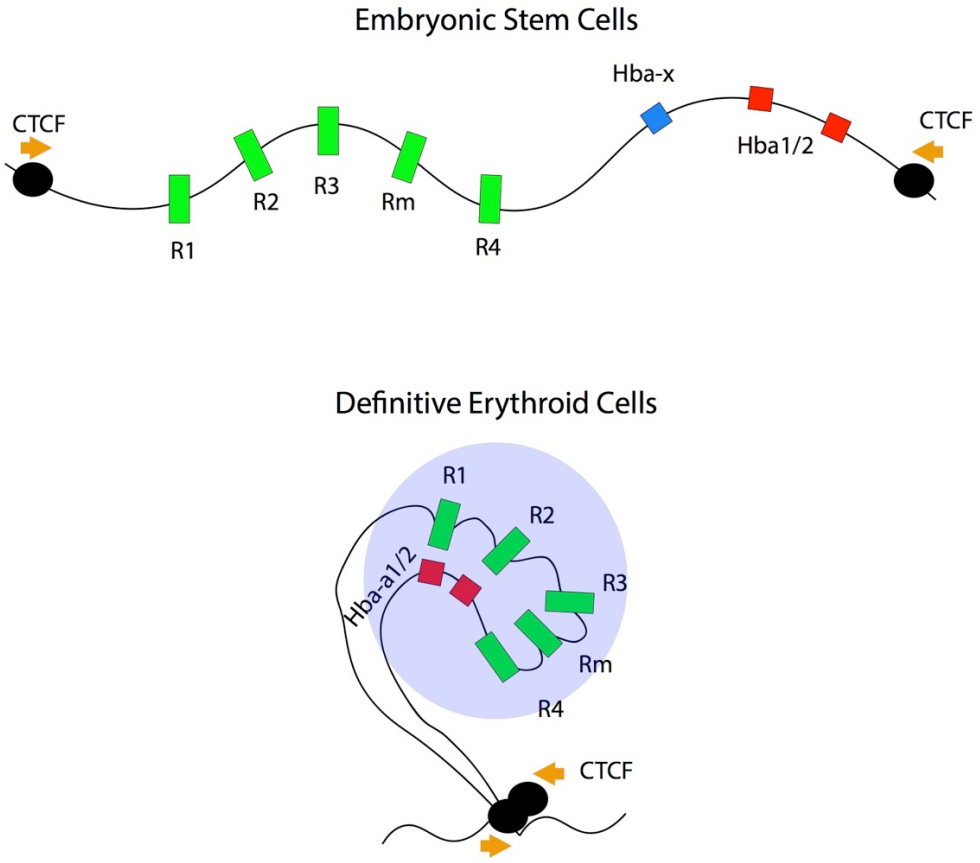


Figure 1.9 An erythroid specific self-interacting domain forms in erythroid cells. In mouse ES cells (top) no sub-TAD, or contact domain is formed. In definitive erythroid cells, a contact domain is formed, flanked by convergent CTCF sites (HS-38/HS-39 and HS44/HS48).

1.12 The regulation of ζ-globin in primitive and definitive erythropoiesis

1.12.1 Structure of the ζ-globin gene

The ζ-globin gene consists of three coding exons which are translated into a 142 amino-acid protein in both mice and humans. It is the first gene in the α-globin cluster in all species (Hughes et al., 2005), in agreement with the findings from other gene clusters that genes tend to be expressed developmentally in the same order as they are found linearly on a chromosome, a phenomenon termed co-linearity (Neijts and Deschamps, 2017). In humans, there is also a ζ-globin pseudogene lying 10.2kb downstream (*HZBP1*) which has presumably arisen via a gene duplication(Hardison, 2012); this differs by only 3bp in the coding exons from *HBZ*, one of which converts amino acid 6 into a premature stop codon (Proudfoot et al., 1982).

The introns of the ζ-globin genes differ from the adult α-globin genes. The introns of the α-globin genes are approximately 100bp long, with the first intron being slightly shorter than the second. By contrast, the first intron of the ζ-globin gene is 886bp long, and contains 12 repeats of a 14bp sequence (ACAGTGGGGAGGGG); intron 2 is 239bp long and contains 35 copies of a 5bp repeat (CGGGG). This results in the GC content of the second intron being 92%, and means there is an intragenic CpG island covering intron 2, and exon 3 of *HBZ* (Proudfoot et al., 1982). In mice, the second intron is much shorter (106bp) and the first intron, whilst being

of a similar size, does not have the same repeat, so it is unclear if these sequences have any functional role.

Comparison of the ζ-globin promoter across species indicates that it is approximately six times larger than the α-globin promoters (Hughes et al., 2005). It seems probable that there are specific binding sites for repressive transcription factors, which do not bind at the α-globin promoters. Transgene experiments provide evidence supporting this hypothesis. Substituting the human *HBA* promoter for the *HBZ* promoter leads to ζ-globin expression; the opposite experiment, where the *HBZ* promoter is substituted for the *HBA* promoter leads to decreased (though not absent) α-globin expression (Liebhaber et al., 1996).

Within the ζ-globin promoter, there are binding sites for GATA1 (Yu et al., 1990), CCAAT box binding proteins (presumably a site for NF-Y) (Yu et al., 1990) and the KLF transcription factor family (Sabath et al., 1996). It is highly likely that there are other, unidentified binding sites in addition to these. A series of deletions in the ζ-globin promoter in transgenes inserted into both animals and cell lines have produced conflicting results about the importance of different regions of the promoter necessary for activation and silencing (Sabath et al., 1995; 1993; Vyas et al., 1992); the precise sequences responsible for driving and repressing ζ-globin expression have consequently not yet been defined.

The 3' region of the ζ-globin gene is hypersensitive when it is transcribed in both K562 cells and in embryonic tissue (Yagi et al., 1986), and is also hypersensitive using the assay for transposase accessible chromatin with high-throughput

sequencing (ATAC-seq) in murine primitive erythroid cells (Songdej; unpublished, outlined in greater detail below). Given there is also a CpG island located at this site in *HBZ* (though not in *Hba-x*) it is possible that this represents a regulatory region. After truncations were introduced into this area in a transgene inserted into mice, ζ-globin expression was de-repressed in fetal and adult tissue (Liebhaber et al., 1996). It is important to note that the 3' region is not conserved between mice and human however, and it is unclear whether it plays a role in its natural chromatin environment.

1.12.2 The influence of cis-acting regulatory elements on ζ-globin expression

A previous student in the laboratory, Duantida Songdej, characterized the accessible chromatin at the α-globin locus and the contribution of each enhancer element to ζ-globin expression in murine primitive erythropoiesis at E10.5 (when ζ-globin accounts for approximately 40% of the total transcript from the α-globin locus).

ATAC-seq showed that all five *cis*-regulatory elements (R1-R4, and Rm) known to be present in definitive erythroid cells (Hay et al., 2016) are also found in primitive erythroblasts (Figure 1.10); no novel embryonic *cis*-elements were identified. Open chromatin was also seen over the ζ-globin gene, consistent with active transcription. A large peak is visible over the ζ-globin promoter and a further peak can be seen extending approximately 1500 bp beyond the 3' region of the gene. It thus appears that there are no embryonic specific *cis*-acting elements other than the promoter of the ζ-globin gene.

Using the same knockout mouse models as previously published (Hay et al., 2016), where the enhancers were removed individually and in two combinations, the contribution of each enhancer element to α and ζ-globin expression in primitive erythropoiesis were defined using Nanostring, which measures RNA transcripts without amplification bias (Geiss et al., 2008) (Figure 1.11).

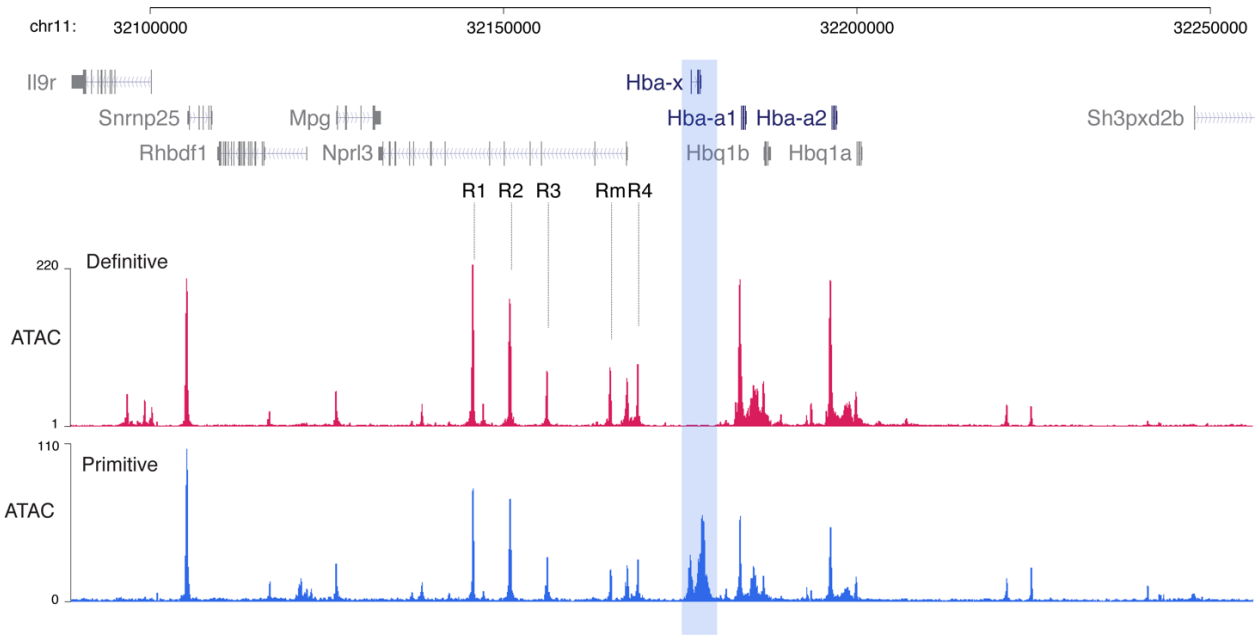


Figure 1.10 Chromatin accessibility at the α-globin locus in definitive and primitive erythropoiesis. The same enhancer elements are seen in primitive and definitive erythropoiesis. The ζ-globin gene (*Hba-x*) becomes accessible with a peak seen at the promoter, and a further large peak seen at the 3' region of the gene (highlighted in blue).

The effects on α-globin expression are very similar to those observed in definitive erythropoiesis (Hay et al., 2016), with R1 and R2 contributing 90% of the transcriptional output. In contrast, knockout of each enhancer element individually has no detectable effect on ζ-globin expression, and it is only when R1 and R2 are

removed together, that there is a significant reduction in ζ-globin expression. This implies that the enhancers act in a redundant manner in primitive cells. It seems probable that the extended ζ-globin promoter is capable of binding activating *trans*-acting factors directly and is therefore less reliant on the enhancers than α-globin (which has a much smaller, less complex promoter).

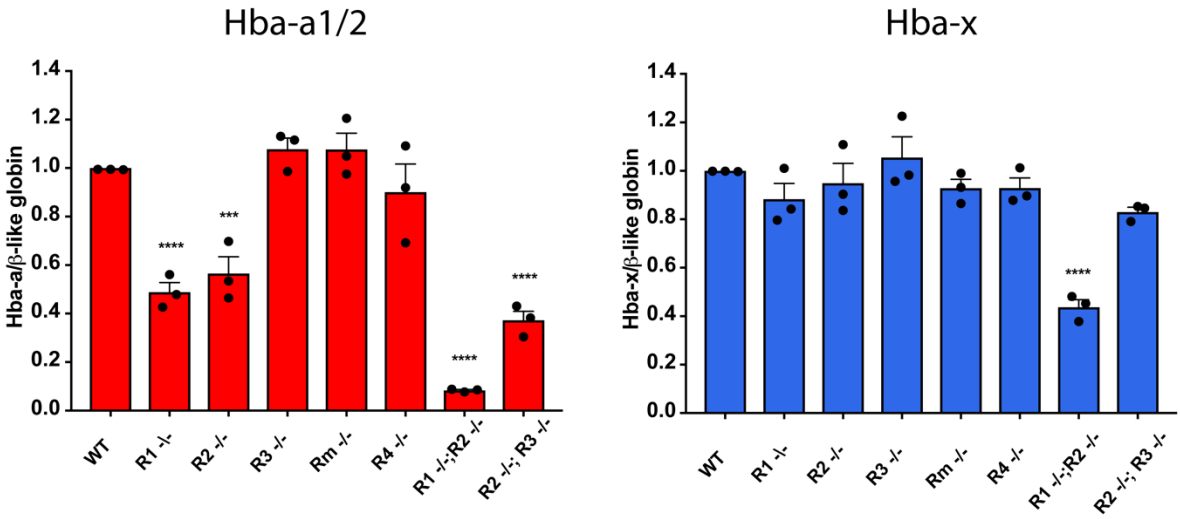


Figure 1.11 Nanostring RNA quantification of the ratio of *Hba-a1/2* and *Hba-x* to β-like globin transcripts in each of the enhancer knockout models. A significant reduction in the α-globin:β-globin-like transcript ratio is observed in $R1^{-/-}$, $R2^{-/-}$ and the $R2^{-/-}$; $R3^{-/-}$ double-knockout embryos in comparison to wild-type. A 90% reduction in the ratio is seen in the $R1^{-/-}$; $R2^{-/-}$ double-knockout embryos. The ζ-globin: β-globin-like transcript ratio is only significantly reduced (by 60%) in the $R1^{-/-}$; $R2^{-/-}$ double-knockout embryos. *** denotes p ≤0.001; **** denotes p≤0.0001; statistical significance was calculated using a one-way ANOVA with Dunnett correction for multiple comparisons.

In humans, the regulation of ζ-globin by the *cis*-acting elements is much less well understood. Transgene experiments in mice indicate that the human ζ-globin gene requires an enhancer element in *cis* for expression (either R2 or HS2 from the β-

globin locus are effective) (Sabath et al., 1993) and R2 is also required for expression of ζ-globin in the erythroleukemia K562 cell line in a transient expression assay (Zhang et al., 1993). The same group used the Bal31 nuclease assay to explore the important binding sequences with the R2 element: a GATA1 site, and two NFE2/AP1 sites along with a GT dinucleotide formed the functional core, necessary for ζ-globin activation (Zhang et al., 1993). The 3' NFE2/AP1 also appears be involved in ζ-globin repression; a single base pair mutation (which impairs NFE2 binding but leaves AP1 binding intact) permits ζ-globin promoter expression in a transgenic mouse model (Huang et al., 1998). The construct used in these experiments consisted of a ζ-globin promoter coupled to human growth hormone (as a readout of transcription) attached to HS-40 (Huang et al., 1998). The artificial nature of all these experiments means it is difficult to draw definitive conclusions, although it seems probable that ζ-globin requires an enhancer in *cis* for detectable expression.

Two observations from patients with deletional α-thalassaemias provide additional insights into ζ-globin silencing and activation. In adult patients with the ζ $--^{SEA}$ deletion on one chromosome, ζ-globin is detectable in small amounts (Chung et al., 1984). In patients who have this on both chromosomes (leading to BHFS), Hb Portland I ($\zeta_2\gamma_2$) makes up about 20% of the detectable haemoglobin (Li et al., 2010). This suggests that in the absence of the α-globin promoters, the enhancers can partly overcome the silencing of ζ-globin in definitive erythroid cells. However, this is not the case for all mutations, with for example, the larger ζ $--^{BRIT}$ and ζ $--^{SA}$ mutations having no detectable ζ-globin (Tang et al., 1993). This perhaps suggests that an active regulatory region necessary for ζ-globin expression in definitive

erythroid cells resides within a 5.2kb region present in the ζ --^{SEA} deletion but absent in the ζ-- ^{BRIT} and ζ -- ^{SA} deletions (Figure 1.5). It should be noted however, that no putative elements have been identified in this region using open chromatin assays.

1.12.3 Trans-acting factors responsible for ζ-globin repression

Over the last five years, some aspects of the transcriptional factor network responsible for embryonic ζ-globin gene repression have been elucidated, based predominantly on coincident work studying the fetal γ-globin to adult β-globin switch at the β-globin locus in both human cells and mice.

1.12.3.1 BCL11A

BCL11A is a multi-zinc finger transcription factor that was initially implicated in mediating the switch between γ-globin and β–globin at the β-globin locus based upon findings obtained from genome-wide association studies. The identified polymorphisms in putative regulatory elements for *BCL11A* are associated with increased levels of γ-globin (Lettre et al., 2008; Menzel et al., 2007; Uda et al., 2008). BCL11A also plays an important role in B-cell development (Yu et al., 2012) within the haematopoietic system, and neurodevelopment, with haploinsufficent individuals having an autism spectrum disorder (Basak et al., 2015; Dias et al., 2016). In adult erythroid cells, BCL11A is expressed as two major isoforms: XL and L at both the protein and RNA level (Sankaran et al., 2008).

Knockdown or knockout of BCL11A in human primary erythroid cells (Sankaran et al., 2008) and animal models (Sankaran et al., 2009; Xu et al., 2011) combined with the study of individuals haploinsufficent for BCL11A (Basak et al., 2015) have confirmed that it is an integral γ-globin silencer. Further work went on to identify an erythroid specific enhancer element regulating *BCL11A* expression in both the human and the mouse, and within this, a specific GATA1 binding motif (Bauer et al., 2013; Canver et al., 2015; Smith et al., 2016). This element offers a rational gene therapy target; if it is removed, it leads to minimal BCL11A expression in erythroid cells, with preserved expression in lymphocytes. Such an approach can now be performed with high efficiency, and can increase γ-globin levels to at least 30-40% in human cells (Wu et al., 2019). This has been shown to ameliorate the phenotype in both sickle cell disease and β-thalassaemia (Wu et al., 2019). Aside from its effects on the β-globin locus, there appears to be some evidence that BCL11A may play a role in ζ-globin expression based upon genome wide transcriptome data (Sankaran et al., 2009) when *Bcl11a* is knocked out in mice.

The consensus binding site for BCL11A, identified from ChIP-seq, CUT&RUN and *in vitro* DNA binding assays is TGACC, which is found in the γ-globin promoter adjacent to a CCAAT box. Disruption of this site leads to γ-globin derepression (Liu et al., 2018; Martyn et al., 2018; Traxler et al., 2016) and direct binding of BCL11A can be seen using CUT&RUN (Liu et al., 2018), and when BCL11A is tagged with an inducible estrogen receptor ligand-binding domain and a V5 epitope (Martyn et al., 2018). This changed the long-standing view that BCL11A mediates its effects via binding to intergenic regions (Xu et al., 2010) rather than directly to gene promoters.

1.12.3.2 LRF

The zinc finger transcriptional repressor LRF (also known as Pokémon, osteoclast-derived ZF, FBI-1, and ZBTB7A) is a member of the BTB/POZ-ZF family; it is known to bind directly to DNA as a homodimer and leads to transcriptional silencing by binding to repressive complexes (Lunardi et al., 2013). Together, LRF and BCL11A appear to be responsible for the entirety of γ-globin repression when knocked out together (Masuda et al., 2016). However, analysis of a conditional LRF knockout mouse model (Maeda et al., 2007) appeared to show that LRF deficient erythroblasts also may have considerable up-regulation of the ζ-globin gene with impaired erythroid differentiation (Masuda et al., 2016).

It is still somewhat uncertain how BCL11A and LRF lead to γ-globin gene repression. The leading hypothesis is that this occurs via recruitment of the NuRD complex. Both BCL11A and LRF co-associate with individual members of the NuRD complex (Masuda et al., 2016; Xu et al., 2010). It appears that BCL11A and LRF act independently based upon comparative ChIP-seq data (Masuda et al., 2016) despite both binding NuRD.

A recent study explored the roles played by each of the 13 possible NuRD subunits. Only five genes, *CHD4*, *GATAD2A*, *HDAC2*, *MBD2* and *MTA2*, were required for HbF repression, suggesting that a sub-complex of NuRD is responsible for repression (Sher et al., 2019). Notably within five paralogous gene families (*CHD*, *GATAD2*, *HDAC*, *MBD* and *MTA*), only one member was required for HbF

repression, this suggests that other members of each family are dispensable. It was possible to target specific sequences within the CHD4 domain which de-repressed γ-globin without leading to significant cellular toxicity (Sher et al., 2019). Clearly, therefore, NuRD plays a significant role in mediating γ-globin silencing.

Although it has previously been noted that LRF and BCL11A may play a role in ζ-globin repression in passing, neither of these observations have been followed up in detail to determine the degree of their effect on ζ-globin transcription.

1.12.3.3 KLF1

KLF1, the founding member of the 17 member family of KLF transcription factors, is a master erythroid transcription factor defined by the presence of three similar C_2H_2 zinc fingers at the C terminus that comprise its DNA binding domain; it is responsible for a wide activation repertoire including genes involved in protein-stabilizing, the heme biosynthetic pathway, the red cell membrane and the cell cycle (Siatecka and Bieker, 2011; Yien and Bieker, 2013). A small number of individuals with a variety of biallelic KLF mutations and elevated ζ-globin expression have been described (Arnaud et al., 2010; Viprakasit et al., 2014a). These cases also have raised γ-globin expression, associated with chronic nonspherocytic haemolytic anaemia. All of these cases have now been shown to harbour bi-allelic mutations in the gene encoding KLF1 causing deleterious effects in DNA binding, thus providing clear *in vivo* evidence that KLF1 plays a role in embryonic globin gene repression.

KLF1 activates both BCL11A (Zhou et al., 2010) and LRF (Norton et al., 2017) and it is therefore possible that at least some of the effects of KLF1 are mediated via these two transcription factors.

KLF1 also transcriptionally activates two transcriptional repressors from the KLF family: KLF3 and KLF8 which bind MCS-R2 and the α-globin promoter and exert their effects via recruitment of the CtBP protein complex (Turner and Crossley, 1998), which in turn causes epigenetic silencing via histone modifications (Shi et al., 2003). A study involving knockout of KLF3 and KLF8 individually and in combination in mouse models demonstrated that both appear to be involved in repressing the embryonic globin genes (including ζ-globin) with each partially compensating for the other when individually knocked out; KLF3 also inhibits *KLF8* expression suggesting a potential fine-tuning mechanism (Funnell et al., 2013). A further option aside from upstream effects of BCL11A, LRF, KL3 and KLF8 is that KLF1 exerts additional effects by binding directly within the α-globin locus (Tallack et al., 2010).

1.12.3.4 SOX6

SOX6, a member of the Sox transcription factor family characterized by the conserved high mobility group domain appears to interact physically as a protein complex with BCL11A; at least at the β–globin locus (Xu et al., 2010). A *Sox6*-null mutant mouse has been shown to express ζ-globin (Yi et al., 2006) and it therefore also appears that Sox6 may also play a role in embryonic globin gene repression.

1.12.3.5 c-myb

c-myb, which is dispensable in primitive erythropoiesis but essential for definitive erythropoiesis, is a known *trans*-activator of *KLF1* (Bianchi et al., 2010), and has been implicated in silencing the embryonic and fetal genes at the β-globin locus (Lettre et al., 2008). The simplest model implied by this is that c-myb acts directly through KLF1 to influence BCL11A and LRF; this however, ignores the multitude of other transcriptional networks affected by c-myb and it is likely there are other effects.

Patients with trisomy 13 (Patau syndrome) have elevated fetal and embryonic globins (Sankaran et al., 2011). Chromosomal mapping of partial trisomy cases has implicated two micro-RNAs (miR-15a and -16-1) as causative: a combination of functional and genetic studies have demonstrated that these down-regulate c-myb, providing additional information on the factors controlling the c-myb-KLF1 pathway.

Given c-myb is a *trans*-activator of *KLF1*, it may also have effects on ζ-globin although this has never been proven.

1.12.3.6 POGZ

Pogz, a zinc finger binding transcription factor, has recently been shown to repress both embryonic and fetal globins (including ζ-globin) in mouse erythroid cells (Gudmundsdottir et al., 2018). Knockout of *Pogz* in mice leads to embryonic lethality at E16.5; analysis of definitive erythroid cells taken from the fetal liver at

E14.5 shows embryonic globin de-repression. Embryonic globin de-repression was separately shown in adult mice with a conditional *Pogz* knockout.

Knockdown of POGZ in CD34 human cells using shRNA also led to fetal globin depression; the effects on *HBZ* weren't reported. BCL11A is decreased when *POGZ* is knocked out suggesting that POGZ activates BCL11A. However, BCL11A over-expression doesn't repress the embryonic globins completely in an adult murine erythroid cell line following Pogz knockdown. This implies that POGZ represses the embryonic globins via additional unknown mechanisms.

1.12.3.7 LIN28B

The childhood leukaemia JMML has a particular sub-type, characterised by high levels of fetal and embryonic globin expression (Helsmoortel et al., 2016; Lee et al., 2013; Papayannopoulou et al., 1991). This sub-type has been shown to have high levels of LIN28B, an RNA binding protein. LIN28B blocks biosynthesis of a micro-RNA called let-7 (Mayr and Heinemann, 2013), which in turn leads to increased translation of let-7 target mRNAs. One study has indicated that this LIN28B-let7 pathway may be important in degrading BCL11A mRNA (Lee et al., 2013) although another study noted no difference in BCL11A expression levels (Helsmoortel et al., 2016). What is clear however, is that both LIN28B and the let7 micro-RNAs play a role in repressing embryonic and fetal globins.

1.12.3.8 GATA1

GATA1 is expressed in primitive and definitive erythroid cells (Fujiwara et al., 1996; Leonard et al., 1993) and is essential for normal erythroid development with *Gata1* null mice dying of anaemia at E10.5 (Fujiwara et al., 1996). Whilst initially identified as a protein which binds to the *cis* elements of the β-globin gene, it has since been shown to regulate a large number of erythroid-specific genes (Ferreira et al., 2005). There are at least two canonical GATA1 binding sites located within the ζ-globin promoter; these have been shown to bind GATA1 *in vitro* (Yu et al., 1990) in K562 cells. GATA1 sites are also found in the α-globin enhancers, and may therefore contribute to ζ-globin expression.

GATA1 can be activating or repressive and it is therefore reasonable to assume that it may play a role in both activating and repressing ζ-globin in primitive and definitive erythropoiesis respectively, dependent upon its binding partners. The exact mechanism by which this may occur however, is unclear. A partner which is normally associated with repression in combination with GATA1 is Friend of GATA1 (FOG1), and this has also been shown to bind and recruit the NuRD complex (Hong et al., 2005). However, disruption of the Fog1/NuRD interaction has no effect on embryonic globin gene silencing (Miccio and Blobel, 2010), at least in mice. LRF also occupies GATA1 bound regions, although the combination of the two factors can be activating or repressive, when looking genome-wide (Yu et al., 2009). Clearly more work is needed to understand this.

1.13 Transcriptional differences between primitive and definitive erythropoiesis

Global gene expression databases of stage-matched, primitive, fetal definitive and adult definitive erythroblasts in mice have been generated using microarrays (Kingsley et al., 2013). This impressive dataset offers additional insights into the transcriptional network governing embryonic globin gene repression. Importantly, only a few of the transcription factors currently known to mediate embryonic globin gene silencing are differentially regulated between primitive and definitive erythropoiesis.

c-myb is up-regulated 22 fold in definitive erythroid cells; this is in keeping with c-myb knockout mouse models, which have preserved primitive erythropoiesis, but absent definitive erythropoiesis (Tober et al., 2008). Similarly, Bcl11a and Sox6 are essentially absent, and therefore dispensable, in primitive erythropoiesis (Kingsley et al., 2013). However, *Klf1*, *Lrf*, *Pogz*, *Klf3*, and *Lin28b* are all expressed in both primitive and definitive erythroid cells (Kingsley et al., 2013), despite playing a clear role in ζ-globin repression.

Clearly then, repression is more complicated than simple expression of a protein with subsequent binding and recruitment of repressive complexes. It is very likely that there are unknown *trans-acting* factors which act in concert with those already described; potentially there may also be differing post-translational modifications of transcription factors in primitive or definitive erythroid cells which affect their function.

1.14 Erythroid cellular models

Experiments in cells and cell lines using *in vitro* culture systems form an integral part of scientific research, allowing rapid manipulation of specific genes or cellular signalling pathways.

For the last four decades, the cellular model systems available for studying erythropoiesis have consisted of erythroleukemia cells lines such as mouse erythroleukemia cells (MEL cells) and K562 cells, derived from a patient with chronic myeloid leukaemia in blast crisis (Lozzio and Lozzio, 1977; 1975). Whilst K562 cells have provided some insights into aspects of human globin gene regulation, they have an extremely abnormal karyotype, and express an embryonic/fetal pattern of globin expression (Benz et al., 1980; Rutherford et al., 1981). Given they are derived from a malignant myeloid clone, are aneuploid and carry multiple mutations these cells do not recapitulate normal erythropoiesis.

Both human embryonic stem (ES) cells and induced pluripotent stem (iPS) cells may be differentiated towards erythroid cells, using a variety of protocols (Chang et al., 2006; Lapillonne et al., 2010; Olivier et al., 2016; 2006). iPS cells in particular offer unique advantages, specifically, the ability to reprogram cells from patients with rare mutations, allowing an essentially unlimited supply of erythroid cells to be generated for subsequent experiments. However, culture protocols are complex, costly and labour intensive with the reprograming of cells and subsequent erythroid differentiation taking months. Erythroid cells derived from both ES cells and iPS cells also typically have an embryonic/fetal transcription profile (Olivier et al., 2016;

Vanuytsel et al., 2018) reducing their use as a model for definitive erythropoiesis and globin switching.

Primary CD34+ cells, derived from peripheral blood, bone marrow, or umbilical cord may be differentiated to reticulocytes, and are considered the gold standard for studying adult or fetal erythroid cells (Migliaccio et al., 2002; Ronzoni et al., 2008). However, their use is limited by restricted access to donors, their expense to procure and their short-term culture potential, the latter meaning that genetic manipulation followed by clonal expansion to large numbers of cells is not feasible (Mettananda et al., 2017a).

Recently, a new definitive erythroid cell model has been described. The human umbilical cord blood derived erythroid progenitor cell line (HUDEP) was established by immortalisation of umbilical cord CD34+ cells using lentiviral delivery of a tetracycline- inducible HPV16-E6/E7 transgene (Kurita et al., 2013). The HPV16 E6 and E7 proteins induce immortalisation by inhibiting the p53 and retinoblastoma (pRb) tumour suppressor genes respectively (Münger and Howley, 2002). Three HUDEP clones were generated from three different cord blood samples; the HUDEP-1 and HUDEP-3 clones had a fetal globin expression pattern, expressing γ-globin, whilst the HUDEP-2 clone had an adult globin expression pattern, expressing β-globin (Kurita et al., 2013). Importantly, the HUDEP-2 cell line has a relatively straightforward differentiation protocol, and cells can be genetically manipulated and then grown clonally.

Given these advantages, the HUDEP-2 cell line was consequently rapidly adopted by the scientific community as a model for definitive erythropoiesis, with at least 18 publications using this cell line since 2013. Despite this, there has been relatively little characterization of the cell line beyond the original publication(Kurita et al., 2013), with for example, the karyotype (which is abnormal) only being reported last year (Vinjamur and Bauer, 2018).

In 2017, a similar cell line to the HUDEP-2 cell line was published. This used the HPV16 E6/E7 transgene to immortalize adult, rather than fetal CD34+ cells, producing a cell line called BEL-A cells (Bristol Erythroid Line-Adult) (Trakarnsanga et al., 2017). This appears to recapitulate adult erythropoiesis, and the cells enucleate, in comparison to HUDEP-2 cells, although their karyotype is again abnormal (personal communication; Jan Frayne). These cells have not yet been used by other laboratories to study globin switching.

1.15 The genome editing revolution

Recent advances in genome editing technology allow scientists to now manipulate almost any sequence in cell culture systems and *in vivo* in model organisms. The broad principle of genome editing uses nucleases targeted to a particular region, which then induce a double stranded break (DSB). This is then repaired via one of three pathways: the error prone non-homologous end-joining (NHEJ), microhomology mediated end joining (MMEJ), or homology directed recombination (HDR) (Adli, 2018; Hsu et al., 2014; Sakuma et al., 2016). NHEJ typically leads to small indels, whereas HDR, with supply of a donor template, allows precise

sequence changes to be introduced. MMEJ harnesses machinery independent of HDR. When short microhomologies exist, both upstream and downstream of the DSB site, the two microhomologies can be annealed by MMEJ, often resulting in deletions of the intervening sequence (Adli, 2018). Efficiencies of HDR, NHEJ and MMEJ vary between different cell types and organisms with NHEJ normally being the most efficient.

In order to create a DSB, DNA must be cut with a nuclease. The first such tools were zinc finger nucleases, which linked several zinc finger domains (Cys_2-His_2) with the FokI restriction enzyme (Kim and Kim, 2014). Limited targeting density, binding site prediction and efficiencies limited the widespread use of these by basic scientists, although, given they were the first useable nucleases, they were adopted by some groups and used successfully in clinical trials (Tebas et al., 2014).

The development of transcription activator like effector nucleases (TALENs), which combined a different class of DNA binding protein (transcription activator like effectors [TALEs] from the plant pathogenic Xanthomonas spp. bacterium) with the *Fok*I nuclease, marked a major improvement in both targeting efficiency and versatility. TALENs are essentially unrestricted in terms of sequences they recognise, and modifications allow them to cleave methylated DNA, meaning any region in the genome can be targeted. However, the construction of DNA segments that encode TALE arrays can be both challenging and time-consuming, which again limited widespread adoption (Kim and Kim, 2014).

In 2012, the first report of the RNA- guided nuclease clustered regularly interspaced short palindromic repeat (CRISPR)- CRISPR-associated (Cas) system was published (Jinek et al., 2012). The most commonly used system remains the wild-type Cas9 protein derived from *S. Pyogenes,* targeted to a genomic location by a single-guide RNA (Hsu et al., 2014) – this is efficient in human cell lines, mouse embryonic stem cells and mouse embryos (as well as most other organisms). It is both straightforward and cheap to use, in comparison to its predecessors (TALENs and ZFNs) – this combination has meant that it has been widely adopted by scientists involved in both basic science and translational research.

There have been a variety of modifications to the wild-type Cas9 system such as improvements in specificity by mutating particular sites of the nuclease (Kleinstiver et al., 2016). Separately, the fusion of a dead Cas9 (dCas9, a protein without nuclease activity) to other enzymes, including those capable of altering histone modifications, DNA methylation and single bases, has provided a wide variety of tools for experimental (and potentially clinical) use (Adli, 2018). CRISPR/Cas9 can also be used to perform knockout screens genome wide with lentivirus to identify genes important for particular processes (Zhou et al., 2014). A similar approach has successfully been used to dissect *cis*-acting elements, identifying the crucial transcription binding factor sites within such elements at high resolution (Canver et al., 2015). Thus, over just seven years, CRISPR/Cas9 has revolutionized molecular biology and permitted major advances in a multitude of fields.

1.16 Summary and outline of the thesis

The ζ-globin gene lies adjacent to one of the strongest enhancer clusters in definitive erythroid cells within an erythroid-specific contact domain where all other genes within that compartment are active. Despite this, it is transcriptionally silent, and the chromatin around the gene is inaccessible. From the perspective of gene regulation, this presents an intriguing conundrum: what is the epigenetic mechanism underlying this silencing, and how is it mediated? Separately, there is a potential clinical benefit to reactivating ζ-globin, in that it could be used as a treatment for patients with severe α-thalassaemia: both BHFS and transfusion dependent HbH disease. This thesis attempts to understand how ζ-globin is silenced, with the ultimate aim of using this knowledge to develop new therapies.

In **Chapter 3** I have defined the epigenetic landscape and self-interacting domain when ζ-globin is active (in primitive erythropoiesis) and in definitive erythropoiesis (when ζ-globin is silenced) using the mouse as a model. I have used ChIP-seq to compare the histone modification signatures and CTCF binding between these two cell types, alongside NG-Capture C. This has allowed me to precisely delineate the erythroid compartment in primitive erythropoiesis and how contact with the enhancer elements and the ζ-globin promoter changes when it is active, and when it is silenced.

I have then studied the transcription factors previously implicated in ζ-globin silencing. In **Chapter 4** I have used mouse models where *Bcl11a* and *Lrf* have been

perturbed. I have also performed ATAC-seq to investigate how the chromatin landscape alters when these transcription factors are removed.

The HUDEP-2 cell line has rapidly become the *in vitro* cellular model of choice for studying globin regulation in human definitive erythroid cells. In **Chapter 5** I have characterised this cell line, and then used HUDEP-2 clones where BCL11A and LRF and have been knocked out to study how this affects the chromatin landscape (using ATAC-seq) and globin expression. I have also studied how the histone modifications change in the presence and absence of these transcription factors, and how DNA methylation at the ζ-globin gene changes.

There are currently only a handful of transcription factors known to be important in activating and repressing ζ-globin. The development of CRISPR library screens means that it is now possible to screen transcription factors involved in mediating a particular process, providing there is a robust phenotypic readout. To facilitate this, I have generated a new mouse model, where a fluorescent protein (mVenus) replaces exon 3 of the endogenous *Hba-x* gene. **Chapter 6** describes the development of this and initial characterisation. Because exon 3 of *Hba-x* is removed, this also knocks out ζ-globin when present in homozygosity; this has separately allowed me to determine whether ζ-globin is essential for development in the mouse. Finally, I have established a drug screening protocol for novel compounds using this mouse model, which will be performed in a collaboration with an industrial partner (AstraZeneca).

2 Materials and Methods

2.1 Cell Culture and related techniques

2.1.1 Culture of mouse embryonic stem cells

E14 mouse ES cells were maintained in GMEM (Life Technologies) supplemented with: FBS (10%, Gibco), sodium pyruvate (1mM, Life Technologies), non-essential amino acids (1mM, Life Technologies), Glutamine (2 mM, Gibco), Penicillin/Streptomycin (P/S; 100IU/ml, Gibco), 1000 IU/ml Leukaemia Inhibitory Factor (LIF, Millipore) and 0.1 mM β-Mercaptoethanol (Sigma-Aldrich) at 37°C with 5% CO_2 incubation on gelatin (0.1%, Sigma-Aldrich) coated flasks.

2.1.2 Differentiation of mouse embryonic stem cells

To differentiate ES cells into embryoid bodies (EBs), cells were initially passaged into IMDM supplemented with monothioglycerol (MTG, 0.15mM, Sigma-Aldrich), FCS (15%), LIF (1000IU/ml) and P/S (100IU/ml) for 24 hours. Cells were then plated into 10cm culture plates (Corning) at densities of 30,000-50,000 per plate. ES cells were suspended in IMDM supplemented with: 15% FCS (heat-inactivated), P/S (100IU/ml), MTG (0.45mM), ascorbic acid (5µg/ml, Sigma-Aldrich), glutamine (2mM), holotransferrin (300µg/ml, Sigma-Aldrich) and 5% Protein-Free Hybridoma II (Life Technologies). Differentiation proceeded for 7-8 days prior to analysis.

2.1.3 Primary Primitive Erythroblasts

Primary primitive erythroblasts were isolated from yolk sacs at E10.5 and used immediately for experiments. Because of the semi-synchronous nature of primitive erythroid maturation, these primary primitive erythroid cells were highly homogeneous at the morphological stage of intermediate erythroblasts; <5% of cells are EMP by size morphologically. Morphological examination of cells for each experiment was undertaken to ensure homogeneity.

2.1.4 Definitive erythroid cells derived from fetal liver

Definitive erythroid cells were obtained from a well-established fetal liver culture system (Drissen et al., 2005). Fetal livers were dissected from E12.5 mouse embryos, and the dissociated cell suspension was expanded for 5–7 d in StemPro (Invitrogen) supplemented with Epo (1 U/ml, OUH pharmacy), SCF (50 ng/ml, Peprotech) and dexamethasone (1 μM, Sigma-Aldrich). After expansion of the cultures, mature red blood cells were removed by negative selection for Ter119 using anti-Ter119 magnetic beads (Miltenyi Biotec) and an LD MACS column, to obtain a population of erythroid precursors, which was then purified by CD44 (IM7, BD Biosciences, 1:100) positive cell sorting, before being placed into expansion media to recover for 6 hours. Cultures were then switched to medium containing a high Epo concentration (5 U/ml) to induce differentiation into mature erythroid cells and harvested after 24 hours. Differentiation was confirmed by flow cytometry analysis for Ter119 and CD71, and morphological examination of cytospins.

2.1.5 Isolation and selection of definitive erythroid cells from spleens

Definitive erythroid cells were obtained from phenylhydrazine-treated mice (Spivak et al., 1973). Spleens were mechanically disrupted to a single cell suspension in ice-cold PBS. To isolate late-stage erythroid cells, cells from a single spleen were resuspended in 5mL of cold PBS/10% FCS and stained with 50uL PE anti-Ter119 antibody (RUO, BD Biosciences) at 4°C for 30 min in the dark. After washing with 40mL of ice-cold PBS/10% FCS, cells were resuspended in 800 µL MACS buffer (PBS, 2 mM EDTA, 0.5% BSA) and 200 µL of anti-PE magnetic beads (Miltenyi Biotec) and incubated for 20 min at 4°C in the dark. Ter119 positive cells were isolated via auto-magnetic-activated cell sorting (MACS) and then processed for downstream applications. Purity of cells was routinely confirmed by flow cytometry.

2.1.6 HUDEP-2 cell culture

WT HUDEP-2 cells, LRF$^{-/-}$, BCL11A $^{-/-}$ and DKO cells were obtained as a gift from Dan Bauer (Masuda et al., 2016); all clones had previously been confirmed as knockout by Western Blot by the donating laboratory. Cells were maintained in the expansion phase using StemSpan SFEM medium (Stem Cell Technologies) supplemented with: hSCF (50 ng/ml, Peprotech), erythropoietin (3 IU/ml, OUH hospital pharmacy), dexamethasone (10^{-6}M, OUH Hospital pharmacy), glutamine (2mM, Life Technologies), Penicillin/Streptomycin (P/S, 100IU/ml, Life Technologies) and doxycycline (2µg/ml, Sigma-Aldrich) at 37°C with 5% CO_2 incubation. Cell concentration was maintained between 0.5 and 1x10^6 cells/ml. The

media was changed every 48 hours whilst cells were in expansion. Cells were kept for a maximum of three weeks.

To induce differentiation, cells were transferred to IMDM (Life Technologies) supplemented with: Solvent/Detergent pooled AB plasma (2.5%, NHSBT), hSCF (50ng/µl), Epo (3IU/ml) holotransferrin (330µg/ml, Sigma-Aldrich), insulin (10 µg/ml, Sigma-Aldrich), heparin (2IU/ml, JR hospital pharmacy), P/S (100IU/ml) glutamine (2mM) and doxycycline (2µg/ml) at 37°C with 5% CO_2 incubation. Media was changed every 48 hours. Cell concentration was maintained between 0.5 and $1x10^6$ cells/ml. After seven days, doxycycline was removed from the media and cells were then cultured for a further 3 days prior to experiments. Adequate differentiation was confirmed using flow cytometry and cytospin.

2.1.7 Primary erythroid culture from $CD34^+$ cells

50-100ml of whole blood from healthy donors was collected into EDTA tubes. Full blood counts were performed (Pentra ES60, Horiba) on all samples to ensure adequate counts. Blood was diluted with PBS and overlaid onto Histopaque-1077 (Sigma-Aldrich) and centrifuged for 30 min at 630 *g*. Harvested peripheral blood mononuclear cells were washed in PBS and MACS buffer (PBS, 2 mM EDTA, 0.5% BSA) and stained with Human CD34 Microbeads kit (Miltenyi Biotec) following the manufacturer's instruction before column separation. Cells were frozen in $1x10^5$ cell aliquots with 90% FCS, 10% DMSO.

$CD34^+$ cells were differentiated to primary erythroid cells using a three-phase protocol requiring a common base media IMDM (BioScience UK Limited) containing 3% AB Serum (NHSBT), 10 µg ml^{-1}, insulin, 3Uml^{-1} heparin (all from Sigma-Aldrich-Aldrich) and 2% fetal bovine serum (Gibco) (Scott et al., 2019). In Phase I (expansion), cells were maintained at 1-3x10^5 cells/mL in base media supplemented with EPO 3U/mL (OUH hospital pharmacy), SCF 10ng/mL (Peprotech), IL-3 1ng/mL (Peprotech) and holotransferrin 200µg/mL (Sigma-Aldrich) for 7 days. Phase II (differentiation) media was the same as Phase I, with the omission of IL-3 and lasted 4 days; cells were maintained at 3x10^5/mL. Phase III media (enucleation) was supplemented with EPO 3u/mL, SCF 10ng/ML and holotransferrin 500µg/mL; cells were maintained at 1x10^6/mL. Cells were taken for experiments at Day 13, when the majority of cells are intermediate/late erythroblasts (Scott et al., 2019).

2.1.8 B16f10 cell culture

B16f10 mouse melanoma cells were maintained in RPMI (Life Technologies) with 10% FCS glutamine (2mM, Life Technologies) and Penicillin/Streptomycin (P/S, 100IU/ml, Life Technologies) and split 1:10 every second day.

2.2 Mouse methods and husbandry

2.2.1 Mouse maintenance and breeding

C57BL/6 were used as WT controls.

Breeding pair Lrf$^{Flox/Flox}$ mice were obtained as a gift from Pier Pandolfi (Maeda et al., 2007). These were crossed to a WT background to obtain heterozygous mice, then crossed with a *Gata1*-cre mouse line (Mao et al., 1999), which has leaky expression and thus causes efficient germ-line excision of floxed regions. *Lrf*$^{Flox/-}$ were then bred to produce litters with a combination of *Lrf*$^{Flox/Flox}$ (WT), *Lrf*$^{Flox/-}$(+/-) and *Lrf* $^{-/-}$ (-/-) litters.

A mouse with the composite enhancer block for Bcl11a (corresponding to 10kb) deleted (*Bcl11a*$^{\Delta 10kb}$) (Smith et al., 2016) was imported as a gift from Stuart Orkin. This was crossed onto a WT background to breed out a transgenic β-YAC human β-globin gene locus (Porcu et al., 1997).

Mice were housed in ventilated cages and all husbandry and procedures performed under Home Office Project License approval, in accordance with European Union Directive 2010/63/EU and the UK Animals (Scientific Procedures) Act, 1986.

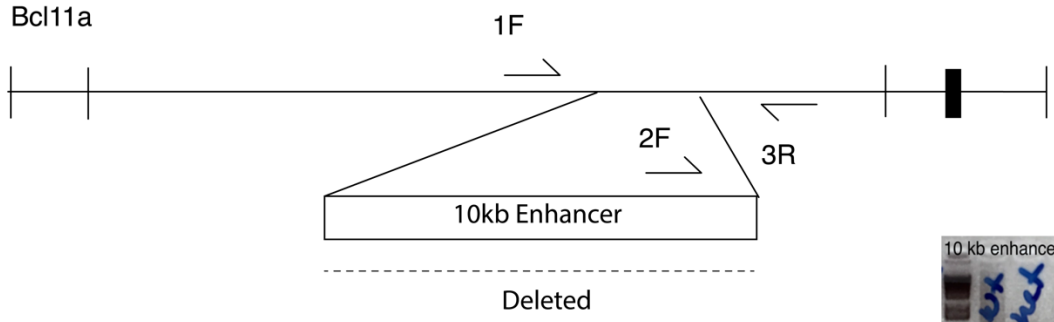
2.2.2 Phenylhydrazine treatment of mice

Phenylhydrazine treatment was carried out by Jacqueline Sharp or Jacqueline Sloane-Stanley. Adult mice were given 3 doses of freshly neutralised

phenylhydrazine (40mg/g body weight) by intraperitoneal injection 12 hours apart. Spleens were harvested on day 5 following the treatment.

2.2.3 Genotyping

Ear biopsies from mice were lysed in 95 µL DirectPCR Lysis Reagent (Ear) (Viagen) with 5 µL of PCR grade Proteinase K (20mg/mL; Thermo Scientific) by incubating overnight at 55°C, then 85°C for 45 min to inactivate Proteinase K. 1 µL of the lysate was used directly as a template for PCR reactions, with 12.5 µL 2x Platinum Hotstart Mastermix (Thermo Scientific), primers at 0.2µM concentration, DMSO 5% and DNAse free H_2O (Sigma-Aldrich) to make reactions up to 25 µL. PCR reaction conditions were performed according to the manufacturer's instructions with the extension time varied depending upon the PCR product size. The annealing temperature was optimised for each primer pair by running a temperature gradient. Primer sets used, and example gels are displayed below. All gels were run on 2% Agarose/TAE gel. Nucleic acid was visualized with 200 µg/ml ethidium bromide; images were taken using GelDoc equipment (Biorad). Primers were ordered from Sigma-Aldrich.



Primers

1F AGTGGTGTTTCCCAATGCTC
2F GGCCTCAGAATCTGGTCTCA
3R GAAAAGCAAGCCTCAACTGG

WT band 651bp
Enhancer knockout 393bp

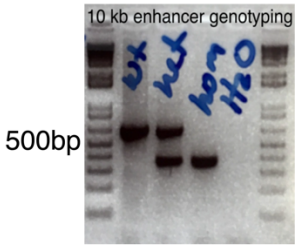


Figure 2.1 *Bcl11a*$^{\Delta 10kb}$ genotyping strategy

The mice imported had a transgenic human β-globin locus. Because experiments were only required on the endogenous β-globin murine locus, this was bred out to avoid confounding human globins. Primers for β-YAC were directed against the human HBG gene; primers for another locus (in this case the endogenous murine Hbb-b1 gene) was also added to ensure an absent band wasn't due to failure of the PCR reaction.

Primers

4F (HBG) GCAAGAAGGTGCTGACTTCC
5R (HBG) ATATGCCTTCTGCCTGCATC
6F (Hbb) ACGATCATATTGCCCAGGAG
7R (Hbb) ATGGCCTGAATCACTTGGAC

DNA control band 769bp (mouse Hbb)
β-YAC band 489bp (human HBG)

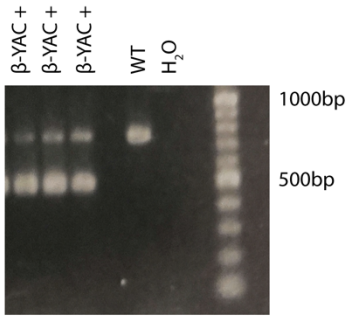


Figure 2.2 β-YAC genotyping strategy

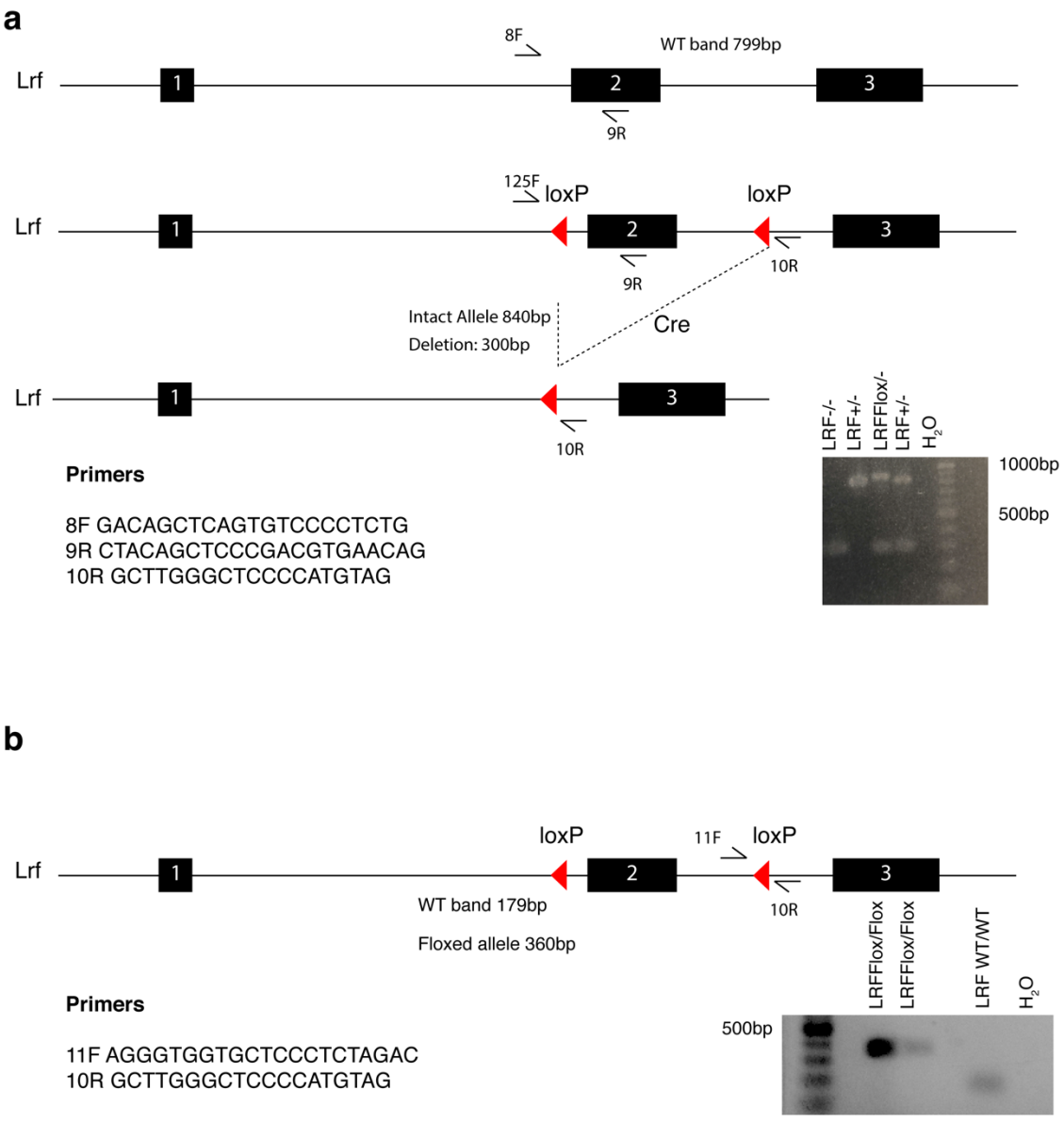


Figure 2.3 *LRF* flox and knockout genotyping strategy. Because the LRF^{Flox} and LRF^{WT} bands are difficult to resolve in (a), a different PCR was used in addition to determine whether the Floxed allele was present (b).

2.2.4 Haematological analysis

Haematological analysis was carried out by Jacqueline Sharpe. Mice were backcrossed on a C57BL/6 genetic background. Reticulocyte preparations were generated using BCB staining, and counts were made by two independent blinded assessors.

2.3 Genome engineering

2.3.1 CRISPR/Cas9 construct cloning

pSpCas9(BB)-2A-Puro (PX459) V2.0 was obtained as a gift from Dr Feng Zhang (Broad Institute) obtained via Addgene (plasmid # 62988). A pair of 24-mer oligonucleotides (Table 2.1, Sigma-Aldrich) containing the appropriate gRNA sequence were annealed with each other and cloned into the BbsI-BbsI site in PX459 V2.0. Plasmids were transformed into One-Shot Top10 Chemically Competent E.coli (Invitrogen) following the manufacturer's instructions. The following day, individual bacterial colonies were picked and grown in LB supplemented with ampicillin 50µg/ml overnight shaking at 220rpm at 37°C. Minipreps were performed using the Qiagen Plasmid Miniprep kit following the manufacturer's instructions. Correct integration was confirmed via Sanger Sequencing, performed by the WIMM core facility with the following primer: 5'-GACTATCATATGCTTACCG. Four different gRNAs were tested all targeting regions within exon3 of the *Hba-x* gene; the targeted sequence was obtained from the mm9 genome.

Oligonucleotide denoted name	Sequence
AKZetagRNA1F	5'-**CACC**GGAGTCGGCGGGAAAGCGTG
AKZetagRNA1R	5'-**AAAC**CACGCTTTCCCGCCGACTCC
AKZetagRNA2F	5'-**CACC**GGCTCAGGGGTGAAGTCGGC
AKZetagRNA2R	5'-**AAAC**GCCGACTTCACCCCTGAGCC
AKZetagRNA3F	5'-**CACC**GGCCTGAGGTCCACGAAGCC
AKZetagRNA3R	5'-**AAAC**GGCTTCGTGGACCTCAGGCC
AKZetagRNA4F	5'-**CACC**GGCCGCCGACTTCACCCCTG
AKZetagRNA4R	5'-**AAAC**CAGGGGTGAAGTCGGCGGCC

Table 2.1 Oligonucleotide primers used for CRISPR/Cas9 construct cloning. Highlighted sequences are those compatible with the BbsI restriction enzyme. 2Gs are placed before the guide RNA sequence because of the U6 promoter present in the PX459 plasmid.

2.3.2 Surveyor Assay (IDT)

B16f10 mouse melanoma cells were used to test guides due to their rapid growth and high transfection rate in RPMI (Gibco) with 10% FCS. $1x10^5$ cells were suspended in 1mL of RPMI and transfected with 500ng of PX459 via Lipofectamine 2000 (used according to manufacturer's instructions). Genomic DNA was extracted after three days and the genomic regions surrounding the CRISPR target site were PCR amplified. The following primers were used: forward 12F5'-AAGATCATGACCGCCGTAGGGGATGCG and Reverse 13R5' GAGCCTTCGCTTTGAACTCGCAGAACC. 10µl of PCR product with wild-type genomic DNA were annealed following the protocol specified by IDT. Following this, the 20µl product was added to 2 µl $MgCl_2$, 1 µl Nuclease S, 1 µl Enhancer S to form a final volume of 24 µl. After incubation in a thermocycler at 42°C for 50 min,

2.5µl of Stop Solution was added, the DNA run on a 1% agarose/TAE gel. gRNA demonstrated the best cutting (~10%), although B16f10 cells are tetraploid so this is likely to be an underestimate.

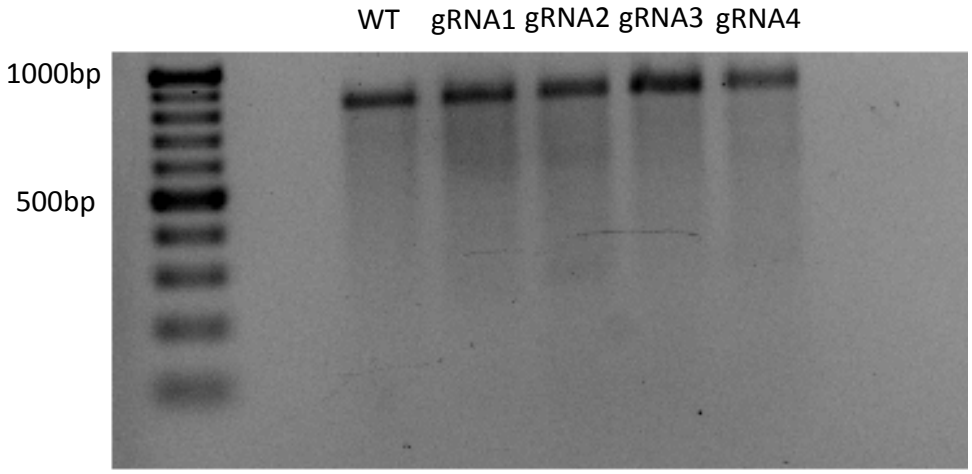


Figure 2.4 Surveyor assay. gRNA2 demonstrates the best cutting.

2.3.3 Cloning of the Hba-x-mVenus donor construct

This strategy was designed with Dr Philip Hublitz, of the core genome engineering facility in the WIMM. A construct containing homology arms of 600bp at the 5' end and 1.2kb at the 3'end, the 2A peptide, restriction enzyme sites for fluorescent protein/localizing signal insertion (SalI/EcoRI) and RE sites for an frt-kanamycin-frt (FKF, MfeI/MfuI) was ordered from GeneArt in a pMA plasmid. An FKF selection cassette was initially inserted by conventional RE cloning using MfeI/MfuI by Dr Hublitz. The mVenus/NLS sequence was inserted by PCR cloning from a previous plasmid in the lab (donated by Danuta Jeziorska) (using a high-fidelity enzyme) following cutting of the vector with SalI/EcoRI. The plasmid was transformed into One-Shot Top10 Chemically Competent E.coli (Invitrogen) and sanger sequenced in its entirety.

2.3.4 Transfection of mouse embryonic stem cells

Passage 14 E14Tg2a cells were plated at a density of 150,000 onto a gelatin (0.1%) coated six well plate. Plasmid HDR donor/PX459 circular DNA in a 1:1 molar ratio were added to 230µl OptiMEM (Life Technologies) with 15µl TransIT-LT1 (Mirus) and incubated at room temperature for 15 mins, prior to being added drop-wise to cells. After 24 hours culture, cells were split in a range of limiting dilutions (1:10-1:160) over six well plates. Puromycin selection (Sigma-Aldrich) was added at 1.5µg/ml 48 hours after transfection and left for 48 hours. G418 (Life Technologies) was then added at 400µg/ml for 72 hours. Media was changed at least every 48 hours whilst cells were undergoing antibiotic selection. WT cells were also cultured with antibiotics as a control to ensure efficacy. Colonies were picked into a gelatinised 96 well plate after 72 hours G418 selection and grown until confluent. Following isolation of a correct clone using the PCR strategy and Southern blot strategy outlined in Chapter 4, the cells were Sanger sequenced. Primer sequences and southern blotting probe sequences are shown below (Table 2.2). Probes were prepared by Helena Ayyub. Following conformation that a targeted clone was correct, it was karyotyped by the transgenic facility in the WIMM and injected into blastocysts to make chimeras by Jaqueline Sharpe; further details and figures are included in Chapter 6.

Sequence 5’ to 3’	Function
F ACGTAAACGGCCACAAGTTC	Forward primer Venus
R CGTTGTGGCTGTTGTAGTTG	Reverse primer Venus
F TCTGATAACAGGCTCCACC	Outside 5’ Homology Arm
F TAAACCATGGGATCGGCCATTG	Within Homology construct
R ATTCAGCCTCTAACTCTACTTGC	Outside 3’ Homology Arm
F ACAGGGATCTGCCATGTAGAC	Sequencing
F GAAGCTAGAGGCATCAAATC	Sequencing
R CTTCCATTGCTCAGCGGTG	Sequencing
R CGTGCAATCCATCTTGTTCAATG	Sequencing
R CAGCTTGCCGTAGGTGGCATC	Sequencing
F GACGGCGAGGATCTCGTCGTGAC	Sequencing
F CAATACGCCCGCGTTTCTTCC	Sequencing
F GTCCCTGTCCCTAAAGGATAAG	Sequencing
F ATCCGGACCACCCATGTGCAG	Sequencing
F CCATGTGCAGATTTTACAAGCTC	Sequencing
F GGCATCGACTTCAAGGAGGA	Forward Screen for Neo removal
R CGAAGCGATTTGGGAAGGTG	Reverse Screen for Neo removal

Table 2.2 Primers used for generation of ζ-Venus Mouse

Southern blot probe sequences

5’ probe

GCAAGCGTAGGTGCCATTAGAAGCTAGAGGCATCAAATCCCCTTGCATCTTAGTTACGGACA
GTAGTGCATTCTCTCCAGGGCAGTAAGTGCCATCTCTCCGAATGCTTTTTGTATTATTCAATGT
ACATTTTTTTGCATTTCATTTCCTTAAATACTATGTAAACCACTACTTTGTTTTCTGTCTGGTAG
GTCTTTAAATTTGTATATCCCTTACATTTTGAGCCCAAGGCAAGAGCCAGTGTCCTAGCCCTCT
CCCTAACCAATACTTACTTGTAAAACTAACGCTTTCTCCCGGCCCCTAGGCTCTTCTGCAGCT
ACCCCCAGACGAAGACCTACTTCCCCCACTTCGACCTGCACCATGGGTCTCAGCAGTTGCGG
GCCCACGGCTTCAAGATCATGACCGCCGTAGGGGATGCGGTTAAGAGCATCGACAACCTCTC
TAGTGCTTTGACTAAGCTGAGCGAGCTGCATGCCTACATCCTGCGTGTGGATCCGGTCAACTT
CAAGGTGGGTTTATGAGAAGGGAGGGAGAGGTCAGCAGAGAAAGAGACAAGGAAAAAAAGC
GGGTGGGGTCTCTGCTGCACTGAGCACTGAAATTCCCTTCCCTCCGCAGCTCCTGTCCCACT
GTCTGCTGGTCACAATGGCCGCACGCTTTCCCGCCGACTTCACCCCTGAGGTCCACGAAGCC
TGGGACAAGTTCATGTCTATCCTGTCTTCTATCCTGACTGAGAAGTACCGCTAAGCTCCACCG
CCACGACCCCCATGACCAGCTATGATCCCTCTCCTCCCCTTTATTGAACCCTCCCTCAGGTCT
TTCCCTGAATCCCTAATAAATGGTTGAAGAAGGAAATAACTGGTCCTGTGTATTCGGTCTCAG
AGGGAAACAGGGATGGGGACACTACGGGATGCACTGAGGTTGGCAGGAACACTGCGGTTAT
GCACTGCCTAGAGGACATTCCTAAGGTGGCTAGGGGAGGGTCTGGAGAACCCCAGAGTCAC
AGCTGGCCAGAAGCAGAGCCGCCCACAGTCAGACCTGCCTGCGACTGGAGTTCTCACCTTC
CCAAATCGCTTCGCACAGTGTGGTGGGTGGGCGTCAGGTATTCCTCAAAGTCTGATGTCATA
ACTTCTGCCACCCTGGCAGGTAGGGGGTGTCCCTGTCCCTAAAGGATAAGAAGACTTCCGGT
AACTGTCTTCATAGGAGTCATTACAGGTTCTGCGAGTTCAAAGCGAAGGCTCGCACTCACACG
ACTTAGTCTATGACAGAAGTCCCATCCTTTGGCCTCTCTCTGGGCATGTTTCTGCGGACACGC
GAAGGGCTCTGTGTTTCTGCTGAGTTCAGCGAATCAAGG

3' Probe

CATGTAAGAATCTCAAGGAGGAGAGACTCAAACAGTATCCAGAAAGAGGGGAGCCCTTCCCT
GAAACTCCACCCCCAAGGAGCAGCCCACCACCTCTACCACAGCCAGGGGCAGCCATTCCTTC
TACCTCTAAGTCCCCCAACATCTTATCCTTCTTGTCCTCAGAGCCAGTACGAAGCCTGAGTTG
GGGGTTGCATAGACGTTAGGCACAGACTAGGCGAAATGATGCGGGACAGTTCGGCCAAAGG
CTGAGCTTCTGTGCCCCAGACAAAGGCCCAAGTTTTCTTTCCAGACACCTGTCTCTAGGGGA
GCAGGAGCAGGAAAGGGGACGAACCATCTGAACCTAAGGCTGGGGTCAGTGAACAAGGTGA
TGCCAGCTCACCATGGACACTACCCTGCGGTGTAAGCTGGAAGTAGCCAAGGATGTGGCCAA
GCTGAGAGTTGATGCCCTCACACATGCCCACCCTACCTAGATATAAATTGCCCAGTTCTGTGT
GCATCCAGCAATCTGGTACTGAAGCACACTCTAGTAATGAGCCCACTAGGGTACTTACTAAAG
AAGATAAGTCCTGATCAGGACAAACAGTACCTATGTTAGCACAGAAGCCCTGGAGAGGCAAG
GGAACCTTTCTGCCCTGCCCAGTCTCCAGAGGGCAAAGCTGCCCTTCCCTCATCCTCTGCCC
CTGAACTTTATTCTTTCTGAAGACACTCCACAGTTTTGCCTCCTACAAGACCTACTTCCTTAAC
TAAGACATGAGTCCCTCTCTGCCCTGGTAGCAGCCCAAGGCAAGAGGTGGCATATACATTCA
ACAACTGGGCAGGCCACCTAGGTAACATTTGCACTGTCATGTCAGCCCAGAGTTTCCTCCTGT
CCAGGCGCTACATGAGGACCCCATCAACCCTGAGGTGACTGGTGGCTGGGCAGGGGAGTGA
GCAGGGCTCTGAGGTATTGTTGGTCCTTCACCTAATACGAGAGAACCTCCTGTGGTGTATACA
GCCAC

2.4 Cell biology techniques

2.4.1 Flow cytometry – human cells

All cells were analysed by flow cytometry using the Cyan Flow Cytometer (DakoCytomation) running the software Summit v4.2 or Attune NxT instrument and proprietary software (Thermo Fisher) (the institute changed its main flow cytometry instruments after two years of my DPhil). Data were analysed using flowJo version 10.3.

After washing HUDEP-2 cells twice with FACS staining buffer (PBS, 10% FCS, Life technologies), cells were resuspended in 100 µL of FACS buffer and stained with the following antibodies at 4°C for 30 min in the dark: PE Mouse anti-human CD235a (GA-R2, BD Biosciences 1:100), FITC mouse anti-human Band 3 (BRIC 6, NHSBT 1:50), APC mouse anti-human CD49d (9F10, BD Biosciences 1:40), PE/Cy7 CD34 (BioLegend 1:100), APC/Cy7 CD36 (BioLegend 1:100),

PerCP/Cy5.5 CD71 (BioLegend 1:100). Antibodies were titrated to obtain the optimum signal. Cells were washed twice with FACS buffer, then resuspended in FACS buffer containing 1ng/µl of Hoechst 33342 (Invitrogen). Negative, single colour controls and fluorescent minus one (FMO) samples were prepared for calibration.

Intracellular staining of HUDEP-2 cells was performed using the Intracellular Fixation & Permeabilization Buffer Set (eBioscience) in accordance with the manufacturer's instructions. Prior to fixation, cells were stained with the LIVE/DEAD® Fixable Aqua Dead Cell Stain Kit (Life Technologies) in accordance with the manufacturer's instructions. Hybridoma supernatant was used as the primary antibody (15 µl in 100 µl) (Luo et al., 1988), PE goat anti-mouse IgG (SouthernBiotech) was used as the secondary antibody. A mouse IgG1 (MOPC-21 BD Biosciences, 1:200) was used as an isotype control. HbF staining was carried out with FITC mouse anti-human HbF (2D12, BD Biosciences) with a FITC mouse IgG1 Isotype control (MOPC-21, BD Biosciences).

To determine whether the ζ-globin antibody was specific to ζ-globin as opposed to other globins using flow cytometry, two experiments were performed.

Firstly, blood from a Bart's Hydrops terminated fetus at 18 weeks (obtained from Ali Ahmed, Toronto Hospital for Sick Children with appropriate local consent) and adult blood from a healthy donor was subject to flow cytometry using the antibody. Children with BHFS still express ζ-globin in significant amounts at this point. 20 µL of whole blood was fixed with 1mL of Glutaraldehyde 0.05% (Sigma-Aldrich), then

permeabilised with 0.5% Triton X-100 (Sigma-Aldrich) 0.1% BSA (Sigma-Aldrich) in PBS. Staining was carried out with the primary ζ-globin antibody for 30 min, then with the secondary PE goat anti-mouse IgG antibody. Following washing, cells were analysed on the flow cytometer (Figure 2.5) and the above isotype control used. A strong signal was seen in the terminated fetus as opposed to adult blood. This excluded cross reactivity with α-globin, β-globin and ε-globin (which has been silenced by 18 weeks).

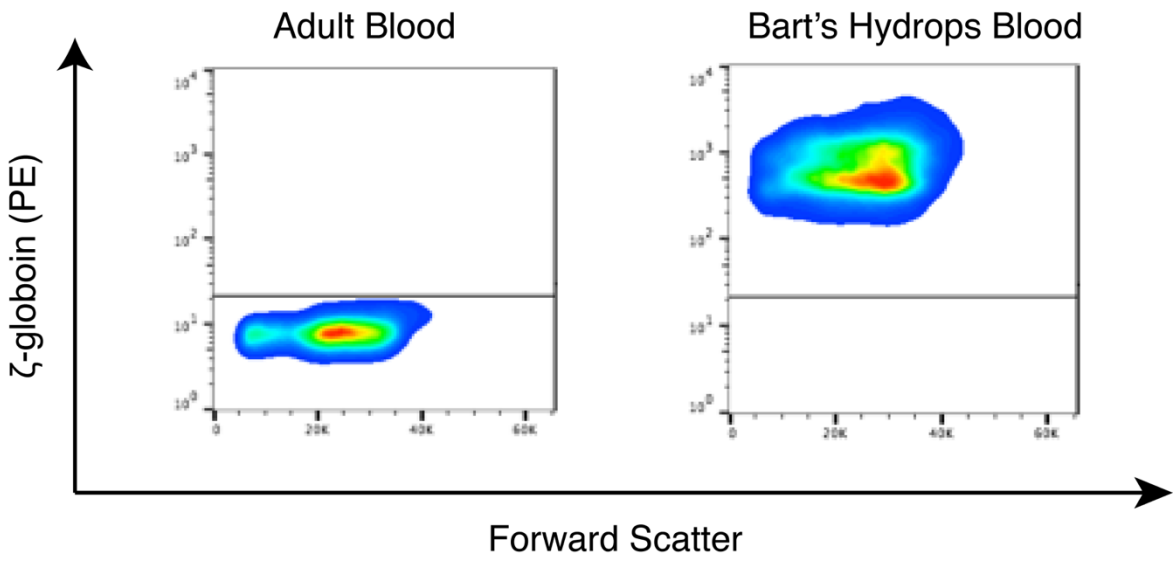


Figure 2.5 Flow cytometry on adult blood and blood from a terminated fetus with Hb Bart's hydrops fetalis.

To exclude cross-reactivity with γ-globin, WT HUDEP-2 cells, which have small numbers of ζ-globin positive and HbF positive cells were FACS sorted after staining on a BD FACSAria Fusion machine with a 100 μM nozzle then RNA extracted and measured by RT-qPCR (Figure 2.6). This showed negligible amounts of ζ-globin RNA in the HbF positive fraction, and negligible amounts of HBG in the ζ-globin positive fraction. This confirmed that the ζ-globin antibody is specific.

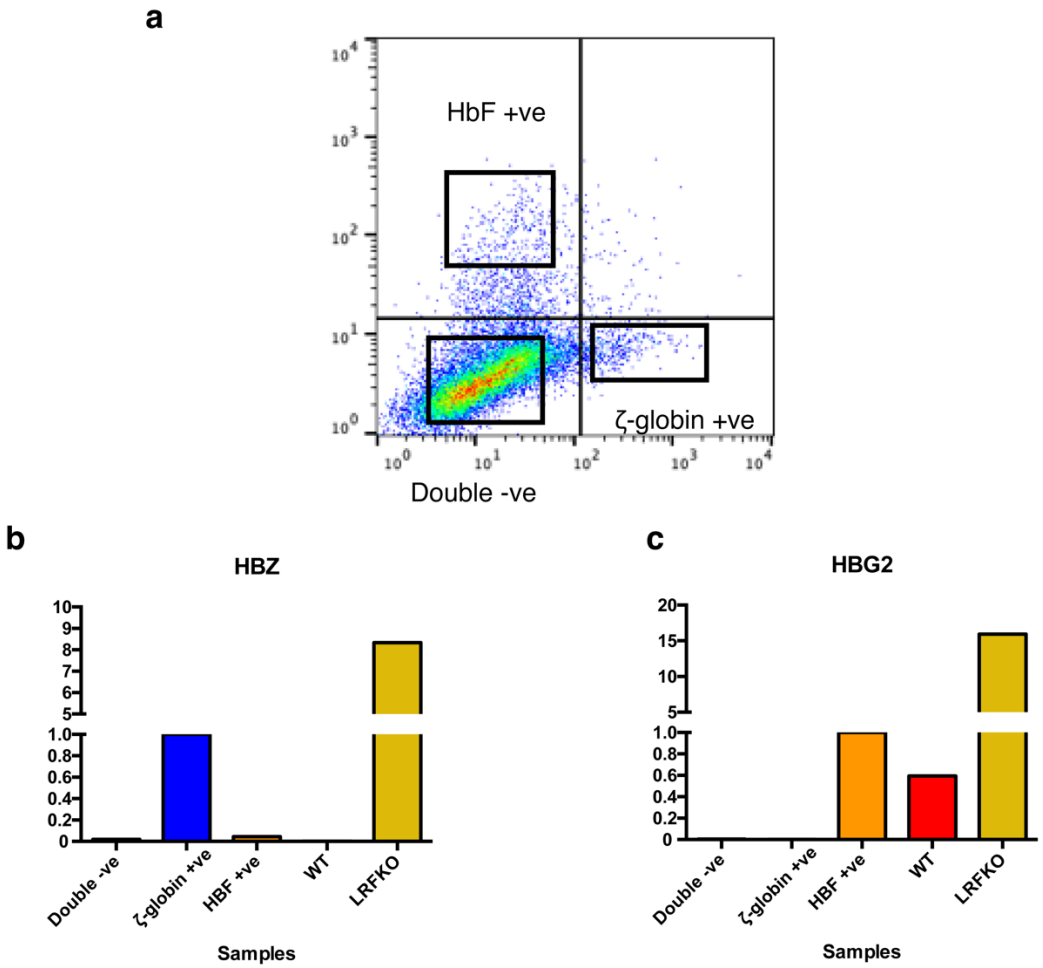


Figure 2.6 RNA expression from different sorted fractions of HUDEP-2 cells after antibody staining with ζ-globin and γ-globin. RNA was normalised to RPL13A, then expressed as a ratio against the ζ-globin +ve and HbF +ve fractions for HBZ and HBG respectively (n=1). WT HUDEP-2 and LRF$^{-/-}$ cells were also included as controls.

2.4.2 Flow cytometry – murine cells

To assess differentiation from the fetal liver culture, cells were suspended in FACS buffer (PBS/10% FCS) containing the following antibodies at 4°C for 30 min which had been titrated: PE anti mouse Ter119 (Ly-76) (1:200, BD Biosciences) and FITC

anti-mouse CD71 (C2, BD Biosciences, 1:100). After washing twice in FACS buffer, cells were resuspended in FACS buffer containing 1ng/μL of Hoechst 33342 (Invitrogen). Single colour stains and FMOs were prepared for calibration.

Flow cytometry on primary primitive erythroid cells from the ζ-Venus mouse was performed at E10.5. Staining was performed with 1ng/μL Hoechst and YFP (in the FITC gate). WT littermate controls were used to define negative populations.

2.4.3 FACS sorting – primary fetal livers

To isolate cells of different stages of differentiation directly from primary fetal livers at E12.5-E13.5, 5-15 fetal livers were pooled together, mechanically dissociated in staining buffer (PBS, 0.2% BSA, 5mM Glucose) and strained through a 30-μM filter. Cells were stained at 4°C with rabbit IgG (200μg/mL, Jackson Laboratories) to block Fc receptors. Cells were then incubated with PE-Cy7 conjugated anti-CD71 (RI7217 BioLegend, 1:2000), APC conjugated anti-Ter119 (RUO, BD Biosciences; 1:200), BB700 conjugated CD117 (104D2, BD Biosciences, 1:100), PE conjugated Annexin V (BioLegend, 1:100), a panel of 5 FITC-conjugated lineage antibodies (anti-CD41, anti-CD45R, anti-CD3e, anti-CD11b and anti-Ly-6G/6C, all at 1μg/mL, all BD Biosciences), Cells were then resuspended in FACS running buffer (staining buffer with the addition of 2mM EDTA). 0.66 μg/mL Hoechst (Invitrogen) was added prior to sorting, which was performed by Kevin Clarke or Sally-Ann Clark on a BD FACSAria Fusion machine with a 100 μm nozzle. The sorting protocol is shown in Figure 2.7. Annexin V was only added for the experiments on the $LRF^{-/-}$ mice. Single colour controls and FMOs were used for calibration.

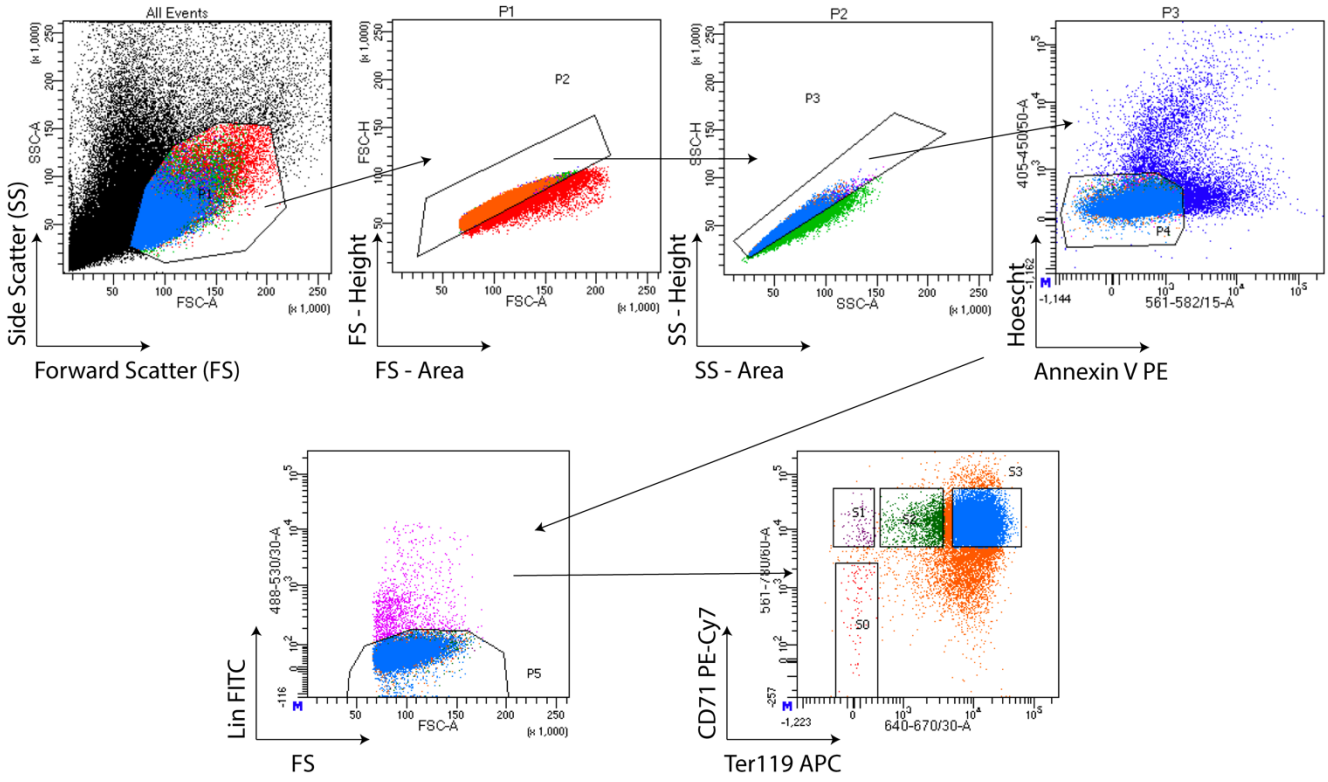


Figure 2.7 Sorting Protocol for primary fetal livers. After gating on FS/SC, two gates were applied to isolate single cells only. Live cells and lineage gated cells were then gated out. For ATAC-seq, S3 populations were used (shown in the furthermost plot).

2.4.4 FACS sorting – primitive erythroid cells

To isolate pure populations of primitive erythroid cells from E9.5 to E11.5 blood was isolated directly from yolk sacs from timed matings. The same protocol was used as above for sorting. On FS vs SS, only the largest cells were taken, and again, S3 cells were used for ATAC-seq.

2.4.5 Wright stain of cytospins

Cytospins were prepared by centrifugation of $1x10^5$ cells at 400 rpm onto a glass slide. The slides were air-dried, then stained with a modified Wright stain using an automated Hemateck 03564647 stainer. Slides were then imaged on a Nikon inverted microscope.

2.5 Routine molecular biology/protein techniques

2.5.1 DNA extraction

Genomic DNA for analysis was extracted using the PureLink™ Genomic DNA Mini Kit (Thermo Scientific) according to the manufacturer's instructions and quantified using the NanoDrop Spectrophotometer.

Plasmid DNA was extracted using the MiniPrep or MaxiPrep kits (depending upon volume) according to manufacturer's instructions.

To isolate genomic DNA from 96 well plates, 50 µL lysis buffer (100 mM Tris Base, 5 mM EDTA, 200 mM NaCl, 0.2% SDS, pH adjusted to 8.0 with HCl and freshly added 100 µg/ml Proteinase K) was added to wells (after spinning and removing supernatant), then incubated at 37°C for at least 48 hours. The contents were then transferred to a V bottom 96 well plate and precipitated with 50 µL ice cold propranolol. Plates were then shaken for 20 min on a plate-shaker (2500rpm), washed twice with 70% ethanol, then resuspended in TE buffer. An adhesive film

was placed over the plate to avoid evaporation, then the plate was placed in a iEMS shaker/heater (Thermo Scientific) set at 60°C to dissolve the DNA.

DNA was only used for downstream analysis with a 260/280 and 260/230 ratio >1.8. DNA was stored at -20°C for long-term storage.

2.5.2 Sanger Sequencing

Sequencing reactions contained 20-450ng of DNA depending upon the template and was carried out by the WIMM sequencing facility. Sequence data was visualised using Snapgene version 4.2.6.

2.5.3 RNA extraction and cDNA synthesis

Isolation of RNA was performed by lysing cells in TRI reagent® (Sigma-Aldrich) according to the manufacturer's instructions. For long-term storage, cells were lysed in TRI reagent®, then stored at -80°C. To isolate RNA from lysed cells, 200 µL chloroform was added and incubated at room temperature after vortexing. Samples were then spun for 15 min (12,000 rpm, 4°C). The aqueous phase was then added to 500 µL of isopranolol and incubated for 5 min at room temperature. Samples were then spun for 10 min (12,000 rpm, 4°C) and the pellet containing RNA washed once with 70% ethanol, then air dried. RNA was suspended in 44 µL DPEC treated H_20 and incubated at 55°C for 10 min. genomic DNA was removed with the DNA-free DNA removal kit (Ambion). RNA samples were stored at -80°C. RNA quality was confirmed using the Agilent Tapestation. 1 µg of RNA was used

for cDNA synthesis using the Superscript III first-strand synthesis SuperMix (Invitrogen) kit according to the manufacturer's instructions.

2.5.4 RT-qPCR analysis

Commercially validated TaqMan assay (Applied Biosystems) were used, detailed in Table 2.3. *RPL13A* was used as the reference gene as previously published(Merryweather-Clarke et al., 2011) in humans, and *Rps18* in mice. RT-qPCR was carried out in 10 µL reaction volumes using the TaqMan Universal PCR Master Mix (Thermo Scientific), following the manufacturer's protocol. Reactions were carried out in triplicate in 96-well plates with an ABI Prism 7500 sequence detection thermocycler (Applied Biosystems). Samples were diluted 1/10 prior to analysis with DNAse free water. Both no template and no RT (non-reversed transcribed RNA) were included in all qPCR reactions amplifying cDNA. The ΔΔCt method was used for relative quantitation.

For experiments on embryoid bodies where Venus expression was quantified, custom primers were designed and used with Fast SyBR green Mastermix (Thermo Scientific) according to the manufacturer's instructions. Custom primers are shown below (Table 2.4).

Target	Catalogue Number
HUMAN	
HBA (HBA1/2)	Hs00361191_g1
HBB	Hs00747223_g1
HBZ	Hs00923579_m1
HBE	Hs00362216_m1
HBG (HBG1/2)	Hs00361131_g1
RPL13A	Hs04194366_g1
MOUSE	
Hba-a (Hba-1/2)	Mm02580841_g1
Hbb-b	Mm01611268_g1
Hba-x	Mm00439255_m1
Hbb-y	Mm00433936_g1
Hbb-bh1	Mm00433932_g1
Bcl11a	Mm00479358_m1
Rps18	Mm02601777_g1

Table 2.3 TaqMan probes used for RT-qPCR analysis. Note that the HBA and Hba-a probe targets both α-globin genes; similarly, the HBG probe targets both γ-globin genes.

Target	Sequence
Hba-x	F ATGCGGTTAAGAGCATCGAC
	R GGGACAGGAGCTTGAAGTTG
Venus	F CAAAGACCCCAACGAGAAGC
	R CTTGTACAGCTCGTCCATGC

Table 2.4 Custom primers (Sigma-Aldrich) used for RT-qPCR analysis.

2.5.5 Multiplex single cell RT-qPCR

Single erythroid cells were isolated by index-FACS from each E12.5-E13.5 embryo (12 S1 cells, 12 S2 cells, 24 S3 cells). Cells were sorted into wells of a 96-well plate containing 5.1 µL lysis buffer (5 µL 2x reaction buffer from SuperScript™ III One-

Step RT-PCR with Platinum™ Taq kit (Thermo Scientific) 0.1 µL SUPERase In™ RNase Inhibitor (Thermo Scientific).

Plates were sealed, briefly centrifuged and frozen at -20ºC. TaqMan assay mastermix was made by pooling 6.6 µL of 43 TaqMan assays and adding 376 µL TE buffer. Plates were defrosted and 4.9 µL amplification mix was added to each well (1.2 µL TE buffer [Invitrogen], 2.5 µL TaqMan assay Mastermix, 1.2 µL RT/Taq enzyme mix from SuperScript™ III One-Step RT-PCR with Platinum™ Taq kit) except for one well containing S3, in which Platinum Taq was substituted for RT/Taq mix to serve as a no-RTase control. cDNA synthesis and sequence-specific preamplification was then carried out (reverse transcription at 50 °C for 15 min; RTase inactivation at 95 °C for 2 min; specific target amplification by 22 cycles of 95 °C for 15 s then 60 °C for 4 min).

An RNA standard curve was prepared using the same Mastermix and amplification program as below. Plates were stored at -20ºC until analysis. Pre-amplified cDNA was diluted 5x with TE buffer and analysed using Universal PCR Master Mix (Applied Biosystems) and individual TaqMan gene expression assays (Life Technologies, Table 2.5), on the Biomark System (Fluidigm) using two 192.24 Dynamic Arrays as per the manufacturer's protocol, except that elongation time was increased from 60s to 75s to accommodate longer custom gene-expression assays for Hba-a1/2 and Hbb-bt/s/1/2 (based upon previous work by Rob Beagrie).

At the same time 5000-10000 cells were taken from the S3 population for ATAC-seq.

Target	Catalogue Number	Target	Catalogue Number
Abcg4	Mm00507247_m1	*Pbx1*	Mm04207617_m1
Aqp1	Mm01326466_m1	*Podxl*	Mm00449829_m1
Aqp8	Mm00431846_m1	*Pogz*	Mm00554475_m1
Arid3a	Mm00492248_m1	*Prtn3*	Mm00478323_m1
Ccr2	Mm00438270_m1	*Rhbdf1*	Mm00711711_m1
Cited2	Mm00516121_m1	*Rhd*	Mm00456910_m1
Cox6c	Mm00835813_g1	*Sec61g*	Mm01613099_g1
Cpeb4	Mm01193440_m1	*Selk*	Mm00785961_s1
Ddx21	Mm00497941_m1	*Smarca4*	Mm01151944_m1
Foxh1	Mm00514851_m1	*Snrnp25*	Mm00547218_m1
H2afy	Mm01337623_gH	*Sox4*	Mm00486320_s1
Hba-a1/2	Custom (IDT)	*Sox6*	Mm01274768_m1
Hba-x	Mm00439255_m1	*Tmem14c*	Mm00481276_m1
Hbb-bh1	Mm00433932_g1	*Vamp5*	Mm00444144_m1
Hbb-bh2	Mm01273444_g1	*Zbtb7a*	Mm00657132_m1
Hbb-bt/bs/b1/b2	Custom (IDT)	*Hbb-b2*	Mm00731743_mH
Hbb-y	Mm00433936_g1	*Ccr9*	Mm02528165_s1
Hbq1a	Mm00731011_s1	*Flt3*	Mm00439016_m1
Hbq1b	Mm02747875_s1	*Ms4a6c*	Mm00459296_m1
Hspe1	Mm00434083_m1	*Pf4*	Mm00451315_g1
Il9r	Mm00434313_m1	*Plek*	Mm00450970_m1
Klf3	Mm00492956_m1	*Ifitm1*	Mm00850040_g1
Lmo4	Mm00495373_m1	*Nrep*	Mm00474047_m1
Mbd2	Mm00521967_m1	*Vwf*	Mm00550376_m1
Mpg	Mm00447872_m1	*Alas2*	Mm00802083_m1
Mpo	Mm01298424_m1	*ApoE*	Mm01307193_g1
Myc	Mm00487804_m1	*Arpc2*	Mm01254383_m1
Myl6	Mm02342525_g1	*Arpc2*	Mm01254383_m1
Nprl3	Mm01193449_m1	*Atp1b2*	Mm00442612_m1
Oaz1	Mm01611061_g1	*CD59b*	Mm02525679_s1
Ddx21	Mm00497941_m1	*Dek*	Mm01351566_m1
Fth1	Mm00850707_g1	*Hnrnpa0*	Mm03809085_s1
Gata2	Mm00492301_m1	*Klf1*	Mm00516096_m1
Gpx1	Mm00656767_g1	*Mpl*	Mm00440310_m1
Ncl	Mm01290600_g1	*Nono*	Mm00834875_g1
Tagln2	Mm00724260_g1	*Slc43a3*	Mm00469627_m1
Trib2	Mm00454876_m1	*Pebp1*	Mm02601848_g1
Sh3pxd2b	Mm00616672_m1	*Hbb-b1*	Mm03646870_gH

Table 2.5 Taqman assays used in single cell experiment

Assay	Left Primer	Right Primer	Internal Oligo
Hba-a1/a2	TGAAGCCCTGGAAAGGATGT	TGAAATCGGCAGGGTGGT	TGGATCCCGTCAACTTCAAGCTCCT
Hbb-bt/bs/b1/b2	CCGATGAAGTTGGTGGTGAG	ACATGCAGCTTGTCACAGTG	CCTGGGCAGGCTGCTGGTTG

Table 2.6 Custom Taqman probes used in single cell experiments

2.5.6 RNA standard curve preparation (performed by Rob Beagrie)

T7 polymerase was used to in vitro transcribe RNA from cDNA clones of Cox6c, Sec61g, Oaz1, Hba-a1, Hba-x, Hbb-bt, Hbb-y and Hbb-bh1. RNA was run on a 6% acrylamide TBE-urea gel (Life Technologies) and bands corresponding to full-length transcript were isolated from the gel. RNA was extracted by incubating crushed gel pieces with elution buffer (500 mM Ammonium acetate, 1 mM EDTA-KOH pH 8.0) for 3 hours at room temperature before ethanol precipitation, resuspension in TE buffer and quantification using the RNA Quantifluor assay (Promega) and RNA Tapestation (Agilent). Each RNA was diluted to 16.8 pM concentration and the 8 transcripts were pooled, aliquoted and stored at -80 °C.

When performing reverse transcription and pre-amplification on sorted single cells, one aliquot of RNA standard was also defrosted, and diluted 10x. Five more serial dilutions were created by diluting 6x in TE buffer, then 2 μl was taken from each of the six standard curve dilutions and pre-amplified using the same mastermix as used for single cells (see above).

2.5.7 Southern blotting

For each restriction enzyme, genomic DNA was prepared from targeted ES cells (10µg) and digested. KpnI-HF (20,000 units/mL NEB) and EcoRV-HF (20,000units/mL NEB) were used. For each reaction, 5 µL of enzyme was added to 5 µL of CutSmart Buffer (NEB), DNA and H_2O to make the reaction up to 50 µL, then incubated overnight at 37°C followed by heat inactivation. Digested DNA was run on a 0.8% Agarose/TAE gel at 50V overnight, then denatured with 1M NaOH for one hour. The gel was then neutralised with neutralisation buffer for 2 hours (1M Tris, 3M NaCl, pH 7.4). DNA was then transferred onto a Zeta-probe membrane (BioRad) by upward capillary transfer in 6x saline-sodium-citrate (SSC). Membranes were UV crosslinked (Stratalinker) and then baked for two hours at 80°C. Prepared probes (50ng) were labeled with the random primers DNA labeling system (Thermo Scientific) according to manufacturer's instructions and cleaned up with Microspin G50 Columns (GE Healthcare), then denatured at 95°C for 5 min. Hybridisation was performed with the PerfectHyb Plus Hybridization Buffer (Sigma-Aldrich) overnight at 68°C. The next day, washes were performed at 42°C in progressively lower salt solutions with sodium dodecyl sulfate (SDS), (2x SSC, 0.1% SDS; 1xSSC 0.1% SDS, 0.5xSSC; 0.1% SDS) for 20 min each time until radioactivity was judged to have fallen to an appropriate level using a Geiger counter. Blots were exposed to films with an intensity screen at -70°C for 72 hours, then developed.

2.5.8 Analysis of haemoglobin

Weak cation exchange high performance liquid chromatography (CE-HPLC) was done using the Bio-Rad VARIANTnbs system (Bio-Rad Laboratories) according to manufacturer's instructions, with Jennifer Eglington of the Haematology Department, Oxford University Hospitals NHS Foundation Trust.

2.6 Validation of the ζ-Venus model

2.6.1 Correlation between *Hba-x* and *Venus* expression and fluorescent signal in embryoid bodies (EBs)

Fluorescence microscopy images of day 7 EBs were captured with an IX51 instrument (OLYMPUS) using a filter for the YFP fluor and overlaid with brightfield images, obtained in parallel, using the Fiji software package(Schindelin et al., 2012).

The mean fluorescence level of this tag for EBs differentiated for 4-7 days was calculated using flow cytometry with an Attune NxT instrument and proprietary software (Thermo Fisher). The mean was chosen as an appropriate metric for comparison to RNA expression levels to account for the fact that only ~40% of EB-derived cells are erythroid in nature.

For expression analyses, $2x10^6$ cells from whole EBs were lysed in TRI Reagent (Sigma-Aldrich) and immediately frozen at -80°C. RNA was then extracted as detailed above. RT-qPCR was performed with Fast SYBR Green Master Mix (Thermo Fisher) and data normalised to the 18S ribosomal gene using the $2^{-\Delta\Delta Ct}$

method. RNA quality was routinely checked by Tapestation analysis (Agilent) and reverse transcriptase enzyme negative controls were included to confirm that DNase treatment was adequate.

2.6.2 Fixing and dissection of embryos

Embryos were fixed in 4% formaldehyde at 4°C (Sigma-Aldrich) for one hour, then embedded in OCT compound (Sakura Finetek) and stored at -80°C until further use. Specimens were sectioned on a cryostat (Leica) at 10 micron thickness onto glass slides. To boost the Venus signal, slides were washed with PBS, then permeabilised with PBS,0.4% Triton-X100 and 2% FCS. Slides were then stained with a primary anti-GFP chick antibody (Abcam ab13970) 1:200, then with an Alex Fluor® 488 secondary antibody (Jackson ImmunoResearch; 703-545-155) 1:200. The anti-GFP antibody also targets YFP. Slides were imaged with a Nikon inverted fluorescence microscope. Pictures of whole embryos from a culled mother were also immediately taken without any further processing on a Nikon Eclipse E600 microscope and DXM1200C camera.

2.6.3 Drug screening optimisation

Fetal livers were dissected from E12.5 embryos and cultured in expansion media as described in section 2.1.4. After 5 days, cells precursors were selected using anti-CD117 beads (Miltenyi Biotec) and an LS column according to the manufacturer's instructions. Adequate purity was confirmed with flow cytometry. 50,000 cells were plated out into 96 well round bottom plates containing 200µL of

differentiation media (detailed in 2.1.4) and cultured for 48 hours in the presence of different concentrations of HC toxin (Sigma-Aldrich-Aldrich). Plates were then analysed on the Attune NxT instrument using an automated plate reader.

2.7 Next generation sequencing techniques

2.7.1 The Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

ATAC-seq was employed to determine regions of open chromatin. This assay relies on a mutated hyperactive transposase (Tn5) which cuts exposed DNA found within open chromatin regions. These include enhancers, promoters, and boundary elements (CTCF sites). After cutting, the transposase ligates adapter sequences into the cute site. This then permits next generation sequencing, allowing genome wide profiles to be generated.

ATAC-seq was performed as published (Buenrostro et al., 2013). 10,000-50,000 cells were used per biological replicate where possible; if cells numbered less than 10,000 an optimised low cell number protocol was used. Cells were spun at 500g for 10 min, then lysed using cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630). Transposition was then performed with Tn5 transposase (Nextera, Illumina) for 30 min at 37°C following the reaction volumes as published. DNA was purified using the Qiagen MinElute kit following the manufacturer's instructions, then eluted in 20µl of elution buffer; samples could be stored at this point at -20°C. Libraries were amplified and barcoded using the

NEBNext 2xMastermix kit with custom primers as published (Buenrostro et al., 2013). ATAC-seq library profiles were visualized using the D1000 tape on the Agilent Tapestation to ensure correct tagmentation. Libraries were quantified using the universal library quantification kit (KAPA Biosystems) prior to massively parallel sequencing using a standard Illumina protocol (NextSeq High Output v2 75 cycle sequencing kit).

The low cell number protocol uses proportionally less lysis buffer and Tn5 transposase (Corces et al., 2017). After spinning at 500g for 10 min and removing the supernatant, cells were permeabilised and tagmented with Dig-transposition buffer (25 µL 2x TD buffer [Illumina], 2.5 µL Tn5 Transposase, 0.5 µL 1% digitonin and 22 µL H_2O) for 30 min. After the transposition step, samples were processed as per larger cell numbers.

2.7.2 Chromatin immunoprecipitation analysis followed by high throughput sequencing (ChIP-seq)

Chromatin immunoprecipitation (ChIP) was performed on 3-5x10^6 cells per replicate using the ChIP Assay Kit (Cat. No. 17-295, Millipore). All stages beyond the fixation stage used kit solutions as per the manufacturer's instructions. A single 10 minute 1% formaldehyde fixation was used for all antibodies. The cross-linking reaction was quenched by addition of glycine (Sigma-Aldrich) to a final concentration of 130mM), then washed once in PBS and snap frozen. Samples were stored at -80°C prior to further processing.

Cells were lysed in SDS lysis buffer with Complete Proteinase Inhibitor Cocktail (PIC, Roche). Samples were then fragmented using the Covaris S220 Focused ultra-sonicator with the following settings (Duty Cycle: 2%, Intensity: 3.0, Cycles/burst: 200, Power mode: Freq. Sweeping, Duration: 120s, Temp: 6°C, Batches: 4). Sonicated chromatin samples were diluted 1:10 in ChIP dilution buffer containing PIC to 2mL. Before addition of the antibody, 100 µL was taken from each 2mL sample to be used as a 5% input. The chromatin was pre-cleared with Protein A agarose for 30 min before immunoprecipitation overnight with the antibodies shown in Table 2.7 at 4°C rotating (the volumes added are also shown). Protein A agarose beads were added the next morning for 60 min at 4°C, after which the beads were sequentially washed in Low Salt Buffer (1x), High Salt Buffer (1x), LiCl Buffer (1x) and TE Buffer (2x) with the final TE buffer wash at room temperature. Bound chromatin fragments were eluted from beads in elution buffer (1% SDS, 0.1M $NaHCO_3$), de-crosslinked by incubating samples at 65°C for 5 hours, then proteinase K treated for one hour at 55°C. DNA was precipitated using phenol/chloroform extraction and ethanol precipitation with GlycoBlue (15mg/mL) (Thermo Scientific) to visualise the pellet prior to dissolving in 55 µL of H_2O.

Target	Manufacturer
CTCF	Millipore 07-729 10 µL
H3K27ac	Abcam ab4729 1µL
H3K4me1	Merck 07-436 0.2µL equivalent
H3K4me3	Abcam ab8580 5µL

Table 2.7 Antibodies used in ChIP experiments

To confirm appropriate enrichment prior to library preparation and next generation sequencing ChIP-qPCR was used with DNA diluted 1/10 and Fast SyBR green

mastermix (Thermo Scientific) according to the manufacturer's instructions. Primers used are shown in Table 2.8.

Target	Sequence	Mark
Intergenic (mm9)	F GTAAGATGCCATTGTCCTGACC R TGTGTTCTCCTCCCTTCCTT	Negative
HS-38 (CTCF mm9)	F TGAGAAGGCTGGCCTTTGAG R CCCAGGGAATGAATGCCAGT	CTCF mouse
Hba-a1 (mm9)	F CTAGCTTCCCCACCACCAAG R CTTCTTGCCGTGACCCTTGA	H3K4me3 H3K27ac
HS-26 (mm9)	F GTACAGAGTCAGAAAGGAAAGGACAA R GGCAAATAAACAGGATGGGAAA	H3K4me1 H3K27ac
Intergenic (hg19)	F GGGTTTCCCTGTTACAGAAATG R CTCTGGTCATCAGGTCAGTATG	Negative
HS-48 (hg19)	F GCCATGTTGGTTCCTCTAACT R AGAGTGCTCTGAGGGAGAAA	H3K27ac
CTCF (hg19) HS-55	F CCACCCTCAGGCAGTATTTCAA R CCAACAGGCCATGGGATCAG	CTCF human
NPRL3 (hg19)	F GTTGCTTTGGTGTGGTGAAC R CATTCGAGATCTGCTCGTTGTA	H3K4me3

Table 2.8 Primers used to check enrichment following ChIP

The concentration of precipitated DNA was determined by the Qubit dsDNA HS assay (Invitrogen), then indexed using the NEBNext Ultra II DNA library prep kit (NEB) following the manufacturer's instructions. Adapter stocks and primers stocks were diluted 1:10. Size selection was not performed. PCR amplifications was typically performed for 11 cycles. Libraries were quantified using the KAPA library quantification kit (KAPA Biosystems) and then sequenced on the NextSeq platform using High-output 75 cycle kits (Illumina), paired end sequencing. Input samples were prepared in the same fashion.

2.7.3 Next Generation (NG) Capture C

Next generation Capture-C was performed as published (Davies et al., 2016). $5x10^6$ cells were fixed in 2% formaldehyde in IMDM at room temperature. The cross-linking reaction was quenched after a 10 minute incubation through addition of ice-cold glycine (final concentration 0.125M). Fixed cells were washed twice in ice cold PBS, the resuspended in 5mL cold lysis buffer (10 mM Tris-HCl pH 8, 10mM NaCl, 0.2% IGEPAL CA630 [Sigma-Aldrich], 1x PIC); incubated for 20 min on ice, then centrifuged (5 min, 500g, 4°C). The pellet containing cell nuclei was snap frozen and stored at -80°C in PBS. After thawing, cells were resuspended in 1mL DpnII buffer and homogenised with a 5mL Dounce homogeniser (55 strokes in total), spun (14,000 rpm, 5 mins, RT) and resuspended in 650 µL od DpnII buffer. The chromatin was then split across three tubes equally (200 µL) and digested by addition of 80 µL of DPNII 10X restriction buffer (NEB), 10 µL of 20% SDS (v/v) and 434 µL of H_2O to which 66 µL of Triton X-100 was added after a 1 hr incubation at 37°C, shaking at 1400 rpm horizontally. After a further 1 hr incubation, three aliquots of 500 U (10 µL) DpnII (NEB) were added over several hours. Incubation was then continued overnight (16-24 hours in total). To control for non-specific degradation, the remaining 50 µL of cross-linked nuclei were treated in an identical fashion, apart from addition of the DpnII enzyme. To assess digestion efficiency, DNA was extraction from the undigested control (in total 200 µL) and a pooled pooled sample was made containing 100 µL from each digestion reaction. Controls were treated with proteinase K and RNAse A, phenol/chloroform extracted, and ethanol precipitated.

The remaining digested material heat inactivated by incubation to 65°C for 20 min, then ligated. To ligate, each (of the three digested) samples were diluted with 500 µL H_2O, after which 133 µL of 10x ligation buffer and 8 µL high concentration T4 DNA ligase (Thermo Scientific; 30U/µL) was added. Samples were then incubated overnight (1400 rpm; horizontal shaking, 16°C). Ligated DNA was reverse cross-linked and proteinase K treated (Thermo Scientific, >600 U/mL) overnight at 65°C. Ligation reactions were then pooled, treated with RNAse (30 µL, Roche) at 37°C for 30 min and then phenol/chloroform extracted and ethanol precipitated. For ethanol precipitation, the 4 mL sample was diluted with 7mL water and precipitated with 1.5 mL of 2M Sodium acetate and 35 mL100% ethanol to improve DTT removal. Ligated 3C libraries and digestion controls were routinely run on a 1% agarose gel to determine digestion and ligation efficiency.

Two 5 µg aliquots of a 3C library were sonicated using a Covaris S220 Focussed ultra-sonicator (Duty Cycle: 10%, Intensity: 5.0, Cycles/burst: 200, Power mode: Freq. Sweeping, Duration: 60s, Temp: 6°C, Batches: 6). Illumina TruSeq adaptors and indices were added using NEBNext reagents (NEB). DNA clean-up steps were performed with Ampure XP beads (Beckman Coulter Genomics) at a 1.8:1 ratio. The Herculase II PCR kit (Agilent) was used to add TruSeq indices in a 7 cycle PCR reaction.

Capture of fragments were performed with Nimblegen SeqCap AX reagents (Roche) using custom biotinylated oliognucleotides targeting genes and regions of interest (Table 2.9) following the manufacturer's protocol. Library fragments hybridized to biotinylated oligos were pulled down using M-270 streptavidin beads

(Life Technologies). DNA fragments were amplified using the HAPA HiFi HotStart ReadyMix and Post-LM_PCR Oligos 1 and 2 following the NimbleGen SeqCap Library SR protocol. The library DNA was recovered with an Ampure XP bead clean-up (1.8x) and quantified using the D1000 Tapestation (Agilent) and the Qubit dsDNA BR assay kit (Invitrogen). A minimum of 500ng of material from the first capture was used as input for the second with a decrease in hybridisation time to 24 hr (from 64-72 hr). Libraries were then quantified with the D1000 Tapestation and the Qubit dsDNA BR assay kit and sequenced on the Illumina MiSeq platform following the manufacturer's instructions (300 bp paired end, Mid-output).

Table 2.9 (next page) Oligonucleotide probes used for Capture-C experiments. All had a 5' biotinylated modification.

Target	Sequence 5' to 3'
Hba-x	**FORWARD**_GATCTTCCCTCAGTGCTAAGTGAGAGG AAT TAC TGC TTCCTTGGAGGGAACCCAAGAGTCACTGACATTAGGTGACAGGGATGGCTCAGACAATTT GTGATTCTGAAATGAGAAAGGACA-3′ **REVERSE**_GACAGAAGTGACAAAGAGCAGTTAGTGAGGAGAAGCCACTGAGAAGGGGTC ATGAGGACAGTATGGTCAGGACAGTAAGGACAGAGGCAGAGAGGACAACGTGGAGAGGA CTGAGAGATC-3′
Hba-a1	**FORWARD**_GATCCATGGTAGCACAGGGCAGCTAAGATGCAAGTCTGAAGGAGGAGTCTG GCGAGCTGCTCCTGCAGTTCCCTGGACCCAGAAGGATGAGCTAGCAGATTCACTTGA GCCAAAGGATTC **REVERSE**_AAACCATGGTGCTCTCTGGGGAAGACAAAAGCAACATCAAGGCTGCCTGGGG GAAGATTGGTGGCCATGGTGCTGAATATGGAGCTGAAGCCCTGGAAAGGTGAGAACAGG ACCTTGATC
Hba-a2	**FORWARD**_GATCCATGGTAGCACAGGGCAGCTAAGATGCAAGTCTGAAGGAGGAGTCTG GCGAGCTGCTCCTGCAGTTCCCTGGACCCAGAAAGATGAGCTAGCAGATTCACTTGAGC CAAAGGATTC **REVERSE**_AAACCATGGTGCTCTCTGGGGAAGACAAAAGCAACATCAAGGCTGCCTGGGG GAAGATTGGTGGCCATGGTGCTGAATATGGAGCTGAAGCCCTGGAAAGGTGAGAACAGG ACCTTGATC
Rhbdf1	**FORWARD**_GATCGGCGGACACGCGCCAGAACGCAAGCTCCGCACTCGCCCGCCCGGAA CCCCGCGCAAAGACCCCGGCCGCCACTCACCGCCGTAGGGTCCTCAGAAGGGTCTGCA CCGGCTGCTGCC **REVERSE**_CAAGAGGGTGGCGTGGGAAGGGACCTAAAAGGCAGATTGAACGCCCCTTCC GAAGGCACGCACTGGCACATCACTGAGGGTCCCGAAGCTCTCACTTGTCCTGTGCACAC GCGCCTGATC
HS-38	**FORWARD**_GATCCCAGTCAAGACCTAGACAAGGAAATGGAAGAGACATCAAAGCTTTATC AAAGGACTGTGACAGTGAAGTCAGGCACAGTCCTAGCACAGATTCCTTCAAACAGCCTCT TAGTAAAG **REVERSE**_CTGCCACATGTCTTCCGCCACAAGAATGAGGGTATCTTGATGACTCCAGGGC TGGGTTGGATAAAAGGGGAGAAGGAGCAGAACAAAAGGGACCCATGTGGAGCGTACCTC ACTTGGATC
R1	**FORWARD**_GATCCCCAGCCTATGTCTTTCCCTCTGAACATTTTGTACTAAGCTGGAGCCT CTTTACAAAGTAAAAAAGCTCTGGCATGGAGACACTGAACACTGCCATTAGGAGTCCCCC GAGTATCA **REVERSE**_CTGTTAATTCAGAGAATCACTTCTTCACACCACAAGGAGTTTATATCCTCATTC CTCCTGCCATTCCTTATCTCTTTAACTGTGGGTGAGCTGTGTTATGTTATAAAGTGAAGGG AGATC
Hbb-b1	**FORWARD**_GATCCACATGCAGCTTGTCACAGTGGAGCTCACTGAGGCTGGCAAAGGTGC CCTTGAGGCTGTCCAAGTGATTCAGGCCATCGTTAAAGGCAGTTATCACTTTCTTGCCATG GGCCTTCA **REVERSE**_TTGCACCAATAACATAGCTTGTCTGTTCAATATTATCTCTTCAATATTGTTTACT ACACAAATAAAAAATTAAAACACATTATACATAAGAAAATTTGCCTGTTCTGTTTTGTTTTGA TC
Hbb-b2	**FORWARD**_GATCCACATGCAGCTTGTCACAGTGGAGCTCACTGAGGCTGGCAAAGGTGC CCTTGAGGCTGTCCAAGTGATTCAGGCCATCGTTAAAGGCAGTTATCACTTTCTTGCCATG GGCCTTCA **REVERSE**_TTTCCCCACAGGCCAGAGACAGCAGCCTTCTCAGCATCAGTCAGGTGCACCA TGATGTCTGTTTCTGGGGTTGTGAGTCAACACAACTATGTCAGAAGCAAATGTGAGGAGC AACTGATC
Hbb-y	**FORWARD**_CAACAACAGGAAGTGACAGTGTAGAGAAGATGGATATAACAAGTTCTTGTTT TGTGATTCTAGAGAGTTCTCTGACCAAACCTTTGTGAACCTTACTCCGTTAGTAATATTGCT TTGATC **REVERSE**_GATCACAGGAGTGTCAGAAGCAAGTATGTACTGCTGCTAGAAGTGGTGGCCT TTTATGCTTCACAGACAATGTTCTGACCCTTTGTTCTGCATTATTCTTTGAAGCTATTGGTC AGCCTG
Hbb-bh1	**FORWARD**_GATCTCAAGTGTGCAAAAGCAAGTGTATTCAAGAGGCTGAAGACCTGTCCTT TTATTCTTCACCCTGGACCTTGCTCCCCTTACCCCTCCCCAGGACTCTGAGGCTATTGGTC AAGAGAG **REVERSE**_TTAGGAATGGGGTTACGAACCTAAGTGATAGAAATATGAGTCTTACTCACAAG CCACGTATAAGAAAGATGTGGTGTGCAAAACCATTGAATGCATTTTCAGTAGGCTTAGGAC ATGATC
Bcl11a	**FORWARD**_GATCCCGGGTCCCTGACTTGGCCTTGGGGGCGCGCTCCCGGCTCTGGAGC CCAGGCCTGCGCCGTCAGCCTCGCCCCAACCACCTCTCTCCAGGCGGCGTCGGTTTCCC CGGGCCCTGAC **REVERSE**_TAGTCACCCCCACAATAGTACGAAAGTGGGACTGTAGAAGAGGGCCCCCGG CGCTTCCCAGTCAGCTGAGTCGGGAGAGGGGCTGCAGCGGCTGGGAAGGCCGGGTGAC CGGGAAGGATC

2.7.4 Bisulfite sequencing

High quality genomic DNA was bisulifte-converted using the EZ DNA Methylation Gold Kit (Zymo Research) following the manufacturer's instructions. Enzymatically methylated human genomic DNA was used as the methylated control (Millipore). Nested primer sets were designed to cover the *HBZ* gene and its promoter (Table 2.9). PCR was performed with the HotStart Taq Polymerase according to the manufacturer's instructions. Products were then run out on a 1% Agarose gel, and gel extracted using the Zymoclean Gel DNA recovery kit (Zymo Research) according to the manufacturer's instructions. Libraries were generated using a NEBNext Ultra DNA Library Prep Kit (NEB) according to the manufacturer's instructions, and sequenced using an Illumina Miseq v2 500 cycle kit.

2.7.5 Array comparative genomic hybridization

Array comparative genomic hybridization (aCGH) data from high quality genomic DNA were generated using an Illumina InfiniumOnmiExpress-24v1-2-A1 Beadchip by Samantha Knight according to the manufacturer's instructions.

Target		Sequence 5' to 3'	Sequence 5' to 3'
HBZ promoter and gene	1A	F GGAAAGTAATATTGTAATGT	R AAATCTAACCACAAAATCTA
	1B	F GTATATGATATTAGGTGAAT	R ATCACTCCTACATTTATAATC
	2A	F GTTGAGTATTATTAAGGTTAG	R CTATCCTCTCCCCACAATAC
	2B	F GTAGGTTTAATTTTAGTGTAG	R CACTATCTCCTCTCCATAATC
	3A	F GATTATGGAGAGGAGATAGTG	R CAATACCCTAATCCCAAATAC
	3B	F GTATTGTGGGGAGAGGATAG	R AACTATAATACCCTAACTAC
	4A	F GTATTTGGGATTAGGGTATTG	R CCTCCCAAACTACCCTAAAC
	4B	F GTATTTGATTTTTTGGGAGG	R ACCCCCCAAACCTCTACAAC
	5A	F GTTGTAGAGGTTTGGGGGGT	R AAAAATAAATCTTAATCTA
	5B	F GTTTAGGGTAGTTTGGGAGG	R AAAAATAAATCTTAATCTA
	6A	F AGTTTTTGTTTTATTGTTTG	R ACTACACATCCCTCTACCTC
	6B	F GTTTTATTGTTTGTTGGTTATTTTG	R CTACCTCTTAAAAACCCTAA
	7A	F TTTATTTGGTTTTTGTTGTT	F CATAACCTAAAAACATACCT
	7B	F GAGGTAGAGGGATGTGTAGT	R CTTATACCCAATAAATATTTC
	8A	F GAAATATTTATTGGGTATAAG	R CAATTAATCCTAATAATAAAC
	8B	F AGGTATGTTTTTAGGTTATG	R CTAAATTTTCAAATCTTTAAC
ACTB	1A	F GAGGGGGTTGGGGTTTGG	R CTACCCCACCCAACCAACT
	1B	F GTTTGTTTTTTATGGTAATAA	R CTCCCCTACCTAATACCTA

Table 2.10 Nested primers used for Bisulfite sequencing. A denotes the outside primer set and B, the internal primer set.

2.8 Computational Analysis

The following software packages and scripts were used:

- Bedtools 2.25.0 (Quinlan et al., 2010)
- ROSE tool (Whyte et al., 2013)
- Samtools 0.1.19 (Li et al., 2009)
- MACS 2.0.10 (Zhang et al., 2008)
- UCSC browser (Kent et al., 2002)
- Bismark v4.1.2 (Krueger and Andrews, 2011)
- UCSC tools 1.0
- SeqMonk 0.24.1

  (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/)
- STAR 2.4.2a (Dobin and Gingeras, 2016)
- Deeptools 2.2.2 (Ramírez et al., 2014)
- GenoSTAN (Zacher et al., 2016)
- FASTQC 0.11.4 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc)
- https://github.com/telenius/captureC/releases
- https://github.com/djdownes/CaptureCompare
- https://github.com/Hughes-Genome-Group/NGseqBasic/releases

2.8.1 Analysis of ATAC-seq and ChIP-seq data

Sequencing quality was assessed by FASTQC. ChIP-seq and ATAC-seq data were aligned to the mm9 (for mouse) and hg19 (for human) data using a customised in house pipeline (Telenius et al., 2018) (script: https://github.com/Hughes-Genome-Group/NGseqBasic/releases). Output BAM files were sorted and indexed using Samtools, then normalised using Deeptools to RPKM (--normalizeUsingRPKM), then converted to a bigwig using USCS tools. Data were visualised in the USCS genome browser. Published datasets were downloaded from the NCBI Gene Expression Omnibus (GEO) database. FASTQ files were then processed as described. Publicly available datasets used in this thesis are shown in Table 2.10.

Cell Type	Data Type	GEO Code	Publication
Ter119 + Splenic erythroblasts	H3K4me1 ChIP-seq H3K27ac ChIP-seq	GSE27921	(Kowalczyk et al., 2012b)
Ter119 + Splenic erythroblasts	CTCF ChIP-seq H3K27me3 ChIP-seq H3K4me3 ChIP-seq	GSE97871	(Hanssen et al., 2017)
WT HUDEP-2	LRF ChIP-seq	GSE74977	(Masuda et al., 2016)
HUDEP-2 BCL11A-ER-V5	BCL11A ChIP-seq	GSE103445	(Martyn et al., 2018)
HUDEP-2 (ΔGγ)	LRF ChIP-seq	GSE103445	(Martyn et al., 2018)
WT HUDEP-2	BCL11A CUT&RUN	GSE104676	(Liu et al., 2018)
Primary Proerythroblasts from $CD34^+$ culture	H3K9me3 ChIP-seq H3K27me3 ChIP-seq	GSE36994	(Xu et al., 2012)

Table 2.11 Publicly available data-sets used for ChIP-seq.

2.8.2 Identifying regions of differential ChIP-seq enrichment

The MACS peak finding software was used to peak call in an unbiased manner, with an FDR (-q) of 10^{-3} on each of the replicates; only peaks which were present in 2 or 3 out of 3 of the replicates were used. An input track processed at the same time was used as background. Intersecting peaks were determined using the intersect feature of bedtools after extending the peaks ±250bp.

2.8.3 Classification of Enhancer Elements and Super-enhancer analysis

Enhancer and promoter annotations were generated by Ron Schwessinger (Computational Research Biology Group, WIMM) using GenoSTAN. All ATAC-seq hypersensitive regions were sub-classified based upon their H3K4me1, H3K4me3, H3K27ac, promoter annotation and CTCF binding.

Super-enhancers were called using the ROSE tool as described(Whyte et al., 2013). Individual enhancers were stitched together to form a single enhancer domain. Stitched enhancers were then ranked for H3K27ac signal, which was normalized to a maximum value of 1.0. Regions ± 250bp of annotated TSS were excluded. Enhancer domains were called as super-enhancers if H3K27ac signal was above a threshold placed where the tangent to the graph ranking enhancers was equal to 1.

2.8.4 Analysis of Capture-C data

Sequencing data was assessed by FASTQC. Custom scripts were then used (https://github.com/telenius/captureC/releases) to generate interaction plots as described (Davies et al., 2016). To determine interactions between definitive and primitive erythroid cells, read counts were normalized to the total number of detected interactions, to find the number of interactions per 100,000 interactions. Data were exported as a BigBed file, then visualized in the UCSC genome browser. Statistically significant changes were calculated on non-normalized read counts per restriction fragment from biological triplicates using the DeSeq2 R package.

For plotting of multiple interaction profiles, Capture-C interactions were binned (into 250bp bins) and a sliding 5kb window used. The mean of three biological replicates and SD were plotted in R using the following script (https://github.com/djdownes/CaptureCompare).

2.8.5 Mapping and visualisation of RNA-seq data

FASTQ files for RNA-seq data were downloaded (GEO accession number: GSE71422) and aligned to the hg19 genome build using STAR. Data were then normalised using the standard recommended parameters: --outFilterType BySJout –outFilterMultimapNmax 20, --alignSJoverhangMin 8, --alignSJDBoverhangMin 1, --outFilterMismatchNmax 999. Deeptools was then used to calculate normalised (RPKM) and strand specific data(RPKM) prior to merging using bedtools and direct visualisation on the USCS genome browser.

2.8.6 Analysis of aCGH data

This was performed by Dr Samantha Knight at the Wellcome Trust Centre for Human Genetics, Oxford University). Briefly, data were processed using GenomeStudioV 2009.2 (Illumina) and analysed using Nexus 8.0 Discovery Edition (BioDiscovery). The following settings for single nucleotide polymorphism (SNP) Rank Segmentation (based on circular binary segmentation statistics) were used: Significance Threshold =$1.0x10^{-6}$, Max Contiguous Probe Spacing (Kbp) = 500.0, Minimum number of probes per segment = 10, High Gain = 0.5, Gain = 0.12, Loss = -0.12, Big Loss = -0.7, 3:1 Sex chromosome gain =1.2, Homozygous Frequency Threshold = 0.98, Homozygous Value Threshold = 0.85, Heterozygous Imbalance Threshold = 0.46, Minimum loss of heterozygosity (LOH) Length (Kb = 500, Minimum SNP Probe Density (Probes/Mb = 0).

2.8.7 Analysis of Bisulfite sequencing data

This was performed by Dr Danuta Jeziorska at the WIMM. Read quality was checked by FASTQC and adapters trimmed. Data were then aligned to a custom hg19 (previous prepared by Simon McGowan) using bismark and seqmonk, then visualised in the UCSC genome browser.

2.8.8 Analysis of single cell RT-qPCR data

This was done by Rob Beagrie (WIMM) using Python (v3.5.2; Matplotlib v2.0.2; Numpy v1.14.5; Pandas v0.23.3; scikit-learn v0.19.2; Scipy v1.1.0). Ct values beyond the limit of detection or marked as "Fail" by the instrument were set to the limit of detection value (Ct = 40). Raw Ct values were transformed into expression space by calculating 2^-Ct.

For the *Lrf*$^{-/-}$ knockout experiments expression of all genes was scaled to a range of 0 to 1, then the mean of all scaled expression values was calculated for each cell and used to normalise expression values. For the experiments on the *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ mice, then three housekeeping genes (Cox6c, Myl6 and Sec61g) were run on both 192.24 chips. The expression of these three genes was scaled to a range of 0 to 1, then the mean of the three scaled expression values was calculated for each cell and used to normalise expression values for the remaining 21 assays on each chip.

2.8.9 Data visualisation and graphs

Graphs were generated using GraphPad Prism v8.0; statistical analysis was performed where relevant with the statistical test used given below each figure in the thesis. Next generation sequencing data was visualised in the UCSC genome browser. Aesthetic modifications only were made with Adobe Illustrator v23.0.1.

3 The epigenetic landscape and self-interacting domain in primitive erythroblasts at the murine α-globin gene locus

3.1 Introduction

In definitive murine erythroblasts, the α-globin gene locus has been extensively characterized. There are five enhancer elements, four of which have sequences that are highly conserved across species (the "multispecies conserved sequences", MCS) with one mouse-specific element (Hughes et al., 2005). The MCS sequences are denoted MCS R1-R4 with the "R" indicating a regulatory element (hereafter simply called R1-R4); the mouse specific element is denoted Rm (Hughes et al., 2005). Four of these enhancer elements lie within the introns of the *Nprl3* housekeeping gene (R1, R2, R3 and Rm) which lies upstream of the globin genes, with R4 lying in an intergenic region. When stitched together, this enhancer cluster meets the definition of a super-enhancer in definitive erythroid cells; indeed, it is the second most highly scoring super-enhancer when ranking enhancers by MED1 occupancy (Hay et al., 2016).

In vivo deletion of each of the enhancer elements individually and informative combinations indicate that these enhancers function additively and independently with respect to α-globin gene expression in definitive erythropoiesis, and are not equivalent (Hay et al., 2016). R1 and R2 account for 40% and 50% of the

transcriptional output respectively (with removal of both enhancer elements leading to a 90% drop in α-globin transcription) with R4 contributing 10% of the transcriptional output. Removal of Rm or R3 does not appear to affect α-globin transcription (Hay et al., 2016).

Activation of α-globin in definitive erythroid cells is associated with a broad domain of histone acetylation and the histone modification H3K4me1 spanning approximately 80kb; this includes the α-globin genes and their associated enhancer elements (Kowalczyk et al., 2012b). In definitive erythropoiesis, this domain is flanked by convergent CTCF sites, and it is these that allow the formation of a sub-TAD or contact domain, probably via loop extrusion (Hanssen et al., 2017). Removal of the flanking upstream CTCF sites allows the enhancer elements to act upon, and activate, the adjacent gene (*Rhbdf1*) normally silenced via the PRC2 complex in definitive erythroid cells (Hanssen et al., 2017); this demonstrates their importance in constraining enhancer activity and mediating appropriate gene expression within the locus.

Previous unpublished work in our laboratory, which is outlined in greater detail in the introduction, has demonstrated that in erythroid cells taken from E10.5 in the mouse (when ζ-globin accounts for approximately 40% of transcriptional globin output from the α-globin locus), the same five enhancer elements (R1-R4 and Rm) are present in primitive erythropoiesis. The contribution of each enhancer element to both α-globin and ζ-globin transcriptional output has also been defined in primitive erythropoiesis, using the same knockout mouse models as those used in definitive erythroid cells (Hay et al., 2016).

The effects on α-globin transcription in the mouse models are broadly similar to those seen in definitive erythropoiesis with the R1 and R2 enhancers being responsible for ~90% of transcriptional output of α-globin in primitive erythroid cells. By contrast, knockout of each enhancer element individually has no detectable effect on ζ-globin expression. Homozygous knockout of R1 and R2 together leads to a significant reduction in ζ-globin to approximately 50% of that seen in WT cells; however, this is fivefold less than the effect of deleting these enhancers on α-globin expression in either primitive or definitive erythropoiesis (where 10% of normal expression is maintained). It thus appears that the enhancer elements act in a redundant fashion on ζ-globin (with one enhancer being capable of compensating for the other).

Outstanding questions related to the epigenetic landscape of the α-globin gene cluster in primitive erythroid murine cells include:

1) CTCF binding throughout the locus (and in particular whether there is an embryonic specific CTCF site acting as a boundary element); the histone modification signatures associated with active promoters and enhancers within the locus (H3K4me3 and H3K4me1 respectively) and the domain of histone acetylation when ζ-globin is active.

2) The contact domain from the perspective of the boundary CTCF elements; the enhancers, the *Hba-x* promoter and the *Hba-a1/2* promoters and how this compares to that seen in definitive erythropoiesis.

3) Chromatin accessibility at the locus during development from E9.5 to E11.5 as ζ-globin is silenced in primary erythroid cells.

Here, I have used a combination of ChIP-seq, ATAC-seq and NG-Capture C to interrogate the α-globin locus and its chromatin environment in primary primitive murine erythroid cells, thereby addressing these issues.

3.2 Results

3.2.1 Source of primary primitive and definitive erythroid cells

Primitive erythroid cells were isolated at E10.5 from mouse embryonic blood. Because primitive erythroblasts mature in a semi-synchronous wave, all cells were at a similar stage of development, judged by morphology (Figure 3.1a). In agreement with previous studies (Kingsley et al., 2006), ζ-globin expression accounted for approximately 40% of total transcriptional globin output at this stage (Figure 3.1b). EMPs make up <5% of cells when performing a differential count by size.

Definitive erythroid cells were obtained from a well characterized fetal liver culture system based upon a previously published protocol (Drissen et al., 2005) (explained in detail in Materials and Methods). Expansion and differentiation yielded a homogenous population of intermediate erythroblasts by both morphology and flow cytometry (Figures 3.1c, d). RT-qPCR confirmed no ζ-globin expression (Figure 3.1e).

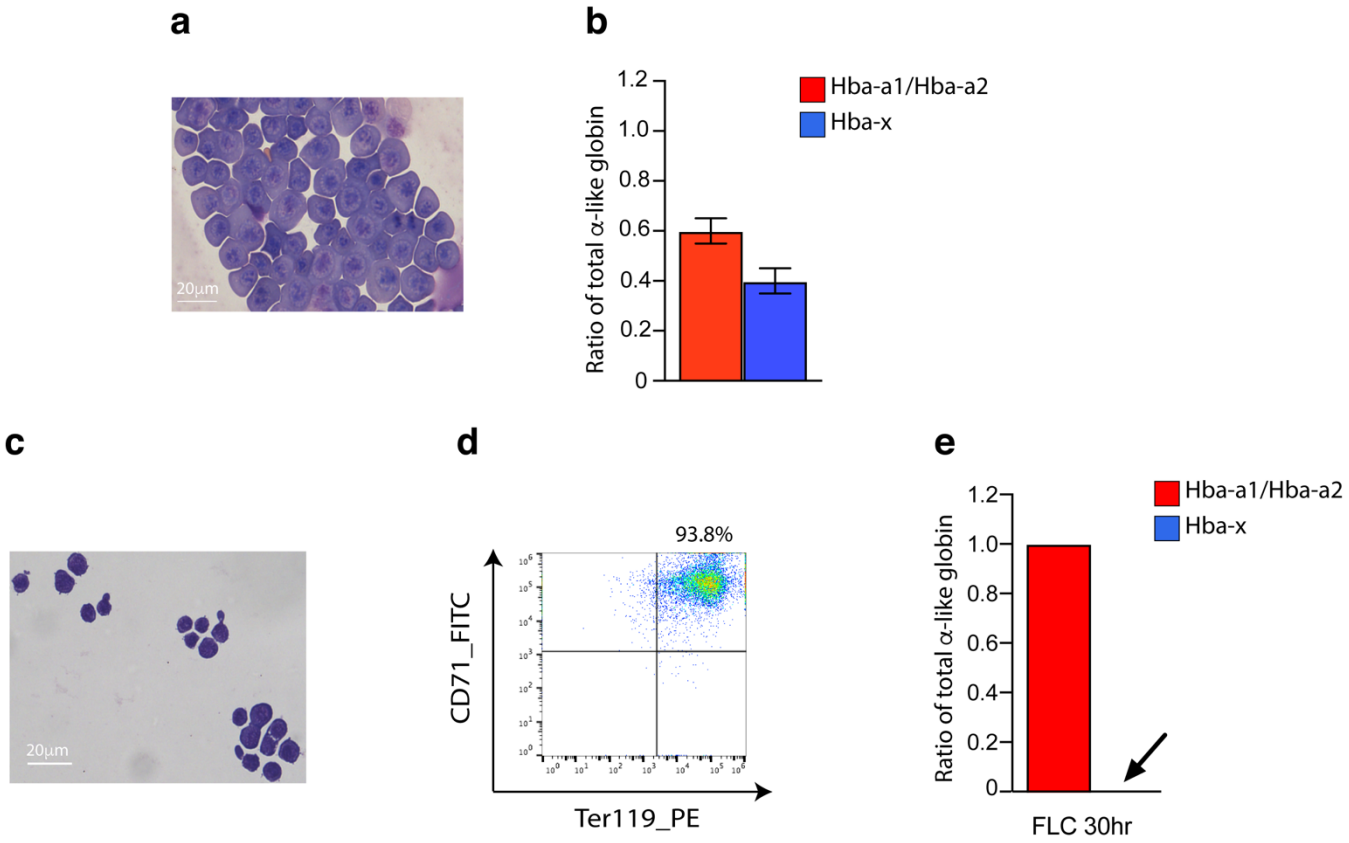


Figure 3.1 Primitive and definitive erythroid cells used for experiments.

Primitive cells were obtained from E10.5 embryos. At this stage, they are a homogenous population of intermediate erythroblasts, considerably larger than definitive cells (a) and express approximately 40% *Hba-x* as a proportion of the total α-like globin on a population basis, measured by RT-qPCR (b). Definitive erythroid cells were obtained from a fetal liver (FL) culture system. At 30 hours, they form a homogenous population of intermediate erythroblasts (c), which are $CD71^{+}Ter119^{+}$ (d) and do not express ζ-globin by RT-qPCR (e). The arrow indicates the lack of *Hba-x* expression.

3.2.2 In primitive erythroid cells, the *Hba-x* gene is accessible and is marked by histone acetylation

ChIP-seq experiments were performed for CTCF, histone acetylation (H3K27ac) and the histone modifications associated with active enhancers (H3K4me1) and promoters (H3K4me3) in primitive erythroid cells, alongside an assay for open chromatin, ATAC-seq. These data at the α-globin locus are shown in Figure 3.2, together with previously published ChIP-seq experiments in definitive erythroid cells shown in Figure 3.3.

Using ATAC-seq, the only significant changes in chromatin accessibility at E10.5 (as assessed by peak calling using MACS2, with an FDR adjusted p-value of 0.01) are seen at the *Hba-x* promoter, its 3´ region and the *Rhbdf1* gene promoter. *Rhbdf1* produces an intramembrane serine protease involved in cell signalling which is widely expressed in most tissues, though not in definitive erythroid cells. Both *Hba-x* and *Rhbdf1* are marked by the histone signature associated with active promoters (H3K4me3) in primitive erythroid cells, in contrast to that seen in definitive erythropoiesis.

The ATAC-seq track confirms that the same five enhancers elements seen in definitive erythroid cells are also present during primitive erythropoiesis. These enhancers all have the classical enhancer histone signature H3K4me1, which extends somewhat closer to the *Hba-x* gene promoter than in definitive erythroid cells.

H3K27ac ChIP-seq shows a continuous domain of histone acetylation extending across the locus, and including the *Hba-x* gene promoter and gene body. This contrasts with that seen in definitive erythropoiesis, where there is localized histone deacetylation over the *Hba-x* gene. In definitive erythroid cells, the H3K27me3 mark associated with PRC2 repression is absent over the *Hba-x* gene promoter, indicating that PRC2 plays no role in its silencing; the *Rhbdf1* gene serves as a positive control for the assay.

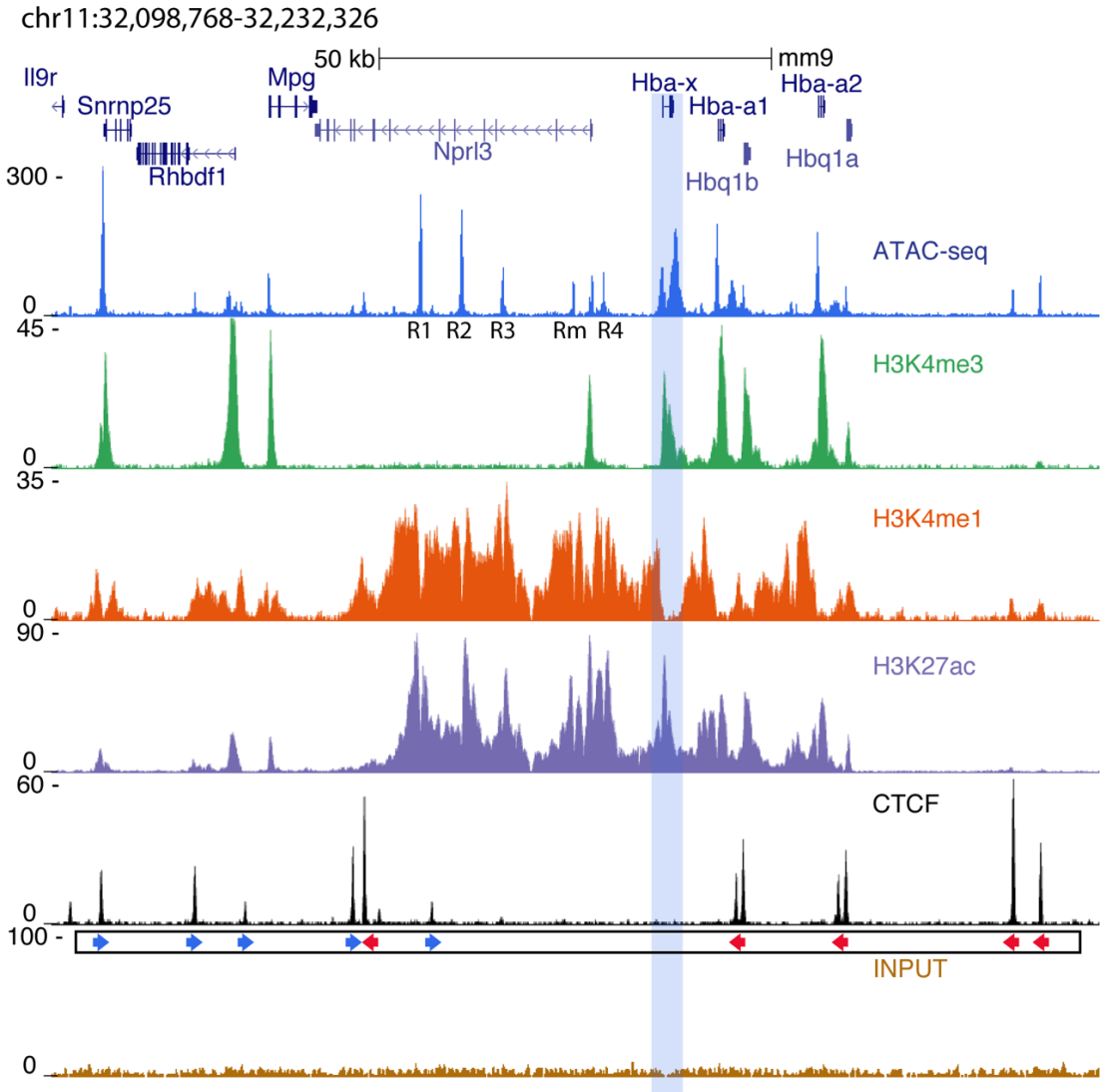


Figure 3.2 (previous page) ChIP-seq for H3K4me3, H3K4me1, H3K27ac and CTCF in primitive erythroid cells at the α-globin locus. An ATAC-seq track is shown for orientation. The *Hba-x* gene is highlighted in blue. The orientation of CTCF sites is shown below the CTCF track. Data represent normalized (RPKM) tracks from three biological replicates.

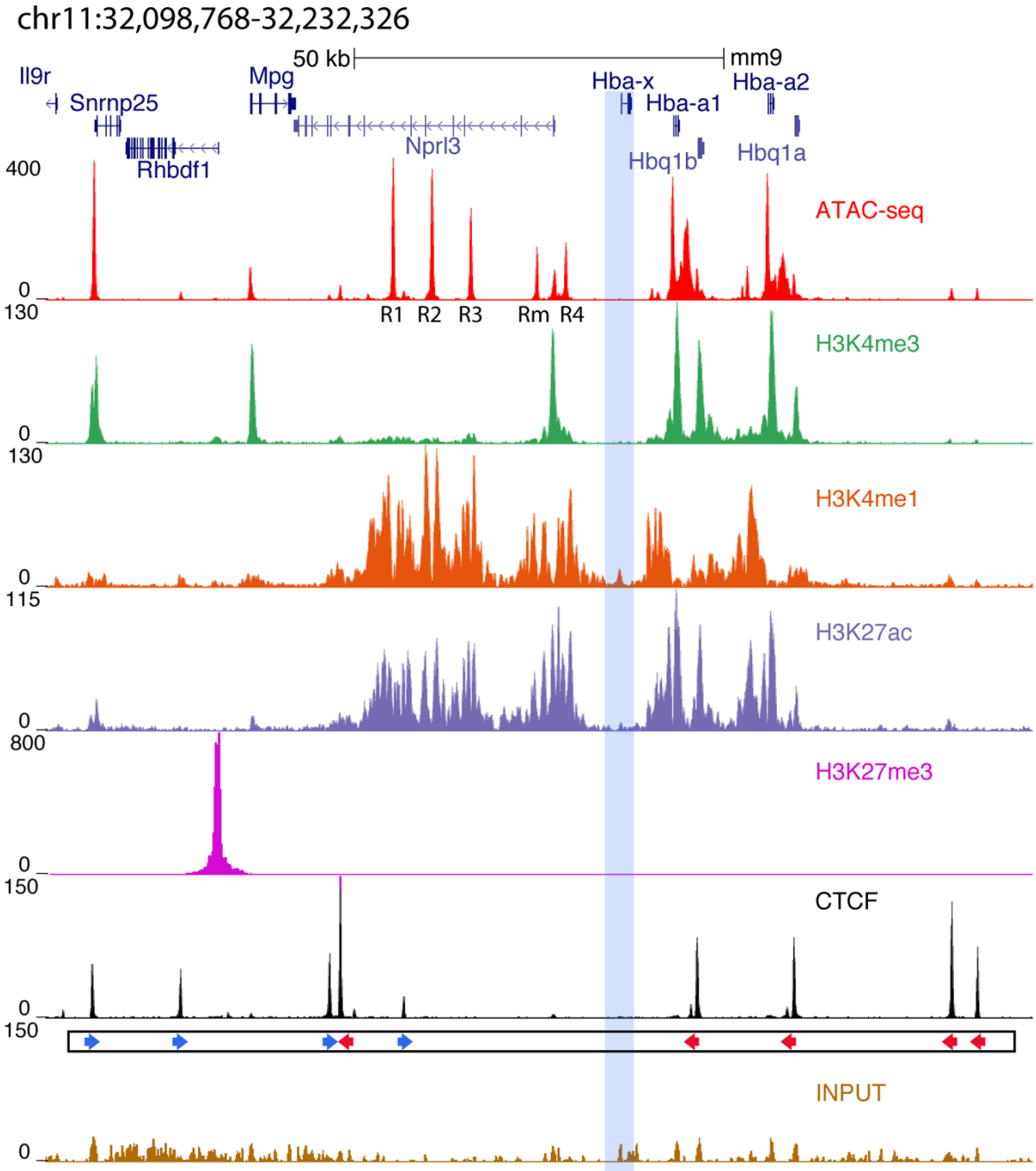


Figure 3.3 (previous page) ChIP-seq for H3K4me3, H3K4me1, H3K27ac, H3K27me3 and CTCF in definitive erythroid cells at the α-globin locus. An ATAC-seq track is shown for orientation. The *Hba-x* gene is highlighted in blue. The orientation of CTCF sites is shown below the CTCF track.

3.2.3 CTCF binding throughout the α-globin locus is broadly unchanged between primitive and definitive erythroid cells

ChIP-seq for the architectural protein CTCF in primitive erythroid cells shows binding to the same sites within the locus as seen in adult erythroid cells. In addition, adjacent to the *Rhbdf1* gene, a CTCF binding site specific to primitive erythroid cells can be seen (Figure 3.4). Binding at this site is also seen in embryonic stem (ES) cells.

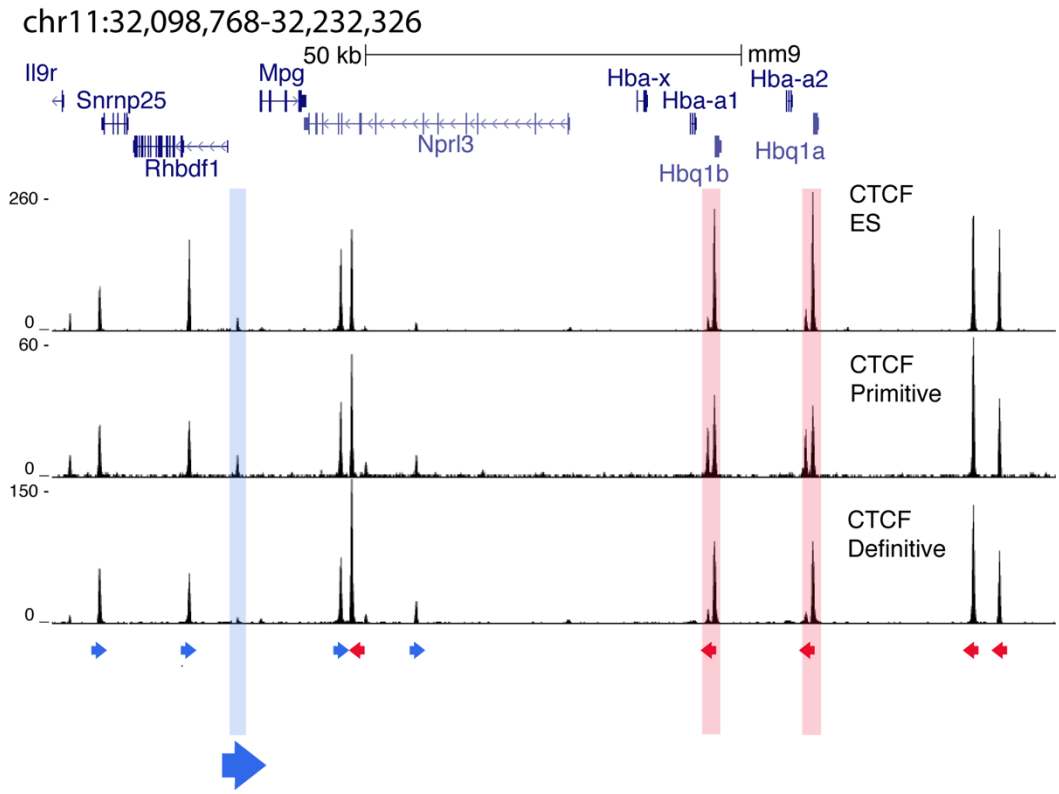


Figure 3.4 (previous page) CTCF binding throughout the locus in embryonic stem (ES) cells, primitive, erythroid cells and definitive erythroid cells. An additional CTCF site is present adjacent to the *Rhbdf1* promoter in ES and primitive cells (highlighted in blue, along with its determined orientation below). The binding of CTCF to one of the sites adjacent to the θ-globin gene appears increased in primitive cells when compared to definitive and ES cells (highlighted in red).

Previous analysis of the CTCF binding orientation at the α-globin cluster has revealed a striking pattern of orientation, with two domains of convergent CTCF sites flanking the α-globin enhancer cluster, shown in Figure 3.2 (Hanssen et al., 2017). The CTCF orientation, as determined bioinformatically, confirms that the peak unique to ES and primitive erythroid cells obeys this observation, facing towards the enhancer cluster.

Separately the CTCF binding at the θ-globin genes appears to have a different profile. There are two adjacent CTCF sites at this location, which are also seen in both ES and definitive erythroid cells. In primitive erythroid cells, the upstream peak adjacent to the θ-globin genes appears to have increased binding (highlighted in red in Figure 3.4), the significance of which is uncertain.

Given the similarities in CTCF binding at the α-globin locus, the data were then analysed genome-wide to determine the number of sites where CTCF occupancy changes between primitive and definitive erythropoiesis (Figure 3.5).

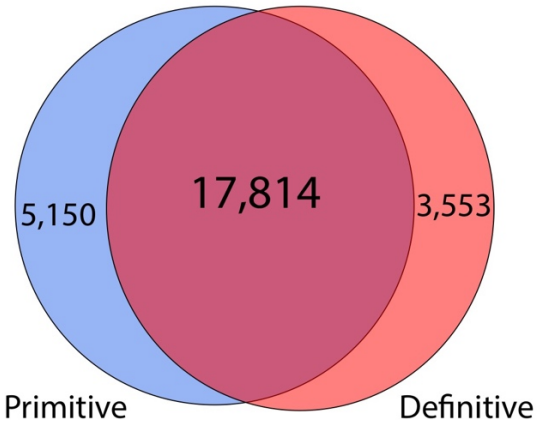


Figure 3.5 CTCF occupancy overlap in primitive and definitive erythropoiesis. Peaks called using MACS 2 with an FDR corrected p value of <0.01.

The majority of CTCF sites (~80%) remain unchanged between the two cell types.

3.2.4 The α-globin enhancer cluster classifies as a super-enhancer (SE) in primitive erythroid cells

In definitive erythroid cells, the five α-globin enhancer elements taken together meet the definition of a SE (Hay et al., 2016).

Given the generation of genome-wide data above, it was possible to assign accessible elements identified in primitive erythroid cells using ATAC-seq as putative enhancers, promoters or CTCF elements, and then use this together with the ChIP-seq for H3K27ac to identify super-enhancers.

Together with assistance from Ron Schwessinger (a D.Phil student in the Centre for Computational Biology), 10 chromatin states were identified using GenoSTAN (Zacher et al., 2017) based upon the presence of CTCF, H3K4me1, H3K4me3 and H3K27ac (Figure 3.6). The enhancer states were then taken forward to determine the elements classifying as super-enhancers.

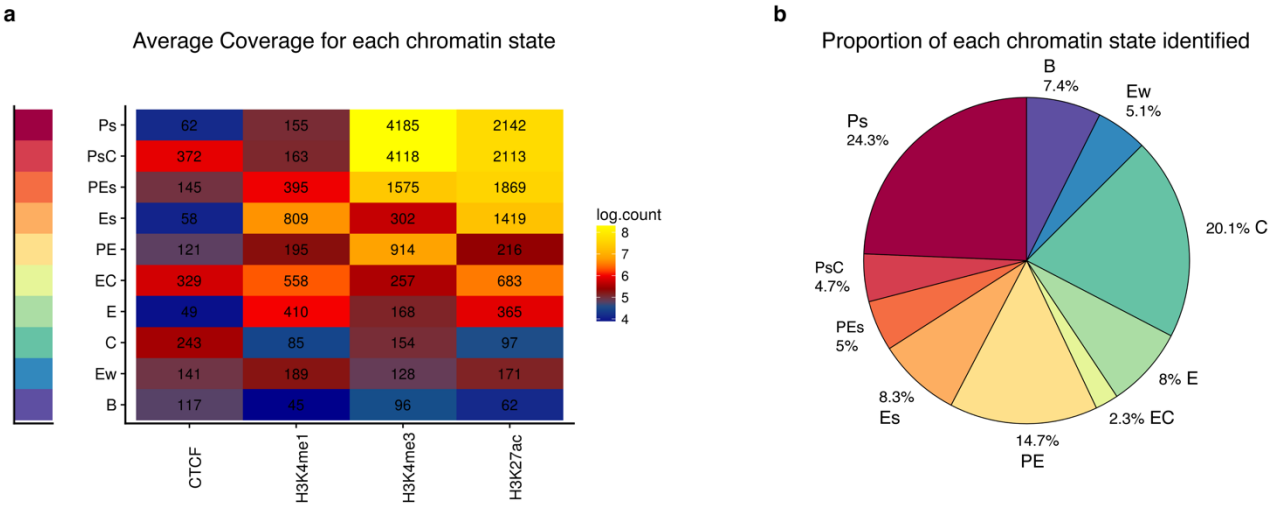


Figure 3.6 Chromatin states in primitive erythroid cells determined using GenoSTAN. a. 10 Chromatin states were identified using the read count for CTCF, H3K4me1, H3K4me3 and H3K27ac: Ps (strong promoters), PsC (strong promoters with CTCF binding), PEs (Strong promoters with an enhancer mark), Es (strong enhancers), PE (promoters with an enhancer mark), EC (enhancers with a CTCF mark), E (enhancers), C (CTCF sites), Ew (weak enhancers) and B (background). **b.** The proportion of each chromatin state identified.

The enhancer states (EC, Es, E and Ew) were combined and then used as input for the ROSE tool (Whyte et al., 2013). Individual enhancer regions lying within 12.5kb of each other were stitched together to form a single genomic region and then ranked based upon H3K27ac signal (Whyte et al., 2013); (Figure 3.7).

As in definitive erythroid cells, both the α and β-globin enhancer clusters classify as super-enhancers, and are the top two super-enhancers when ranking by H3K27ac signal.

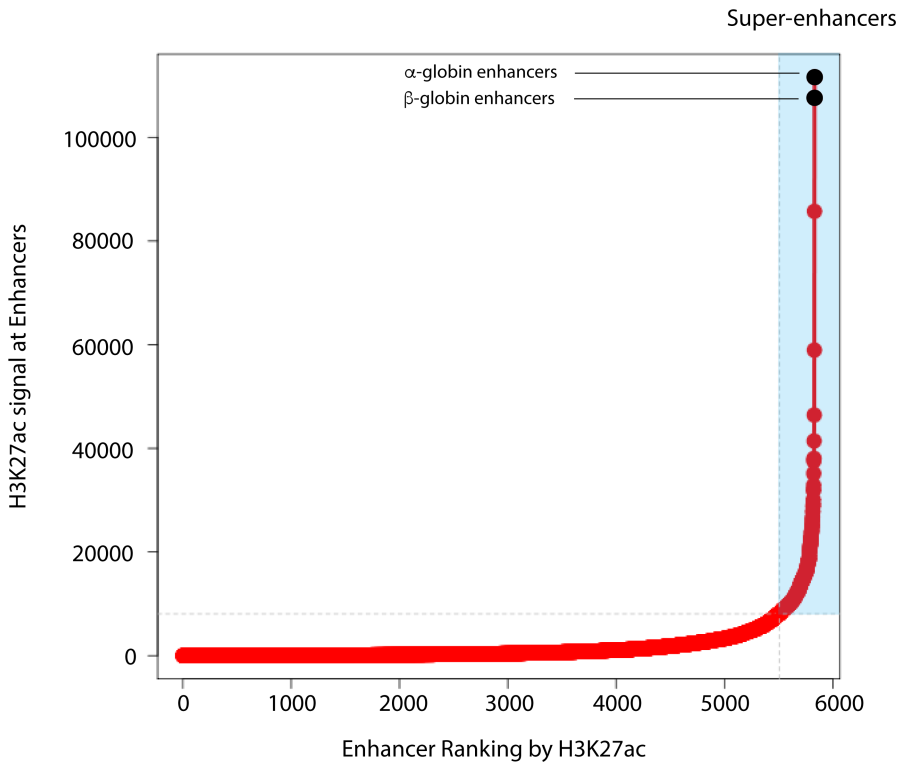


Figure 3.7 Super-enhancers in primitive erythroid cells. 327 super-enhancers were identified in primitive erythroid cells. The two top-ranking super-enhancers are the α-globin and β-globin enhancer cluster.

3.2.5 The chromatin compartment formed in primitive erythropoiesis is almost identical to that found in definitive erythroid cells

In definitive erythroid cells, a contact domain, or sub-TAD, is formed and delimited by flanking CTCF sites located at positions annotated as HS-38 and HS-39 (forming

the upstream boundary) and sites annotated as HS44 and HS48 (forming the downstream boundary) (Hanssen et al., 2017), with the numbers indicating the distance from the transcriptional start site of *Hba-x* in kb. To investigate whether the same contact domain is formed in primitive erythroid cells, NG Capture C was performed from three viewpoints: the *Rhbdf1* gene promoter (which lies outside the contact domain), the proximal boundary CTCF element (HS-38) of the contact domain, and the R1 erythroid enhancer (which lie inside the contact domain), with the help of Dr Damien Downes (Figure 3.8). Experiments were performed in primary primitive erythroid, definitive erythroid and mouse ES cells, the latter serving as a control when both the globin genes are inactive, and no contact domain is formed (Hanssen et al., 2017).

The interactions from the *Rhbdf1* promoter illustrate that, as in definitive erythroid cells, the genes lying upstream of this region are excluded from this domain, despite the *Rhbdf1* gene being active in primitive erythropoiesis. The interactions from the HS-38 CTCF element form the upstream boundary, interacting with the HS44 and HS48 CTCF elements and the CTCF elements adjacent to the θ genes. This viewpoint also shows subtle, but increased interactions with the CTCF element located within the intron of *Rhbdf1* (HS-59) in primitive erythroid cells in comparison to definitive.

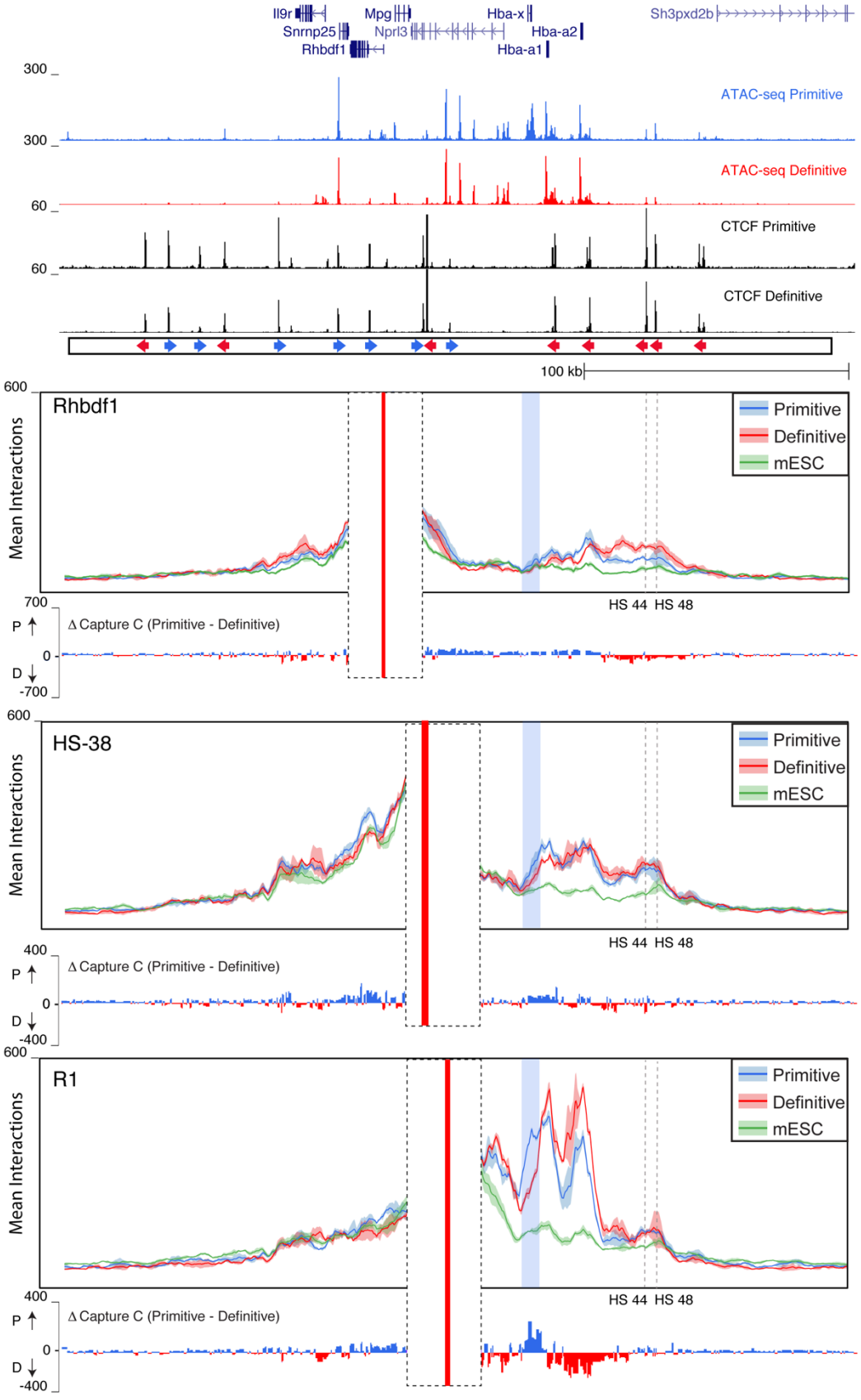
Assessment of chromatin interactions from within the contact domain (the R1 viewpoint) confirm that the interactions within the compartment are constrained by the HS-38 and HS44/48 CTCF sites, with interactions dropping steeply beyond these elements. This viewpoint also shows significantly increased interactions

between R1 and the *Hba-x* promoter in primitive erythroid cells compared to definitive; these appear to occur at the expense of interactions between the enhancer and the *Hba-a1/a2* promoters.

In summary, these NG-Capture C data show that the compartment formed in primitive erythroid cells is almost identical to that formed in definitive erythropoiesis, with the creation of a clear self-interacting domain containing the α-globin enhancers and the active globin genes, demarcated by the HS-38/-39 upstream CTCF sites, and the CTCF site adjacent to *Hba-a2*. HS44 and HS48 downstream CTCF sites. Chromatin interactions from within the compartment however, are significantly altered between primitive and definitive erythroid cells (Figure 3.8).

Figure 3.8 (next page) NG Capture-C profiles from the *Rhbdf1* promoter, the HS-38 CTCF element and the R1 erythroid enhancer. Overlaid, normalized Capture-C data. The viewpoints are indicated by vertical red bars, and the dashed lines indicate the exclusion zones (which lie too close to the capture viewpoint to judge meaningful interactions). The mean, plus and minus one standard deviation (s.d.), of sliding 5 kb windows are visualized. Differential tracks (ΔCapture-C) show a subtraction (Primitive (P) – Definitive (D)) of the mean number of meaningful interactions per restriction fragment. The experiments represent three biological replicates. Normalised CTCF ChIP-seq and ATAC-seq tracks for primitive and definitive erythroid cells are also shown. The orientation of the CTCF sites is also shown. The HS+44 and HS+48 CTCF sites are highlighted; the *Hba-x* gene is also highlighted in blue throughout the figure.

chr11:32,000,000-32,300,000



3.2.6 The *Hba-x* gene promoter contacts the enhancers at a lower frequency in definitive erythroid cells although it remains within the self-interacting domain

Given the self-interacting chromosomal domain as a whole remains unchanged, but interactions from within the compartment at the R1 enhancer viewpoint are significantly altered in primitive erythropoiesis in comparison to definitive, it was next investigated how interactions change from the perspective of the globin genes in these two cell types (Figure 3.9).

NG Capture-C was performed from the viewpoint of the *Hba-x* gene promoter and the *Hba-a1/2* promoters, again in primitive erythroid, definitive erythroid and mouse ES cells (Figure 3.9).

Whilst interactions between the ζ-globin promoter and the enhancers are significantly decreased between primitive and definitive erythroid cells, it can clearly be seen that the contact frequency does not return to the “control” baseline, of mouse ES cells, where both globin genes are inactive, and the contact domain doesn’t form.

Interactions from the perspective of the α-globin promoters also differ significantly between primitive and definitive erythroid cells. During primitive erythropoiesis, interactions between all the enhancers and the α-globin promoters are decreased; this is particularly marked for the R4 and Rm elements.

3.2.7 The embryonic globin genes at the murine β-globin locus together with an intergenic region form a small self-interacting region in contrast to the changes seen at the α-globin locus

Next, chromatin re-organization at the murine β-globin locus during development was studied, to see whether this is similar to the observations made at the α-globin locus. At the β-globin locus, there is a more defined developmental embryonic to adult globin switch with four genes being expressed during development: *εy, βH1, β1* and *β2*. *βH1* is expressed in the yolk sac at E8.5, with a switch from βH1- to εy-globin expression occurring between E8.5 and E12.5; β1- and β2-globin transcripts are substantially up-regulated between E10.5 and E12.5 (Kingsley et al., 2006). All these genes are driven by a well-defined cluster of enhancers termed the locus control region (LCR) (Blobel et al., 2015).

NG Capture C was performed from the viewpoints of each of the globin genes' promoters (*Hbb-bh1, Hbb-y, Hbb1/2*) in primitive erythroid cells from E10.5, definitive erythroid cells and mouse ES cells (Figure 3.10).

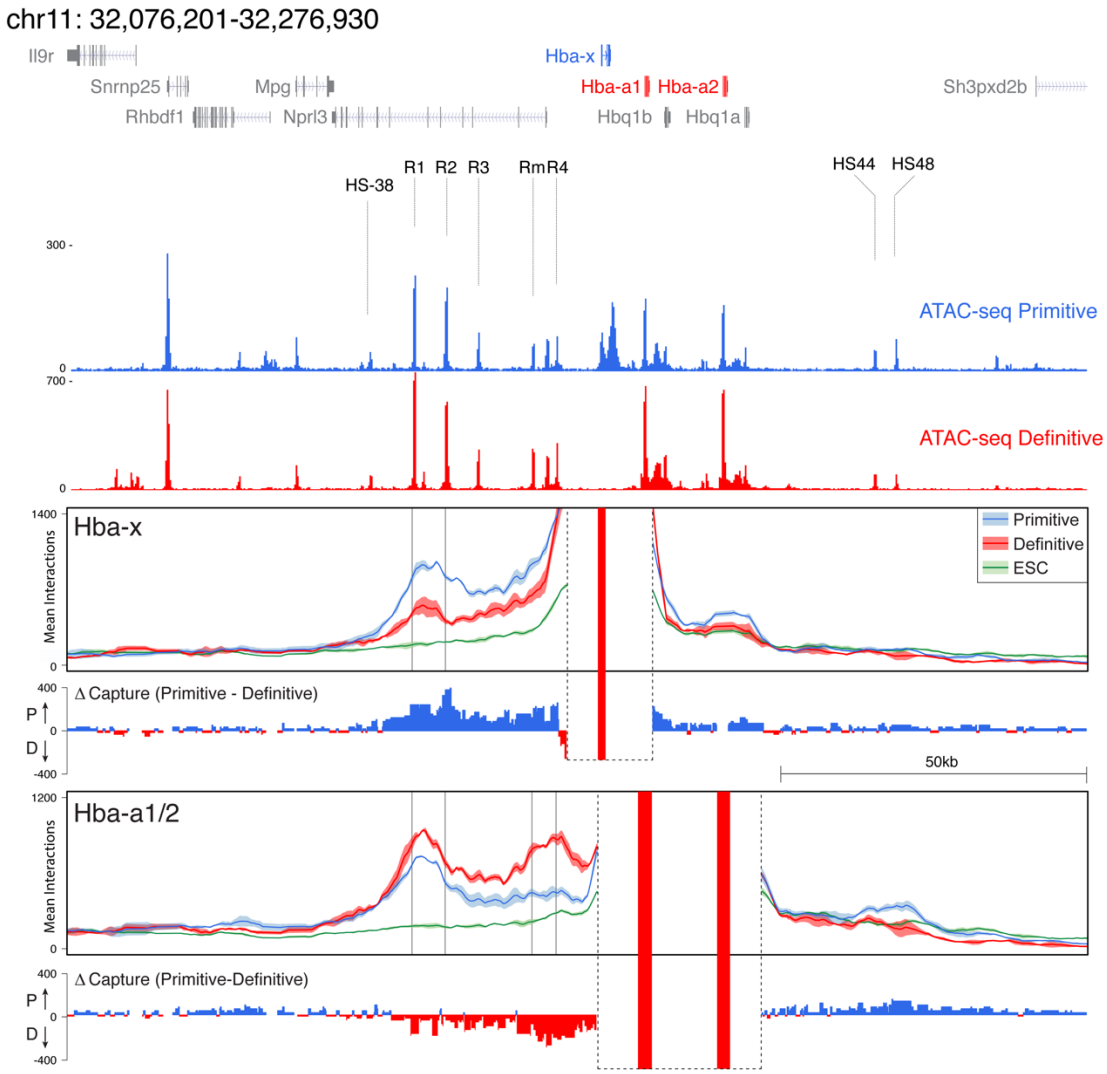


Figure 3.9 NG Capture-C profiles from the Hba-x promoter and the Hba-a1/2 promoters. Overlaid, normalized NG Capture-C data. The viewpoints are indicated by vertical red bars and the dashed lines indicate the exclusion zones (which lie too close to the capture viewpoint to judge meaningful interactions). The mean, plus and minus one standard deviation (s.d.), of sliding 5 kb windows are visualized. Differential tracks (ΔCapture-C) show a subtraction (Primitive (P) – Definitive (D)) of the mean number of meaningful interactions per restriction fragment. The experiments represent three biological replicates. The erythroid enhancers and flanking CTCF sites are shown.

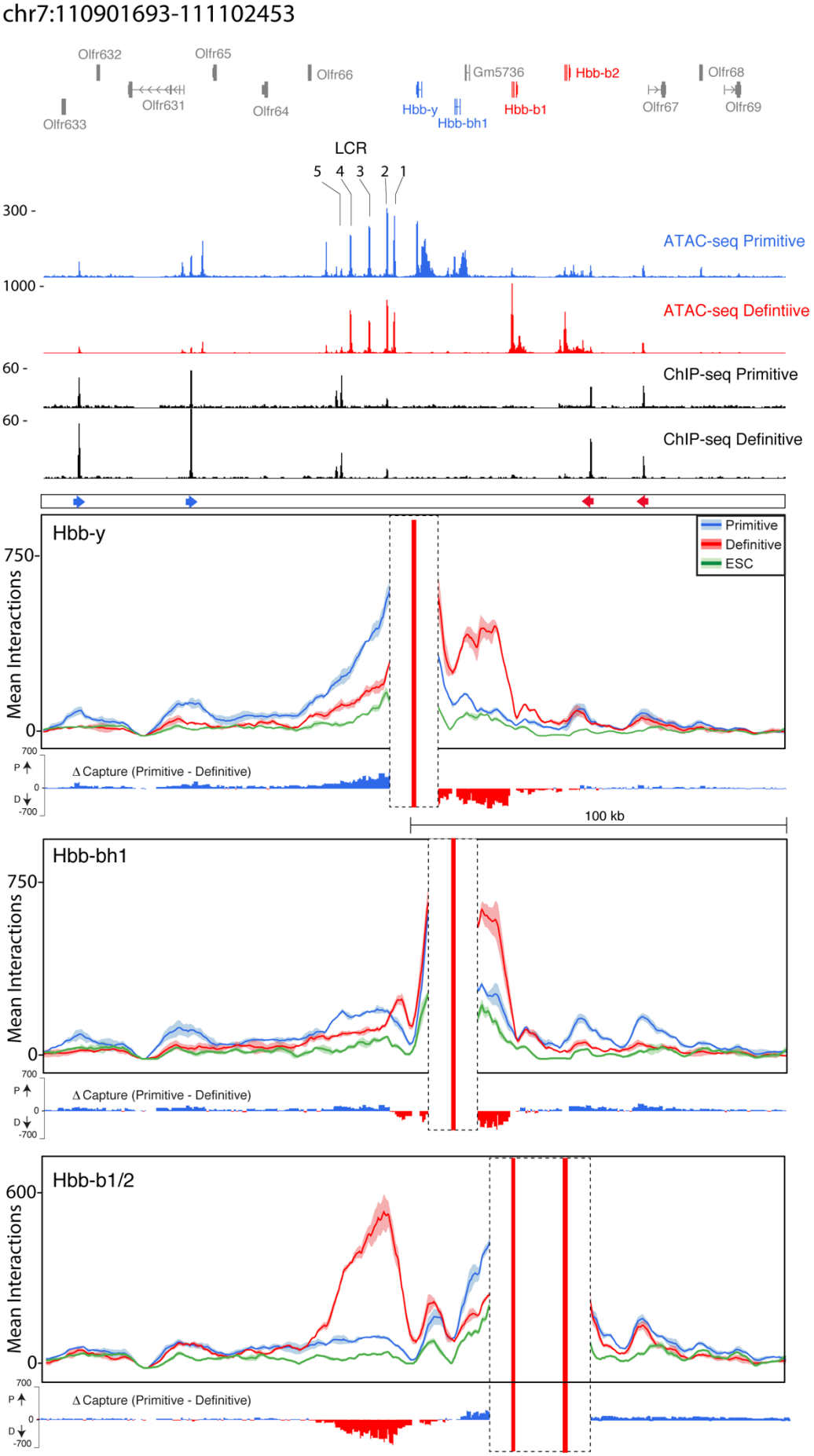

chr7:110901693-111102453
Olfr632
Olfr633
Olfr631
Olfr65
Olfr64
Olfr66
Gm5736
Hbb-y
Hbb-bh1
Hbb-b2
Hbb-b1
Olfr67
Olfr68
Olfr69
LCR
5 4 3 2 1
300 -
ATAC-seq Primitive
1000 -
ATAC-seq Defintiive
60 -
ChIP-seq Primitive
60 -
ChIP-seq Definitive
Hbb-y
Primitive
Definitive
ESC
750
Mean Interactions
0
700
P
0
D
-700
Δ Capture (Primitive - Definitive)
100 kb
Hbb-bh1
750
Mean Interactions
0
700
P
0
D
-700
Δ Capture (Primitive - Definitive)
Hbb-b1/2
600
Mean Interactions
0
700
P
0
D
-700
Δ Capture (Primitive - Definitive)

Figure 3.10 (previous page) Overlaid, normalized NG Capture-C data at the β-globin locus. The viewpoints are indicated by vertical red bars and the dashed lines indicate the exclusion zones (which lie too close to the capture viewpoint to judge meaningful interactions). The mean, plus and minus one standard deviation (s.d.), of sliding 5 kb windows are visualized. Differential tracks (ΔCapture-C) show a subtraction (Primitive (P) – Definitive (D)) of the mean number of meaningful interactions per restriction fragment. The experiments represent three biological replicates. Normalised CTCF ChIP-seq and ATAC-seq tracks for primitive and definitive erythroid cells are shown with the orientation of the CTCF sites below.

These data show that the chromatin re-organization at the β-globin locus during development differs from the changes at the α-globin locus. In primitive cells taken from E10.5, the LCR interacts with the *ϵy* (*Hbb-y*) gene, and in definitive erythroid cells, these interactions move to the adult *β1/β2* genes (*Hbb-b1/2*). Interactions between the *βH1* (*Hbb-bh1*) gene and the LCR are increased in primitive erythroid cells, but are less than those seen for *ϵy*; this may reflect the partial silencing of *βH1* which has already occurred by E10.5.

The most striking observation is that the embryonic globin genes (*ϵy* and *βH1*) in definitive erythroid cells interact together strongly and with the intergenic region between *Hbb-bh1* and the adult globin genes. This implies that they are excluded together in a small loop, away from the enhancers (although interactions with the enhancers do occur at a slightly increased frequency above the control mouse ES cells). Such a well-defined excluded region is not seen in the NG Capture-C data from the α-globin locus.

3.2.8 A novel hypersensitive element is seen within the α-globin locus at E11.5

The data presented so far are from unsorted primitive erythroid cells derived from E10.5 day embryos, when ζ-globin expression is maximal. This day was chosen as it gives sufficient cells numbers for NG-Capture C and ChIP-seq experiments. It does mean however, that the chromatin landscape as ζ-globin is activated and then silenced has not been studied. To address this, ATAC-seq was performed on E9.5, E10.5 and E11.5 cells. In order to ensure a pure population, primitive erythroid cells were sorted by flow cytometry (outlined in the Materials and Methods). ATAC-seq was also performed on sorted primary definitive erythroid cells from E12.5 fetal liver cells, when ζ-globin is silenced. These data are shown in Figure 3.7.

Overall, chromatin accessibility changes throughout the α-globin locus during development are fairly minor. The α-globin gene promoters show increased accessibility between E9.5 and E11.5, corresponding to increased expression through these time points. Similarly, the 3' peak associated with the ζ-globin gene becomes smaller between E9.5 and E11.5 as expression decreases. The most interesting observation is a novel hypersensitive site apparent on day E11.5 just 3´ of the ζ-globin gene (highlighted in blue). This sequence appears to be conserved in humans (though not across all vertebrates and mammals) and contains putative binding sites for important erythroid transcription factors including GATA1, KLF1 and BCL11A. It also lies in a region marked by the histone enhancer signature H3K4me1 in both primitive (E10.5) and definitive erythroid cells. It is consequently

plausible that this represents a stage specific repressive element which may transiently bind transcription factors necessary for ζ-globin silencing.

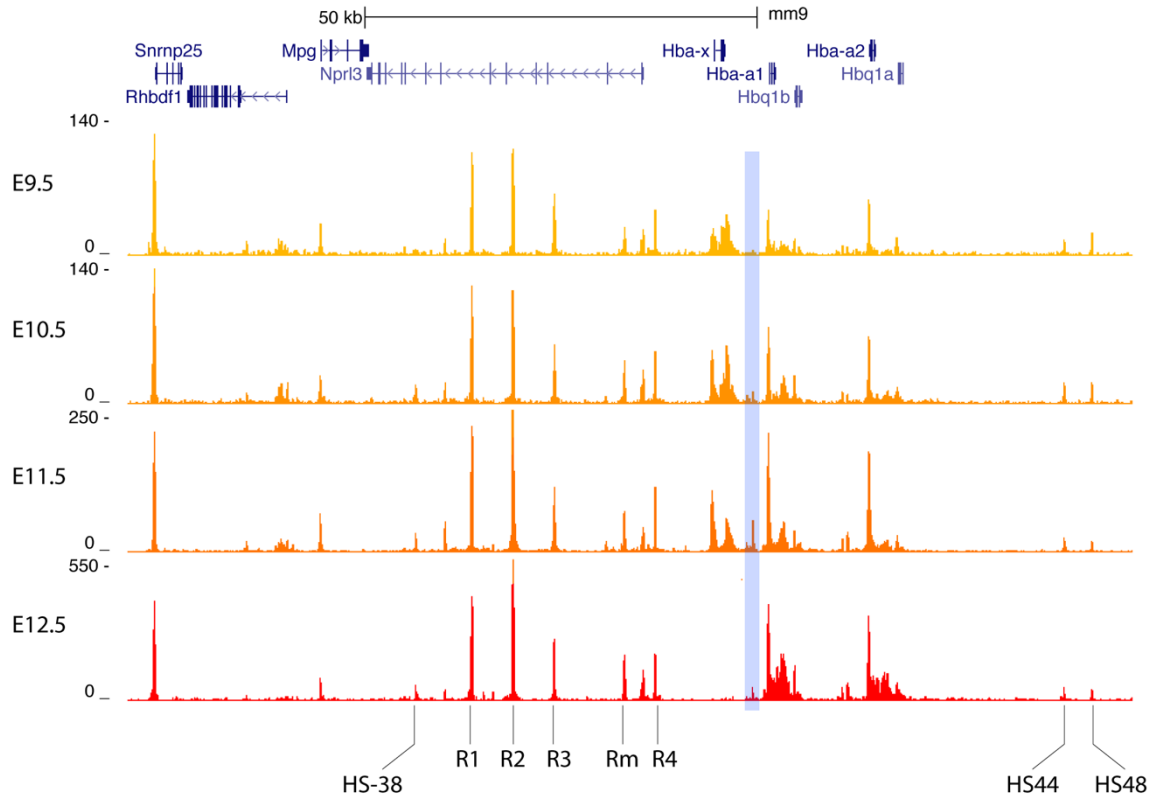


Figure 3.11 Chromatin accessibility at the α-globin locus during development. ATAC-seq tracks are shown for sorted primitive erythroid cells (E9.5-11.5) and definitive erythroid cells (E12.5). The boundary CTCF elements and α-globin enhancers are shown. A novel hypersensitive element is apparent on day E11.5. Data represent single normalized (RPKM) experiments.

3.3 Discussion

In this chapter, it is demonstrated that the α-globin locus in primitive erythroid cells is contained within the same 80kb self-interacting chromosomal domain as that found in definitive erythroid cells, with the pattern and insulation strength of CTCF mediated loops seemingly unchanged. The same CTCF sites form the boundary elements (with the H3-38/HS-39 CTCF sites forming the upstream boundary and the HS 44 and 48 sites forming the downstream boundary) and there is no embryonic specific CTCF insulator site responsible for reconfiguring the locus and mediating ζ-globin expression. Identical CTCF binding is also seen between primitive and definitive erythroid cells at the β-globin locus (shown in Figure 3.4) and genome wide, CTCF occupancy is similar in primitive and definitive erythroid cells, with 80% overlap in these two cell types.

Whilst examples of differential CTCF occupancy important for governing enhancer promoter interactions and driving tissue specific gene expression exist (Hark et al., 2000; Bell and Felsenfeld, 2000; Magbanua et al., 2015) such reports are rare. It would consequently be surprising were an embryonic specific CTCF site seen within the α-globin locus. CTCF occupancy genome wide does show some plasticity when multiple cell types are compared, with 30-60% of sites being variable (Kim et al., 2007; Wang et al., 2012). However, when CTCF occupancy is compared across just two cell types, there is generally reasonably high concordance (around 70%) (Wang et al., 2012), which is in keeping with the data from primitive/definitive erythroid cells presented here.

In terms of CTCF binding to the sites within the locus, there appears to be increased binding to the CTCF site located next to the θ globin genes in primitive erythroid cells compared to definitive erythroid cells. The significance of this is uncertain; the interaction profiles from the HS-38 viewpoint and the θ-globin sites remain similar, suggesting that there may not be a clear functional role. Fortuitously, *in vivo* mouse models where one of the paired CTCF sites adjacent to the θ-globin gene have been deleted are being generated in our laboratory. Careful analysis of these models in both primitive and definitive erythropoiesis will thus definitively determine whether these CTCF sites influence ζ-globin or α-globin expression.

A CTCF site specific to primitive erythroid cells was noted outside the self-interacting domain at the *Rhbdf1* promoter. This is also present in mouse ES cells, and correlates with *Rhbdf1* expression: when *Rhbdf1* is expressed, and the chromatin around the promoter is accessible, CTCF binds. CTCF binding to promoters *per se* isn't rare, with between 7 and 20% of CTCF binding sites lying within promoters genome-wide (Cuddapah et al., 2009; Kim et al., 2007; Wang et al., 2012); indeed CTCF was initially discovered via its binding to the promoter proximal regions of the *c-myc* gene (Filippova et al., 1996; Lobanenkov et al., 1990). In the data generated here, 4.7% of promoters have CTCF binding when chromatin states are assigned using classical histone signatures. Aside from its clearly defined role as an insulator protein, CTCF binding at promoters has been implicated in a variety of processes, including transcriptional repression (Lobanenkov et al., 1990), transcriptional activation (Vostrov and Quitschke, 1997) and RNA polymerase II pausing (Paredes et al., 2013).

When the HS-38 and H3-39 boundary elements are removed, this allows *Rhbdf1* to be activated by the normally constrained α-globin enhancers in definitive erythroid cells, although it still retains the H3K27me3 mark associated with PRC2 (Hanssen et al., 2017). CTCF doesn't appear to bind to the *Rhbdf1* promoter under such a scenario when reviewing the data (Hanssen et al., 2017) providing some evidence that CTCF binding isn't necessary for the gene to be activated. It may therefore be that once *Rhbdf1* recruits PRC2, the chromatin becomes inaccessible to transcription factors including CTCF. Although methylation may repress CTCF binding (Wang et al., 2012), the CpG island associated with the *Rhbdf1* promoter presumably isn't methylated in definitive erythroid cells given the presence of H3K27me3, which is normally antagonized by DNA methylation (Bartke et al., 2010).

In any case, what is clear from the NG Capture-C data generated here, is that the CTCF site located at the *Rhbdf1* promoter has no influence on the formation of the self- interacting domain in primitive erythroid cells. Interestingly, despite this, it still obeys the general observation whereby there are convergent CTCF sites focused on the α-globin enhancers, with its orientation determined as facing towards the enhancers.

Each of the enhancers present in definitive erythroid cells is also seen in primitive erythroid cells. When ranking by H3K27ac signal, the α-globin enhancer classifies as a super-enhancer; indeed, it is the most highly ranked super-enhancer out of the 327 identified. As outlined in the introduction, previous unpublished work from our laboratory indicates that removal of each enhancer element has no effect on ζ-

globin expression, and it is only when two enhancers are removed together (R1 and R2) , that ζ-globin expression falls to 40% (see Figure 1.11).

The characterisation of super-enhancers as a new paradigm in transcriptional regulation remains controversial (Pott and Lieb, 2015); previous work assessing the effect of individual and combinatorial enhancer removal at the α-globin locus implies an additive model, whereby there are no synergistic effects between enhancers and their effects on transcription are simply the sum of the parts (Hay et al., 2016). Mathematical modelling from this data set, and another independent enhancer cluster and gene, supports this (Dukler et al., 2016). It is therefore interesting to note the apparent deviation from an additive model in primitive erythropoiesis.

The concept of enhancer redundancy (or “shadow enhancers”), as observed in the enhancers’ effects on ζ-globin expression, is not novel. Reports exist in both vertebrates and invertebrates dating back twenty years (reviewed in Barolo, 2012), and a recent publication suggested that this is a widespread phenomenon in the mammalian genome (Osterwalder et al., 2018). This is the first report however, of a “super-enhancer” having differing effects on transcription on genes within the same cluster. As to why the globin genes evolved, such that there is enhancer redundancy for ζ-globin expression but not α-globin expression, is an open question.

Whilst the self-interacting domain formed in primitive erythroid cells is identical to that found in definitive erythroid cells, contacts within the domain between the globin gene promoters and the α-globin enhancers are significantly altered. In primitive

erythroid cells, there are clearly increased interactions between the ζ-globin promoter and the enhancer cluster compared to definitive; likewise, there are decreased interactions between the α-globin promoters and enhancers in primitive compared to definitive erythroid cells. This latter observation presumably occurs due to competition between the ζ-globin and α-globin promoters in primitive erythroid cells for the enhancers. Humans with both α-globin genes deleted from one chromosome express small amounts of ζ-globin (Chung et al., 1984), suggesting that competition from the α-globin promoters also plays a role in the complete silencing of ζ-globin in definitive erythroid cells.

It is also clear from the NG Capture-C data that despite ζ-globin being completely silenced in definitive erythroid cells, it still contacts the enhancer cluster, albeit at a lower frequency than seen in primitive erythropoiesis. This demonstrates that rather than enhancers and promoters forming discrete, stable interactions, there is a compartmentalised domain formed where all the enhancers and promoters can interact with each other. Within primitive erythroid cells, the preferred interaction is presumably between ζ-globin, α-globin and the enhancer cluster whereas within definitive erythroid cells, the preferred interaction is between α-globin and the enhancer cluster. Nevertheless, in definitive erythroid cells, contact between ζ-globin and the enhancers remains possible and does occur.

When primitive erythroid cells are compared to definitive, interactions between the Rm/R4 and the α-globin promoters are decreased more than those interactions between R1/R2 and the α-globin promoters. This perhaps suggests that the preferred interaction in primitive erythropoiesis is one whereby *Hba-x* loops in

between Rm and R4 and the α-globin gene promoters to contact R1 and R2. The preferred interaction in definitive erythropoiesis would be one where *Hba-x* loops out, thereby allowing Rm and R4 in particular to contact the α-globin promoters at a greater frequency (Figure 3.12).

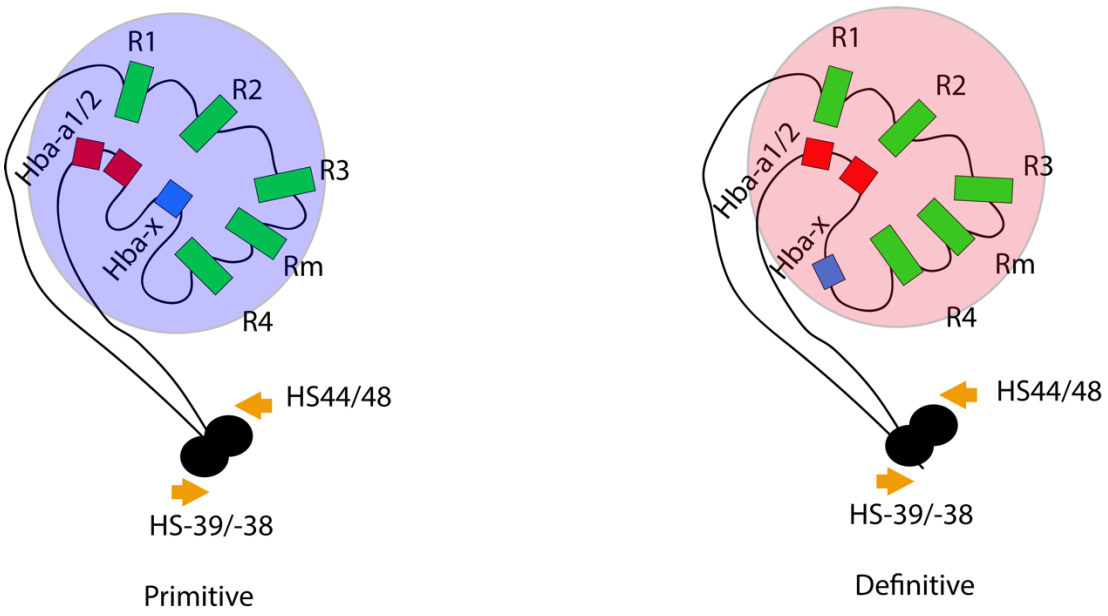


Figure 3.12 Proposed configuration of the α-globin locus in primitive and definitive erythropoiesis based upon the NG Capture-C data. In primitive erythropoiesis, the *Hba-x* gene loops in to interact with R1 and R2, and therefore reduces contact frequency between Rm, R4 and the α-globin promoters. In definitive erythropoiesis, it moves out to allow the α-globin promoters to fully interact with all the enhancers. Enhancers are shown in green, CTCF sites in black, *Hba-a1/2* in red and *Hba-x* in blue, with the self-interacting domain indicated by pale blue and pale red.

Analysis of the β-globin locus shows that, whilst the self-interacting domain remains similar between primitive and definitive erythropoiesis, reconfiguration of interactions within the domain is more dramatic than those seen at the α-globin

locus. In definitive erythropoiesis, the NG Capture C from the globin gene viewpoints show that the embryonic globin genes (*Hbb-bh1* and *Hbb-y*) form a small self-interacting loop, where they interact with each other at a much higher frequency than with either the definitive globin genes or the LCR. These observations generally agree with those reported previously using simpler 3C experiments (Palstra et al., 2003) although the data generated here are of a much higher resolution. Presumably it is the presence of two silenced genes in the β-globin locus (*Hbb-bh1* and *Hbb-y*), rather than just one in the α-globin locus (*Hba-x*), that allows such a discrete self-interacting loop containing silenced genes to form.

Given that the ζ-globin promoter remains in contact with its enhancer elements in definitive erythroid cells, its silencing is clearly due to more than simply removing the promoter away from the enhancer. The ChIP-seq data for histone acetylation (H3K27ac) show that in primitive erythropoiesis, there is a broad region of acetylation across the entirety of the 80kb self-interacting domain. In definitive erythropoiesis by contrast, there is localized deacetylation encompassing the ζ-globin gene and its coding regions. Broad hyperacetylated domains have been recognized for over a decade at various loci, including the globin loci (Bulger, 2005) and are not a universal phenomenon throughout the genome; rather they are associated with highly expressed tissue-specific genes, typically controlled by distal enhancers.

Histone acetylation of lysines on histone proteins' highly basic tail region via histone acetyltransferases leads to neutralization of their positive charge and increased chromatin accessibility (Verdone et al., 2006); thereby allowing activating *trans*-acting factors to bind. Removal of histone acetylation, mediated by histone

deacetylation (HDAC) proteins, causes increased chromatin compaction and repression of gene activity (Seto and Yoshida, 2014). Given the localized histone deacetylation over the ζ-globin gene associated with chromatin inaccessibility in definitive erythroid cells, it seems probable that this plays an important role in its silencing.

As HDACs are incapable of binding DNA directly, they associate with multiprotein complexes including the nucleosome and remodeling deacetylation complex (NuRD) and the Sin3A complex (Verdone et al., 2006). The two main transcription factors responsible for repressing γ-globin (the human β-like fetal globin), BCL11A and LRF, both associate with the NuRD complex and therefore recruit HDAC proteins where they bind (Xu et al., 2013; Masuda et al., 2016). Given this, and the localized histone deacetylation seen over the ζ-globin gene associated with silencing, the effects of BCL11A and LRF knockout on ζ-globin expression and the chromatin landscape at the α-globin locus were determined; this is investigated in the next chapter within this thesis.

Aside from the marked increase in chromatin accessibility over the ζ-globin promoter and 3' region, the chromatin landscape at the α-globin locus appears unchanged between E10.5 and definitive erythroid cells. As ζ-globin expression is maximal at this time point, it wasn't possible to exclude a transient hypersensitive site occurring with either ζ-globin activation (at E9.5) or repression (at E11.5 as ζ-globin is silenced) due to a "hit and run" transcription factor (Shah et al., 2019). ATAC-seq was therefore performed sequentially on E9.5 through E11.5 of embryogenesis, and also on E12.5 primary definitive erythroid cells. Interestingly, a hypersensitive site becomes apparent at E11.5, 5.2kb from the TSS of the ζ-

globin gene, which then isn't present in definitive erythropoiesis. Unfortunately, this lies within the exclusion zones for both the *Hba-x* and *Hba-a1/2* capture view-points, so it isn't possible to see whether this contacts either of the globin gene promoters. However, it lies in a region marked by the enhancer signature H3K4me1 in both primitive and definitive erythroid cells; it also contains binding sites for important erythroid transcription factors including GATA1 and KLF1 and is conserved in humans. Taken together, this all suggests that it may have a functional role.

There is a precedent for an embryonic specific enhancer element from the murine β-globin locus. A hypersensitive element termed HS-E1 was discovered in primitive, but not definitive cells, located 3´ of the embryonic *Hbb-y* gene (Fromm et al., 2011). It is notable that the position of this is similar to the hypersensitive element identified at E11.5 at the α-globin locus, lying downstream of the embryonic gene. HS-E1 functions as an enhancer in luciferase reporter assays and its removal lead to a specific decrease in the embryonic, but not adult, β-like globins (Fromm et al., 2011). In my independent genome-wide data, I can also see this element at the β-globin locus, although it is present throughout E9.5-E11.5. Clearly further work is required to characterize the putative enhancer/repressor element in the α-globin locus, including enhancer assays and deletion via genome editing.

In conclusion, the data presented in this chapter demonstrate that the 80kb self-interacting domain containing the α-globin enhancer cluster and the globin genes is developmentally conserved between primitive and definitive erythroid cells. The α-globin enhancer cluster classifies as a super-enhancer in primitive erythroid cells, with the same five enhancers being active. Within the domain, interactions between the enhancer cluster and the ζ-globin gene decrease between primitive and

definitive erythropoiesis; however, even when the ζ-globin gene is silenced, it still contacts the enhancers. Silencing of the ζ-globin gene is associated with histone deacetylation and inaccessible chromatin. Separately, a novel hypersensitive site has been identified as ζ-globin is silenced which requires further investigation.

4 The effect of Lrf and Bcl11a on the endogenous globin loci in mouse

4.1 Introduction

Two transcription factors, LRF and BCL11A, are together responsible for repression of γ-globin at the β-globin locus in humans in adult life, with knockout of both transcription factors leading to almost 100% γ-globin expression in the HUDEP-2 cell line (Masuda et al., 2016). Both LRF and BCL11A associate with the NuRD complex (Xu et al., 2013; Masuda et al., 2016) and appear to interact independently at this locus.

Transcriptome data from a mouse model where *Bcl11a* had been knocked out in the erythroid lineage demonstrated up-regulation of *Hba-x.* Separately, knockout of *Lrf* in a mouse model demonstrated up-regulation of *Hba-x* in Lrf deficient erythroblasts in RNA-seq data (Masuda et al., 2016). Finally, the data presented in the previous chapter suggest that histone deacetylation may be an important mechanism in silencing; this could potentially be mediated by both these transcription factors via the NuRD complex.

Given these observations, the effect of knockout of *Lrf* and *Bcl11a* individually on *Hba-x* expression and the chromatin landscape was investigated *in vivo* in the mouse.

Two erythroid populations were studied where *Hba-x* is normally silenced: definitive erythroid cells from fetal liver taken at E12.5 and either *in vitro* cultured E12.5 fetal erythroid cells (which were cultured for 8 days) or definitive erythroid cells from adult mice. *Bcl11a* knockout leads to embryonic globin gene derepression at the β-globin locus in both adulthood and fetal liver erythropoiesis whether that is the human γ-globin gene (which behaves as an embryonic gene in the mouse) or the normal murine embryonic globin genes. However, the changes are more marked in definitive cells from the early fetal liver (Sankaran et al., 2009; Smith et al., 2016). Assuming BCL11A and LRF silence *Hba-x* using the same mechanism as the embryonic β-globin genes, we hypothesised that there would be a similar difference in *Hba-x* expression between definitive fetal erythroid cells and late fetal or adult cells when Lrf and Bcl11a are removed.

4.2 Mouse models: choice and validation

4.2.1 Lrf knockout

Knockout of *Lrf* leads to embryonic lethality due to severe anaemia at E16.5 (Maeda et al., 2007). To circumvent this, a conditional mutant mouse line was previously created with conventional genome engineering strategies, where loxP sites flank Exon 2 (Figure 1.3) (Maeda et al., 2007). Removal of Exon 2 has been proven to lead to absent *Lrf* expression via assessment of both protein in multiple publications (Lunardi et al., 2013; Maeda et al., 2007; Masuda et al., 2016). We therefore imported the *Lrf*$^{Flox/Flox}$ mouse as a gift from Pier Pandolfi.

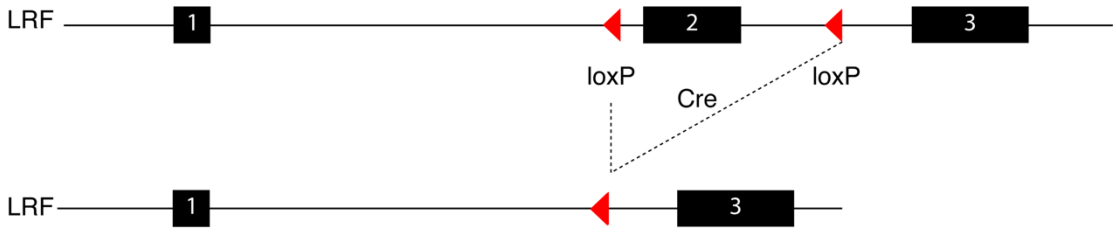


Figure 4.1 LRF floxed mouse model. Cre expression leads to removal of Exon 2 and LRF knockout.

We crossed the *Lrf*$^{Flox/Flox}$ with a *Gata1*-Cre mouse line(Mao et al., 1999), which has germline expression and thus causes germline excision of floxed regions. *Lrf*$^{Flox/-}$ were then bred to produce litters with a combination of *Lrf*$^{Flox/Flox}$ (WT), *Lrf*$^{Flox/-}$(+/-) and *Lrf* $^{-/-}$ (-/-) for E12.5 day experiments. -/- mice weren't viable beyond E16.5 as previously published (Maeda et al., 2007).

4.2.2 *Bcl11a* enhancer knockout

Bcl11a is essential for development in the mouse with knockout of *Bcl11a* leading to death three to four hours after birth from uncharacterised causes (Liu et al., 2003). A *Bcl11a*$^{flox/flox}$ mouse has previously been published with loxP sites flanking exon 4 (Yu et al., 2012); in order to achieve erythroid specific knockout, this would require crossing with an ErGFPcre mouse (Heinrich et al., 2004) however, which isn't available in our laboratory. To obviate the need for breeding conditional mice with Cre expressing lines we chose to study a mouse model where the erythroid-specific enhancer for *Bcl11a* has been removed (Smith et al., 2016). This has a 10kb deletion, encompassing two hypersensitive sites (denoted +58 and +62), and has been shown to lead to erythroid specific down-regulation of *Bcl11a* expression (Smith et al., 2016). The mouse is viable, with normal development, and no clear phenotype in comparison to wild-type, aside from up-regulation of the β-like embryonic globins. Importantly, *Bcl11a* expression is preserved in brain tissue and B cells (Smith et al., 2016).

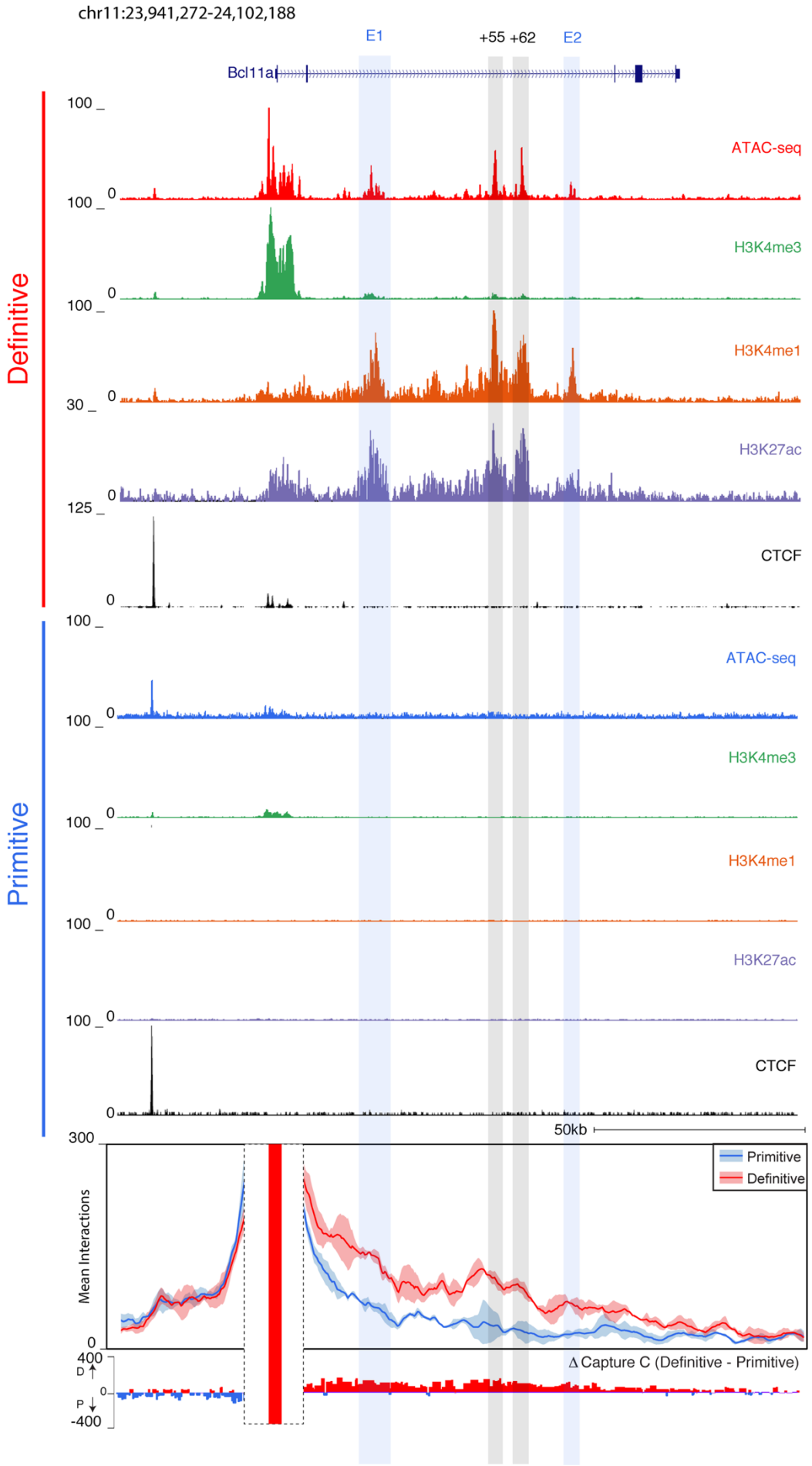
Previous characterization of the putative erythroid enhancers for *Bcl11a* was performed in MEL cells (Bauer et al., 2013). We thought it would be of interest to confirm these data *in vivo* from primary murine erythroid cells. NG Capture-C was therefore performed from the viewpoint of the *Bcl11a* promoter in primitive and definitive primary erythroid cells and aligned with the ChIP-seq data from definitive and primitive erythroid cells (Figure 4.2).

The two enhancers previously identified as important for *Bcl11a* expression (orthologous to the human +55 and +62 *Bcl11a* enhancers) (Bauer et al., 2013) are highlighted in grey. It should be noted that the distance from the transcriptional start site (TSS) and these hypersensitive elements does not correspond to 55kb or 62kb in the mouse; however, they have been denoted as this in previous publications (Bauer et al., 2013; Canver et al., 2015; Smith et al., 2016), given their sequence conservation with the human enhancers, located at that position. These sites are all accessible by ATAC-seq, and are decorated with both H3K4me1 and H3K27ac, whilst they are negative for H3K4me3. There are also two additional putative enhancer elements highlighted in light blue, and marked as E1 and E2. The importance of these *in vitro* or *in vivo* has not been investigated.

In primitive erythroid cells, there is minimal hypersensitivity over the *Bcl11a* promoter and there are no hypersensitive elements located within the intron; ChIP-seq for H3K4me1 and H3K27ac confirms that the enhancers aren't active.

The interaction profile from the viewpoint of the *Bcl11a* promoter shows increased interactions throughout the second intron corresponding to the enhancer cluster in definitive erythroid cells when compared to primitive. These data corroborate previous publications (Canver et al., 2015; Smith et al., 2016), and demonstrate that the enhancers are involved in driving *Bcl11a* expression in definitive erythroid cells.

Figure 4.2 (next page) Chromatin accessibility, histone marks and interaction profile at the *Bcl11a* gene locus in murine primitive and definitive erythroid cells. The putative enhancers are shown. Overlaid, normalized NG Capture-C data is also shown from the viewpoint of the *Bcl11a* promoter (indicated by a red line). The mean, plus and minus one standard deviation (s.d.), of sliding 5 kb windows are visualized. Differential tracks (ΔCapture-C) show a subtraction (Primitive (P) – Definitive (D)) of the mean number of meaningful interactions per restriction fragment. Primitive data are ≥3 biological replicates from E10.5 erythroid cells; definitive tracks are previously published (Kowalczyk et al., 2012b).


chr11:23,941,272-24,102,188
E1
+55 +62
E2
Bcl11a
Definitive
ATAC-seq
H3K4me3
H3K4me1
H3K27ac
CTCF
Primitive
ATAC-seq
H3K4me3
H3K4me1
H3K27ac
CTCF
50kb
Primitive
Definitive
Mean Interactions
Δ Capture C (Definitive - Primitive)
D
P

Although the orthologous +62 enhancer *in vitro* was demonstrated as the most important in cell lines (Canver et al., 2015), there also appears to be a contribution to *Bcl11a* expression from the orthologous +55 enhancer *in vivo* (Smith et al., 2016). We therefore imported the mouse with the composite enhancer block (corresponding to 10kb) deleted (*Bcl11a*$^{\Delta 10kb}$) (Smith et al., 2016) as a gift from Stuart Orkin.

This mouse was bred to homozygosity and the transgenic β-globin locus was bred out (the mice had previously been bred with a mouse containing a transgenic β-YAC human β-globin gene locus (Porcu et al., 1997)) to avoid confounding human globins. The PCR based genotyping strategy is outlined in the Material and Methods. *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ mice are viable and appear phenotypically normal in terms of development, behaviour and health. Haematological analysis however, demonstrated that homozygous mice had a probable mild haemolytic anaemia, with an increased reticulocyte count and MCV (Table 4.1 and Figure 4.3). Spleen size was not significantly different between the various genotypes.

	Hb	WCC	MCV	Ret %	Spleen
***Bcl11a*$^{WT/WT}$**	15.5±0.35	8.1±0.67	50+0.45	2.02±0.28	0.26±0.07
***Bcl11a*$^{\Delta 10kb/WT}$**	14.5±0.66	9.42±2.3	51±1.52	2.68±1.24	0.24±0.04
***Bcl11a*$^{\Delta 10kb/\Delta 10kb}$**	13.8±0.71**	10.6±3.5	54.0±1.38***	3.08±0.36***	0.32±0.13

Table 4.1 Hematologic parameters for adult mice for WT, heterozygous and homozygous *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ models. Spleen size is expressed as a percentage of body weight. Haemoglobin (Hb g/dL), White Cell Count (WCC; x10^9/L), Mean Cell Volume (MCV; fL) and Reticulocyte percentage (Ret %) are shown. ** denotes $p<0.01$; *** denotes $p<0.001$ by one way ANOVA.

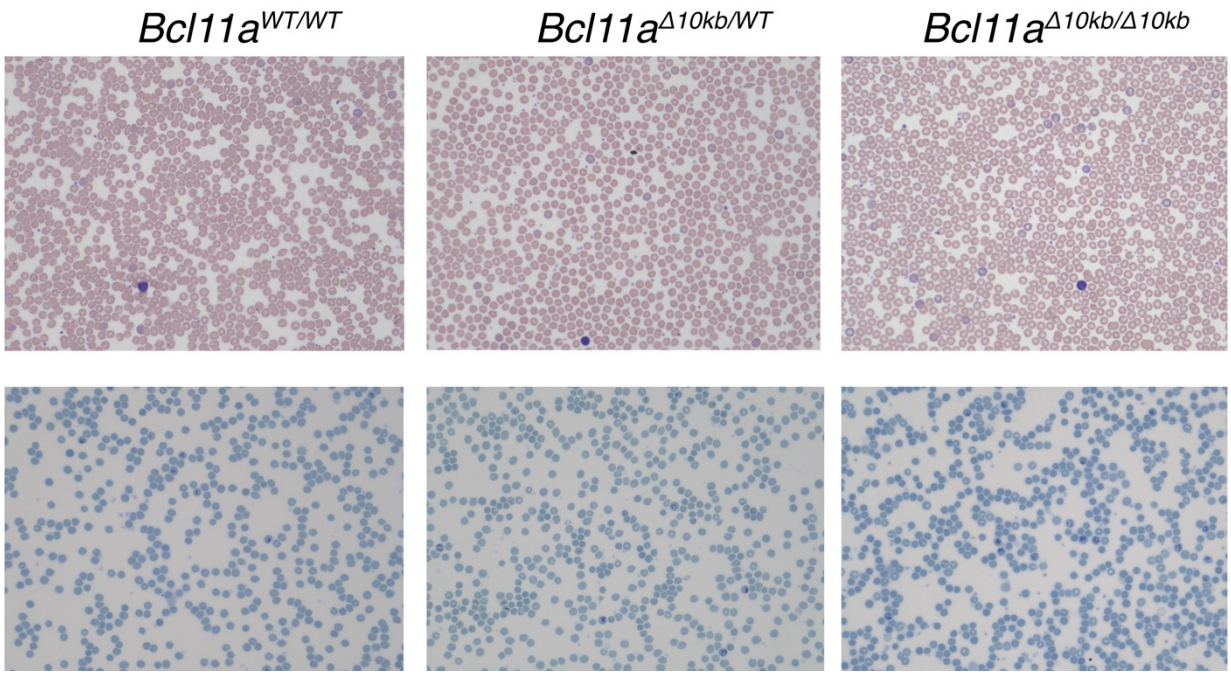


Figure 4.3 Representative blood smears (top panels) and reticulocyte staining for WT, heterozygous and homozygous mice for the *Bcl11a*$^{\Delta 10kb}$ deletion.

A subtle increase in reticulocytes is seen in the homozygous mice.

Following establishment of homozygous *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ mice, definitive erythroid cells from adult spleens were harvested and expression analysis performed for *Bcl11a* using RT-qPCR alongside ATAC-seq to assess the chromatin accessibility at the *Bcl11a* locus (Figure 4.4).

Knockout of the +55 and +62 orthologous mouse enhancers reduces *Bcl11a* expression to that found in primitive erythroid cells taken from E10.5, where it is just at the limit of detection of the assay.

Removal of the +55 and +62 enhancer leads to changes in chromatin accessibility throughout the *Bcl11a* locus, with loss of the E1 and E2 putative enhancers, and a reduction in chromatin accessibility over the *Bcl11a* promoter (Figure 4.4a). *Bcl11a*

appears to have multiple enhancer elements, although the enhancer cluster doesn't meet the arbitrary definition of a super-enhancer in definitive erythroid cells when ranking by Med1 or H3K27ac using our previously published data-set (Hay et al., 2016). Nevertheless, it is interesting to note that removal of two of the four putative enhancers also causes loss of the other two hypersensitive elements.

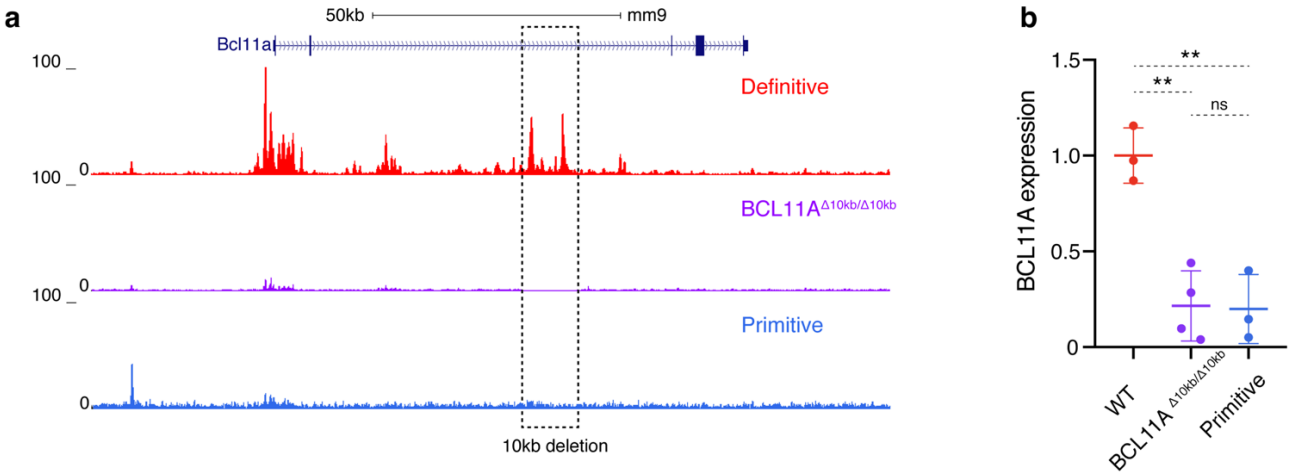


Figure 4.4 a. Chromatin accessibility and b. *Bcl11a* expression in WT definitive, primitive (E10.5) and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ definitive erythroid cells. *Bcl11a* expression was performed in bulk populations on steady-state RNA, normalized to 18S rRNA. Experiments were performed in ≥3 biological replicates. Statistical analysis is by one-way ANOVA with a Bonferroni correction for multiple comparisons; the individual data points and SEM are shown. ** denotes $p<0.01$. Definitive erythroid cells used for experiments were Ter119 selected erythroid cells from adult spleens.

4.3 Results

4.3.1 *Lrf* knockout leads to increased apoptosis and a differentiation block

Following labelling with fluorescently labelled antibodies directed against the cell surface markers CD71 and Ter119, fetal liver cells taken from E12.5-E13.5 form a characteristic distribution which allows separation into a series of increasingly mature erythroid cells via gates denoted S0-S5 (Shearstone et al., 2011) (Figure 4.5); see Materials and Methods for detailed sorting strategy.

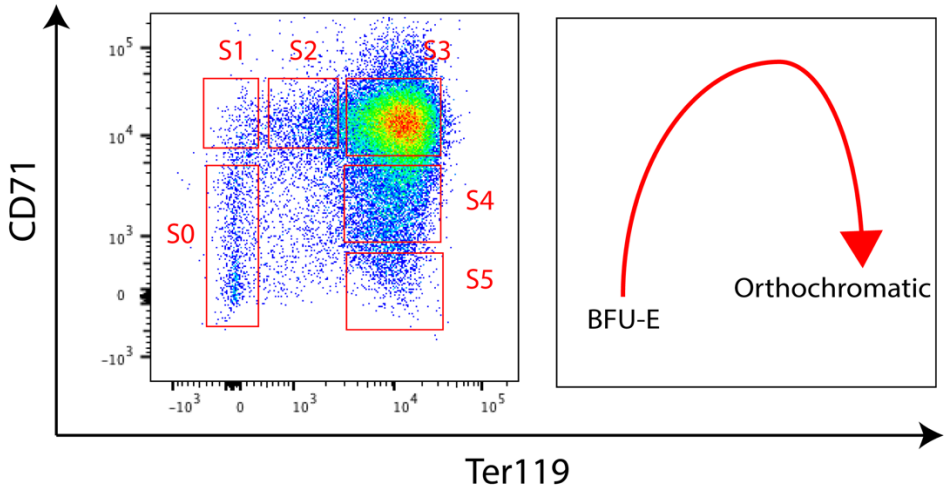


Figure 4.5 Erythroid differentiation in freshly isolated WT fetal liver (E12.5), defined by the CD71/Ter119 flow cytometric profile. Cells can be separated into six sub-sets, corresponding to increasing erythroid differentiation.

Knockout of *Lrf* has previously been reported to cause apoptosis in late erythroid cells and a differentiation block, occurring around the S3 to S4 transition (Maeda et al., 2009). Staining for the apoptotic marker Annexin-V was therefore performed (Wlodkowic et al., 2009) in addition to the usual panel of markers. There is a clear differentiation block in both the $Lrf^{Flox/-}$ and $Lrf^{-/-}$ fetal livers (Figure 4.6a); Annexin V

staining is also increased. Annexin V appears to be increased throughout all stages of differentiation rather than being confined to a particular subset (Figure 4.6b).

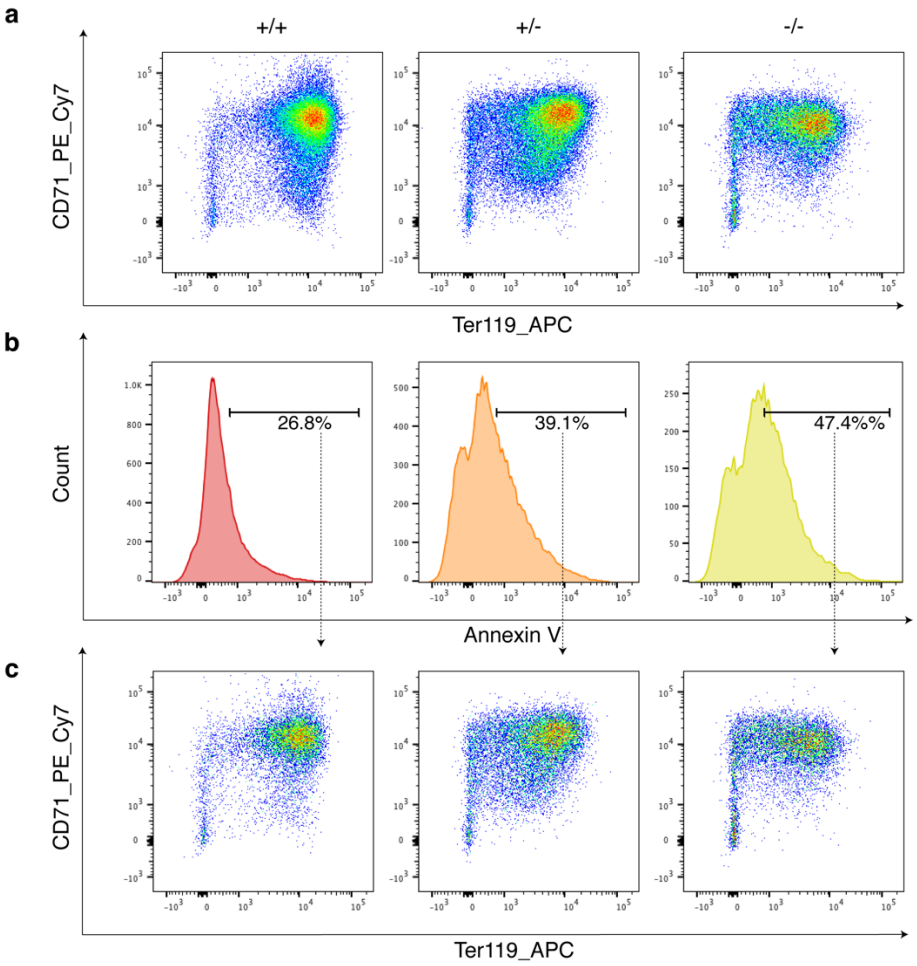


Figure 4.6 a. Representative plots of erythroid differentiation in *Lrf*$^{Flox/Flox}$(+/+), *Lrf*$^{Flox/-}$ (+/-) and *Lrf*$^{-/-}$(-/-) from ~E12.5 fetal livers. There is a differentiation block in the *Lrf*$^{Flox/-}$ and the *Lrf*$^{-/-}$ erythroid cells. **b. Annexin-V staining in *Lrf*$^{Flox/Flox}$, *Lrf*$^{Flox/-}$ and *Lrf*$^{-/-}$.** Annexin-V staining is increased in both the heterozygous and homozygous knockout. The FMO sample without Annexin V was used to define the positive gate **c. CD71 and Ter119 flow cytometric profile of the Annexin-V positive cells.** No clear sub-set of cells appears to be affected.

Given this, Annexin V staining was used to gate out positive cells, to ensure that increased apoptosis in the *Lrf* knockout heterozygous and homozygous fetal livers did not affect the results.

4.3.2 Knockout of *Lrf* leads to *Hba-x* and *Hbb-bh1* globin gene derepression and profound changes in chromatin accessibility in early fetal life

Single cell analysis was performed on cells sorted from S0-S3 using the Biomark system panel on 54 genes, outlined in Table 4.2 below. This panel had previously been developed by Rob Beagrie, a post-doctoral scientist from the WIMM using published single cell data (Tusi et al., 2018). Aside from the globin genes, these were selected to confirm or refute lineage contamination, chosen because they lay next to the α-globin gene cluster or chosen because they change during erythroid differentiation.

PCA analysis was then performed (Figure 4.7). When the globin genes are excluded from the analysis, S0, S1, and S2 and S3 separate (Figure 4.7a). However, there is no separate clustering of the WT, heterozygote and homozygote *Lrf* knockout cells, implying that cells are matched for differentiation. This is also true when just considering the S3 populations (Figure 4.7b).

Globin	Lineage	Differentiation	α-globin cluster
Hba-a1/a2 Hba-x Hbb-b2 Hbb-bh1 Hbb-bh2 Hbb-bs Hbb-y Hbq1a Hbq1b Hbb-bt/bs/b1/b2	Ccr2(Monocyte) Ccr9(Lymphocyte) Flt3(Lymphocyte) Mpo(Granulocyte) Ms4a6c(Monocyte) Pf4(Megakaryocyte) Plek(Megakaryocyte) Prtn3 (Granulocyte) Vwf (Megakaryocyte) Ifitm1(Basophil) Lmo4(Basophil) Nrep(Basophil)	Abcg4 Alas2 αpoe Arpc2 Atp1b2 cd59b ddx21 Dek Fth1 Gata2 Gpx1 H2afy Hnrnpa0 Hspe1 Klf1 Mpl Myc Ncl Nono Sox4 Tagln2 tmem14c Trib2 Vamp5 Slc43a3 Rhd Podxl Pebp1	Mpg Nprl3 Rhbdf1 Sh3pxd2b Snrnp25

Table 4.2 Genes used in the single cell analysis for the *Lrf*$^{-/-}$ erythroblasts. Genes changing during differentiation were mined from Tusi et al. 2018.

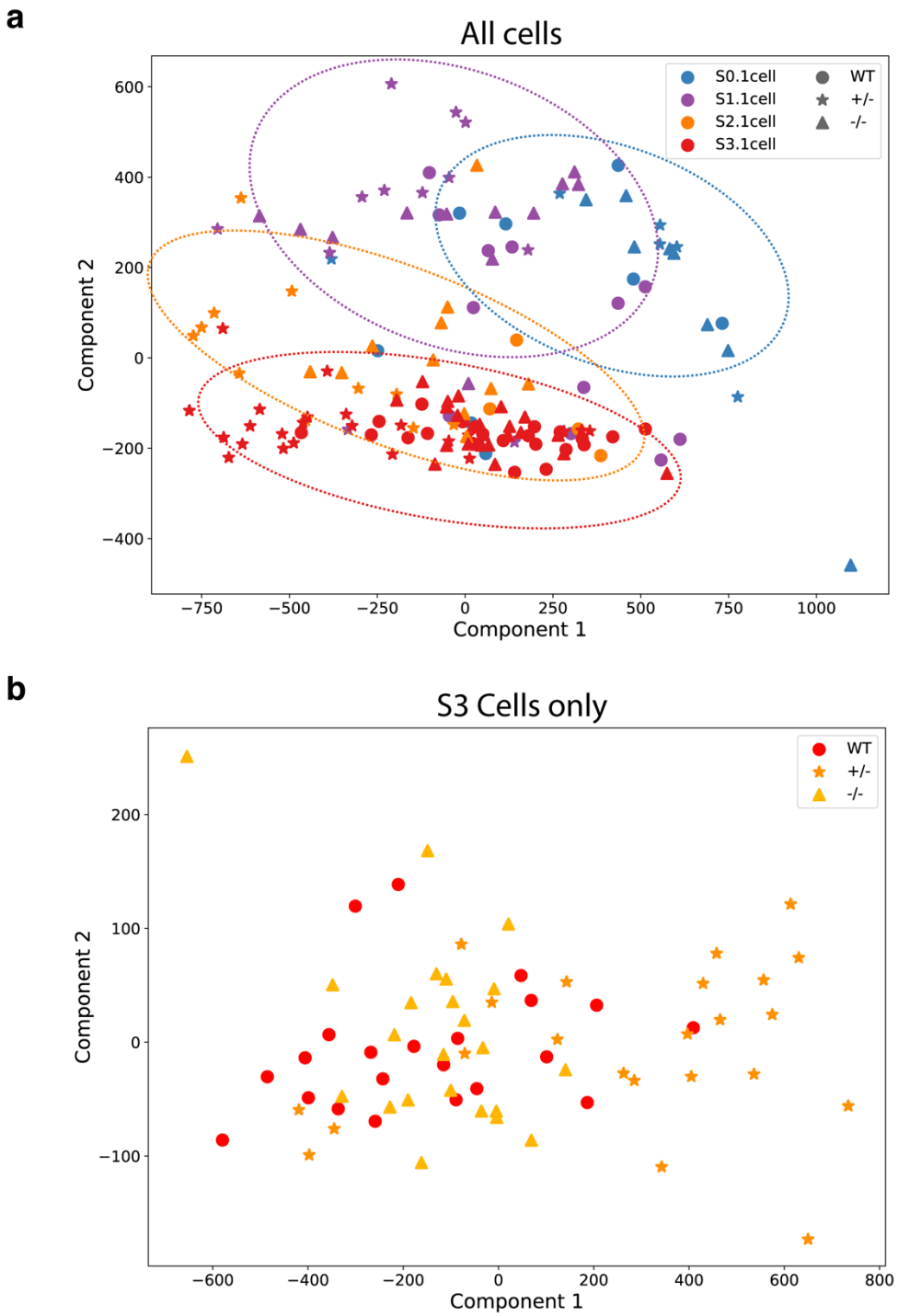


Figure 4.7 PCA Analysis for single erythroid cells from WT, heterozygote (+/-) and homozygote(-/-) *Lrf* knockout cells E12.5 fetal livers. a. PCA analysis excluding the globin genes. Four overlapping clusters can be identified (circled) corresponding to the different S0-S3 cells. No separate clustering based upon genotype occurs**. b. PCA excluding the globin genes for the S3 population.**

Alongside the single cell expression data, ATAC-seq was also performed on 5000 cells from the S3 populations. The data for both the globin single cell expression and ATAC-seq from the S3 populations are shown in Figures 4.8 and 4.9 for the α-globin and β-globin loci respectively.

When *Lrf* is absent, *Hba-x* expression is markedly de-repressed in defintiive erythroid cells taken from E12.5 fetal liver, and chromatin accessibility at the *Hba-x* gene resembles that seen in primitive erythropoiesis. θ-globin gene expression is unaffected, and *Hba-a1/2* expression is modestly decreased. The enhancers' accessibility appears to be unchanged, with the exception of R4, which is perhaps slightly increased in the heterozygote and homozygote knockout erythroid cells.

At the β-globin locus, knockout of *Lrf* has a specific effect on *Hbb-bh1* with *Hbb-y* remaining repressed, as previously reported (Masuda et al., 2016). This is mirrored by chromatin accessibility changes over the *Hbb-bh1* gene, with the *Hbb-y* promoter remaining inaccessible by ATAC-seq. As at the α-globin locus, there is a modest reduction in the adult globins.

In summary, these single cell data demonstrate that LRF is important for repression of both *Hba-x* and *Hbb-bh1* but not *Hbb-y* in fetal life. Knockout leads to marked de-repression of these globins in fetal life. The ATAC-seq data imply that LRF mediates silencing via chromatin remodelling.

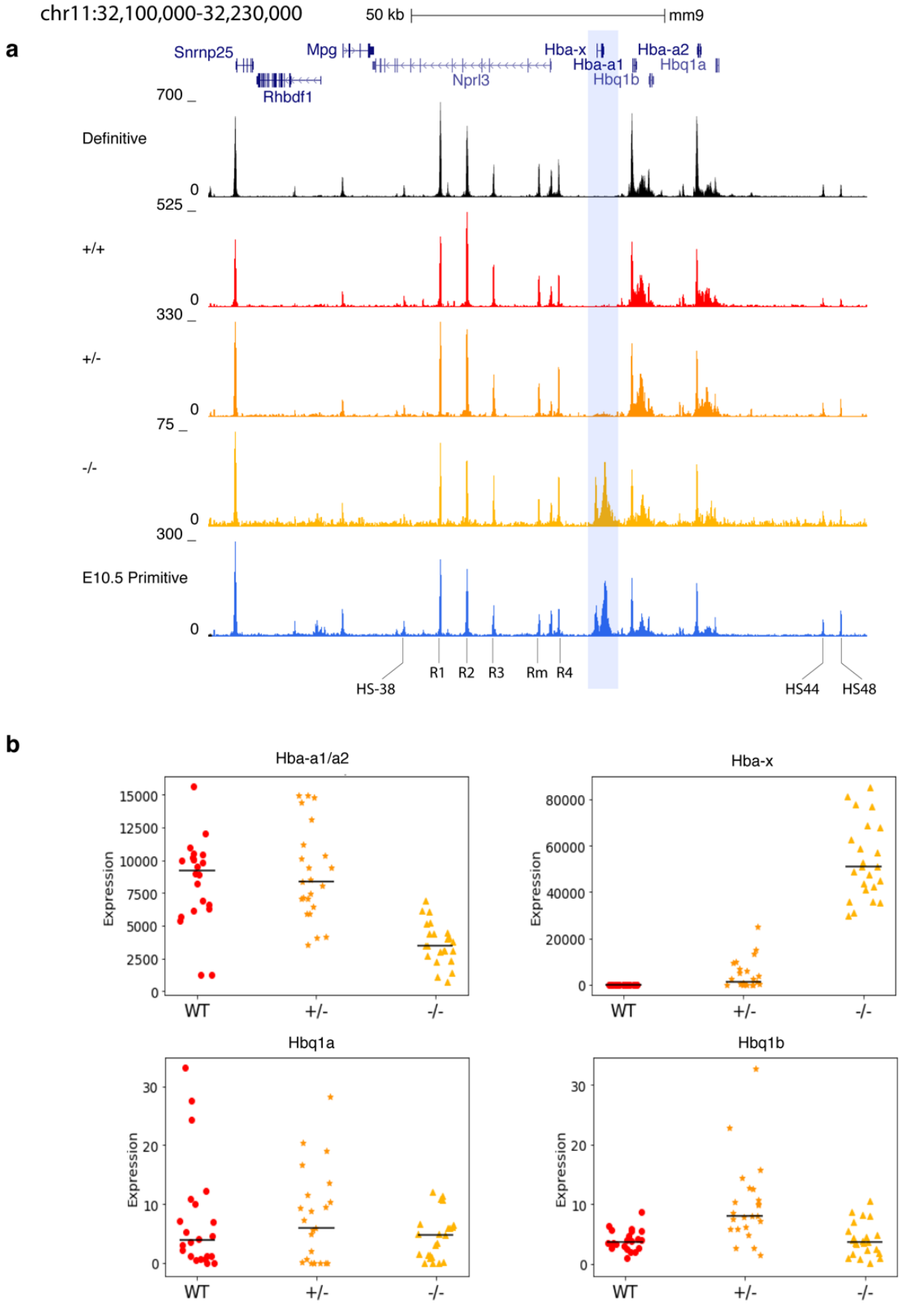

a
chr11:32,100,000-32,230,000
50 kb
mm9
Snrnp25
Mpg
Hba-x
Hba-a2
Hba-a1
Hbq1a
Nprl3
Hbq1b
Rhbdf1
Definitive
+/+
+/-
-/-
E10.5 Primitive
HS-38
R1
R2
R3
Rm
R4
HS44
HS48
b
Hba-a1/a2
Hba-x
Hbq1a
Hbq1b
Expression
WT
+/-
-/-

Figure 4.8 (previous page) a. Chromatin accessibility and b. single cell expression at the α-globin locus in *Lrf*$^{Flox/Flox}$ (+/+), *Lrf*$^{Flox/-}$ (+/-) and *Lrf*$^{-/-}$ (-/-) erythroid cells from the S3 population. Chromatin accessibility in definitive and primitive erythroid cells are also shown. Tracks shown represent single replicates from littermates; *Hba-x* is highlighted in blue. Regarding the expression data, each dot represents a single cell. Lines indicate median expression. The enhancers and boundary CTCF elements are shown.

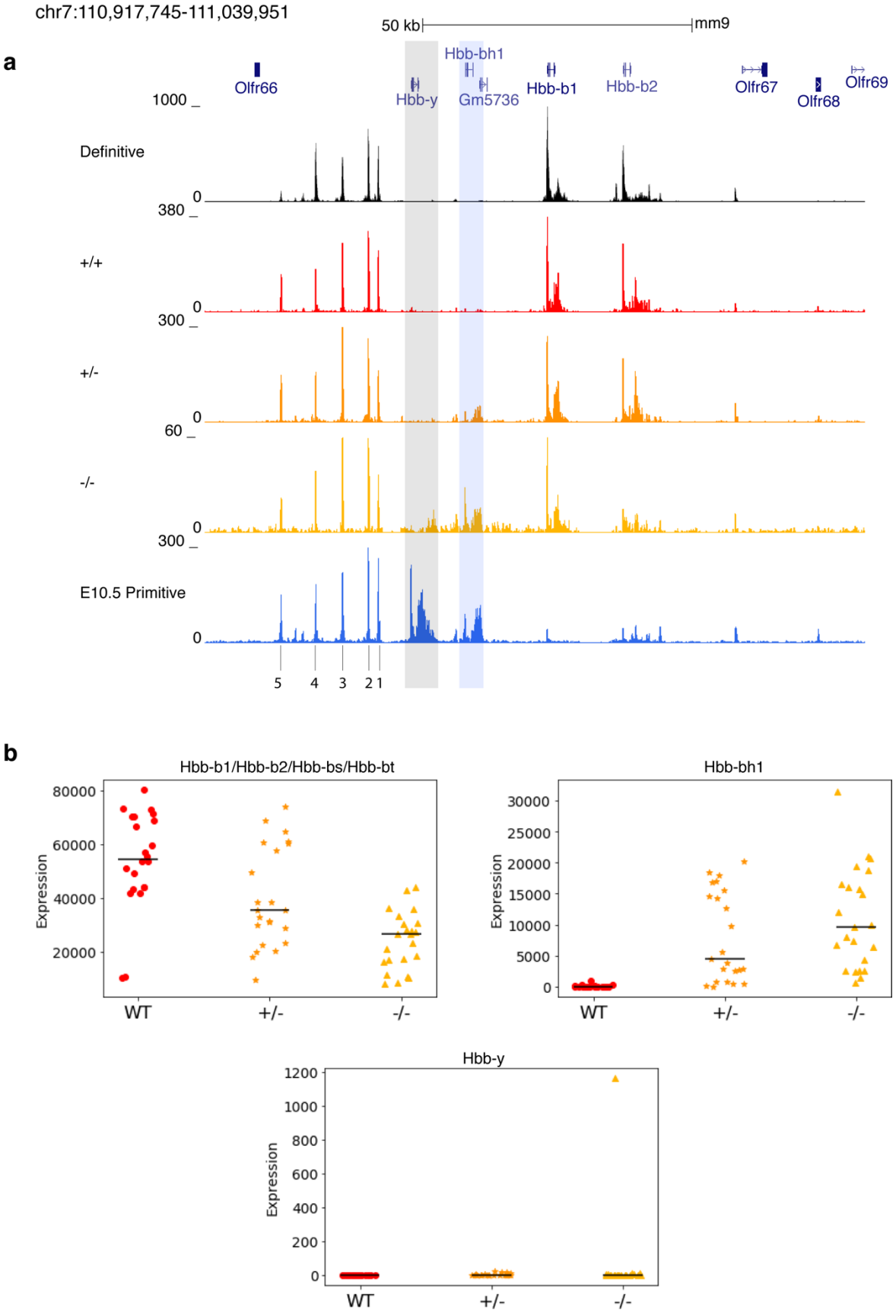

a
chr7:110,917,745-111,039,951
50 kb
mm9
Olfr66
Hbb-y
Hbb-bh1
Gm5736
Hbb-b1
Hbb-b2
Olfr67
Olfr68
Olfr69
Definitive
+/+
+/-
-/-
E10.5 Primitive
1000
380
300
60
0
5
4
3
2
1
b
Hbb-b1/Hbb-b2/Hbb-bs/Hbb-bt
Hbb-bh1
Hbb-y
Expression
WT
+/-
-/-
80000
60000
40000
20000
30000
25000
15000
10000
5000
1200
1000
800
600
400
200

Figure 4.9 (previous page) a. Chromatin accessibility and b. single cell expression at the β-globin locus in *Lrf*$^{Flox/Flox}$ (+/+), *Lrf*$^{Flox/-}$ (+/-) and *Lrf*$^{-/-}$ (-/-) primary murine erythroid cells from the S3 population. Chromatin accessibility in definitive (from the *in vitro* fetal liver culture system) and primitive (E10.5) erythroid cells are also shown; tracks shown represent single replicates from littermates; *Hbb-bh1* is highlighted in blue; *Hbb-y* is highlighted in grey. Regarding the expression data, each dot represents a single cell. A probe targeting all the adult globin genes was used. Contamination with one primitive erythroid cell is seen in the *Hbb-y* expression data.

4.3.3 Fetal definitive *Lrf*$^{-/-}$ erythroid cells undergo further embryonic globin gene silencing when cultured *in vitro* for 8 days

It was surprising that knockout of *Lrf* alone in definitive erythroid cells from E12.5 fetal liver led to chromatin accessibility changes to a level similar to that seen in primitive erythropoiesis. We wondered if this would still be the case in cells taken at a later timepoint. Knockout of *Lrf* has previously been reported to lead to embryonic lethality at E16.5 (Maeda et al., 2007) and we were also unable to generate homozygous viable mice. A well-established erythroid differentiation system from E12.5 fetal livers was therefore used, which involves six days in expansion erythroid media followed by 30 hours differentiation (more details are provided in Materials and Methods) to simulate late stage fetal erythroid cells.

Following six days of *in vitro* expansion and 30 hours of erythroid differentiation, flow cytometry was performed. As expected *Lrf*$^{-/-}$ knockout cells failed to differentiate properly in comparison to *Lrf*$^{Flox/Flox}$ littermates, as demonstrated by a

paler cell pellet, larger size, morphological appearance and their cell surface markers, they also had decreased cell viability (Figure 4.10).

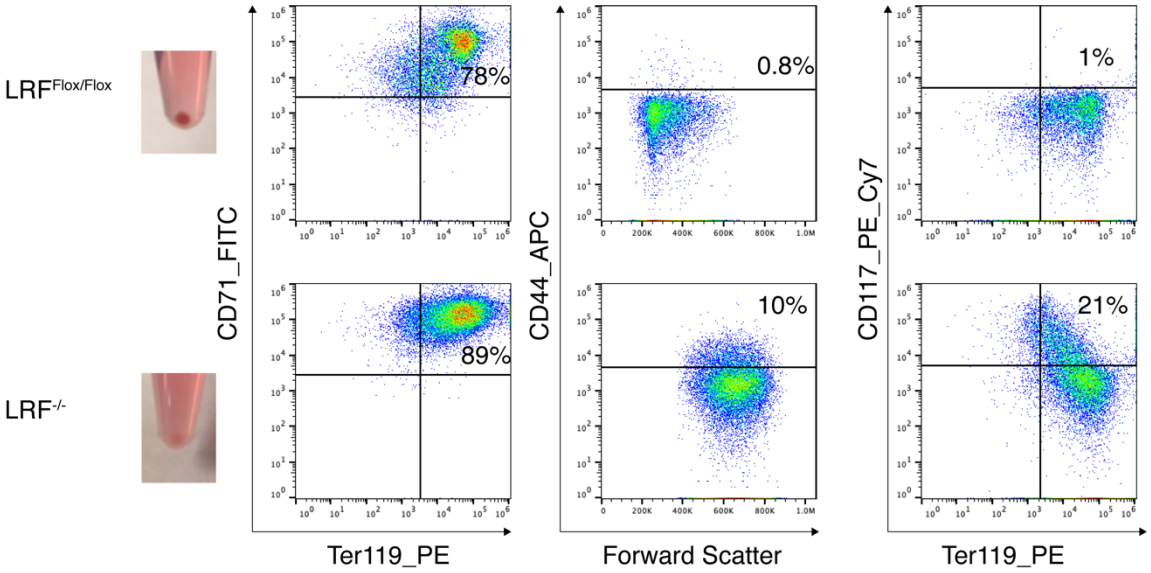


Figure 4.10 Differentiation of *Lrf*$^{Flox/Flox}$ and *Lrf*$^{-/-}$ littermates in an erythroid fetal liver differentiation culture system. *Lrf*$^{-/-}$ have a differentiation block, indicated by a larger size (forward scatter), and a proportion of cells retaining CD44 and CD117 (which are found on early stage erythroid cells). This is also seen morphologically, and macroscopically, with a paler cell pellet.

ATAC-seq was performed on erythroid cells derived from *Lrf*$^{-/-}$ fetal livers following 6 days expansion and 30 hours differentiation. Expression for the α-like and β-like globins was also performed on populations of cells. The results are shown in Figures 4.11-4.12.

At the α-globin locus, the expression data show that there is a small increase in α-globin (when normalised against all the β-globin like transcripts), and a significant increase in ζ-globin at a population level, such that the ζ-globin transcript accounts for ~5% of the α-like transcripts. In contrast to the primary fetal stage cells taken at

E12.5 however, the chromatin is inaccessible at the *Hba-x* gene. The enhancers and gene promoters generally have decreased accessibility in comparison to the WT cells; this is almost certainly due to the impaired differentiation.

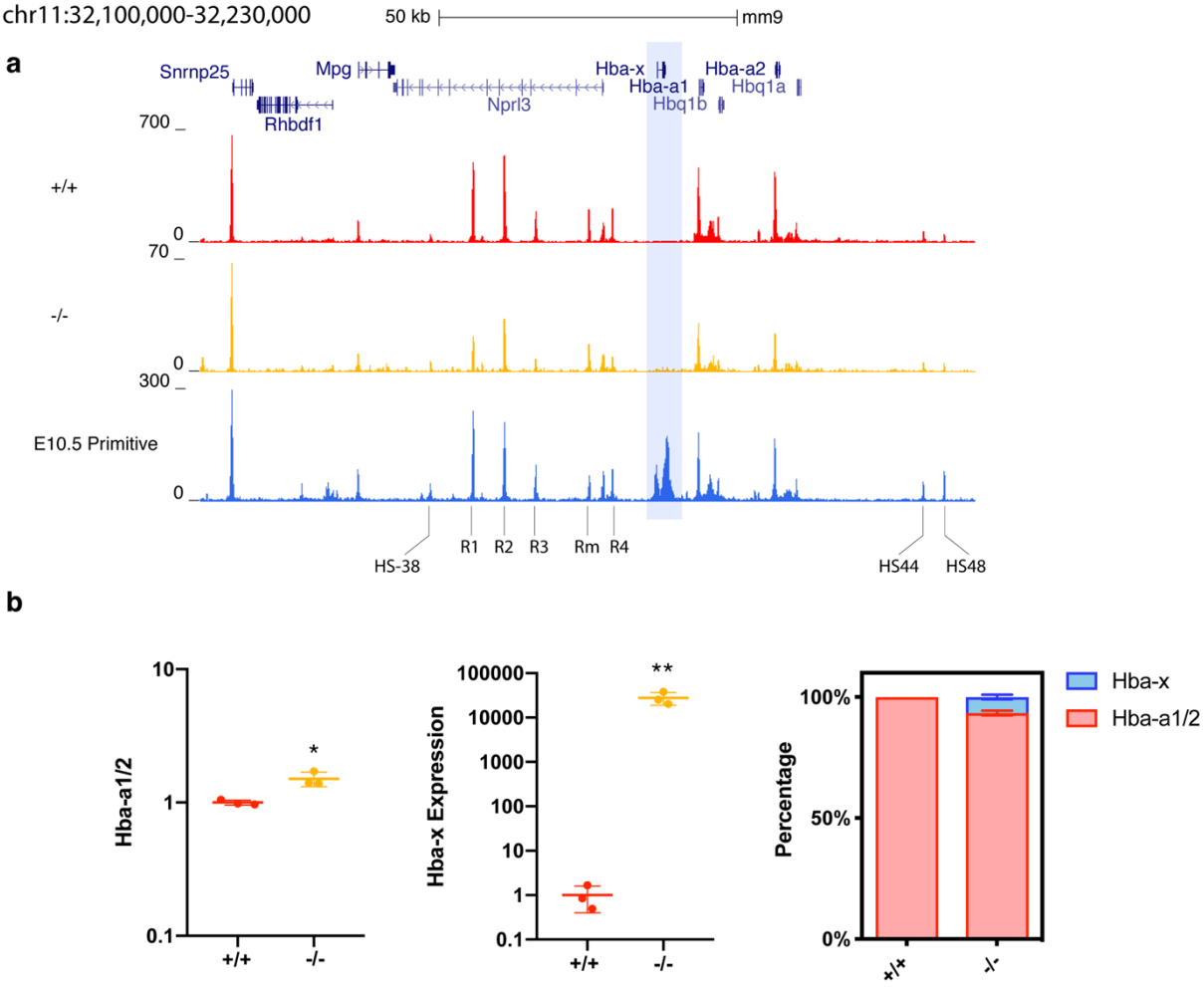


Figure 4.11 a. Chromatin accessibility and b expression of the α-like globins at the α-globin locus in *Lrf*$^{Flox/Flox}$ (+/+) and *Lrf*$^{-/-}$ (-/-) cells from the *in vitro* fetal liver culture system. Chromatin accessibility in primitive erythroid cells is also shown. *Hba-x* is highlighted in blue. **b.** Population expression data was normalized against all the β-like globins; then presented as a ratio over WT expression. The percentage of *Hba-x* expression as a total amount of α-like globin expression is also shown. * denotes p<0.05; ** denotes p<0.01. Error bars denote SEM. Experiments were performed in triplicate.

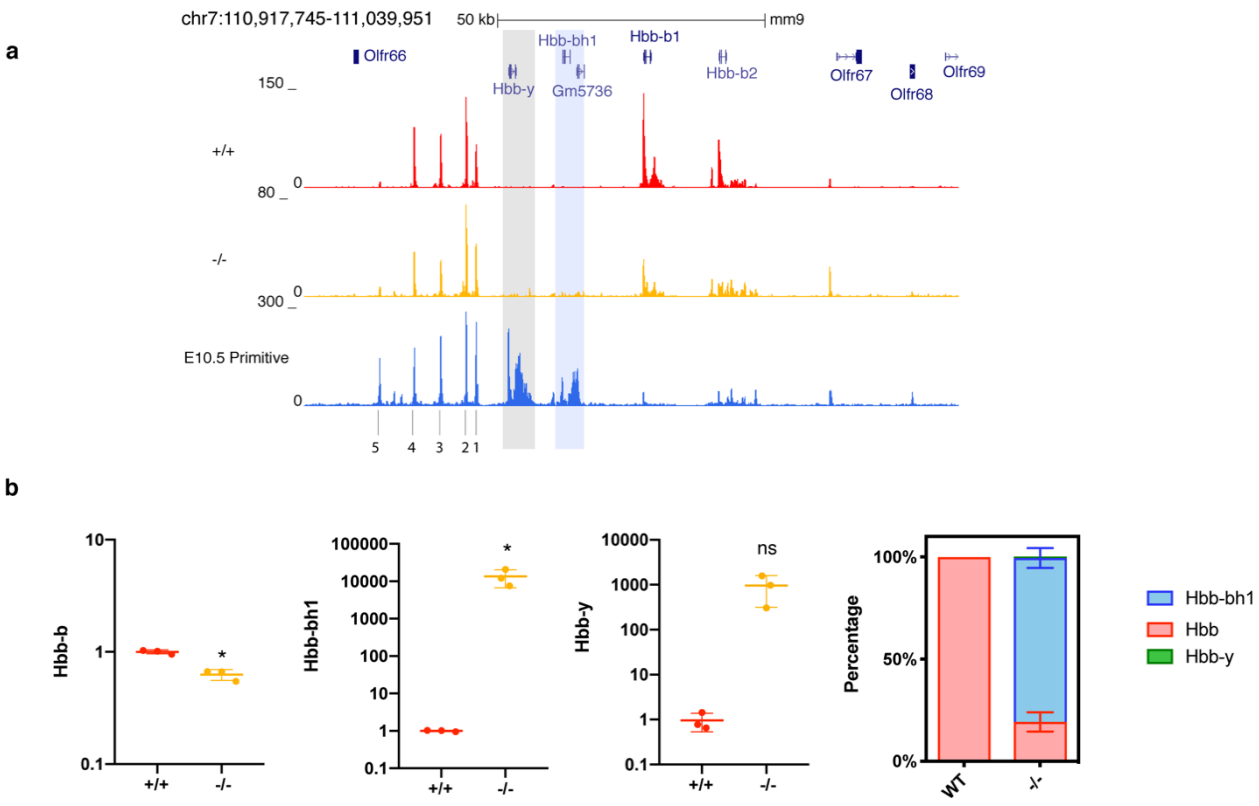


Figure 4.12 a. Chromatin accessibility and b. expression of the β-like globins at the β-globin locus in *Lrf*$^{Flox/Flox}$ (+/+) and *Lrf*$^{-/-}$ (-/-) cells from the in vitro fetal liver culture system. Chromatin accessibility in primitive erythroid cells is also shown. *Hbb-y* is highlighted in grey; *Hbb-bh1* is highlighted in blue. Population expression data was normalized against all the α-like globins; then presented as a ratio over WT expression. The percentage of the embryonic globin expression as a total amount of β-like globin expression is also shown. * denotes $p<0.05$; ** denotes $p<0.01$; ns not significant. Error bars denote SEM. Experiments were performed in triplicate.

At the β-globin locus, β-globin expression is decreased by about 40%. There is a significant increase in *Hbb-bh1* expression, and a non-significant increase in *Hbb-y* expression. The latter is expressed at a low level, close to the limit of detection of

the assay, and there was therefore significant variability in measurement. *Hbb-bh1* accounts for approximately 75% of the β-like transcripts in the *Lrf*$^{-/-}$ cells with *Hbb-y* accounting for <0.5%. The chromatin accessibility is again reduced at both the enhancers and adult β-globin promoters, likely due to the differentiation block. The chromatin is inaccessible over *Hbb-bh1* in contrast to the fetal stage E12.5 erythroid cells.

Use of this *in vitro* fetal liver differentiation system is clearly a very artificial substitute for later stage fetal erythroid cells. The *Lrf*$^{-/-}$ cells also have a differentiation block and decreased cell viability in culture, which generally leads to decreased and variable globin expression, illustrated by their paler pellet and the wide SEM bars in the expression data. It is difficult therefore, to be sure about the changes in globin expression. Despite these significant caveats, the chromatin accessibility data generated here suggest that there are additional cell intrinsic silencing mechanisms which occur in the *Lrf*$^{-/-}$ cells as they mature between E12.5 and birth.

4.3.4 Knockout of the *Bcl11a* composite enhancer leads to embryonic globin derepression and mild chromatin accessibility changes in early fetal erythroid cells

Primary fetal liver erythroid cells were isolated from *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$, *Bcl11a*$^{\Delta 10kb/WT}$ and *Bcl11a*$^{WT/WT}$ cells at E12.5. Flow cytometry from littermates with each genotype demonstrated similar differentiation profiles with *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ cells progressing through differentiation. *Bcl11a* knockout has previously been reported to cause both *Hbb-bh1* and *Hbb*-y de-repression (Smith et al., 2016). Given differentiation is

unperturbed the panel of genes used for the *Lrf* knockout experiments wasn't required. A different panel of genes was therefore used to instead ensure that the effects being seen in the experiment weren't simply due to primitive erythroid contamination (Table 4.3). Genes were selected which were differentially expressed between primitive and definitive erythropoiesis, based upon RNA-seq data generated in our laboratory by a previous DPhil student.

Such an experiment wasn't necessary with the *Lrf* knockout cells as *Hbb-y* remained repressed on a single cell level when *Hba-x* and *Hbb-bh1* were derepressed (and the effects seen can't therefore have been due to primitive erythroid contamination).

Globin	Lineage	Primitive/Definitive	α-globin cluster
Hba-a2 Hba-x Hbb-bh1 Hbb-bh2 Hbb-y Hbq1a Hbq1b Hbb-bt/bs/b1/b2	Ccr2(Monocyte) Mpo(Granulocyte) Ms4a6c(Monocyte) Prtn3 (Granulocyte) Lmo4(Basophil)	Abcg4 Aqp1 Aqp8 Arid3a Bcl11a Cited2 Cpeb4 Ddx21 Foxh1 H2afy Hspe1 Il9r Klf3 Zbtb7a Vamp5 Tmem14c Sox6 Sox4 Smarca4 Selk Rhd Pogz Podxl Pbx1 Oaz1 Myc Mbd2	Mpg Nprl3 Rhbdf1 Snrnp25

Table 4.3 (previous page) Genes used for analysis of *Bcl11a*$^{\Delta 10kb}$ mice

PCA analysis showed that the WT, heterozygous and homozygous *Bcl11a*$^{\Delta 10kb}$ clustered together (Figure 4.13). There was separation of the S3 cells away from the other differentiated cells. *Hba-x* expression and *Hbb-bh1* expression were absent in the WT cells, confirming that definitive erythroid cells, rather than primitive had been analysed.

Single cells were sorted from the S0-S3 populations; ATAC-seq was also performed in 5000 cells from S3. The results are shown in Figures 4.14-4.15.

When the *Bcl11a* enhancer is removed in fetal life, there is *Hba-x* de-repression. However, the chromatin accessibility changes are more minor in comparison to those seen when *Lrf* is knocked out. As in the *Lrf* knockout, θ-globin gene expression is unaffected; *Hba1/a2* expression is also unaffected or even slightly increased. Chromatin accessibility appears to be slightly increased at the R4 enhancer element, as seen in the *Lrf* knockout definitive erythroid cells.

At the β-globin locus, there is de-repression of both *Hbb-y* and *Hbb-bh1* with *Hbb* expression left unchanged. The chromatin accessibility changes in the homozygous *Bcl11a* enhancer knockout are again subtler than those seen in the *Lrf* knockout definitive erythroid cells.

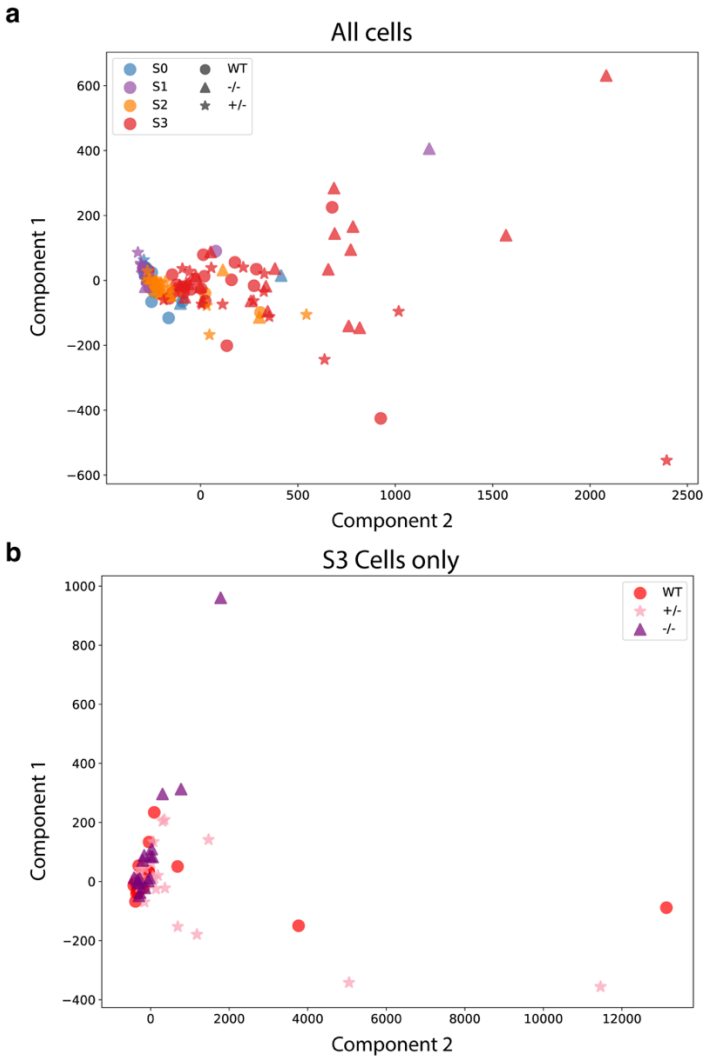
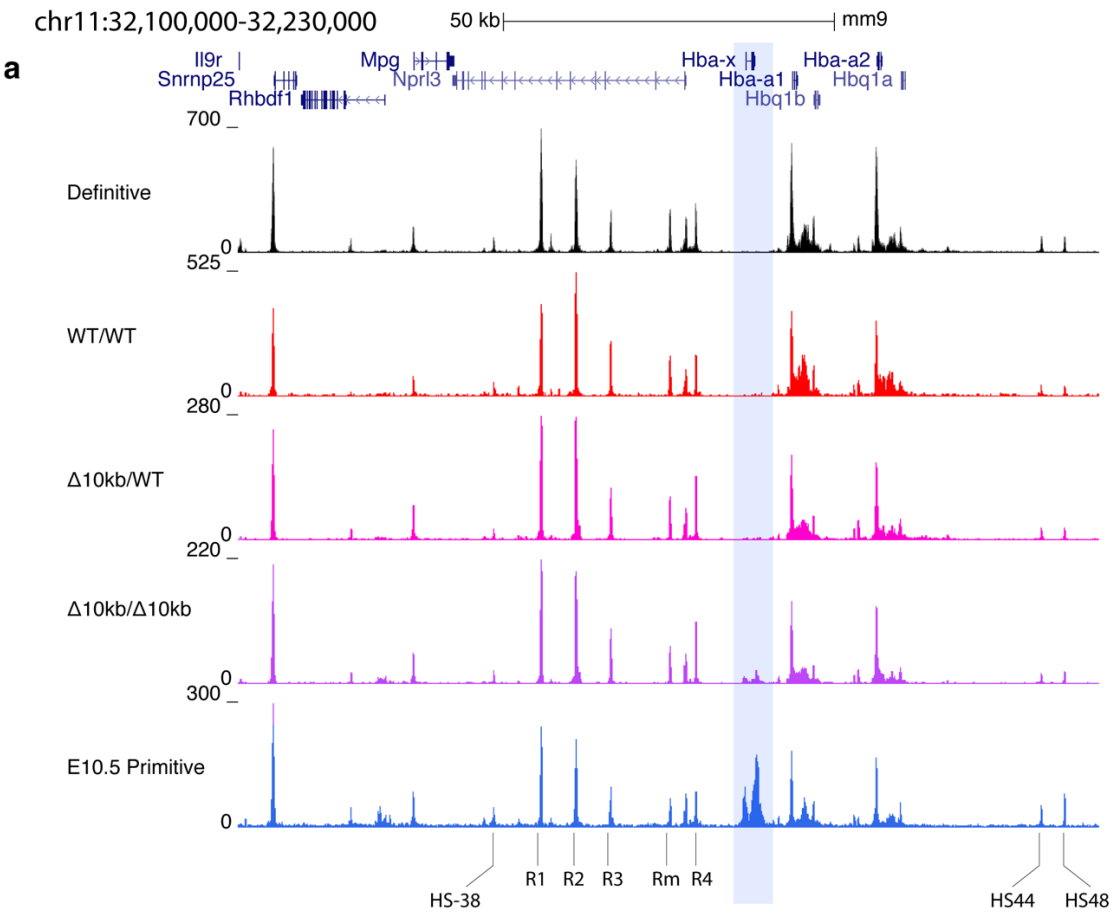
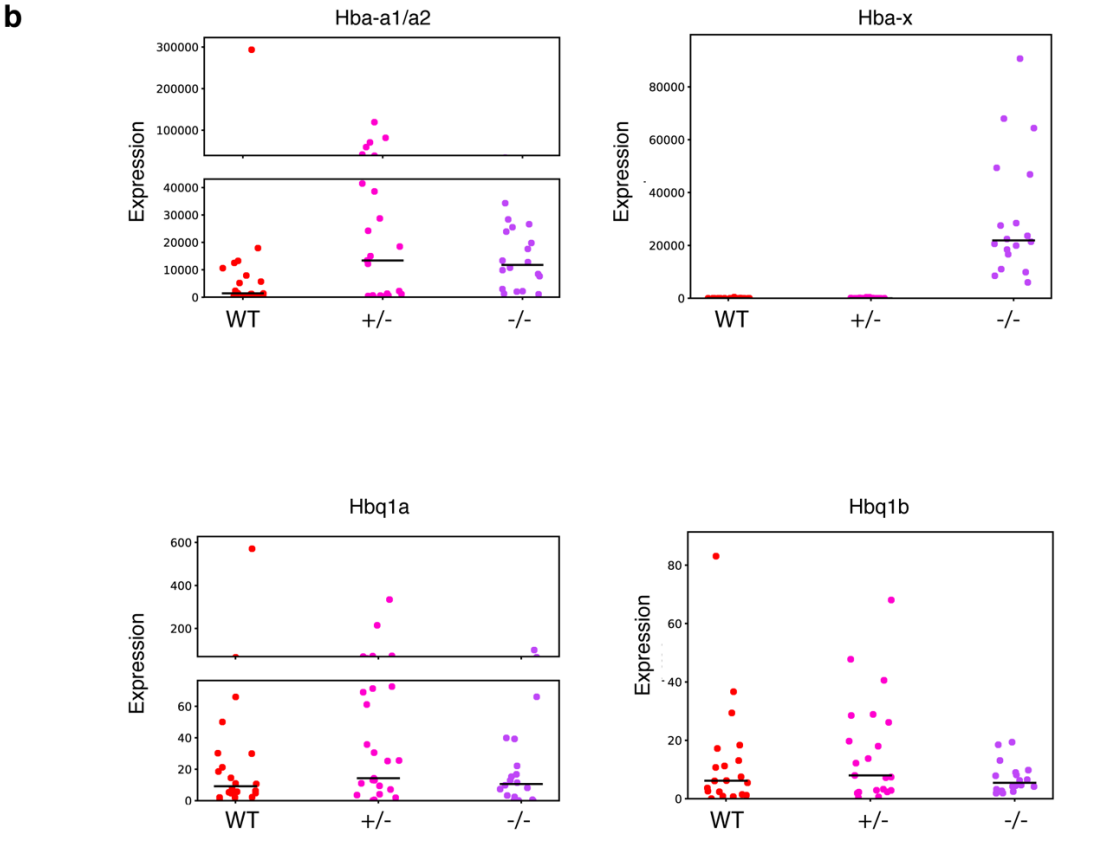
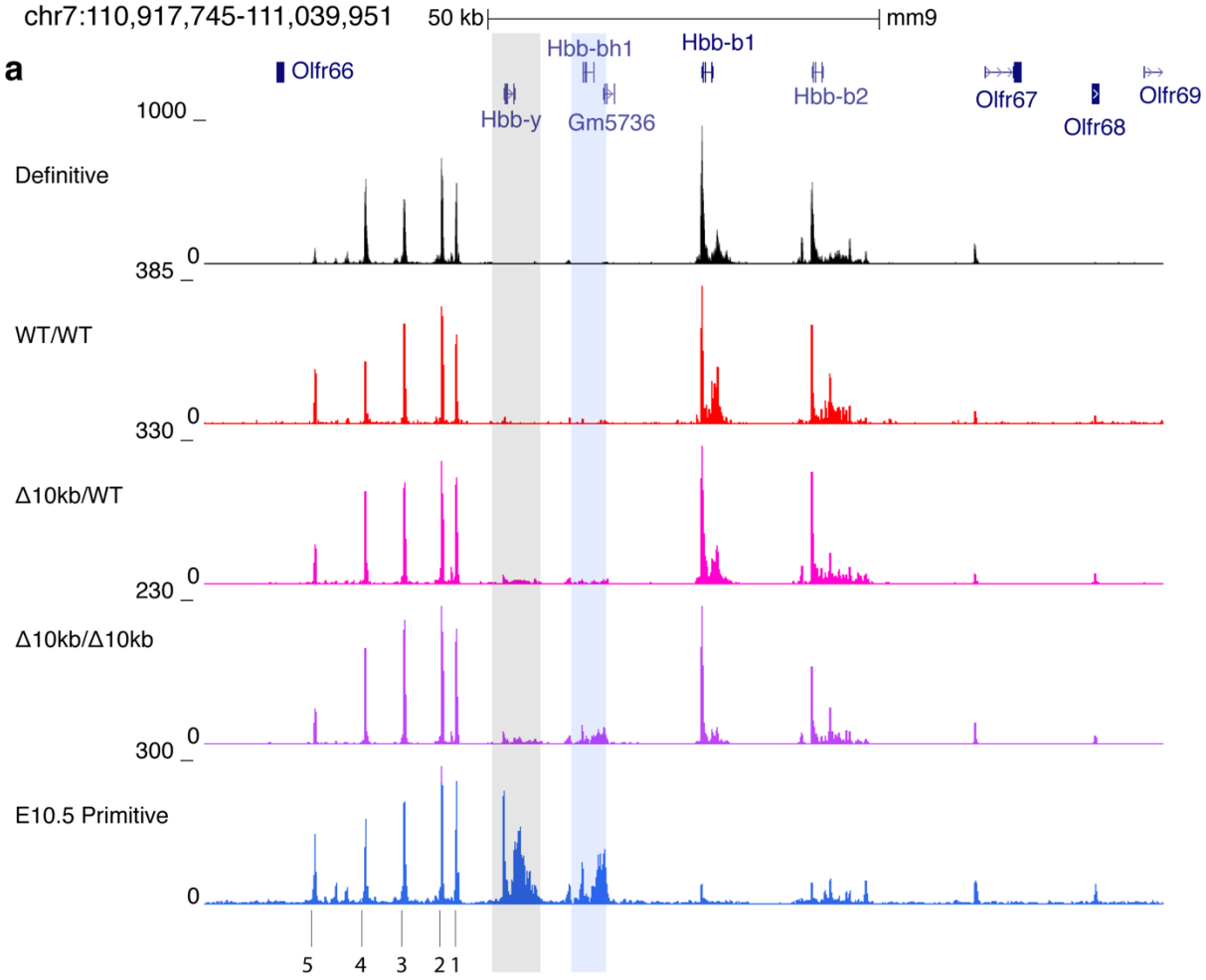


Figure 4.13 a. PCA Analysis for single erythroid cells for all stages (S0-S3) and b. just S3, from WT, heterozygote (+/-) and homozygote (-/-) *Bcl11a*$^{\Delta 10kb}$ knockout cells.

Figure 4.14 (next page) a. Chromatin accessibility and b. single cell expression at the α-globin locus in *Bcl11a*$^{WT/WT}$ (WT/WT), *Bcl11a*$^{\Delta 10kb/WT}$ (Δ10kb/WT) and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ (Δ10kb/Δ10kb) erythroid cells from the fetal liver S3 population. Chromatin accessibility in definitive and primitive (E10.5) erythroid cells are also shown. Tracks shown represent between 1 and 3 replicates; *Hba-x* is highlighted in blue. For expression data, each dot represents a single cell. The line indicates the median value.


a
chr11:32,100,000-32,230,000
50 kb
mm9
Il9r
Snrnp25
Rhbdf1
Mpg
Nprl3
Hba-x
Hba-a1
Hba-a2
Hbq1a
Hbq1b
Definitive
WT/WT
Δ10kb/WT
Δ10kb/Δ10kb
E10.5 Primitive
HS-38
R1
R2
R3
Rm
R4
HS44
HS48


b
Hba-a1/a2
Hba-x
Hbq1a
Hbq1b
Expression
WT
+/-
-/-


chr7:110,917,745-111,039,951
50 kb
mm9
a
Olfr66
Hbb-y
Hbb-bh1
Gm5736
Hbb-b1
Hbb-b2
Olfr67
Olfr68
Olfr69
Definitive
WT/WT
Δ10kb/WT
Δ10kb/Δ10kb
E10.5 Primitive
1000
385
330
230
300
0
5 4 3 2 1

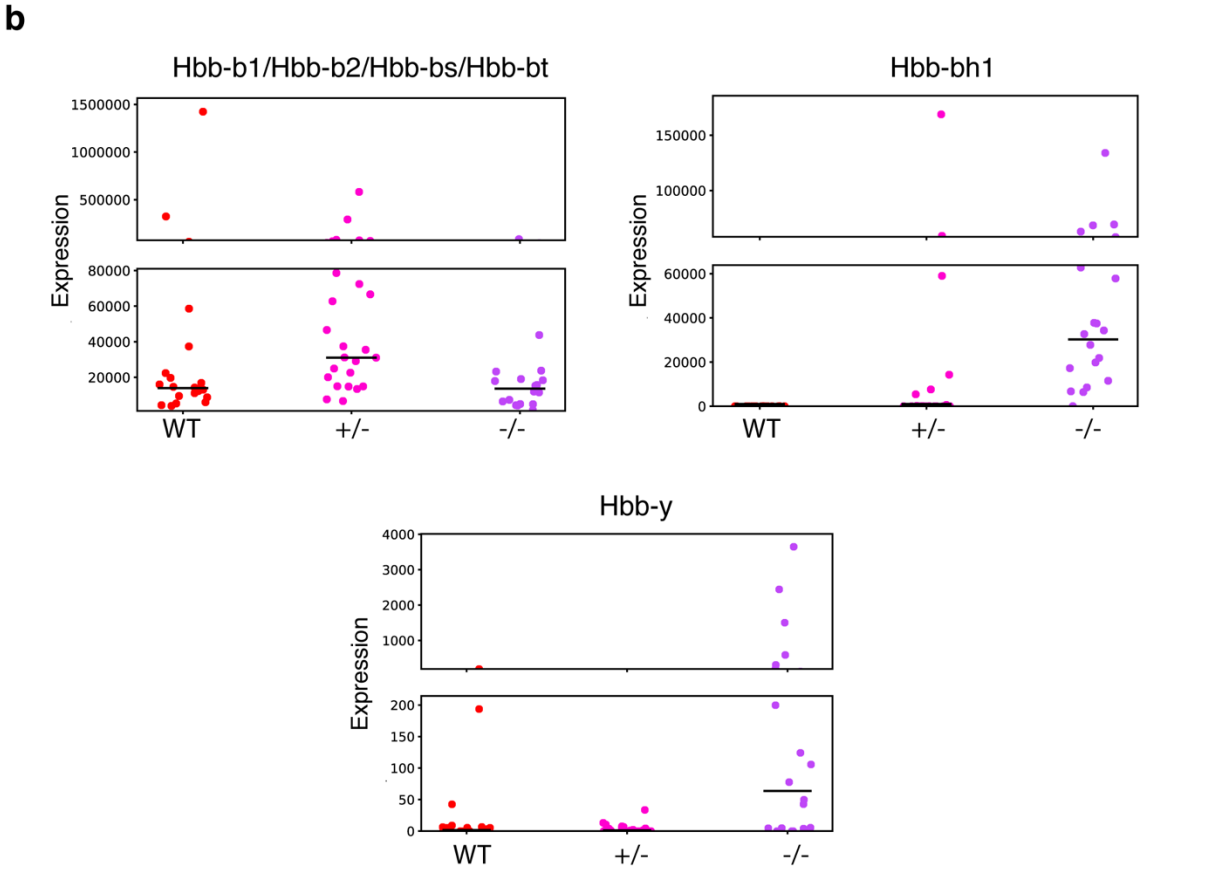

b
Hbb-b1/Hbb-b2/Hbb-bs/Hbb-bt
Hbb-bh1
Hbb-y
Expression
WT
+/-
-/-

Figure 4.15 (previous page) a. Chromatin accessibility and b. single cell expression at the β-globin locus in *Bcl11a*$^{WT/WT}$ (WT/WT), *Bcl11a*$^{\Delta 10kb/WT}$ (Δ10kb/WT) and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ (Δ10kb/Δ10kb) erythroid cells from the fetal liver S3 population. Chromatin accessibility in definitive and primitive erythroid cells are also shown; tracks shown represent single replicates from littermates; *Hbb-bh1* is highlighted in blue; *Hbb-y* is highlighted in grey. Regarding the expression data, each dot represents a single cell. A probe targeting all the adult β-like globin genes was used. The line indicates the median value.

Taken together, these results demonstrate that Bcl11a is important for repression of all the embryonic globin genes in fetal life. Its removal causes de-repression, with chromatin changes seen over the genes and gene bodies. The changes are less pronounced over both *Hba-x* and *Hbb-bh1* than when *Lrf* is knocked out; this suggests that, at least in fetal life, Lrf plays a greater role in *Hba-x* repression.

4.3.5 Knockout of the *Bcl11a* erythroid specific enhancers causes embryonic globin gene de-repression in adult erythroid cells with minimal changes in chromatin accessibility

Bcl11a$^{\Delta 10kb/\Delta 10kb}$ mice are viable and develop normally into adulthood. We wondered if the changes in chromatin accessibility and embryonic globin derepression seen in early fetal definitive erythropoiesis would be as pronounced in adult definitive erythroid cells. It is possible to analyse such cells by isolating Ter119+ erythroblasts from phenylhydrazine treated spleens (Davies et al., 2016; Hanssen et al., 2017; Spivak et al., 1973) and comparing them to WT erythroid cells.

ATAC-seq was performed to determine differences in chromatin accessibility; RNA expression of the globins on populations of cells was also assayed using RT-qPCR. The results are shown in Figures 4.16.4.17.

In adult erythroid cells, removal of the *Bcl11a* composite erythroid enhancer (with subsequent return of *Bcl11a* expression to that seen in primitive erythroid cells) causes minimal changes in chromatin accessibility at the globin loci. There is perhaps, an increase in accessibility seen at R4 at the α-globin locus, and a mild decrease in accessibility seen at the both the α and β globin genes, the significance of which is uncertain. The most striking observation however, is that the chromatin at the embryonic globin genes at both the α and β globin loci remains inaccessible.

Following removal of the *Bcl11a* composite enhancer homozygously, there is a clear and significant increase in embryonic globin expression, with *Hba-x, Hbb-bh1 and Hbb-y* expression all increasing between 100 and 1000 fold in comparison to WT.

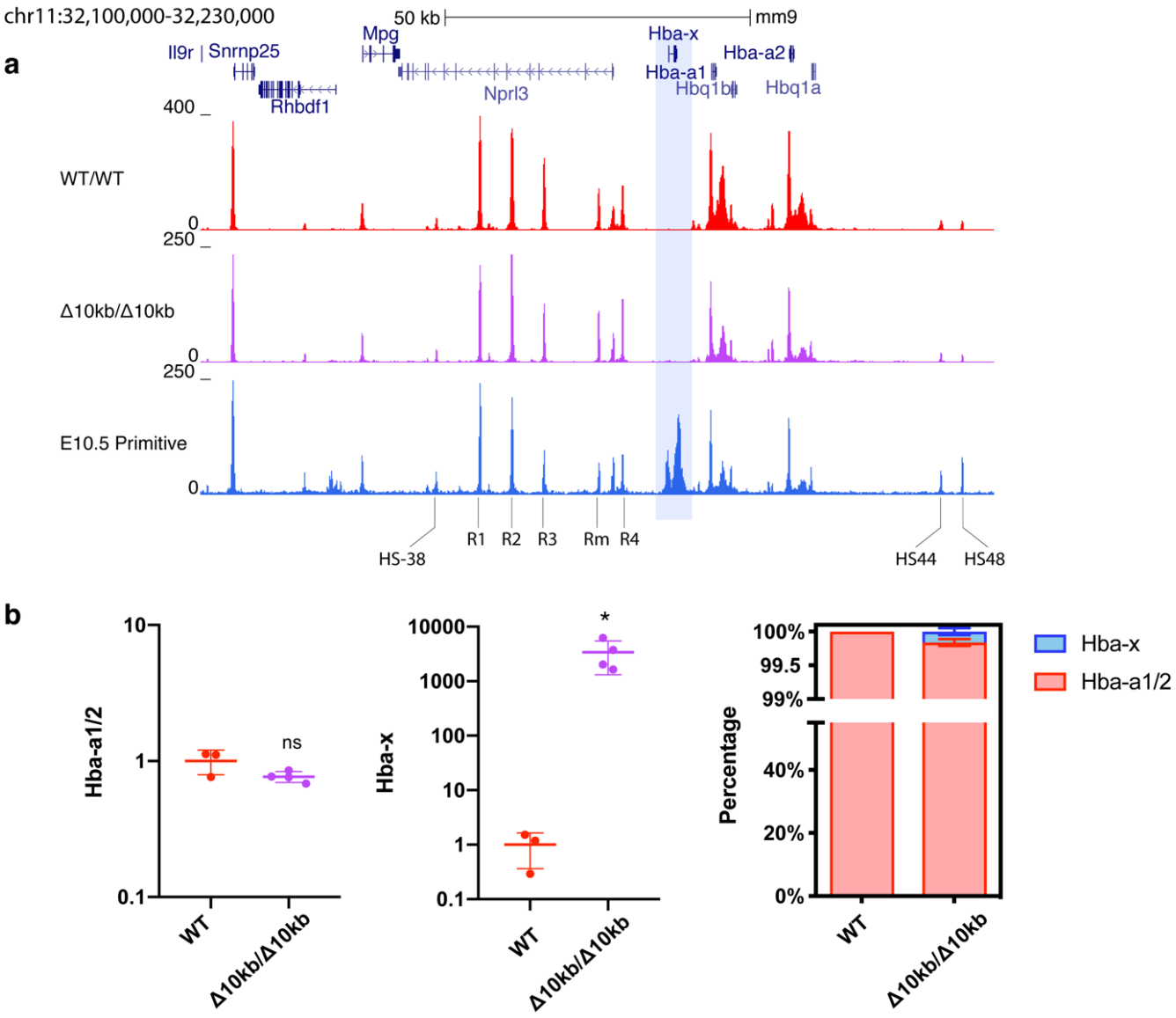


Figure 4.16 a. Chromatin accessibility and b. expression of the α-like globins from the α-globin locus in WT and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ splenic Ter119+ erythroblasts. Chromatin accessibility in primitive erythroid cells is also shown. *Hba-x* is highlighted in blue. Population expression data was normalized against all the β-like globins; then presented as a ratio over WT expression. The percentage of the *Hba-x* globin expression as a total amount of α-like globin expression is also shown. * denotes p<0.05; ns not significant. Error bars denote SEM. 4 animals were used for the *Bcl11a* $^{\Delta 10kb/\Delta 10kb}$ experiments; 3 WT animals were used.

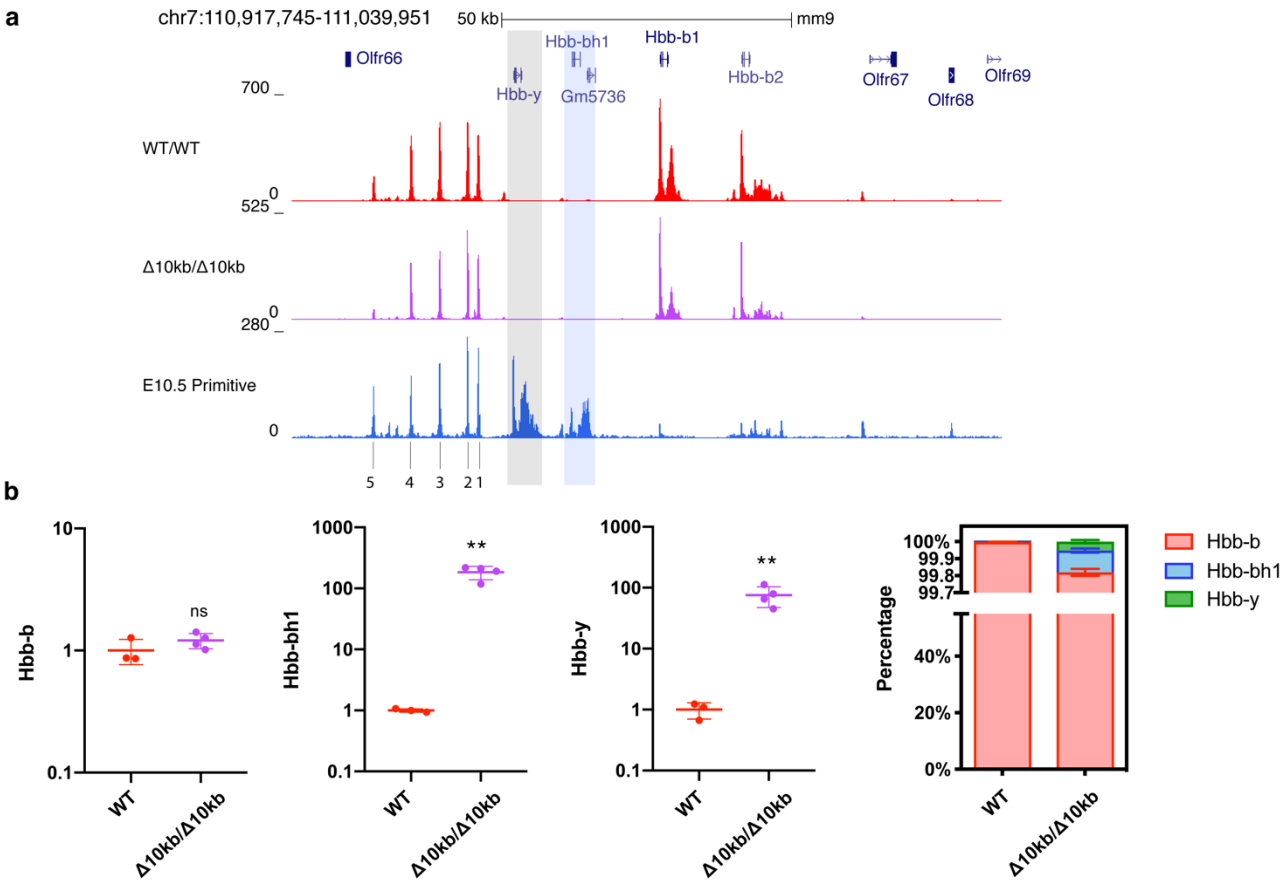


Figure 4.17 a. Chromatin accessibility and b. expression of the β-like globins from the β-globin locus in WT and *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ splenic Ter119+ erythroblasts. Chromatin accessibility in primitive erythroid cells is also shown. *Hbb-bh1* is highlighted in blue; *Hbb-y* is shown in grey. Population mature RNA expression was normalized against all the β-like globins; then presented as a ratio over WT expression. The percentage of the embryonic globins as a total amount of β-like globin expression is also shown. ** denotes p<0.01; ns not significant. Error bars denote SEM. 4 animals were used for the BCL11A $^{\Delta 10kb/\Delta 10kb}$ experiments; 3 WT animals were used.

However, these increases are trivial when considered in the context of the entirety of globin expression, with the embryonic globins accounting for <0.2% of globin expression from both the α and β globin loci.

Together, these results demonstrate that Bcl11a plays a role in embryonic globin gene repression in murine adult erythroid cells and its removal leads to embryonic globin gene de-repression. However, the amount of expression and chromatin accessibility changes are less when compared to that seen in E12.5 definitive erythroid cells.

4.4 Discussion

Here, it is demonstrated that Bcl11a and Lrf play a key role in repressing ζ-globin *in vivo* with their absence leading to ζ-globin de-repression in fetal and adult-type murine erythroid cells.

H3K4me1, H3K4me3 and H3K27ac ChIP-seq data generated in definitive erythroid cells shows that there are four putative enhancer elements for *Bcl11a* located in its introns, when correlated with ATAC-seq (Figure 4.2). Two of these (denoted E1 and E2 here) have not been characterised (or reported) further *in vivo* or *in vitro*. The two other enhancers have previously been denoted +58 and +62, as they are orthologous to the enhancers for the human *BCL11A* gene found at those positions in the human locus (Canver et al., 2015). These have been demonstrated *in vivo* as essential for *Bcl11a* expression in erythroid cells, whilst leaving *Bcl11a* expression in lymphoid and brain tissue unperturbed (Smith et al., 2016). Capture-

C data generated here, demonstrate that the *Bcl11a* promoter interacts with the entire enhancer block in adult erythroid cells when it is expressed.

Reviewing our previously published super-enhancer analysis (Hay et al., 2016), the composite *Bcl11a* enhancer (made up of the four individual enhancer elements) does not meet the arbitrary definition of a super-enhancer, when ranking enhancers by either Med1 or H3K27ac. Nevertheless, given it is a composite enhancer, it is of interest to observe the effects on chromatin accessibility when two enhancer elements both removed (Figure 4.4). There are marked changes, with the promoter and two other putative enhancer elements becoming inaccessible; this is associated with *Bcl11a* expression falling to levels seen in primitive erythroblasts. By contrast, removal of the two strongest enhancer elements at the α-globin locus (R1 and R2), showed no effect on formation of the other enhancer elements (Hay et al., 2016). From these data, it seems that a simple additive model may not apply at all composite enhancers *in vivo*, with some enhancer elements proving essential for formation of a homocomplex. It must be highlighted however, that the two putative enhancer elements (denoted E1 and E2 here) have not been characterised or shown to function as individual enhancers alone.

Previous analysis of the *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ mouse model reported no phenotypic abnormalities, with normal haematopoietic lineages when assessed by flow cytometry, and robust engraftment of *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ haematopoietic cells in lethally irradiated mice (Smith et al., 2016). Whilst the mice analysed here were initially judged as phenotypically unremarkable with normal differentiation observed at E12.5 by flow cytometry, haematological analysis revealed a mild anaemia and

a higher reticulocyte count in adult *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ mice, suggesting a shorter red cell lifespan. Data presented in abstract form last year showed decreased erythroid reconstitution in xenotransplantation models following *BCL11A* enhancer knockout in human CD34^{+} cells, with increased apoptosis markers by flow cytometry, although this remains unpublished. Contrasting reports have recently been published however, showing normal erythroid, myeloid and lymphoid human reconstitution in xenotransplantation models with robust HbF induction following targeting of the GATA1 binding motif in the BCL11A enhancer (Wu et al., 2019), Different targeting strategies may explain this variability, and targeting the *BCL11A* enhancer remains a current therapeutic strategy for β-globin haemoglobinopathies, and will soon enter Phase I clinical trials.

Absent *Lrf* expression *in vivo* leads to a clear differentiation block and increased apoptosis in fetal erythroid cells taken at E12.5-E13.5. No *Lrf*$^{-/-}$ offspring survived to term, as previously reported (Maeda et al., 2009). Previous work has shown that the increased apoptosis is mediated via up-regulation of the pro-apoptotic factor *Bim,* with *Lrf* acting as a direct repressor of *Bim* transcription (Maeda et al., 2009). Apoptosis was observed throughout all stages of erythroid differentiation when staining was performed using Annexin-V in the experiments here.

Because of the differentiation block, single cell gene expression was performed at E12.5 alongside genes which vary in differentiation; in combination with a sorting strategy this ensured that like-for-like cells were compared. Regarding the globin genes, there was up-regulation of the *Hba-x* gene and *Hbb-bh1* gene, without up-regulation of *Hbb-y*; similarly, chromatin accessibility at the globin loci was

increased over those genes. These findings are in complete agreement with those previously published (Masuda et al., 2016). Despite *Hbb-y* being expressed at a very similar timepoint to *Hba-x*, it is still completely silenced when *Lrf* is knocked out.

Removal of the *Bcl11a* composite enhancer homozygously with a consequent reduction in *Bcl11a* expression to those seen in primitive erythroid cells doesn't lead to a differentiation block. At the α-globin locus, there is an increase in ζ-globin expression and chromatin accessibility; however, this is to a lesser extent than that seen in the *Lrf* knockout cells. At the β-globin locus, there is de-repression of both *Hbb-y* and *Hbb-bh1* expression, with concurrent increases in chromatin accessibility. The increase in chromatin accessibility over the *Hbb-bh1* gene is again less than that seen with *Lrf* knockout. Together, these data are supportive of Lrf and Bcl11a operating via different pathways: their knockout leads to differing effects on chromatin accessibility in stage-matched fetal erythroid cells, and *Lrf* knockout affects *Hba-x* and *Hbb-bh1* only in contrast to the *Bcl11a* enhancer knockout which de-represses all the murine embryonic globin genes.

During development, the onset of the third wave of erythropoiesis is heralded by the emergence of haematopoietic stem cells (HSCs) from the major arterial vessels of the embryo (Kumaravelu et al., 2002). These initially seed the developing liver; then around the time of birth, migrate to the bone marrow where definitive erythropoiesis is maintained for life. Derepression of the embryonic β-like globin genes by *Bcl11a* knockout appears to be more marked in early fetal, as opposed to adult life (Sankaran et al., 2009; Smith et al., 2016). In order to see whether similar effects

where observed on ζ-globin expression and chromatin accessibility, experiments were performed in adult erythroid cells from the *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ mice and later stage fetal cells from the *Lrf* knockout mice.

Given knockout of *Lrf* homozygously leads to embryonic lethality at around E16.5, a proxy for late stage fetal erythroid cells had to be used. E12.5 fetal livers were therefore placed through an *in vitro* 8-day culture system. There are clearly significant caveats to this approach, including that the cells fail to differentiate completely, meaning they weren't strictly stage matched against their wild-type litter-mates when compared on a population level. Nevertheless, chromatin accessibility does decrease over the embryonic globin genes after this approach, when compared to primary E12.5 cells. Adult erythroid cells from *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ cells isolated from the spleen also show inaccessible chromatin at the embryonic globin genes in contrast to the early stage fetal erythroid cells. The embryonic globins do remain derepressed, although account for a trivial amount of the total globin expression from each locus.

It therefore appears that there are additional unidentified silencing mechanisms occurring during fetal erythropoiesis, and between fetal and adult erythropoiesis. Although both fetal liver and adult bone marrow derived cells are considered definitive, there are transcriptional differences (Kingsley et al., 2013). This could potentially have important translational implications, namely that any therapies aimed at increasing ζ-globin would have to be confirmed as being effective in adult, bone marrow derived erythroid cells.

The classic dogma is that active regulatory elements within the genome, including enhancers, promoters and boundary elements are characterised by co-operative binding of sequence specific transcriptional regulatory factors; these lead to a remodelled chromatin state which is characterised by marked accessibility to nucleases including DNAse I (Thurman et al., 2012) and Tn5 (Buenrostro et al., 2013). It is consequently of interest that the embryonic globin genes are transcribed in adult-type erythroid cells despite the chromatin appearing inaccessible using a standard ATAC-seq assay. Whilst the embryonic globin transcription as a proportion of the total globin transcription is low (~0.2% of the total α-like transcript in the *Bcl11a*$^{\Delta 10kb/\Delta 10kb}$ is *Hba-x*) this is still a significant amount when considered in the context of the massive levels at which *Hba-a1/2* are transcribed.

In conclusion, the data in this chapter demonstrate that both Lrf and Bcl11a play a role in the silencing of *Hba-x* in definitive erythropoiesis in the mouse with LRF knockout having a greater effect than Bcl11a removal. Their knockout (or removal of their enhancers in the case of Bcl11a) *in vivo* leads to changes in chromatin accessibility in definitive erythroid cells at the α-globin loci, with an associated increase in *Hba-x* expression. Lrf also represses *Hbb-bh1* but doesn't play a significant role in *Hbb-y* silencing; Bcl11a represses both *Hbb-bh1* and *Hbb-y*. Their removal appears to have a greater effect in early fetal, as opposed to late fetal and adult definitive erythroid cells.

5 The chromatin landscape in human erythroid cells at the α-globin locus and the effect of *LRF* and *BCL11A* knockout on ζ-globin expression

5.1 Introduction

The work presented so far in this thesis has characterised the chromatin landscape around the murine ζ-globin gene in primitive and definitive erythropoiesis, demonstrated that two transcription factors, Lrf and Bcl11a, play a key role in its silencing via chromatin remodelling, and shown silencing is associated with histone deacetylation. The tractable nature of mouse models and the easy accessibility of fetal and embryonic tissue mean the mouse is an excellent tool for testing hypotheses and clarifying mechanism. However, given there is a potential translational benefit from de-repressing ζ-globin, it is important that similar experiments are performed in human cells.

For many years, K562 cells formed the "workhorse" for *in vitro* experiments investigating human globin switching. These cells were originally derived from a patient with chronic myeloid leukaemia in blast crisis who had erythroleukemia (Lozzio and Lozzio, 1977; 1975). They are straight-forward to use and differentiate and provide a virtually unlimited supply of material. Like all malignant cell lines however, they have an abnormal karyotype. They also express predominantly fetal/embryonic globins rather than adult-type globins. Because, they are derived

from a malignant clone, it is difficult to know how faithfully they recapitulate normal erythropoiesis.

More recently, human ES cells and iPS cells have been differentiated towards erythroid cells, the latter following reprogramming (Chang et al., 2006; Lapillonne et al., 2010; Olivier et al., 2006; 2016). iPS cells offer the opportunity to reprogram cells from patients with rare mutations; this is clearly of significant benefit, yielding a potentially unlimited supply of material for subsequent experiments whilst only requiring a single donation from any one individual. They also usually have a normal karyotype. However, it has proved challenging to force iPS cells to differentiate through to adult definitive erythroid cells with their transcriptional profile seemingly representing a yolk sac erythromyeloid progenitor program rather than adult, bone marrow derived haematopoietic stem cells (Vanuytsel et al., 2018).

The gold standard *in vitro* model for experiments on globin expression remains primary $CD34^+$ cells collected from either umbilical cord or adult donors; these are then differentiated to erythroid precursors and reticulocytes using tissue culture protocols lasting for between 14 and 21 days (Migliaccio et al., 2002; Ronzoni et al., 2008). Their end result can be tens of millions of erythroid cells from a single adult donor donating approximately 100mls of peripheral whole blood. However, on a clonal level, they have limited proliferative potential, being capable of expanding to a few thousand cells only following genome editing (Mettananda et al., 2017a). This makes conventional cell population-based experiments such as ChIP-seq impossible following genome editing.

Given these limitations, the globin switching community has needed a new human cellular erythroid model. Ideally, this would be immortal (thereby providing an unlimited supply of cells); under normal circumstances recapitulate definitive erythropoiesis and be straight forward to genetically manipulate and expand from a single clone to millions of cells. There was therefore considerable interest in the publication in 2013 of the HUDEP and HiDEP cell lines (Kurita et al., 2013). These were originally developed to provide an additional source for allogeneic blood transfusion. Umbilical cord CD34+ (HUDEP) and iPS cells (HiDEP) were immortalised using lentiviral delivery of a tetracycline- inducible HPV16-E6/E7 transgene; this inhibits the p53 and retinoblastoma genes, hence leading to on-going cell growth in perpetuity, providing a tetracycline is present in the culture medium (Münger and Howley, 2002). Whilst the majority of the clones derived in the original publication had a fetal and/or embryonic pattern of globin expression, one, the HUDEP-2 clone, only expressed adult α and β globin. Subsequent publications demonstrated that genetic engineering of this line was possible and that genetically edited clones could be grown up from a single cell (Moir-Meyer et al., 2018). The development of a Cas9 expressing HUDEP-2 cell line also facilitated CRISPR knockout screens (Canver et al., 2015).

DNA methylation, whereby cytosine residues are methylated in the context of CpG residues in a reversible manner is a major epigenetic modification in multicellular organisms (Jones, 2012). More than half of the genes in vertebrates contain short (~0.5-2kb) CpG rich regions known as CpG islands, the majority of which are unmethylated; the rest of the genome is depleted for CpGs (Jones, 2012). At transcription start sites, there is excellent evidence that methylation of CpG islands

leads to gene repression (Jones, 2012). The role of DNA methylation in promoters which are CpG deplete is less clear, with conflicting reports: some suggest an inverse relationship between CpG methylation and transcription (Han et al., 2011) whilst others report no association (Weber et al., 2005). In humans, the ζ-globin gene promoter has only 6 CpGs, although there is a CpG island located at the 3' region of the gene. The methylation profile across the promoter and this CpG island have not yet been studied and it isn't therefore known whether DNA methylation plays a role in ζ-globin regulation.

Given the widespread scientific interest in erythropoiesis and globin switching, genome wide data sets have been published and made freely available for several histone marks in primary erythroid cells derived from $CD34^+$ cells (Xu et al., 2012). Re-analysis of these in combination with novel data generated here allows a detailed understanding of the epigenetic landscape at the α-globin locus, and consequently potential repressive mechanisms for ζ-globin.

In this chapter, I have initially characterised the HUDEP-2 cell line. I have then used three separate genetically modified HUDEP-2 lines where transcription factors have been perturbed to understand the contribution of BCL11A and LRF to ζ-globin silencing in human haematopoiesis. I have also studied DNA methylation across the ζ-globin gene when it is repressed. Finally, I have analysed previously published data to understand whether other repressive mechanisms including H3K9me1/2, H3K9me3 and H3K27me3 play a role in ζ-globin repression in primary definitive erythroid cells.

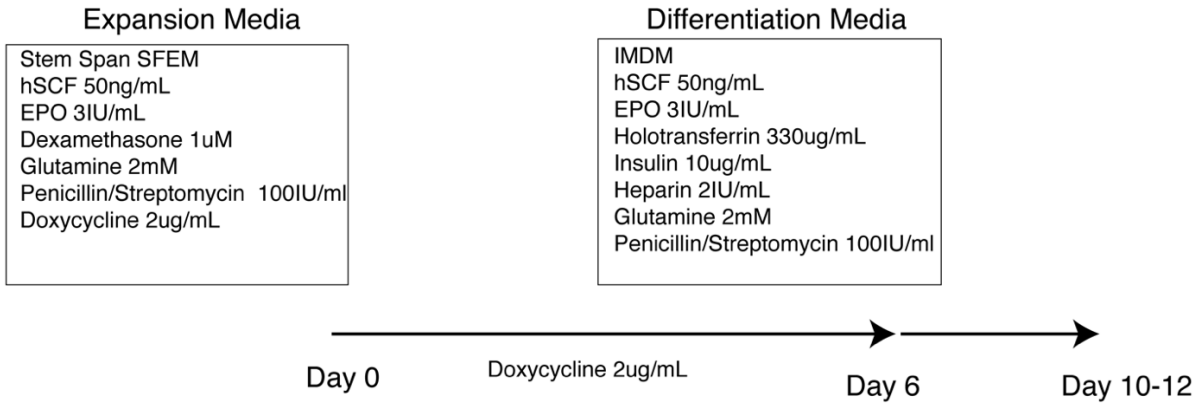
5.2 Results

5.2.1 Characterisation of the HUDEP-2 cell line

5.2.1.1 Differentiation Protocol

A 10-12 day differentiation protocol was developed based upon the previous literature (Canver et al., 2015; Kurita et al., 2013; Masuda et al., 2016) (Figure 5.1). There is clear morphological differentiation with the cell pellet macroscopically becoming red.

a



b

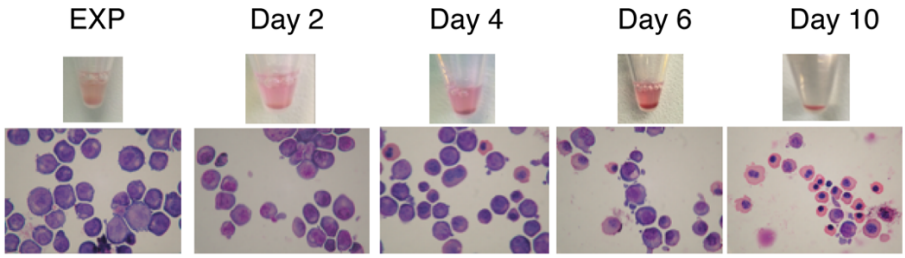


Figure 5.1 (previous page) a. Differentiation protocol for HUDEP-2 cells. Terminal differentiation requires removal of doxycycline at Day 6. **b. Macroscopic and microscopic appearance of HUDEP-2 cells during differentiation.** The cell pellet becomes red as the cells differentiation; by Day 10-12 of differentiation the cells are mainly orthochromatic erythroblasts.

Terminal differentiation requires removal of doxycycline for the last 3-5 days. This is presumably because the natural process of terminal erythroid differentiation necessitates cell cycle exit (Dzierzak and Philipsen, 2013), and the continuous signal to proliferate provided by the doxycycline induced E7/E8 expression hampers this. Very few of the cells enucleate, as reported by others (Vinjamur and Bauer, 2018).

5.2.1.2 Cell surface flow cytometry changes during differentiation

Staining with antibodies against CD34, CD71, CD44, Band 3 (CD233), CD49d (α4-intergrin) and CD235a (Glycophorin A) during differentiation is shown in Figure 5.2 as single antibody stains; no antibody negative controls also shown. As erythroid cells differentiate, they acquire higher CD235a and Band 3 expression and lose CD71 expression, CD36 expression and CD49d expression. They are negative for CD34.

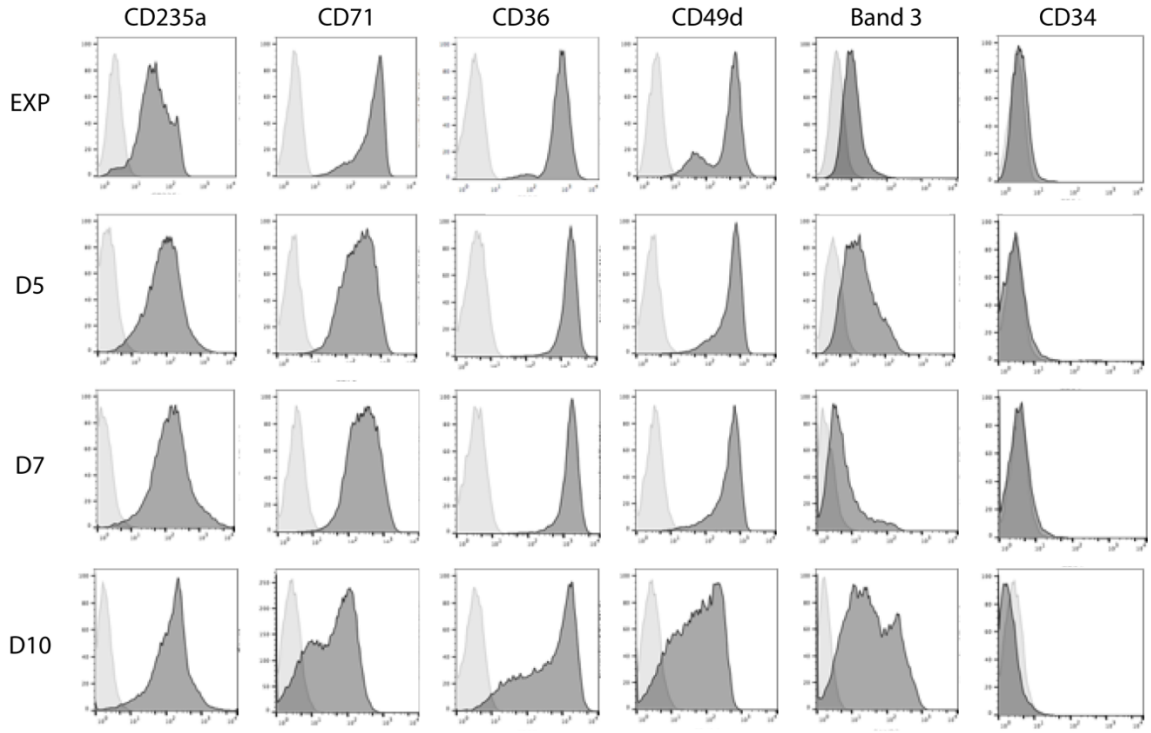


Figure 5.2 Flow cytometric analysis of expression of cell surface membrane proteins during HUDEP-2 differentiation. The y-axis demonstrates the number of cells expressing the fluorescent intensity on the x-axis. Note the increase in CD235a and Band 3 intensity and the decrease in CD71, CD36 and CD49d during differentiation. No antibody controls are shown in pale grey.

The change in expression of cell surface markers during terminal differentiation allows a sorting strategy that can be used to obtain pure populations of cells (Figure 5.3) based upon a previous publication in primary erythroid cells (Hu et al., 2013).

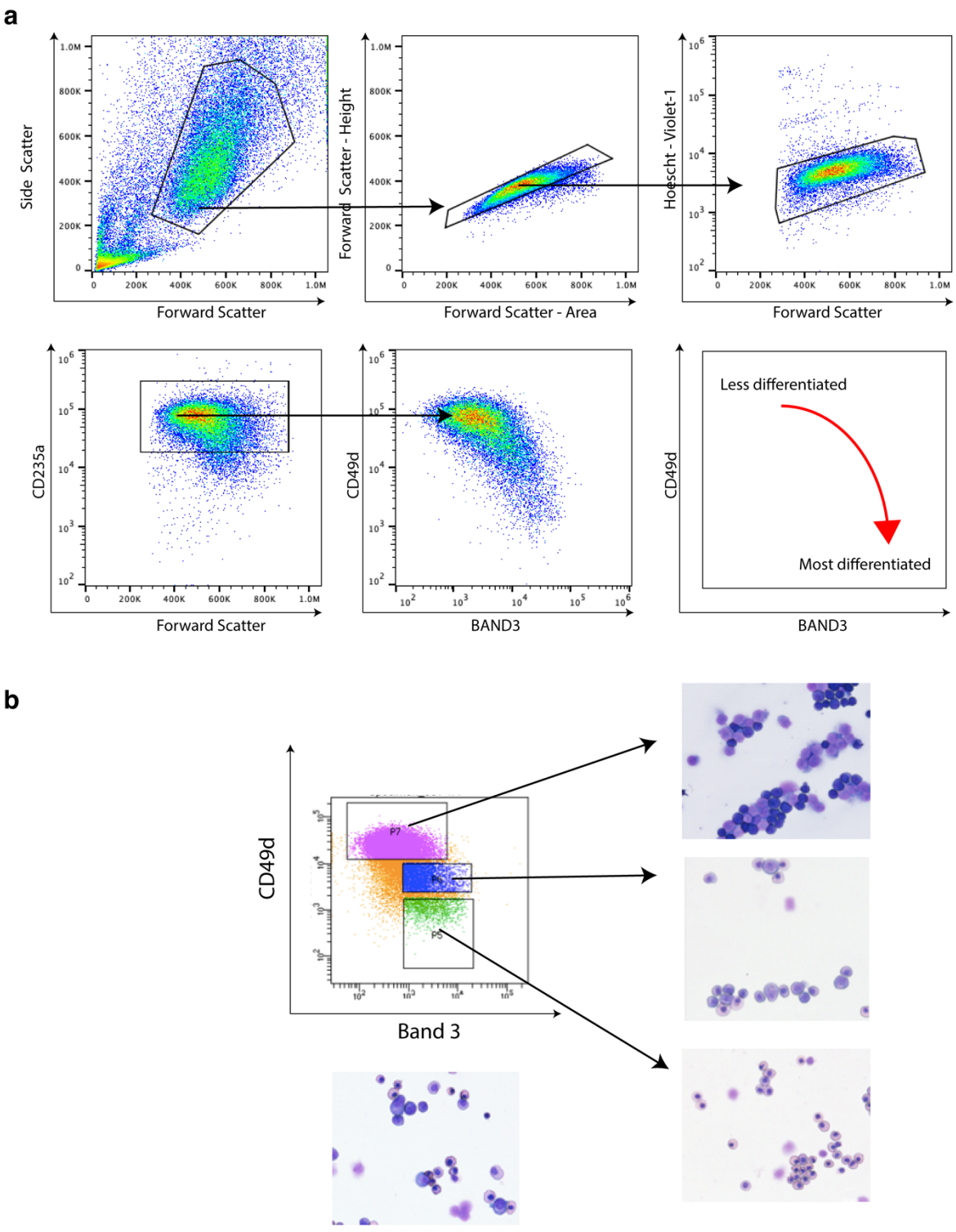


Figure 5.3 a. Sorting protocol which can be used to separate HUDEP-2 cells at Day 10 of differentiation; cells are sorted on CD235a, then by Band3 against CD49d. **b. Homogenous populations can be obtained corresponding to late basophilic, polychromatic and orthochromatic erythroblasts.**

5.2.1.3 Globin expression and haemoglobin production

Expression analysis was performed on HUDEP-2 cells during expansion phase and differentiation using RT-qPCR. Cells express predominantly adult-type haemoglobins (*HBA*, *HBB* and *HBD)* in the expansion phase although small amounts of *HBZ* and *HBE* are detectable (Figure 5.4a). The differentiation protocol leads to significantly increased expression of all globins, with the exception of *HBZ* although the expression of *HBE, HBD and HBG* remains trivial (Figure 5.4a). This is illustrated by isoelectric focusing (IEF) with the only apparent haemoglobin being HbA ($\alpha_2\beta_2$) (Figure 5.4b).

Flow cytometry was also performed using a well-established antibody directed against HbF (Masuda et al., 2016) and an antibody against ζ-globin (Luo et al., 1988). The latter was characterised prior to use to ensure it was specific to ζ-globin (details are contained in the Materials and Methods chapter). Flow cytometry showed that the majority of cells are double negative for HbF and ζ-globin; 5% are positive for HbF only, 2.5% of cells are positive for ζ-globin only, and 0.5% are double positive.

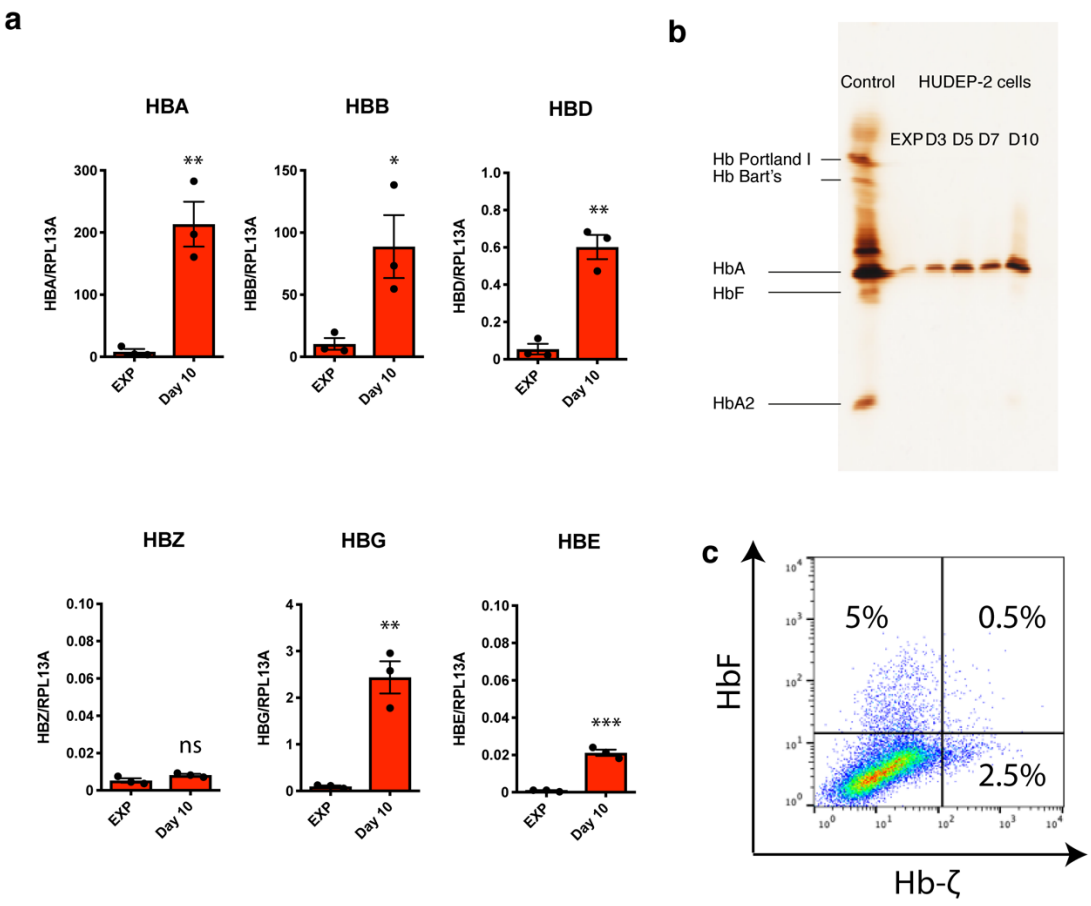


Figure 5.4 a. Expression analysis for HBA, HBB, HBD, HBZ, HBG and HBE from HUDEP-2 cells in expansion phase and at Day 10 of differentiation, normalized to RPL13A. * denotes p<0.05, ** denotes p<0.01; *** denotes p<0.001. Significance was determined by a two-tailed Student's t-test. **b. Representative IEF of HUDEP-2 cells during differentiation.** The only haemoglobin visible is HbA ($\alpha_2\beta_2$). **c. Representative flow cytometry on HUDEP-2 cells for HbF and ζ-globin.**

5.2.1.4 Chromatin Accessibility

To assess chromatin accessibility, ATAC-seq was performed on three sorted populations from Day 10 HUDEP-2 cells: late basophilic erythroblasts,

polychromatic erythroblasts, and orthochromatic erythroblasts (Figure 5.5 and Figure 5.6). The α and β-globin loci are shown, alongside chromatin accessibility (as assessed by ATAC-seq) from primary erythroid cells derived from a $CD34^+$ cell differentiation system.

Chromatin accessibility for all HUDEP-2 stages at the α-globin and β-globin loci is very similar to that seen from primary erythroid cells derived from a $CD34^+$ selected erythroid culture system (data generated by Damien Downes). There is a slight increase in accessibility seen as cells enter the polychromatic erythroblast stage, and a decrease seen as cells enter the orthochromatic stage; this is more prominent at the β-globin locus. At the β-globin locus, the chromatin at the γ-globin genes in the primary erythroid cells is accessible (though to a much lower level than the adult genes); this is consistent with low level γ-globin expression. γ-globin expression is known to occur with erythroid cells derived from $CD34^+$ cells, even when cells are collected from adult peripheral blood collect so this observation is unsurprising (Mettananda et al., 2017b). Given chromatin accessibility was very similar at the α-globin locus and it proved difficult to obtain large numbers of cells for experiments from cell sorting for ChIP-seq, further experiments were performed on unselected HUDEP-2 differentiated cells at Day 10/11.

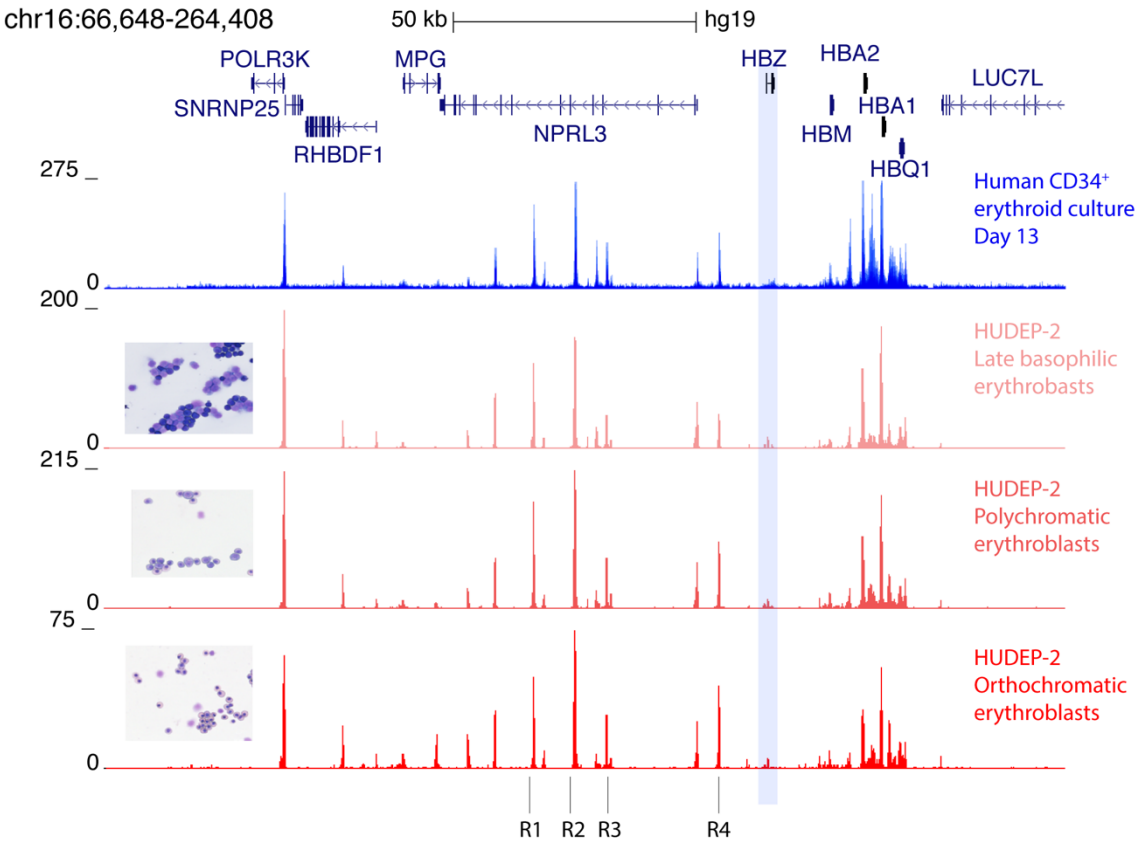


Figure 5.5 Chromatin accessibility at the α-globin locus in HUDEP-2 cells. ATAC-seq is shown for erythroid cells derived from a human CD34+ culture system (corresponding broadly to polychromatic/orthochromatic erythroblasts), and sorted HUDEP-2 cells for three different populations (cytospins are shown for those different populations). The *HBZ* gene highlighted by a blue rectangle. The enhancers are shown (R1-R4). Data are normalized by RPKM and combined ≥3 separate differentiations.

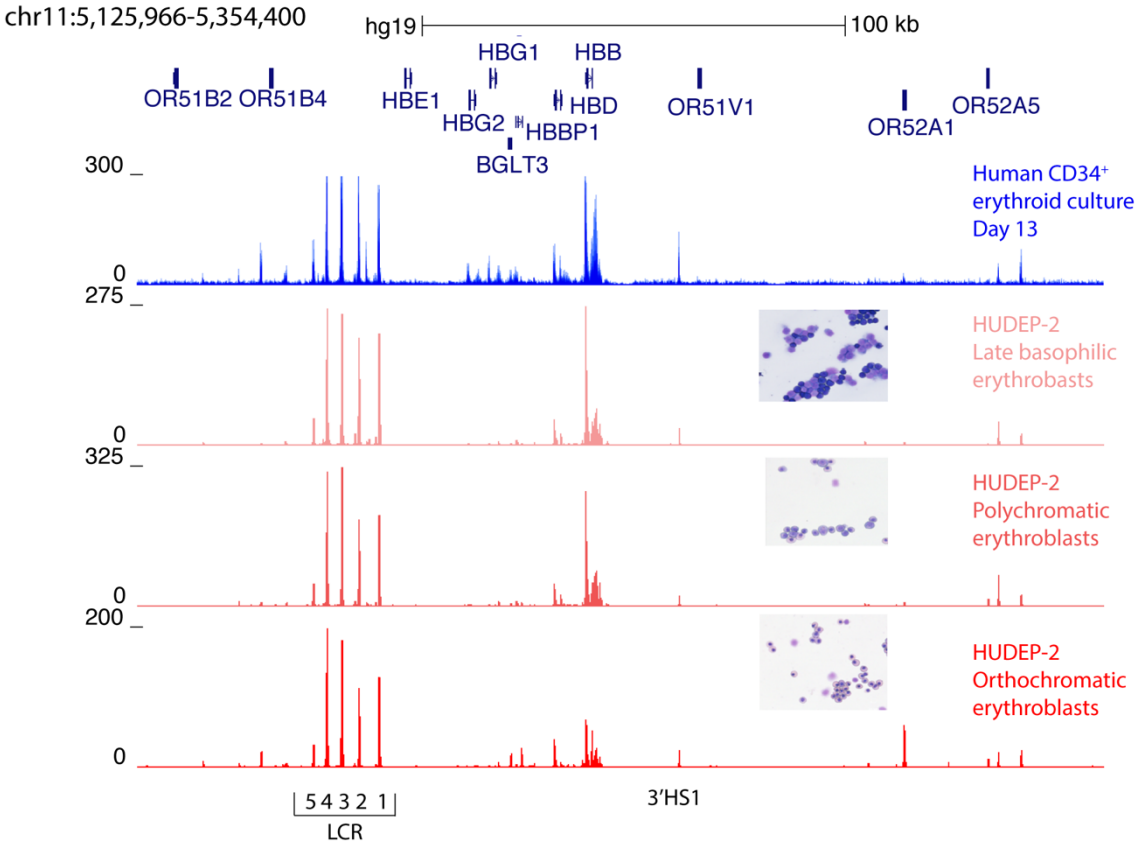


Figure 5.6 Chromatin accessibility at the β-globin locus in HUDEP-2 cells.

ATAC-seq is shown for erythroid cells derived from a human CD34+ culture system (corresponding broadly to polychromatic/orthochromatic erythroblasts), and sorted HUDEP-2 cells for three different populations. The locus control region (LCR) is shown. Data are normalized by RPKM and combined ≥3 separate differentiations.

5.2.1.5 Histone marks and CTCF binding

ChIP-seq for H3K27ac, H3K4me3 and CTCF was performed on HUDEP-2 cells from Day 10 of differentiation and compared to primary erythroid cells derived from a $CD34^+$ selected erythroid culture system across both globin loci. The histone marks and CTCF binding are almost identical across the two cell types at the α-globin locus (Figure 5.7). The ζ-globin gene is deacetylated, and doesn't mark for H3K4me3. At the β-globin locus the profiles are very similar between primary erythroid cells derived from $CD34^+$ cells (Figure 5.8). However, histone acetylation and the H3K4me3 mark are present over the γ-globin genes in primary erythroid cells, consistent with low level expression.

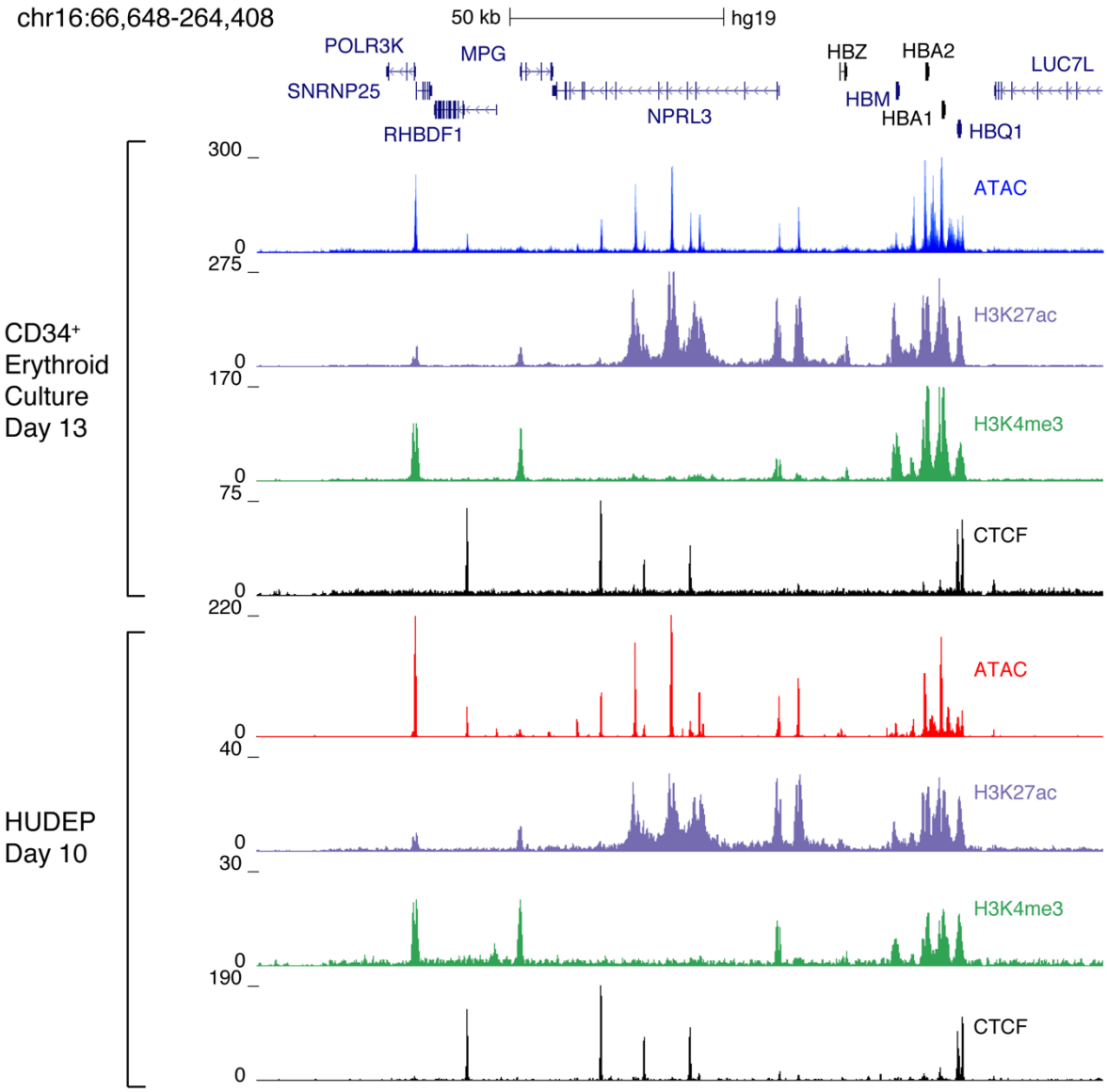


Figure 5.7 Histone marks and CTCF binding for primary erythroid cells and HUDEP-2 cells at the α-globin locus. Primary erythroid cell data are from a CD34+ selected erythroid culture system and were generated with Damien Downes. HUDEP-2 ChIP-seq data consist of a single replicate performed on Day 10 of the HUDEP-2 culture system; ATAC-seq data are in triplicate. Data are normalized by RPKM.

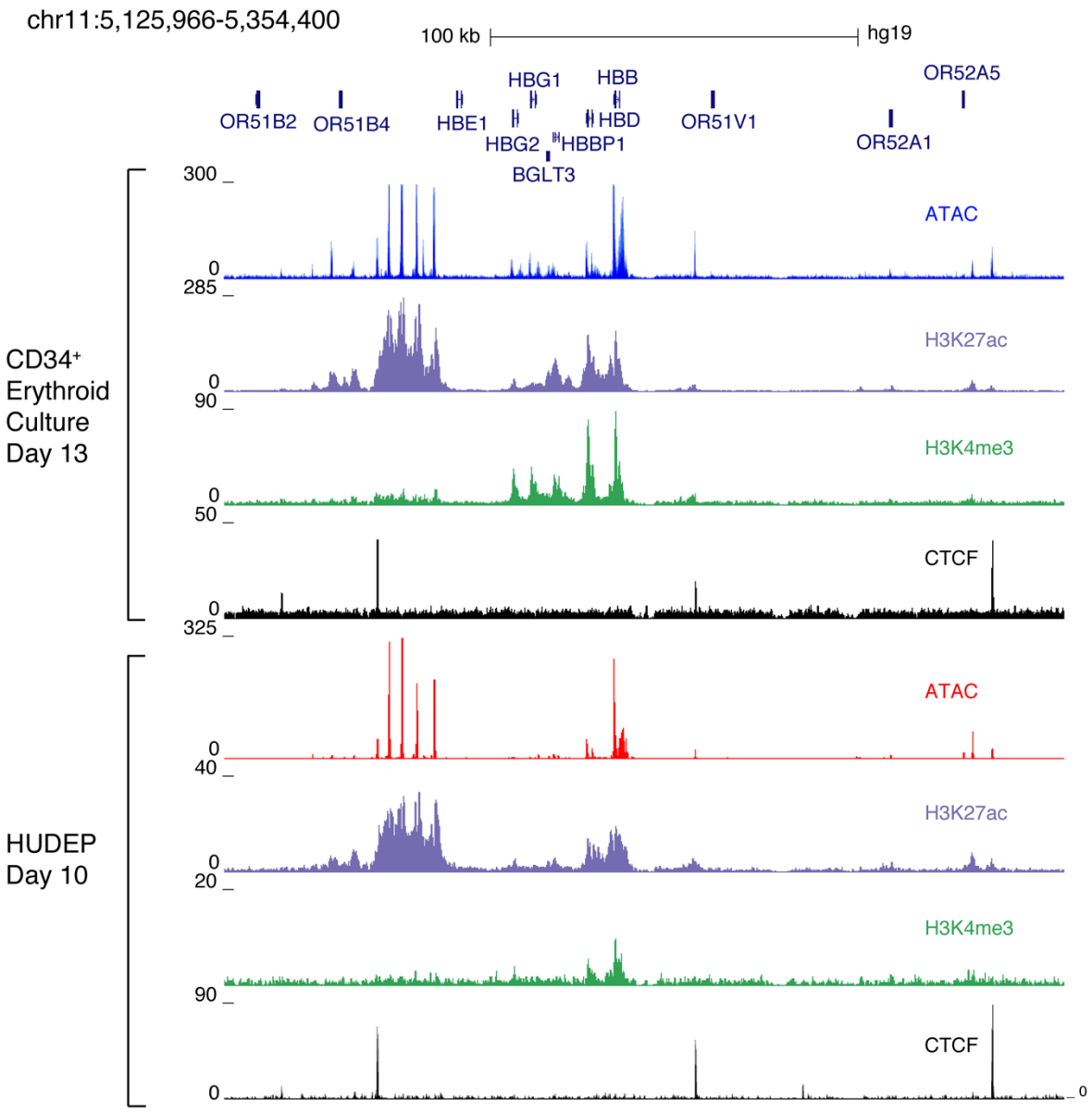


Figure 5.8 Histone marks and CTCF binding for primary erythroid cells and HUDEP-2 cells at the β-globin locus. Primary erythroid cell data are from a CD34^{+} selected erythroid culture system and were generated with Damien Downes. ChIP-seq data consist of a single replicate performed on Day 10 of the HUDEP-2 culture system; ATAC-seq data are in triplicate. Data are normalized by RPKM.

5.2.1.6 HUDEP-2 karyotype

Acquisition of genetic mutations and chromosome abnormalities frequently occurs in immortalized and transformed cell lines, even when the starting cells have a normal diploid state (Rebuzzini et al., 2015). Prior to utilizing the HUDEP-2 cell line, karyotyping and array comparative genomic hybridization (aCGH) was performed in collaboration with Samantha Knight. Chromosome counts on WT lines obtained from different sources revealed a modal chromosome count of 51 XY with a range of 49-53 chromosomes (an example is shown in Figure 5.7). aCGH demonstrates trisomies of chromosomes 6, 17 (partial ~30–80 Mb), 18 (partial ~18–78 Mb), 19 and 21 (partial ~14–48 Mb). There is also variable gain of material along the length of chromosome 8 and smaller regions of loss of heterozygosity (Table 5.1)

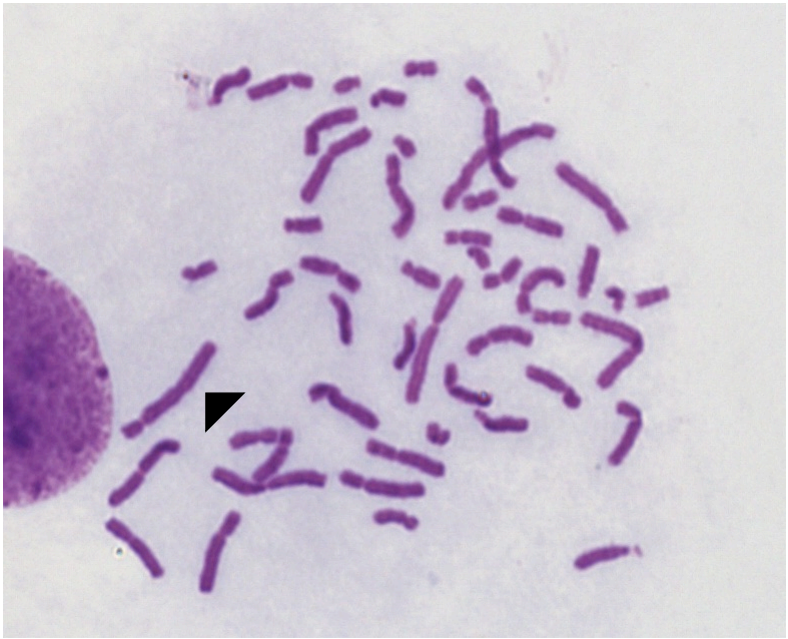


Figure 5.9 Karyotype of a WT HUDEP-2 cell line. In addition to being hyperdiploid (50 chromosomes in this case), structural abnormalities are present (for example, the chromosome marked with an arrowhead); performed with Jill Brown.

Chromosome	Start (hg18 *)	End (hg18)	Event	Length	Cytoband	Count of Genes
3	46,990,391	50,479,675	ROH	3,489,285	p21.31	121
3	156,276,396	159,447,085	ROH	3,170,690	q25.31–q25.33	24
6	0	171,115,067	CN Gain	171,115,068	p25.3–q27	1,433
8	0	83,821,458	CN Gain	83,821,459	p23.3–q21.13	597
8	83,821,459	110,503,487	CN Gain	26,682,029	q21.13–q23.2	165
8	110,503,488	125,424,290	CN Gain	14,920,803	q23.2–q24.13	65
8	125,424,291	138,306,862	CN Gain	12,882,572	q24.13–q24.23	76
8	138,306,863	146,364,022	CN Gain	8,057,160	q24.23–q24.3	141
14	37,393,981	37,510,754	CN Loss	116,774	q13.3	2
17	30,301,223	81,195,210	CN Gain	50,893,988	q11.2–q25.3	1,022
18	0	15,375,878	CN Loss	15,375,879	p11.32–p11.21	119
18	18,561,020	78,077,248	CN Gain	59,516,229	q11.1–q23	310
19	0	59,128,983	CN Gain	59,128,984	p13.3–q13.43	1,872
21	14,368,320	48,129,895	CN Gain	33,761,576	q11.2–q22.3	363

* Genome build NCBI36/hg18 was used. ROH: Region of homozygosity; CN: copy number.

Table 5.1 aCGH on the WT HUDEP-2 cell line.

Taken together, these data show that the HUDEP-2 cell line is an adequate model system with which to study silencing of ζ-globin. The cells follow a similar differentiation profile to primary erythroid cells in terms of morphology and differentiation assessed by cell surface markers (Hu et al., 2013). RNA expression of the globin genes shows predominantly adult-type globin expression, with minimal embryonic and fetal globin expression; assessment of protein production by both flow cytometry and IEF is congruent with gene expression. The chromatin accessibility profile, histone marks and CTCF binding across both globin loci are very similar to that seen in primary erythroid cells. The cytogenetic abnormalities affecting the HUDEP-2 cells do not affect the chromosomes where the globin genes are located appear normal.

5.2.2 Knockout of *LRF* and *BCL11A* separately leads to embryonic and fetal globin de-repression in the HUDEP-2 cell line

HUDEP-2 clones where *LRF* and *BCL11A* have been knocked out (confirmed by western blot in the donating laboratory and published in Masuda et al., 2016) have previously been generated to study their effects on γ-globin expression. These cell lines were imported into the laboratory and characterized as a collaborative project.

After 10 days differentiation, the *LRF*$^{-/-}$ clone was consistently more differentiated than the WT clone and the *BCL11A*$^{-/-}$ was consistently less differentiated by both morphology and flow cytometry (Figure 5.10). RNA expression and haemoglobin production were then assessed in three separate differentiations (Figure 5.11).

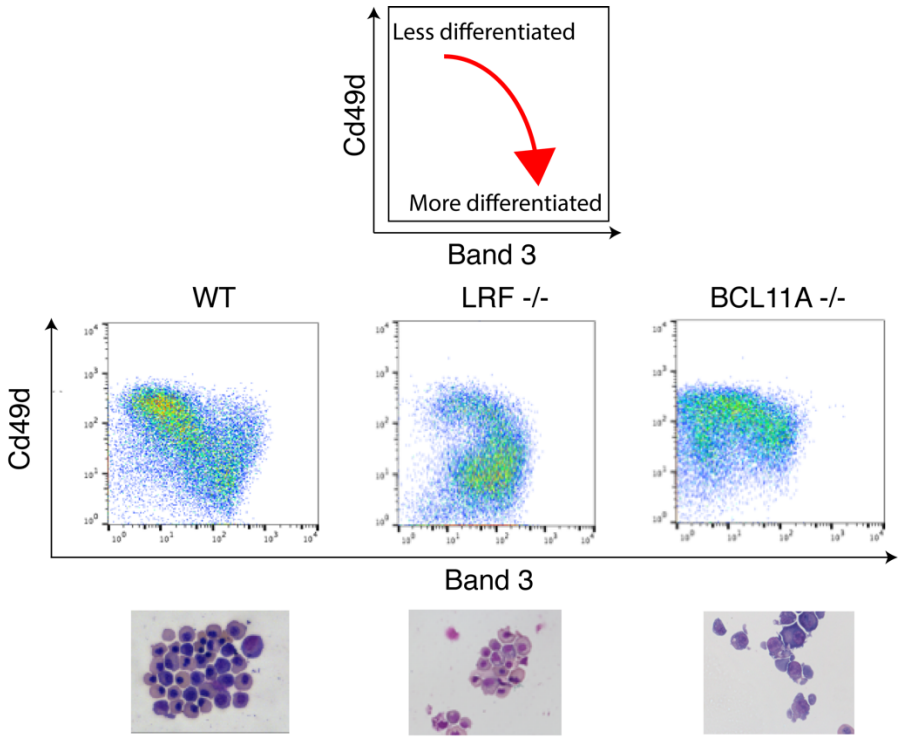


Figure 5.10 Day 10 differentiation of a. WT, *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ clone; representative flow cytometry plots and cytospins are shown.

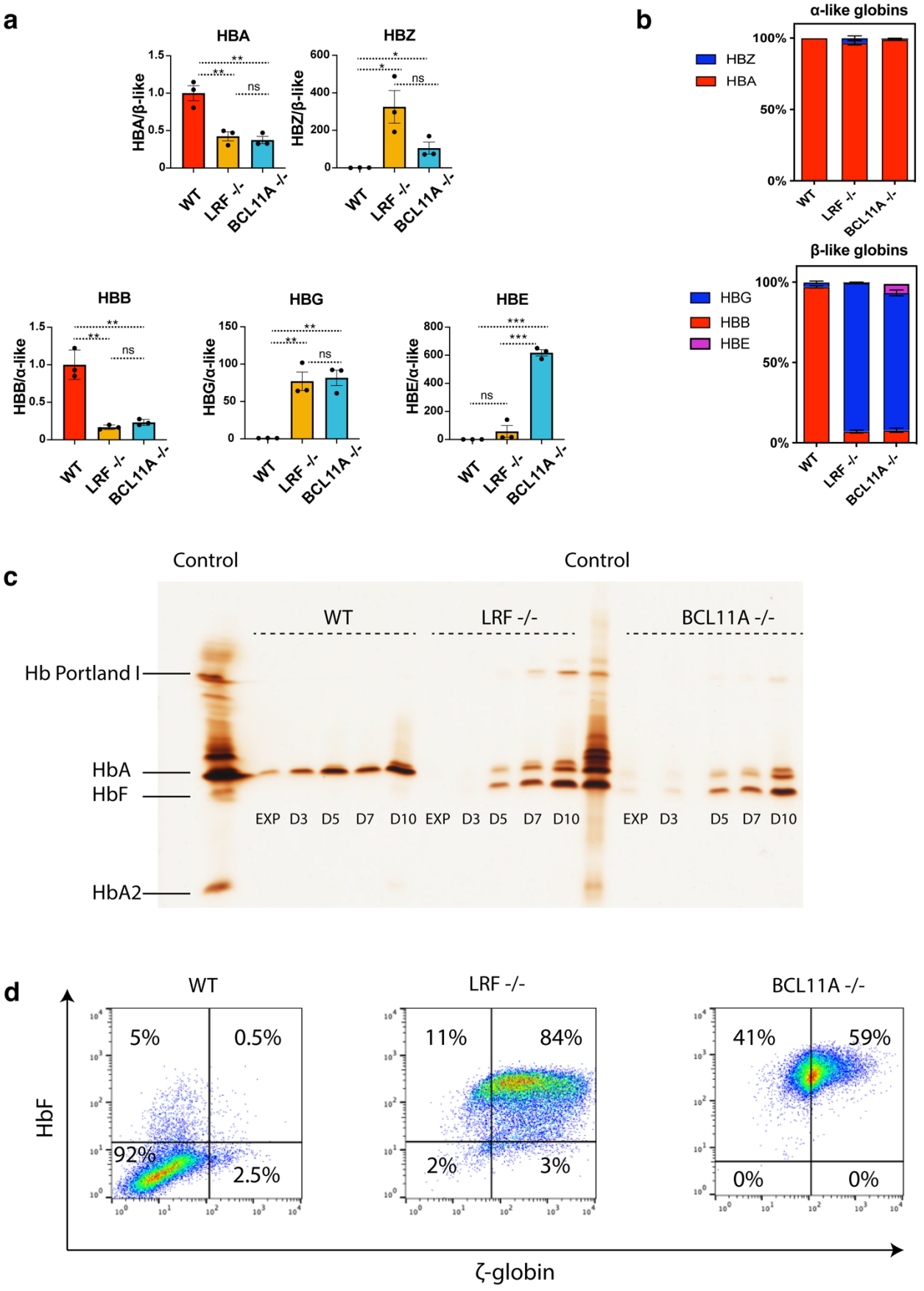

a
HBA
HBZ
HBA/β-like
HBZ/β-like
WT
LRF -/-
BCL11A -/-
HBB
HBG
HBE
HBB/α-like
HBG/α-like
HBE/α-like
ns
b
α-like globins
HBZ
HBA
β-like globins
HBG
HBB
HBE
100%
50%
0%
c
Control
Control
WT
LRF -/-
BCL11A -/-
Hb Portland I
HbA
HbF
HbA2
EXP D3 D5 D7 D10 EXP D3 D5 D7 D10 EXP D3 D5 D7 D10
d
WT
LRF -/-
BCL11A -/-
HbF
ζ-globin
5%
0.5%
92%
2.5%
11%
84%
2%
3%
41%
59%
0%
0%

Figure 5.11 (previous page) Expression and protein analysis globins in WT, *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ HUDEP-2 clones. a. Expression data. * denotes p<0.05, ** denotes p<0.01; *** denotes p<0.001. Significance was determined by a one-way ANOVA with Bonferroni correction for multiple comparisons. SEM and data points from each differentiation are shown. Data were normalized to β-like (for *HBA* and *HBZ*) and α-like (for *HBB*, *HBG* and *HBE*) globins to control for differences in differentiation, then presented as a proportion of WT. b. Bar graphs showing proportions of *HBZ* to total α-like globin and proportions of *HBG*, *HBB* and *HBE* to total β-like globin in WT, *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ HUDEP-2 cells. c. IEF of WT, *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ cells during differentiation. D. Flow cytometry using an antibody directed against HbF and ζ-globin. The percentage of positive cells in each quadrant is shown. Negatives were defined by isotype controls for each sample.

The effects of *LRF* and *BCL11A* knockout on the β-locus are striking. Knockout of either gene leads to a statistically significant increase in γ-globin expression and a fall in β-globin expression, such that the majority of the β-like globins produced in those cells is now γ-globin (Figure 5.11b). Knockout of *BCL11A* also leads to a significant increase in ε-globin expression (though the. total amount of ε-globin produced is small) with knockout of *LRF* leading to a small (though not statistically significant) rise. These effects are also seen when assessing protein production by either IEF or flow cytometry: the major haemoglobin produced in these cells is HbF, and almost all the cells are F positive in both KO models. These findings are in keeping with those previously published (Masuda et al., 2016).

In contrast, the effects on the α-globin locus are subtler. There is a fall in α-globin expression, and an increase in ζ-globin expression (with a greater increase when *LRF* is knocked out). The total amount of ζ-globin produced however remains small, accounting for ~3% of the total α-like globin in the *LRF*$^{-/-}$ cell line, and ~1% of the total globin in the *BCL11A*$^{-/-}$ cell line. This increase in ζ-globin can be seen using IEF, with Haemoglobin Portland I (ζ2γ2) becoming visible in both knockout cell lines, with a greater amount visible in the *LRF* knockout line. More of the *LRF*$^{-/-}$ cells also become ζ-globin positive when compared to the *BCL11A*$^{-/-}$ cells. In summary, both *LRF* and *BCL11A* knockout lead to an increase in ζ-globin expression and protein production, with *LRF* knockout having a greater effect than *BCL11A* knockout.

5.2.3 Knockout of *LRF* and *BCL11A* separately leads to marked changes in chromatin accessibility and histone acetylation at the β-globin locus with minimal changes seen at the α-globin locus

Next, the effect of knockout of *LRF* and *BCL11A* on chromatin accessibility and histone acetylation at the globin loci in HUDEP-2 cells was studied (Figures 5.12 and 5.13).

At the α-globin locus, there are minimal changes in comparison to the WT cells, with only a subtle increase in histone acetylation across the ζ-globin gene. The chromatin remains inaccessible by ATAC-seq.

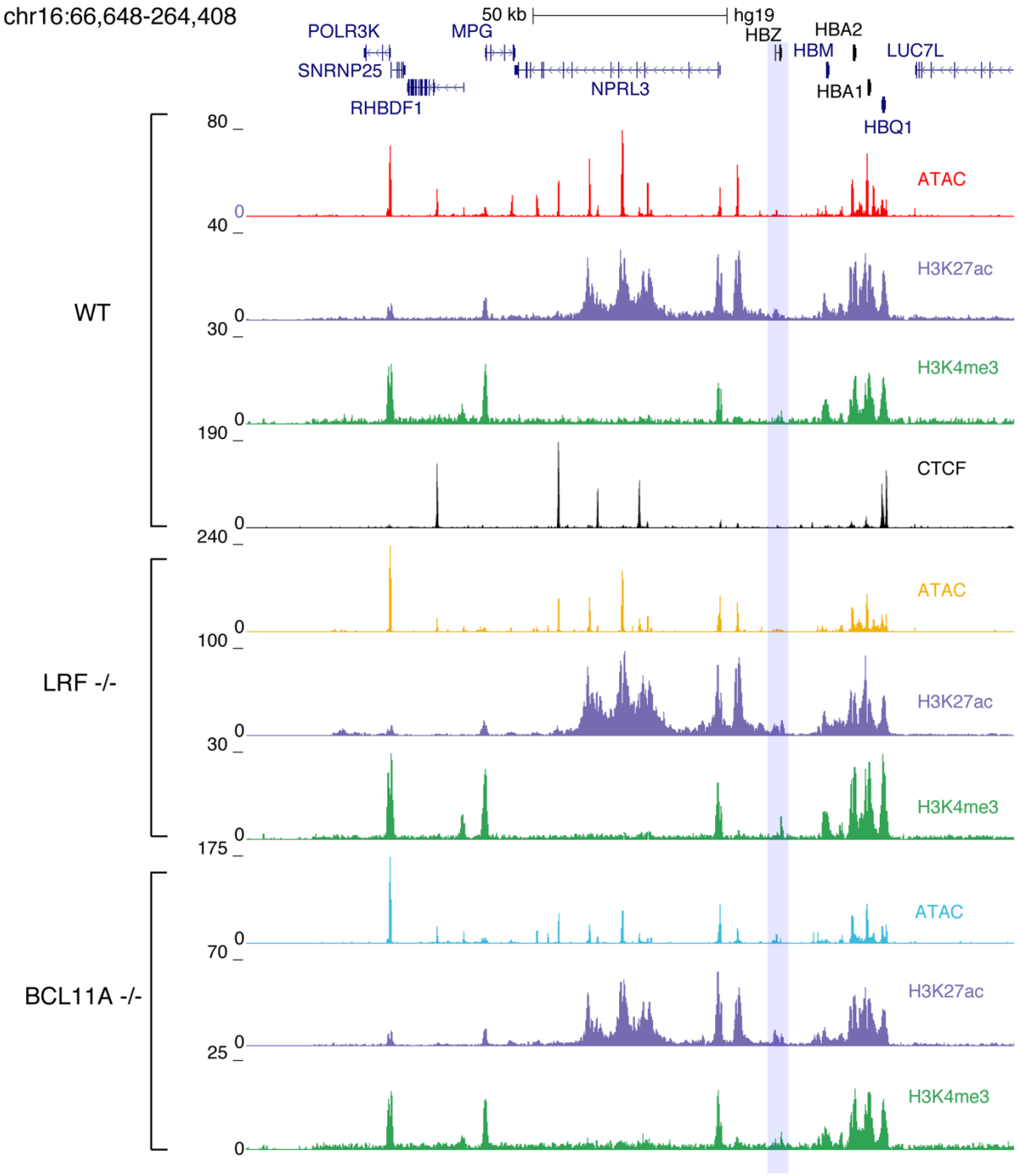


Figure 5.12 Chromatin accessibility and histone marks for: WT, *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ HUDEP-2 cells at the α-globin locus. ChIP-seq data consist of a single replicate performed on Day 10 of the HUDEP-2 culture system; ATAC-seq data are in triplicate. Data are normalized by RPKM. The ζ-globin gene is highlighted by a blue rectangle.

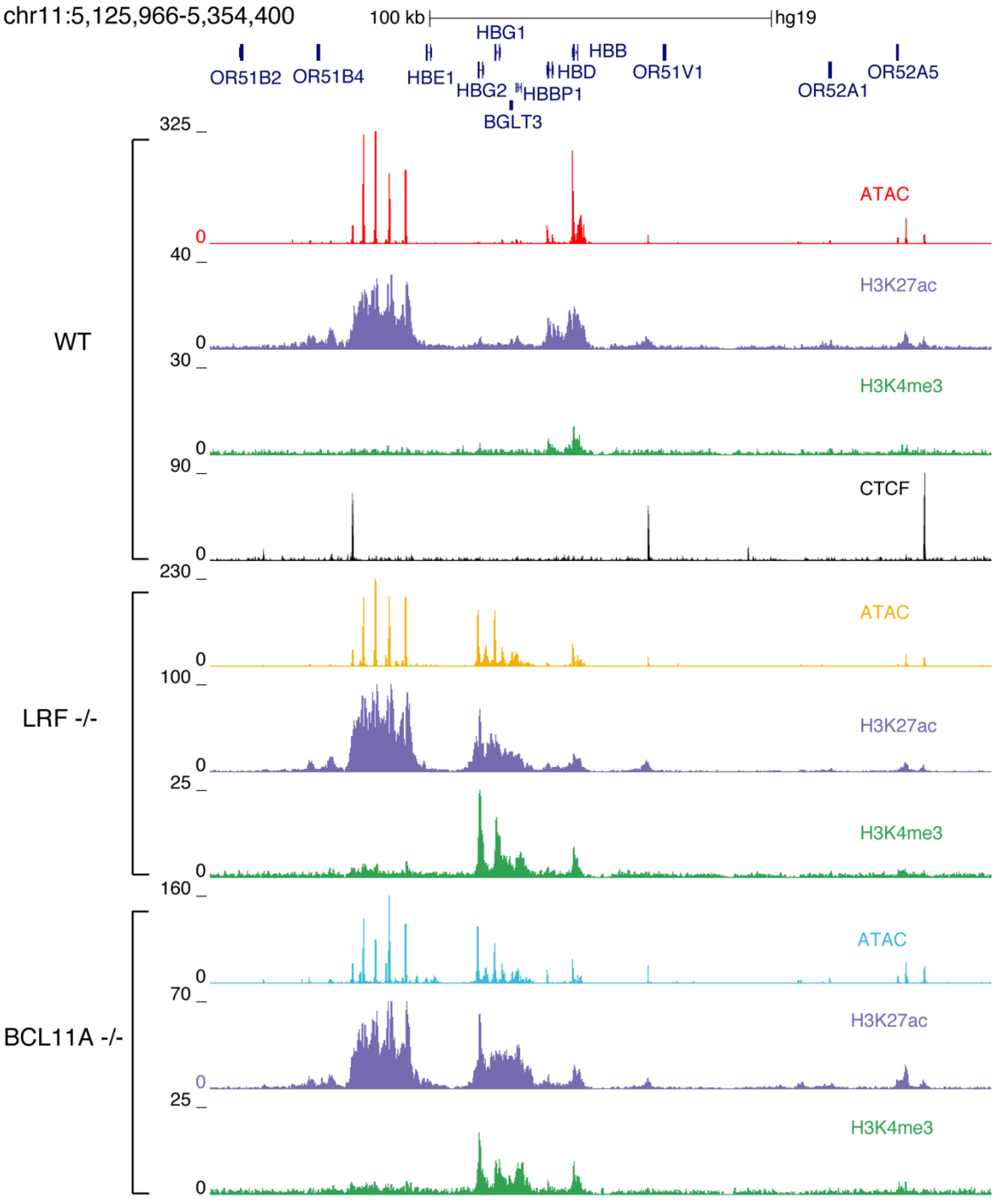


Figure 5.13 Chromatin accessibility and histone marks for: WT, *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ HUDEP-2 cells. ChIP-seq data consist of a single replicate performed on Day 10 of the HUDEP-2 culture system; ATAC-seq data are in triplicate. Data are normalized by RPKM.

In contrast, there are marked changes in chromatin accessibility and the histone marks at the β-globin locus. The paired γ-globin genes in both the *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ become accessible, with associated histone acetylation and acquisition of the H3K4me3 mark. In the *BCL11A*$^{-/-}$ knockout line, the ε-globin gene also has a small increase in chromatin accessibility and histone acetylation. These results extend and corroborate those previously reported (Masuda et al., 2016).

5.2.4 LRF and BCL11A bind directly to the ζ-globin gene

In order to determine whether LRF and BCL11A exert their effects by binding directly to the *HBZ* gene promoter, ChIP-seq and CUT&RUN data from three recent publications (Liu et al., 2018; Martyn et al., 2018; Masuda et al., 2016) were downloaded and re-analysed (Figure 5.14).

LRF ChIP-seq data shows binding of LRF across the α-globin locus, including the four MCS-R elements, the adult α-globin genes and the ζ-globin gene. The ChIP-seq data for BCL11A shows binding at the R2, R4 element, the ζ-globin gene and the adult α-globin genes; the CUT&RUN data shows binding at all hypersensitive elements and the ζ-globin gene.

The consensus site for BCL11A was identified as TGACC (Liu et al., 2018; Martyn et al., 2018) and that for LRF is GACCC (C/T) (Masuda et al., 2016); these are both found in the ζ-globin gene promoter (Table 5.2).

Taken together, these data support binding of both transcription factors directly to the ζ-globin promoter. It does also appear that there is binding at other locations within the locus and it seems likely that some of the signal in the LRF and CUT&RUN tracks may represent technical artefact rather than true binding.

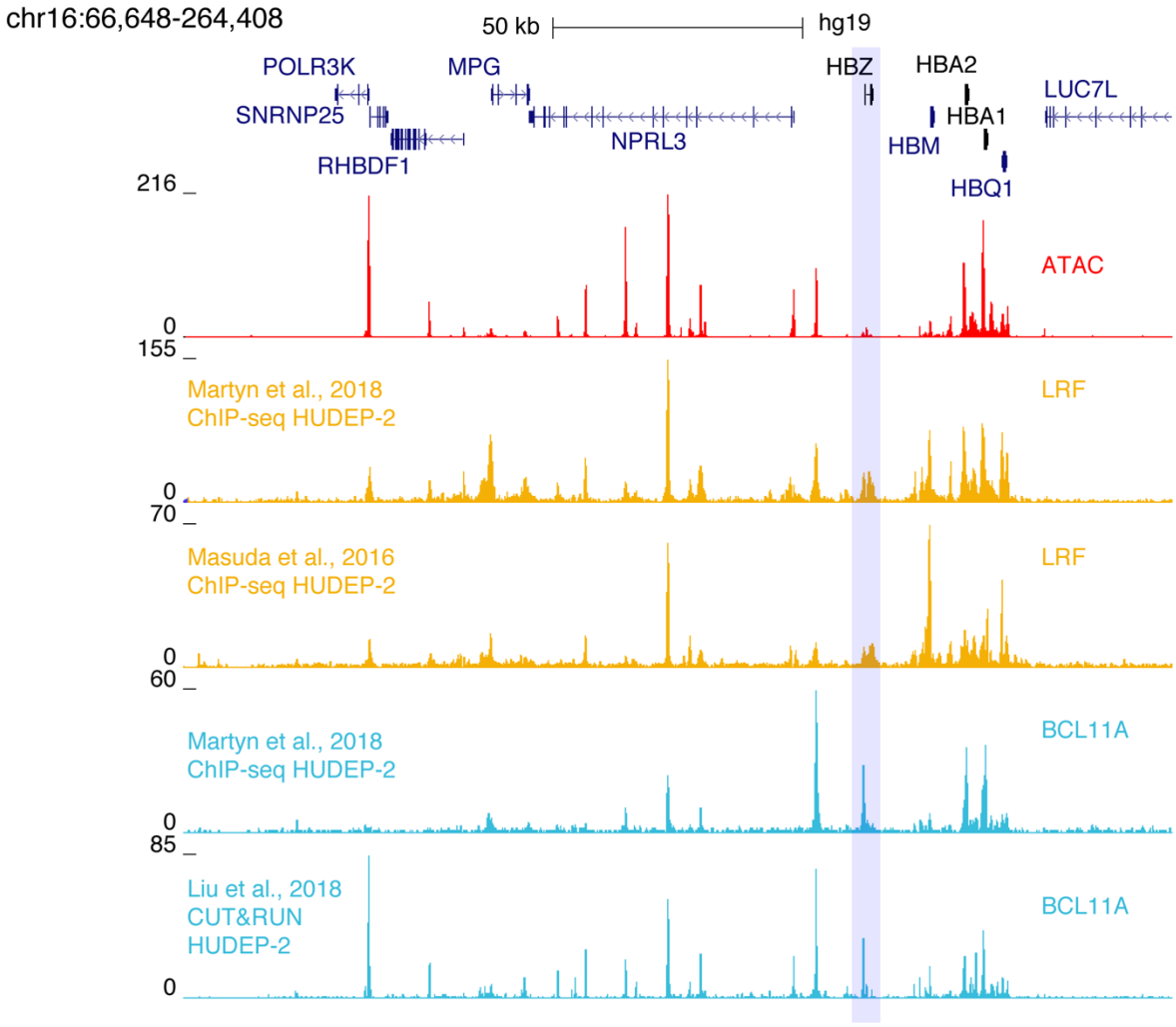


Figure 5.14 ChIP-seq and CUT&RUN data for LRF and BCL11A binding across the α-globin locus. The relevant publication, cell type and technique used are shown. An ATAC-seq track for HUDEP-2 cells is also shown.

Transcription factor	Sequence	Position relative to TSS
BCL11A	TGT**GGTCA**GACTCTG	-211 *
BCL11A	CC**TGACC**AATGGCCA	-66
LRF	TAAGGG**GACCCT**GGG	-16

Table 5.2 Consensus sites for BCL11A and LRF in the human ζ-globin promoter. The distance from the *HBZ* transcriptional start site (TSS) is also shown. The consensus sites are shown in bold and highlighted. * denotes reverse strand.

5.2.5 Knockout of *BCL11A* and *LRF* together leads to greater embryonic and fetal globin de-repression than knockout of each individual transcription factor in HUDEP-2 cells

To determine whether *BCL11A* and *LRF* knockout together would lead to a greater degree of ζ-globin de-repression than either transcription factor alone, a cell line was imported where both had been knocked out, confirmed by western blot shown in the publication (Masuda et al., 2016). In comparison to WT HUDEP-2 cells, differentiation was consistently impaired in the $LRF^{-/-}$/$BCL11A^{-/-}$ (DKO) clone (Figure 5.15) by Day 10.

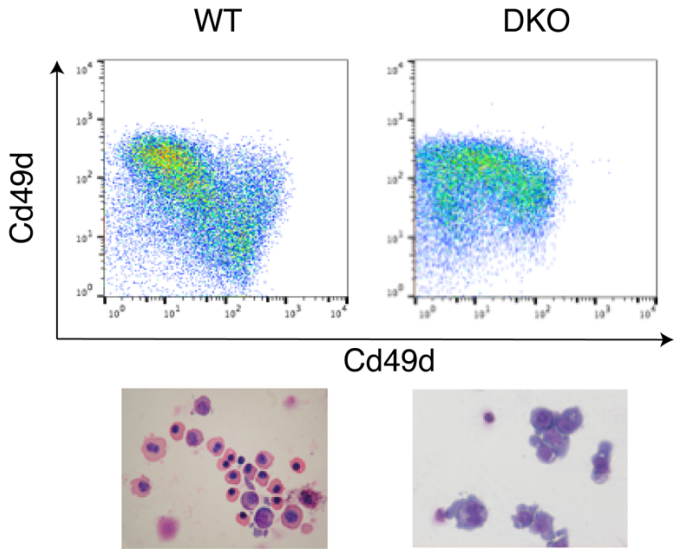


Figure 5.15 Day 10 differentiation of a WT and double *LRF/BCL11A* knockout clone; representative flow cytometry plots and cytospins are shown.

RNA expression and haemoglobin production was then assessed (Figure 5.16). Knockout of both *BCL11A* and *LRF* leads to a greater degree of de-repression of ζ-globin than knockout of each individual transcription factor, with ~15% of the total α-like globin being ζ-globin in the DKO clone. This is also visible using IEF, with a greater degree of haemoglobin being Hb Portland I (ζ2γ2) than in the *LRF*$^{-/-}$ and *BCL11A*$^{-/-}$ knockout clones. At the β-globin locus, almost all of the β-like globin is γ-globin and ε-globin.

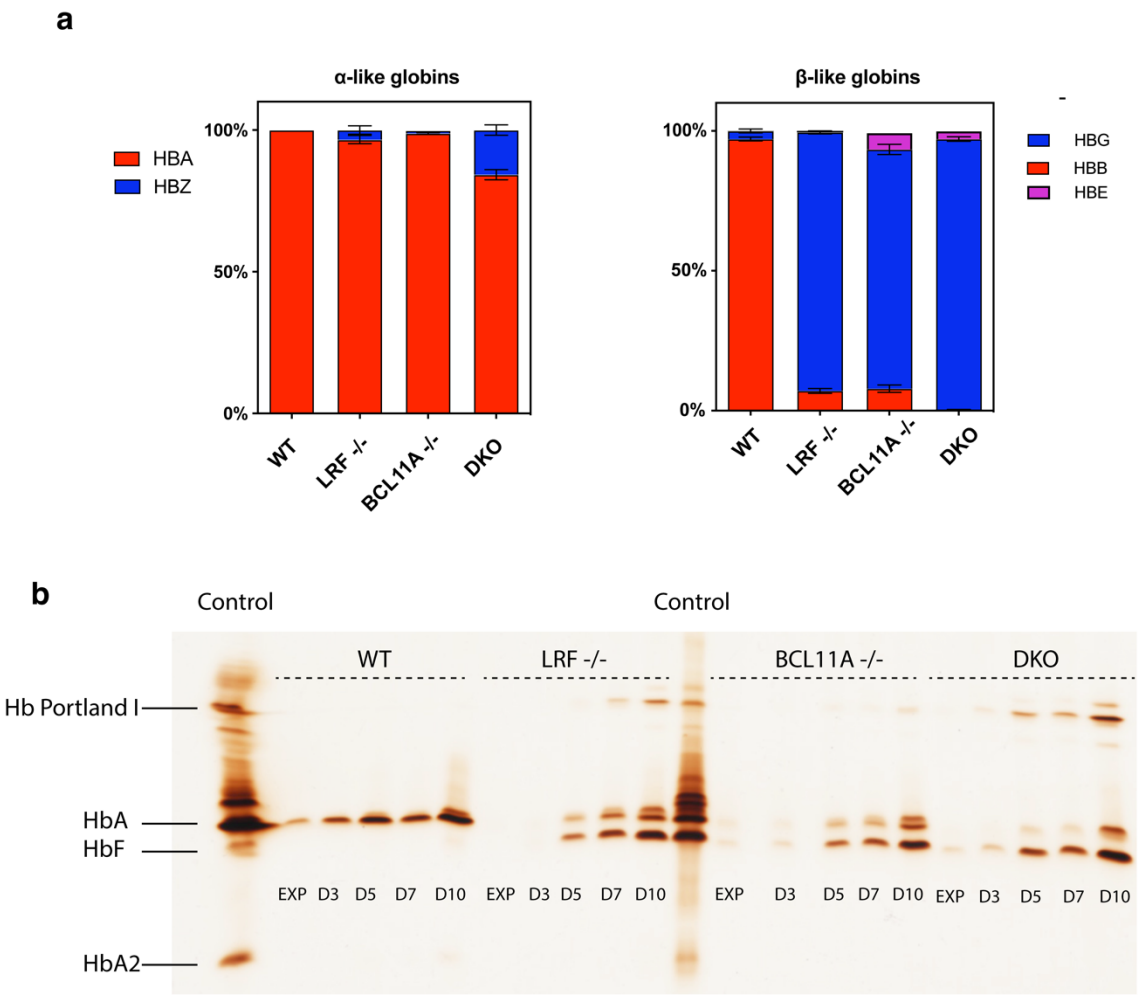


Figure 5.16 Expression and haemoglobin production from the LRF/BCL11A double knockout clone. **a.** Bar graphs showing proportions of *HBZ* expression to total α-like globin and proportions of *HBG*, *HBB* and *HBE* to total β-like globin in WT, *LRF*$^{-/-}$, *BCL11A*$^{-/-}$ and DKO HUDEP-2 cells. Three separate differentiations were performed, and the mean and SD are shown. **b.** IEF of WT, *LRF*$^{-/-}$, *BCL11A*$^{-/-}$ and DKO cells during differentiation.

5.2.6 Knockout of BCL11A and LRF together leads to an increase in chromatin accessibility over the ζ-globin gene

To assess chromatin accessibility, ATAC-seq was performed on DKO cells (Figure 5.17).

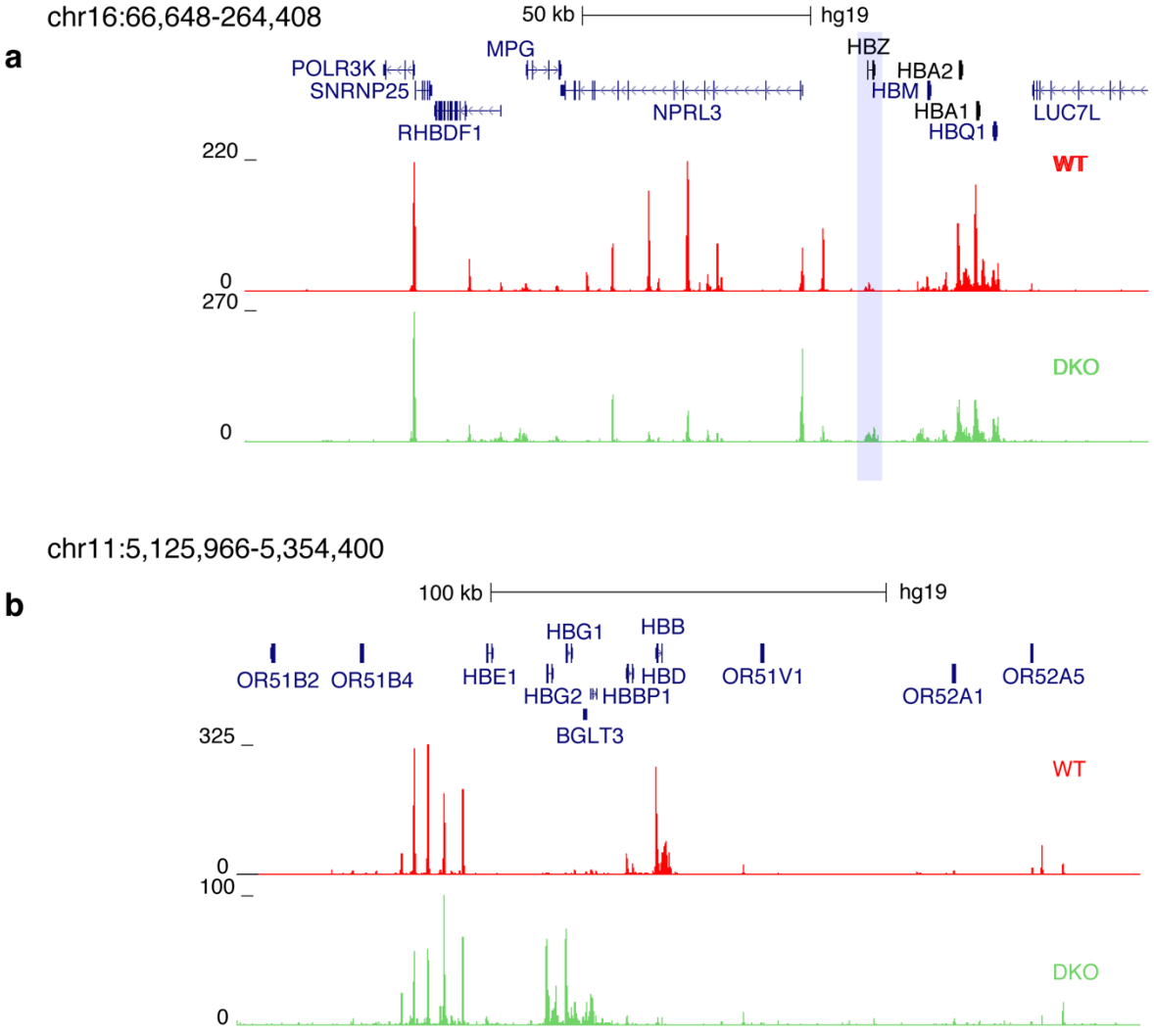


Figure 5.17 a. chromatin accessibility, assessed by ATAC-seq at the a α-globin locus and b. β-globin locus in WT and DKO cells. The ζ-globin gene is highlighted in blue.

Over the ζ-globin gene, chromatin accessibility is mildly increased in the DKO cells in comparison to WT (particularly when directly compared to the accessibility over the α-globin genes). The reduced accessibility over the enhancers and α-globin genes in the DKO clone probably reflects that the cells are less differentiated. However, accessibility over the ζ-globin gene does not increase to the level seen over the α-globin genes (as it is in primitive erythroid cells), suggesting that other transcription factors are responsible for repressing ζ-globin, in addition to LRF and BCL11A.

At the β-globin locus, the chromatin accessibility over the adult globin genes becomes inaccessible in the DKO cells and is markedly accessible over the γ-globin genes. It therefore appears that LRF and BCL11A are responsible for the entirety of γ-globin repression with their removal establishing a fetal-like globin expression profile.

5.2.7 DNA methylation is not necessary for repression of ζ-globin and methylation of the CpG island located at the 3' end becomes methylated as transcription increases

In order to understand whether DNA methylation plays a role in ζ-globin repression, focused bisulfite sequencing was performed across the ζ-globin gene in primary erythroid cells derived from $CD34^+$ cells and each of the HUDEP-2 cell lines (Figure 5.18).

In primary erythroid cells, when *HBZ* is silenced, four CpGs are methylated (>50% methylation) and two appear unmethylated (<50% methylation). The CpG island located in the 3' region of the gene are generally unmethylated. In HUDEP-2 cells, all the CpG islands have <50% methylation including the WT HUDEP-2 cells when ζ-globin expression is trivial. The number of reads across the *LRF*$^{-/-}$ promoter was very low for technical reasons (due to inadequate DNA being isolated), so cannot be interpreted. The CpGs within the 3' region of the gene become increasingly methylated from WT through to the DKO cells as transcription through the gene increases. Based upon the observation that ζ-globin transcription is low in the WT HUDEP-2 cells despite the CpGs in the promoter being unmethylated, it appears that methylation of the CpGs within the ζ-globin gene promoter does not play an important role in gene silencing.

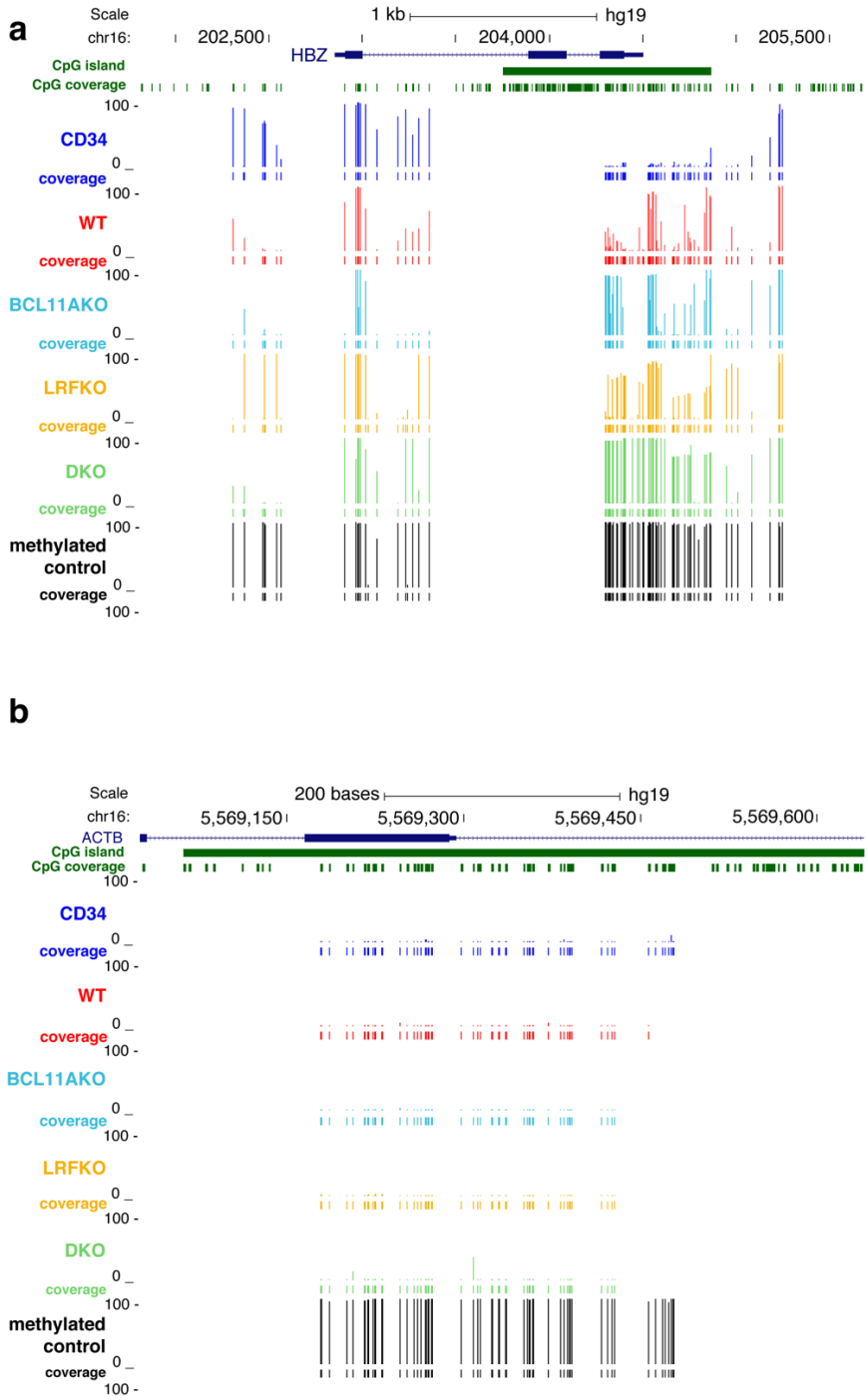


Figure 5.18 a. DNA Methylation across the *HBZ* gene and b. the *ACTB* gene in primary erythroid cells derived from a CD34+ cell culture system at Day 13 (CD34) and WT, *LRF*$^{-/-}$, *BCL11A*$^{-/-}$ and DKO HUDEP-2 cells at Day 10 of differentiation. The CpG islands, CpGs, gene bodies and a methylation control using methylated DNA are also shown. The *ACTB* gene is a transcribed house-keeper gene, and therefore serves as a negative control.

5.2.8 The ζ-globin gene is not marked by H3K9me3 or H3K27me3 in primary definitive erythroid cells and inhibition of EHMT1/EHMT2 does not lead to ζ-globin de-repression.

Finally, previously published data were analysed to determine whether repressive chromatin marks are associated with ζ-globin repression in definitive human erythroid cells. ChIP-seq data for H3K9me3 and H3K27me3 (Xu et al., 2012) were downloaded, re-analysed and aligned to the genome browser. Data from cells differentiated to the erythroblast stage from fetal and adult CD34^{+} cells are shown. Both H3K9me3 and H3K27me3 are not enriched at the ζ-globin gene (Figure 5.19).

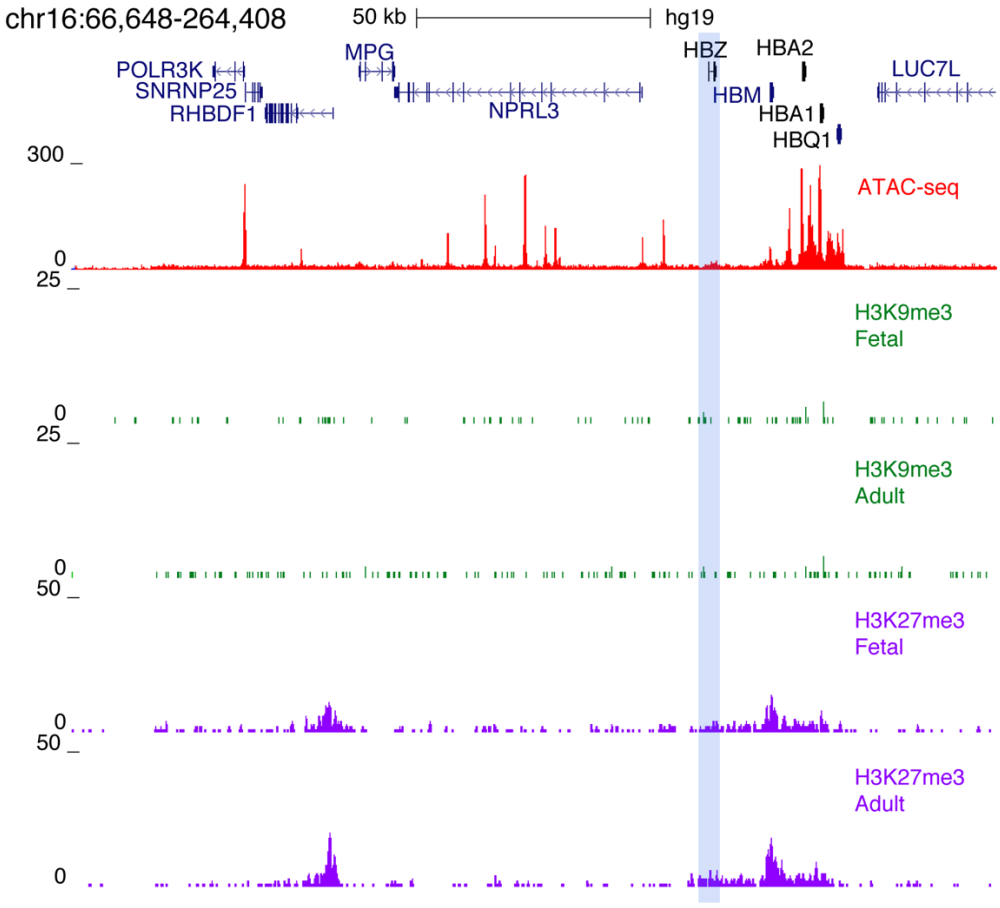


Figure 5.19 Normalised ChIP-seq data for H3K9me3 and H3K27me3 in fetal and adult erythroblasts. An ATAC-seq track is also shown. The ζ-globin gene is highlighted in blue. Data were single replicates.

To determine whether H3K9me2 is involved in ζ-globin repression, RNA-seq data from primary erythroid cells cultured with a specific inhibitor of EHMT1/EHMT2 (Renneville et al., 2015) for seven days were downloaded, analysed and visualised. EHMT1/2 catalyses mono and dimethylation of H3K9, and their inhibition leads to γ-globin derepression (Renneville et al., 2015). Visualisation of the ζ-globin gene shows no increase in expression following culture with the inhibitor in contrast to the γ-globin gene (Figure 5.20).

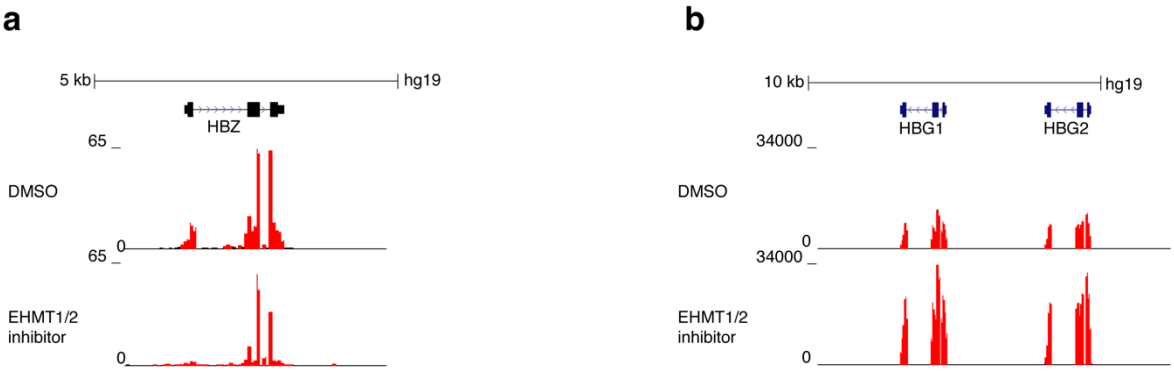


Figure 5.20 RNA-seq data from primary erythroid cells derived from a CD34+ culture system following addition of a vehicle control (DMSO; top) and a specific inhibitor of EHMT1/2 (UNC0638). a. The ζ-globin gene b. The γ-globin genes. Data are normalised (to RPKM) and combined from three replicates.

Together, these data imply that H3K9me2, H3K9me3 and H3K27me3 do not play a significant role in repression of ζ-globin.

5.3 Discussion

Here, it is shown that the HUDEP-2 cell line serves as a tractable model for human definitive erythropoiesis. WT HUDEP-2 cells following differentiation express trivial amounts of fetal and embryonic globins. Their surface markers change in a similar fashion to erythroid cells derived from a $CD34^+$ culture system as they differentiate (Hu et al., 2013). Chromatin accessibility, histone acetylation, CTCF binding and H3K4me3 marks are all very similar between HUDEP-2 cells and primary erythroid cells at the globin loci. Previously, it has been demonstrated that bulk RNA-seq data from differentiated HUDEP-2 cells shows a transcriptome profile similar to primary basophilic erythroblasts (Masuda et al., 2016) further supporting their use as a model for definitive human erythropoiesis.

However, there are issues with HUDEP-2 cells. Like many cell lines, they have an abnormal karyotype, characterised here in detail. In our laboratory we have also noted clonal variability and phenotypic “drift”, with some cells grown from a single clone having unexpectedly high levels of fetal haemoglobin or changes in differentiation without genetic manipulation. Ideally, multiple clones with the same genetic modification should be generated and compared to WT HUDEP-2 cells isolated in parallel. When loss of function is performed through gene editing (or other approaches) a functional rescue via “reverse” genome engineering or reintroduction of the gene via transfection would significantly strengthen conclusions, if it leads to a similar reversal in phenotype. The data presented here are from one clone only (although with multiple separate differentiations) which haven’t been rescued; this is clearly a potential criticism of the work.

That being said, the results shown here following knockout of *LRF* and *BCL11A* are entirely consistent with the *in vivo* data generated in mice in the previous chapter. *LRF and BCL11A* knockout separately lead to a small amount of ζ-globin derepression, with more ζ-globin being produced following knockout of *LRF* than *BCL11A.* At the β-globin locus, knockout of *LRF* leads to marked γ-globin derepression (the orthologous gene to *Hbb-bh1* in the mouse), whereas it has almost no effect on ε-globin expression. Knockout of *BCL11A* leads to both γ and ε-globin expression. These findings at the β-globin locus are also consistent with those previously published (Masuda et al., 2016). It is interesting to note the minimal effects of *LRF* knockout on ε-globin expression, in a similar manner to there being no effect on *Hbb-y* expression in the mouse. It thus appears that despite ε-globin and ζ-globin having a similar developmental pattern of expression, their silencing appears to be mediated via different mechanisms.

The effects on the chromatin landscape at the α-globin locus are minor following *LRF* and *BCL11A* knockout, with only a subtle increase in histone acetylation over the ζ-globin gene. In contrast, there are marked changes in chromatin accessibility at the β-globin locus, with a decrease in chromatin accessibility over the definitive globin genes, and a marked increase over γ-globin. There is also a slight increase in chromatin accessibility over the ε-globin gene in *BCL11A* knockout cells. This is accompanied by histone acetylation and acquisition of the H3K4me3 mark. These findings complement, and extend those already published (Masuda et al., 2016).

Knockout of *BCL11A* and *LRF* together leads to a full switch to γ-globin in HUDEP-2 cells. However, only ~15% of the α-like globin is ζ-globin after knockout of both transcription factors. This is supported by the ATAC-seq experiments, which show a small increase in chromatin accessibility over the ζ-globin gene in the DKO clone, but not to the level of that seen over the adult α-globin gene. Clearly then, there are additional repressive transcription factors aside from LRF and BCL11A responsible for silencing ζ-globin.

Exactly how BCL11A and LRF mediate gene silencing remains unclear. The prevailing view that BCL11A mediates γ-globin silencing by binding to intergenic regions (Xu et al., 2010) was disproved recently by two papers which showed direct binding to the γ-globin gene promoter (Liu et al., 2018; Martyn et al., 2018) using ChIP-seq (in a HUDEP-2 line where BCL11A had a ER-V5-tag) and CUT&RUN. Recent papers have highlighted a key role for NuRD which bind to both LRF and BCL11A (Masuda et al., 2016; Sher et al., 2019). Re-analysis of those genome wide data and ChIP-seq for LRF (Martyn et al., 2018; Masuda et al., 2016) also show direct binding to the ζ-globin promoter. There is however, apparent binding throughout the locus at both the enhancers and the definitive α-globin genes in all data sets; the LRF ChIP and BCL11A CUT&RUN also shows binding at other open chromatin regions. It may well be that some of these signals are erroneous; similar widespread binding for both transcription factors is also seen at the β-globin locus and both proteins are difficult to ChIP.

The consensus sites identified in those ChIP-seq and CUT&RUN data sets (Liu et al., 2018; Martyn et al., 2018) can both be found in the ζ-globin promoter. Of

particular interest is the TGACC site adjacent to a CCAAT box; when this sequence has point mutations in humans, it leads to HPFH. Genome editing to produce a similar mutation leads to γ-globin derepression *in vitro* (Liu et al., 2018; Martyn et al., 2018; Traxler et al., 2016). We are currently attempting to mutate this site in the ζ-globin promoter in both primary erythroid and HUDEP-2 cells in the laboratory.

In order to see whether other epigenetic mechanisms played a role in ζ-globin silencing, DNA methylation was examined in primary erythroid cells derived from $CD34^+$ cells, and the four HUDEP-2 cell lines. There are only 6 CpGs in the ζ-globin promoter, although there is an intragenic CpG island encompassing the second and third exon and 3'UTR, the functional relevance of which was unknown. In primary erythroid cells derived from adult $CD34^+$ cells, four CpGs are methylated and two are unmethylated, whereas in HUDEP-2 cells (which still express trivial amounts of ζ-globin), all but one is unmethylated. As to why these are different is unclear and may relate to aberrant methylation due to HUDEP-2 immortalisation, or that HUDEP-2 cells are derived from fetal CD34+ cells rather than adult. However, given ζ-globin is still effectively silenced in WT HUDEP-2 cells, DNA methylation at the promoter does not appear to play a silencing role.

As ζ-globin transcription increases the gene body CpG island becomes methylated: in the DKO cells almost all the CpGs within the island are methylated in comparison to the WT HUDEP-2 and primary erythroid cells where most aren't. Intragenic CpG island methylation is known to correlate with gene transcription in *cis* (Jeziorska et al., 2017). Given the ζ-globin gene's intragenic CpG island's methylation correlates with transcription in agreement with this, it clearly isn't important in its silencing. It

also suggests that ζ-globin is never transcribed in adult $CD34^+$ cells as they differentiate to erythroid cells *in vitro,* as once a gene has been transcribed, methylation in intragenic CpG islands is faithfully maintained through cell division (Jeziorska et al., 2017).

Having excluded DNA methylation as a silencing mechanism, the role of other histone modifications in ζ-globin silencing were examined by mining previous data-sets published in primary definitive erythroid cells. H3K9me3 is a repressive histone mark, normally associated with pericentric and facultative chromatin (Becker et al., 2016)) and H3K27me3 is laid down by PRC2. Neither of these marks are present in definitive erythroid cells at the ζ-globin gene based upon re-analysis of previously published data, and this agrees with a previous publication using ChIP-on-chip (Garrick et al., 2008). Following culture of primary erythroid cells with an inhibitor against EHMT1/2 (which is responsible for catalysing the mono and dimethylation of H3K9) ζ-globin expression is not increased. It therefore seems that H3K9me1 and H3K9me2 are not required for ζ-globin silencing, in contrast to γ-globin.

In conclusion, ζ-globin silencing in definitive erythroid cells is reliant on histone deacetylation and chromatin remodelling, and this appears to be maintained by transcription factors (including LRF and BCL11A), with their perturbation leading to gene de-repression. Other epigenetic silencing mechanisms do not appear to play a role. Knockout of both LRF and BCL11A together does not lead to restoration of ζ-globin to levels found in primitive erythroid cells, so other transcription factors must also contribute. The final chapter in this thesis describes development of a novel

mouse model to facilitate screening for these transcription factors and potential silencing complexes.

6 Generation and characterisation of a ζ-globin reporter mouse

6.1 Introduction

The work presented so far has demonstrated that two transcription factors, BCL11A and LRF play an important role in the silencing of ζ-globin in fetal and adult life. It is almost certain however, that additional transcription factors and complexes are involved in activating and silencing ζ-globin based upon the work presented so far in this thesis. Work by a previous student in the laboratory has generated a candidate list of transcription factors that are differentially expressed (DE) between primitive and definitive erythropoiesis in mouse using RNA-seq (Duantida Songdej, unpublished); similar data-sets have also been published using microarray technology (Kingsley et al., 2013). Among these DE genes there are a number of transcription factors that are specifically upregulated in primitive or definitive erythropoiesis and as such are candidates for positive or negative regulators of zeta-globin transcription.

Generation of a mouse with ζ-globin partially replaced and tagged by a stable fluorescent protein would serve several functions. Firstly, it would facilitate high-throughput screening of candidate transcription factors in an unbiased fashion using primary erythroid cells, as well as allowing screening for *cis*-acting sequences responsible for repression. It would also permit high throughput drug screening for

compounds capable of de-repressing ζ-globin. Thirdly, it would allow a screen of mouse tissue to determine whether ζ-globin is expressed in cells other than erythroid cells, as has been reported for α-globin (Straub et al., 2012). Finally, it would allow the importance of ζ-globin during mammalian development to be determined, as ζ-globin can be rendered non-functional by the design of the reporter.

Multiple mouse lines have been generated where the β-like globin genes express a fluorescent reporter. Two methods have been commonly used. The first involves insertion of transgenes via random genomic integration, typically with an embryonic or fetal globin promoter, a truncated erythroid specific enhancer, and a fluorescent protein such as green fluorescent protein (GFP) (Fraser et al., 2007; Vacaru et al., 2013). Whilst the generation of these reporters is straight forward and rapid, each line must be carefully characterised to ensure that the transgene is expressed at sufficiently high levels and in the expected spatiotemporal pattern. This is because the transgene may be driven ectopically by neighbouring genomic sequences to the integration site, may have a variable copy number, and may be inappropriately silenced by the surrounding chromatin environment (Garrick et al., 1998; Fraser et al., 2010). Transgenic mouse lines are normally generated by microinjection of DNA into the male pronucleus of a fertilized egg; hence such lines can be rapidly established.

The second approach utilised a bacterial artificial chromosome (BAC) which contains the entire globin locus including its regulatory elements (Papadopoulos et al., 2012). This then allows the genes to be driven by their endogenous regulatory

elements, providing a more accurate expression pattern of the fluorescent protein. The use of BACs is still prone to the problems with transgene insertions however (copy number differences, complex mapping problems and position effects). Conventional homologous recombination techniques can be utilised with BACs to modify endogenous genes in mouse ES cells (Hay et al., 2016) which can then be injected into blastocysts to generate chimeric mice. However, such approaches are time-consuming, require considerable laboratory expertise and typically have a low efficiency.

Recent advances in gene editing technology mean that manipulation of DNA sequences in their natural chromosome environment has become more tractable. In particular, the use of the CRISPR/Cas9 nuclease has become routine, allowing generation of novel cell lines and organisms rapidly and with high efficiency (Hsu et al., 2014). This technology has been used to generate fluorescent reporters via both direct zygote injection and ES cell manipulation (Yang et al., 2013) with high efficiency, low off-target effects and scarless integration of the fluorescent reporter sequence into the correct locus. The high success rate of CRISPR/Cas9 associated with its ease and ability to allow targeting of endogenous loci, make it an attractive choice for targeting a fluorescent reporter to the ζ-globin gene.

The correct choice of fluorescent protein is critical. The range of fluorescent proteins is extensive (Cranfill et al., 2016; Shaner et al., 2005), with novel fluorescent proteins with differing properties and emission spectra being published regularly. Haemoglobin absorbs (quenches) fluorescent light meaning that in red cells, fluorescent proteins which have particular emission spectra should be avoided

(Figure 6.1). Based upon this, the most sensible choice is one of the far-red fluorescent proteins such as mCardinal (Chu et al., 2014). However, to detect this requires specialized infrared microscopy and flow cytometry filters, which adds additional expense and complexity to experiments. YFP has an emission spectrum which sits around 540nm; this has the lowest absorption by haemoglobin of the routinely visible wavelengths. It is also easily detectable using standard microscopy and flow cytometry. Consequently, YFP was selected as the fluorescent protein for this project.

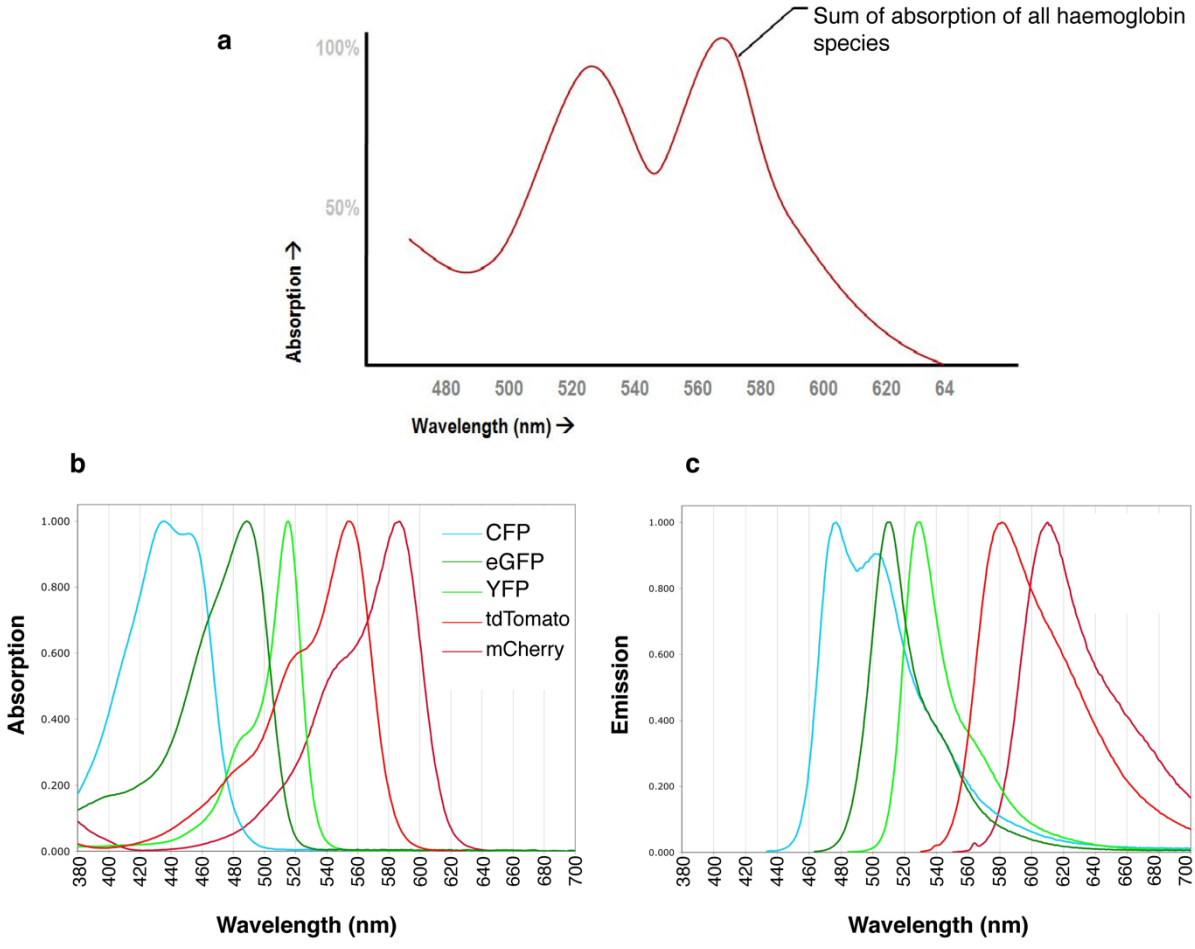


Figure 6.1 a. Absorption spectra of haemoglobin. b. Excitation spectra of commonly used fluorescent proteins. c. Emission spectra of commonly used fluorescent proteins. Adapted from (Chu et al., 2014; Shaner et al., 2005).

To minimize quenching from haemoglobin, it is possible to attach a localizing signal to the fluorescent protein which then moves it away from the cytoplasm. Typically, this would either be an amino-acid sequence that tags the protein for import into the nucleus (a nuclear localizing signal; NLS) (Fraser et al., 2010), or a protein which is inserted on the cell surface (for example glycophorin A) (Papadopoulos et al., 2012). An NLS was selected here, explained in more detail in the results section.

It remains uncertain whether ζ-globin is required for normal mammalian development. No mutations have been reported in humans (Fei et al., 1989; Felice et al., 1986; Winichagoon et al., 1982) although deletions do occur heterozygously (in 0.4-2.8% of populations studied). In certain mouse models where the ζ-globin gene has been homozygously deleted (with a neomycin selection gene left in place), some mice are viable dependent upon strain background (Leder et al., 1997). This implies that, at least with some genetic backgrounds, ζ-globin is not essential for life.

In this chapter, I describe the production and initial validation of a ζ-globin reporter mouse which will be used for both genome wide CRISPR knockout screening and drug screening. As part of the design, exon 3 of the ζ-globin gene was deleted. This then allowed the necessity of ζ-globin for viability to be studied.

6.2 Results

6.2.1 ζ-globin reporter design

Based upon the excitation and emission spectra of the different fluorescent proteins and the ease of their detectability using routine laboratory equipment, the mVenus variant of YFP was deemed the most appropriate fluorescent protein to use. YFP was initially modified in 2002 via a mutation to increase its brightness and stability (Venus); a mutation was then introduced to make it monomeric (mVenus) (Gert-Jan Kremers et al., 2006; Nagai et al., 2002).

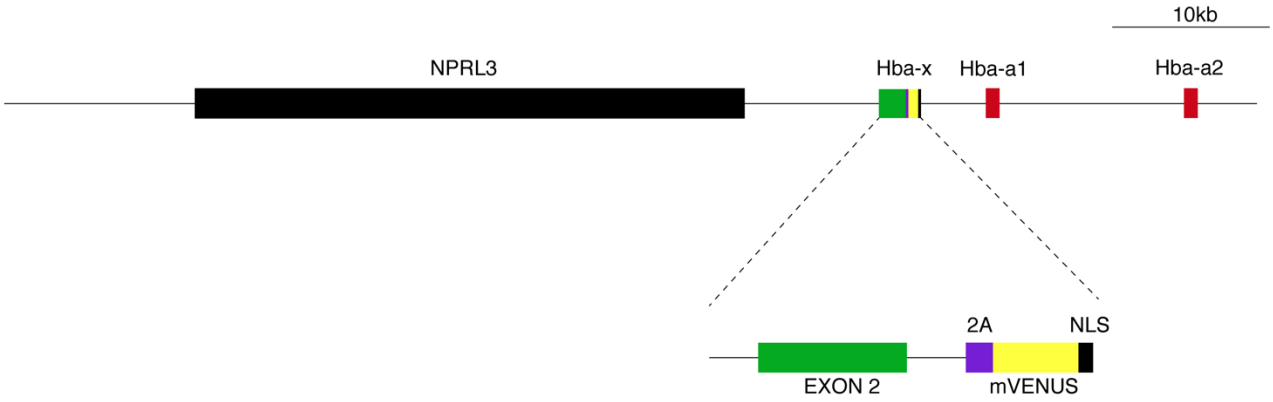


Figure 6.2 Design of the ζ-globin reporter. The coding globin genes are shown, with the modifications made to the ζ-globin and the addition of the reporter. NLS, nuclear localising signal.

To minimise fluorescent quenching from the haemoglobin molecule, two steps were taken. Firstly, a 2A peptide from the foot and mouth virus was introduced to allow cleavage of the mVenus from the truncated ζ-globin moiety in case this had a

quenching effect (Ryan et al., 1991). Secondly, a monopartite SV40 NLS was introduced at the C terminal end of the mVenus protein to move the fluorescent protein to the nucleus (Lange et al., 2007). Although red cells enucleate, which could in theory lead to ejection of the fluorescent signal primitive erythroid cells remain nucleated, at least until E12.5 (Kingsley et al., 2004). The *in vitro* culture systems where a CRISPR/cas9 screen or drug screen would be performed also generate nucleated definitive erythroid cells (Zhao et al., 2014). Using glycophorin A was considered, but this is 168 amino-acids, which would consist of 504bp of DNA; this may therefore have interfered with contact between the α-globin promoters and enhancers and/or fluorescent protein function.

In order to generate a loss of function ζ-globin allele, exon 3 of the *Hba-x* gene was deleted. The design of the reporter used is shown in Figure 6.2.

6.2.2 Targeting Strategy

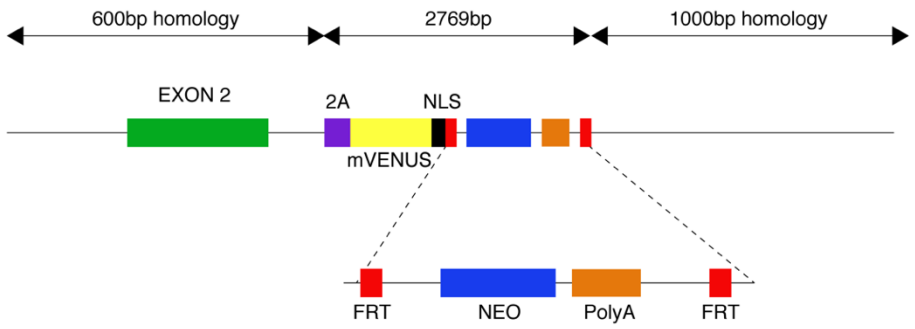
Mouse ES cells were targeted according to the strategy illustrated in Figure 6.3. Targeting mouse ES cells allowed validation of the reporter protein prior to mouse generation via blastocyst injection. A donor template was used in combination with a DSB mediated via CRISPR/cas9; this would then incorporate the desired sequence via homology directed repair.

In order to increase targeting efficiency, a neomycin cassette was included in the construct design to allow selection at the level of integration (Limaye et al., 2009). This was flanked by *FRT* sites to allow later removal via Flp mediated recombination

(Bouabe and Okkenhaug, 2013). Homology arms of 600bp and 1000bp were included; a shorter 5' arm was used to avoid repetitive elements in the first intron of *Hba-x*. Generally, at least 500-1000bp of homology are recommended to increase HDR efficiency (Oji et al., 2016; Yang et al., 2013).

Exon 3 of the *Hba-x* gene was cut with a WT cas9 delivered on the PX459 plasmid (Ran et al., 2013) in conjunction with the guide sequence shown. This guide RNA was validated using the Surveyor™ assay (Qiu et al., 2018) prior to use to confirm robust cutting efficiency (full details and results are included in section 2.3.2).

a



b

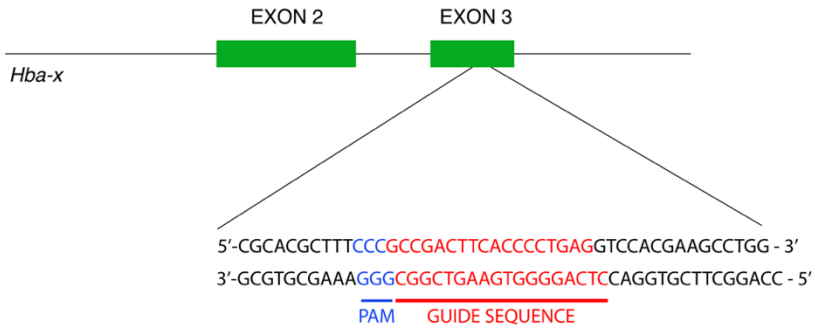


Figure 6.3 Construct used for initial ES cell targeting (previous page). a *FRT*, flippase recognition target; Neo, neomycin. b guide RNA used to target exon 3. The guide sequence is shown. PAM, protospacer adjacent motif.

Mouse ESCs were targeted using the protocol outlined in Figure 6.4. The 18 nucleotide gRNA was cloned into the PX459 plasmid (Ran et al., 2013) which also encodes a puromycin resistance gene and the Cas9 protein. Following synthesis, the targeting construct was cloned into a separate pMA plasmid. The cells were cultured under puromycin, then neomycin selection prior to isolation of clonal lines. Clones were then screened using a PCR and Southern blotting strategy outlined below. Technical details are included in the Materials and Methods chapter.

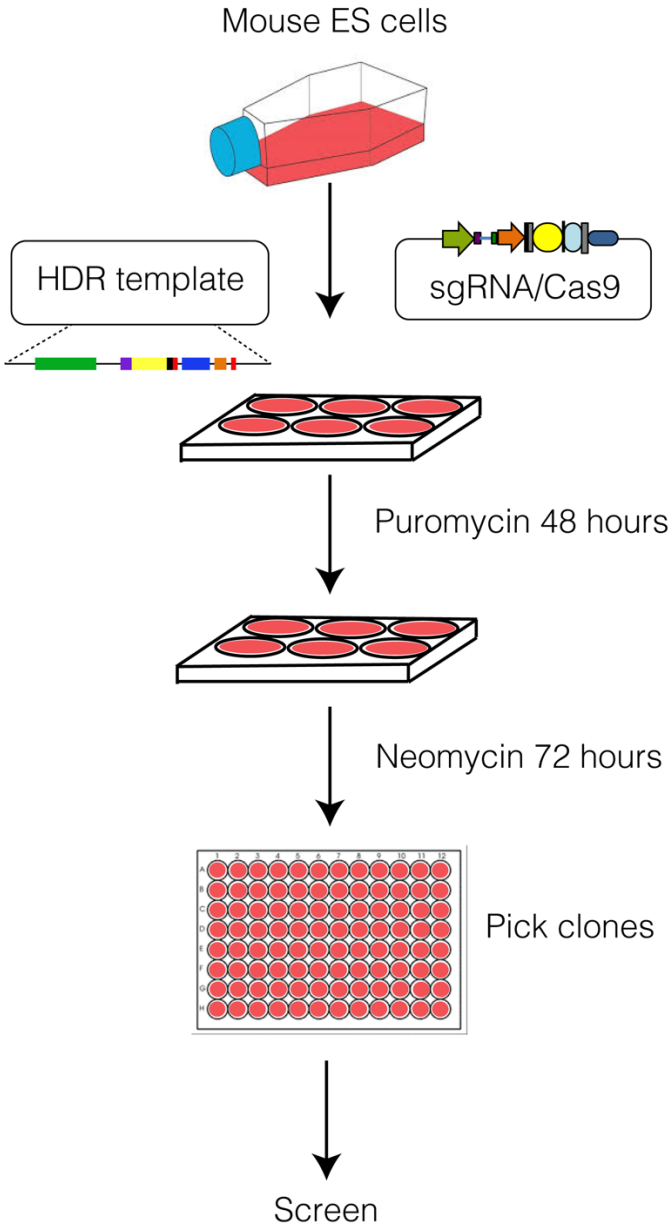


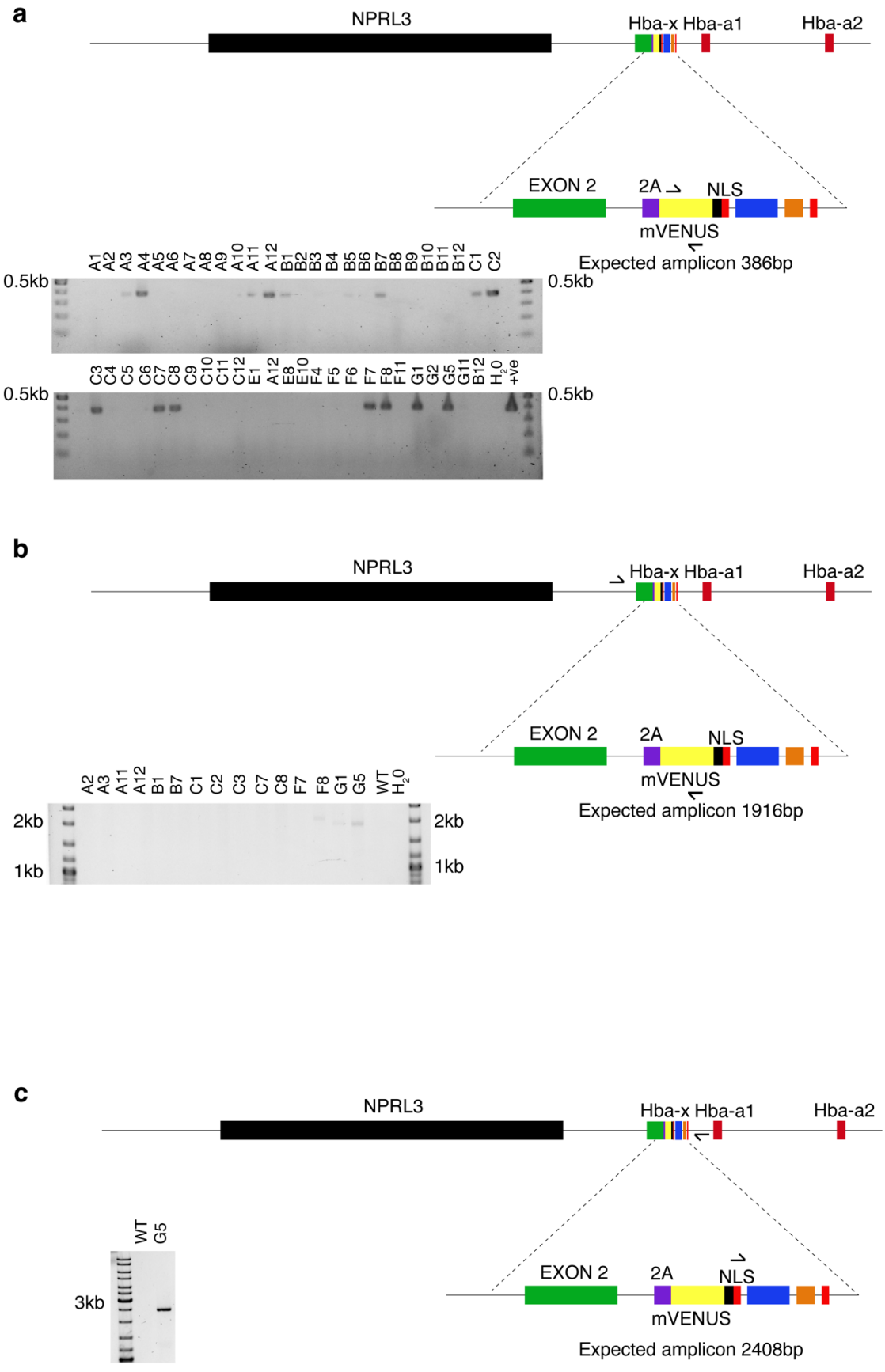
Figure 6.4 Targeting of mouse ESCs. ESCs were transduced with two plasmids: one encoding the gRNA, the Cas9 protein and a puromycin resistance gene, and one encoding the HDR template. Following sequential selection under puromycin and neomycin, clonal lines were isolated and screened.

6.2.3 Screening of targeted ESCs

52 clones were picked, expanded and DNA extracted. Targeted cells were initially screened using a PCR based strategy (Figure 6.5). Complete details, including oligonucleotide sequences can be found in the Materials and Methods.

Screening primers, denoted by arrows, were initially placed within Venus. 15/52 clones were positive. A PCR reaction was performed with the forward primer upstream of the homology arm and the reverse primer within the Venus sequence. This identified one positive clone (G5); presumably, the other clones identified by the initial PCR screening had either random integration of the construct or incomplete integration. The G5 clone was then screened with a forward primer within Venus and the reverse primer outside of the 3' homology arm; this confirmed correct integration. To ensure the locus remained intact, Southern blotting was then performed (Figure 6.6).

Figure 6.5 (next page) PCR screening of targeted clones. Primers are denoted by forward and reverse arrows. **a Screening for primers within Venus** (14F and 15R)**.** The results are shown in the adjacent gel (1% Agarose), with a positive control (using DNA from the targeting plasmid) and negative (H_2O) control**. b Screening with a primer outside of the 5' homology arm and within Venus** (16F and 15R)**.** Only one positive clone was identified, G5 (2% Agarose Gel). **c Screening with a primer within the construct and outside of the 3' homology arm** (17F and 18R)**.** G5 had an appropriately sized band identified (2% Agarose gel).


a
NPRL3
Hba-x Hba-a1
Hba-a2
EXON 2
2A
NLS
mVENUS
Expected amplicon 386bp
A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 C1 C2
0.5kb
0.5kb
C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 E1 A12 E8 E10 F4 F5 F6 F7 F8 F11 G1 G2 G5 G11 B12 H2O +ve
0.5kb
0.5kb
b
NPRL3
Hba-x Hba-a1
Hba-a2
EXON 2
2A
NLS
mVENUS
Expected amplicon 1916bp
A2 A3 A11 A12 B1 B7 C1 C2 C3 C7 C8 F7 F8 G1 G5 WT H2O
2kb
1kb
2kb
1kb
c
NPRL3
Hba-x Hba-a1
Hba-a2
EXON 2
2A
NLS
mVENUS
Expected amplicon 2408bp
WT G5
3kb

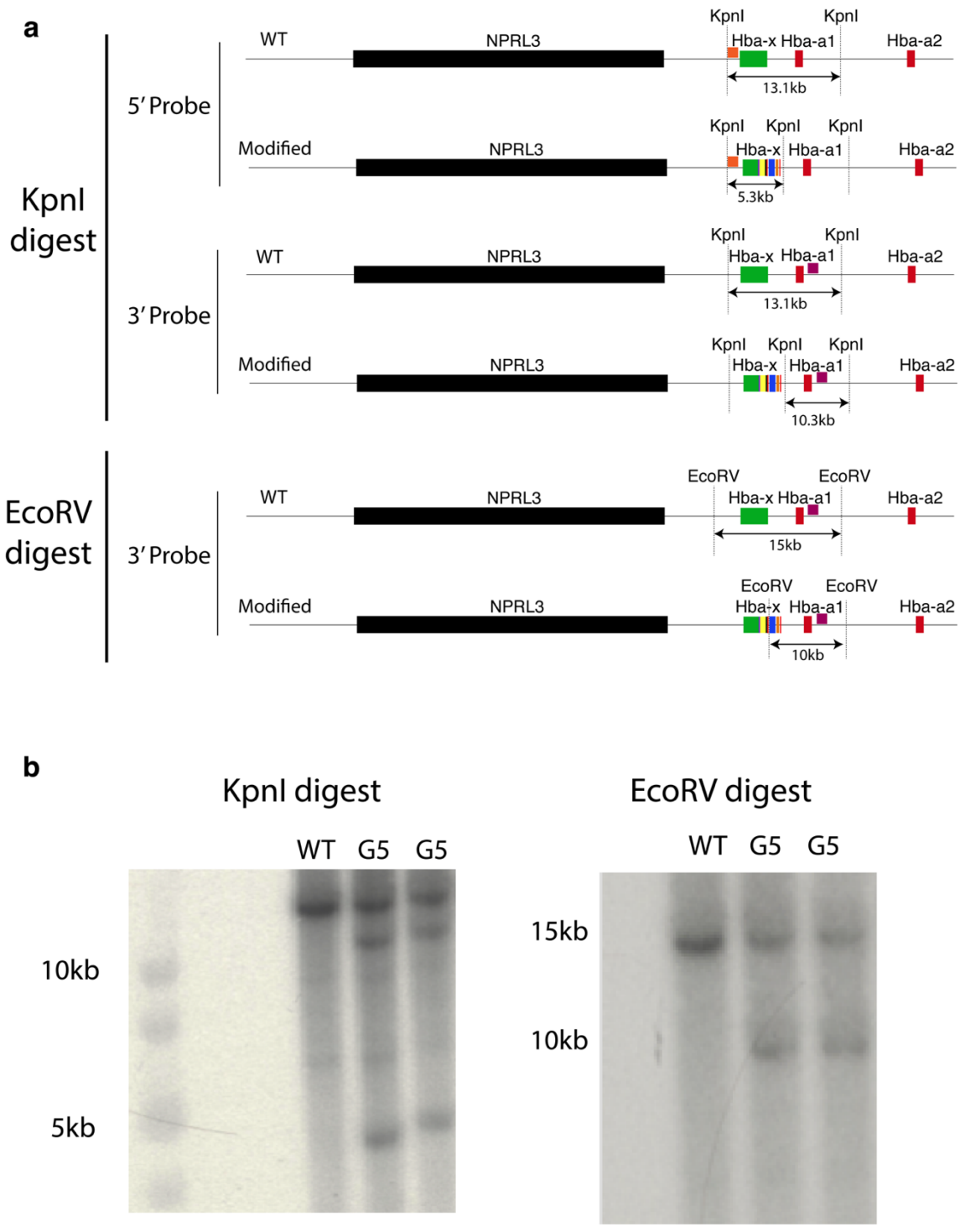


Figure 6.6 a. Southern blotting strategy with b. results.

A 5' probe, denoted by the orange rectangle, was used with a KpnI digest. A 3' probe, denoted by the purple rectangle was also used on the same blot. DNA from

the G5 clone was run twice on adjacent lanes, alongside a WT control. It can be seen that there is a WT band, a band at approximately 10.3kb, and a band at approximately 5.3kb confirming that there are no major rearrangements at the locus. The first G5 sample has run slightly faster than the second; hence the bands are lower.

A second blot was performed with the 3' probe using an EcoRV digest (again denoted with a purple box); this was done to confirm the locus lying 3' of the targeted region was intact, as this wasn't covered with the PCR strategy outlined in Figure 6.5. DNA from the G5 clone was again run twice on adjacent lanes. The correct 10kb band can be seen in the G5 clone with reduced intensity of the WT 15kb band.

To exclude the possibility of indels too small to detect using Southern blotting, the entire locus was also Sanger sequenced from outside each homology arm in the G5 clone. This confirmed that there were no unexpected mutations introduced into the sequence during the genome engineering process.

Together, these strategies confirmed that the G5 clone had the correct sequence incorporated, and that the rest of the α-globin locus remained intact.

6.2.4 Validation of the ζ-Venus reporter in the EB culture system

ES cells can be differentiated into embryoid bodies (EBs) (Chagraoui et al., 2018) using a 7 day protocol. EBs contain erythroid cells which mirror primitive erythropoiesis (Helena Francis; in preparation) and consequently express ζ-globin. Prior to blastocyst injection, the G5 clone was therefore differentiated to EBs to

ensure an adequate fluorescent signal, and that ζ-globin expression (measured from the other allele with the reverse primer in the third exon) correlated with this signal (performed in collaboration with Helena Francis). The results are shown in Figure 6.7.

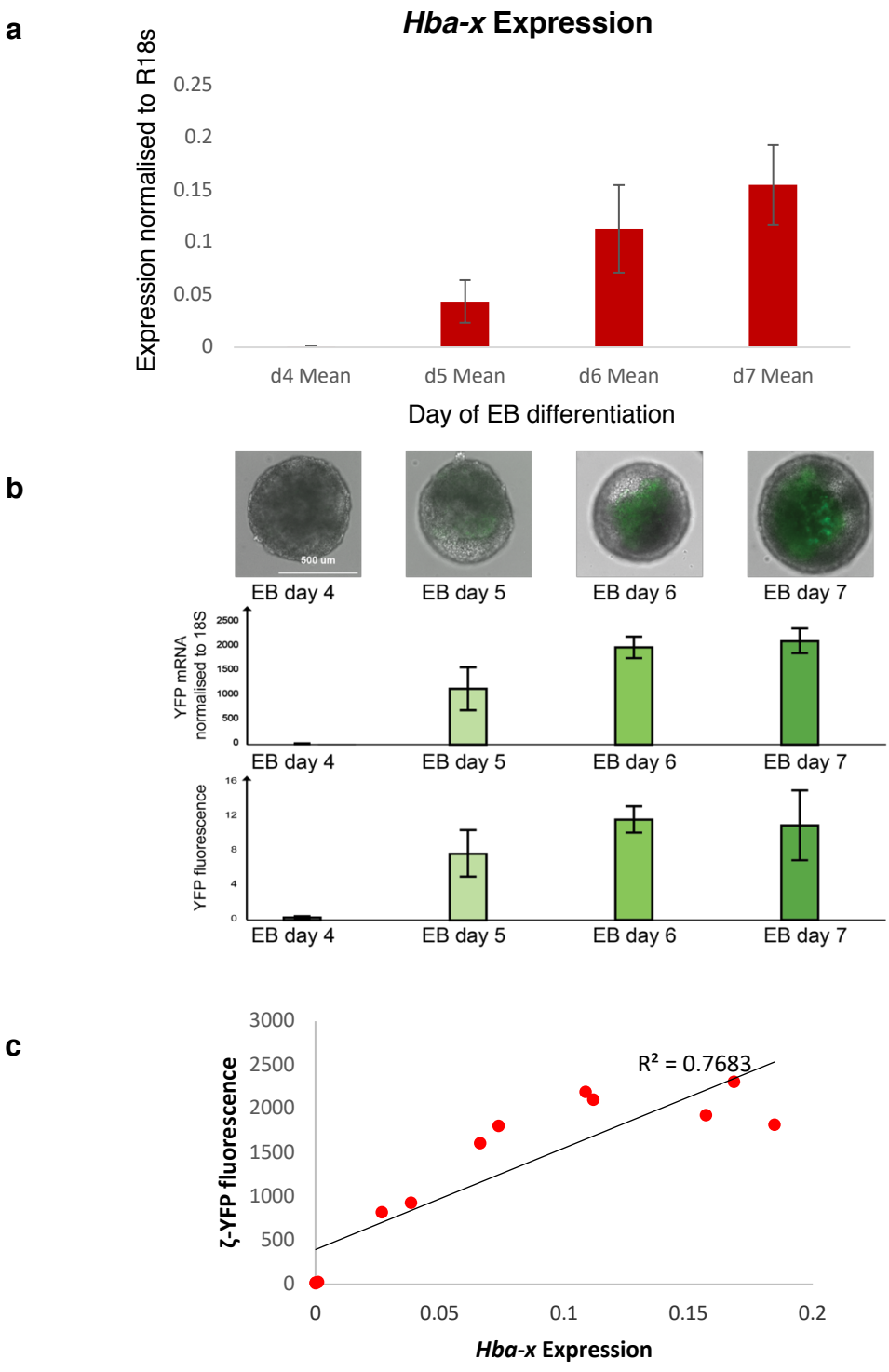


Figure 6.7 (previous page) Correlation between fluorescent signal and expression. a. ζ-globin (*Hba-x*) RNA expression measured using RT-qPCR normalised to R18s b. Microscopy and YFP (Venus) expression/fluorescence c Correlation between YFP (Venus) fluorescence and ζ-globin expression.

During EB differentiation, ζ-globin expression increases (Figure 6.7a). EBs begin emitting a fluorescent signal around Day 4 of differentiation, and this is clearly visible using light microscopy (Figure 6.7b). Quantifying this signal by flow cytometry shows that it correlates with Venus transcription and with ζ-globin (*Hba-x*) expression with a coefficient of determination (R^2) for Venus fluorescence and *Hba-x* expression of 0.77 (Figure 6.7c).

Together, these data provided evidence that the fluorescent tag in the G5 clone was functional, omits a robust fluorescent signal when ζ-globin is expressed, and that the degree of fluorescence correlates with ζ-globin expression.

6.2.5 Mouse generation

The G5 clone was karyotyped (with a normal karyotype in 90% of the cells examined in metaphase) (Jackie Sharpe, Weatherall Institute of Molecular Medicine). It was then injected into Bl6 blastocysts, and chimera offspring generated (Figure 6.8). Following germline transmission of the ζ-Venus-Neo allele (confirmed using the PCR strategy shown in Figure 6.5) mice were crossed with a mouse constitutively expressing the Flp enzyme. This removed the neomycin cassette, leaving a single

FRT site. This step was done *in vivo* to avoid unnecessary manipulation of ESCs, therefore maximising the chance of germline transmission.

Removal of the neomycin cassette was confirmed with a PCR genotyping strategy (an example gel is shown in Figure 6.8) and this product was then Sanger sequenced as confirmation.

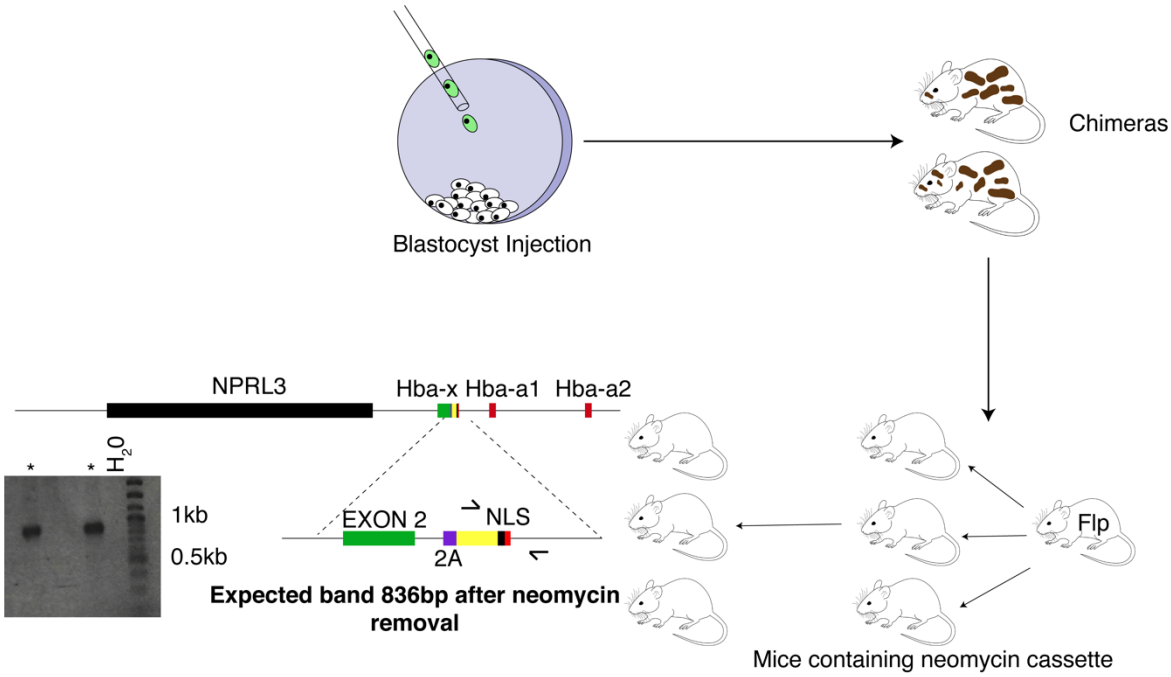


Figure 6.8 Mouse generation. The G5 clone was injected into blastocysts to generate chimeras. Following confirmation of germline transmission, mice were crossed with a mouse expressing Flp and then genotyped; primers are indicated with arrows. An example gel with two mice which have had their neomycin cassette removed is shown; the lanes containing DNA from these mice are indicated with a asterix. An 836bp band was only seen if the neomycin cassette had been removed.

Heterozygous ζ-Venus mice were then crossed to generate homozygous ζ-Venus mice. Initially, heterozygote mice were used for experiments to avoid undue time delays whilst awaiting breeding.

6.2.6 Validation of ζ-Venus *in vivo*

In order to validate the ζ-Venus tag *in vivo*, embryos were taken at E10.5 days. Microscopy and flow cytometry were performed using heterozygous ζ-Venus embryos and WT littermate controls (Figure 6.9) when ζ-globin accounts for ~40% of the total α-like globin (Figure 6.9a). A clear fluorescent signal was visible under the YFP filter for the ζ-Venus heterozygous embryo, in comparison to wild-type. This was also detectable by flow cytometry.

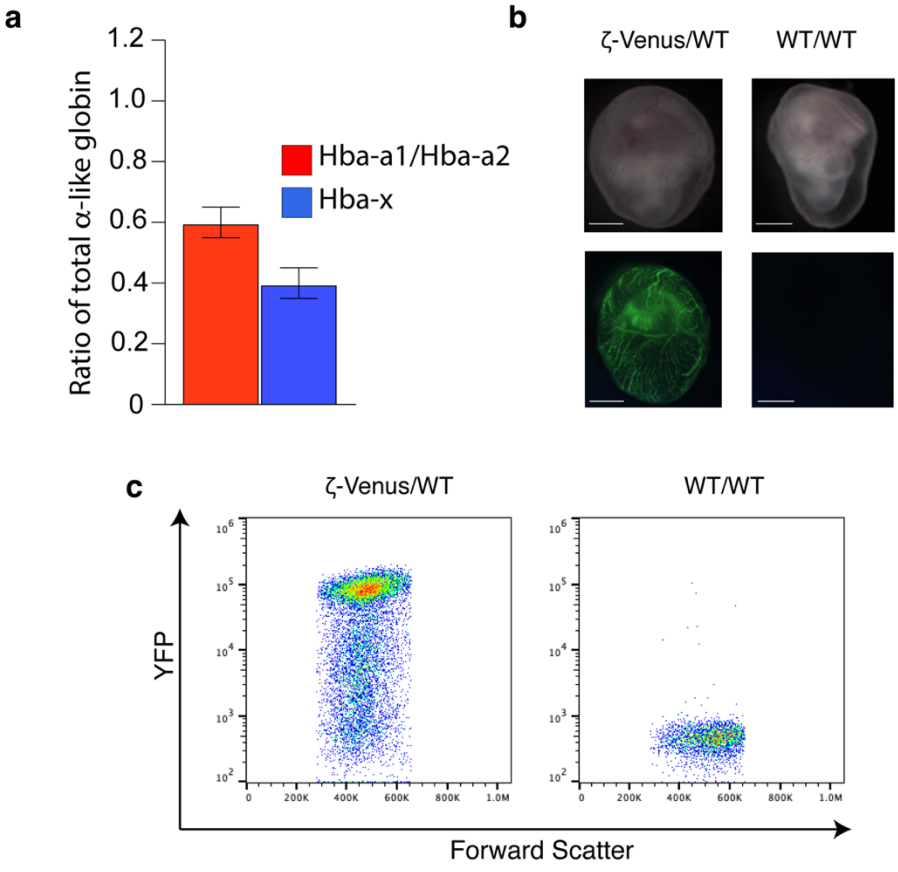


Figure 6.9 Fluorescence at E10.5. (previous page) a. Expression of ζ-globin (*Hba-x*) at E10.5 (WT cells). b. Microscopy of ζ-Venus/WT and WT/WT littermate embryos under light (top panel) and a YFP filter (bottom panel). A scale bar indicates 1mm. c. Representative flow cytometry plots for erythroid cells derived from E10.5 embryos. A clear fluorescent signal is seen in the ζ-Venus heterozygous embryos.

To examine the fluorescent signal in definitive erythroid cells, the fetal liver *ex vivo* culture system discussed in previous chapters, and outlined in detail in the Materials and Methods was used (Figure 6.10). Cells were differentiated for 48 hours and their fluorescent signal then assessed by flow cytometry.

Whilst the majority of cells in the ζ-Venus/WT model are YFP negative, there is a small population (<1%) which are positive. On back-gating, the YFP positive cells are no larger than the YFP negative cells, meaning it is unlikely they represent primitive erythroid contamination. It is plausible that a few definitive erythroid cells express small amounts of ζ-globin which is not apparent on bulk analysis; this has previously been reported in humans (Luo et al., 1999).

In order to examine whether fluorescence was observed in cells other than erythroid cells, embryos were fixed, sectioned and then examined under light microscopy and a YFP filter (Figure 6.11). Only erythroid cells in vessels were fluorescent.

Together, these data show that *in vivo* the fluorescent tag omits a detectable signal when ζ-globin is transcriptionally active, there is minimal fluorescent signal when ζ-globin is silenced, and that the signal appears to be confined to erythroid cells.

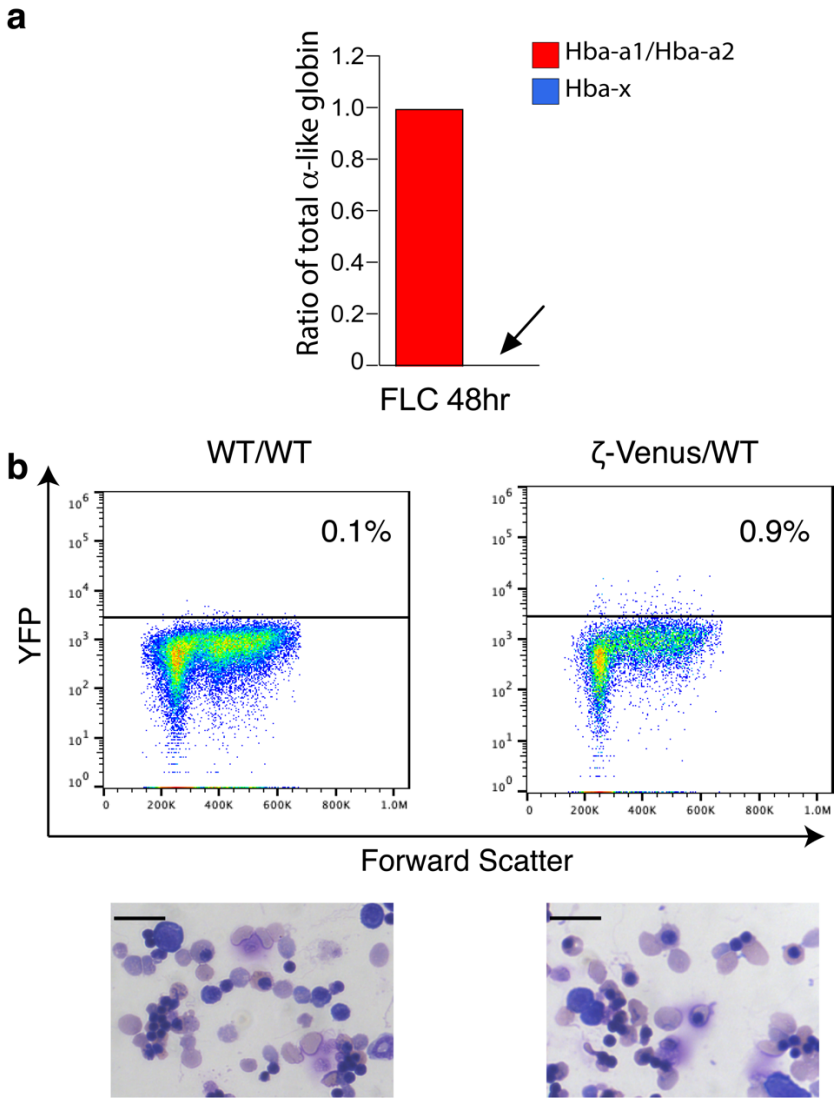


Figure 6.10 Fluorescent signal in definitive erythroid cells. **a**. After 48 hours culture, ζ-globin (*Hba-x*) expression is not detectable by RT-qPCR in WT cells. **b.** Representative flow cytometry plots and erythroblasts stained with a modified Wright's procedure (scale bar indicates 20μm). In the ζ-Venus/WT mouse, there is a small increase in YFP positive cells. The erythroid morphology is identical, showing that at least morphologically, differentiation is unperturbed.

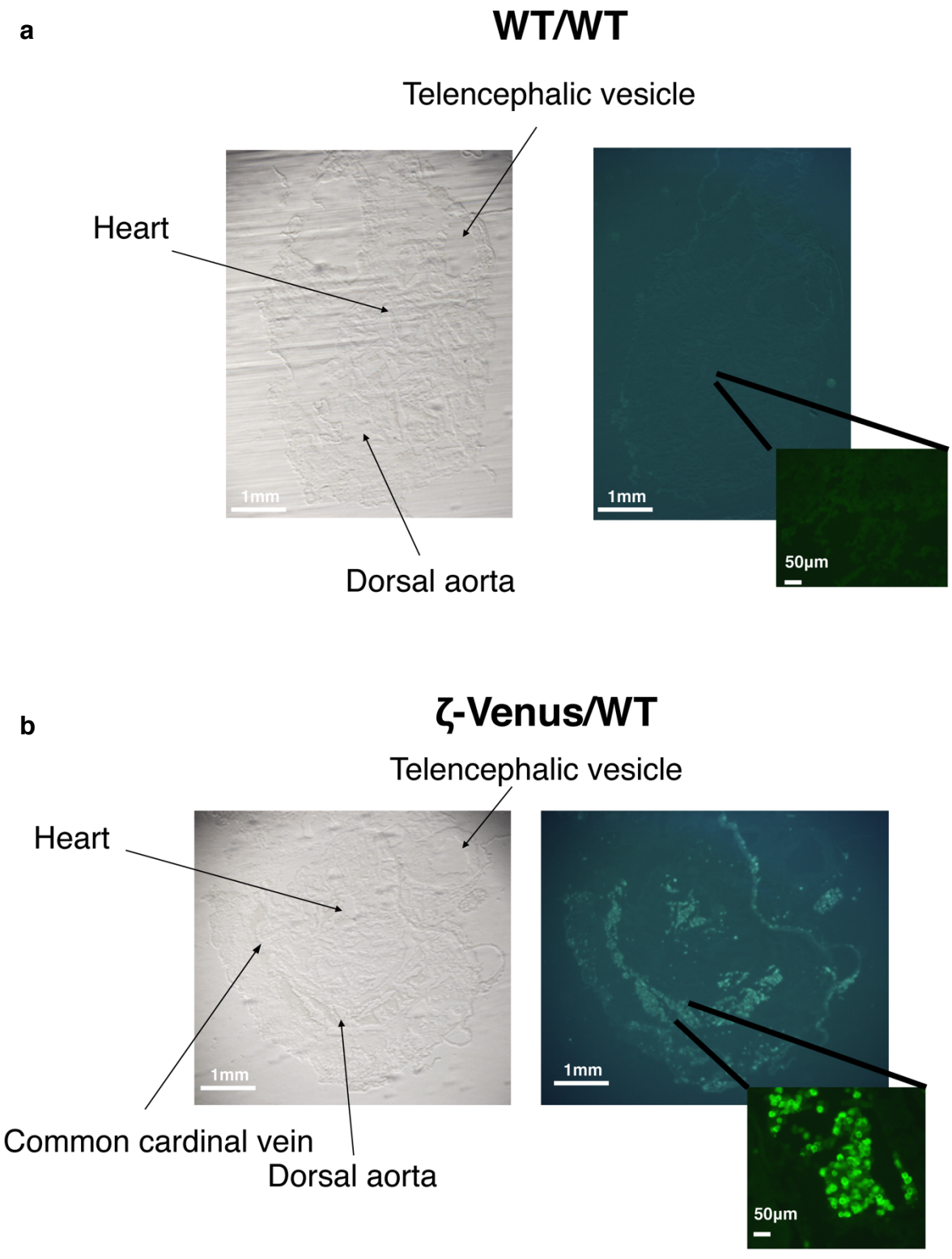


Figure 6.11 Sections of a. WT/WT and b. ζ-Venus/WT embryos. The left panel shows the section under light microscopy; the right shows the same view under a YFP filter.

6.2.7 Validation of ζ-Venus erythroid cells for drug screening

Part of the rationale for developing the ζ-Venus model was to use primary erythroid cells to screen for compounds capable of de-repressing ζ-globin in a high-throughput manner; these compounds could potentially be taken forward clinically to treat severe α-thalassaemia. In order to develop a workflow to undertake this, a positive control was required. Previously, it has been shown that HDAC inhibitors led to ζ-globin depression (Bradner et al., 2010). HC toxin, a known HDAC inhibitor, was therefore added to heterozygous ζ-Venus erythroid cells for 48 hours and the ζ-globin expression measured (via YFP signal). This leads to an increase, in a dose dependent fashion (Figure 6.12).

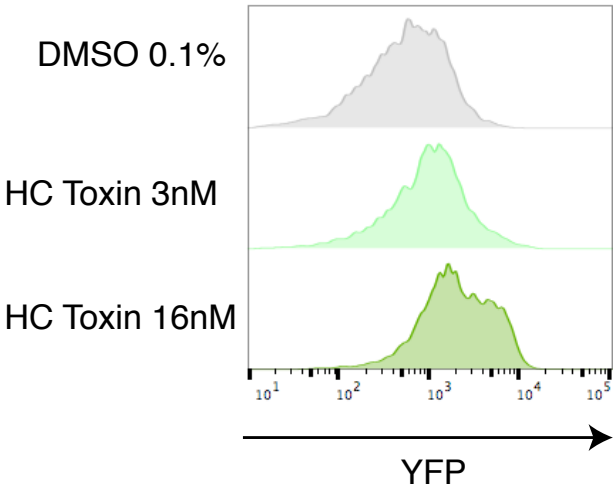


Figure 6.12 The effect of adding an HDAC inhibitor to erythroid cells. Adding an HDAC inhibitor (HC Toxin) leads to an increase in ζ-globin expression (as measured by YFP fluorescence in the ζ-Venus model).

After *ex vivo* expansion for 6 days, fetal liver cells derived from E12.5 embryos were sorted for early progenitors (selecting for CD117 using a MACS column; yielding

approximately 5 million cells); 50,000 cells were then plated out in each well of a 96 well plate (Figure 6.13). The effect of adding HC Toxin at a dose of 3nM and 16nM was confirmed across 10 separate plates, with 96 wells in each plate, that were analysed at different times. No edge effect was seen. The robust Z score (RZ') was robust at 0.94, with a signal to background ratio of 22.4 at the higher drug concentration. Such a strategy could be adapted to screening in a 384-well format with appropriate reductions in the number of cells. These data formed the basis for a successful MRC/AstraZeneca grant application where ζ-Venus primary erythroid cells will be screened with hundreds of thousands of compounds.

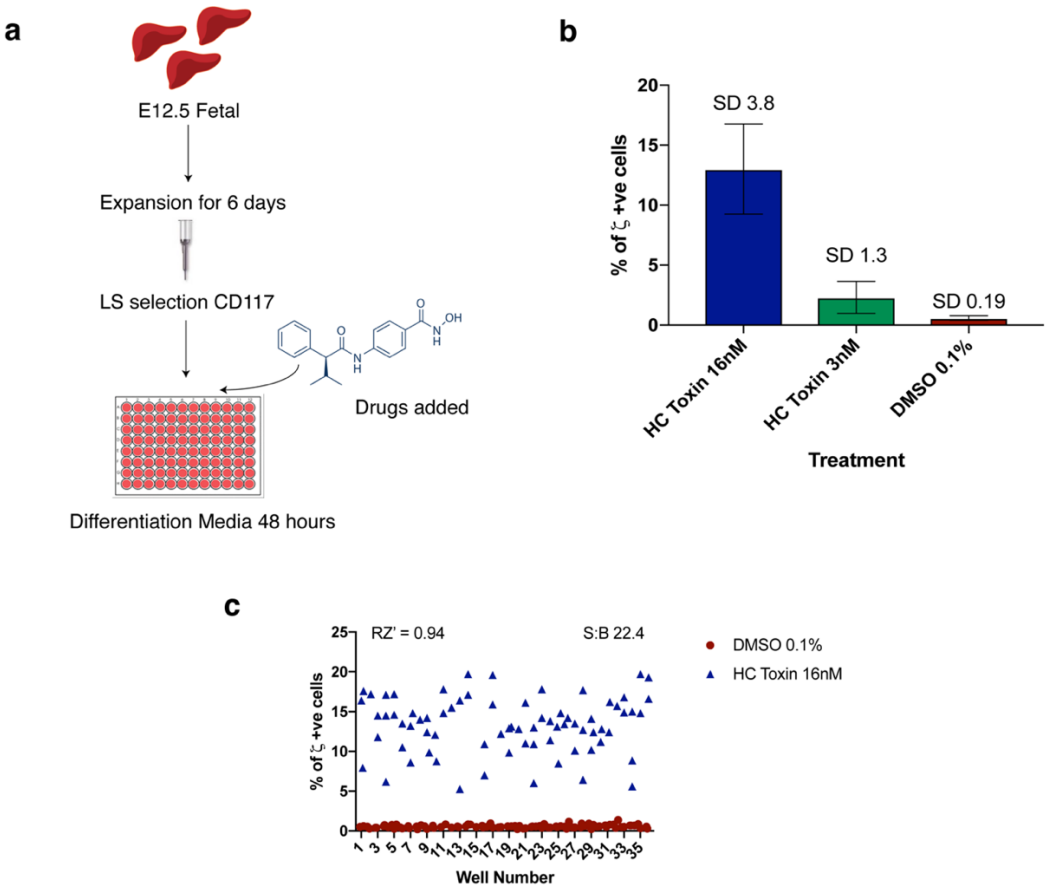


Figure 6.13 High throughput drug screening protocol. **a.** Workflow**.** **b**. A significant increase in fluorescence is observed with the positive HDAC inhibitor control across one hundred replicates. **c.** A scatter plot for the higher drug concentration and the negative control (DMSO 0.1%).

6.2.8 The effect of ζ-globin knockout on murine survival

The integration of the ζ-Venus reporter removed the third exon of ζ-globin in order to render the protein non-functional and thereby investigate the effect of homozygous ζ-globin knockout. Heterozygous (het) mice were initially crossed to produce homozygous (hom) off-spring; these then rely on α-globin transcription alone to sustain them during embryonic life.

Hom ζ-Venus mice surviving to birth are viable and appear phenotypically normal without developmental problems and so far, a normal lifespan (Figure 6.14).

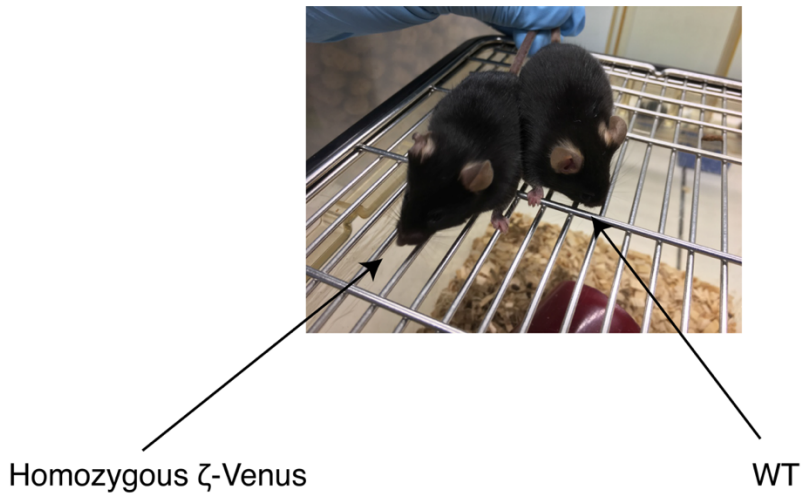


Figure 6.14 A homozygous ζ-Venus mouse with its WT littermate.

Following multiple het pairings, it became apparent however, that whilst some hom mice do survive (and that those that survive are phenotypically normal), the actual number of live births was distorted against hom mice. Hom mice were born with a

frequency of 0.07 against an expected frequency of 0.25 (p value=$1.5x10^{-4}$; binomial probability); homxhet pairings were also tried with a similar though not statistically significant result (a frequency of 0.4 against an expected frequency of 0.5; binomial probability=0.06) (Table 6.1). The mean litter size (with standard deviation) for hetxhet pairings was 4.3±2.8 (range 0-10) for hetxhet pairings; the mean litter size for homxhet pairings was 5.8±1.5 (range 4-8).

Pairing	Number of litters	Total number mice born	Number of each genotype (N; freq)		
			WT	HET	Hom
Het x Het	14	61	20 (0.33)	37 (0.61)	4 (0.07)
Hom x Het	6	35	N/A	21 (0.6)	14 (0.4)

Table 6.1 Number of each genotype of adult mice born for WT, heterozygous (Het) and homozygous (Hom) ζ-globin knockout.

Viable embryos were then studied at E12.5 from a combination of Hom x Het and Hom x Hom pairings. It appears that at this gestational age, there is no reduction in the expected number of Hom embryos with normal numbers produced from Hom x Hom pairings, and a frequency of 0.7 (against an expected frequency of 0.5) for Hom x Het pairings (Table 6.2). The mean litter size for Hom x Hom pairings at this age was 6.5±1.3 (range 4-8); the mean litter size for Hom x Het pairings was 5±0.8 (range 4-6).

Pairing	Number of litters	Total number of embryos	Number of each genotype (N; freq)		
			WT	HET	Hom
Hom x Hom	4	26	N/A	N/A	26 (1.0)
Hom x Het	4	20	N/A	6 (0.3)	15 (0.7)

Table 6.2 Number of each genotype for WT, Het and Hom ζ-globin knockout at E12.5.

In summary, these data show that knockout of ζ-globin leads to a decrease in viable births with no effect heterozygously. However, it is still possible for healthy, phenotypically normal mice to be born without ζ-globin. Whilst the live birth-rate is decreased, the number of embryos at E12.5 appears to be normal.

6.3 Discussion

The complexity of the ζ- to α-globin switch means there is a need for a model with an accurate readout of ζ-globin transcription in order to allow application of high-throughput screening techniques. This chapter describes the generation and validation of such a model where ζ-globin has been tagged; this is the first model where a globin gene has been tagged in its endogenous locus.

Previous models at the β-globin locus have used either simple transgene approaches or BACs (Fraser et al., 2007; Vacaru et al., 2013). These models, however, were all made in the pre-CRISPR/cas9 era. Utilising CRISPR/cas9 here facilitated addition of a fluorescent protein to the endogenous ζ-globin (*Hba-x*) gene;

indeed, the correct modification (confirmed by a combination of Sanger sequencing, PCR and southern blotting) was achieved in the first round of mouse ES cell targeting. Modification of the endogenous locus avoids the potential issues with variable copy number, and inappropriate gene activation/silencing by the local chromatin environment where integration occurs.

The development of fluorescent reporter models for the globin genes has historically been hampered by haemoglobin quenching, whereby haemoglobin absorbs fluorescence omitted by proteins. A modified YFP protein (mVenus) was chosen as its emission (~530nm) is where haemoglobin absorbs the least light (aside from the far-red proteins); the emission of haemoglobin fluorescence is also low here. mVenus is ~150% as bright as eGFP and has a half-life of approximately 24 hours in cells meaning it accumulates as transcription occurs. This, in combination with a nuclear localising signal to remove the protein away from globin molecules, means that there is a bright signal when ζ-globin is fully transcriptionally active at E10.5 *in vivo*. A fluorescent signal could also be detected when ζ-globin expression was ~1% of the control R18s gene in erythroid cells derived from EBs, confirming it is sensitive down to low levels of transcription. Using far-red proteins (such as mNeptune or mCardinal) (Chu et al., 2014) would be another option to avoid issues with haemoglobin quenching; however these require specialized microscopy and flow cytometry emission filters in order to detect them properly which would have added additional expense and complexity to experiments.

One of the main reasons for developing this model was to allow a CRISPR/cas9 library screen for transcription factors and repression complexes involved in repression of ζ-globin with a proposed work flow shown in Figure 6.15.

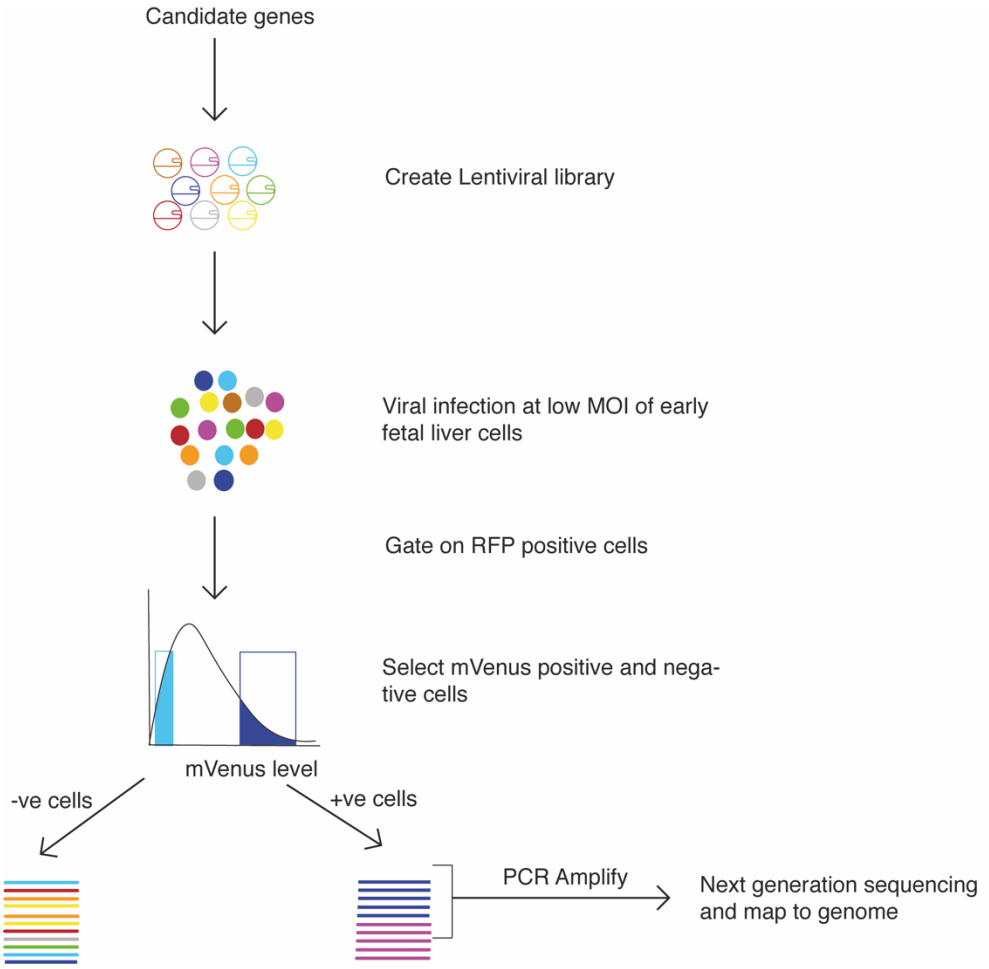


Figure 6.15 Workflow of CRISPR/Cas9 lentiviral screen. Candidate genes are cloned into lentiviral particles. ζ-Venus cells are then transduced at a low multiplicity of infection (MOI), cultured for 5-7 days, then selected for (based upon a fluorescent protein contained within the lentiviral construct) and sorted into mVenus high and low populations. DNA can then be extracted, and sequenced to determine the guide RNAs (and therefore the genes) represented in the high mVenus populations.

Initial optimisation experiments were unsuccessful and have not therefore been presented here. Whilst it is straightforward to transduce WT primary fetal liver

definitive erythroid cells (which do not express ζ-globin) with small control lentiviral particles expressing GFP, larger ones, which contained a Cas9, guide RNA (directed against BCL11A as a positive control) and red fluorescent protein (RFP) (to allow isolation by flow cytometry of transduced cells) could not be transduced. This was almost certainly due to the total size of the insert (~9kb for the Cas9/guide RNA/RFP). A potential way of circumventing this is to cross the ζ-Venus mouse with a line that constitutively expresses Cas9 in the fetal liver; such an approach was successfully used to perform a CRISPR/Cas9 screen in order to study genetic mutations involved in the development of a myeloid leukaemia (Labuhn et al., 2019). Whilst the Cas9 mouse used in those experiments had a GFP tag which would interfere with reading the YFP signal (Platt et al., 2014), another mouse is locally available without a fluorescent tag (Ben Davies, Wellcome Trust Centre, Oxford). This line will be imported into the lab and its Cas9 expression determined in primary erythroid cells from fetal livers. Assuming Cas9 is expressed at reasonable levels with fetal livers, further optimisation experiments with guide RNAs directed against known positive controls (*Bcl11a* and *Lrf*) will be performed.

Whilst work on the CRISPR/Cas9 lentiviral screen is on-going, drug screening using primary fetal liver cells is ready to be carried out with an industrial partner. It is hoped that the identification of compounds capable of reactivating ζ-globin may provide insights into how ζ-globin is repressed, as well as providing new therapeutic avenues for patients with severe α-thalassaemia.

A further purpose served by the ζ-Venus mouse is that it allows a screen for non-erythroid tissues which express ζ-globin. There is a growing literature reporting

expression of globins in human tissues, the best characterised of which is α-globin in endothelial cells where it plays a role in nitric oxide (Straub et al., 2012; 2014). Initial sectioning and examination of E10.5 embryos revealed no obvious fluorescence in tissues apart from erythroid cells in blood vessels. It will be important to confirm these findings at earlier and later time points (E8.5, E9.5, E11.5, E12.5) and also potentially in adult mice.

Orthologues of the ζ-globin genes are found in many tetrapods, chickens and all mammals suggesting that they confer some survival advantage on the organism (Hardison, 2012). As part of the design of the ζ-globin reporter mouse, exon 3 of the gene was removed. This renders the gene non-functional, allowing the question of whether ζ-globin is necessary for normal development to be studied. Previously, it has been demonstrated that ζ-globin deletion (via insertion of a PGK-Neo selection cassette in place of the ζ-globin gene in mice) does still allow viable embryos to be born, although there was a decreased live birth rate (Leder et al., 1997). The genetic background of the mice was important for survival of the ζ-null embryos however, with those on an in-bred 129/SvEv background having no viable births. One flaw with this experiment (which was unavoidable due to the genome engineering technology in use at the time) was that the selection gene (neomycin) with a promoter was left in place of the ζ-globin gene; this diverts the α-globin enhancers away from the α-globin genes, leading to an α-thalassaemia.

The data presented here, show that there is a statistically decreased birth rate for live ζ-null (ζ-Venus/ζ-Venus) mice from hetxhet pairings. Those that do survive however, are phenotypically indistinguishable from their WT littermates. All the

mice here have been bred onto a C57/BL/6 background, so the question of whether survival depends upon the genetic background could not be addressed. Of note, survival at E12.5 does not appear to be affected, although the number of litters analysed so far is relatively small which may affect the accuracy of the results. This result is presumably due to an insult during embryonic life caused by lack of ζ-globin leading to fetal death after E12.5.

It therefore appears that there is some genetic redundancy between the ζ-globin and adult α-globin genes during development and that it isn't essential for normal mice to be born. This is perhaps unsurprising, given α-globin expression is always found at reasonable levels in mice, even in the earliest primitive erythroid cells. Nevertheless, there is a survival advantage in having at least one functional ζ-globin gene and this helps explain why it is so strongly conserved across evolution. It remains unclear as to how this survival advantage is conferred: presumably the oxygen carrying properties of the embryonic haemoglobins (or even functions unrelated to their gas exchange) are advantageous very early in development although this has never been proven.

In summary, this chapter reports the development of a robust *in vivo* reporter for ζ-globin which can be used for drug screening, and, in the future, may be used for CRISPR/Cas9 library screening. It also reports on the effects of knockout of ζ-globin on mouse survival.

7 General discussion and conclusions

This thesis explores how the ζ-globin gene is regulated, focusing on how it is silenced. From the perspective of gene regulation, ζ-globin is a particularly interesting gene to study, in that it lies next to a strong enhancer cluster in the well characterised α-globin gene locus, yet is completely silenced in definitive erythroid cells. There is also a potential translational aspect given reactivation of ζ-globin would be of therapeutic benefit in severe α-thalassaemia. Here, I summarise the main conclusions from this thesis and highlight future avenues for study.

7.1 ζ-globin is advantageous but not essential for normal development

Prior to this work, it was uncertain whether ζ-globin was necessary for life although it is strongly conserved throughout evolution (Hughes et al., 2005). As part of the design of a ζ-Venus reporter mouse, the third exon of the ζ-globin gene was removed making it non-functional. Knockout homozygously led to a statistically significant decrease in viable births, although viable births of phenotypically normal, healthy mice is still possible (**Chapter 6**). It therefore appears that α-globin alone can in principle, serve the needs of a developing embryo. ζ-globin does however, provide a clear survival advantage. The number of embryos at E12.5 is within expected limits, although this may reflect that the death of the developing ζ-null fetus occurs between E12.5 and E20 following an insult earlier in embryogenesis.

Presumably, there is an advantage when ζ-globin is incorporated into haemoglobin tetramers during very early embryogenesis. This may relate to its protein binding characteristics with the β-like embryonic globins, or the different gas exchange properties of the haemoglobins produced when ζ-globin is present instead of α-globin (He and Russell, 2001; Manning et al., 2007). Future work with the generated mouse model will investigate the full blood counts in both embryonic and adult ζ-null mice to determine the degree, timing and relevance of any anaemia. The importance of ζ-globin *in vivo* could be proven by substituting the codons of α-globin for ζ-globin in another mouse model. Assuming expression and silencing are promoter driven, this would then lead to developmental expression of additional α-globin in place of ζ-globin, allowing the importance of the ζ-globin protein *per se* to be studied.

7.2 There are decreased contacts between the enhancers and the ζ-globin promoter when definitive and primitive erythroid cells are compared

In definitive and primitive erythroid cells, the same CTCF sites and enhancer elements are found across the locus. As in definitive erythropoiesis, the five enhancer elements classify as a super-enhancer in primitive erythroid cells (**Chapter 3**). Interaction data from the upstream HS-38 CTCF element show that the self-interacting domain formed in primitive and definitive erythroid cells is almost identical. However, chromatin interactions from within the domain are significantly altered, with decreased contacts between the ζ-globin promoter and the enhancers in definitive erythroid cells. Of note, interactions between the ζ-globin promoter and

enhancers are still increased in definitive erythroid cells when compared to embryonic stem cells, when the compartment is not formed; this means that ζ-globin is still contained in the same self-interacting domain in definitive erythroid cells and that it simply has a lower contact frequency with its enhancers.

It is interesting to consider how these data can be reconciled with the phase separation model of transcriptional regulation (Hnisz et al., 2017). With such a model, the ζ-globin gene and its surrounding chromatin would need to be isolated in definitive erythroid cells, either within its own discrete small droplet within a broader activating domain, or alternatively on the edge of the activation domain away from the majority of the transcription factors and co-factors. In primitive erythroid cells by contrast, the gene would be in the heart of the activation domain. It would be of significant interest though challenging, to try and design experiments to see whether such a process is happening.

7.3 Silencing of ζ-globin is mediated by chromatin remodelling and histone deacetylation

ζ-globin gene silencing appears to be predominantly mediated via histone deacetylation, presumably also with a contribution from chromatin remodellers which alter nucleosome occupancy. PRC2, DNA methylation and H3K9me2/3 do not play a significant role, assuming a conserved mechanism between mice and human. There are several lines of evidence for this presented within the thesis:

- In primitive erythropoiesis in mice, there is a broad domain of histone acetylation encompassing the ζ-globin gene; in definitive erythroid cells there is a localised area of histone deacetylation, associated with chromatin inaccessibility (**Chapter 3**).
- There are only 6 CpGs in the ζ-globin promoter in humans and when these are unmethylated (as in WT HUDEP-2 cells), ζ-globin expression is still repressed. Methylation of the intergenic CpG island found in the human ζ-globin gene also correlates with transcription through the gene, rather than lack of expression (**Chapter 5**).
- ChIP-seq data for H3K27me3 does not show enrichment in definitive erythroid cells in either mice or human, excluding PRC2 as a silencing mechanism (**Chapter 3, Chapter 5**).
- ChIP-seq data for H3K9me3 in human does not show enrichment in definitive erythroid cells. Inhibition of the enzyme responsible for laying down H3K9me2 also doesn't lead to an increase in ζ-globin expression (**Chapter 5**).
- Knockout of the transcription factors LRF and BCL11A in both mouse and human cells leads to ζ-globin de-repression (**Chapter 4, Chapter 5**). These transcription factors associate with the NuRD complex, and thereby lead to chromatin remodelling and histone deacetylation.

A proposed model for ζ-globin silencing based upon this and the NG Capture-C data is shown in Figure 7.1.

There are a still a number of questions which remain unanswered however.

Although histone deacetylation and chromatin remodelling are clearly important, it isn't entirely clear which silencing complexes are responsible. The NuRD complex is known to associate with both BCL11A (Sankaran et al., 2008; Xu et al., 2013) and LRF(Masuda et al., 2016) in proteomic affinity screens; furthermore a binding site for a component of NuRD has been identified within the N-terminus of BCL11A (Xu et al., 2013).

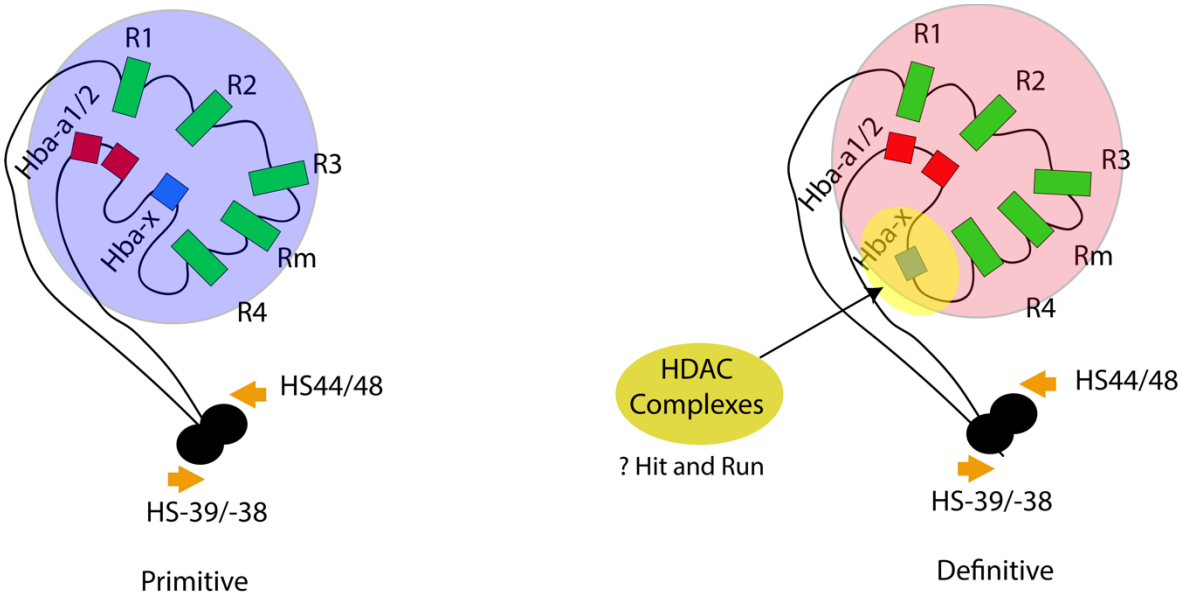


Figure 7.1 Proposed model explaining silencing of ζ-globin. Localised histone deacetylation and chromatin remodelling facilitated by transcription factors leads to decreased contact between the enhancers and the ζ-globin promoter, despite the same contact domain being formed.

It therefore does seem that NuRD plays a role in ζ-globin repression. However, binding at the ζ-globin promoter has never been shown with ChIP for any member of the NuRD complex. Importantly BCL11A has also been shown to associate with

other complexes including SIN3A and CoREST (Xu et al., 2013) so it is possible there may be more than one silencing complex responsible.

A further issue is whether the complex is recruited to the ζ-globin promoter throughout erythropoiesis, or transiently binds at a particular stage of development (or cell cycle) and then dissociates, with the epigenetic modification then faithfully passed on throughout differentiation. The latter "hit and run" model (Shah et al., 2019) is difficult to detect with traditional ChIP-seq methods, which provides a static viewpoint of occupancy at the timepoint of the experiment, in large numbers of cells. The development of low cell number techniques to detect protein occupancy, in particular CUT&RUN (Skene and Henikoff, 2017) with FACS sorting, should allow the different complexes' binding to be studied throughout the stages of differentiation. However, conventional ChIP of NuRD, SIN3A and CoREST sub-units is not straight forward, so this may prove challenging.

Finally, as discussed below, it is almost certain that there are additional transcription factors responsible for recruiting HDAC complexes in addition to LRF and BCL11A which are currently unknown.

7.4 LRF and BCL11A play an important role in ζ-globin repression

Lrf and Bcl11a play a critical role in repression of ζ-globin in fetal and adult erythropoiesis in mice and in an adult human erythroid model **(Chapters 4 and 5).** Their knockout has a greater effect in early fetal stage, as opposed to later

fetal/adult stage erythroid cells (**Chapter 4**) implying additional silencing mechanisms which may be specific to late fetal or adult erythropoiesis. When both are knocked out in a human definitive erythroid cell line, the chromatin accessibility over the ζ-globin gene and level of ζ-globin expression does not return to that seen in primitive erythropoiesis, implying that other transcription factors are involved (**Chapter 5**). In both mouse and human, LRF knockout consistently has a greater effect on ζ-globin de-repression than BCL11A removal.

Based upon this work and the work of others outlined in detail in the introduction (Arnaud et al., 2010; Viprakasit et al., 2014b, Yi et al., 2006; Gudmundsdottir et al., 2018; Funnell et al., 2013; Bianchi et al., 2010; Sankaran et al., 2011), it is possible to construct a transcriptional factor network (Figure 7.2) responsible for repression of ζ-globin.

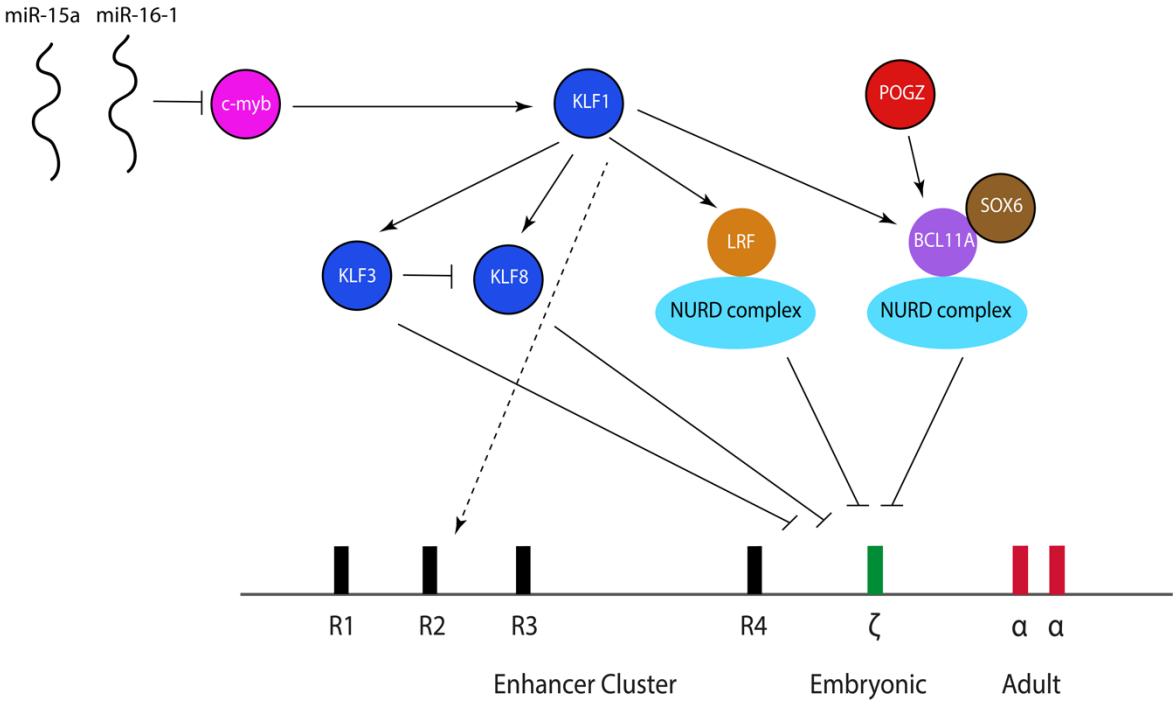


Figure 7.2 Transcriptional network responsible for repressing ζ-globin.

Exactly how this network functions is still incompletely understood. It should be noted that the majority of the transcription factors so far identified as potentially important in ζ-globin regulation operate through LRF or BCL11A. It may be (and is perhaps likely) that there is another complete transcriptional network which hasn't yet been discovered, or additional downstream targets of the known transcription factors.

An unexplained and intriguing observation is that some transcription factors implicated in switching (including LRF, KLF1 and POGZ) are not differentially expressed between primitive and definitive erythropoiesis when considering mRNA level (Kingsley et al., 2013) in contrast to others such as BCL11A. It may therefore be that there are post-translational modifications of some transcription factors which affect their function or additional co-factors which haven't yet been identified. This requires further work.

Detecting where BCL11A and LRF bind to exert their effects has been challenging. Whilst the *in vitro* data identifying their binding sequences are convincing, and the link with HPFH elegant (Liu et al., 2018; Martyn et al., 2018), I find the ChIP-seq and CUT&RUN data difficult to trust, with binding seen across the α-globin locus, including the adult globin genes. Binding of both proteins is also seen at the adult β-globin genes and hasn't been satisfactorily explained. My own experiences attempting to determine BCL11A binding using both CUT&RUN and ChIP have been unsuccessful (and they therefore haven't been presented in this thesis); this highlights the technical challenges identifying their binding *in vivo* and I think the story remains incomplete.

As with silencing complexes, transcription factors may only bind transiently during differentiation. This has been proposed for BCL11A (Martyn et al., 2018). As when detecting the silencing complexes, detecting "hit and run" transcription factors is challenging (Shah et al., 2019). The use of inducible proteins, cell sorting, low number novel techniques such as CUT&RUN and small molecules that arrest the cell cycle at particular stages may all be necessary to visualise direct transcription factor occupancy of repressed genes (Shah et al., 2019).

It is important to consider the "hit and run" model for transcription factor silencing when designing CRISPR screens on primary cells. The ζ-Venus model in combination with the fetal liver culture system is a valuable tool (**Chapter 6**); however, at least some HSCs from this system will have to be targeted to ensure all stages of differentiation are covered. HSCs capable of repopulating the entire murine haematopoietic system are found in the fetal liver (Ema and Nakauchi, 2000) and this should therefore be possible.

Aside from screening transcription factors in a high-throughput manner, it would also be of great interest to screen the ζ-globin promoter in definitive erythroid cells using saturation mutagenesis in a similar fashion to studies focusing on the BCL11A enhancers (Canver et al., 2015). This should provide sequences necessary for repression of ζ-globin, and potentially therefore provide insights into repressive transcription factors. Such an approach has already been employed successfully in primary erythroid cells derived from $CD34^+$ cell cultures (Alistair Funnell, unpublished) examining the γ-globin locus.

Finally, much of the work on ζ-globin so far has focused on silencing rather than activating mechanisms. The characterization of the EB differentiation system from mouse ES cells in our laboratory shows that the erythroid cells derived from this have a primitive-like signature using both single cell RNA and ATAC-seq (Hele Francis; in preparation). This system therefore provides a way of screening for activating transcription factors and performing saturation mutagenesis of the ζ-globin promoter in primitive erythropoiesis.

7.5 Potential therapeutic strategies to de-repress ζ-globin

The work presented here offers rational targets for therapies to de-repress ζ-globin in order to treat patients with severe α-thalassaemia. Drugs targeting chromatin remodellers and HDACs may lead to ζ-globin de-repression and a collaboration has been established with an industrial partner to screen for potential molecules (**Chapter 6**).

Separately, the same gene therapy approaches proposed for reactivation of γ-globin will also reactivate ζ-globin. The BCL11A consensus binding site (TGACC) is located within the ζ-globin promoter and could therefore be mutated, potentially alongside the LRF (or other transcription factors') binding site. Alternatively, the erythroid specific enhancer within BCL11A could be targeted, as has now successfully been employed in human HSCs (Wu et al., 2019).

Whilst joint BCL11A and LRF knockout here led to ζ-globin levels which would probably be sub-therapeutic, this is in the presence of both α-globin genes. It is

well known that patients with the ζ--^{SEA} mutation express increased levels of ζ-globin (presumably due to increased enhancer contact with the ζ-globin promoter in the absence of the α-globin promoters) (Chung et al., 1984), and it may therefore be that removal of two repressive transcription factors (or more preferably their binding sites) would up-regulate ζ-globin to potentially therapeutic levels.

A complicating factor in targeting the ζ-globin promoter is the presence of the ψζ gene, which has almost complete homology with the ζ-globin gene and lies 10.2kb 3´ (Proudfoot et al., 1982). Targeting the ζ-globin gene promoter with a WT CRISPR/Cas9 would also cause a double strand break at the ψζ pseudogene and potentially lead to deletions, gene fusion between ζ and ψζ, and no functional ζ-globin. Preliminary experiments which are not presented in the thesis, have confirmed this. The development of base-editors (where a catalytically inert Cas9 protein is fused to a cytosine or adenine deaminase) could circumvent this issue, by mutating CC bases without inducing a dsDNA break (Komor et al., 2016). There have however, been recent concerns about off-target effects from cytosine deaminases (Zuo et al., 2019) and such an approach would therefore require careful validation in pre-clinical models.

I suspect it will always be preferable to offer termination or pre-implantation genetics to parents at risk of BHFS. However, if this is unacceptable to parents, it could also be envisaged in the future that affected children could be diagnosed early in utero (potentially using non-invasive circulating free foetal DNA (Cutts et al., 2019)). They could then be salvaged with aggressive intra-uterine transfusion and have their cord blood banked for modification of autologous stem cells to reactivate ζ-globin, which

could be reinfused at a later date following conditioning chemotherapy or antibody-based ablation of the bone marrow. Small molecules targeting epigenetic repressors could also be used to reduce or obviate the need for transfusion.

7.6 Summary

In summary, the data presented in this thesis contribute significantly to the understanding of the mechanisms underlying how the embryonic ζ-globin gene is silenced. Histone deacetylation and chromatin remodelling play a crucial role, mediated in part, by two transcription factors: LRF and BCL11A. Other transcription factors must also contribute, and a novel mouse model has therefore been generated with a fluorescent protein attached to ζ-globin to facilitate high-throughput screening approaches.

8 References

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat Commun *9*, 1911.

Amid, A., Chen, S., Brien, W., Kirby-Allen, M., and Odame, I. (2016). Optimizing chronic transfusion therapy for survivors of hemoglobin Barts hydrops fetalis. Blood *127*, blood–2015–10–673889–1211.

Anguita, E., Johnson, C.A., Wood, W.G., Turner, B.M., and Higgs, D.R. (2001). Identification of a conserved erythroid specific domain of histone acetylation across the alpha-globin gene cluster. Pnas *98*, 12114–12119.

Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Pnas *90*, 11995–11999.

Arnaud, L., Saison, C., Helias, V., Lucien, N., Steschenko, D., Giarratana, M.-C., Prehu, C., Foliguet, B., Montout, L., de Brevern, A.G., et al. (2010). A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. Am. J. Hum. Genet. *87*, 721–727.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. Bioessays *34*, 135–141.

Baron, M.H., Vacaru, A., and Nieves, J. (2013). Erythroid development in the mammalian embryo. Blood Cells Mol. Dis. *51*, 213–219.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell *129*, 823–837.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell *143*, 470–484.

Basak, A., Hancarova, M., Ulirsch, J.C., Balci, T.B., Trkova, M., Pelisek, M., Vlckova, M., Muzikova, K., Cermak, J., Trka, J., et al. (2015). BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations. J. Clin. Invest. *125*, 2363–2368.

Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L., et al. (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science *342*, 253–257.

Beaudry, M.A., Ferguson, D.J., Pearse, K., Yanofsky, R.A., Rubin, E.M., and Kan, Y.W. (1986). Survival of a hydropic infant with homozygous alpha-thalassemia-1. J. Pediatr. *108*, 713–716.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends in Genetics *32*, 29–41.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.

Benz, E.J., Murnane, M.J., Tonkonow, B.L., Berman, B.W., Mazur, E.M., Cavallesco, C., Jenko, T., Snyder, E.L., Forget, B.G., and Hoffman, R. (1980). Embryonic-fetal erythroid characteristics of a human leukemic cell line. Pnas *77*, 3509–3513.

Bianchi, E., Zini, R., Salati, S., Tenedini, E., Norfo, R., Tagliafico, E., Manfredini, R., and Ferrari, S. (2010). c-myb supports erythropoiesis through the transactivation of KLF1 and LMO2 expression. Blood *116*, e99–e110.

Bird, A.P., Taggart, M.H., Nicholls, R.D., and Higgs, D.R. (1987). Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. Embo J. *6*, 999–1004.

Blatnik, J.S., Schmid-Schönbein, G.W., and Sung, L.A. (2005). The influence of fluid shear stress on the remodeling of the embryonic primary capillary plexus. Biomech Model Mechanobiol *4*, 211–220.

Blazsek, I., Chagraoui, J., and Péault, B. (2000). Ontogenic emergence of the hematon, a morphogenetic stromal unit that supports multipotential hematopoietic progenitors in mouse bone marrow. Blood *96*, 3763–3771.

Blobel, G.A., Bodine, D., Brand, M., Crispino, J., de Bruijn, M.F.T.R., Nathan, D., Papayannopoulou, T., Porcher, C., Strouboulis, J., Zon, L., et al. (2015). An international effort to cure a global health problem: A report on the 19th Hemoglobin Switching Conference. Exp. Hematol. *43*, 821–837.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842–1855.e16.

Bouabe, H., and Okkenhaug, K. (2013). Gene Targeting in Mice: A Review. In Virus-Host Interactions, (Totowa, NJ: Humana Press, Totowa, NJ), pp. 315–336.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell *132*, 311–322.

Bradner, J. E., Mak, R., Tanguturi, S. K., Mazitschek, R., Haggarty, S. J., Ross, K., 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 of the United States of America, *107*, 12617–12622.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Meth *10*, 1213–1218.

Bulger, M. (2005). Hyperacetylated chromatin domains: lessons from heterochromatin. J. Biol. Chem. *280*, 21689–21692.

Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature *527*, 192–197.

Chagraoui, H., Kristiansen, M.S., Ruiz, J.P., Serra-Barros, A., Richter, J., Hall-Ponselé, E., Gray, N., Waithe, D., Clark, K., Hublitz, P., et al. (2018). SCL/TAL1 cooperates with Polycomb RYBP-PRC1 to suppress alternative lineages in blood-fated cells. Nat Commun *9*, 5375–17.

Chan, W.Y., Leung, A.W., Luk, C.W., Li, R.C., Ling, A.S., and Ha, S.Y. (2018). Outcomes and morbidities of patients who survive haemoglobin Bart's hydrops fetalis syndrome: 20-year retrospective review. Hong Kong Med J *24*, 107–118.

Chang, K.-H., Nelson, A.M., Cao, H., Wang, L., Nakamoto, B., Ware, C.B., and Papayannopoulou, T. (2006). Definitive-like erythroid cells derived from human embryonic stem cells coexpress high levels of embryonic and fetal globins with little or no adult globin. Blood *108*, 1515–1523.

Charache, S., Catalano, P., Burns, S., Jones, R.T., Koler, R.D., Rutstein, R., and Williams, R.R. (1985). Pregnancy in carriers of high-affinity hemoglobins. Blood *65*, 713–718.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Chu, J., Haynes, R.D., Corbel, S.Y., Li, P., González-González, E., Burg, J.S., Ataie, N.J., Lam, A.J., Cranfill, P.J., Baird, M.A., et al. (2014). Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. Nat Meth *11*, 572–578.

Chui, D.H., and Waye, J.S. (1998). Hydrops fetalis caused by alpha-thalassemia: an emerging health care problem. Blood *91*, 2213–2222.

Chung, S.W., Wong, S.C., Clarke, B.J., Patterson, M., Walker, W.H., and Chui, D.H. (1984). Human embryonic zeta-globin chains in adult patients with alpha-thalassemias. Pnas *81*, 6188–6191.

Coelho, A., Picanço, I., Seuanes, F., Seixas, M.T., and Faustino, P. (2010). Novel large deletions in the human alpha-globin gene cluster: Clarifying the HS-40 long-range regulatory role in the native chromosome environment. Blood Cells Mol. Dis. *45*, 147–153.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Meth *14*, 959–962.

Cranfill, P.J., Sell, B.R., Baird, M.A., Allen, J.R., Lavagnino, Z., de Gruiter, H.M., Kremers, G.-J., Davidson, M.W., Ustione, A., and Piston, D.W. (2016). Quantitative assessment of fluorescent proteins. Nat Meth *13*, 557–562.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.-Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Research *19*, 24–32.

Cutts, A., Vavoulis, D.V., Petrou, M., Smith, F., Clark, B., Henderson, S., and Schuh, A. (2019). A method for non-invasive prenatal diagnosis of monogenic autosomal recessive disorders. Blood.2019002099.

Davies, J.O.J., Oudelaar, A.M., Higgs, D.R., and Hughes, J.R. (2017). How best to identify chromosomal interactions: a comparison of approaches. Nat Meth *14*, 125–134.

Davies, J.O.J., Telenius, J.M., McGowan, S.J., Roberts, N.A., Taylor, S., Higgs, D.R., and Hughes, J.R. (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat Meth *13*, 74–80.

De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., and Higgs, D.R. (2007). Tissue-specific histone modification and transcription factor binding in α globin gene expression. Blood *110*, 4503–4510.

De Gobbi, M., Viprakasit, V., Hughes, J.R., Fisher, C., Buckle, V.J., Ayyub, H., Gibbons, R.J., Vernimmen, D., Yoshinaga, Y., de Jong, P., et al. (2006). A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science *312*, 1215–1217.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Derderian, S.C., Jeanty, C., Walters, M.C., Vichinsky, E., and MacKenzie, T.C. (2014). In utero hematopoietic cell transplantation for hemoglobinopathies. Front Pharmacol *5*, 278.

Dias, C., Estruch, S.B., Graham, S.A., McRae, J., Sawiak, S.J., Hurst, J.A., Joss, S.K., Holder, S.E., Morton, J.E.V., Turner, C., et al. (2016). BCL11A Haploinsufficiency Causes an Intellectual Disability Syndrome and Dysregulates Transcription. Am. J. Hum. Genet. *99*, 253–274.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dobin, A., and Gingeras, T.R. (2016). Optimizing RNA-Seq Mapping with STAR. In Data Mining Techniques for the Life Sciences, (New York, NY: Humana Press, New York, NY), pp. 245–262.

Drissen, R., Lindern, von, M., Kolbus, A., Driegen, S., Steinlein, P., Beug, H., Grosveld, F., and Philipsen, S. (2005). The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability. Mol. Cell. Biol. *25*, 5205–5214.

Dukler, N., Gulko, B., Huang, Y.-F., and Siepel, A. (2016). Is a super-enhancer greater than the sum of its parts? Nature Publishing Group *49*, 2–3.

Dzierzak, E., and Philipsen, S. (2013). Erythropoiesis: development and differentiation. Cold Spring Harb Perspect Med *3*, a011601–a011601.

Ema, H., and Nakauchi, H. (2000). Expansion of hematopoietic stem cells in the developing liver of a mouse embryo. Blood *95*, 2284–2288.

Fei, Y.J., Kutlar, F., Harris, H.F., Wilson, M.M., Milana, A., Sciacca, P., Schiliro, G., Masala, B., Manca, L., and Altay, C. (1989). A search for anomalies in the zeta, alpha, beta, and gamma globin gene arrangements in normal black, Italian, Turkish, and Spanish newborns. Hemoglobin *13*, 45–65.

Felice, A.E., Cleek, M.P., Marino, E.M., McKie, K.M., McKie, V.C., Chang, B.K., and Huisman, T.H. (1986). Different zeta globin gene deletions among black Americans. Hum Genet *73*, 221–224.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic Reprogramming in Plant and Animal Development. Science *330*, 622–627.

Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. Mol. Cell. Biol. *25*, 1215–1227.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V.V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol. Cell. Biol. *16*, 2802–2813.

Flint, J., Hill, A.V., Bowden, D.K., Oppenheimer, S.J., Sill, P.R., Serjeantson, S.W., Bana-Koiri, J., Bhatia, K., Alpers, M.P., and Boyce, A.J. (1986). High frequencies of alpha-thalassaemia are the result of natural selection by malaria. Nature *321*, 744–750.

Fraser, S.T., Isern, J., and Baron, M.H. (2007). Maturation and enucleation of primitive erythroblasts during mouse embryogenesis is accompanied by changes in cell-surface antigen expression. Blood *109*, 343–352.

Fraser, S.T., Isern, J., and Baron, M.H. (2010). Use of transgenic fluorescent reporter mouse lines to monitor hematopoietic and erythroid development during embryogenesis. Meth. Enzymol. *476*, 403–427.

Fromm, G., Cadiz-Rivera, B., de Vries, C., Getman, M., McGrath, K.E., Kingsley, P.D., Fields, J., Fiering, S., and Bulger, M. (2011). An embryonic stage-specific enhancer within the murine β-globin locus mediates domain-wide histone hyperacetylation. Blood *117*, 5207–5214.

Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C., and Orkin, S.H. (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Pnas *93*, 12355–12358.

Funnell, A.P.W., Mak, K.S., Twine, N.A., Pelka, G.J., Norton, L.J., Radziewic, T., Power, M., Wilkins, M.R., Bell-Anderson, K.S., Fraser, S.T., et al. (2013). Generation of mice deficient in both KLF3/BKLF and KLF8 reveals a genetic interaction and a role for these factors in embryonic globin gene silencing. Mol. Cell. Biol. *33*, 2976–2987.

Garrick, D., Fiering, S., Martin, D.I., and Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. Nature Genetics *18*, 56–59.

Garrick, D., De Gobbi, M., Samara, V., Rugless, M., Holland, M., Ayyub, H., Lower, K., Sloane-Stanley, J., Gray, N., Koch, C., et al. (2008). The role of the polycomb complex in silencing alpha-globin gene expression in nonerythroid cells. Blood *112*, 3889–3899.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., et al. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature Biotechnology *26*, 317–325.

Gert-Jan Kremers, Joachim Goedhart, Erik B van Munster, A., and Theodorus W J Gadella, J. (2006). Cyan and Yellow Super Fluorescent Proteins with Improved Brightness, Protein Folding, and FRET Förster Radius†,‡. Biochemistry *45*, 6570–6580.

Gibbons, R.J. (2012). α-Thalassemia, Mental Retardation, and Myelodysplastic Syndrome. Cold Spring Harb Perspect Med *2*, a011759–a011759.

Gibson, B.A., Doolittle, L.K., Schneider, M.W.G., Jensen, L.E., Gamarra, N., Henry, L., Gerlich, D.W., Redding, S., and Rosen, M.K. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. Cell.

Goh, S.-H., Lee, Y.T., Bhanu, N.V., Cam, M.C., Desper, R., Martin, B.M., Moharram, R., Gherman, R.B., and Miller, J.L. (2005). A newly discovered human alpha-globin gene. Blood *106*, 1466–1472.

Gómez-Marín, C., Tena, J.J., Acemel, R.D., López-Mayorga, M., Naranjo, S., la Calle-Mustienes, de, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E., et al. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of

ancestral topological associating domains borders. Proc. Natl. Acad. Sci. U.S.a. *112*, 7542–7547.

Gudmundsdottir, B., Gudmundsson, K.O., Klarmann, K.D., Singh, S.K., Sun, L., Singh, S., Du, Y., Coppola, V., Stockwin, L., Nguyen, N., et al. (2018). POGZ Is Required for Silencing Mouse Embryonic β-like Hemoglobin and Human Fetal Hemoglobin Expression. Cell Rep *23*, 3236–3248.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.

Hajdu, S.I. (2003). A note from history: The discovery of blood cells. Ann Clin Lab Sci *33*, 237–238.

Han, H., Cortez, C.C., Yang, X., Nichols, P.W., Jones, P.A., and Liang, G. (2011). DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. Hum. Mol. Genet. *20*, 4299–4310.

Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Biggs, D., Preece, C., Downes, D.J., Gosden, M., Sharpe, J.A., Sloane-Stanley, J.A., Hughes, J.R., et al. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. Nat. Cell Biol. *19*, 952–961.

Hardison, R.C. (2012). Evolution of Hemoglobin and Its Genes. Cold Spring Harb Perspect Med *2*, a011627–a011627.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486–489.

Harteveld, C.L., and Higgs, D.R. (2010). α-thalassaemia. Orphanet Journal of Rare Diseases 2010 5:1 *5*, 13.

Hatton, C.S., Wilkie, A.O., Drysdale, H.C., Wood, W.G., Vickers, M.A., Sharpe, J., Ayyub, H., Pretorius, I.M., Buckle, V.J., and Higgs, D.R. (1990). Alpha-thalassemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster. Blood *76*, 221–227.

Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Sharpe, J.A., Suciu, M.C., et al. (2016). Genetic dissection of the [alpha]-globin super-enhancer in vivo. Nature Genetics *48*, 895–903.

He, Z., and Russell, J.E. (2001). Expression, purification, and characterization of human hemoglobins Gower-1 (ζ2ε2), Gower-2 (α2ε2), and Portland-2 (ζ2β2) assembled in complex transgenic–knockout mice. Blood *97*, 1099–1105.

He, Z., and Russell, J.E. (2004). Antisickling effects of an endogenous human |[alpha]|-like globin. Nature Medicine *10*, 365–367.

Heinrich, A.C., Pelanda, R., and Klingmüller, U. (2004). A mouse model for visualization and conditional mutations in the erythroid lineage. Blood *104*, 659–666.

Helsmoortel, H.H., Bresolin, S., Lammens, T., Cavé, H., Noellke, P., Caye, A., Ghazavi, F., de Vries, A., Hasle, H., Labarque, V., et al. (2016). LIN28B overexpression defines a novel fetal-like subgroup of juvenile myelomonocytic leukemia. Blood *127*, 1163–1172.

Higgs, D.R., Wood, W.G., Jarman, A.P., Sharpe, J., Lida, J., Pretorius, I.M., and Ayyub, H. (1990). A major positive regulatory region located far upstream of the human alpha-globin gene locus. Genes Dev. *4*, 1588–1601.

Higgs, D.R., Engel, J.D., and Stamatoyannopoulos, G. (2012). Thalassaemia. Lancet *379*, 373–383.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13–23.

Hong, W., Nakazawa, M., Chen, Y.-Y., Kori, R., Vakoc, C.R., Rakowski, C., and Blobel, G.A. (2005). FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. Embo J. *24*, 2367–2378.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

Hu, J., Liu, J., Xue, F., Halverson, G., Reid, M., Guo, A., Chen, L., Raza, A., Galili, N., Jaffray, J., et al. (2013). Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. Blood *121*, 3246–3253.

Huang, B.L., Fan-Chiang, I.R., Wen, S.C., Koo, H.C., Kao, W.Y., Gavva, N.R., and Shen, C.K. (1998). Derepression of human embryonic zeta-globin promoter by a locus-control region sequence. Pnas *95*, 14669–14674.

Hughes, J.R., Cheng, J.-F., Ventress, N., Prabhakar, S., Clark, K., Anguita, E., De Gobbi, M., de Jong, P., Rubin, E., and Higgs, D.R. (2005). Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc. Natl. Acad. Sci. U.S.a. *102*, 9830–9835.

HUTTON, J.J., BISHOP, J., SCHWEET, R., and RUSSELL, E.S. (1962). Hemoglobin inheritance in inbred mouse strains. II. Genetic studies. Pnas *48*, 1718–1724.

Ivanovs, A., Rybtsov, S., Welch, L., Anderson, R.A., Turner, M.L., and Medvinsky, A. (2011). Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. J. Exp. Med. *208*, 2417–2427.

Jatavan, P., Chattipakorn, N., and Tongsong, T. (2018). Fetal hemoglobin Bart's hydrops fetalis: pathophysiology, prenatal diagnosis and possibility of intrauterine treatment. J. Matern. Fetal. Neonatal. Med. *31*, 946–957.

Jeziorska, D.M., Murray, R.J.S., De Gobbi, M., Gaentzsch, R., Garrick, D., Ayyub, H., Chen, T., Li, E., Telenius, J., Lynch, M., et al. (2017). DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. Proc. Natl. Acad. Sci. U.S.a. *114*, E7526–E7535.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Johnson, K.D., Christensen, H.M., Zhao, B., and Bresnick, E.H. (2001). Distinct Mechanisms Control RNA Polymerase II Recruitment to a Tissue-Specific Locus Control Region and a Downstream Promoter. Mol. Cell *8*, 465–471.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics *13*, 484–492.

Kelly, R.D.W., and Cowley, S.M. (2013). The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. Biochem. Soc. Trans. *41*, 741–749.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Research *12*, 996–1006.

Kim, H., and Kim, J.-S. (2014). A guide to genome engineering with programmable nucleases. Nature Reviews Genetics *15*, 321–334.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231–1245.

Kingsley, P.D., Greenfest-Allen, E., Frame, J.M., Bushnell, T.P., Malik, J., McGrath, K.E., Stoeckert, C.J., and Palis, J. (2013). Ontogeny of erythroid gene expression. Blood *121*, e5–e13.

Kingsley, P.D., Malik, J., Emerson, R.L., Bushnell, T.P., McGrath, K.E., Bloedorn, L.A., Bulger, M., and Palis, J. (2006). “Maturational” globin switching in primary primitive erythroid cells. Blood *107*, 1665–1672.

Kingsley, P.D., Malik, J., Fantauzzo, K.A., and Palis, J. (2004). Yolk sac-derived primitive erythroblasts enucleate during mammalian embryogenesis. Blood *104*, 19–25.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Keith Joung, J. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends Biochem. Sci. *31*, 89–97.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

Kowalczyk, M.S., Hughes, J.R., Babbs, C., Sanchez-Pulido, L., Szumska, D., Sharpe, J.A., Sloane-Stanley, J.A., Morriss-Kay, G.M., Smoot, L.B., Roberts, A.E., et al. (2012a). Nprl3 is required for normal development of the cardiovascular system. Mamm. Genome *23*, 404–415.

Kowalczyk, M.S., Hughes, J.R., Garrick, D., Lynch, M.D., Sharpe, J.A., Sloane-Stanley, J.A., McGowan, S.J., De Gobbi, M., Hosseini, M., Vernimmen, D., et al. (2012b). Intragenic enhancers act as alternative promoters. Mol. Cell *45*, 447–458.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Kumaravelu, P., Hook, L., Morrison, A.M., Ure, J., Zhao, S., Zuyev, S., Ansell, J., and Medvinsky, A. (2002). Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. Development *129*, 4891–4899.

Kurita, R., Suda, N., Sudo, K., Miharada, K., Hiroyama, T., Miyoshi, H., Tani, K., and Nakamura, Y. (2013). Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells. PLoS ONE *8*, e59890.

Kühn, R., Schwenk, F., Aguet, M., and Rajewsky, K. (1995). Inducible gene targeting in mice. Science *269*, 1427–1429.

Labuhn, M., Perkins, K., Matzk, S., Varghese, L., Garnett, C., Papaemmanuil, E., Metzner, M., Kennedy, A., Amstislavskiy, V., Risch, T., et al. (2019). Mechanisms of Progression of Myeloid Preleukemia to Transformed Myeloid Leukemia in Children with Down Syndrome. Cancer Cell.

Lai, A.Y., and Wade, P.A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nat Rev Cancer *11*, 588–596.

Lai, K., Li, S., Lin, W., Yang, D., Chen, W., Li, M., Pang, L., and Chen, P. (2018). Invasive prenatal diagnosis of α-thalassemia to control Hb Bart's hydrops fetalis syndrome: 15 years of experience. Arch. Gynecol. Obstet. *73*, 1081–1085.

Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E., and Corbett, A.H. (2007). Classical nuclear localization signals: definition, function, and interaction with importin alpha. J. Biol. Chem. *282*, 5101–5105.

Lapillonne, H., Kobari, L., Mazurier, C., Tropel, P., Giarratana, M.-C., Zanella-Cleon, I., Kiger, L., Wattenhofer-Donzé, M., Puccio, H., Hebert, N., et al. (2010). Red blood cell generation from human induced pluripotent stem cells: perspectives for transfusion medicine. Haematologica *95*, 1651–1659.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Laugesen, A., and Helin, K. (2014). Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. Cell Stem Cell *14*, 735–751.

Leder, A., Daugherty, C., Whitney, B., and Leder, P. (1997). Mouse zeta- and alpha-globin genes: embryonic survival, alpha-thalassemia, and genetic background effects. Blood *90*, 1275–1282.

Lee, Y.T., de Vasconcellos, J.F., Yuan, J., Byrnes, C., Noh, S.-J., Meier, E.R., Kim, K.S., Rabel, A., Kaushal, M., Muljo, S.A., et al. (2013). LIN28B-mediated expression of fetal hemoglobin and production of fetal-like erythrocytes from adult human erythroblasts ex vivo. Blood *122*, 1034–1041.

Lemmens-Zygulska, M., Eigel, A., Helbig, B., Sanguansermsri, T., Horst, J., and Flatz, G. (1996). Prevalence of α-thalassemias in northern Thailand. Hum Genet *98*, 345–347.

Leonard, M., Brice, M., Engel, J.D., and Papayannopoulou, T. (1993). Dynamics of GATA transcription factor expression during erythroid differentiation. Blood *82*, 1071–1079.

Lettre, G., Sankaran, V.G., Bezerra, M.A.C., Araújo, A.S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., Costa, F.F., Hirschhorn, J.N., et al. (2008). DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc. Natl. Acad. Sci. U.S.a. *105*, 11869–11874.

Leung, S.-O., Proudfoot, N.J., and Whitelaw, E. (1987). The gene for |[thetas]|-globin is transcribed in human fetal erythroid tissues. Nature *329*, 551–554.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, T.K.T., Leung, K.Y., Lam, Y.H., Tang, M.H.Y., and Chan, V. (2010). Haemoglobin level, proportion of haemoglobin Bart"s and haemoglobin Portland in fetuses affected by homozygous α0-thalassemia from 12 to 40 weeks" gestation. Prenatal Diagnosis *30*, 1126–1130.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Liebhaber, S.A., Wang, Z., Cash, F.E., Monks, B., and Russell, J.E. (1996). Developmental silencing of the embryonic zeta-globin gene: concerted action of the

promoter and the 3'-flanking region combined with stage-specific silencing by the transcribed segment. Mol. Cell. Biol. *16*, 2637–2646.

Limaye, A., Hall, B., and Kulkarni, A.B. (2009). Manipulation of Mouse Embryonic Stem Cells for Knockout Mouse Production. Current Protocols in Cell Biology *44*, 19.13.1–19.13.24.

Liu, N., Hargreaves, V.V., Zhu, Q., Kurland, J.V., Hong, J., Kim, W., Sher, F., Macias-Trevino, C., Rogers, J.M., Kurita, R., et al. (2018). Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. Cell *173*, 430–442.e17.

Liu, P., Keller, J.R., Ortiz, M., Tessarollo, L., Rachel, R.A., Nakamura, T., Jenkins, N.A., and Copeland, N.G. (2003). Bcl11a is essential for normal lymphoid development. Nat. Immunol. *4*, 525–532.

Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene *5*, 1743–1753.

Lozzio, B.B., and Lozzio, C.B. (1977). Properties of the K562 cell line derived from a patient with chronic myeloid leukemia. Int. J. Cancer *19*, 136.

Lozzio, C.B., and Lozzio, B.B. (1975). Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. Blood *45*, 321–334.

Lunardi, A., Guarnerio, J., Wang, G., Maeda, T., and Pandolfi, P.P. (2013). Role of LRF/Pokemon in lineage fate decisions. Blood *121*, 2845–2853.

Luo, H.Y., Clarke, B.J., Gauldie, J., Patterson, M., Liao, S.K., and Chui, D.H. (1988). A novel monoclonal antibody based diagnostic test for alpha-thalassemia-1 carriers due to the (-SEA/) deletion. Blood *72*, 1589–1594.

Luo, H.Y., Liang, X.L., Frye, C., Wonio, M., Hankins, G.D., Chui, D.H., and Alter, B.P. (1999). Embryonic hemoglobins are expressed in definitive cells. Blood *94*, 359–361.

Maeda, T., Merghoub, T., Hobbs, R.M., Dong, L., and Maeda, M. (2007). … Shigematsu H, Akashi K, Teruya-Feldstein J, Cattoretti G, pandolfi PP: Regulation of B versus T lymphoid lineage fate decision by the proto-oncogene LRF (Science).

Maeda, T., Ito, K., Merghoub, T., Poliseno, L., Hobbs, R.M., Wang, G., Dong, L., Maeda, M., Dore, L.C., Zelent, A., et al. (2009). LRF is an essential downstream target of GATA1 in erythroid development and regulates BIM-dependent apoptosis. Dev. Cell *17*, 527–540.

Magbanua, J.P., Runneburger, E., Russell, S., and White, R. (2015). A variably occupied CTCF binding site in the ultrabithorax gene in the Drosophila bithorax complex. Mol. Cell. Biol. *35*, 318–330.

Manning, L.R., Russell, J.E., Padovan, J.C., Chait, B.T., Popowicz, A., Manning, R.S., and Manning, J.M. (2007). Human embryonic, fetal, and adult hemoglobins have different subunit interface strengths. Correlation with lifespan in the red cell. Protein Science *16*, 1641–1658.

Mao, X., Fujiwara, Y., and Orkin, S.H. (1999). Improved reporter strain for monitoring Cre recombinase-mediated DNA excisions in mice. Pnas *96*, 5037–5042.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Martyn, G.E., Wienert, B., Yang, L., Shah, M., Norton, L.J., Burdach, J., Kurita, R., Nakamura, Y., Pearson, R.C.M., Funnell, A.P.W., et al. (2018). Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. Nature Publishing Group *50*, 498–503.

Masuda, T., Wang, X., Maeda, M., Canver, M.C., Sher, F., Funnell, A.P.W., Fisher, C., Suciu, M., Martyn, G.E., Norton, L.J., et al. (2016). Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. Science *351*, 285–289.

Maurer, H.S., Behrman, R.E., and Honig, G.R. (1970). Dependence of the oxygen affinity of blood on the presence of foetal or adult haemoglobin. Nature *227*, 388–390.

Mayr, F., and Heinemann, U. (2013). Mechanisms of Lin28-mediated miRNA and mRNA regulation--a structural and functional perspective. Int J Mol Sci *14*, 16532–16553.

McGrath, K.E., Frame, J.M., Fromm, G.J., Koniski, A.D., Kingsley, P.D., Little, J., Bulger, M., and Palis, J. (2011). A transient definitive erythroid lineage with unique regulation of the β-globin locus in the mammalian embryo. Blood *117*, 4600–4608.

Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., Foglio, M., Zelenika, D., Boland, A., Rooks, H., et al. (2007). A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nature Publishing Group *39*, 1197–1199.

Merryweather-Clarke, A.T., Atzberger, A., Soneji, S., Gray, N., Clark, K., Waugh, C., McGowan, S.J., Taylor, S., Nandi, A.K., Wood, W.G., et al. (2011). Global gene expression analysis of human erythroid progenitors. Blood *117*, e96–e108.

Mettananda, S., Fisher, C.A., Hay, D., Badat, M., Quek, L., Clark, K., Hublitz, P., Downes, D., Kerry, J., Gosden, M., et al. (2017a). Editing an α-globin enhancer in primary human hematopoietic stem cells as a treatment for β-thalassemia. Nat Commun *8*, 424.

Mettananda, S., Fisher, C.A., Sloane-Stanley, J.A., Taylor, S., Oppermann, U., Gibbons, R.J., and Higgs, D.R. (2017b). Selective silencing of α-globin by the histone demethylase inhibitor IOX1: a potentially new pathway for treatment of β-thalassemia. Haematologica *102*, e80–e84.

Miccio, A., and Blobel, G.A. (2010). Role of the GATA-1/FOG-1/NuRD pathway in the expression of human beta-like globin genes. Mol. Cell. Biol. *30*, 3460–3470.

Migliaccio, G., Migliaccio, A.R., Petti, S., Mavilio, F., Russo, G., Lazzaro, D., Testa, U., Marinucci, M., and Peschle, C. (1986). Human embryonic hemopoiesis. Kinetics of progenitors and precursors underlying the yolk sac----liver transition. J. Clin. Invest. *78*, 51–60.

Migliaccio, G., Di Pietro, R., di Giacomo, V., Di Baldassarre, A., Migliaccio, A.R., Maccioni, L., Galanello, R., and Papayannopoulou, T. (2002). In vitro mass production of human erythroid cells from the blood of normal donors and of thalassemic patients. Blood Cells Mol. Dis. *28*, 169–180.

Moir-Meyer, G., Cheong, P.L., Olijnik, A.-A., Brown, J., Knight, S., King, A., Kurita, R., Nakamura, Y., Gibbons, R.J., Higgs, D.R., et al. (2018). Robust CRISPR/Cas9 Genome Editing of the HUDEP-2 Erythroid Precursor Line Using Plasmids and Single-Stranded Oligonucleotide Donors. Methods and Protocols 2018, Vol. 1, Page 28 *1*, 28.

Münger, K., and Howley, P.M. (2002). Human papillomavirus immortalization and transformation functions. Virus Res. *89*, 213–228.

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nature Biotechnology *20*, 87–90.

Neijts, R., and Deschamps, J. (2017). At the base of colinear Hox gene expression: cis-features and trans-factors orchestrating the initial phase of Hox cluster activation. Dev. Biol. *428*, 293–299.

Norton, L.J., Funnell, A.P.W., Burdach, J., Wienert, B., Kurita, R., Nakamura, Y., Philipsen, S., Pearson, R.C.M., Quinlan, K.G.R., and Crossley, M. (2017). KLF1 directly activates expression of the novel fetal globin repressor ZBTB7A/LRF in erythroid cells. Blood Advances *1*, 685–692.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L.A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc. Natl. Acad. Sci. U.S.a. *115*, E6697–E6706.

Oji, A., Noda, T., Fujihara, Y., Miyata, H., Kim, Y.J., Muto, M., Nozawa, K., Matsumura, T., Isotani, A., and Ikawa, M. (2016). CRISPR/Cas9 mediated genome editing in ES cells and its application for chimeric analysis in mice. Sci Rep *6*, 31666.

Olivier, E.N., Marenah, L., McCahill, A., Condie, A., Cowan, S., and Mountford, J.C. (2016). High-Efficiency Serum-Free Feeder-Free Erythroid Differentiation of Human Pluripotent Stem Cells Using Small Molecules. Stem Cells Transl Med *5*, 1394–1405.

Olivier, E.N., Qiu, C., Velho, M., Hirsch, R.E., and Bouhassira, E.E. (2006). Large-scale production of embryonic red blood cells from human embryonic stem cells. Exp. Hematol. *34*, 1635–1642.

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243.

Oudelaar, A.M., Hanssen, L.L.P., Hardison, R.C., Kassouf, M.T., Hughes, J.R., and Higgs, D.R. (2017). Between form and function: the complexity of genome folding. Hum. Mol. Genet. *26*, R208–R215.

Palis, J., Chan, R.J., Koniski, A., Patel, R., Starr, M., and Yoder, M.C. (2001). Spatial and temporal emergence of high proliferative potential hematopoietic precursors during murine embryogenesis. Pnas *98*, 4528–4533.

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

Palis, J. (2014). Primitive and definitive erythropoiesis in mammals. Front Physiol *5*, 3.

Palstra, R.-J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nature Genetics *35*, 190–194.

Papadopoulos, P., Gutiérrez, L., van der Linden, R., Kong-A-San, J., Maas, A., Drabek, D., Patrinos, G.P., Philipsen, S., and Grosveld, F. (2012). A Dual Reporter Mouse Model of the Human β-Globin Locus: Applications and Limitations. PLoS ONE *7*, e51272.

Papayannopoulou, T., Nakamoto, B., Anagnou, N.P., Chui, D., Dow, L., and Sanders, J. (1991). Expression of embryonic globins by erythroid cells in juvenile chronic myelocytic leukemia. Blood *77*, 2569–2576.

Paredes, S.H., Melgar, M.F., and Sethupathy, P. (2013). Promoter-proximal CCCTC-factor binding is associated with an increase in the transcriptional pausing index. Bioinformatics *29*, 1485–1487.

Peschle, C., Mavilio, F., Care, A., Migliaccio, G., Migliaccio, A.R., Salvo, G., Samoggia, P., Petti, S., Guerriero, R., Marinucci, M., et al. (1985). Haemoglobin switching in human embryos: asynchrony of |[zeta]| |[alpha]| and |[epsi]| |[gamma]|-globin switches in primitive and definitive erythropoietic lineage. Nature *313*, 235–238.

Piel, F.B., and Weatherall, D.J. (2014). The α-Thalassemias. NEJM. Http://Dx.Doi.org/10.1056/NEJMra1404415 *371*, 1908–1916.

Piel, F.B., Steinberg, M.H., and Rees, D.C. (2017). Sickle Cell Disease. NEJM. Http://Dx.Doi.org/10.1056/NEJMra1510865 *376*, 1561–1573.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440–455.

Porcu, S., Kitamura, M., Witkowska, E., Zhang, Z., Mutero, A., Lin, C., Chang, J., and Gaensler, K.M. (1997). The human beta globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. Blood *90*, 4602–4609.

Pott, S., and Lieb, J.D. (2015). What are super-enhancers? Nature Publishing Group *47*, 8–12.

Proudfoot, N.J., Gil, A., and Maniatis, T. (1982). The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell *31*, 553–563.

Qiu, P., Shandilya, H., D'Alessio, J.M., O'Connor, K., Durocher, J., and Gerard, G.F. (2004). Mutation detection using Surveyor™ nuclease. Biotechniques 36, 702-707.

Quinlan, A.R., Clark, R.A., Sokolova, S., Leibowitz, M.L., Zhang, Y., Hurles, M.E., Mell, J.C., and Hall, I.M. (2010). Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Research *20*, 623–635.

Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: How is it established, and why does it matter? Dev. Biol. *339*, 258–266.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res *42*, W187–W191.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature Protocols *8*, 2281–2308.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

Rebuzzini, P., Zuccotti, M., Redi, C.A., and Garagna, S. (2015). Chromosomal Abnormalities in Embryonic and Somatic Stem Cells. Cytogenet. Genome Res. *147*, 1–9.

Renneville, A., Van Galen, P., Canver, M.C., McConkey, M., Krill-Burger, J.M., Dorfman, D.M., Holson, E.B., Bernstein, B.E., Orkin, S.H., Bauer, D.E., et al. (2015). EHMT1 and EHMT2 inhibition induces fetal hemoglobin expression. Blood *126*, 1930–1939.

Ronzoni, L., Bonara, P., Rusconi, D., Frugoni, C., Libani, I., and Cappellini, M.D. (2008). Erythroid differentiation and maturation from peripheral CD34+ cells in liquid culture: cellular and molecular characterization. Blood Cells Mol. Dis. *40*, 148–155.

Rose, N.R., and Klose, R.J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. Biochim. Biophys. Acta *1839*, 1362–1372.

Russell, J.E., and Liebhaber, S.A. (1998). Reversal of Lethal □- and β-Thalassemias in Mice by Expression of Human Embryonic Globins. Blood *92*, 3057–3063.

Rutherford, T., Clegg, J.B., Higgs, D.R., Jones, R.W., Thompson, J., and Weatherall, D.J. (1981). Embryonic erythroid differentiation in the human leukemic cell line K562. Pnas *78*, 348–352.

Ryan, K., Bain, B.J., Worthington, D., James, J., Plews, D., Mason, A., Roper, D., Rees, D.C., la Salle, de, B., Streetly, A., et al. (2010). Significant haemoglobinopathies: guidelines for screening and diagnosis. British Journal of Haematology *149*, 35–49.

Ryan, M.D., King, A.M.Q., and Thomas, G.P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. Journal of General Virology *72*, 2727–2732.

Sabath, D.E., Koehler, K.M., Yang, W.Q., Patton, K., and Stamatoyannopoulos, G. (1995). Identification of a major positive regulatory element located 5' to the human zeta-globin gene. Blood *85*, 2587–2597.

Sabath, D.E., Spangler, E.A., Rubin, E.M., and Stamatoyannopoulos, G. (1993). Analysis of the human zeta-globin gene promoter in transgenic mice. Blood *82*, 2899–2905.

Sabath, D.E., Koehler, K.M., and Yang, W.-Q. (1996). Structure and Function of the ζ-globin Upstream Regulatory Element. Nucleic Acids Res *24*, 4978–4986.

Safo, M. K., Ko, T.-P., Schreiter, E. R., & Russell, J. E. (2015). Structural basis for the antipolymer activity of Hb ζ2βs2 trapped in a tense conformation. *Journal of Molecular Structure*, *1099*, 99–107.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nature Reviews Molecular Cell Biology *7*, 437–447.

Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.-I.T., and Yamamoto, T. (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. Nature Protocols *11*, 118–133.

Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. U.S.a. *112*, E6456–E6465.

Sankaran, V.G., Menne, T.F., Šćepanović, D., Vergilio, J.-A., Ji, P., Kim, J., Thiru, P., Orkin, S.H., Lander, E.S., and Lodish, H.F. (2011). MicroRNA-15a and -16-1 act

via MYB to elevate fetal hemoglobin expression in human trisomy 13. Proc. Natl. Acad. Sci. U.S.a. *108*, 1519–1524.

Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K.A., Hirschhorn, J.N., Cantor, A.B., and Orkin, S.H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science *322*, 1839–1842.

Sankaran, V.G., Xu, J., Ragoczy, T., Ippolito, G.C., Walkley, C.R., Maika, S.D., Fujiwara, Y., Ito, M., Groudine, M., Bender, M.A., et al. (2009). Developmental and species-divergent globin switching are driven by BCL11A. Nature *460*, 1093–1097.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Meth *9*, 676–682.

Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nature Reviews Molecular Cell Biology *12*, 799–814.

Scott, C., Downes, D.J., Brown, J.M., Babbs, C., Olijnik, A.-A., Gosden, M., Beagrie, R., Schwessinger, R., Fisher, C.A., Rose, A., et al. (2019). Modelling erythropoiesis in congenital dyserythropoietic anaemia type I (CDA-I). bioRxiv *185*, 744367.

Seto, E., and Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb Perspect Biol *6*, a018713–a018713.

Shah, M., Funnell, A.P.W., Quinlan, K.G.R., and Crossley, M. (2019). Hit and Run Transcriptional Repressors Are Difficult to Catch in the Act. Bioessays *41*, e1900041.

Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005). A guide to choosing fluorescent proteins. Nat Meth *2*, 905–909.

Sharpe, J.A., Chan-Thomas, P.S., Lida, J., Ayyub, H., Wood, W.G., and Higgs, D.R. (1992). Analysis of the human alpha globin upstream regulatory element (HS-40) in transgenic mice. Embo J. *11*, 4565–4572.

Shearstone, J.R., Pop, R., Bock, C., Boyle, P., Meissner, A., and Socolovsky, M. (2011). Global DNA demethylation during mouse erythropoiesis in vivo. Science *334*, 799–802.

Sher, F., Hossain, M., Seruggia, D., Schoonenberg, V.A.C., Yao, Q., Cifani, P., Dassama, L.M.K., Cole, M.A., Ren, C., Vinjamur, D.S., et al. (2019). Rational targeting of a NuRD subcomplex guided by comprehensive in situ mutagenesis. Nature Publishing Group *51*, 1149–1159.

Shi, Y., Sawada, J.-I., Sui, G., Affar, E.B., Whetstine, J.R., Lan, F., Ogawa, H., Luke, M.P.-S., Nakatani, Y., and Shi, Y. (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. Nature *422*, 735–738.

Siatecka, M., and Bieker, J.J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. Blood *118*, 2044–2054.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*, 576.

Smith, E.C., Luc, S., Croney, D.M., Woodworth, M.B., Greig, L.C., Fujiwara, Y., Nguyen, M., Sher, F., Macklis, J.D., Bauer, D.E., et al. (2016). Strict in vivo specificity of the Bcl11a erythroid enhancer. Blood *128*, 2338–2342.

Songdej, D., Babbs, C., and Higgs, D.R. (2017). An international registry of survivors with Hb Bart9s Hydrops Fetalis Syndrome. Blood *129*, blood–2016–08–697110–1259.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics *13*, 613–626.

Spivak, J.L., Toretti, D., and Dickerman, H.W. (1973). Effect of Phenylhydrazine-induced Hemolytic Anemia on Nuclear RNA Polymerase Activity of the Mouse Spleen. Blood *42*, 257–266.

Stanya, K.J., and Kao, H.-Y. (2009). New insights into the functions and regulation of the transcriptional corepressors SMRT and N-CoR. Cell Div *4*, 7–8.

Steinberg, M.H., and Rodgers, G.P. (2015). HbA2 : biology, clinical relevance and a possible target for ameliorating sickle cell disease. British Journal of Haematology *170*, 781–787.

Straub, A.C., Butcher, J.T., Billaud, M., Mutchler, S.M., Artamonov, M.V., Nguyen, A.T., Johnson, T., Best, A.K., Miller, M.P., Palmer, L.A., et al. (2014). Hemoglobin α/eNOS Coupling at Myoendothelial Junctions Is Required for Nitric Oxide Scavenging During Vasoconstriction. Arterioscler. Thromb. Vasc. Biol. *34*, 2594–2600.

Straub, A.C., Lohman, A.W., Billaud, M., Johnstone, S.R., Dwyer, S.T., Lee, M.Y., Bortz, P.S., Best, A.K., Columbus, L., Gaston, B., et al. (2012). Endothelial cell expression of haemoglobin α regulates nitric oxide signalling. Nature *491*, 473–477.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev. *19*, 815–826.

Tallack, M.R., Whitington, T., Yuen, W.S., Wainwright, E.N., Keys, J.R., Gardiner, B.B., Nourbakhsh, E., Cloonan, N., Grimmond, S.M., Bailey, T.L., et al. (2010). A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome Research *20*, 1052–1063.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. Cell *146*, 1016–1028.

Tang, W., Cai, S.P., Eng, B., Poon, M.C., Waye, J.S., Illum, N., and Chui, D.H. (1993). Expression of embryonic zeta-globin and epsilon-globin chains in a 10-year-old girl with congenital anemia. Blood *81*, 1636–1640.

Tebas, P., Stein, D., Tang, W.W., Frank, I., Wang, S.Q., Lee, G., Spratt, S.K., Surosky, R.T., Giedlin, M.A., Nichol, G., et al. (2014). Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. N. Engl. J. Med. *370*, 901–910.

Telenius, J., Consortium, T.W., and Hughes, J.R. (2018). NGseqBasic - a single-command UNIX tool for ATAC-seq, DNaseI-seq, Cut-and-Run, and ChIP-seq data mapping, high-resolution visualisation, and quality control. bioRxiv *100*, 393413.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tober, J., McGrath, K.E., and Palis, J. (2008). Primitive erythropoiesis and megakaryopoiesis in the yolk sac are independent of c-myb. Blood *111*, 2636–2639.

Trakarnsanga, K., Griffiths, R.E., Wilson, M.C., Blair, A., Satchwell, T.J., Meinders, M., Cogan, N., Kupzig, S., Kurita, R., Nakamura, Y., et al. (2017). An immortalized adult human erythroid line facilitates sustainable and scalable generation of functional red cells. Nat Commun *8*, 14750.

Traxler, E.A., Yao, Y., Wang, Y.-D., Woodard, K.J., Kurita, R., Nakamura, Y., Hughes, J.R., Hardison, R.C., Blobel, G.A., Li, C., et al. (2016). A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. Nature Medicine *22*, 987–990.

Tuchinda, S., Nagai, K., and Lehmann, H. (1975). Oxygen dissociation curve of haemoglobin Portland. FEBS Lett. *49*, 390–391.

Tufarelli, C., Stanley, J.A.S., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G., and Higgs, D.R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nature Genetics *34*, 157–165.

Turner, J., and Crossley, M. (1998). Cloning and characterization of mCtBP2, a co-repressor that associates with basic Krüppel-like factor and other mammalian transcriptional regulators. Embo J. *17*, 5129–5140.

Tusi, B. K., Wolock, S. L., Weinreb, C., Hwang, Y., Hidalgo, D., Zilionis, R., et al. (2018). Population snapshots predict early haematopoietic and erythroid hierarchies. Nature, *555*, 54–60.

Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V.G., Chen, W., Usala, G., Busonero, F., Maschio, A., Albai, G., et al. (2008). Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc. Natl. Acad. Sci. U.S.a. *105*, 1620–1625.

Vacaru, A.M., Isern, J., Fraser, S.T., and Baron, M.H. (2013). Analysis of primitive erythroid cell proliferation and enucleation using a cyan fluorescent reporter in transgenic mice. Genesis *51*, 751–762.

Van Handel, B., Prashad, S.L., Hassanzadeh-Kiabi, N., Huang, A., Magnusson, M., Atanassova, B., Chen, A., Hamalainen, E.I., and Mikkola, H.K.A. (2010). The first trimester human placenta is a site for terminal maturation of primitive erythroid cells. Blood *116*, 3321–3330.

Vanuytsel, K., Matte, T., Leung, A., Naing, Z.H., Morrison, T., Chui, D.H.K., Steinberg, M.H., and Murphy, G.J. (2018). Induced pluripotent stem cell-based mapping of β-globin expression throughout human erythropoietic development. Blood Advances *2*, 1998–2011.

Varga-Weisz, P. (2001). ATP-dependent chromatin remodeling factors: Nucleosome shufflers with many missions. Oncogene *20*, 3076–3085.

Verdone, L., Agricola, E., Caserta, M., and Di Mauro, E. (2006). Histone acetylation in gene regulation. Brief Funct Genomic Proteomic *5*, 209–221.

Vernimmen, D. (2014). Uncovering Enhancer Functions Using the α-Globin Locus. PLoS Genet. *10*, e1004668.

Vernimmen, D., Lynch, M.D., De Gobbi, M., Garrick, D., Sharpe, J.A., Sloane-Stanley, J.A., Smith, A.J.H., and Higgs, D.R. (2011). Polycomb eviction as a new distant enhancer function. Genes Dev. *25*, 1583–1588.

Vernimmen, D., Marques-Kranc, F., Sharpe, J.A., Sloane-Stanley, J.A., Wood, W.G., Wallace, H.A.C., Smith, A.J.H., and Higgs, D.R. (2009). Chromosome looping at the human alpha-globin locus is mediated via the major upstream regulatory element (HS -40). Blood *114*, 4253–4260.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep *10*, 1297–1309.

Vinjamur, D.S., and Bauer, D.E. (2018). Growing and Genetically Manipulating Human Umbilical Cord Blood-Derived Erythroid Progenitor (HUDEP) Cell Lines. Methods Mol. Biol. *1698*, 275–284.

Viprakasit, V., Ekwattanakit, S., Riolueang, S., Chalaow, N., Fisher, C., Lower, K., Kanno, H., Tachavanich, K., Bejrachandra, S., Saipin, J., et al. (2014a). Mutations in Kruppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression. Blood *123*, 1586–1595.

Viprakasit, V., Ekwattanakit, S., Riolueang, S., Chalaow, N., Fisher, C., Lower, K., Kanno, H., Tachavanich, K., Bejrachandra, S., Saipin, J., et al. (2014b). Mutations in Krüppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression. Blood *123*, 1586–1595.

Viprakasit, V., Harteveld, C.L., Ayyub, H., Stanley, J.S., Giordano, P.C., Wood, W.G., and Higgs, D.R. (2006). A novel deletion causing alpha thalassemia clarifies

the importance of the major human alpha globin regulatory element. Blood *107*, 3811–3812.

Vostrov, A.A., and Quitschke, W.W. (1997). The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. J. Biol. Chem. *272*, 33353–33359.

Vyas, P., Sharpe, J.A., Watt, P., Higgs, D.R., and Wood, W.G. (1992). Regulation of human embryonic globin genes zeta 2 and epsilon in stably transformed mouse erythroleukemia cells. Blood *80*, 1832–1837.

Wajcman, H., and Galacteros, F. (1996). Abnormal hemoglobins with high oxygen affinity and erythrocytosis. Hematol Cell Ther *38*, 305–312.

Wallace, H.A.C., Marques-Kranc, F., Richardson, M., Luna-Crespo, F., Sharpe, J.A., Hughes, J., Wood, W.G., Higgs, D.R., and Smith, A.J.H. (2007). Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. Cell *128*, 197–209.

Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Research *22*, 1680–1688.

Ware, R.E., de Montalembert, M., Tshilolo, L., and Abboud, M.R. (2017). Sickle cell disease. The Lancet.

Weatherall, D.J., Clegg, J.B., and Naughton, M.A. (1965). Globin Synthesis in Thalassaemia: An in vitro Study. Nature *208*, 1061–1065.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature Genetics *37*, 853–862.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796–801.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wilkie, A.O., Buckle, V.J., Harris, P.C., Lamb, J., Barton, N.J., Reeders, S.T., Lindenbaum, R.H., Nicholls, R.D., Barrow, M., and Bethlenfalvay, N.C. (1990). Clinical features and molecular analysis of the alpha thalassemia/mental retardation syndromes. I. Cases due to deletions involving chromosome band 16p13.3. The American Journal of Human Genetics *46*, 1112–1126.

Winichagoon, P., Higgs, D.R., Goodbourn, S.E., Lamb, J., Clegg, J.B., and Weatherall, D.J. (1982). Multiple arrangements of the human embryonic zeta globin genes. Nucleic Acids Res *10*, 5853–5868.

Witt, R.G., Kreger, E.M., Buckman, L.B., Moradi, P.W., Ho, P.T., Derderian, S.C., Tsai, P., Baker, C., Schramm, N., Cleary, R., et al. (2018). Systemic multilineage engraftment in mice after in utero transplantation with human hematopoietic stem cells. Blood Advances *2*, 69–74.

Wlodkowic, D., Skommer, J., and Darzynkiewicz, Z. (2009). Flow cytometry-based apoptosis detection. Methods Mol. Biol. *559*, 19–32.

Wu, Y., Zeng, J., Roscoe, B.P., Liu, P., Yao, Q., Lazzarotto, C.R., Clement, K., Cole, M.A., Luk, K., Baricordi, C., et al. (2019). Highly efficient therapeutic gene editing of human hematopoietic stem cells. Nature Medicine *25*, 776–783.

Xu, J., Bauer, D.E., Kerenyi, M.A., Vo, T.D., Hou, S., Hsu, Y.-J., Yao, H., Trowbridge, J.J., Mandel, G., and Orkin, S.H. (2013). Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proc. Natl. Acad. Sci. U.S.a. *110*, 6518–6523.

Xu, J., Peng, C., Sankaran, V.G., Shao, Z., Esrick, E.B., Chong, B.G., Ippolito, G.C., Fujiwara, Y., Ebert, B.L., Tucker, P.W., et al. (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science *334*, 993–996.

Xu, J., Sankaran, V.G., Ni, M., Menne, T.F., Puram, R.V., Kim, W., and Orkin, S.H. (2010). Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev. *24*, 783–798.

Xu, J., Shao, Z., Glass, K., Bauer, D.E., Pinello, L., Van Handel, B., Hou, S., Stamatoyannopoulos, J.A., Mikkola, H.K.A., Yuan, G.-C., et al. (2012). Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. Dev. Cell *23*, 796–811.

Yagi, M., Gelinas, R., Elder, J.T., Peretz, M., Papayannopoulou, T., Stamatoyannopoulos, G., and Groudine, M. (1986). Chromatin structure and developmental expression of the human alpha-globin cluster. Mol. Cell. Biol. *6*, 1108–1116.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370–1379.

Yi, Z., Cohen-Barak, O., Hagiwara, N., Kingsley, P.D., Fuchs, D.A., Erickson, D.T., Epner, E.M., Palis, J., and Brilliant, M.H. (2006). Sox6 directly silences epsilon globin expression in definitive erythropoiesis. PLoS Genet. *2*, e14.

Yien, Y.Y., and Bieker, J.J. (2013). EKLF/KLF1, a tissue-restricted integrator of transcriptional control, chromatin remodeling, and lineage determination. Mol. Cell. Biol. *33*, 4–13.

Yu, C.Y., Chen, J., Lin, L.I., Tam, M., and Shen, C.K. (1990). Cell type-specific protein-DNA interactions in the human zeta-globin upstream promoter region: displacement of Sp1 by the erythroid cell-specific factor NF-E1. Mol. Cell. Biol. *10*, 282–294.

Yu, M., Riva, L., Xie, H., Schindler, Y., Moran, T.B., Cheng, Y., Yu, D., Hardison, R., Weiss, M.J., Orkin, S.H., et al. (2009). Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. Mol. Cell *36*, 682–695.

Yu, Y., Wang, J., Khaled, W., Burke, S., Li, P., Chen, X., Yang, W., Jenkins, N.A., Copeland, N.G., Zhang, S., et al. (2012). Bcl11a is essential for lymphoid development and negatively regulates p53. J. Exp. Med. *209*, 2467–2483.

Zacher, B., Michel, M., Schwalb, B., Cramer, P., Tresch, A., and Gagneur, J. (2017). Accurate Promoter and Enhancer Identification in 127 ENCODE and Roadmap Epigenomics Cell Types and Tissues by GenoSTAN. PLoS ONE *12*, e0169249.

Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. Science *328*, 916–919.

Zhang, Q., Reddy, P.M., Yu, C.Y., Bastiani, C., Higgs, D., Stamatoyannopoulos, G., Papayannopoulou, T., and Shen, C.K. (1993). Transcriptional activation of human zeta 2 globin promoter by the alpha globin regulatory element (HS-40): functional role of specific nuclear factor-DNA complexes. Mol. Cell. Biol. *13*, 2298–2308.

Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications – writers that read. EMBO Reports *16*, 1467–1481.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, 1–9.

Zhao, B., Mei, Y., Yang, J., and Ji, P. (2014). Mouse Fetal Liver Culture System to Dissect Target Gene Functions at the Early and Late Stages of Terminal Erythropoiesis. JoVE (Journal of Visualized Experiments) e51894.

Zhou, D., Liu, K., Sun, C.-W., Pawlik, K.M., and Townes, T.M. (2010). KLF1 regulates BCL11A expression and [gamma]- to [beta]-globin gene switching. Nature Genetics *42*, 742–744.

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature *509*, 487–491.

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L.M., Li, Y., and Yang, H. (2019). Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. Science *364*, 289–292.