Understanding regulation of zeta-globin transcription as the first step towards embryonic globin induction in patients with severe alpha-thalassemia

Duantida Songdej

Green Templeton College

A thesis submitted for the degree of

Doctor of Philosophy

Of the

University of Oxford

MRC Molecular Haematology Unit
Weatherall Institute of Molecular Medicine
University of Oxford

Trinity Term 2015

This thesis is dedicated to my parents

For their endless love and support

Abstract

Duantida Songdej
Green-Templeton College

Doctor of Philosophy
Trinity Term 2015

It is estimated by the World Health Organization that 250,000 individuals with severe hemoglobinopathies are born each year. A significant proportion of these suffer from α-thalassemia, which is one of the most common human monogenic disorders known with a carrier rate of >1% among all tropical and subtropical populations that have been studied. A common cause of alpha-thalassemia is an intra chromosomal deletion, termed the Southeast Asian (SEA) deletion, which removes both adult alpha-globin genes leaving the embryonically expressed zeta-globin gene intact. The SEA deletion is very common in some areas of the world, including Northern Thailand, where it is present at an allele frequency of approximately 15%. Individuals homozygous for the SEA deletion die of severe anaemia and tissue hypoxia in the third trimester of pregnancy, a condition termed Hb Bart's Hydrops Fetalis Syndrome (BHFS).

Although BHFS has hitherto been considered a universally fatal disorder, an increasing number of patients have survived because of prenatal and immediate postnatal blood transfusion. The first aim of this work is to fully document the natural history and clinical outcomes of long-term survivors with the BHFS to gain insight into whether this disease should now be considered manageable and to assess whether the burden of treatment is too great. To achieve this, I have initiated a BHFS survivor registry and recruited 60 cases. Analysis suggests that as many as 82% of the BHFS survivors have favorable long-term neurodevelopmental outcomes. However, 50% of the patients suffer from severe growth retardation and 14% have inoperable limb defects. In addition, the majority (83%) have the burden of life-long transfusion dependence. Previous work suggests that zeta-globin can functionally substitute for alpha globin in adulthood, therefore there is a need for improved understanding of the regulation of the embryonic zeta-globin gene to allow development of targeted therapeutic approaches for embryonic hemoglobin induction to ameliorate BHFS.

The second aim of this work is to investigate the *cis*- and *trans*-regulation of the zeta-globin gene with the ultimate aim of preventing silencing of this gene or reactivating its expression in definitive hematopoiesis. I have investigated the zeta-globin *cis*-regulatory network during murine primitive erythropoiesis using a DNaseI hypersensitivity assay coupled with an approach (termed Capture-C) to determine the *cis*-acting regulatory landscape of the zeta-globin gene. Interestingly, the data show that all five previously characterized DNaseI Hypersensitive Sites (DHSs) 5' of the α-globin cluster are present in primitive erythroid cells and that no extra sites are present at this developmental stage. These DHSs interact with both the zeta-globin and alpha-globin gene promoters in primitive erythroid cells as determined by Capture-C. I have also identified differential contributions of the individual alpha-globin *cis*-acting elements on alpha- and zeta-globin expression during primitive erythropoiesis using mouse lines harboring specific *cis*-element deletions.

To identify novel *trans*-acting regulators of zeta-globin I have compared the transcriptomes of primitive and definitive erythroblasts using high-throughput sequencing and gene expression arrays. I have prioritized differentially expressed genes, using ontology analysis, for future functional testing using a Cas9/CRISPR library screen. In the last part of this work, I have integrated information on the *cis*- and *trans*-regulation of zeta-globin in an attempt to gain insight into an unusual BHFS survivor, who remarkably survived the first year of life with minimal transfusion, most likely because of persistent expression of zeta-globin.

Acknowledgements

I would like to thank to my first supervisor Professor Doug Higgs who gave me the opportunity to work in the lab and introduced me to this exciting project. He is my role model not only as a great scientist but also as an incredibly supportive teacher. Thanks for his expert guide to the work presented in this thesis over the last three years. I am also truly indebted to Dr. Chris Babbs who is my second supervisor and who has patiently introduced me to molecular biology and human genetics. He taught me laboratory work from scratch and has been available for help and advice throughout my time in the lab. This thesis would never have been written without him. I am also grateful to Professor Veronica Buckle who has allowed me to take up a working space in her lab and looked after me over these years. Performing this work has been an absolute joy.

I am thankful to the transgenic group, Jackie Sloane-Stanley, Jackie Sharpe and Sue Butler for taking care of the mice and providing endless advice on cell culture. I would also like to thank to Stephen Taylor, Simon McGowan, Emmanouela Repapi and Jelena Telenius for their bioinformatics help. Analysis of all the high-throughput sequencing data I have performed in this work would not be possible without them.

I also thank our clinical collaborators who kindly provided clinical information of survivors with the Hb Bart's hydrops fetalis syndrome (BHFS) (names given in Chapter 3, Table 3.2, 3.3). Of all the collaborators, I would like to give a special thank to Dr. Vip Viprakasit, who introduced the remarkable boy with

the BHFS presented in Chapter 6 to the department and provided us with the patient's clinical samples. Multiple deliveries of the samples would only be possible with helps from Dr. Vip's lab manager, Suchada Riolueang, to whom I am also grateful.

This thesis has been informed by fruitful discussions with members of the Higgs, Gibbons and Buckle laboratories: Jim Hughes, Maria Suciu, James Davies, Nigel Roberts, Sachith Mettananda, Yavor Bozhilov, Bryony Graham, Mira Kassouf, Julia Truch, Deborah Hay, Andrew Smith, Helena Ayyub, Jill Brown, Gemma Moir-Meyer and Noemi Roy to name but a few. I also have to thank the members of my thesis committee, Professor Sir David Weatherall and Associate Professor Marella De Bruijn for their helpful suggestions and advice in general.

I am greatly thankful to Ramathibodi Hospital, Mahidol University, Thailand, who has funded for my DPhil study. Thanks to my colleagues in the Hematology-Oncology Unit, Department of Pediatrics, Ramathibodi Hospital, who always believed in me and who have worked hard when I am away studying. I promise to do my best to take care of our patients upon my return.

I also thank all my friends both in the UK and Thailand for providing the support and friendship that I need. I wholeheartedly thank my mom and dad who have dedicated their lives to me and provided unconditional love and care. I know I always have my family to count on when times are rough and I love both of you so much. I hope you are proud of me.

Abbreviations

AA	Amino acid
aCGH	Array comparative genomic hybridization
APE	Apurinic/apyrimidinic endonuclease
ASD	Atrial septal defect
ATAC	Assay for Transposase-accessible chromatin
BHFS	Hb Bart's hydrops fetalis syndrome
bp	Base pair
BSA	Bovine serum albumin
BW	Birth weight
CBRG	Computational Biology Research Group
CD	Cluster of differentiation
CFC	Colony forming cell
CHB	Han Chinese in Beijing
CHF	Congestive heart failure
ChIP	Chromatin immunoprecipitation
CMP	Common myeloid progenitor
CN	Copy number
CPAP	Continuous positive airway pressure
CPR	Cardiopulmonary resuscitation
CRISPR	Clustered regularly interspaced short palindromic repeats
CS	Caesarean section
DAVID	The database for annotation, visualization and integrated discovery
DE	Differential expressed

DEPC	Diethylpyrocarbonate
DGV	Database of genomic variants
DHS	DNaseI hypersensitive site
DIC	Disseminated intravascular coagulation
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DQ	Developmental quotient
DRED	Direct repeat erythroid definitive
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMP	Erythroid-myleloid progenitor
EPO	Erythropoietin
EPOR	Erythropoietin receptor
EryD	Definitive erythroid progenitor
EryP	Primitive erythroid progenitor
ES	Embryonic stem
EVS	Exome variant server
F	Forward primer
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FDR	False discovery rate
FLCE	unsorted fetal liver culture
FU	Follow up

GA	Gestational age
GEDI	Gene expression dynamic inspector
GFP	Green fluorescent protein
GH	Growth hormone
GI	Gastrointestinal
GPA	Glycophorin A
GVHD	Graft versus host disease
GWAS	Genome-wide association study
HDAC	Histone deacetylase
HFOV	High frequency oscillatory ventilation
HLA	Human leukocyte antigen
HPFH	Hereditary persistent of fetal hemoglobin
HS	Hypersensitive site
HSC	Hematopoietic stem cell
HT	Hypertension
IEF	Isoelectric focusing
IMDM	Iscove's Modified Dulbecco's Medium
iPSC	Induced pluripotent stem cell
IVH	Intraventricular hemorrhage
JPT	Japanese in Tokyo
kb	Kilobase
LCR	Locus control region
LIC	Liver iron concentration
LOH	Loss of heterozygosity
LPLC	Low pressure liquid chromatography

LT-HSC	Long-term hematopoietic stem cell
Mb	Megabase
MCA-PSA	Peak systolic velocity in the middle cerebral artery
MCHC	Mean corpuscular hemoglobin concentration
MCS	Multispecies conserved sequences
MCV	Mean corpuscular volume
MEL	Mouse erythroleukemia
MGI	Mouse genome informatics
MLPA	Multiplex ligation-dependent probe amplification
MRI	Magnetic resonance imaging
MTA	Mouse transcriptome array
MTG	Monothioglycerol
NA	Data not available
NEG	Negative
NICU	Neonatal intensive care unit
NO	Nitric oxide
NPCPAP	Nasopharyngeal continuous positive airway pressure
PBE	Phosphate buffered saline, Bovine serum albumin and Ethylenediaminetetraacetic acid
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDA	Patent ductus arteriosus
PDS	Plasma-derived serum
PFHM	Protein-free hybridoma medium

PIC	Pre-initiation complex
PMSF	Phenylmethanesulfonyl fluoride
POS	Positive
PPHN	Persistent pulmonary hypertension of the newborn
PRC	Polycomb repressive complex
PRE	Pyrimidine-rich element
PV	Per vaginal
qPCR	Quantitative PCR
RDS	Respiratory distress syndrome
RIN	RNA integrity
RNA	Ribonucleic acid
RPKM	Reads per kilobase of exon per million reads mapped
RPMI	Roswell Park Memorial Institute
RSB	Receive signaling buffer
RT	Reverse transcriptase
SCD	Sickle cell disease
SCF	Stem cell factor
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEA	South East Asian
SFL	Sorted fetal liver culture
shRNA	Short-hairpin RNA
SND	Smallest natural deletion
SNV	Single nucleotide variant
TD	Tagmentation DNA

UCSC	University of California – Santa Cruz
UDG	Uracil-DNA glycosylase
URE	Upstream regulatory element
URE-BF	Upstream regulatory element binding factor
US	United States of America
UTR	Untranslated region
UV	Ultraviolet
WGS	Whole genome sequencing
WIMM	Weatherall Institute of Molecular Medicine
WT	Wild type
WTCHG	Wellcome trust center for human genetics

Table of contents

List of Tables

List of Figures

Chapter 1 : Introduction

1.1 α-Thalassemia and the Hemoglobin Bart's Hydrops Fetalis Syndrome

1.1.1 Pathophysiology and clinical presentation of α-thalassemia

Hemoglobin is the oxygen-carrying molecule of red blood cells, at all stages of development, human hemoglobin is a tetramer made up of two α-like and two β-like globin chains. Thalassemia, one of the most common human monogenic disorders, is the term given to an imbalance in the levels of α- and β-globin chains in red blood cells, excess of either chain damages the red cell precursors and causes ineffective production of erythroblasts. Both the α- and β-globin loci harbor genes encoding globin chains that are specifically expressed only in the first trimester of gestation (termed embryonic globins) in addition to the genes encoding the adult globin chains that are expressed throughout adult life. The β-globin locus also encodes a β-like chain that is expressed throughout the second and third trimesters of development, this is termed γ-globin and complexes with α-globin to form fetal hemoglobin (HbF) (Figure1.1, Table 1.1). The primary defect in individuals with α-thalassemia is the absence or underproduction of α-globin chains. As a result, excessive β-like globin chains aggregate to form tetramers, in fetal life this is termed Hb Bart's (γ_4) and in adult life this is termed termed HbH (β_4). Excessive β-like globin chains also damage the maturing erythroid precursors, giving rise to intramedullary hemolysis or so called "ineffective erythropoiesis". Both Hb

Bart's and HbH have decreased P50 values and therefore have a high oxygen affinity and this further contributes to tissue hypoxia. Moreover, symptoms of affected individuals can be worsened by extravascular hemolysis resulting from the premature destruction of mature red cells carrying Hb Bart's and HbH.

The most common causes of α-thalassemia are complete or partial deletions of the α-globin locus: normal individuals have four α-globin genes, arranged as linked pairs at the tip of each copy of chromosome 16, written as αα/αα. α-Thalassemia most commonly results from the deletion of one (-α) or both (--) α-globin genes. Carriers of α-thalassemia (-α/αα) and (--/αα) and homozygotes for the mild haplotype –α (-α/-α) have mild hypochromic microcytic anemia. Compound heterozygotes (--/-α) have hemolytic anemia, while homozygotes for the -- haplotype have a lethal condition referred to as Hb Bart's Hydrops Fetalis syndrome (BHFS). Other causes of α-thalassemia include point mutations of the α-globin coding sequence which, in concert with the -- haplotype, cause severe α-thalassemia. The severity of alpha-thalassemia correlates well with the degree of reduction in α-globin synthesis, resulting from the different types of mutation.

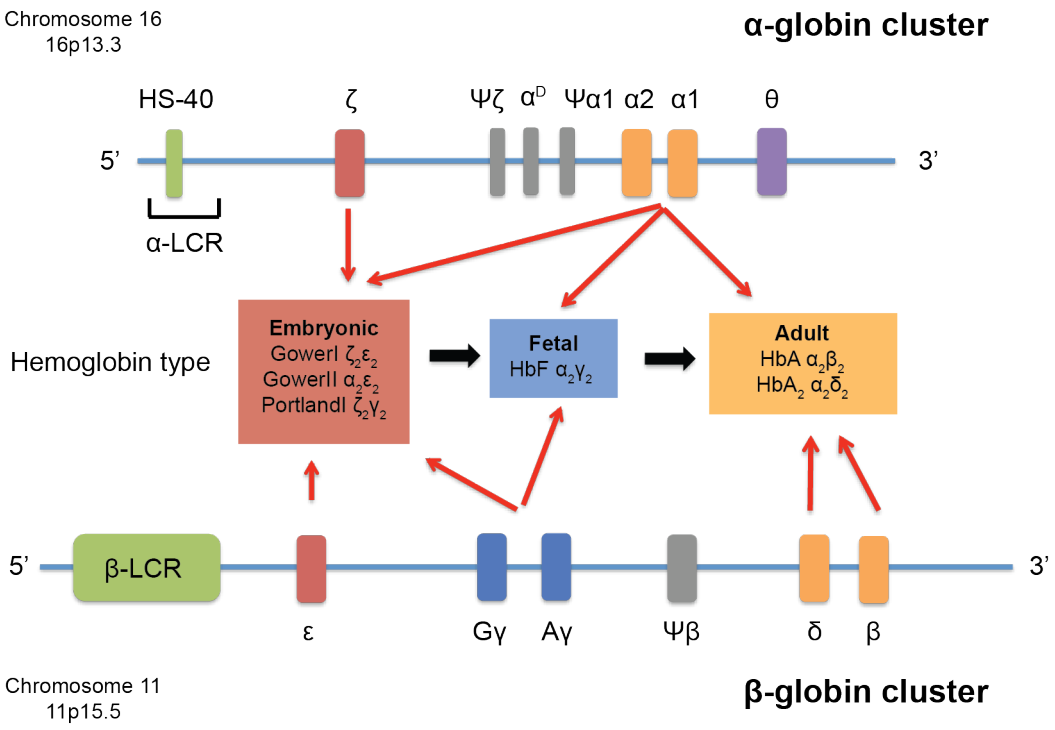


Figure 1.1 Globin tetramers and developmental globin switching

The upper panel shows genes within the α-globin locus and the α-globin 5′ regulatory region (α-LCR) on human chromosome 16. The lower panel shows the β-like globin genes and the β-globin locus control region (β-LCR) on chromosome 11. Non-functional genes are shaded in grey. The middle panel shows normal hemoglobins at each developmental stage. Red arrows indicate the contribution of each α- and β-like globin gene to form developmental-stage specific globin tetramers.

1.1.2 Epidemiology of α-thalassemia

α-Thalassemia is one of the most common human monogenic disorders known. The carrier rate of α-thalassemia varies, however, it is greater than 1% among all tropical and subtropical populations that have been studied (Harteveld and Higgs, 2010). In some parts of the world, where malaria is or has been endemic, the carrier rate is especially high, for example the allele frequency of the most common deletion that removes both α-globin paralogues, termed the South East Asian deletion (SEA), is up to ~15% in Northern Thailand (Lemmens-Zygulska et al., 1996; Weatherall and Williams, 2009). This high frequency is likely to be because the carrier state of α-

thalassemia offers a degree of protection against *falciparum* malaria although the underlying mechanism remains unclear.

In places where carriers of α-thalassemia are common, the clinically important diseases HbH disease and the BHFS occur in compound heterozygotes and homozygotes. Due to regional variation of pathogenic α-globin alleles, HbH disease is mainly seen in South East Asia, the Middle East and the Mediterranean, whereas the BHFS is predominantly seen in South East Asia (Fischel-Ghodsian et al., 1988; Fucharoen and Winichagoon, 1992; Liao et al., 2007; Suwanrath-Kengpol et al., 2005; Waye et al., 1992; Winichagoon et al., 1992).

1.1.3 Molecular basis of α-thalassemia

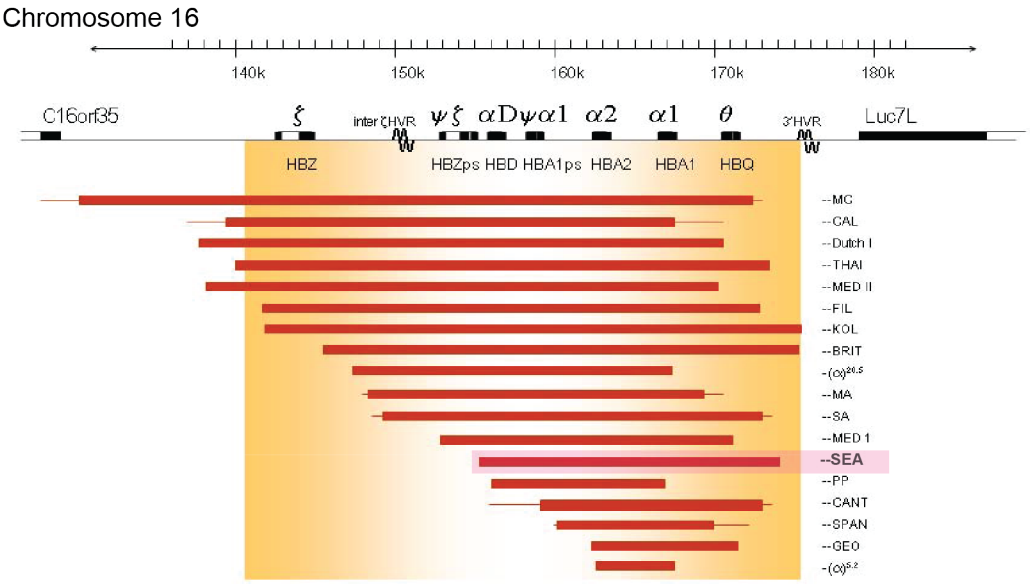
The α-globin cluster is highly conserved in mammals, in humans it is located in a gene-dense region close to the telomere of chromosome 16 (in band 16p13.3). The alpha-like genes are embedded along the cluster in sequence from 5′ to 3′ (5′-ζ-ψζ-α^{D}-ψα1-α2-α1-θ-3′) of which only the embryonically expressed ζ-globin genes and the paralogous adult expressed α2 and α1 are functional. All three α-like genes are expressed in primitive erythroblasts emerging from the blood islands of the yolk sac (Peschle et al., 1985). It has been demonstrated in mouse that as primitive erythroblasts mature, there is a maturational switch from ζ- to α-globin expression (Kingsley et al., 2006) and α-globin is exclusively expressed in definitive erythropoiesis which starts in the liver at ~E10.5. It is likely that a similar maturational switch occurs in humans (Olivier et al., 2006). In humans, definitive erythropoiesis starts in the liver from 6 weeks of gestation concurrent with silencing of ζ-globin in

primitive erythropoiesis. The ζ-globin gene is quiescent at birth in humans and mice.

In normal individuals α-globin is produced by the two paralogous α-globin genes that reside on each copy of chromosome 16. Expression of the α-globin genes depends upon four Multispecies Conserved Sequences termed MCS-R1 to R4 that lie several kb 5′ of the α-globin genes (Hughes et al., 2005). Most frequently, α-thalassemia results from deletion of one or both α-globin genes (written as –α and --). Occasionally, α-thalassemia arises from point mutations in critical regions of the α2, or α1 genes (termed non-deletional α-thalassemia). Far less frequently, α-thalassemia occurs as a result of deletion of one or more of the regulatory elements. In such circumstances MCS-R2 is always removed suggesting its role as the major regulatory element (summarized in (Higgs et al., 1998)). When a mutation gives rise to a complete depletion of α-globin produced from both alleles, this is called α^0-thalassemia and when a mutation only partially downregulates α-globin expression, this is called α^+-thalassemia. To date ~80 mutations resulting in α^+-thalassemia and ~40 mutations associated with α^0-thalassemia have been reported (Giardine et al., 2007; Higgs, 2009a). In this chapter, only the molecular basis of α^0-thalassemia due to deletions of the structural α-globin genes will be reviewed.

Humans heterozygous for deletions of both α-globin paralogs on one allele survive and develop normally, however, individuals lacking all four adult α-like globin genes have the BHFS (see Chapter 1, Section 1.1.4). While all of the deletions associated with α^0-thalassemia either completely or partially remove both α-globin genes, many deletions also remove the ζ-globin gene (Figure

1.2). Individuals lacking all functional α-like globin genes do not survive the early embryonic stages of gestation, although heterozygotes appear to develop normally. This is because without the ζ- and α-globin genes neither embryonic ($\zeta_2\gamma_2$, $\zeta_2\varepsilon_2$, $\alpha_2\varepsilon_2$) nor fetal globin ($\alpha_2\gamma_2$) can be formed. Among deletions causing α^0-thalassemia, the SEA deletion ($--^{SEA}$) is the most prevalent and occurs predominantly in the South East Asia and China. In some parts of the world such as Northern Thailand, the allele frequency of the SEA deletion is up to ~15% (Weatherall and Williams, 2009). The deletion removes both α-globin paralogs but leaves the ζ-globin gene intact. This makes it theoretically possible to reactivate ζ-globin in the absence of all four α-globin genes in individuals homozygous for the SEA deletion. Previous studies show that heterozygotes for the SEA deletion almost always have a very small amount of ζ-globin expression in both fetal (Chui et al., 1989) and adult life (Chui et al., 1986; Tang et al., 1992). The question that remains to be answered, however, is how the ζ-globin gene can be reactivated in definitive erythropoiesis to the extent that it could functionally substitute for α-globin to ameliorate the BHFS.



Adapted from Harteveld CL and Higgs DR, Orphanet J Rare Dis, 2010

Figure 1.2 Deletions that remove both α-globin paralogs with breakpoints lying within the α-globin cluster.

The upper panel shows genes within the α-globin cluster. The lower panel shows the extent of each deletion indicated by red bars. The most prevalent α^0-thalassemia deletion, termed the South East Asian (SEA) deletion, is highlighted in pink.

1.1.4 The Hb Bart's hydrops fetalis syndrome (BHFS)

Hydrops fetalis is a clinical syndrome characterized by generalized edema, ascites, pleural and pericardial effusion in the developing fetus. The condition can result from various causes giving rise to severe heart failure *in utero*. The most common cause of hydrops fetalis in South East Asia (up to 90% of cases) is α^0-thalassemia (Ko et al., 1991; Liang et al., 1985; Lie-Injo Luan, 1959; Suwanrath-Kengpol et al., 2005; Tan et al., 1989; Thumasathit et al., 1968), this is called the Hb Bart's hydrops fetalis syndrome. In normal individuals the α-globin genes are expressed throughout gestation, contributing to Hb Gower 2 ($\alpha_2\varepsilon_2$) in the embryonic period, Fetal Hb (HbF, $\alpha_2\gamma_2$) in the fetal period and the adult Hb (HbA, $\alpha_2\beta_2$) that is continuously produced throughout adult life. Affected fetuses with the BHFS produce a significant amount of non-functional Hb Bart's (γ_4) and suffer from severe

anemia *in utero*, which in turn causes severe tissue hypoxia, heart failure and the abnormalities associated with hydrops fetalis. In circumstances where deletions leave at least one copy of the ζ-globin gene intact, a small amount of functional embryonic Hb Portland ($\zeta_2\gamma_2$) is produced. This allows sufficient tissue oxygenation for such infants to survive until the third trimester of pregnancy when they usually die *in utero* or shortly after birth. Affected fetuses with deletions spanning the ζ- and α-genes on both alleles die during early gestation and thus do not demonstrate the classical hydrops fetalis syndrome.

Hb type	Globin chains
HbGower I	$\zeta_2\varepsilon_2$
HbGower II	$\alpha_2\varepsilon_2$
HbPortland I	$\zeta_2\gamma_2$
HbPortland II	$\zeta_2\beta_2$
HbF	$\alpha_2\gamma_2$
HbA	$\alpha_2\beta_2$
HbA_2	$\alpha_2\delta_2$
HbBart's	γ_4
HbH	β_4

Table 1.1 Composition of globin chains in different types of hemoglobin

In Southeast Asia, the most common cause of the BHFS is homozygosity for the SEA deletion ($--^{SEA}/--^{SEA}$). The BHFS also manifests in compound heterozygotes for the $--^{SEA}$ and two less common α^0-thalassemia deletions, --

^{FIL} and --^{THAI} (deletions spanning the ζ- and α-globin genes), (Figure 1.2) (Fischel-Ghodsian et al., 1988). Although α^0-thalassemia is far less common in other parts of the world, rare cases of the BHFS have been reported in individuals of Greek (--^{MED}/--^{MED})(Diamond et al., 1965; Kattamis et al., 1980; Pressley et al., 1980; Sharma et al., 1979) and Sardinian (--^{MED}/--^{MED}) origin (Galanello et al., 1990). Predominant hematological findings in infants with the BHFS include severe macrocytic hypochromic anemia with a mean hemoglobin level of 6.5 g/dl (Vaeusorn et al., 1985), increased reticulocyte count and nucleated red cells in peripheral blood (Wasi et al., 1974). Affected infants also have hyperplastic bone marrow and various sites of extramedullary erythropoiesis including liver, spleen and kidney.

As a result of excessive β-like globin chains, in the absence of α-globin, Hb Bart's (γ_4) accounts for ~70-90% of the total hemoglobin (Vaeusorn et al., 1985; Wasi et al., 1974), while small amount of HbH (β_4) can be detected in fetal blood towards the end of gestation. Both Hb Bart's and HbH display high oxygen affinity and do not demonstrate any heme-heme interaction (Benesch et al., 1961; Horton et al., 1962), leading to severe tissue hypoxia. In affected individuals, Hb Portland I ($\zeta_2\gamma_2$) and to a lesser extent Portland II ($\zeta_2\beta_2$) make up ~10-20% of the fetal blood (Higgs, 2009b). Hb Portland exhibits a normal pattern of oxygen dissociation (Tuchinda et al., 1975), therefore can deliver oxygen to the developing fetus. The mechanisms underlying ζ-globin regulation are still unclear and are yet to be extensively studied.

Infants with the BHFS usually survive until a gestational age of ~33 weeks (range 23-43 weeks) when they frequently die *in utero*, or perinatally (Chui and Waye, 1998; Liang et al., 1985; Nakayama et al., 1986; Thumasathit et

al., 1968; Vaeusorn et al., 1985). Apart from the hydropic features, an increase in placental to fetal weight ratio, ranging from 0.37-1.16 (mean 0.68) compared to the normal ratio of 0.15-0.25, is also noted (Liang et al., 1985). Autopsy findings also report a progressive decrease in brain weight relative to that expected for gestational age from about 8 weeks of gestation (Higgs, 2009b). Congenital abnormalities have been reported in up to 17% of cases (Abuelo et al., 1997; Adam et al., 2005; Bizzarro et al., 2003; Carr et al., 1995; Chitayat et al., 1997; Dame et al., 1999b; Fung et al., 1999; Guy et al., 1985; Harmon et al., 1995; Lam et al., 1997; Lee et al., 2009; Liang et al., 1985; Lucke et al., 2005; Nakayama et al., 1986; Ng et al., 1998; Singer et al., 2000; Vaeusorn et al., 1985). These include urogenital abnormalities (hypospadias, ambiguous genitalia and undescended testes), limb defects, hydrocephaly, microcephaly, cardiac defects, gastrointestinal (GI) defects and pulmonary hypoplasia. Among these associated congenital abnormalities, urogenital defects are most frequently noted. The cause of the common genital abnormalities found in affected individuals with the BHFS is currently unclear. It is also unclear why infants with the BHFS show considerable variation in the extent of abnormal development and clinical severity despite having the same genotype ($--^{SEA}/--^{SEA}$).

The BHFS has long been regarded as a universally fatal disorder. However, with ever-improving neonatal screening and intensive care, there are an increasing number of reports of infants who have been treated with blood transfusion *in utero* or immediately after birth and who have survived with the BHFS (Beaudry et al., 1986; Bianchi et al., 1986; Bizzarro et al., 2003; Carr et al., 1995; Chik et al., 1998; Chmait et al., 2014; Dame et al., 1999b; Dwinnell

et al., 2011; Fung et al., 1999; Fung et al., 1998; Gumuscu et al., 2013; Hayward et al., 1998; Jackson D.N. et al., 1990; Joshi et al., 2004; Kou et al., 2013; Lam et al., 1992; Lee et al., 2007; Lee et al., 2009; Leung et al., 2002; Liu et al., 2002; Lucke et al., 2005; Naqvi A et al., 1997; Ng et al., 1998; Singer et al., 2000; Sohan et al., 2002; Thornley et al., 2003; Wang and Ryan, 2009; Westgren et al., 1996; Yi et al., 2009; Zhou et al., 2001). The BHFS survivors are good candidates for hematopoietic stem cell transplantation and future novel curative therapies. However, given the extreme phenotypic variability regarding both physical and mental development, there is currently a need for an extensive review of long-term outcomes of the BHFS survivors to ascertain whether this should be regarded as a treatable disease. In addition to maternal risk factors such as obstetric complications (pre-eclampsia, poly- or oligo-hydramnios and pre-mature labor), this information is crucial for counseling couples at risk of having affected child with the BHFS.

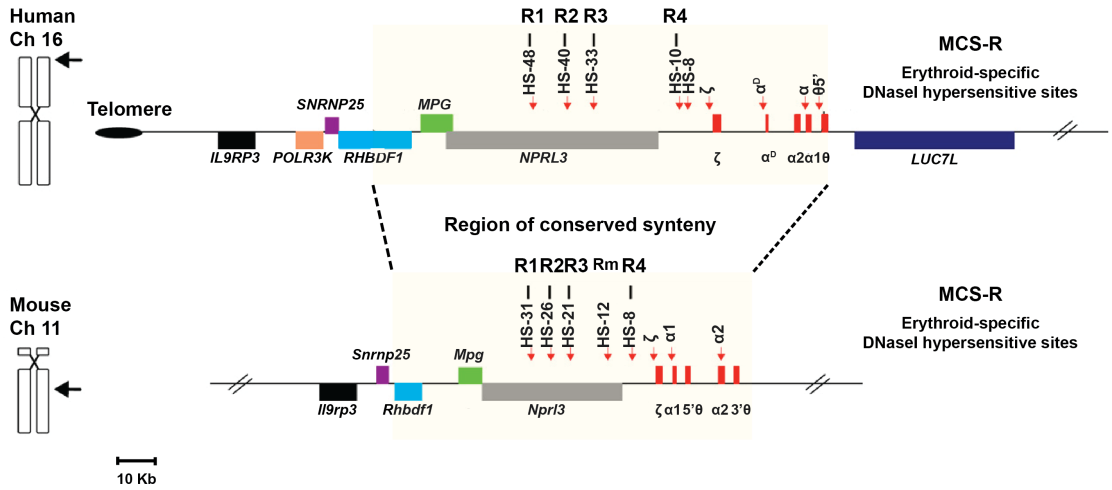
1.2 Regulation of α-globin expression during erythropoiesis

1.2.1 Control of α-globin expression by cis-acting regulatory elements

The epigenetic status and conformation of the α-globin cluster and the mechanisms controlling α-globin expression have been extensively studied in definitive erythropoiesis. The DNA sequence underlying the α-globin cluster shows more than 500 million years of evolutionary conservation (Hughes et al., 2005). The region of conserved synteny is ~135 kilobases (kb) and not only includes the α-globin cluster, but also several ubiquitously expressed

flanking genes (Hughes et al., 2005) (Figure 1.3). Although this 135 -kb region lies close to a telomere in many species (Hughes et al., 2005), including human chromosome 16 (16p13.3, ~150 kb from the telomere), in the mouse it lies at an interstitial region of chromosome 11 (11qA4) (Figure 1.3). This finding suggests that although translocations have arisen during evolution, the 135 kb segment contains all the *cis*-regulatory elements required for complete developmental regulation of α-globin expression.

In both human and mouse the α-globin gene cluster is located in a GC-rich, repeat-rich, gene-dense segment of DNA. The promoters of all genes within this region are marked by CpG-rich islands. Expression of the α-globin genes is controlled by four upstream, conserved, non-coding regulatory sequences, known as Multispecies Conserved Sequences or MCS-R1-R4 (Hughes et al., 2005). These MCSs correspond to DNaseI hypersensitive sites (DHSs) (Birney et al., 2007; Higgs et al., 1990), which in human are termed HS-48 (MCS-R1), HS-40 (MCS-R2), HS-33 (MCS-R3) and HS-10 (MCS-R4), according to the distance in kb 5′ of the ζ-globin mRNA cap site (Figure 1.3). In mouse the orthologous elements are termed HS-31 (R1), HS-26 (R2), HS-21 (R3) and HS-8 (R4). There is also an additional non-conserved mouse-specific hypersensitive site termed HS-12 also thought to perform a regulatory function (Hughes et al., 2005; Kielman et al., 1996) (Figure 1.3). Three of these remote regulatory elements, MCS-R1-R3, are situated within introns of an adjacent gene *Nprl3*, the function of which is required for normal development of the cardiovascular system (Kowalczyk et al., 2012).



Adapted with permission from Wallace HA et al., Cell, 2007 (Wallace et al., 2007)

Figure 1.3 The chromosomal localization and region of conserved synteny of the human and the mouse α-globin cluster

The left side of the figure shows that the human α-globin cluster lies in the sub-telomeric region of chromosome 16 (16p13.3), whereas the mouse cluster lies at an interstitial chromosomal position (11qA4). The right hand side shows a detailed view of the α-globin cluster in humans and mice. The globin genes are shown as red rectangles. Other human genes and their mouse orthologs in the region are annotated below or above relevant colored rectangles. *POLR3K* and α^D are found only in the human cluster. Orange shading shows the region of conserved synteny extending from the RHBDF1 gene to the θ-globin gene. Red arrows above the clusters show the known erythroid-specific DNaseI hypersensitive sites (HS). Black grids at the top of each cluster indicate the position of multispecies conserved regulatory elements (MCS-R; R1-4) corresponding to the HSs, which regulate the α-globin genes. The mouse-specific regulatory element is annotated as Rm.

Several lines of evidence, from transgenic experiments (Lung et al., 2000; Moreau-Gaudry et al., 2001; Sharpe et al., 1992; Sharpe et al., 1993a; Sharpe et al., 1993b; Tang et al., 2007) and homologous recombination (Wallace et al., 2007) suggest that when removed, MCS-R2 (HS-40) is the only element to reduce expression of the α-globin genes to a great extent and consequently it has been regarded as the major and crucial distal regulatory element. This observation is supported by reports of naturally occurring deletions that remove upstream elements and cause α-thalassemia. (summarized in (Higgs and Wood, 2008b)). However, more recently a report describes a male Portuguese patient carrying homozygous deletion of HS-40 and who presented with HbH disease (Hb level 10.3 g/dl), demonstrating significant reduction, but not a complete absence of α-globin production (Coelho et al.,

2010). However, nothing is currently known about the contribution of the α-globin MCS-Rs to ζ-globin gene regulation during primitive erythropoiesis.

1.2.2 Transcriptional activation of the α-globin genes

The ability to expand and isolate erythoid progenitors, coupled with advanced development of chromatin immunoprecipitation assays (ChIP) allows us to monitor dynamic changes of transcription factor binding and the associated chromatin modifications along the α-globin locus throughout erythropoiesis in human and mouse (De Gobbi et al., 2007). In multipotent cells (CMP), the cluster is primed by binding of protein complexes containing SCL, NF-E2 and GATA2 to MCS-R2. As the proerythroblast stage approches, additional remote regulatory elements are bound by multiprotein complexes containing various combinations of the SCL pentameric complex, NF-E2 and GATA1 replacing GATA2. At this stage, the α-globin promoter is also occupied by a combination of factors including NF-Y and the chromatin associated with both the MCS elements is decorated with epigenetic marks associated with active chromatin and the promoter also becomes transcriptionally active. However, there is little globin synthesis observed at this point. As transcription starts in early and intermediate erythroblasts, the pre-initiation complex (PIC), including PolII, is recruited to both the enhancers and the promoter in a cooperative but independent manner. KLF1 is also recruited, independently of the MCS elements, to the promoter. At this final stage, physical interaction of the enhancers and the promoters is thought to occur via the PIC thereby initiating high levels of transcription. This process is dependent on the presence of the MCS-R sequences (particularly MCS-R2) and is thought to occur through a mechanism of chromatin looping (summarized in (Higgs and

Wood, 2008b), also see (Anguita et al., 2004; De Gobbi et al., 2007; Sharpe et al., 1993b; Vernimmen et al., 2009)). The looping mechanism is supported by recent studies using chromosomal conformation capture (3C) (Vernimmen et al., 2007) and Capture-C (Hughes et al., 2014) techniques that show direct interaction of the α-globin promoters and MCS-R sequences. MCS-R2 plays a critical role in RNA polymerase II (PolII) recruitment, since removal of MCS-R2 appears to specifically affect the recruitment of PolII, severely downregulating α-globin expression (Vernimmen et al., 2007). An experiment using interspecific MEL hybrid cells carrying an abnormal human chromosome 16 containing a deletion of all the MCS elements showed that neither PolII nor the general transcription factors (e.g. TFIIB and TFIID) are recruited to the promoter in the absence of the MCS elements, thus completely silencing α-globin transcription (Vernimmen et al., 2007). These findings suggest that MCS-R1, 3 and 4 (with or without MCS-R2) may normally play a role in recruiting the basal transcriptional machinery whereas MCS-R2 is important for recruiting PolII at the promoter (Higgs and Wood, 2008b).

The α-globin locus is highly conserved between mouse and human, however, two significant differences in transcription factor binding of the orthologous MCS-R elements have been shown (De Gobbi et al., 2007). First, the mouse α-globin promoters contain a consensus GATA sequence that is bound by GATA1, whereas the human promoter does not. Second, the human corresponding sequence to mouse HS-12 (hoHS-12) is not marked by a DNaseI HS and contains none of the transcription factor binding sites found at the mouse *cis*-element HS-12 (De Gobbi et al., 2007). ChIP experiments

performed in mouse erythroid cells demonstrated that this element binds GATA1, the SCL pentameric complex and NF-E2 and contains epigenetic marks indicative of enhancer activity during erythropoiesis. However, no such recruitment has been seen at the equivalent position in the human α-globin locus. These observations suggest that, despite minor differences, the human and mouse α-globin genes are regulated in a similar fashion.

Significant changes in the chromatin environment of the α-globin locus have been observed during erythroid differentiation. Once the multipotent progenitor (CMP) becomes committed to erythropoiesis and essential erythroid-enriched transcription factors bind the MCS elements, histone modifications (e.g. acetylation) and an increase in associated chromatin accessibility occur (Higgs et al., 2005). In non-erythroid cells the α-globin genes, which are surrounded by ubiquitously expressed genes in a gene dense region, are silenced very early in development via recruitment of the polycomb repressive complex 2 (PRC2). PRC2 is recruited to the α-globin promoter CpG islands, which in turn promotes histone methylation (H3K27me3) and transcriptional silencing is mediated by the localized activity of histone deacetylases (HDACs) (Garrick et al., 2008). When hematopoietic progenitors differentiate to form erythroid cells, the PRC2/HDAC complex is removed and the H3K27me3 chromatin mark disappears through passive and active mechanisms including demethylation by the specific histone lysine demethylase enzyme JMJD3 (Vernimmen et al., 2011).

α-Globin transcription reaches its maximum level in intermediate (polychromatophillic) erythroblasts and is subsequently downregulated as the erythroid cells mature (Brown et al., 2006). However, translation of globin

mRNA continues for 2-3 days in reticulocytes from which the nucleus has been extruded. This process is critical for normal red cell production and is made possible by the long half-life (>24 hours) of globin mRNA (Higgs, 2009a). Stabilization of α-globin mRNA is achieved by an RNA binding complex, termed the α-globin poly(C) binding protein (αCP), which controls deadenylation of its polyA tail and protects the RNA from an erythroid enriched endoribonuclease (Waggoner and Liebhaber, 2003).

1.2.3 Regulation of ζ-globin expression

It remains unclear how expression of the ζ-globin gene is controlled, however, several lines of evidence suggest it might be regulated independently of the α-globin genes. Firstly, interspecies comparison of the globin promoter DNA sequences suggests that the ζ-globin gene promoter is larger and more complex than those of the α-globin genes (see Section 1.5). This implies that a greater number of *trans*-acting factors may bind this region and that ζ-globin may be less reliant on distal *cis*-acting regulatory elements. Secondly, a mouse mutant homozygous for a targeted deletion of three critical α-globin regulatory elements, including MCS-R1 (HS-31), R2 (HS-26) and R3 (HS-21), and the intervening genetic material (termed the smallest natural deletion or SND allele) die of anemia at ~E14.5, shortly after the ζ- to α-globin switch normally occurs (Higgs D, personal communication). The survival of these mice to this stage suggests that they are capable both of expression and repression of ζ-globin in the absence of the α-globin regulatory elements. Thirdly, in transgenic mice the human ζ-gene, when attached to the HS-40 element shows developmental specificity as it is expressed almost exclusively at the embryonic stage (Vyas et al., 1992). This suggests that the ζ-globin

promoter and the HS-40 element together contain all the *cis*-acting regulatory elements required to achieve developmental stage-specific expression and silencing. Lastly, induced pluripotent stem cells (iPS cells) derived from patients with complete loss of the α-globin genes (all five α-globin upstream *cis*-regulatory elements are intact) and trans-differentiated into erythroblasts suggest that ζ-globin is transcribed and repressed in a similar fashion to control cells indicating an independent (non-competitive) means of repression (Chang and Bouhassira, 2012).

1.3 Ontogeny of erythropoiesis

1.3.1 Emergence of primitive erythropoiesis

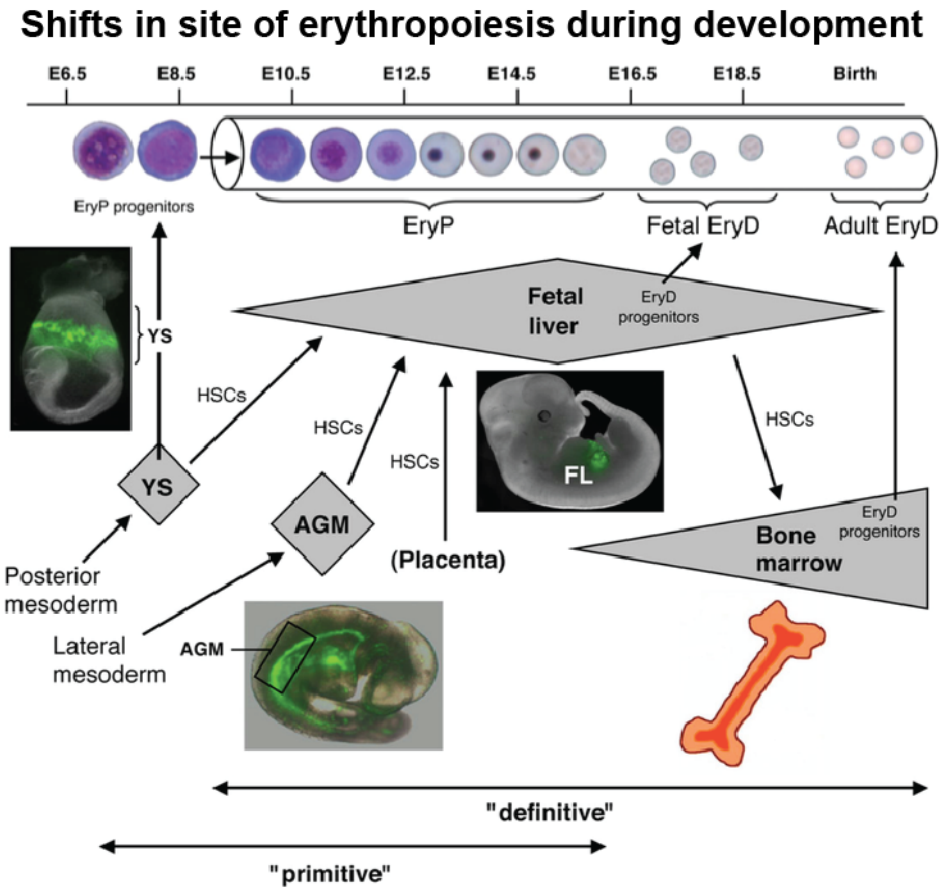
The first wave of mammalian hematopoietic cells arises in the yolk sac and is predominantly erythropoietic, a process known as “primitive erythropoiesis”, however, it also consists of megakaryocytes and macrophages (Bertrand et al., 2005; Ferkowicz and Yoder, 2005; Palis et al., 1999; Tavian and Peault, 2005; Tober et al., 2007). The term “primitive erythropoiesis” originates from the observation that primitive erythroid cells are characterized by large red cells, formed in the yolk sac and circulated as nucleated erythroblasts and are therefore similar to the red cells of non-mammalian embryos (Palis et al., 2010). It has now been clearly identified in the mouse fetus that primitive erythroid cells are eventually enucleated between E12.5-E16.5 (Fraser et al., 2007; Kingsley et al., 2004; McGrath et al., 2008). These large enucleated primitive red cells continue to circulate in the fetal blood and can last up to 5 days after birth (Kingsley et al., 2004). The critical role of primitive erythroid progenitors (EryPs) during embryonic development is not only to carry out

oxygen and carbon dioxide exchange, but also to facilitate vascular formation (Lucitti et al., 2007).

The EryPs identified by *in vitro* assays as EryP-colony forming cells (EryP-CFCs) (Baron, 2013; Palis et al., 1999; Wong et al., 1986), are derived from posterior mesodermal cells soon after the onset of gastrulation and first emerge in “blood Islands” within the yolk sac at E7.25, giving rise to the first blood cells to circulate in the mammalian conceptus (Figure 1.4). Following emergence at E7.25, the EryP-CFCs reach their maximum number at E8.25 and disappear by E9.0 (Palis et al., 1999), therefore they are considered to be a transient wave of progenitors. EryP-CFCs can be cultured *in vitro* and with the addition of erythropoietin (EPO) require 5 days to form colonies containing several hundred mature primitive erythroid cells. However, primitive erythroid progenitors, unlike hematopoietic stem cell (HSC)-derived definitive erythroid progenitors (EryDs) in the fetal liver and postnatal bone marrow, lack the capacity to self-renew when cultured *ex vivo*, even with the presence of EPO, stem cell factor (SCF) and dexamethasone, this likely results from differential expression of receptors for the latter two factors (England et al., 2011).

The observation that primitive erythroblasts emerge in close spatial and temporal proximity with the vascular endothelial cells of the yolk sac suggests these two cell types might share a common progenitor, this is termed the hemangioblast (reviewed in (Choi et al., 1998)). A number of studies support this hypothesis, a study in mouse embryos using clonal assays demonstrates that these bipotential (endothelial and hematopoietic) precursors originate from the primitive streak of gastrulating mouse embryos (Huber et al., 2004). It has also been shown that EryPs are derived from mesodermal cells

expressing the endothelial/hematopoietic markers Flk1, CD31 and Tie-2 (Ema et al., 2006) suggesting their origin is from hemangioblasts. In addition, studies of mouse and human embryonic stem cell cultures when differentiated towards the erythroid lineage also demonstrate that the population of hemangioblasts arises before the onset of hematopoietic progenitor activity (Choi et al., 1998; Kennedy et al., 2007; Zambidis et al., 2005). It is possible that the intracellular environment of EryPs and EryDs influences their later differences in globin expression, however, this has yet to be investigated.



Reprinted with permission from Baron MH, Stem Cells, 2013

Figure 1.4 Shifts in site of erythropoiesis during mouse development

The figure summarizes the shift in sites of hematopoiesis during mouse development, which is highly analogous to human hematopoiesis. Primitive erythropoiesis begins as EryPs, which are derived from posterior mesoderm and develop in the yolk sac. These EryPs subsequently enter circulation where they mature in a semi-synchronous manner. The second and transient

wave of erythropoiesis emerges in the yolk sac as erythroid-myeloid progenitors (EMPs). EMPs migrate from the yolk sac soon after their emergence and seed in the fetal liver where they mature. These EMPs are capable of producing definitive erythroid cells, megakaryocytes and myeloid cells. Definitive hematopoiesis arises from HSCs that emerge from the AGM region. HSCs initially migrate to the fetal liver, this destination shifts to the bone marrow around time of birth where they self-renew and underlie definitive erythropoiesis throughout adult life. After enucleation, mature definitive erythroid cells enter the circulation. The photographs, from left to right, show transgenic mouse embryos that express a green fluorescent protein (GFP) reporter in the EryP lineage, in HSCs and in definitive erythroid progenitors (EryD) respectively.

1.3.2 Terminal maturation of primitive erythroid cells

Although fully functional circulation is established in the mouse embryo after E10, small numbers of primitive erythroid precursors first migrate from the yolk sac and enter the embryo proper at E8.25, concomitant with the proposed onset of cardiac function (McGrath et al., 2003). The number of primitive erythroid precursors expands rapidly in the embryo proper at E9.5, when they continuously mature in the blood stream in a synchronized manner between E9.5-E12.5 (Fraser et al., 2007; Kingsley et al., 2004). Terminal maturation of primitive erythroid cells is similar to that of their definitive counterparts, involving expansion of erythroblast numbers through a limited set of symmetric cell divisions, accumulation of hemoglobin, decrease in cell size, nuclear pyknosis, decrease in RNA content and expression of cell adhesion proteins (reviewed in(Palis, 2014), (Fraser et al., 2007)). Maturing EryPs also demonstrate progressive loss of CD71 and upregulation of Ter-119 expression, similar to that observed for HSC-derived definitive erythroblasts (Fraser et al., 2007; Zhang et al., 2003). When primitive erythroblasts have matured to an orthochromatic stage by E12.5, cell division stops. Primitive erythroblasts have the unique feature of embryonic globin gene expression, from both the α- and β-globin clusters. During primitive

erythropoiesis, maturational “globin switching” occurs (Kingsley et al., 2006) and this will be discussed in detail in Section 1.4.

Studies using specific antibodies for the embryonic globins have shown that primitive erythroblasts in the mouse embryo enucleate between E12.5-E16.5 (Kingsley et al., 2004). As HSC-derived fetal liver definitive erythropoiesis (see section 1.3.4) becomes active, definitive erythrocytes are released into the bloodstream between E11.5 and E12.5 and eventually outnumber primitive erythrocytes (McGrath and Palis, 2005), however, these enucleated primitive erythrocytes can be isolated in the bloodstream of mice up to several days after birth (Kingsley et al., 2004). Another recent study using mice containing a GFP transgene driven by the human ε-globin promoter to identify enucleated primitive erythrocytes also confirms this finding (Fraser et al., 2007). Furthermore, the total number of primitive erythroid cells does not change between E12.5 and E16.5, consistent with the cessation of cell division in late-stage primitive erythroblasts and the enucleation of the entire population of primitive erythroid cells (Kingsley et al., 2004). When bone marrow-derived definitive erythroid cells enucleate two cell populations emerge, the first are enucleate reticulocytes that continue to terminally mature by removal of internal organelles and reorganization of the cytoskeleton, second the condensed nucleus surrounded by a thin rim of cytoplasm and an intact cell membrane, known as a pyrenocyte, which is subsequently ingested by a macrophage (McGrath et al., 2008; Yoshida et al., 2005). Consistently, primitive pyrenocytes can be detected in the bloodstream of mouse embryos between E12.5-E16.5, during the time of enucleation (McGrath et al., 2008). The enucleation process of primitive erythroblasts is thought to occur in the

erythroblastic islands of the fetal liver, where primitive erythroblasts can be isolated during the corresponding embryonic stages and this happens in association with macrophages (Isern et al., 2008; McGrath et al., 2008).

1.3.3 Yolk-sac-derived definitive erythro-myeloid progenitors (EMPs)

In the mammalian embryo hematopoietic function is required prior to the existence of multipotent HSCs, which emerge in fetal and adult hematopoiesis (Fujiwara et al., 1996; Kumaravelu et al., 2002; Muller et al., 1994) (see Section 1.3.4). Additionally, evidence in the mouse also demonstrates that primitive erythropoiesis alone is not sufficient to support embryonic survival up to the time when HSCs are functional (Chen et al., 2011). Therefore, there is a necessity for the second wave of hematopoiesis as an alternative source of transient hematopoietic progenitors bridging between yolk-sac-derived primitive hematopoiesis (mainly erythropoiesis, see Section 1.3.1) and HSC-derived definitive hematopoiesis (see Section 1.3.4).

In the mouse embryo, this second transient hematopoietic wave emerges in the yolk sac at E8.25-E9.5 and consists of progenitors, termed erythro-myeloid progenitors (EMPs) that are capable of producing definitive erythorid, megakaryocyte, myeloid (mainly neutrophil), but not lymphoid cells (Baron, 2013; McGrath et al., 2015) (Figure 1.4). Thousands of definitive erythroid progenitiors found in the murine fetal liver prior to HSC colonization were thought to have arisen from these EMPs (reviewed in (McGrath et al., 2015)). In human embryos EMPs are found in the yolk sac by ~ 4 weeks and in the fetal liver at ~ 5-6 weeks of gestation (Frame et al., 2013; Migliaccio et al., 1986). A recent study demonstrated using a unique immunophenotype that

EMPs arise in the yolk sac, have erythroid and broad myeloid potential and then migrate to the fetal liver at E10.5-E11.5 where they rapidly differentiate, including the production of circulating neutrophils by E11.5 (McGrath et al., 2015). The proportion of remaining EMPs as compared both to primitive erythroid cells in the yolk sac and to peripheral blood is less than 5% at E10.5 (McGrath et al., 2015). The same study also shows that, unlike HSCs, EMPs lack long-term potential but are capable of providing transient adult-like red blood cell reconstitution.

1.3.4 Aorta-gonad-mesonephros (AGM)-derived definitive erythropoiesis

The third wave of hematopoiesis arises mainly from the dorsal aorta area of AGM region within the mammalian embryo. This site is a potent hematopoietic site from which HSCs first emerge at E10.5-E11.5 (Baron, 2013; de Bruijn et al., 2000). HSCs are defined as a population of cells that are capable of long-term reconstitution of the entire hematopoietic system (including the lymphoid lineage) of an irradiated adult recipient (reviewed in (de Bruijn et al., 2000)) and are responsible for the continuous production of all mature blood cells throughout adult life (Lemischka, 1991; Spangrude et al., 1991). In humans, HSCs arise in the AGM region around 5 weeks of gestation (reviewed in (Baron, 2013)).

Definitive erythropoiesis occurs in the fetal liver and postnatal bone marrow when AGM-derived HSCs were seeded (~E12.5 in mouse fetal liver) and is characterized by the progression of lineage-committed cells through progenitor, precursor and mature RBC compartments (Palis, 2014) (Figure 1.4). The first two populations of lineage-committed definitive erythroid

progenitors are termed burst-forming unit erythroid (BFU-E) and colony-forming unit erythroid (CFU-E), the later are only 4-5 cell divisions upstream of mature RBCs. Unlike their primitive counterpart, cellular expansion and terminal maturation of definitive erythorid cells occurs extravasculary in erythroblastic blood islands within the fetal liver and postnatal bone marrow. These compose of erythroblasts physically attached to central macrophage cells, these macrophages are thought to play a role in promoting erythroblast proliferation and maturation, especially in stress erythropoiesis (reviewed in (Palis, 2014)). Terminally maturing definitive erythroid precursors, similar to their primitive counterparts, are morphologically distinguishable at each stage from proerythroblasts to basophilic, polychromatophilic and orthochromatic erythroblasts before they enucleate and go on to become reticulocytes and mature red blood cells. Under normal circumstances, only enucleated definitive erythroid cells enter the blood stream. Recently, a number of flow cytometry-based approaches have been developed for both mice and humans to identify and isolate these subpopulations for further studies (reviewed in (Palis, 2014)).

1.3.5 Characteristics of primitive and definitive erythroid cells

Primitive and definitive erythroid cells are distinct erythroid lineages. While primitive erythropoiesis originates from yolk sac blood islands, definitive erythroid cells are derived from two types of progenitors, yolk sac-derived EMPs and AGM-derived HSCs (see Section 1.3.1, 1.3.3, 1.3.4). Although primitive red cells have gone through progressive morphological and developmental stages, similar to their definitive counterparts, primitive red cells are six-fold larger than definitive red cells and contain six-fold more

hemoglobin when fully mature (Steiner and Vogel, 1973). Additionally, primitive erythroblasts progressively mature in a semi-synchronous fashion within the circulation (see Section 1.3.2), unlike their definitive counterparts that mature extravascularly, mainly in the fetal liver and bone marrow spaces (see Section 1.3.4, 1.3.5). While the cytoskeletal network of definitive red cells is well characterized, how and when the cytoskeletal proteins of primitive red cells are assembled is not clear. In the membrane of mature definitive red cells an elastic network of spectrin-based cytoskeletal proteins is anchored to the cholesterol/phospholipid plasma membrane with actin and Protein 4.1 at their junction (reviewed in (Mohandas and Gallagher, 2008)). The anchoring of the cytoskeleton to the lipid bilayer occurs at these junctions through glycophorin C. In addition, interactions through ankyrin-band 3 stabilize the cytoskeleton to the lipid bilayer (reviewed in (Palis, 2014)). This cytoskeletal protein assembly ensures that mature red cells contain enough elastic potential to repeatedly pass through the small capillary network and yet maintain their biconcave shape. All of these cytoskeletal proteins accumulate during definitive erythroid maturation (Chen et al., 2009). In primitive erythroblasts, band 3 and glycophorin A expression are highly upregulated during early differentiation (Isern et al., 2011), whereas spectrin and ankyrin transcripts have been identified during terminal maturation (Peters et al., 1992). Primitive erythroid cells undergo a loss of surface area and volume during terminal maturation, however, unlike their definitive counterparts this process occurs irrespective of whether the cells are enucleated (Waugh et al., 2001). This suggests that the maturational processes of membrane

remodeling and enucleation are uncoupled in terminally maturing primitive erythroid cells (Palis, 2014).

Primitive and definitive erythroid cells express different combinations of α-like and β-like globin genes. In mouse primitive erythroid cells, all of the functional α-like genes are expressed (embryonic ζ- and adult α1, α2,), however, in contrast only the embryonic β-like (βH1, εγ) globin genes are expressed. Whereas in HSC-derived definitive erythroid cells, only adult α_1 and α_2, and the adult β1 and β2 genes are expressed (see Section 1.4), the transient wave of EMP-derived definitive erythroid cells have a unique globin gene expression pattern, consisting of adult α1- and α2-, adult β1- and β2- and low levels of embryonic βH1-, but not εy-, globin transcripts (McGrath et al., 2011). The transcriptional network underlying these differences in globin gene expression profiles is poorly understood. However, this is a challenging problem because the manifold differences between EryPs and EryDs detailed in this section are also like to stem from differences in transcriptional regulation.

There has also been a debate on whether primitive and definitive erythropoiesis are differentially dependent on the EPO signaling pathway. Erythropoietin signaling is absolutely necessary for definitive erythroid differentiation (Lin et al., 1996; Wu et al., 1995). Early studies showed that EPO failed to increase heme synthesis in cultures of whole mouse embryos containing primitive erythroid cells (Cole and Paul, 1966). However, following studies demonstrated that targeted disruption of the EPO receptor (EPOR) causes a marked reduction of primitive erythroblasts by E10.5-E11.5 and a profound anemia by E12.5 (reviewed in (Palis, 2014)). A more recent analysis

of primitive erythroid cells in EPOR-null mouse embryos also demonstrates the importance of EPOR signaling in the terminal maturation of primitive erythroid cells (Malik et al., 2013).

In this thesis, from this section onward, “primitive” erythropoiesis refers to yolk sac-derived primitive erythroid cells and “definitive” erythropoiesis refers to HSC-derived definitive erythroid cells, unless specified otherwise.

1.3.6 Transcriptional differences between primitive and definitive erythropoiesis

Transcription factors are key regulators of lineage-specific cellular maturation. Global gene expression databases of both single-lineage erythroid cells (Isern et al., 2011; Merryweather-Clarke et al., 2011; Miller, 2004; Redmond et al., 2006) and similarly staged “primitive *vs* definitive” erythroid cells (Kingsley et al., 2013) have been made available. These data as well as comparative expression analyses of specific genes (e.g. *Gata1*, *Klf1*) in primitive *vs* definitive erythroid cells lead to better understanding of the similarities and differences between these two lineages of erythroblasts. Core erythroid genes including *Gata1, Klf1* and *Nfe2* are not differentially expressed (Kingsley et al., 2013) suggesting that these genes are indispensable for both primitive and definitive erythropoiesis and that their functional differences, if any, are not transcriptionally regulated. Targeted inactivation of *Gata1* results in early maturational arrest at the proerythroblast stage in both primitive and definitive erythropoiesis (Fujiwara et al., 1996; Pevny et al., 1995). However, lineage-specific transgenic rescue of a Gata1-null mouse revealed that different functional domains of Gata1 are required to regulate target genes in primitive and definitive erythroid cells (Shimizu et al., 2001). In addition, targeted

disruption of other transcription factors of the GATA1 transcriptional complex, including LDB1, SCL/TAL1, LMO2, and FOG1, results in the arrest of primitive and definitive erythropoiesis (Li et al., 2010; Robb et al., 1995; Shivdasani et al., 1995; Tsang et al., 1998; Warren et al., 1994). KLF1 (EKLF), another key transcriptional regulator of erythropoiesis, plays a role in both primitive and definitive lineages (Drissen et al., 2005; Hodge et al., 2006; Isern et al., 2010; Nuez et al., 1995; Perkins et al., 1995). Functions of KLF1 involve transcriptional regulation of several erythroid-specific genes including the adult and embryonic globins (see Section 1.4), α-hemoglobin stabilizing protein, heme biosynthetic enzymes, several other transcription factors, as well as red cell membrane proteins and commonly expressed blood group antigens (reviewed in (Palis, 2014), (Basu et al., 2007; Hodge et al., 2006; Nilson et al., 2006)). In primitive erythropoiesis, however, complete terminal erythroid maturation seems to be more sensitive to *Klf1* dosage, since haploinsufficiency of *Klf1* in the embryo results in marked abnormalities in the red cell membrane and surface (Isern et al., 2010). Moreover, *Klf1* and the closely related gene *Klf2* have been shown to have overlapping roles in primitive erythropoiesis. Mice with a double knockout of *Klf1/Klf2* demonstrate a more severely anemic phenotype as compared to those with single-knockout of either *Klf1* or *Klf2 (Basu et al., 2007)*. Embryonic lethality in double-knockout mice occurs by E11.5, which is 2-4 days earlier than when either gene is deleted alone (Basu et al., 2007). A recent study suggests that *Klf1* and *Klf2* regulate primitive erythropoiesis through synergistic transcriptional regulation of *c-Myc* (Pang et al., 2012). In humans, definitive erythroid cells are also sensitive to KLF1 such that haploinsufficiency can

cause de-repression of fetal hemoglobin, the condition known as hereditary persistence of fetal hemoglobin (HPFH) (Borg et al., 2010) (see Section 1.4).

While primitive and definitive erythroblasts share many features in common as detailed above, there are also many differences in their structure and biochemistry in addition to their different globin expression profiles. A recent study elegantly compared the expression profiles of stage-matched primitive and definitive erythroblasts using microarrays (Kingsley et al., 2013). This study reported 3,024 genes to be commonly expressed between primitive and definitive erythropoiesis (definitive data included fetal liver and adult bone marrow) and 658 genes to be differentially expressed between primitive erythropoiesis and definitive erythropoiesis in fetal liver. However, the cell-type specific functions of many of these differentially expressed genes in this valuable dataset have yet to be elucidated. The majority of the differentially expressed genes are categorized as transcription factors, of which 3 have been previously shown to be involved in hemoglobin switching. Mice harboring homozygous deletion of *c-Myb*, which is upregulated in definitive erythroid cells by 22-fold (Kingsley et al., 2013), demonstrated an absence of definitive erythroid cells whereas normal primitive erythropoiesis was maintained (Tober et al., 2008). In addition, the same study also showed that *c-Myb* was not expressed in primitive erythroid precursors. Homozygous *c-Myb* mutant mice became progressively anemic by E15 and died by day E15.5 supporting the hypotheses that primitive erythropoiesis, which functions as a source of erythroid cells to support survival of the mouse embryo before that stage, is independent of c-Myb (Mucenski et al., 1991; Palis, 2014). Additionally, a recent study of patients with Trisomy 13, a chromosomal

disorder associated with persistent embryonic and fetal hemoglobin expression demonstrated that the MYB function is dysregulated, suggesting that MYB may play a role in hemoglobin switching (see Section 1.4, (Sankaran et al., 2011a)). Furthermore, *Sox6* and *Bcl11a* have been shown to negatively regulate embryonic and fetal globin gene expression ((Xu et al., 2010; Yi et al., 2006), see details in Section 1.4). To protect the early embryos from exogenous hydrogen peroxide (H_2O_2), primitive and definitive erythroblasts also express different Aquaporin gene family members (Kingsley et al., 2013). These members of the Aquaporin gene family differentially transport glycerol and other small uncharged molecules such as CO_2, urea and H_2O_2. Aquaporins 3 and 8 are expressed in primitive erythroid cells, whereas Aquaporins 1 and 9 are expressed in adult definitive erythroid cells. This differential usage of Aquaporin family members correlates with the ability of primitive erythroid cells (using Aquaporins 3 and 8) to accumulate higher levels of free oxygen radicals, as compared to adult definitive erythroid cells which express Aquaporins 1 and 9. This is because under the hypoxic conditions that exist *in utero,* increased levels of exogenous hydrogen peroxide are produced. Therefore, primitive erythroid cells are likely to have developed this capacity to accumulate oxygen radicals in order to protect the early embryo (Kingsley et al., 2013). Given the manifold differences between primitive and definitive erythroblasts, it is unsurprising that there are so many differentially expressed transcription factors. However, a few of those studied in further detail cause relatively small levels of embryonic and fetal hemoglobin de-repression when mutated and so are unlikely to be solely

responsible for hemoglobin switching. Therefore there is a need for further studies in this area to elucidate the pathways responsible for regulation of the embryonic and fetal hemoglobin genes.

1.3.7 Limitation of the study of human primitive erythropoiesis

While human bone marrow cells can be accessed for the study of HSC-derived definitive erythropoiesis, the possibility to obtain appropriate tissue for the study of human primitive erythropoiesis is far more difficult. This is due to ethical issues related to physical access to human embryos. Consequently, the majority of studies of primitive erythropoiesis have been conducted in mouse, while little is known about human primitive erythropoiesis. Nevertheless, previous studies demonstrate that primitive erythroblasts are first observed in the blood islands of the yolk sac at ~18-20 days of gestation (reviewed in (Luckett, 1978; Palis, 2014)) and become the only circulating erythroid cells in human embryos from 3 to 6 weeks of gestation. Nucleated primitive erythroid cells, however, are found to circulate in the fetus throughout the first trimester (reviewed in (Palis, 2014)). A recent study has shown that ζ-globin positive enucleated primitive erythroid cells were found at a high number (higher than those found in the circulation) in human placental villi between 5-7 weeks of gestation (Van Handel et al., 2010). Reticulocytes and pyrenocytes were also identified in placental villi during the same period and these extravascular primitive red cells are associated with placental macrophages, which contained ingested nuclei (Van Handel et al., 2010). This suggests human primitive erythroid cells also enucleate, like their murine counterparts.

A study of primitive and definitive erythroid cells in 25 human embryos and 6 fetuses revealed that human primitive erythroid cells express the embryonic ζ- and ε-globin genes (Peschle et al., 1985). Surprisingly, unlike mouse primitive erythroblasts, in which the adult α-globin genes are expressed at a higher level than ζ-globin at all stages, in human primitive erythroblasts ζ-globin accounts for ~82% of total α-like globin during early development. However, mirroring their murine counterparts, a maturational globin switch occurs within the human primitive lineage and the ζ- to α-globin switch occurs between 5-7 weeks of gestation preceding the ε- to γ- switch, which occurs at the beginning of definitive erythropoiesis in the fetal liver at 8 weeks of gestation (Peschle et al., 1985).

In vitro differentiation of human embryonic stem (ES) cells into the erythroid lineage can serve as a study model for the early stage of erythroid development. A recent study demonstrated, using a complex immunophenotyping approach, that differentiating ES cells mimic hematopoietic development in the yolk sac. Specifically this consists of a wave of primitive hematopoiesis followed by the emergence of distinct populations of EMP-derived definitive progenitors and a small population of cells with B cell potential (McGrath et al., 2015). The ES cell derived human primitive erythroblasts have similar properties to their primary counterparts, for example they undergo a maturational globin switch from Hb GowerI ($\zeta_2\varepsilon_2$) to Hb GowerII ($\alpha_2\varepsilon_2$) (Qiu et al., 2008). Recently, primitive erythroblasts derived from human ES cells have also been shown to require EPO for their terminal maturation as was demonstrated in primary primitive erythroid cells. Over the past few years reprogramming and differentiation of human “induced

pluripotent stem cells" (iPSs) *in vitro* has become increasingly common for disease modeling and study of developmental processes, including erythropoiesis. iPS cells are very similar to ES Cells (Carey et al., 2011; Chang et al., 2011; Maherali and Hochedlinger, 2008; Pera, 2008; Rossant, 2011). However, when differentiated to the erythroid lineage, unlike mouse ES cells, human iPS cell derived red blood cells express little or no adult β-globin and enucleation is generally incomplete (Chang et al., 2011; Chang et al., 2010; Lapillonne et al., 2010). Thus, differentiation of iPS cells *in vitro* produces erythroid cells that are phenotypically most similar to embryonic/fetal stages of ontogeny.

1.4 Hemoglobin switching

The change in the combination of globin gene expression in a developmentally regulated manner, resulting in sequential assembly of different globin tetramers, is known as hemoglobin switching. Over the past four decades, there has been a significant expansion of research attempting to identify underlying mechanisms. This is mainly because reactivation of functional fetal/embryonic globin in adult stage erythropoiesis may provide a potential long-term cure for individuals suffering from hemoglobinopathies including thalassemia and sickle cell disease (SCD).

1.4.1 β-globin switching

In humans, eight functional globin genes from the α- and β-globin clusters give rise to six different globin tetramers, embryonic Hb Gower I ($\zeta_2\varepsilon_2$), Gower II ($\alpha_2\varepsilon_2$), and Hb Portland I ($\zeta_2\gamma_2$), fetal hemoglobin (Hb F; $\alpha_2\gamma_2$) and adult Hb A ($\alpha_2\beta_2$) and Hb A_2 ($\alpha_2\delta_2$). In some cases, a very small amount (< 0.15%) of

embryonic ζ-globin is detected in fetal and newborn blood (Chui et al., 1989). When this occurs in the presence of β-globin chains, a 7th type of globin tetramer, Hb Portland II ($\zeta_2\beta_2$) can be detected. The β-globin locus lies on human chromosome 11 and the genes within the cluster are arranged in developmental order, 5′-ε-Gγ-Aγ-ψβ1-δ-β-3′ (Figure 1.1). Two stages of β-globin switching occur during development, the first being the embryonic to fetal switch, in which there is a change from ε- to Gγ- and Aγ-globin gene expression and the second being the fetal to adult switch, in which a change from Gγ-, Aγ- to β- and δ-globin expression occurs. The ε- to γ-globin switch begins very early in gestation, as HbF is readily detected in 5-week-old human embryos and it is completed well before the 10th week of gestation (Gale et al., 1979; Hecht et al., 1966; Huehns et al., 1964). This was attributed mainly to the switch from primitive erythropoiesis in yolk sac to fetal definitive erythropoiesis in the liver. Similarly, the γ- to β-globin switch occurs with the shift from fetal to adult bone marrow erythropoiesis: β-globin synthesis increases to approximately 10% of total hemoglobin by 30-35 weeks of gestation and HbF accounts for 60-80% of total hemoglobin at birth. By 2 years of age, HbA comprises the majority of hemoglobin and HbF decreases to the level of 0.5-1% of total hemoglobin (reviewed in (Stamatoyannopoulos et al., 2009)).

The human ε-globin gene is orthologous to murine εy and the human γ-globin gene is orthologous to βh1 whereas the β-globin gene is orthologous to the murine β minor and β major genes. However, the difference in ontogeny of erythropoiesis between human and mouse needs to be taken in to consideration when studying globin switching using the mouse as a model.

Unlike in humans, there is an absence of fetal-stage erythropoiesis in mouse. Thus, only the embryonic to adult globin switch occurs, in this process the εy and βh1 genes become repressed and β-genes become upregulated.

If switching at the α- and β-globin loci could be reversed, then this may offer alternative therapies for many of the individuals affected with thalassemia and SCD. This is becoming an increasingly important issue because although there has been a continuous improvement of clinical outcomes, stem cell transplantation is considered the only potentially curative treatment for such individuals (Michlitsch and Walters, 2008). However, this treatment is not widely accessible, particularly in the developing world (Sankaran and Nathan, 2010). Therefore there is an increasing need for clinical and basic research for alternative treatments for these transfusion-dependent patients. A variety of clinical research has demonstrated that the severity of β-thalassemia is ameliorated when production of HbF is increased. Observations have been made that infants with β-thalassemia only begin to show symptoms after the expression of HbF declines in the months following birth and that β-thalassemia patients with elevated production of HbF showed a milder clinical course (Weatherall, 2001; Weatherall and Clegg, 2001). Large epidemiological studies have shown that there are several genetic modifiers that increase HbF and ameliorate the symptoms of β-thalassemia (Galanello et al., 2009; Nuinoon et al., 2010; Premawardhena et al., 2005). As a result, there have been long-standing efforts to develop approaches to reactivate HbF with the aim of aiding those affected with severe forms of β-thalassemia. A range of clinical compounds have been investigated and some have been shown to have HbF induction potential both in *vivo* and in clinical trials.

However, the exact molecular mechanisms by which these compounds elevate HbF are not well understood and their potential toxicities are of major concern. This includes the DNA methylation inhibitor 5-azacytidine ((DeSimone et al., 1982; Ley et al., 1982; Ley et al., 1983), reviewed in (Sankaran and Orkin, 2013)), an S-phase inhibitor (hydroxyurea) (Platt, 2008; Platt et al., 1984) and β-hydroxybutyrate a short-chain fatty acid that functions as a histone deacetylase (HDAC) inhibitor (Fathallah et al., 2007; Perrine et al., 1993; Perrine et al., 1994; Sankaran and Nathan, 2010; Sher et al., 1995). More recently, the majority of studies in this field have been focusing on more effective and specifically targeted approaches for HbF induction. These studies have emerged following an improved knowledge of the molecular mechanisms by which the fetal-to-adult globin switch is controlled. Important clues to identify which transcription factors are responsible for such processes came about from studies of naturally occurring genetic variations in human populations. Genome-wide association studies (GWAS) of genetic variation in individuals with elevated HbF levels led to identification of three genomic loci harboring polymorphisms that influence HbF levels. These reside in a region within the *BCL11A* gene, an intergenic region between the *HBS1L* and *MYB* genes and a region within the β-globin locus (Galarneau et al., 2010; Lettre et al., 2008; Menzel et al., 2007; Thein et al., 2007; Uda et al., 2008). These studies were critical in showing that BCL11A and MYB are key regulators of the fetal to adult hemoglobin switch. Apart from a small number of individuals with *KLF1* mutations (Arnaud et al., 2010; Viprakasit et al., 2014), no individuals with elevated ζ-globin have been identified to allow equivalent studies to identify novel regulators of ζ-globin.

The high-HbF phenotype is associated with reduced *BCL11A* expression. Furthermore, it has been shown that elevated levels of γ-globin expression in primitive and fetal definitive erythroid cells is associated with low expression or absence of the full length isoforms of BCL11A suggesting this protein acts as a developmental stage-specific repressor of the γ-globin gene (Sankaran et al., 2008). This is supported by the observation of robust γ-globin expression when *BCL11A* mRNA is depleted (by short-hairpin RNA, shRNA) in primary adult erythroid cells (Sankaran et al., 2008). Although there is a divergence of normal ontogeny of globin expression observed in the human and the mouse model, an evolutionarily conserved role for BCL11A in globin gene silencing and switching has been shown in mice (McGrath et al., 2011; Sankaran et al., 2009). Mice harboring a human β-globin locus transgene and lacking Bcl11a appear to have normal erythropoiesis, but fail to fully silence the embryonic globin genes in definitive erythroid cells and allow persistent expression of γ-globin from the intact human β-globin locus (reviewed in (Sankaran and Orkin, 2013)). A recent study using high-resolution chromatin immunoprecipitation (ChIP) analysis demonstrates that BCL11A binds the locus control region (LCR) of the β-globin locus, the ε-globin gene and the intergenic regions between the γ- and δ-globin genes in human erythroid progenitors (Xu et al., 2010). This study suggests that BCL11A silences γ-globin through interactions with SOX6, a transcription factor that binds chromatin at the proximal γ-globin promoter, as well as through long-range chromosomal interactions throughout the β-globin gene cluster (Xu et al., 2010). Moreover, a previous study demonstrates, using a proteomic approach, that BCL11A interacts with the NuRD chromatin remodeling and

repressor complex containing HDACs 1 and 2, the erythroid transcription factors GATA1 and FOG1 and the nuclear matrix component Matrin-3 in erythroid progenitors (Sankaran et al., 2008). However, SOX6 interacts with BCL11A, GATA1 and FOG1, but not the NuRD component MTA2 and MBD3 (Xu et al., 2010). It is thought that the physical interaction between BCL11A and SOX6 may help recruit BCL11A and NuRD repressor complexes to the proximity of γ-globin gene (Xu et al., 2010). Another study mapping a variety of deletions within the human β-globin locus that either result in δβ-thalassemia or HPFH, with increased HbF production illustrates that a 3-kb region 5′ of the δ-globin gene is necessary for silencing of the γ-globin genes. This region contains binding sites for BCL11A and its partners such as GATA1 and HDAC1, thus providing insight into mechanisms by which BCL11A functions to silence HbF (Sankaran et al., 2011b).

The expression of *BCL11A* appears to be regulated by KLF1. *KLF1* encodes a zinc finger protein, which binds differentially to embryonic, fetal and adult globin gene promoters during development. Decreased *Klf1* levels resulting from knockout of a *Klf1* enhancer (termed EHS1), reduced the level of *Bcl11a* and thereby elevated embryonic globin gene expression (εγ) and led to persistent expression of γ-globin in transgenic mice containing a human β-globin locus (Zhou et al., 2010). The study also showed that a similar effect occurs in primary human erythroid cells when using an shRNA approach. Furthermore, the investigators demonstrated that elevation of embryonic and fetal β-like globins in transgenic mice with a hypomorphic allele of *Klf1* occurs through both the direct effects of Klf1 at the β-globin locus and also via indirect effects mediated through reduced expression of Bcl11a (Zhou et al.,

2010). Furthermore, KLF1 can function as a transcriptional regulator of BCL11A, as shown by a study of a *KLF1* nonsense mutation in the family with an unlinked form of HPFH (Borg et al., 2010). This study showed reduced levels of both KLF1 and BCL11A in primary erythroid progenitors derived from these patients compared to the levels found in primary erythroblasts derived from unaffected controls. Although rare variants in *KLF1* are indeed associated with elevation in HbF, this does not appear to occur consistently or to the same extent even with similar mutations (Gallienne et al., 2012). Two additional Klf family members (Klf2 and Klf8) have been shown to be associated with developmental stage-specific globin expression. Firstly *Klf2* when knocked out in concert with *Klf1* further reduces εy- and βh1-globin levels demonstrating that these two Klf family members act in concert to positively regulate embryonic β-like globin expression (Pang et al., 2012). This reduced expression has been suggested to result from the interactions of Klf1 and Klf2 with c-Myc. Embryonic (*Hba-x, Hbb-y, Hbb-bh1*), but not adult globin genes are de-repressed in *Ter119*$^{+}$ E13.5 fetal liver cells of mice harboring homozygous deletion of both *Klf3* and *Klf8* suggesting they also play roles in silencing embryonic globin expression during development (Funnell et al., 2013).

In addition to BCL11A, the intergenic region between the genes *HBS1L* and *MYB* on chromosome 6 has been identified in GWAS studies of individuals with persistent expression of HbF (Lettre et al., 2008; Menzel et al., 2007; Thein et al., 2007; Uda et al., 2008). Although overexpression of *HBS1L* did not appear to affect γ-globin expression in K562 erythroleukemia cells, overexpression of *MYB* was observed to affect the level of γ-globin produced

in these cells (Jiang et al., 2006). Consistent with these findings, the HBS1L-MYB intergenic region identified in the GWAS studies contains a variety of regulatory elements that have been suggested to have an important role in regulating expression of *MYB* in erythroid progenitors (Mukai et al., 2006; Stadhouders et al., 2012; Wahlberg et al., 2009). Direct knockdown of MYB in primary adult erythroid progenitors resulted in a marked increase in γ-globin production (Sankaran et al., 2011a). The mechanism by which MYB regulates HbF levels remains unclear. This may be due to an effect on the kinetics of erythropoiesis or may also occur as a result of a direct effect within the β-globin locus ((Higgs and Wood, 2008a), reviewed in (Sankaran and Orkin, 2013)).

Several other molecules have been identified to play roles in γ-globin gene regulation in studies using cell culture approaches or mouse models. In mouse and human definitive erythroid cells TR2 and TR4, the orphan nuclear hormone receptors binding to the repeat elements (DRED) within the γ-globin gene promoters, have been suggested to act as repressors of embryonic/fetal β-like globin gene expression (Cui et al., 2011; Tanabe et al., 2002; Tanabe et al., 2007). TR2 and TR4 are thought to function with an array of transcriptional co-repressors, including DNMT1, the NuRD, LSD1/CoREST repressor complex, HDAC3 and TIF1β (Cui et al., 2011). The mechanisms underlying these findings are yet to be established and further human genetic evidence for their involvement in normal hemoglobin switching may provide further insight into the role of these proteins. Furthermore, molecules including COUP-TFII, FOP1 and NF-E4 have also been identified to be associated with embryonic/fetal β-like globin gene expression in either mouse or human cell

culture studies, however further work is needed to elucidate their mechanism of action (Filipe et al., 1999; Jane et al., 1995; van Dijk et al., 2010; Zhao et al., 2006).

The insights into the molecular mechanisms underlying β-globin switching obtained from these studies provide novel candidates for the development of specific targeted therapies and gene therapies to ameliorate the clinical course of individuals affected with β-thalassemia.

1.4.2 α-globin switching

In mice and humans the α-globin gene cluster (located on chromosome 11 in mice and 16 in humans) contains three functional globin genes: ζ, α2 and α1. All three α-like genes are expressed in primitive erythroblasts, which originate in the blood islands of the yolk sac, with ζ-globin accounting for approximately 50% of the total α-like (α+ζ) transcription at E8.5 in mice. At E10.5 this ratio appears unchanged in maturing primitive erythroblasts (Kingsley et al., 2006). By E12.5 the ζ-globin gene is down-regulated in primitive erythroblasts and accumulated ζ-globin transcripts account for ~10% of the α-like (α+ζ) transcripts suggesting maturational switching in this lineage is initiated before this stage. At E10.5 the site of active erythropoiesis migrates to the fetal liver, where HSC-derived definitive erythropoiesis begins. The ζ-globin gene is constitutively silenced in this lineage while the 2 α-globin genes remain transcriptionally active and continue to express at high levels into adulthood (Trimborn et al., 1999). In humans, definitive erythropoiesis starts in the liver from 6 weeks of gestation onward (Peschle et al., 1985) concurrently with switching of genes within the α-globin cluster. ζ-globin is quiescent at birth in humans and mice. However, very small amounts of ζ-globin mRNA are

present throughout fetal life (Hill et al., 1985) and ζ-globin (as Hb Portland) can be detected in the cord blood of non-thalassemic newborns (Chui et al., 1989).

In contrast to fetal globin, until a report of a gain of function mutation in the transcription factor KLF1 (Arnaud et al., 2010) and a recent report of patients harboring loss of function mutations in this gene (Viprakasit et al., 2014), no naturally occurring mutations causing persistent expression of ζ-globin have been identified. Therefore, there is no such large cohort study identifying naturally occurring variants that might underlie expression of ζ-globin. Study of a transgenic mouse model has demonstrated that embryonic ζ-globin can functionally substitute for α-globin in adult erythroid cells (Russell and Liebhaber, 1998). This suggests that reactivation of ζ-globin expression could ameliorate the clinical symptoms of patients with severe α-thalassemia and thereby be considered as a potential therapeutic approach. However, unlike β-globin switching, the pathway underlying embryonic globin expression and the ζ- to α-globin switch is poorly understood. Klf1 and BCL11A have both been shown to bind specific sites in the α-globin locus in mice (Tallack et al., 2010) and humans (Jawaid et al., 2010), however, their effect on switching the α-globin genes is unclear. Evidence that KLF1 may also affect regulation of the ζ-globin gene comes from recent cohort study of 8 patients with chronic hemolytic anaemia or chronic non-spherocytic hemolytic anemia who were found to be compound heterozygotes for *KLF1* mutations. In addition to raised levels of fetal hemoglobin, these patients were also found to have increased levels of the embryonic ζ- and ε-globins (Viprakasit et al., 2014). Further evidence that KLF1 may affect expression of ζ-globin comes from a case of

congenital dyserythropoietic anemia type IV in which a p.E325K change in *KLF1,* leading to a gain of function, was also found to underlie expression of embryonic ζ- and ε-globin (Arnaud et al., 2010). It has been reported in a recent study that the 2 other members of the Klf family, Klf3 and Klf8, may play roles in regulating embryonic globin gene expression, since knockout of these 2 transcription factors in fetal liver definitive erythroid cells de-repressed embryonic globin (ζ, ε, γ) synthesis (Funnell et al., 2013). Because expression of *Klf3* and *Klf8* is driven by Klf1, it is possible that disruption of Klf1 reduces Klf3 and Klf8 levels, thereby leading to de-repression of the embryonic globins. Although expression of the embryonic globin genes was normal in *Klf1* null mutant embryos during primitive erythropoiesis, owing to lethality it is not clear whether ζ-globin is ectopically expressed in these mice in the adult stage (Nuez et al., 1995; Perkins et al., 1995).

More recently another candidate gene, *Lrf,* has been implicated in the globin-switching process. Lrf (leukemia/lymphoma related factor), also known as Pokemon and Zbtb7a, is a POZ and Krüppel (POK)-type transcription factor that has multiple functions in hematopoietic development, oncogenesis and humoral immunity (reviewed in (Lee et al., 2013)). This study reported the necessity of Lrf for hematopoietic stem cell (HSC) maintenance. Provocatively, *Hba-x* and *Hbb-bh1* were also identified as being dramatically upregulated in *Lrf*-decifient long-term HSC (LT-HSC) (Lee et al., 2013), suggesting a role for *Lrf* as a repressor of the α-like as well as the β-like embryonic globin genes. However, more detailed studies are required to further investigate this finding.

Further study of the mechanisms underlying embryonic globin expression and ζ- to α-globin switching is of great interest as it may lead to the development of therapeutic approaches that could ameliorate severe α-thalassemias. Previous studies of a transgenic mouse model containing the entire human α-globin locus demonstrated that the human ζ-globin, as well as the mouse ζ-globin gene, switches off normally in a developmentally regulated manner (Sharpe et al., 1993b). These data suggest that the mouse serves as a good study model for regulation of the ζ-globin gene and ζ- to α-globin switching.

1.5 The ζ-globin gene and its promoter

1.5.1 The structure of ζ-globin gene

The ζ-globin gene resides at the 5′ end of the highly conserved α-globin cluster. The gene comprises of 3 coding exons, which are translated to a 142 amino acid protein in both mice and humans. Approximately 8 kb 3′ of the ζ-globin gene is the non-expressed Ψζ1 which is a pseudogene and which has a gene-like structure with sequence homology and an exon-intron structure similar to the actively expressed ζ-globin gene (Zhang and Gerstein, 2004). This Ψζ1 gene, similar to other pseudogenes in the α- and β-globin cluster (Ψα1, Ψβ1), appears to have arisen by gene duplication events (Forget and Hardison, 2009). The Ψζ1-gene differs from the ζ-gene by only 3 bp in the sequence corresponding to the exons, one of which converts the codon for amino acid 6 into a stop codon (Proudfoot et al., 1982).

In humans, the pattern of intron sizes of the ζ-globin gene and Ψζ1 gene differs from that of the other α-like globin genes. The introns of the adult α-genes are ~100 bp in length with the first intron being slightly shorter than the

second. Contrastingly, the ζ-globin gene intron 1 is 886 bp long and contains 12 iterations of a simple 14 bp sequence (ACAGTGGGGAGGGG), while intron 2 is 239 bp long and contains 35 copies of a 5 bp repeat sequence (CGGGG) (Proudfoot et al., 1982). The highly repetitive pattern of the repeat sequence in intron 2 results in a base composition of 92% G/C and the presence of 46 CpG dinucleotides (Proudfoot et al., 1982). In mice, the structure of the ζ-globin gene mirrors that of the human, however, there is no evidence of the iterative repeats in intron1.

Because the α-globin gene is expressed throughout all developmental stages, it is possible that there may be some functional redundancy between α- and ζ-globin in early development. Mice homozygous for a deletion of the ζ-globin gene show variable viability that is dependent on strain background, however, the deleted allele contained the neomycin gene as a result of the targeting process and this may confound the result (Leder et al., 1997). Furthermore, no coding variants of the ζ-globin gene have been reported, this high level of conservation suggests a critical role for this protein. Human studies have also indicated that deletions and duplications in the ζ-gene are not associated with any defined phenotype, interestingly no individuals homozygous for ζ-globin gene deletions were reported (Fei et al., 1989; Felice et al., 1986). Therefore the question of whether ζ-globin is required for normal mammalian development and viability remains unanswered. Large deletions spanning the ζ- and α-globin genes cause early embryonic lethality, presumably because no functional hemoglobin tetramer can be synthesized (Forget and Hardison, 2009).

1.5.2 The ζ-globin promoter

Interspecies comparison of α- and ζ-globin promoter sequences suggests that the ζ-globin gene promoter is larger and more complex than that of the α-globin paralogs (Figure 1.5, 1.6). This implies that a greater number of *trans*-acting factors bind this region and ζ-globin may be less reliant on distal *cis*-acting regulatory elements. Previous studies performed in K562 cells (a human ζ-globin expressing erythroleukemia cell line) demonstrated using DNaseI and gel mobility shift assays that there are binding sites for GATA1 (previously named NF-E1), Sp1 and CCAAT box-binding factors (possibly TFCP2) located in the ζ-globin promoter (Yu et al., 1990). Subsequent studies have also identified potential functional domains within the ζ-globin promoter. Deletion experiments in K562 cells demonstrated that the ζ-globin upstream regulatory elements (URE) are positioned within a 60 bp region located between 220 and 279 bp 5′ of the ζ-globin transcription start site (Sabath et al., 1996). This study also indicated the presence of multiple transcription factor binding sequences in this region. It has been demonstrated using electrophoretic mobility shift assays that GATA-1 binds at position -230, and Sp1 and an unidentified factor bind a CCACC site at -240 (Sabath et al., 1995). This unidentified factor is distinct from two other CCACC binding factors, EKLF (KLF1) and BKLF (KLF3) (Sabath et al., 1996). The GATA-1 site at -230 is only partially responsible for URE activity, since deletion of sequences from -417 to -207 5′ of the ζ-globin mRNA cap site decreases promoter activity by 90% in K562 cells, while point mutation of a GATA-1 site at -230 alone decreases promoter activity to ~50% (Sabath et al., 1995). In addition, mutation of the CCACC site at -240 results in no decrease

in promoter activity, suggesting some other elements are necessary for full URE function (Sabath et al., 1996). A subsequent study from the same group identified a novel DNA-binding protein, annotated as URE binding factor (URE-BF), which interacts with a site in the -269 to -255 region (Sabath et al., 1996). URE-BF appears to interact with a GATA factor, since formation of the URE-BF complex can be prevented by the presence of unlabelled oligonucleotides containing GATA sites (Sabath et al., 1996). In addition to the 60 bp URE region, a previous study demonstrated that a positive regulatory element located between 207 and 417 bp 5′ to the transcription start site is necessary for high-level ζ-globin promoter activity in transiently transfected K562 cells (Sabath et al., 1995). However, this element does not seem to regulate ζ-globin expression during development, since it can be deleted without disrupting embryonic-specific expression in transgenic mice (Pondel et al., 1992; Sabath et al., 1993).

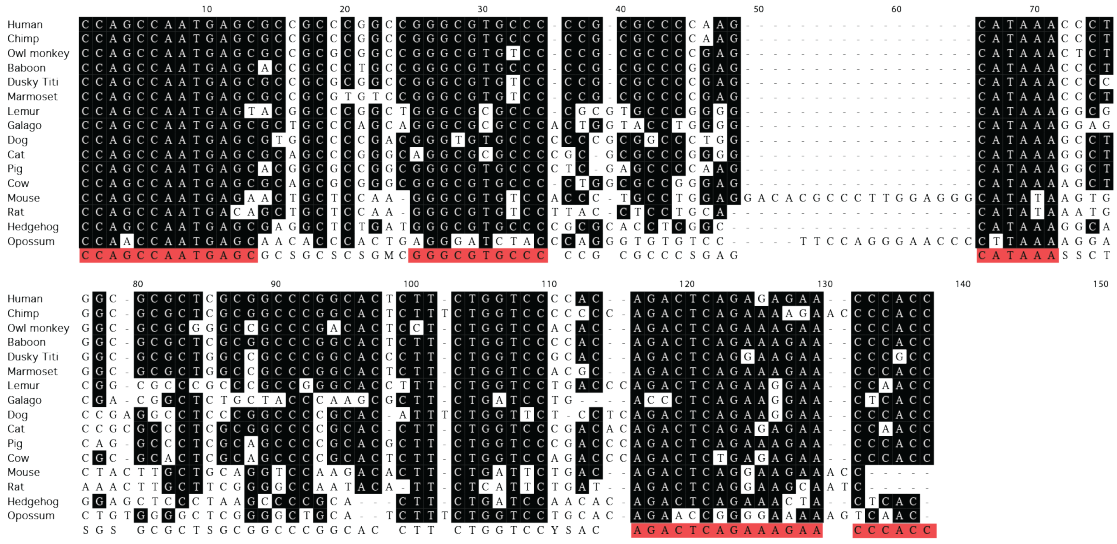


Figure from Jim Hughes

Figure 1.5 The α-globin gene promoter

Multispecies alignment of the α-globin promoter sequences. Conserved bases are shaded and predicted transcription factor binding sites are underlined in red.

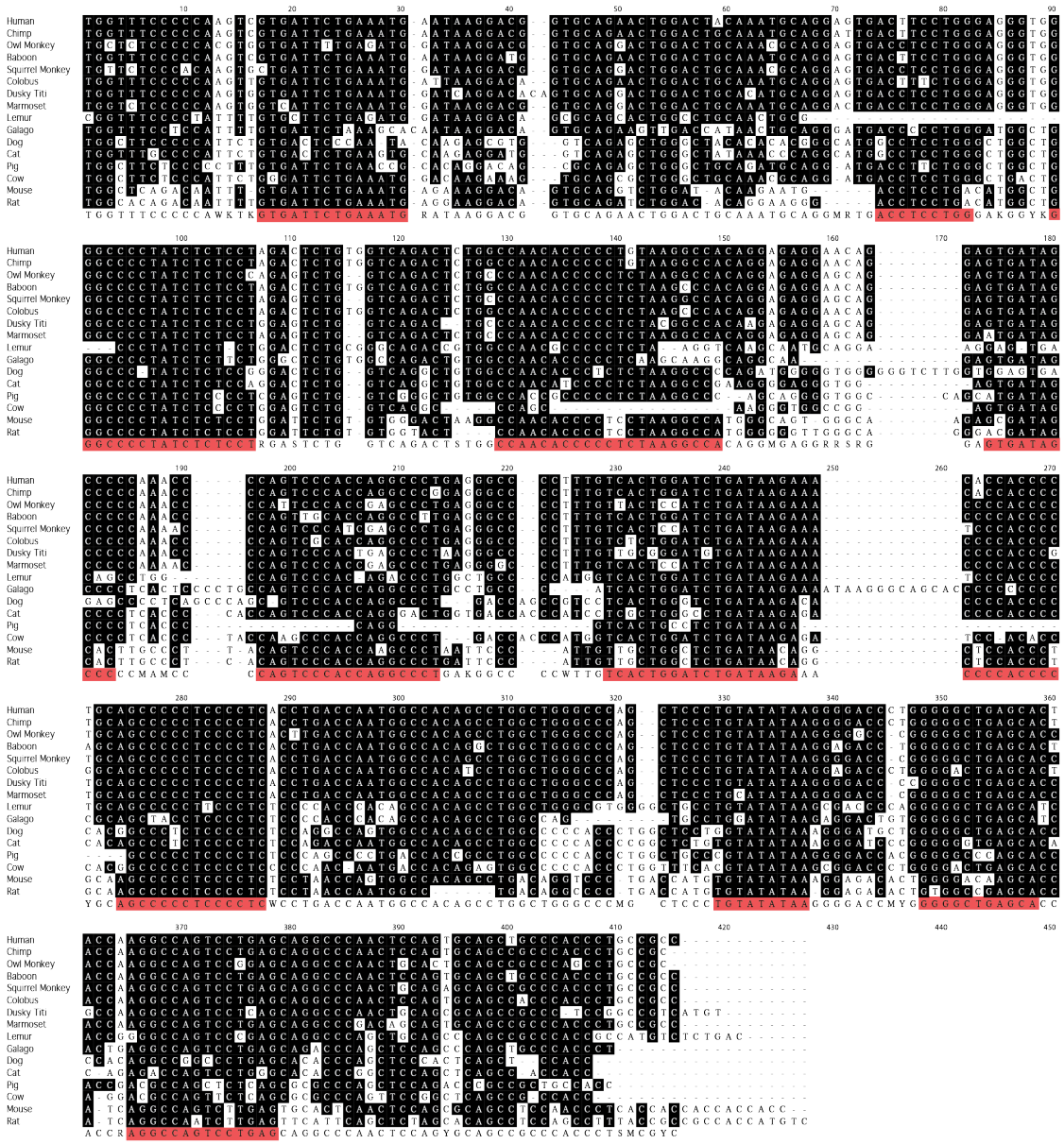


Figure from Jim Hughes

Figure 1.6 The ζ-globin gene promoter

Multispecies alignment of the ζ-globin promoter sequences. Conserved bases are shaded and predicted transcription factor binding sites are underlined in red. The ζ-globin gene promoter is much more complex than the α-globin gene promoter suggesting a greater number of *trans*-acting factors bind this region and ζ-globin may be less reliant on the distal *cis*-acting regulatory elements.

1.5.3 Regulation of ζ-globin expression by cis-acting regulatory elements

Several lines of evidence suggest that transcription of the ζ-globin gene is also mediated by the major α-globin upstream regulatory element HS-40 (MCS-R2). HS-40 was initially demonstrated as a classical enhancer for the ζ-globin as well as the α-globin promoters (Pondel et al., 1992; Ren et al., 1993; Zhang et al., 1993). The ζ-globin gene is expressed in transgenic mice containing the human ζ-, α_2, α_1 genes and the HS-40 sequence (Gourdon et al., 1994). This finding is also supported by a transient expression assay in K562 cells, in which a high-level of ζ-globin promoter activity is observed in the presence of the HS-40 element (Zhang et al., 1993). Site-directed mutagenesis also demonstrated that the interaction between the HS-40 enhancer and the ζ-globin promoter is mediated by the two GATA-1 binding motifs located at 230 and 104 bp 5′ to the ζ-globin mRNA cap site (Zhang et al., 1993). The same study also mapped the functional domains of HS-40 and demonstrated that one GATA-1 motif and two NF-E2/AP1 motifs together with a conserved GT dinucleotide form the functional core of HS-40 in erythroid-specific activation of the ζ-globin promoter. Furthermore, a targeted mutation illustrated that the enhancer function of one of the two NF-E2/AP1 motifs of HS-40 is mediated through its binding to NF-E2, but not the AP1 transcription factor (Zhang et al., 1993). Subsequent systematic analysis of site-directed mutagenesis revealed that HS-40 does not only exert enhancer function on the ζ-globin promoter, but also negatively regulates ζ-globin expression in K562 cells. This occurs through a dual positive and negative regulatory effect of the 3′ NF-E2/AP1 motif within the HS-40 element on ζ-globin promoter

activity and this dual function appears to be modulated by differential binding of the ubiquitous AP1 factor and the erythroid-enriched NF-E2 factor (Rombel et al., 1995; Zhang et al., 1995). As a result, it has been shown in an experiment using fetal and adult transgenic mice that developmental silencing of the human ζ-globin gene is, in conjunction with *cis*- interaction with the ζ-globin promoter, in part modulated by the competitive DNA binding of different members of the NF-E2/AP1 transcription factor family (Huang et al., 1998).

The roles of the remaining α-globin upstream regulatory elements on ζ-globin expression are poorly understood. However, a recent study suggests that MCS-R4 may be involved in ζ-globin expression as K562 cells, which express ζ-globin, have increased acetylation of this gene in addition to increased acetylation of the MCS-R4 element compared to human primary definitive erythroblasts (De Gobbi et al., 2007).

1.6 Post-transcriptional processing of ζ-globin mRNA

A previous study has suggested that post-transcriptional control of the ζ-globin mRNA, leading to accelerated clearance of the transcripts, may contribute to full developmental silencing of the ζ-globin gene (Russell et al., 1998). In transgenic mice with constitutive expression of α- and ζ-globin (Russell et al., 1998), it was demonstrated that single nucleotide differences in pyrimidine-rich elements (PREs) within the 3' UTRs of the α- and ζ-globin genes contributed to a 6-fold reduction in the affinity of the ζ-PRE for α-CP, a cytoplasmic mRNA-binding protein that contributes to the high stability of α-globin mRNA (Kiledjian et al., 1995; Russell et al., 1998). Importantly, although ζ-globin mRNA stability increased, it was not stabilized to the same extent as α-globin mRNA by exchange of the ζ-PRE for the corresponding α-

globin element, suggesting there may be other mRNA stabilizing factors. A recent study, using cultured erythroid cells, postulated that this post-transcriptional control region is located in a tetranucleotide 3′ UTR motif of the ζ-globin gene (He et al., 2014). This ζ-globin mRNA stabilizing motif is targeted by AUF1, a ubiquitous RNA-binding protein that enhances the half-life of adult β-globin mRNA, suggesting commonalities in the post-transcriptional process between the α-like and β-like genes (He et al., 2014). Transgenic human ζ-globin can also be produced at sufficient levels to support viability in α-globin null mice (Russell and Liebhaber, 1998), indicating that the half-life of the encoding mRNA is sufficiently long, although less stable than the α-globin mRNA, to ensure that biologically relevant levels of ζ-globin are produced (reviewed in (He et al., 2014)).

1.7 Functional properties of ζ-globin

ζ- together with embryonic ε- and fetal γ-globin chains give rise to two distinct embryonic globin tetramers, Hb Gower I ($\zeta_2\varepsilon_2$) and Hb Portland I ($\zeta_2\gamma_2$), which are synthesized in human yolk sac-derived primitive erythroid cells between 4 and 14 weeks of gestation (Steinberg and Nagel, 2009). Subsequently, these hemoglobins start to be replaced by HbF, which can be detected in fetal blood (AGM-derived definitive erythroid cells), but some ζ- and ε-globin chains can be found in definitive erythrocytes (Luo et al., 1999). Hb Portland I ($\zeta_2\gamma_2$) exhibits a higher O_2 affinity and a lower rate of O_2 dissociation as compared to HbA. The presence of the a glycine at position 38 of ζ-globin has been reported to make a major contribution in producing higher O2 affinity within the ζ-chain-containing human embryonic hemoglobins (Steinberg and Nagel, 2009). Previous experiments, in which the ζ-globin gene was constitutively

expressed in definitive erythroid cells, demonstrate that ζ-globin assembles with adult β-globin to form Hb Portland II ($\zeta_2\beta_2$), which exhibits O_2-binding and allosteric properties that differ modestly from those of Hb A, but that remain fully compatible with normal adult physiology (He et al., 2000; He and Russell, 2001). This finding further suggests that ζ-globin (as Hb Portland) can functionally substitute for α-globin (Hb A).

1.8 Summary and outline of the thesis

The BHFS is a major health problem in some areas of the world where a high prevalence of deletional α-thalassemia is observed, including the South East Asia subcontinent. The BHFS is regarded as a universally fatal disorder. However, over the last decade, with ever-improving neonatal screening and intensive care, there have been an increasing number of reports of infants who have been treated with blood transfusion and who have survived with BHFS. There is currently a need to review all long-term survivors with BHFS to understand the full clinical spectrum and the outcome of treating this condition using intervention with blood transfusion and HSC transplantation.

These patients also raise important issues that are both of basic scientific interest and potential clinical importance. The embryonic ζ-globin gene is unaffected by most of the α^0-thalassemia deletions, including the most common, termed the Southeast Asian (SEA) deletion (Figure 1.2) making it theoretically available for reactivation in adults with α-thalassemia. In addition, we have recently become aware of an exceptional case of BHFS who survived gestation and early life in a largely transfusion independent manner, most likely through persistent expression of embryonic globin. This suggests that expression of ζ-globin might mitigate clinically significant α-thalassemia

and may be of value to developing fetuses with otherwise lethal BHFS. At present very little is known about how the embryonic genes are regulated and the majority of previous studies have been performed only in transgenic models or cell lines.

Aims of this thesis

1. To fully document the natural history of long-term survivors of the BHFS throughout the world and to gain insight into whether this disease should be considered salvageable
2. To investigate how the ζ-globin genes are normally activated in primitive erythropoiesis and repressed in definitive erythropoiesis

Aims of each results chapter

Chapter 3 – To fully document the natural history and outcomes of long-term survivors of the BHFS throughout the world, with the ultimate aim of providing evidence-based advice for parents who wish to consider continuing with BHFS pregnancies and to optimize management for BHFS patients.

Chapter 4 – To characterize the *cis*-regulatory landscape of the ζ-globin gene in murine primary primitive erythroid cells.

Chapter 5 – To identify novel *trans*-acting factors regulating ζ-globin expression by comparing the transcriptomes of primitive and HSC-derived definitive erythroid cells.

Chapter 6 – To integrate information on the *cis*- and *trans*-regulation of ζ-globin in an attempt to gain insight into an exceptional case of the BHFS who remarkably survived his first year of life with minimal blood transfusion, most

likely because of persistent ζ-globin expression. By studying this patient, I also aim to gain more insight into how the ζ-globin gene is regulated.

Chapter 2 : Materials and Methods

2.1 The BHFS survivor registry

The BHFS survivors included in this registry were divided into 2 categories:

2.1.1 Published cases

All literature reporting survival of the BHFS were reviewed. Patients with the BHFS who survived beyond one month of age at the time of report were included in this registry. I contacted the corresponding author of each report via email to request updated clinical information of each of the patients.

2.1.2 Unpublished cases

Unpublished cases of the BHFS survivors have come to our attention by means of personal communication. Most of them have been brought to the attention of Professor Douglas Higgs by our clinical collaborators. I have requested the physicians taking care of the cases to complete the clinical record form. Patient information sheets and informed consent forms were provided to those physicians who have requested them. Ethical approval for the studies presented here was provided by the Oxfordshire Research Ethics Committee (reference: MREC 03/8/097) to Professor Douglas Higgs.

2.2 DNA procedures

2.2.1 Polymerase chain reaction (PCR)

Routine PCR amplification was carried out in 20 µl reactions containing 20 ng of DNA template, 1×FastStart PCR reaction buffer (Roche), 1.5 mM $MgCl_2$,

primers at 0.5 µM, and dNTPs at 200 µM (final concentrations), with 0.75 U of FastStart Taq DNA polymerase (Roche). PCR conditions consisted of an initial incubation at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 30 sec, annealing for 30 sec, and extension at 72°C for 30 sec, with a final extension for 10 min. In most circumstances, an annealing temperature of 63°C was used, however this was varied as necessary. DMSO was added to 10% of final volume and cycling conditions were varied where optimization was required.

2.2.2 Gel electrophoresis

Amplification products and other DNA preparations were analyzed using agarose gel of appropriate concentration (generally between 0.8-3%). Nucleic acid was visualized with 200 µg/ml ethidium bromide using long or short wave transillumination as necessary, images were taken using GelDoc equipment (Biorad).

2.2.3 Sanger sequencing

Sequencing reactions contained 200-500 ng of amplified DNA template and used BigDye Terminator mix version 3.1 (Applied Biosystems), and 3.2 pmol of the relevant sequencing primer. Reaction products were precipitated and washed with 70% ethanol before analysis. Analysis of DNA fragments was performed using an ABI PRISM-3730 DNA sequencer. Sequence data were visualized using Sequencher software 5.0.1 (Gene Codes Corporation).

2.3 RNA procedures

2.3.1 RNA extraction

Primitive erythroblasts, fetal liver-derived definitive erythroblasts and erythroblasts differentiated from the patient-derived iPSCs were lysed and homogenized at room temperature in 1 ml of TRIzol reagent (Ambion) per 10^7 cells before being frozen on dry ice and stored at -80°C. Subsequently, homogenized cells were thawed on ice and 1/5 volumes of chloroform was added. Mixtures were shaken vigorously, incubated at room temperature for 2-3 min and centrifuged at 13,000 rpm in a bench top microfuge at 4°C for 20 min. The upper aqueous phase containing RNA was carefully aspirated and transferred to a fresh tube where an equal volume of isopropanol was added. Solutions were mixed by inversion, incubated at room temperature for 10-15 min and centrifuged at 13,000 rpm at 4°C for 20 min. Supernatant was carefully aspirated and discarded. Pellets were washed with 1 ml of 70% ethanol and centrifuged at 8,000 rpm at 4°C for 20 min. Supernatant was aspirated and tubes were briefly centrifuged and any remaining ethanol removed. RNA pellets were air-dried for 1-5 min until a transparent glassy appearance was observed. Dry RNA pellets were dissolved in DEPC-treated water (Ambion) and either stored at -80°C or immediately processed for downstream application. RNA concentrations were determined with a Nanodrop spectrophotometer (Thermo Scientific) prior to further use or storage.

2.3.2 DNase I treatment

All extracted RNA was DNaseI treated to eliminate genomic DNA contaminantion prior to downstream applications. A typical DNaseI reaction contained 1 unit of recombinant DNaseI (Roche) per 1 µg of RNA, 2 µl of 10 x DNaseI buffer, 0.5 µl of RNase Inhibitor (Roche) and DEPC-treated water (Ambion) to make up final reactions to 20 µl. Reactions were incubated at 37°C for 30 min. Subsequently, 1 µl of 50 mM EDTA, pH 8.0 was added and reactions were incubated at 75°C for 10 min to inactivate DNase I enzyme.

2.3.3 Reverse transcription

RNA was reverse transcribed to cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). In brief, 3 µl of 10 x RT buffer, 1.2 µl of 25 x dNTP mix, 3 µl of 10 x random primers, 1 µl of RNase inhibitor and 1.5 µl of Multiscript reverse transcriptase were added to DNase I-treated RNA. Mixtures were brought up to final volume of 30 µl with DEPC-treated water and transferred to a thermocycler for incubation at 25°C for 10 min at 37°C for 2 hr, and 85°C for 5 min. Relative quantification of cDNA products was performed using RT-qPCR as described below.

2.3.4 Quantitative real-time PCR

Quantitative real-time PCR (qPCR) reactions were performed using Taqman primers and probes with the Taqman universal PCR mastermix (Applied Biosystems), unless indicated otherwise. A typical 10 µl reaction included 3.5 µl water, 5 µl 2 x Universal Taqman PCR mastermix, 0.5 µl primer and probe mix, and 1 µl of either template cDNA or genomic DNA. The amount of template was titrated to yield amplification products that fell within the

dynamic range of each reaction. All qPCR reactions were performed in triplicate with an ABI Prism 7500 sequence detection thermocycler (Applied Biosystems). To exclude the possibility of genomic DNA contamination, both no template and no RT (non-reverse transcribed RNA) were included in all qPCR reactions amplifying template cDNA. TaqMan primers and probes employed in this thesis are listed below (Table 2.1 and 2.2).

Gene name	Catalog number
HBA	Hs00361191_g1
HBB	Hs00747223_g1
HBZ	Hs00923579_m1
HBE	Hs00362216_m1
HBG2	Hs00361131_g1
RPL13A	Hs04194366_g1
PRS18	Hs01375212_g1
PABPC1	Hs00743792_s1

Table 2.1 Human TaqMan primers and probes employed in this thesis

Gene name	Catalog number
Hba-a	Mm02580841_g1
Hbb-b	Mm01611268_g1
Hba-x	Mm00439255_m1
Hbb-y	Mm00433936_g1
Hbb-bh1	Mm00433932_g1
Pabpc1	Mm00849569_s1
Rps18	Mm02601777_g1

Table 2.2 Mouse TaqMan primers and probes employed in this thesis

2.4 The nCounter® expression assay

2.4.1 Principles of the assay

The nCounter® expression assay (NanoString® technology) is based on digital detection and direct molecular barcoding of target molecules through the use of a color-coded probe pair. The probe pair consists of a reporter probe, which carries the signal on its 5′ end and a capture probe, which carries a biotin on the 3′ end. The color codes carry six positions and each position can be one of four colors, thus allowing for a large diversity of tags that can be mixed together in a single well for direct hybridization to target and yet still be individually resolved and identified during data collection.

2.4.2 Material preparation

Total RNA was quantified using a Qubit® Fluorometer (Qubit® RNA BR Assay Kit), DNase-treated and purified using the RiboMinus Concentration Module Kit (Invitrogen) according to the manufacturer's protocol. RNA quality was assessed by RNA integrity score (RIN) using a Bioanalyzer (Agilent Technologies). Only total RNA samples with a RIN score > 8 were selected for the nCounter® expression assay.

2.4.3 Capture probe design

Capture probes included in the assay were designed to target all adult and embryonic globin genes, erythroid specific genes, genes for use as expression controls and genes lying up to 2 Mb from either 5′ or 3′ end of the α-globin locus (termed surrounding genes). The target sequences and descriptions are shown in Table 2.3.

Gene symbol	Target sequence (shown 5′- 3′)	Target description
Alas2	GCCCAACACTTCTCCGAGGCATCTATGGCATCAAAGGATGT TTCTGTTTGGTGTAGTAATGACTATTTGGGCATAAGCAGACA CCCTCGTGTCTTGCAGG	Erythroid specific gene
Bod1	ACCAGAACCTAAGGCAGAAGGTGGATAATTTCGTGTCAACA CATCTGGACAAACAGGAATGGAATCCTGCAATGAACAAGAA CCAGTTGCGAAATGGGCT	Surrounding gene
Btf3	TGGTGCAGACAGCCTGACTAGTTTAAGGAGACTGGCTGAA GCTCTGCCCAAACAATCTGTGGATGGAAAAGCACCCCTTGC TACTGGAGAGGATGATGAT	Expression control
Car2	TGCCCAGCATGACCCTGCCCTACAGCCTCTGCTCATATCTT ATGATAAAGCTGCGTCCAAGAGCATTGTCAACAACGGCCAC TCCTTTAACGTTGAGTTT	Expression control
Chac2	GTGTATGGGGTGTTGCTTACAAACTACCAGTAGGAAAAGAA GAGGAAGTAAAAACATACCTTGACTTCAGAGAGAAAGGAGG CTACAGAACTACGACAGT	Surrounding gene
Cpeb4	GCATCGTCCCTTACTGGTTTCAGTAACTGGTCAGCAGCGAT AGCACCTTCTTCCTCCACTATAATCAATGAAGATGCAAGTTT CTTTCACCAGGGAGGGG	Surrounding gene
Eef2	TGGACCAGTTCCTTGTGAAGACGGGGACCATCACTACCTTT GAGCACGCTCACAACATGCGCGTGATGAAGTTCAGCGTCA GCCCTGTCGTCCGTGTGGC	Expression control
Fth1	TGGGAGAGCGGGCTGAATGCAATGGAGTGTGCACTGCACT TGGAAAAGAGTGTGAATCAGTCACTACTGGAACTGCACAAA CTGGCTACTGACAAGAATG	Erythroid specific gene
Gypa	GGCTTCAACTGTAGGTAACCCAAATCAGCATTCAGCAACAA TGTCAACACCAGCTATTCATGTCTCAACTTATCACACGGCC CCTACTGAAGTGTCTGCT	Erythroid specific gene
Hba-a1	GAAGAAACCATGGTGCTCTCTGGGGAAGACAAAAGCAACAT CAAGGCTGCCTGGGGGAAGATTGGTGGCCATGGTGCTGAA TATGGAGCTGAAGCCCTGG	Globin gene
Hba-x	CTGGTCACAATGGCCGCACGCTTTCCCGCCGACTTCACCC CTGAGGTCCACGAAGCCTGGGACAAGTTCATGTCTATCCTG TCTTCTATCCTGACTGAGA	Globin gene
Hbb-bh1	CACAAGTACCATTAAACTCCATTCCAGTACACTGGCAATCC CATGTGTCTATGATGCCCTCTTTGAGTCCATGGGGACTGCA TTGAGAGCACAAATTTTG	Globin gene
Hbb-bh2	GGTGTTGAGTCCCTGGAAAGAATACTGACGGTTTATCCACA CACCAAGAGATACTTTGACCACTTTGGAGACTTCTCCTTCTG TGCTGCCACTGAAGATA	Globin gene
Hbb-b1	ACCTATCCTCTGCCTCTGCTATCATGGGTAATGCCAAAGTG AAGGCCCATGGCAAGAAAGTGATAACTGCCTTTAACGATGG CCTGAATCACTTGGACAG	Globin gene
Hbb-y	GTACCACTGAGCCCTCTCTCTAGCTGTCCAGCAATCCTGTG TGTCCGCTATGCCTCCTTTCTGCACATGAATACTGGACTGTT CCTTGAAAGCACATCAT	Globin gene
Hbq	TCACGTGACTTCGGCACTGGTCTCCAAATATCGCTGAATTG GGGATACAGGGTTCTATAGTCTCCACCCATGCCCTTCCTAA AGCTCACAATGTTTGAGT	Globin gene
Il9r	TGACTGAAATCAAACACAAATGCACCTTCTGGGACAGTATG TGTACCCTGGTGCTGCCTAAAGAGGAGGTGTTCTTACCTTT TGACAACTTCACCATCAC	Surrounding gene
Mpg	AAAGAGAGACTCCTGTCGACCCCGGGCCTCCGGCGGAGTA TCTACTTCTCCAGCCCAGAGGACCATTCTGGCCGGCTAGGA CCAGAGTTTTTTGACCAGC	Surrounding gene
Npm1	CTGTTCTGTGGAACAGGAGGCAGTTGTTTTCCGTCCGGCTT CTCCCACACCGAAGTGCGCGCCTCCACCTCATGGAAGACT CGATGGATATGGACATGAG	Surrounding gene
Nprl3	TCCACTATGCAGCTTCAAGTCTGATTCCTCCTGAGGCTATT GAACGGAGCCTGAAAGCCATCCGCCCGTACCATGCCTTGC TACTTCTCAGTGACGAGAA	Surrounding gene
Pabpc1	TGATGAAAATGGCTCCAAGGGCTATGGATTTGTACACTTTG AAACACAGGAAGCAGCTGAAAGAGCTATTGAAAAAATGAAT GGGATGCTTCTAAATGAT	Expression control
Paip2	AAGTCGCAGCAGTACTAGCCCAAGCATCATCAATGACGATG TGATTATTAACGGTCATTCTCATGAAGAGGATAATCCATTTG CAGAGTACATGTGGATG	Expression control
Rhbdf1	GGGTGCTTCCTGGACACTGACCTCTTGTGCCTTGCTCGCTC CGGTTGTACGCTTACACTGCTGGGCATTTAGTACATGAGTT CCCATGATAACCTCCTAA	Surrounding gene
Rpl38	GAGGAGATCAAGGACTTTCTGCTGACAGCCCGGCGGAAGG ATGCCAAGTCTGTCAAGATCAAGAAGAACAAGGATAATGTG AAGTTCAAGGTTCGCTGCA	Expression control

Gene symbol	Target sequence (shown 5′- 3′)	Target description
Rps18	CAGCACACCAAGACCACTGGCCGCAGGGGCCGAACTGTGG GTGTATCCAAGAAGAAATGAGTCTCTGGGCCTTTGCTGTTA ATAAATAGTTTATATACCT	Expression control
Sh3pxd2b	ATGGTCCTGGAGCAGTATGTGGTAGTGGCAGACTATCAGAA GCAGGAGAGCTCAGAGATCAGCCTCAGTGTGGGCCAAGTA GTGGACATCATTGAGAAGA	Surrounding gene
Shmt2	CCGACTTGTTTTTGAGACGGGGGTCTCACTTAGCTGTGAGT GGCCTTGAACTGACCCTTCCACCTTCATTTCCCAAGTGCTG GGGTTATAGGTTTGCACC	Expression control
Slc4a1	GGCAGCGCACTGGATAGGGCTGGAGGAAAACCTTCGAGAG GATGGTGTATGGGGTCGCCCACATCTATCTTACCTGACCTT CTGGAGCCTTCTAGAACTG	Erythroid specific gene
Snrnp25	CGGAGGACGACGAGGATGATGAGGAGACATTGCCCCATTC CGAGGCTGTGGACGTGTTCCAAGAAGGTCTCGCCATGGTG GTGCAAGACCCGCTGCTCTG	Surrounding gene
Stk10	TGTCGTCGCGCTCCTCATCCAGCATGGCTTTCGCCAATTTC CGCCGCATCTTGCGGTTATCCACCTTCGAGAAGAGAAAGTC CCGTGAATATGAGCACGT	Surrounding gene
Ubtd2	CGTAACCAGCCTTTGAAAAAGGAGAAACCAAAATGGAAAAG TGACTATCCCATGACAGATGGACAACTGCGCAGCAAGAGG GATGAATTCTGGGACACTG	Surrounding gene

Table 2.3 Capture probes included in nCounter® expression assay

2.4.4 Hybridization reaction

Hybridization reactions were prepared according to the manufacturer's protocol. Briefly, 100 ng of sample RNA was transferred to a tube containing 10 µl of Reporter CodeSet and 10 µl of hybridization buffer. Subsequently, 5 µl of Capture ProbeSet was added to and gently mixed with each reaction before immediately placing at 65°C in a pre-heated thermocycler for 12 hr.

2.4.5 Post-hybridization processing

Immediately after the hybridization step, samples were transferred to the nCounter® Prep Station (Nanostring® Technologies) for post-hybridization processing and subsequently to the nCounter® Digital Analyzer (Nanostring® Technologies) for data collection.

2.4.6 Data analysis

Data analysis including quality control of assays, normalization and target gene expression was performed using Expression Console and Transcriptome Analysis Software (Nanostring® Technologies).

2.5 Isolation of primitive erythroid cells from mouse embryos

2.5.1 Mouse strain

All mice, except those harboring deletions of the α-globin hypersensitive sites (HS), were of C57BL6/J background. Mice harboring the series of deletions of the alpha upstream HS were from a mixed genetic background including 129SvEv, C57BL6/J, CBA and BalbC.

2.5.2 Isolation of primitive erythroid cells

Mice were heparinized and sacrificed by cervical dislocation at 10.5 days observation of a vaginal plug. Embryos and yolk sacs were dissected from the uterus into sterile tubes, blood was washed three times with 1 ml sterile PBS and transferred into the same tubes. Washes were centrifuged at 10,000 rpm for 2 min and pellets were resuspended with appropriate solutions and volume for further applications.

2.5.3 DNA isolation from mouse embryos for genotyping

The whole embryo was used for genotyping whenever needed. Digestion buffer was prepared from 2.5 ml of 1 M Tris pH8.0, 0.1 ml of 0.5 M EDTA pH 8.0, 250 µl of Tween-20 and water to a total volume of 50 ml. To each embryo 500 µl of digestion buffer and 10 µl of Proteinase K (20 mg/ml) (Sigma) were added before incubation at 42°C overnight. Tubes were transferred to a 95°C heat block for 10 min for protein denaturation. Supernatant containing DNA was aspirated into a new tube and stored at -20°C.

2.5.4 Genotyping of mouse embryos

Genotyping was performed for embryos obtained from heterozygote crosses of mice harboring double-deletions of HS-31 and HS-26. A typical 20 µl reaction consisted of 1×FastStart PCR reaction buffer (Roche), 1.5 mM $MgCl_2$, forward and reverse primers at 0.5 µM, and dNTPs at 200 µM (final concentrations), with 0.75 U of FastStart Taq DNA polymerase (Roche). Reactions were initially incubated in a thermocycler at 94°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 63°C for 30 sec, and extension at 72°C for 1 min. Final extension was performed at 72°C for 10 min. Positive and negative controls were included in each reaction. Primer sequences for genotyping are listed below (Table 2.4).

Primer name	Primer sequence (shown 5′- 3′)
26 del positive_F	CCT AGC CTA CCT ATA TAC ATG CTG GCT GAC
26 del positive_R	TGC TCT CTG TGG TGT TCT CTG TTT CCA CTA G
26 WT_F	GGA CAA ATG GTA CCA CTG ATT AGG ACC TCT GA
26 WT_R	CCC TTT AGT GTA TGT CAT TGA GCC TTA TGG TGT G
31 del positive_F	GGG AAA AAG CAA GAG AGG AAA GGG TTA CAC TG
31 del positive_R	CTG TTA ACT GCT CAA GAG CAA GAA CAA GCC TGA
31 WT_F	TGG AGA CAC TGA ACA CTG CCA TTA GGA GTC
31 WT_R	CCC TTT TTC TTC CCT TCT CTC CAC AGA CGT

Table 2.4 Primer sequences for genotyping of mouse embryos obtained from litters of heterozygous crosses of double- knockout HS-31/HS-26

2.6 Primitive erythroid culture

2.6.1 Primitive erythroid culture conditions

The protocol for primitive erythroid culture was adapted from the primitive erythroid 2-step maturation protocol published in Greenfest-Allen *et al*, 2013 (Greenfest-Allen et al., 2013). A litter of E8.5 murine embryos and yolk sacs was dissected and dissociated with 200 µl of 0.25% Trypsin EDTA, phenol red (Gibco) to a single cell suspension by gentle pipetting every 1-2 min for 5 min. Trypsin was subsequently inactivated with 200 µl of bovine plasma-derived serum (PDS, First Link (UK), Ltd.) and pellets obtained by centrifugation at 1000 rpm for 5 min at room temperature (RT). Cells were resuspended in 2 ml of complete culture media and seeded to 1 well of a 12-well plate coated with 0.1% gelatin (~7 x 10^5 - 1x 10^6 cells/well) before incubation at 37°C with 5% CO_2 for 2-3 days. Cells were counted daily and media was added to maintain a cell concentration of less than 1 x 10^6 cells/well.

Complete primitive erythroid culture media contained Iscove's Modified Dulbecco's Medium (IMDM, Life Technologies) supplemented with 10% knock-out serum replacement (Invitrogen), 10% PFHM-II (Invitrogen), 1% PDS (First Link (UK), Ltd.), 2 mM Glutamax (Invitrogen), 150 µM Thioglycerol (MTG, Sigma) and 1 unit/ml recombinant human erythropoietin (EPO, Janssen). Complete media was filter sterilized with a 0.45 µM filter prior to use.

2.6.2 Harvest of Primitive erythroid cells

After 48-72 hr of culture, primitive erythroid cells were harvested. Cell suspension was aspirated and transferred to a fresh tube. Wells were washed with PBS twice and washes were combined with the tube containing the cell suspension before being centrifuged at 1000 rpm at room temperature for 5 min. Supernatant was aspirated and pellets resuspended with an appropriate volume of PBS for cell counts and an aliquot taken for morphological analysis. Remaining cell suspensions were centrifuged at 1000 rpm at room temperature for 5 min and 1 ml of TRIzol (Ambion) per 10^7 cells was transferred to the tube for cell resuspension and homogenization. TRIzol suspensions were immediately stored at -80°C for subsequent RNA extraction.

2.7 Fetal liver culture

2.7.1 Fetal liver cell expansion

Mouse erythroblasts were expanded and differentiated according to a protocol adapted from Marieke Von Lindern (Drissen et al., 2005). For each replicate, one E12.5 mouse fetal liver was dissected, dissociated into a single cell suspension and resuspended in 200 µl StemPro medium (Invitrogen). Cell suspensions were seeded in to one well of 12-well plate containing StemPro medium supplemented with 1 unit/ml of EPO (Janssen), 50 ng/ml of Stem cell factor (SCF, Peprotech), 1 µM of dexamethasone and 1x L-Glutamine. 24 hours after dissection fetal liver cells in suspension were transferred to a new plate containing the same media at a cell concentration of ~1 million cells/ml. Cells were counted daily and maintained at a concentration of ~ 1 million

cells/ml for up to 5-7 days. The numbers of early erythroblasts obtained at the end of expansion phase range from 5 x 10^7 to 10^8 cells from a single mouse fetal liver.

2.7.2 Ter119 depletion

At the end of the cell expansion phase, mature red cells were removed using Ter119 depletion: FACS buffer (PBS + 10% FCS) and PBE (PBS + 0.5% BSA, Sigma + 2 mM EDTA, pH 8.0, Ambion) were freshly prepared. All centrifugations were performed at 1000 rpm for 5 min at room temperature. Cells obtained from the expansion phase were centrifuged and resuspended in 4 ml FACS buffer before being divided into 1ml aliquots. Anti-Ter119-PE antibody (BD Bioscience, cat no. 553673) was added at a volume of 10 µl per tube and incubated with the cells at 4°C for 20 min in the dark. All cell suspensions were washed twice with FACS buffer. Supernatant was removed and cells were resuspended with 1 ml PBE and transferred to fresh tubes. 100 µl of anti-PE microbeads (Miltenyl Biotec, cat no. 130-048-801) was added to each tube and incubated at 4°C in the dark. After 15 min, cell suspensions were transferred to a fresh tube and washed twice with PBE. Supernatant was removed and cells were resuspended with 1 ml PBE before being passed through a 30 µM filter (Miltenyl Biotec, cat no. 130-041-407), to PBE-equilibrated LS columns (Miltenyl Biotec, cat no. 130-042-401) mounted on a magnet. Magnetic columns were washed 3 times with 3 ml of PBE and flow through containing Ter119-negative cells was collected for the next step. Magnetic columns were removed from the magnetic stand and 5 ml PBE was pushed through the column using a syringe to elute Ter119-positive cells for purity analysis by flow cytometry.

2.7.3 CD 44-hi cell sorting

Previous work demonstrated that the adhesion molecule termed CD44 exhibited a progressive and marked decrease from proerythroblast to reticulocyte and can therefore be used to distinguish each stage of erythroid maturation (Chen et al., 2009). To maximize synchronization of the differentiation phase of the fetal liver-derived definitive erythroid cell culture, CD44-hi erythoid progenitors, which are most likely proerythroblasts, were used as starting material. Ter119-negative cells were counted, centrifuged and resuspended at concentration of 2 x 10^7 cells/ml of FACS buffer. To each 1 ml aliquot of cell suspension, 10 µl of CD44-APC antibody (BD Bioscience, cat no. 561862) was added and reactions were incubated at 4°C in the dark. After 20 min, cell suspensions from each tube were combined in a fresh tube and washed twice with FACS buffer. Supernatant was removed and cells were resuspended in 2 ml of FACS buffer, to which 20 µl of 1:100 dilution of Hoechst (Invitrogen) was added for live-cell gating. Cell sorting was performed by the WIMM FACS facility. CD44-hi cells were sorted using a FACSAria Fusion cell sorter (BD Biosciences) into collection tubes containing fetal liver cell expansion medium. Pre- and post-sorting, cell samples were handled on ice at all times.

2.7.4 Erythoid maturation

Erythroid differentiation medium was freshly prepared from StemPro medium (Invitrogen) supplemented with 5 units/ml of EPO (Janssen), 0.5 mg/ml of transferrin (Sigma), and 1 x L-Glutamine. Immediately after CD44-hi cell sorting, cells were centrifuged at 1000 rpm for 5 min and resuspended with pre-warmed erythroid differentiation media at a concentration of 1 x 10^6

cells/ml. Cells were transferred to either T25 or T75 according to cell suspension volume and were incubated at 37°C with 5% CO_2.

2.7.5 Erythroid cell harvest

After 30 hours of culture in erythroid differentiation medium, maturing erythroblasts were harvested. Cells were centrifuged at 1000 rpm for 5 min, washed with PBS, resuspended and homogenized with 1 ml of TRIzol per 1 x 10^7 cells before storage at -80°C for subsequent RNA extraction.

2.7.6 Flow cytometric analysis of erythroid cells

Cell surface antigen analysis of pre- and post- CD44hi-sorted erythroid cells was performed using Cyan Flow Cytometer (Beckman Coulter). Cells (1 X 10^5 cells/ml in 2% BSA (Sigma) in PBS) were stained with 10 µl of mouse anti-Ter119-PE (BD Bioscience, cat no. 553673), 10 µl of anti-CD44-APC (BD Bioscience, cat no. 561862) and 10 µl of anti-CD71-FITC (BD Bioscience, cat no. 553266) antibodies per 1 ml of cell suspension. Excess antibodies were washed twice and stained cells were resuspended in 500 µl of PBS. Live cells were differentiated from dead cells by adding 5 µl of 1:100 dilution of Hoechst (Invitrogen) to each cell suspension prior to flow cytometric analysis.

2.8 iPSCs maintenance and erythroid differentiation

2.8.1 Human iPSCs maintenance

Wild type human iPS cell lines C19 (a kind gift from Lee Carpenter) and patient-derived iPSCs (“BH” line; reprogrammed by Lee Carpenter) were cultured at 37°C, with 5% CO_2 on Matrigel (BD Biosciences)-coated plates. Complete medium was made up from mTeSR1 Basal medium and mTeSR1

supplement (Stem Cell Technologies). Cells were passaged when confluence reaches 70-80% either mechanically by colony picking, or chemically using 1 ml of EDTA (Lonza) per well of 6-well plates. In cases where EDTA was used for passaging, iPSCs were incubated with EDTA in 5% CO2, 37°C for 4 min. EDTA was aspirated and replaced with 1 ml of fresh complete mTeSR1 media (StemCell Technologies) per well. iPSC colonies disrupted by EDTA were gently scraped off the well with a cell scraper (BD Falcon) and split 1:6 to 1:8 into new a Matrigel-coated plate. Complete mTeSR1 media was added to make a total of 3 ml per well of a 6-well plate. Freezing medium (2X) was prepared from a mixture of 20% DMSO, 20% Knockout DMEM (Gibco), 60% ES-cell tested FBS (Gibco). 70-80% confluent iPSCs in each well of a 6-well plate were divided and transferred into 2 cryovials for long-term storage in vapor phase liquid nitrogen. Each iPSC-containing cryovial was seeded into one well of a 6-well plate when required. iPSCs used for erythroid differentiation were between passages 15 and 20. 10 μM of ROCK inhibitor (Sigma) was added when cells were re-seeded from frozen.

2.8.2 Feeder-based erythroid differentiation from iPSCs

Erythroid differentiation was performed based on methods published in Chang CJ, *et al.*(Chang and Bouhassira, 2012). Undifferentiated iPSCs (1X 6-well plate per experiment) were dissociated from mTeSR1-coated plate with freshly prepared 1mg/ml of collagenase type IV (Invitrogen) and transferred onto irradiated FH-B-hTERT (a kind gift from Dr. E Bouhassira) feeder layers for co-culture during the first 14 days. During this phase, co-cultured cells were maintained in RPMI 1640 (Invitrogen) supplemented with 15% FBS (Invitrogen), 2 mM L-glutamine, 1% MEM-nonessential amino acids, 50 μg/ml

ascorbic acid, 100 μM monothioglycerol (MTG), 100 units/ml penicillin and 100 μg/ml streptomycin. Media changes were performed every 2-3 days. At the end of the co-culture phase, differentiated iPSCs on FH-B-hTERT were dissociated using collagenase IV followed by treatment with trypsin/EDTA (Invitrogen) supplemented with 5% chick serum (Invitrogen). CD34+ve cells were selected using a CD34 microbead kit (Miltenyi Biotec) according to the manufacturer's protocol. CD34+ve cells obtained were subsequently seeded in liquid culture as described below.

The liquid culture consists of 4 steps:

Step 1: amplification of progenitors

CD34+ve cells obtained were cultured for 7 days at a concentration of 50,000 cells per 1ml of StemSpan medium (StemCell Technologies) supplemented with cytokine cocktails, including 10^{-6} M of hydrocortisone, 13 ng/ml of BMP4, 13 ng/ml of IL3, 33 ng/ml of Flt3L, 100 ng/ml of SCF, and 2.7 U/ml of erythropoietin (EPO).

Step 2: differentiation

After day 7 of liquid culture, cells were grown and differentiated in StemSpan medium supplemented with 10^{-6} M of hydrocortisone, 13 ng/ml of BMP4, 13 ng/ml of IL3, 40 ng/ml of IGF-1, 40 ng/ml of SCF and 3.3 U/ml of EPO for 7 days. Cell density was maintained below 1 million cells/ ml at all times by adding fresh medium every 2-3 days.

Step 3: maturation

Differentiating erythroid cells were then transferred into the maturation step, in which they are co-cultured with MS-5 stromal cells (a kind gift from Dr. L Carpenter) in StemSpan containing 3.3 U/ml of EPO for 3 days.

Step 4: final maturation

In this final step of erythroid maturation, medium was switched to StemSpan only for 7 days. Medium was added every 2-3 days as required.

Cultured cells were incubated at 37°C in 5% CO2 throughout the protocol. Samples were collected at days 7, 10, 14, 17, 21 and 24 for morphological analysis using Modified Wright's stain, RNA extraction and cell surface marker analysis, including CD34, CD71 and Glycophorin A (GPA), using flow cytometry.

2.8.3 Flow cytometric analysis of erythroid cells

Cell surface antigen analysis of human iPSC-derived differentiating erythroblasts was performed using a Cyan Flow Cytometer (Beckman Coulter). Cells (~2-5 X 10^4 cells) were washed with PBS and resuspended in 200 µl of 2% BSA (Sigma) before being stained with 4 µl of human anti-CD34 (Miltenyi Biotech, cat no. 130-090-954), 4 µl of anti-CD36 (BD Biosciences, cat no. 555454), 4 µl of anti-CD71 (BD Bioscience, 555536) and 4 µl of anti-GPA (BD Biosciences, cat no. 340947) antibodies. Cell suspension and antibodies were incubated on ice, in the dark, for 20 min. Excess antibodies were washed twice with PBS and stained cells were resuspended in 500 µl of PBS for flow cytometric analysis. All centrifugation was performed at 1,000 rpm for 5 min, at room temperature.

2.9 DNaseI hypersensitivity assay

2.9.1 Sample preparation

Primitive erythroid cells (~ 3 X 10^7 cells), obtained from 4 litters of E10.5 yolk sacs, were pooled and centrifuged at 1000 g. Cell pellets were resuspended

in 10 ml ice-cold PBS, centrifuged at 1000 g, at 4°C for 5 min before removal of supernatant.

2.9.2 Cell lysis reactions

To lyse the cell membrane while leaving the nuclear envelope intact, pellets were resuspended in 600 µl of a pre-mixed lysis buffer containing 972 µl high salt lysis buffer (50 mM KCl,10 mM $MgSO_4$, 3mM DTT, 5 mM Hepes, 0.05% NP-40), 20 µl of 50X Protease Inhibitor (Roche), 5 µl of 200 mM PMSF and 3 µl of 1 M DTT per ml. Reactions were incubated at room temperature for 1 min and transferred to a 1.5 ml tube during this time. Tubes were placed on ice to stop the lysis reaction and centrifuged at 1000 g at 4°C for 5 min. At this stage, red supernatant was discarded leaving a white pellet containing only nuclei. Pellets were resuspended in 1 ml cold RSB (10 mM NaCl, 10 mM Tris pH 7.4 and 3 mM $MgCl_2.6H_2O$) and transferred to a 2 ml tube containing 1 ml RSB. Suspensions were centrifuged at 1000 g, at 4°C for 5 min and washed with 3 times with 2 ml of cold RSB.

2.9.3 DNaseI digestion

Following removal of the final wash, pellets were resuspended in 1.6 ml cold RSB and 200 µl of the resultant solution was dispensed into eight 1.5 ml tubes containing 800 µl of cold RSB. On ice, 100 µl, 80 µl, 60 µl, 40 µl, 20 µl, 10 µl and 5 µl of a 1:10 dilution of recombinant DNase I (Roche) (10 units/µl) were added to tubes. All tubes were immediately transferred to a 37°C heat block for incubation. The viscosity of DNA samples in each tube were continuously checked by pipetting of 10 µl aliquots of each DNA sample, starting from tubes with the two highest amounts of DNaseI in a 10 µl drop of

general lysis buffer (10 mM Tris, 10 mM EDTA, 10 mM NaCl, 0.5% SDS) on parafilm. All tubes were placed on ice at the point where DNA samples from tubes containing 80 µl and 100 µl of DNaseI became watery. Each reaction was transferred to a fresh tube containing 3 ml of general lysis buffer and 20 µl of Proteinase K (NEB). Reactions were incubated at 37°C overnight in a water bath.

2.9.4 RNase treatment and Phenol/Chloroform extraction

Residual RNA was removed by incubation of each sample, at 37°C for 30 minutes with 7 µl of RNase (Roche, 500 ng/µl). Digested DNA was recovered by Phenol/Chloroform extraction and the aqueous upper phase was mixed with 400 µl of 4 M NaCl and 8 ml of 100% chilled EtOH. The mixture was stored at -20°C overnight and tubes were centrifuged at 3,500 rpm, at 4°C for 30 min. Supernatant was removed and pellets were washed with 1.5 ml of 70% ethanol. After centrifugation at 13,000 rpm at 4°C for 30 min, supernatant was completely removed and pellets were air-dried for 5 min. 100 µl of nuclease-free water was added to DNA pellets and tubes were incubated at 4°C on a rotor for 24-48 hr to allow full resuspension of digested DNA.

2.9.5 DNaseI fade selection

DNA samples from four different DNase I concentrations were selected for library preparation on the basis of visualizing the degree of digestion using 1% agarose gel electrophoresis and the molecular weight was determined using a Genomic DNA Screen Tape (Agilent Technologies). Quantification of selected DNA fades was performed using Nanodrop spectrophotometer (Thermo Scientific).

2.9.6 DNaseI-Seq library preparation

DNaseI library preparation was performed using NEBNext DNA Library Prep Master Mix Set for Illumina (NEB). To perform end-repair 5 µg of fragmented DNaseI treated material per selected digestion was mixed with 10 µl of 10X NEBNext End Repair Reaction buffer, 5 µl of NEBNext End Repair enzyme mix and nuclease-free water (Sigma) to make volume to 100 µl. The reaction was incubated in a thermal cycler at 20°C for 30 min, subsequently purified using the MinElute PCR purification kit (Qiagen) and resuspended in 51 µl of elution buffer. To add polyadenylated sequences (termed dA-tailing), 25 µl of end-repaired DNA was transferred to a fresh tube and mixed with 5 µl of 10X NEBNext dA-Tailing reaction buffer, 3 µl of Klenow fragment and 17 µl of nuclease-free water. Mixtures were incubated at 37°C for 30 min, cleaned up using the MinElute PCR purification kit (Qiagen) and resuspended in 32 µl of elution buffer. Sequencing adaptors were ligated to dA-tailed DNA by incubation, at 20°C for 15 min, of a reaction containing 10 µl of dA-tailed DNA, 10 µl of 5X NEBNext quick ligation reaction buffer, 5 µl of 1:20 dilution of 15 µM NEBNext adaptor (NEB), 5 µl of Quick T4 DNA Ligase (NEB, E6047A) and 15 µl of nuclease-free water. Following adaptor ligation, the U-shape adaptors ligated in previous stage were cleaved with USER enzyme (NEB) according to the manufacturer's protocol. Reactions were re-purified using the MinElute PCR purification kit (Qiagen) and adaptor-ligated DNA was resuspended in 32 µl of nuclease-free water. To reduce amplification bias, library enrichment was performed in parallel by two rounds of PCR amplification per sample. Each reaction consisted of 10 µl of adaptor ligated DNA, 1 µl of 1:10 Universal PCR Primer (NEB, E7335L), 1 µl of 1:10 Index

Primer (NEB, E7335L), 25 µl of NEBNext High-Fidelity 2X PCR master mix and 13 µl nuclease-free water. Reactions were transferred to a thermal cycler, incubated at 98°C for 30 sec, followed by 10 cycles of 98°C 10 sec, 65°C 30 sec, 72°C 15 sec and the final extension was performed at 72°C for 5 min.

2.9.7 Size selection of library

Size selection was achieved using 2% Certified Low Range Ultra Agarose (Bio-Rad) gel electrophoresis. Gel fragments containing DNA ranging in size from ~150 bp to ~500 bp was excised under UV light and DNA purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research) according to the manufacturer's protocol. DNaseI libraries from each fade were combined at this stage.

2.9.8 Quality control and quantification of size-selected DNaseI library

To ensure appropriate size of recovered DNA, a quality check was performed using the DNA1000 Screen Tape (Agilent Technologies). Quantification of libraries was performed using a Qubit dsDNA HS Assay kit (Invitrogen) and confirmed by qPCR using the KAPA library quantification kit (KAPA Biosystems) according to the manufacturer's protocol.

2.9.9 High throughput sequencing and analysis of DNaseI-Seq

Four libraries were multiplexed at equimolar ratio and this mix was loaded at a concentration of 4 nM onto a single lane of a MiSeq sequencing maching (Illumina). The MiSeq 150 cycle V3 kit (Illumina) was used for sequencing reactions. Reads were mapped to the NCBI37/mm9 genome (this build was

used for all analysis presented in this thesis) and bioinformatics analysis was performed by Maria Suciu using custom scripts.

2.10 Assay for Transposes-Accessible Chromatin (ATAC)

This method was originally published by Buenrostro, et al. (Buenrostro et al., 2013) and is based on direct *in vitro* transposition of sequencing adaptors into native chromatin.

2.10.1 Cell Preparation

150,000 mouse primitive erythroblasts, obtained from E10.5 yolk sacs, were used for each biological replicate. Cells were washed with PBS in a 1.5 ml tube and centrifuged at 500 g for 5 min, 4°C. Supernatant was removed and cells resuspended in 500 µl PBS before being equally divided into 2x250 µl in fresh tubes, each being treated as a technical replicate (~75,000 cells each). Tubes were centrifuged at 500 g for 10 min at 4°C and placed on ice.

2.10.2 Transposition reaction and purification

Supernatant was removed and cell pellets resuspended by gentle pipetting with 50 µl of cold lysis buffer containing 10 mM Tris-HCl, pH 7.4, 10 mM NaCL, 3 mM MgCl2, 0.1% IGEPAL CA-630 (Sigma) before being centrifuged at 500 g for 10-15 min at 4°C. Tubes were placed on ice and supernatant was discarded. At this stage, the pellet color changed from red to white as only cell nuclei remained. Transposition reaction mix containing 2 µl of 2x TD Buffer (Illumina), 2.5 µl Tn5 Transposes (Illumina) and 22.5 µl nuclease-free water (Sigma) was added to nuclei pellets for resuspension. Transposition reactions

were mixed by gentle pipetting before brief centrifugation and incubation at 37°C for 30 min. Digested DNA was purified using the MinElute Kit (Qiagen) according to the manufacturer's instructions and was eluted in 20 µl elution buffer (10 mM Tris, pH 8). Purified digested DNA was stored at -20°C.

2.10.3 ATAC library preparation

To amplify transposed DNA fragments, 10 µl of fragmented DNA was combined with 10 µl nuclease-free water (Sigma), 2.5 µl of 25 µM customized Nextera PCR Primer 1 (Epicentre), 2.5 µl of 25 µM customized Nextera PCR Primer 2 (Barcode, Epicentre) and 25 µl NEBNext High-Fidelity 2 x PCR Master Mix (NEB) to make up total of 50 µl. Reactions were incubated at 72°C for 5 min, 98°C for 30 sec, then 11 cycles of 98°C for 10 sec, 63°C for 30 sec and 72°C for 1 min, before being held at 4°C. Amplified ATAC libraries were purified using the MinElute PCR Cleanup Kit (Qiagen) according to the manufacturer's instructions and eluted in 20 µl of elution buffer (10 mM Tris buffer, pH 8).

2.10.4 Nucleosomal patterning and library quantification

ATAC libraries were checked for an iterative pattern of nucleosomal digestion by running amplified transposed samples on a DNA 1000 Screen Tape (Agilent Technologies) according to the manufacturer's protocol. Libraries in which the characteristic nucleosomal phasing could be observed were selected and quantified by qPCR in comparison with 6 DNA standards using the KAPA library quantification kits (Kapa Biosystems).

2.10.5 High throughput sequencing and analysis of ATAC-seq

Equimolar multiplexed mixes of three ATAC libraries were loaded at a concentration of 4 nM onto one lane of a MiSeq device (Illumina) for sequencing. The MiSeq 150 cycle V3 kit was used for sequencing reactions. Reads were mapped and bioinformatics analysis was performed by Maria Suciu using custom scripts.

2.11 Capture-C

The Capture-C technique was performed to investigate the landscape of chromosomal interactions from the viewpoints of the *Hba-x* and *Hba-a1*, *Hba-a2* promoters. The method was originally published by Hughes J, *et al* (Hughes et al., 2014) and subsequently modified by Davies J, *et al* (Davies et al., 2016).

2.11.1 Formaldehyde fixation

An aliquot of 7 x 10^6 E10.5 yolk sac-derived primitive erythroid cells and ES cells, was collected with 7 ml of PBS. Protein-DNA cross-linking was performed by adding 378 µl of 37% formaldehyde to the cell suspension (final concentration 2%) and incubation of the mixture on a roller, at room temperature for 10 min. Fixed cells were quenched by adding 1.05 ml of 1 M cold glycine (Final concentration 125 mM) and were centrifuged at 1000 rpm at 4°C for 5 min. Pellets were washed in 7 ml of cold PBS and centrifuged at 1000 rpm at 4°C for 5 min before resuspension in 3.5 ml of cold lysis buffer containing 10 mM Tris, pH 8, 10 mM NaCl, 0.2% NP-40 (Igepal) and 1 tablet of Complete protease inhibitor cocktail (Roche, Cat no. 11873580001) per 50 ml of lysis buffer. Lysis reactions were incubated on ice for 20 min and

centrifuged at 1800 rpm at 4°C for 5 min in a bench top centrifuge and supernatants were discarded. Pellets were resuspended in 1 ml of lysis buffer and transferred to fresh 1.5 ml tubes prior to snap freezing on dry ice and storage at -80°C.

2.11.2 Restriction enzyme digestion

Aliquots of formaldehyde fixed cells, each containing 7 x 10^6 cells, were defrosted on ice, centrifuged at 14,000 rpm at room temperature for 5 min and supernatants were carefully removed. Pellets were resuspended in 1 ml of nuclease-free water (Sigma) using wide bore tips. Homogenization of pellets was performed on ice using a Dounce homogenizer for 2 x 25 strokes with the pestle and homogenized solution was transferred to a fresh 1.5 ml tube. The homogenizer was rinsed with 400 µl of nuclease-free water and 10 more strokes performed before combining the wash with the existing material. Homogenized material was centrifuged at 14,000 rpm in a microfuge at RT for 5 min. Supernatants were removed and pellets resuspended in 450 µl of nuclease-free water. From this volume 2 digestion reactions and 1 undigested control reaction were performed. Each digestion reaction comprised 10 µl 20% SDS, 80 µl 10x restriction enzyme buffer (NEB), 414 µl nuclease-free water and 200 µl of suspended homogenized material (chromatin). Control reactions containing no enzyme comprised 2.5 µl 20% SDS, 20 µl 10X restriction enzyme buffer, 111 µl nuclease-free water and 50 µl of suspended homogenized chromatin. Both digestion reactions and undigested control reactions were shaken at 1400 rpm at 37°C for 1 hr using an Eppendorf Thermomixer before addition of 66 µl and 16 µl of 20% Triton X 100 to digested and undigested reactions respectively. From this point, all reactions

were shaken continuously at 1400 rpm at 37°C until the next step. 10 µl of high concentration (50,000 U/ml) *Dpn*II enzyme (NEB) was added to each digestion reaction after 1 hour of shaking, again at the end of the day and a third time on the following morning followed by a minimum of 3 hours of further digestion. Good mixing of enzyme and digestion reactions was ensured by gently pipetting and inverting the tubes.

2.11.3 Ligation of digested chromatin and de-crosslinking of controls

Aliquots of 100 µl from each digestion reaction were made and 3 µl of proteinase K (NEB) was added, Proteinase K was also added to undigested reactions, these tubes were transferred to a 65°C heat block for overnight incubation and stored at -80°C awaiting for further treatment and later analysis. To prepare the remaining DNA for ligation, digestion reactions were incubated at 65°C for 20 min to inactivate the *Dpn*II enzyme. On ice, 133 µl of 10x Ligation buffer (Thermo Scientific), 500 µl of nuclease-free water and 8 µl of T4 DNA Ligase (30 unit/µl, Thermo Scientific) was added to digestion reactions. Ligation reactions were transferred to a pre-cooled Thermomixer for overnight incubation at 16°C with shaking at 1400 rpm.

2.11.4 De-crosslinking of ligation reactions

5 µl of Proteinase K (NEB) was added to ligation reactions followed by overnight incubation at 65°C.

2.11.5 DNA extraction of ligation and control materials

Ligation reactions were pooled and all samples were RNase-treated by addition of 30 µl, 4 µl and 6 µl of RNase (Roche, 500ng/µl) to pooled ligation

reactions, undigested controls and digested controls respectively before incubation for 30-min at 37°C. DNA was extracted using phenol/chloroform and chloroform alone before being recovered by mixing with 25 ml of 100% ethanol, 5 ml of nuclease-free water and 800 µl of 3 M NaOAc. DNA was recovered from undigested controls by mixing with 600 µl of 100% ethanol, 24 µl of 3 M NaOAc and 1 µl of Glycogen while DNA in digested controls was recovered by mixing with 900 µl of 100% ethanol, 36 µl of 3 M NaOAc and 1 µl of Glycogen. Ligation reactions and controls were incubated at -80°C for at least 2 hr prior to centrifugation at 4°C at 3,200 rpm for 45 min for ligation reactions and at 13,000 rpm for 10 min for controls. Pellets were washed in 70% ethanol and air-dried at room temperature. Pellets containing ligated DNA were resuspended in 300 µl of nuclease-free water and undigested control pellets and digested control pellets were resuspended in 50 µl and 25 µl of nuclease-free water respectively.

2.11.6 Assessment of digestion and ligation

Quality control was performed to ensure high enzymatic digestion efficiency and effective ligation of chromatin. This was assessed using 1% agarose gel electrophoresis. Digestion efficiency was also determined by qPCR comparison of amplicons across a *Dpn*II cut site to amplicons within a *Dpn*II cut site on the α-globin locus (Davies et al., 2016).

2.11.7 Sonication

Ligated DNA was quantified using a Qubit fluorometer (Invitrogen). A total volume of 120 µl containing 5-6 µg of ligated DNA was transferred to a Covaris microtube. DNA was sheared using a Covaris S220 focused

ultrasonicator using frequency sweeping mode, a duty cycle of 10%, intensity of 5, 200 cycles per burst and 6 cycles of 60 seconds each. This setting was selected to shear the DNA to a size of 200 bp. Sonicated DNA was purified using Ampure XP SPRI beads (Beckman Coulter) according to the manufacturer's protocol. In brief, DNA was captured using a 180 µl of beads. Beads were washed twice with 500 µl of fresh 80% ethanol on the magnetic stand and air-dried at room temperature. Sonicated DNA was eluted in 30 µl of nuclease-free water (Sigma) and assessed with the DNA1000 Screen Tape (Agilent Technologies).

2.11.8 Capture-C library preparation

Library preparation was performed using the NEBNext DNA library prep reagent set for Illumina using NEBNext Multiplex oligos (NEB) in accordance with the manufacturer's instructions. 27 µl of sonicated DNA was mixed with 10 µl of 10x NEBNext End Repair Reaction Buffer and 5 µl of NEBNext End Repair Enzyme Mix for an end repair reaction. Sterile nuclease-free water was added to the mixture to make up a total volume of 100 µl before incubation in a thermal cycler for 30 min at 20°C. The reaction was cleaned up using 1.8 volumes of Ampure XP beads (according to the manufacturer's instructions) and resuspended in 43 µl of nuclease-free water. Subsequently, 42 µl of end-repaired DNA fragments were dA-tailed by adding 5 µl of 10x NEBNext dA-Tailing Reaction Buffer and 3 µl of Klenow Fragment. The 50 µl dA-tailing reaction was incubated at 37°C for 30 min in a thermal cycler. The reaction was cleaned up with 1.8x volume of Ampure XP beads and resuspended in 25 µl of nuclease-free water before being processed to the adaptor ligation step. Adaptor ligation of dA-tailed DNA was performed by

adding 25 µl of dA-tailed DNA to 10 µl of 5x Quick Ligation Reaction Buffer, 10 µl of U-Shape NEBNExt Adaptor and 5 µl of Quick T4 DNA Ligase to make up a total volume of 50 µl. Reactions were incubated at 20°C for 15 min in a thermal cycler. To cleave the U-loop of ligated adaptors, 3 µl of USER^{TM} enzyme (NEB) was added and mixed by gentle pipetting and reactions were incubated at 37°C for 15 min. DNA was recovered using 1.8 volumes of Ampure XP beads and resuspended in 53 µl of nuclease-free water. To minimize amplification bias, enrichment of adaptor-ligated DNA was performed with 2 PCR reactions run in parallel. For each PCR reaction, 20 µl of adaptor-ligated DNA library was mixed with 2.5 µl of index primer (25 µM), 2.5 µl of universal primer (NEB, E7335L) and 25 µl of NEBNext High-Fidelity 2x PCR Master Mix. PCR amplifications were carried out in a thermal cycler at 98°C, 30 sec for initial denaturation, 98°C 10 sec for denaturation, 65°C 30 sec for annealing and at 72°C 30 sec for extension. Denaturation, annealing and extension steps were repeated for a total of 8 cylces before final extension of the PCR products at 72°C for 5 min. Each PCR amplification was cleaned up separately with 1.8x volume of AMPure XP Beads and resuspended in 50 µl nuclease-free water. Two PCR amplifications of the same library were combined at this step. Quality control of amplified libraries was performed using a DNA1000 Tape Station (Agilent Technologies) and DNA concentration was measured using Qubit Fluorometer (Invitrogen).

2.11.9 Capture oligonucleotides

Sequences of biotinylated oligonucleotides used for the Capture-C experiment are given in Table 2.5. Oligonucleotides were reconstituted in water and diluted to a concentration of 2.89 µM and combined into an equimolar mix.

Oligonucleotide name	Capture oligonucleotide sequences
Hba-a1 F	5'biotinylated-GAT CCA TGG TAG CAC AGG GCA GCT AAG ATG CAA GTC TGA AGG AGG AGT CTG GCG AGC TGC TCC TGC AGT TCC CTG GAC CCA GAA GGA TGA GCT AGC AGA TTC ACT TGA GCC AAA GGA TTC-3′
Hba-a1 R	5'biotinylated-AAA CCA TGG TGC TCT CTG GGG AAG ACA AAA GCA ACA TCA AGG CTG CCT GGG GGA AGA TTG GTG GCC ATG GTG CTG AAT ATG GAG CTG AAG CCC TGG AAA GGT GAG AAC AGG ACC TTG ATC-3′
Hba-a2 F	5'biotinylated-GAT CCA TGG TAG CAC AGG GCA GCT AAG ATG CAA GTC TGA AGG AGG AGT CTG GCG AGC TGC TCC TGC AGT TCC CTG GAC CCA GAA AGA TGA GCT AGC AGA TTC ACT TGA GCC AAA GGA TTC-3′
Hba-a2 R	5'biotinylated-AAA CCA TGG TGC TCT CTG GGG AAG ACA AAA GCA ACA TCA AGG CTG CCT GGG GGA AGA TTG GTG GCC ATG GTG CTG AAT ATG GAG CTG AAG CCC TGG AAA GGT GAG AAC AGG ACC TTG ATC-3′
Hba-x F	5'biotinylated-GAT CTT CCC TCA GTG CTA AGT GAG AGG AAT TAC TGC TTC CTT GGA GGG AAC CCA AGA GTC ACT GAC ATT AGG TGA CAG GGA TGG CTC AGA CAA TTT GTG ATT CTG AAA TGA GAA AGG ACA-3′
Hba-x R	5'biotinylated-GAC AGA AGT GAC AAA GAG CAG TTA GTG AGG AGA AGC CAC TGA GAA GGG GTC ATG AGG ACA GTA TGG TCA GGA CAG TAA GGA CAG AGG CAG AGA GGA CAA CGT GGA GAG GAC TGA GAG ATC-3′
Mitoferrin F	5'biotinylated-GAT CGG ACG GCT CAG AGC CAC AGC CAC CTC CGC GCA CTG AGT GCT CAC CAC TTG TCC CCC ACG ACC ACG ACA GAG ACA GAA GTG CAA GAC TCC CTG GCG GAA TCT AGG GGT ACT CCC CCG-3′
Mitoferrin R	5'biotinylated-GGG AGA GTC AGC AGA AGG GAT ATG CTC AGA GCT GCA AGA GCA GGA GAA GGG GCT GTC CCC AGA ATC GGT AGT TTT GCA GAC ACT AAT CAA CCA GCT TTA GCA TTC CGA AGG CTG GGG ATC-3′

Table 2.5 Capture oligonucleotide sequences

2.11.10 Hybridization of Capture oligonucleotides

Hybridization reactions were performed using the Nimblegen SeqCap SR kit (Roche). A mixture of 1.5-2 µg of Capture-C library with adaptors and index primers, 5 µl of COT DNA (Invitrogen), 1 nM TS-HE universal oligo and 1 nM TS-HE index oligo was prepared and samples were dried in a vacuum centrifuge at 50^{o}C. Careful reconstitution of DNA by pipetting and vortexing was performed with a mixture of 7.5 µl of 2x hybridization buffer and 3 µl of hybridization component A. Samples were heat denatured at 95°C for 10 min and briefly centrifuged. Reactions were then added to 4.5 µl of 2.89 µM

pooled biotinylated oligonucleotides pre-heated to 47°C in a separate 0.2 ml PCR tube. The mixture was centrifuged briefly and incubated at 47°C for 64-72 hr.

2.11.11 Washing and recovering captured material

This step was performed according to SeqCap EZ library SR user's guide v4.2 (Roche). In brief, wash buffers and 100 µl of capture beads per sample were prepared in accordance with the Manufacturer's instructions. Hybridization samples were thoroughly mixed with the Streptavidin beads. This mixture was heated at 47°C on a Thermomixer at 600 rpm for 45 min. Removal of unbound DNA was performed by adding and mixing 100 µl of 1x wash buffer I pre-heated to 47°C to samples containing capture beads and bound DNA. Tubes were placed in the DynaMag-2 device to bind the beads. Sequential washes were performed with 2X 200 µl of 1X wash buffer preheated to 47°C and 200 µl each of room temperature 1X wash buffer I, II and III. After the application of wash buffer III, tubes were removed from the DynaMag-2 device and 42 µl of nuclease-free water (Sigma) added to resuspend bead-bound captured samples.

2.11.12 Amplification of captured multiplex DNA sample using PCR

This step was performed using SeqCap EZ accessory kit V2 (Roche) in accordance with V4.2 of the User's Guide. Amplification of captured multiplex DNA samples, bound to the capture beads was achieved using a total of two PCR amplications per sample, subsequently combined, to minimize PCR bias. A total of 60 µl of post-capture PCR master mix was prepared from 50 µl

of KAPA HiFi HotStart Ready Mix (Kapa Biosystems) and 10 µl of 5 µM Post-LM-PCR Oligos. Aliqouts of 30 µl of post-capture PCR master mix were added to and mixed well with each 20 µl aliquot of bead-bound captured DNA. PCR amplification was performed at 98°C for 45 sec, 13 cycles of at 98°C for 15 sec, at 60°C for 30 sec and at 72°C for 30 sec, and at 72°C for 1 min before holding at 4°C.

2.11.13 Post-hybridization amplification clean-up

Each post-hybridization amplification sample was purified separately using Ampure XP beads in accordance with the manufacturer's protocol. In brief, 1.8 volumes of homogenous beads was added to each amplified library and mixture were incubated at room temperature for 5 min. Tubes were placed in a magnetic stand to bind the beads and the beads were washed twice with 500 µl fresh 70% ethanol. Tubes were removed from the magnetic stand and samples air dried at room temperature for 5 min. Amplified post-hybridization DNA was eluted with 30 µl of nuclease-free water and transferred to a fresh 1.5-ml tube. Amplified samples were assessed using a DNA 1000 Screen Tape (Agilent Technologies).

2.11.14 Double Capture

A second capture was performed to increase effectiveness of the capture by repeating procedure detailed in Sections 2.11.10 - 2.11.13. A pool from 25 µl each of materials from the 1st capture was used from this step. The hybridization of capture oligonucleotides was performed for only 24 hours for the second capture.

2.11.15 Library quantification for the MiSeq and analysis of Capture-C

The 4 nM solution of double-captured library was prepared separately from each of the two amplification reactions, using concentrations measured with the Qubit dsDNA BR assay kit (Invitrogen) and the fragment size obtained from the DNA 1000 Screen Tape (Agilent Technologies). 4nM samples were pooled and re-measured with the Qubit dsDNA HS assay kit (Invitrogen) for accuracy. Samples were sequenced using a MiSeq device (Illumina). Reads were mapped and bioinformatics analysis of Capture-C performed in primitive erythroblasts were carried out by James Davies using custom scripts (Davies et al., 2016) and those of ES cells were performed using a Capture-C analysis pipeline created by Jelena Telenius of the Computational Biology Research Group (CBRG).

2.12 RNA-sequencing

2.12.1 Preparation of total RNA samples

Total RNA was extracted from E10.5 yolk sac-derived primitive erythroblasts, *in vitro* cultured primitive erythroblasts and fetal liver culture-derived definitive erythroblasts according to the method described in RNA procedures (Section 2.3.1). All samples were DNaseI treated using recombinant DNaseI (Roche) at a concentration of 1 unit per 1 µg of RNA with an incubation time of 30 min at 37°C. Total RNA was purified using the RNeasy Minikit (Qiagen), according to the manufacturer's protocol, prior to concentration measurement using a Nanodrop spectrophotometer (Thermo Scientific). RNA quality was assessed using a Bioanalyzer RNA 6000 Nano kit (Agilent Technologies). Only total

RNA samples with a RIN (RNA integrity) score of > 8 were selected for RNA sequencing.

2.12.2 Globin and ribosomal depletion

Globin and ribosomal depletion as well as RNA sequencing were performed by the high-throughput sequencing service at the Wellcome Trust Centre for Human Genetics (WTCHG). To maximize the breadth of coverage, either of two different globin depletion processes was employed. The first method involved depletion of α- and β-globin transcripts and ribosomal RNA (Globin Zero Gold, Epicentre) prior to fragmentation of RNA by adaptive focused acoustics (Covaris). Subsequently, reverse transcription and construction of cDNA libraries was performed. The second method also involved depletion of α- and β-globin transcripts, ribosomal transcripts and included custom oligonucleotides designed to deplete embryonic ζ-, γ-, βH1-globin transcripts (NuGEN). The NuGEN globin and ribosomal depletion procedure, however, was performed after reverse transcription and construction of cDNA libraries.

2.12.3 Sequencing of the libraries and data analysis

Libraries were sequenced using the Illumina Hiseq2000 platform as 2 x 100 bp paired end reads. Bioinformatic analysis of RNA-seq data was performed by Emmanouela Repapi of the CBRG. The reads were aligned using Tophat (Trapnell et al., 2009) against the mouse genome assembly (mm9 UCSC). Non-uniquely mapped reads and reads that were identified as PCR duplicates using Samtools (Li et al., 2009) were discarded. Gene expression levels were quantified as read counts using the featureCounts function (Liao et al., 2014) from the Subread package (http://subread.sourceforge.net; SourceForge

package version 1.4.5) with default parameters. Read counts were used for the identification of global differential gene expression between specified populations using the edgeR package (Robinson et al., 2010). RPKM values were also generated using the edgeR package. Genes were considered differentially expressed between populations if they had an adjusted *p*-value (FDR) of less than 0.05. Sequence data were visualized and assessed using the UCSC genome browser (Kent et al., 2002). Differentially expressed (DE) genes were annotated using the DAVID functional annotation tool (Huang et al., 2007a; Huang et al., 2007b).

2.13 Whole genome sequencing

Whole genome sequencing was performed by the Wellcome Trurst Centre for Human Genetics. DNA was isolated from a lymphoblastoid cell line derived from the proband prepared by phenol chloroform extraction and libraries were sequenced on an Illumina GAIIx platform with 100 bp paired-end reads. Bioinformatic analysis of WGS was performed by Simon McGowan of the CBRG. Mapping was carried out with STAMPY software (Lunter and Goodson, 2011) to the hg19 genome. Variants were called with the SAMtools program (Li et al., 2009) and their effects on coding genes determined by ANNOVAR (Wang et al., 2010).

2.14 Mouse Transcriptome Arrays

2.14.1 Preparation of total RNA samples

Total RNA was extracted from E10.5 yolk sac-derived primitive erythroblasts, unsorted fetal liver cultured definitive erythroblasts and CD44-hi sorted fetal

liver cell culture-derived definitive erythroblasts according to the method described in RNA procedures (Section 2.3.1). These total RNA samples were used for both RNA-sequencing and array analyses. All samples were DNaseI treated using recombinant DNaseI, RNase-free (Roche) at a concentration of 1 unit per 1 µg of RNA with an incubation time of 30 min at 37°C. Total RNA was purified using the RNeasy Minikit (Qiagen), according to the manufacturer's protocol, prior to concentration measurement using a Nanodrop spectrophotometer (Thermo Scientific). Quality of total RNA was assessed using the RNA TapeStation kit and analyzer (Agilent Technologies). Only total RNA samples with RIN score of > 8 were selected for the transcriptome array experiment.

2.14.2 Preparation of pre-hybridized sense-stranded cDNA (ss-cDNA)

Single-stranded cDNA (ss-cDNA) was prepared from total RNA using the Whole transcriptome Plus reagent kit (Affymetrix) according to the manufacturer's protocol. In brief, 100 ng of total RNA was primed with oligonucleotides containing a T7 promoter sequence and was reverse transcribed resulting in ss-cDNA with a T7 promoter sequence at the 5′ end. Second-strand cDNA was synthesized to create double-stranded cDNA. Subsequently, cRNA was synthesized and amplified overnight by *in vitro* transcription (IVT) of the ss-cDNA template using T7 RNA polymerase before being purified using beads. Purified cRNA acted as a template in the reverse transcription reaction using the 2nd-cycle primers provided resulting in sense-stand cDNA, which contained dUTP at a fixed ratio relative to dTTP. Sense-strand cDNA was purified using purification beads and template cRNA was

hydrolyzed using RNase H. The fragmentation procedure was performed using uracil-DNA glycosylase (UDG) and apurinic/apyrimidinic endonuclease 1 (APE 1) to degrade the sense-strand cDNA at the dUTP residues. The fragmented cDNA was labeled by terminal deoxynucleotidyl transferase (TdT) using an proprietary DNA labeling reagent covalently linked to biotin.

2.14.3 WT Array Hybridization

Pre-hybridization reactions were prepared according to the manufacturer's protocol. Fragmented and labeled sense-strand cDNA was hybridized to Mouse Transcriptome Array (MTA) 1.0 GeneChips (Affymetrix) for 16 hr, at 45°C, at rotation speed 60 rpm. Subsequently, GeneChips were washed and stained using a Fluidic Station (Affymetrix) prior to the scanning process.

2.14.4 Analysis of MTA 1.0

MTA 1.0 GeneChips were scanned using the Affymetrix GeneChip scanner and GeneChip Command Console program®. Data analysis was performed using Expression Console® and Transcriptome Analysis Console® software (Affymetrix). Genes were considered differentially expressed between populations if they had an adjusted *p*-value (FDR) of less than 0.05.

2.15 Array Comparative Genomic Hybridization (aCGH) analysis

Array CGH was performed by the Clinical Diagnostic Facility in the John Radcliffe Hospital. DNA extracted from a patient-derived lymphoblastoid cell line was analyzed using the Illumina Human OmniExpressExome platform (Illumina®), containing over 273,000 functional exonic markers. Initial copy

number variation calls were performed according to the manufacturer instructions by Pauline Robbe (Diagnostic Laboratory, John Radcliffe Hospital) using the Nexus copy number analysis software. The human genome hg19 sequence release was used for this analysis. Copy number variation was assayed, for overlap with variants present in the database of genomic variants (DGV) (www.projects.tcag.ca/variation/) as a part of the prioritization pipeline.

Chapter 3 : The BHFS survivor registry

3.1 Introduction

The Hb Bart's Hydrops Fetalis Syndrome (BHFS), caused by α^0-thalassemia, has until recently been considered a universally fatal disorder. However, over the last two decades ever-improved antenatal diagnosis, intrauterine intervention and postnatal intensive care have resulted in increasing numbers of long-term survivors of the BHFS. Several of these cases have been published, however, we have become aware of many more unreported survivors, who are under care of and who have received a variety of treatments from the local obstetricians and hematologists.

Affected fetuses with the BHFS may be maintained with intrauterine intervention and/or postnatal transfusion, however, many questions remain regarding their long-term clinical outcome and their quality of life. The first question regards how well their neurological development has proceeded. This is a major clinical concern since affected infants with the BHFS experience intrauterine tissue hypoxia for considerable periods of time, from the point when ζ-globin expressing primitive erythropoiesis is replaced with fetal definitive erythropoiesis (~6-8 weeks of gestation) to the time of diagnosis and initiation of treatment. Moreover, autopsies of affected fetuses show there is a progressive decrease in brain weight relative to that expected for gestational age after ~8 weeks of gestation (Higgs, 2009b). A previous review of 22 infants with the BHFS who survived naturally until birth and received neonatal intensive care reported that at least half of the infants who

survived for months or years suffered from abnormal neurological development (Lee et al., 2007). The second question concerns how frequently the BHFS survivors are born with associated congenital abnormalities, their severity and how these abnormalities affect their quality of life. Congenital abnormalities, including hydrocephaly, microcephaly, abnormal limb development and urogenital abnormalities, have previously been reported in up to 17% of all hydropic fetuses from ultrasonographic findings and analysis of a small number of survivor cases (reviewed in (Higgs, 2009b)). A larger review of BHFS survivors also reported that urogenital defects (hypospadias being the most common) are the most frequently associated congenital abnormalities and are found in 5 out of the 12 BHFS survivors included in the study (Singer et al., 2000). However, many more cases need to be reviewed to provide a realistic prevalence of these co-morbidities. The third question addresses how many of the BHFS survivors have eventually undergone curative treatments and how many have the burden of life-long transfusion. Improved knowledge of these three issues is likely to provide answers to three important questions: “Could the BHFS ever be considered salvageable disease?” and if it could “how should we best treat the patients?” and lastly “how should we provide counseling to couples at risk of having a child born with the BHFS”.

A major aim of this thesis is to fully document the natural history and clinical outcomes of long-term survivors of the BHFS throughout the world, with the ultimate aim of providing guidelines for management of these patients. To achieve this, I have initiated a BHFS survivor registry. The registry currently

contains clinical information from 60 cases, of which 28 have not been previously reported.

3.2 Results

3.2.1 General clinical characteristics of the BHFS survivors

Of a total of 32 previously reported long-term survivors with the BHFS in this registry, 4 had been reported between 1986 and 1994, 18 from 1995 to 2004, and 10 since 2005. Among this group of patients, the clinical information from 7 survivors has been updated in the current registry, by means of personal communication with the corresponding authors or the local hematologists providing clinical care. We have also become aware of a further 28 unpublished survivors with the BHFS, of which 11 cases have been followed up at two different centers in Canada, 5 are from two different centers in Hong Kong, 4 are from a single center in Thailand, 2 from two different centers in Australia, 2 from two different centers in the US, 2 from a single center in Malaysia and the remaining 2 cases are from Sweden and Poland.

At the time of report or last communication sixteen patients have reached beyond the age of 10, seventeen patients are between 5 and 10 years of age and 26 patients are ≤ 5 years old. The oldest BHFS survivor in the registry is currently 31 years old. One patient had suffered from severe mental retardation and died at the age of 11 due to respiratory failure. Gender has been reported in 50 BHFS survivors and male to female ratio is 1.63:1 (Male = 31, Female = 19) in this registry. However, given that survival of the BHFS is a very rare condition resulting in small number of patients recruited, it might not be possible to formally demonstrate the preferential occurrence of this

condition in males. Of a total of 60 long-term survivors with the BHFS, 33 had undergone at least one episode of intrauterine intervention. The majority of those 27 cases who had survived naturally until birth received transfusion within the first few hours after they were born and all except one received transfusion within the first day of life. The exceptional case received his first transfusion on the second day of life and is of special interest (this case will be discussed in detail in Chapter 6).

The α-globin genotype has been reported in 43 BHFS survivors, 42 of which are homozygous for the South East Asian deletion ($--^{SEA}/--^{SEA}$), whereas the remaining case has compound heterozygosity for the South East Asian and the Filipino deletions ($--^{SEA}/--^{Fil}$). The general clinical characteristics of long-term survivors of the BHFS are summarized in Table 3.1. All survivors for whom ethnicity has been reported are of South East Asian heritage including Filipino, Thai, Vietnamese, Laotian, Hmong, and Chinese.

Clinical characteristics	Patients (total n=60)
Age (years)	
0-5	26
> 5-10	17
>10	16
NA	1
Gender	
M	31
F	19
NA	10
GA at birth (weeks)	
≥ 37	15
≥ 30-36+	34
< 30	8
NA	3
First transfusion	
***In utero* at GA (weeks)**	33
≥ 30	6
≥ 20-29+	19
< 20	2
NA	6
Postnatal	27
Current status	
Transfusion-dependent	49
Transplanted	10
Death	1

Table 3.1 General clinical characteristics of the long-term BHFS survivors
Total of 60 survivors are included in this study. GA = gestational age, NA = data not available

Case	Reference/Referring physician	Age at time of report/last evaluation	Gender	Delivery/GA(w)/Wt(kg) /hydropic?	1st Hb level (g/dl)	Immediate postnatal intervention		Hb Portland (%) at birth	Neonatal intervention & course	Congenital malformation	Long-term treatment	Growth	Development
						Transfusion	Ex Transfusion						
1	Beaudry 1986, Jackson 1990 (Beaudry et al., 1986; Jackson D.N. et al., 1990)	(6 y)	M	CS/33/2.3/hydropic	9.7	+	+	20	Apgar 6, 8 (at 1, 5 min), ventilated-29d, convulsion x1, pulmonary hemorrhage	Hypospadias	Transfusion	Wt 5th centile, Ht 10th centile	Delayed speech & hearing
2	Bianchi 1986, Fischel-Ghodsian 1987, Jackson 1990 (Bianchi et al., 1986; Fischel-Ghodsian et al., 1987; Jackson D.N. et al., 1990), FU data (Bauer D)	31y	F	CS/28/1.1/non-hydropic	~ 9.6	+	+	19	Apgar 3, 4 (at 1, 5 min), ventilated- 3w,	None	Transfusion	"Normal"	"Normal"
3	Jackson 1990 (Jackson D.N. et al., 1990)	10 mo	NA	PV/28/NA/hydropic	NA	+	+	7	NICU 64 d, paracentesis	NA	NA	Wt 5th centile, Ht 5th centile	"Mild delayed"
4	Lam 1992 (Lam et al., 1992), FU data (Li CK)	11 y	F	CS/31/1.6/non-hydropic	7.9	+	−	Present	Apgar 3, 6 (at 1, 5 min), ventilated- 8d, pericardial effusion, during the 3 mo of life suffered from septicemia and heart failure	None	Transfusion	< 3rd centile	"Severely delayed", spastic quadriplegia, died at 11 y of age due to respiratory failure
5	Chik 1998 (Chik et al., 1998), FU data (Chan V)	20 y	F	CS/35/1.9/hydropic	6.3	+	−	Present	Apgar0, 1 (at 1, 5 min), CPR, ventilated	None	Transplanted at 21 mo	3rd centile	Delayed development in early childhood, now enter main-stream school with mild learning difficulty
6	Fung 1999 (Fung et al., 1999)	6 y	M	PV/35/1.9/NA	NA	+	+	15	NA	Hypospadias	NA	NA	"Normal"
7	Singer 2000 (Singer et al., 2000)	2.5 y	F	CS/34/2.2/hydropic	7.4	+	+	5.2	Ventilated- 6d, metabolic acidosis	ASD	Transfusion	Wt 15th centile, Ht 5th centile	"Normal"
8	Liu 2002 (Liu et al., 2002)	< 5 y?	M	NA/32/2.3/hydropic	6	+	?	NA	Apgar 3,6 (at 1, 5 min), acidosis	NA	NA	NA	NA
9	Zhou 2001 (Zhou et al., 2001), FU data (Ha SY)	18 y	F	PV/24/1/mild hydropic	8.7	+	−	19.3	Apgar 2, 7 (at 1, 5 min), ventilate- 2 mo	PDA (requiring ligation)	Transplanted at 20 mo	Wt 50th centile, Ht 10th centile	Mild learning difficulty

Table 3.2 Natural history and clinical outcomes of the long-term BHFS survivors who had survived naturally until birth

NA= Data not available, GA= Gestational age, w= weeks, mo= months, y= years, Wt= Body weight, Ht= Height, FU= Follow-up, Ex transfusion= Exchange

Case	Reference/Referring physician	Age at time of report/last evaluation	Gender	Delivery/GA(w)/Wt(kg) /hydropic?	1st Hb level (g/dl)	Immediate postnatal intervention		Hb Portland (%) at birth	Neonatal intervention & course	Congenital malformation	Long-term treatment	Growth	Development
						Transfusion	Ex Transfusion						
10	Lee 2007 (Lee et al., 2007)	16 y	M	NA/35/1.9/mild hydropic	7.8	−	+	NA	Ventilated- 4d	NA	NA	NA	"Normal"
11	Lee 2009 (Lee et al., 2009)	2 y	F	PV/35/1.1/non-hydropic	9	−	+	"Small amount"	Apgar 4,6 (at 1, 5 min), ventilated- 14 d, PPHN (requiring HFOV, NO)	Hypoplasia/ aplasia of digits, jejunal atresia	Transfusion	NA	"Mild delayed"
12	Gumuscu 2013 (Gumuscu et al., 2013), FU data (Coccia P)	10 y	F	CS/30/1.2/hydropic	NA	+	−	25.4	Ventilated- 2 mo	ASD	Transplanted at 44 months	"Normal"	"Normal"
13	Pongtanakul 2013 (Pongtanakul et al., 2013)	6 y	F	CS/32/1.3/hydropic	6.8	−	+	None	Apgar 1, 4, 7 (at 1, 5, 10 min), ventilated, abdominal and pleural paracentesis	None	Transplanted at age 19 mo	50th centile	Mullen Scale of Early Learning: normal
14	Howarth S & Cole C	2 y	M	CS/33/1.9/hydropic	9.7	−	+	17	Ventilated?, metabolic instability, cortical brain infarction	Inguinal hernia	Transfusion	NA	"Mild delayed"
15	Bowden DK	5 y	F	PV/32/1.6/hydropic	7	+	?	10	NA	Asymmetrical hand size	NA	"Severely retarded"	"Normal"
16	Gallagher P	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17	Viprakasit V (KD)	6 y	M	PV/39/2.3/non-hydropic	11.2	Few days after birth	−	NA	Ventilated, PPHN with pneumothorax (requiring HFOV), sepsis, PDA with CHF, neonatal jaundice, multiple admission with recurrent pnuemonia and poor growth during 1st year of life	Hypospadias	Transfusion	< 3rd centile	6-12 mo. delayed
18	Viprakasit V (KT)	13 y	M	NA/32/1.8/non-hydropic	NA	−	+	NA	Ventilated-1 mo., neonatal jaundice	None	Transfusion	"Normal"	"Normal"

Table 3.2 Natural history and clinical outcomes of the long-term BHFS survivors who had survived naturally until birth (continued)

Case	Reference/Referring physician	Age at time of report/last evaluation	Gender	Delivery/GA(w)/Wt(kg) /hydropic?	1st Hb level (g/dl)	Immediate postnatal intervention		Hb Portland (%) at birth	Neonatal intervention & course	Congenital malformation	Long-term treatment	Growth	Development
						Transfusion	Ex Transfusion						
19	Viprakasit V (BS)	14 y	M	CS/28/1.5/non-hydropic	NA	+	?	NA	Ventilated?	None	Transfusion	"Normal"	"Normal"
20	Viprakasit V (RS)	7 y	M	PV/27/1/non-hydropic	NA	−	+	NA	Apgar 3,6,7 (at 1,5, 10 min), CPR, ventilated-9 d, CPAP 8-d, jaundice	Toes deformities, hydrocele	Transfusion	"Normal"	"Normal"
21	Merson L & Nancy (A)	2 y	M	CS/39/3.5/NA	NA	−	+	NA	NA	Hypospadias	NA	NA	NA
22	Merson L & Nancy (C)	4 y	M	CS/38/2.5/NA	NA	−	+	NA	NA	Undescended testes	NA	NA	NA
23	Yeoh SL (A)	6 y	M	PV/32/1.4/non-hydropic	9	+	?	NA	Ventilated- 7d, CPAP- 3d, severe neonatal jaundice, bilateral undescended testes, pulmonary hemorrhage	Bilateral undescended testes	NA	< 3rd centile	"Normal"
24	Yeoh SL (B)	6 y	M	PV/32/1.2/non-hydropic	10.7	+	?	NA	Ventilated-3wk, CPAP- 10d, pulmonary hemorrhage	Hypospadias, hydronephrosis, hydrocele, undescended testes	NA	< 3rd centile	"Normal"
25	Ha SY (B)	5 y	M	CS/33/1.4/non-hydropic	6.9	+	?	Present	NA	Ambiguous genitalia, micropenis, hypospadias, undescended testes, microcephaly	Transfusion	Wt 25-50th centile, Ht 50th centile	"Normal"
26	Chan V (KWH2)	4 y	NA	NA/NA/NA/NA	6.6	+	?	Present	NA	NA	Transfusion	NA	NA
27	Chan V (KY4)	17 y	NA	NA/23/NA/NA	NA	NA	NA	NA	NA	NA	Transplanted	NA	DQ: normal to mild delayed, "learning difficulty"

Table 3.2 Natural history and clinical outcomes of the long-term BHFS survivors who had survived naturally until birth (continued)

Case	Reference/Referring physician	Age at time of report/last evaluation	Gender	Delivery/GA(wk)/Wt(kg) /hydropic?	Prenatal intervention			GA(w) at 1st intrauterine intervention	Neonatal intervention & course	Congenital malformation	Long-term treatment	Growth	Development
					Intrauterine transfusion	Intrauterine Ex Transfusion	Intrauterine Stem cells infusions						
28	Carr 1995 (Carr et al., 1995)	2 y	M	CS/34/2.2/non-hydropic	+	+	−	26	Apgar score 8,9 at 1,5 min, CPAP-3 d, transfusion at d1 of life,	Missing 1/3 of foot, syndactyly, hand defects, hypospadias, undescended testes	Transfusion	< 3rd centile	Psychological and motor delay ~ 5 mo.
29	Westgren 1996 (Westgren et al., 1996)	3 y	M	CS/37/2.4/non-hydropic	+	−	+	15	Apgar 8,9 at 1,5 min, minor bleeding from GI and thrombocytopenia, require 1st blood transfusion at 6 w	None	Transfusion	NA	Development =24 mo at age 27 mo
30	Naqvi 1997 (Naqvi A et al., 1997)	3 y	NA	CS/37/NA/NA	+	−	−	32	Ventilated-1d, left portal vein obstruction, thrombocytopenia	None	Transfusion	NA	Neurologically "normal"
31	Hayward 1998 (Hayward et al., 1998)	9 mo	M	NA/36/NA/non-hydropic	+	−	+	13	NA	Hypospadias	Transfusion	NA	"Normal"
32	Ng 1998 (Ng et al., 1998)	3 mo	M	CS/29/1.3/non-hydropic	+	−	−	23	Ventilated-11d, IVH gr I, surfactant therapy for RDS	None	Transfusion	NA	NA
33	Dame 1999 (Dame et al., 1999b)	2 y	M	CS/37/2.5/hydropic	+	−	−	29	Transfusion immediate after birth, ex transfusion d7, ventilated-2d	Hypospadias, bifid scrotum	Transfusion	10th centile	"Normal"
34	Fung 1998 (Fung et al., 1998), FU data (Li CK)	17 y	M	CS/29/1.4/non-hydropic	−	+	−	23	NA	Hypospadias	Transfusion	Short stature	"Normal"
35	Leung (A) 2002 (Leung et al., 2002)	<5 y?	NA	CS/37/2.5/NA	+	−	−	31	Ex transfusion after birth, otherwise unremarkable	NA	Transfusion	NA	NA

Table 3.3 Natural history and clinical outcomes of the long-term BHFS survivors who received intrauterine treatment

NA= Data not available, GA= Gestational age, w= weeks, mo= months, y= years, Wt= Body weight, Ht= Height, FU= Follow-up, Ex transfusion= Exchange transfusion, CS= Caesarean section, PV= Per vaginal, NICU= Neonatal intensive care unit, CPR= Cardiopulmonary resuscitation, PPHN= Persistent pulmonary hypertension of the newborn, HFOV= High Frequency Oscillatory Ventilation, NO= Nitric Oxide, CPAP= Continuous Positive Airway Pressure, NPCPAP= Nasopharyngeal Continuous Positive Airway Pressure, PDA= Patent ductus arteriosus, ASD= Atrial septal defect, CHF= Congestive heart failure, IVH= Intraventricular hemorrhage, RDS= Respiratory distress syndrome

Case	Reference/Referring physician	Age at time of report/last evaluation	Gender	Delivery/GA(wk)/Wt(kg) /hydropic?	Prenatal intervention			GA(w) at 1st intrauterine intervention	Neonatal intervention & course	Congenital malformation	Long-term treatment	Growth	Development
					Intrauterine transfusion	Intrauterine Ex Transfusion	Intrauterine Stem cells infusions						
36	Leung (B) 2002 (Leung et al., 2002)	< 5 y?	NA	PV/34/1.6/NA	+	−	−	31	Ex transfusion after birth, otherwise unremarkable	NA	Transfusion	NA	NA
37	Sohan 2002 (Sohan et al., 2002), FU data (Roberts I)	12 y	F	PV/38/3.2/NA	+	−	−	21	Apgar score 8,10,10 (at 1,5, 10 min), 2 x exchange tx in 48 hr, otherwise uneventful	Mild lobster claw deformity of the right foot	Transplanted in Feb 2012	"Poor"	"Normal"
38	Bizzarro 2003 (Bizzarro et al., 2003), FU data (Bhandari V)	11 y	M	PV/34/1.8/non-hydropic	− (fetal paracentesis was performed)	+	−	30	Apgar 5, 7 (at 1, 5 min), pulmonary hypoplasia, ventilated-11d, PPHN (requiring NO & HFOV, surfactant therapy), transfusion at d6 NPCPAP-2d	Hypospadias, cryptorchidia, pulmonary hypoplasia	Transfusion	Short stature	Reported as below average in all subjects at school
39	Thornley 2003 (Thornley et al., 2003)	5 y	M	PV/34/2/non-hydropic	+	−	−	26	No neonatal complication	Hypospadias	Transplanted at age 23 mo	Normal growth velocity	Appropriate developmental milestones
40	Joshi 2004 (Joshi et al., 2004)	3 y	M	PV/34/1.7/non-hydropic	+	−	−	27	Apgar 1,5 (at 1,5 min), transfusion afterbirth, ventilated, NICU 2 mo	None	Transfusion	Short stature	Age appropriate
41	Lucke 2005 (Lucke et al., 2005)	6.5 y	M	CS/34/2/hydropic	+	−	−	After 20	Ventilated- first days of life, pulmonary HT	Hypoplasia of phalanges, hypospadias	Transfusion y	Wt 15th centile, Ht 30th centile	Columbia Mental Maturity Scale- 1 y delayed of intelligence and cognition, Peabody Picture Vocabulary Test- 1 y delayed, Kaufman Assessment Battery for Children- 1 y delayed of intellectual level
42	Wang and Ryan 2009 (Wang and Ryan, 2009)	< 5 y?	F	PV/36/NA/NA	+	−	−	NA	NA	NA	Transfusion	NA	NA
43	Yi 2009 (Yi et al., 2009)	7 y	M	CS/36/2.6/non-hydropic	+	−	−	25	Unremarkable	None	Transplanted at 18 mo old	Short stature	Age appropriate school grade (special education)

Table 3.3 Natural history and clinical outcomes of the long-term BHFS survivors who received intrauterine treatment (continued)

Case	Reference/Referring physician	Age at time of report/last evaluation	Gender	Delivery/GA(wk)/Wt(kg) /hydropic?	Prenatal intervention			GA(w) at 1st intrauterine intervention	Neonatal intervention & course	Congenital malformation	Long-term treatment	Growth	Development
					Intrauterine transfusion	Intrauterine Ex Transfusion	Intrauterine Stem cells infusions						
44	Dwinnell 2011 (Dwinnell et al., 2011)	15 mo	F	PV/35/2.3/non-hydropic	+	+	−	NA	Apgar 9,10 (at 1,5 min), exchange tx after birth, otherwise uneventful	Hand and foot deformities	Transfusion	Wt 3rd centile, Ht 10-25th centile	Mild gross motor delay
45	Kou 2013 (Kou et al., 2013)	< 5 y?	NA	CS/31/NA/NA	+	−	−	NA	NA	NA	Transfusion	NA	NA
46	Chmait 2014 (Chmait et al., 2014)	1.5 y	NA	NA/NA/NA/NA	+	−	−	NA	NA	NA	Transfusion	NA	Battelle Developmental Inventory, DQ-normal
47	Teng J	6 wk	M	CS/36/NA/non-hydropic	+	−	−	NA	None	Hypospadias	Transfusion	"Normal"	"Normal"
48	Merson L & Nancy (B)	8 y	F	CS/37/3.1/NA	+	−	−	27	NA	None	Transfusion	NA	NA
49	Merson L & Nancy (D)	4 y	F	PV/33/1.6/NA	+	−	−	23	Ex transfusion at birth, otherwise-NA	None	Transfusion	NA	Mild developmental delay
50	Ha SY	6 y	F	CS/38/2.1/non-hydropic	+	−	−	21	Pulmonary hypertension, thrombocytopenia (ventilated?), prolonged hospitalization	ASD, congenital stenosis of pulmonary valve	Transplanted at agr 21 mo	Wt 3-10th centile, Ht < 3rd centile	~6 mo globally delayed (Griffiths Mental Development Scale
51	Turowski P	1 y	F	CS/33/1.6/non-hydropic	+	−	−	25	Ventilated	None	Transfusion	< 3rd centile	NA

Table 3.3 Natural history and clinical outcomes of the long-term BHFS survivors who received intrauterine treatment (continued)

Case	Reference/Referring physician	Age at time of report/last evaluation	Gender	Delivery/GA(wk)/Wt(kg) /hydropic?	Prenatal intervention			GA(w) at 1st intrauterine intervention	Neonatal intervention & course	Congenital malformation	Long-term treatment	Growth	Development
					Intrauterine transfusion	Intrauterine Ex Transfusion	Intrauterine Stem cells infusions						
52	Chan V (KY2)	2 y	NA	CS/32/NA/NA	+	−	−	NA	CPR, ventilated	None	Transfusion	NA	2-3 mo delayed
53	Amid A (EC)	15 y	F	NA/33/1.6/NA	+	−	−	31	Ventilated-13d, Transfusion after birth, sepsis, IVH gr II	None	Transfusion	Wt & Ht 5th centile	General intelligence-average
54	Amid A (JA)	13.5 y	M	NA/38/3.5/NA	+	−	−	25	Ventilated-3d, PPHN, thrombocytopenia, transfusion after birth	Hypospadias	Transfusion	Wt & Ht 5th centile	General intelligence-average
55	Amid A (NS)	15.5 y	M	NA/37/2.5/NA	+	−	−	31	Ventilated?, Bilat pneumothorax, transfusion after birth	None	Transfusion	Wt & Ht 3rd centile	General intelligence-below average
56	Amid A (JF)	19.2 y	F	NA/37/3.1/NA	+	−	−	26	Ventilated?, Portal vein thrombosis, transfusion at d 5	None	Transfusion	Wt & Ht 5th centile	General intelligence-average
57	Amid A (JT)	6.6 y	M	NA/38/3.3/NA	+	−	−	28	Transfusion at d 17, otherwise-unremarkable	ASD, undescended testes	Transplanted	Wt 3rd centile, Ht 10th centile	NA
58	Amid A (AA)	5.1 y	F	NA/36/2.3/NA	+	−	−	23	Transfusion at d 28, none	None	Transfusion	Wt & Ht 3rd centile	NA
59	Amid A (AT)	5.6 y	M	NA/37/2.5/NA	+	−	−	24	Ventilated?, tranfusion at d 2, PPHN	Hypospadias, undescended testes, rib deformity	Transfusion	Wt & Ht 3rd centile	NA
60	Bauer D	13 y	M	PV/34/1.8/non-hydropic	+	−	−	23	Ventilated-21d	Hypospadias (repaired), undescended testes (repaired), hip dysplasia	Transfusion	Short stature	Learning difficulties

Table 3.3 Natural history and clinical outcomes of the long-term BHFS survivors who received intrauterine treatment (continued)

3.2.2 Prenatal diagnosis and perinatal intervention

Seventy percent (7/10 cases) of cases born before 1995 had survived naturally until birth without receiving intrauterine treatment, whereas 63% (19/30 cases) and 58% (11/19 cases) of those born between 1995-2004 and between 2005-2015 respectively underwent either intrauterine transfusions or stem cell infusions. These data show the advances in prenatal diagnosis and management over the past two decades. Of those treated *in utero*, 27 cases received blood transfusion alone, 4 cases underwent exchange transfusion, and 2 cases underwent hematopoietic cell transplant. Initiation of the interventions has been reported up to 13 weeks after the initial prenatal diagnosis, this reflected the time taken to confirm the molecular diagnosis and for family counseling. The earliest gestational age at which intrauterine intervention was performed was 13 weeks. Details of individual BHFS survivors are shown in Tables 3.2 and 3.3.

Blood transfusions via the umbilical vein were initiated at the gestational ages of 20-29 weeks in 19 cases and from 30 weeks in 6 cases. Intrauterine transfusions were employed as frequently as every week (case 33, Table 3.3) and were carried out a maximum of 6 times before birth (case 47, Table 3.3). The clinical parameters used for fetal monitoring and indication for subsequent transfusions include umbilical vein blood sampling to allow measurement of hemoglobin levels and hemoglobin component analysis by electrophoresis, ultrasonographic assessment of fetal hydrops and peak systolic velocity in the middle cerebral artery (MCA-PSV) as this is a widely accepted indicator of fetal anemia. Intrauterine exchange transfusions were

performed in 4 cases (cases 28, 34, 38 and 44, Table 3.3), of which three had fetal ascites as a predominant symptom (cases 28, 34 and 38, Table 3.3). Other reasons reported for favoring exchange transfusion over intrauterine transfusion alone were removal of non-functional Hb Bart's (γ4) (cases 34, 38 and 44, Table 3.3) and avoidance of fetal volume overload that may aggravate existing poor cardiac function (case 34, Table 3.3). In one case, intrauterine exchange transfusion improved fetal growth, which was severely restricted after prior treatment with intrauterine transfusion alone (case 44, Table 3.3). Intrauterine exchange transfusion, however, is a more complicated and a time-consuming procedure and thus may increase risk of intrauterine infection as reported in case 34.

Intrauterine hematopoietic stem cells (HSC) transplantation was performed in 2 cases, both of which were experimental in nature and part of a research study. Although intrauterine HSC transplantation in the human fetus has previously been reported to provide a cure in primary immunodeficiency syndrome (Touraine et al., 1989), the outcomes were not convincing in the cases of the BHFS. Fetal liver cells derived from legally aborted fetuses were used as a source of HSCs in one case (case 29, Table 3.3) and haploidentical paternal CD34+ve cells were used in the other (case 31, Table 3.3). In both cases, the HSC infusion was initiated early in the second trimester of pregnancy (GA 15 weeks in case 29 – Table 3.3, GA 13 weeks in case 31 – Table 3.3). However, fetal hydrops persisted in both cases resulting in requirement for intrauterine transfusion support throughout pregnancy. In addition, these two infants have continued to require transfusion since birth consistent with unsuccessful engraftment.

3.2.3 Maternal complications

A total of 12 mothers of patients in the BHFS registry are reported to have a history of at least one stillbirth or neonatal death. The gestational age at which the stillbirths occurred ranges from 25 to 33 weeks. Two neonatal deaths were reported at 1 and 5 hours after birth. Maternal history of spontaneous abortion is not uncommon and is documented in 5 cases at gestational age of 8 to 16 weeks. The current data, similar to previously published reports (Liang et al., 1985; Nakayama et al., 1986), also suggest a high incidence of serious maternal complications. In the antenatal period, the majority of problems arise in the third trimester of pregnancy with preterm delivery being the most common obstetric complication. A total of 8 affected infants (14%) were delivered from a gestational age of 23 to 29 weeks and 34 infants (60%) from a gestational age of 30 to 36 weeks, the times at which either maternal life-threathening conditions, including pre-eclampsia and/or fetal distress were detected. This review strongly suggests that intrauterine intervention has prolonged the course of BHFS prenancies as the frequency of term delivery increases from 12% in pregnancies without intrauterine treatment to 38% when intrauterine interventions are employed. The median gestational age at birth of affected fetuses treated *in utero* is 36 weeks and this is significantly higher than those who have survived naturally until birth (median GA at birth 32 weeks, p=0.001). Other antepartum complications include polyhydramnios, oligohydramnios and intrauterine growth restriction, all of which were detected by routine ultrasound investigation, predominantly in the third trimester of pregnancy (GA 22 to 34 weeks). These three antepartum complications occur in similar frequencies irrespective of whether the mothers received

intrauterine treatment. Despite the benefits, intraueterine intervention resulted in intrauterine infection and termination of pregnancy in 2 cases (in case 32 after a uterine transfusion and in case 34 after an exchange transfusion, Table 3.3). Toxemia of pregnancy (pre-eclampsia) was reported in a total of 3 cases, one of which resulted in abruptio placenta (case 41, Table 3.3) and all of which lead to emergency caesarean sections.

Malpresentation was reported in 4 cases (~13% of cases where relevant data are available). These include 3 cases of breech and 1 case of footling presentations. Although the frequency of malpresentation is higher in the BHFS pregnancies included in this study as compared to non-hydrops pregnancies (< 5%, reviewed in (Higgs, 2009b)), it is almost three times less frequent than the 37% malpresentation reported in a previous review of 46 BHFS pregnancies (Liang et al., 1985). There are reports of delivery methods in 45 mothers of BHFS survivors, of which the majority (60%) required caesarean section and 4% required assisted vaginal delivery. Postpartum hemorrhage is documented in one case and this severe event resulted in maternal disseminated intravascular coaggulation (DIC) and histerectomy (case. 44, Table 3.3). Interestingly, two multiple pregnancies are reported in this registry: one being a triplet pregnancy resulting in only one affected infant (case 4, Table 3.2) and the other being a twin pregnancy resulting in two neonates with the BHFS (case 23 and 24 – Table 3.2). Maternal complications in the BHFS are summarized in Table 3.4.

	BHFS % *	Non hydrops % **
Antepartum		
Polyhydramnios	16	< 1
Oligohydramnios	10	
Intrauterine infection	3 ♯	
Pre-eclampsia	10	7
Abruptio placenta	3	
Premature delivery		5-10
< 30 weeks	14	
≥ 30-36^{+} weeks	60	
≥ 37 weeks	26 &	
Delivery and Postpartum		
Malpresentation	13	< 5
Assisted vaginal delivery	4	5
Caesarean section	60	2-3
Postpartum hemorrhage	3	5

Table 3.4 Maternal complications in the BHFS

* calculated according to number of cases with available data
** according to a review in Higgs, 2009 (Higgs, 2009b)
♯ occurred in only 2 cases, to which intrauterine interventions were given
& twelve out of 15 cases were born to mothers who received intrauterine interventions

3.2.4 Neonatal course of the BHFS survivors

Infants affected with BHFS and whose abnormalities were not detected during the prenatal period, usually presented with hydrops (52%) and anemia (mean Hb level 7.7 ± 1.4 g/dl) that was observed immediately after birth. Interestingly, a significant number of affected infants who survived without *in utero* transfusion and are therefore likely to have suffered from prolonged fetal anemia and severe tissue hypoxia *in utero*, did not present with hydrops (48%). Almost all of these infants, however, were anemic at birth, as confirmed by immediate postnatal hemoglobin levels (mean Hb level 9.2 ± 1.5 g/dl, not significantly different from that of hydropic BHFS infants). Other common clinical presentations include respiratory distress and

hepatosplenomegaly. In this review, the presence of clinical hydrops does not correlate with increased GA at birth (mean GA at birth is 31.6 or weeks 32.4 weeks in affected infants born with or without hydropic features respectively, p=0.65). Since the available data reporting Hb Portland ($\zeta_2\gamma_2$) levels at birth are limited, the relationship between elevated Hb Portland and the presence of hydrops remains unclear. In most affected infants who did not receive prenatal diagnosis, BHFS was confirmed by hemoglobin electrophoresis, showing Hb Bart's (γ_4) to be the major Hb component, from a few hours to a few days after birth. The presence of Hb Portland, indicating persistent upregulation of the ζ-globin gene, was reported in at least 14 cases in this group of survivors (Table 3.2). The proportion of Hb Portland ranged from 7% to 25.4% of total globin constituents (median 17%). Ninety-two percent of the BHFS survivors who received intrauterine treatment were born with a birth weight (BW) of ≥ 10^{th} centile according to GA, whereas at least 19% of affected infants without prior intrauterine treatment had low BW of < 10^{th} centile. Sixty percent of affected babies without intrauterine treatment who were born with hydrops had BW of ≥ 10-50^{th} centile. This means that their actual BW was lower than 10-50^{th} centile range. Taken together, these data suggest that intrauterine treatment can improve intrauterine fetal growth.

Immediate newborn cardiopulmonary resuscitation was reported in one case among the infants who received intrauterine treatment and in two of the cases that received no treatment. Apgar scores, which are a commonly used measure of fetal distress and recovery post-partum by assessment of heart rate, respiration, muscle tone, skin color and reflex of new born babies, are noted in 9 cases among a total of 27 survivors who had survived naturally

until birth. Median scores in this group are relatively low, indicating relatively high fetal distress and poor treatment response: at 1 minute the median score is 3 (range 0-6) and at 5 minutes the median score is 6 (range 1-8). Apgar scores are noted in 6 out of 33 BHFS survivors who had received intrauterine treatments, with median scores of 8 (range1-9) and 9 (range 5-10) at 1 and 5 minutes respectively. Considering the available data, there is a statistically significant increase in Apgar scores at each time point ($p=0.03$ at 1 and $p=0.02$ at 5 minutes) in the BHFS survivors who were treated *in utero* suggesting better response to initial resuscitation, thus less chance of severe birth asphyxia among this group of patients.

A higher proportion of BHFS survivors without prior intrauterine treatment (84.2%) required mechanical and/or assisted ventilation during the neonatal period compared to those survivors treated *in utero* (59.3%), however, statistical significance was not observed. The median duration of mechanical and/or assisted ventilation required in affected infants who did not receive prior intrauterine treatment is 19 days (range 4-60 days) and is significantly longer than that required in affected infants who received intrauterine treatment (3 days, range 1-21 days, $p=0.01$). Infants who received blood transfusion *in utero* often experienced a delay in initiation of postnatal transfusion. Sixty-five percent of affected infants in this group received postnatal transfusion and/or exchange transfusion within 24 hours after birth and 82% received it within the first week of life. In three cases, the affected infants had survived in a transfusion-free manner until the 17^{th} day, the 28^{th} day and the 6^{th} week afterbirth (cases 57, 58 and 29, Table 3.3). However, all of the BHFS survivors in this group became transfusion-dependent. In the

group of the BHFS survivors who survived naturally until birth, almost all required blood transfusion and/or exchange transfusion within the first day of life and subsequently became transfusion-dependent (Table 3.2).

Other serious neonatal complications reported in the BHFS survivors who received intrauterine treatment include 1 case of respiratory distress syndrome (RDS) requiring surfactant therapy, 1 case of mild intraventricular hemorrhage (IVH), 2 cases of portal vein thrombosis which may have been related to umbilical catheterization, 1 case of bilateral pneumothorax and 3 cases of persistent pulmonary hypertension, of which 1 case required nitric oxide and high frequency ventilation (Table 3.3). Two affected infants who had survived naturally until birth and were born hydropic required abdominal and/or pleural paracentesis immediately after birth. Other serious complications noted in infants who survived without *in utero* treatment include 1 case who presented with convulsions, 1 case of cortical brain infarction, 1 case of pericardial effusion (non-hydropic), 2 cases of pulmonary hemorrhage and 1 case of persistent pulmonary hypertension requiring nitric oxide and high frequency ventilation (Table 3.2). Taken together these data suggest that intrauterine treatment results in a less stormy neonatal period.

3.2.5 Associated congenital abnormalities

A high frequency of congenital abnormalities has been observed in the BHFS survivors (a total of 9 defects in 30 cases, Table 3.2, 3.3). The most common congenital abnormalities found are urogenital abnormalities, which occur in 45% of the survivors in the registry. These urogenital defects occur exclusively in males and include hypospadias, which is the most common defect found among the male BHFS survivors. In fact, hypospadias alone was

present in 61% (19/31 cases) of the male survivors and considered together urogenital defects identified were reported in 71% (22/31 cases) of the male survivors. Other urogenital abnormalities reported include undescended testes, ambiguous genitalia, bifid scrotum, micropenis and hydrocele (Tables 3.2 and 3.3). This finding may be of significance in elucidating novel developmental processes underlying urogenital development. A number of reports have suggested that the occurrence of homozygous α-thalassemia (BHFS) together with hypospadias might be functionally linked (Abuelo et al., 1997; Dame et al., 1999a; Fung et al., 1999). The etiology of the hypospadias as well as the other disruptions of urogenital organ formation in these patients, however, is still unclear. It has previously been speculated that hypospadias in homozygous α-thalassemia might result from severe anemia, vascular occlusion or abnormal mechanical force in early gestation (Abuelo et al., 1997). However, a high incidence of hypospadias has not been observed in other conditions causing hydrops and fetal anemia, such as severe Rhesus hemolytic disease or non-immunological hydrops caused by other factors such as intrauterine infection. Moreover, there is also a high incidence of urogenital abnormalities in patients with ATR-X syndrome, and α-thalassemia metal retardation syndrome (ATR-16) (Dame et al., 1999a). Individuals affected with ATR-X syndrome have reduced transcription of the α-globin genes and have a clinical presentation similar to that of HbH disease. Since affected fetuses with ATR-X syndrome have lower levels of Hb Bart's (γ_4), this makes it less likely that intrauterine anemia and hypoxia causes disruption of urogenital organ formation in these patients. One possibility may be that α-globin is playing a role in a hitherto unknown developmental process. Two

studies have tried to identify a genetic determinant within the α-globin cluster (Chr16p13.3), which may be linked to urogenital organ formation (Dame et al., 1999a; Utsch et al., 2001). However, these approaches were not successful and the link between α-thalassemia and urogenital abnormalities remains an open question. Although urogenital defects occur in the majority of male BHFS survivors, most of these defects are correctable with minor operations.

Limb abnormalities of varying degrees of severity are the second most common congenital abnormality reported and are present in 14% (7/60 cases) of the survivors. These defects range from relatively mild deformities, such as asymmetrical hand size, to major abnormalities such as complete absence of the distal elements of the foot, that may remain as life-long co-morbidities (Table 3.2, 3.3). The majority of the limb abnormalities found in the BHFS survivors result from disruption of distal limb formation and these mostly occur unilaterally suggesting this is more likely to result from severe fetal anemia or vascular occlusion by abnormal erythrocytes rather than resulting from a patterning defect (Abuelo et al., 1997).

Atrial septal defect (ASD) was reported in 4 cases (8%), of which one case also has concurrent congenital stenosis of the pulmonary valve. The frequency of ASD in the general child population has been reported to vary widely between populations ranging from 1:2500 to 1:8500 live births (reviewed in (Seldon et al., 1962)). Although the frequency of ASD observed in the BHFS survivors is higher than that found in general population, the low numbers involved prevent direct association of homozygous α-thalassemia and ASD. Other congenital malformations reported in this study include microcephaly, jejunal atresia, patent ductus ateriosus (PDA) requiring ligation,

rib deformity and hip dysplasia (Table 3.2, 3.3), most of which can be managed or treated with minor surgery.

3.2.6 Growth

To assess growth, I have considered the patients as two separate groups: 1) Those who had survived naturally until birth; 2) Those who had received intrauterine treatment. There are growth data available in 17 out of a total of 27 cases who survived naturally until birth. Among these 17 cases, the latest growth records were made at age ≤ 2 years (infant to toddler) in 1 case, age ≥ 2 to 5 years (pre-school) in 3 cases and at age > 5 years (school age) in 13 cases. Of the 13 cases who have survived longest, 54% (7/13) have body weight (Wt) recorded as "normal" or > 10^{th} centile, 8% (1/13) of cases have Wt falling between the 3-10^{th} centile and 38% (5/13) of cases has Wt recorded as "severely retarded" or ≤ 3^{rd} centile (Figure 3.1A, Table 3.2). The height (Ht) in this group of patients mirrors the Wt data with 46% (6/13) having Ht recorded as "normal" or > 10^{th} centile, 15.5% (2/13) having Ht falling in between 3-10^{th} centile, and 38.5% (5/13) with Ht recorded as "severely retarded" or ≤ 3^{rd} centile (Figure 3.1B, Table 3.2). This finding indicates that among the long-term BHFS survivors (who did not receive intrauterine treatment) who have reached school age, nearly 40% experience unfavorable growth. Of the 3 BHFS survivors who had survived naturally until birth and whose latest growth records were given at age > 2 to 5 years, 1 case has Wt and Ht of > 10^{th} centile, 1 case has Wt of > 10^{th} centile and Ht falling in between 3-10^{th} centile and the remaining case has severe growth retardation (Wt and Ht). In a single case of the BHFS who had survived naturally until

birth, growth records were available as early as 10 months of age and in this case both Wt and Ht fall in between 3-10th centile (Figure 3.1A,B, Table 3.2).

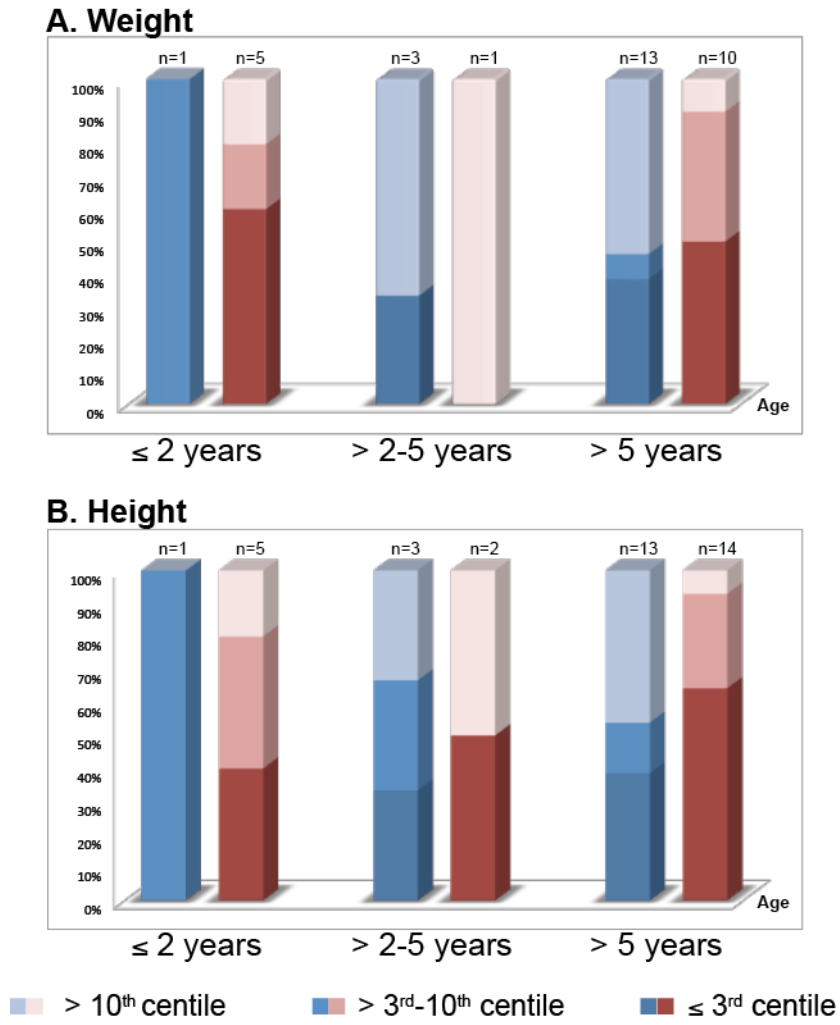


Figure 3.1 Weight (A) and Height (B) of the long-term BHFS survivors

Blue bars represent patients who had survived naturally until birth. Red bars represent patients who received intrauterine treatment. The X-axis indicates age at the time of report. Y-axis shows proportion of the BHFS survivors in percent (%). Color codes below the bar graphs indicate weight and height levels as described. Numbers of patients with available reports are shown on the top of each bar.

To explore further what might be influencing long-term growth of the BHFS survivors who did not receive transfusion *in utero*, several factors have been considered. Affected individuals in this group, who have long-term Wt and Ht of > 3rd centile or are documented as having “normal” growth (11 cases) were born significantly more prematurely (median GA at birth 30 weeks, p=0.048)

than those who experience severe growth retardation (6 cases, median GA at birth 33 weeks, p=0.048). This finding suggests that premature delivery allowing intensive extrauterine intervention and life support including blood transfusion, thereby preventing hypoxic organ damage *in utero*, may contribute to better long-term growth. At least 5 of the 6 survivors with severe growth retardation were documented as having stormy neonatal periods. They all required mechanical ventilation for a period of time ranging from 8 to 31 days. In addition, one case required CPR at birth (case 5, Table 3.2) and another case experienced septicemia and congestive heart failure during the first 3 months of life (case 4, Table 3.2). However, there are 6/11 BHFS survivors who were untransfused *in utero* who also have favorable long-term growth (Wt and Ht > 3rd centile) and who experienced as difficult a neonatal course as those found in the BHFS survivors with severe growth retardation (cases 1, 2, 3, 9, 12, 13 – Table 3.2). Two of these patients were transplanted early in life (at 19 and 20 months of age) and have become transfusion-independent, which might be the reason for their improved growth. I speculate that the transfusion regime is also critical for the long-term growth of these patients. The frequency of blood transfusion given to the patients varied from institution to institution and it is highly dependent on the local physicians, availability of blood components and available capacity of each center. In this registry, there are not enough reported details to propose what transfusion regimen might be suitable to promote optimal growth of the long-term BHFS survivors. However, a single institution in Thailand, which holds 4 of the transfusion-dependent survivors in this registry, reported significant improvement of weight and height after initiation of hypertransfusion (V.

Viprakasit, personal communication). These patients are transfused to maintain pre-transfusional functional hemoglobin (Hb A + Hb F) levels of 10 g/dl. By this means, 3 of the 4 patients have long-term "normal" growth (cases 18-20) and the remaining patient has improving growth (case 17, Table 3.2). Taken together these observations suggest that the BHFS survivors who require a longer period of neonatal critical care seem to have more severe long-term growth retardation. However, this situation may be improved by subsequent treatment regimens including optimal blood transfusions and hematopoietic stem cell (HSC) transplantation.

There are growth data (Wt and/or Ht) available in 21 out of 33 long-term BHFS survivors who were transfused *in utero*. Among these 21 cases, the latest growth records were made at age ≤ 2 years (infant to toddler) in 5 cases, at age ≥ 2 to 5 years (pre-school) in 2 cases and at age > 5 years (school age) in 14 cases. Of the 14 cases who have lived beyond 5 years of age, only 10% (1/10 cases) has "normal" or > 10^{th} centile body weight (Wt), 40% (4/10 cases) have Wt falling in between 3-10^{th} centile and 50% (5/10 cases) have Wt recorded as "severely retarded", "poor", or ≤ 3^{rd} centile (Figure 3.1A, Table 3.2). Considering height (Ht) in this group of patients, only 7% (1/14 cases) have "normal" or > 10^{th} centile Ht, 29% (4/14 cases) have Ht falling in between 3-10^{th} centile and 64% (9/14 cases) have Ht recorded as "severe retarded", "short stature", "poor" or ≤ 3^{rd} centile (Figure 3.1B, Table 3.2). Of the 2 BHFS survivors who were transfused *in utero* and whose latest growth records were given at age > 2 to 5 years, 1 case has "normal" growth while the other suffers from growth retardation (Wt or Ht ≤ 3^{rd} centile). Among 5 cases of the BHFS who received intrauterine treatment with latest growth

recorded at ≤ 2 years old, 3 cases experienced severe growth retardation (Wt and/or Ht ≤ 3rd centile) (Figure 3.1A,B, Table 3.2).

Although the frequency of the BHFS survivors suffering from growth retardation seems to be higher, at any given age group, in the patients who were transfused *in utero* compared to the patients who had survived naturally until birth, statistical significance was not observed. A number of factors may influence long-term growth of the *in utero* transfused survivors. However, the median GA at which intrauterine treatments were initiated was 25 weeks in the group with severe growth retardation (Wt and/or Ht ≤ 3rd centile) while in the less severely retarded group the GA was 26 weeks. The median GA at birth of individuals suffering from severe growth retardation was 35.5 weeks and 36 weeks in the less retarded group. This suggests that there is no correlation between onset of first transfusion or GA at birth with growth retardation. The proportion of the patients requiring mechanical or assisted ventilation during the neonatal period and the duration of ventilation needed are not significantly different between those suffering from severe growth retardation (62%) and those who have better long-term growth (71%), suggesting neither of these two factors are predictive of growth outcomes in survivors treated *in utero*. HSC transplantation was performed in 5 out of 33 BHFS survivors in this group, giving rise to transfusion-independent individuals. Poor growth, however, was not corrected in 4 of the 5 patients, even though transplantations were performed early in life (case 43 at age 18 months, case 50 at age 21 months, Table 3.3). This suggests that improved growth does not necessarily result from curative therapy. Endocrinologic investigations were reported in two BHFS survivors with short stature: one

was diagnosed as having idiopathic short stature that responded to growth hormone (GH) therapy, the second was diagnosed as having GH deficiency and is currently on GH replacement therapy.

In summary, growth retardation is a major long-term adverse outcome affecting a large number of the BHFS survivors, regardless of prenatal history and intrauterine treatment.

3.2.7 Neuropsychological development

Neuropsychological development is one of the major concerns for the BHFS survivors, since a previous review shows that affected fetuses are likely to have reduced brain weight relative to gestational age from ~8 weeks of gestation onwards (Higgs, 2009b). To assess long-term neurodevelopmental outcomes, the patients were divided into two groups according to whether they were previously treated *in utero*, and are evaluated according to age group at the time of report. The severity of neurodevelopmental delay is classified into four categories according to either evidence from commonly used tools for child developmental assessment or rating by referring physicians: 1) severe delay; 2) 6-12 months delay, this category also includes delay of 1-2 life skills and significant intellectual delay (+/-) requiring special education; 3) < 6 months or “mild” delay; 4) normal development.

There is documentation of neuropsychological development in 22 out of 27 patients who survived naturally until birth, of which 3 patients were documented at age of ≤ 2 (infant-toddler) and > 2-5 years (pre-school age). The majority (16/22 or 73%) of the neurodevelopmental reports of the survivors in this group were made at age > 5 years (school age) and as such represent “long-term” developmental outcomes. Sixty-three percent (10/16) of

those patients who survived beyond 5 years of age were assessed as having "normal" neuropsychological development, while 19% (3/16) were identified with only "mild" delay. Twelve percent (2/16) of the patients in this group have 6-12 months developmental delay whereas 6% (1/16) suffers from severely delayed neuropsychological development. The BHFS survivors whose age at the time of their last report is ≤ 5 years also have favorable outcomes. All 3 cases age ≤ 2 years experience only mild developmental delay and all 3 cases whose age is > 2-5 years are assessed as having "normal" neurodevelopment (Figure 3.2, Table 3.2). This finding indicates that even though affected infants had not been supported by transfusion *in utero*, surprisingly, the majority (59%, 13/22) of the survivors developed normally in the long-term.

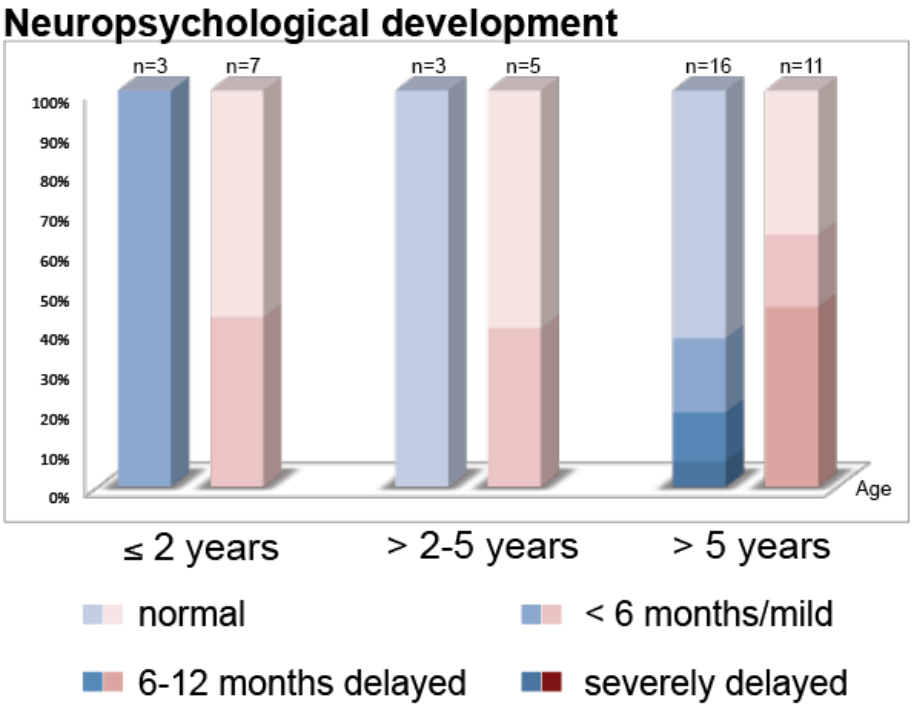


Figure 3.2 Long-term neuropsychological development of the BHFS survivors

Blue bars represent the patients who had survived naturally until birth. Red bars represent the patients who received intrauterine treatment. The X-axis indicates age at the time of report. Y-axis shows proportion of the BHFS survivors in percent (%). Color codes below the bar graphs indicate the level of neuropsychological development as described. Numbers of patients with available reports are shown on the top of each bar graph.

To further investigate possible factors contributing to developmental delay in patients who had survived naturally until birth, those severely delayed and with 6-12 months' delay are grouped and annotated as having "significant neurodevelopmental delay" and < 6 months/mild delay and "normal" are grouped as "normal/near normal" development. Median GA at birth of the survivors who suffer from significant neurodevelopmental delay is not significantly different from that of the survivors who have normal/near normal neurodevelopmental outcomes (33 weeks *vs* 32 weeks). This suggests that, unlike growth, favorable long-term neuropsychological outcomes are not associated with a more premature birth. Rather, long-term neurodevelopmental outcomes might be influenced by the neonatal course and illness occurring in early life. Although all survivors required mechanical or assisted ventilation for different durations during their neonatal period, in the delayed group there were critical events that might be associated with worse long-term outcomes. Of the 3 survivors who suffer from significant long-term neurodevelopmental delay, one developed convulsions during the neonatal period, which possibly lead to hypoxic brain injury (case 1, Table 3.2). The second survivor had pericardial effusion during the neonatal period and suffered from septicemia and heart failure during the first three months of life. This led to prolonged hospitalization and lack of stimulation, which are possible reasons for delayed development (case 4, Table 3.2). The third case remarkably remained without a diagnosis of BHFS until 1st year of age. This is a unique case, who was mildly anemic and non-hydropic at birth. However, he suffered from hypoxia during the neonatal period, was diagnosed as having recurrent pneumonia and heart failure and was admitted multiple times during

the first year of life. Since he was mildly anemic and was undiagnosed, blood transfusions were given infrequently and then only to treat neonatal jaundice. At one year of age, his growth was equal to a one-month old baby. Unsurprisingly, prolonged hospitalization and sub-optimal transfusion in early life led to significant long-term neurodevelopmental delay in this case (case 17, Table 3.2). This BHFS survivor is of particular interest and will be discussed in detail in Chapter 6. The survivors who have normal or near normal neurodevelopmental outcomes may have a critical neonatal period, (cases 3, 5, 9, 11, Table 3.2), however, those with severe delay are more likely to have a difficult neonatal period. Whether HSC transplantation early in life can improve neurodevelopmental outcomes of the patients is unclear. However, one survivor (case 5, Table 3.2), who was reported as suffering from developmental delay in early childhood, underwent HSC transplantation at 21 months of age and has now entered mainstream school with only mild learning difficulty.

Reports of neuropsychological development of the BHFS survivors who had been maintained with intrauterine treatment are available in 23 out of 33 cases. Since intrauterine interventions have been increasingly employed over the last decade, the survivors in this group are generally younger than those who survived naturally until birth. Forty-eight percent (11/23 cases) of the patients have survived beyond 5 years of age (school age) at the time of their report, 22% (5/23 cases) are > 2-5 years of age (pre-school age), and 30% (7/23 cases) are ≤ 2 years old (infant-toddler). Similar numbers of the patients who have survived beyond 5 years were reported to have “normal” (4/11 cases, 36.5%) and 6-12 months’ neurodevelopmental delay (5/11 cases,

45.5%). The two remaining patients in this group were reported as having "mild delay" (2/11 cases, 18%) (Figure 3.2, Table 3.3). Of the 5 survivors whose neurodevelopmental outcomes were noted at the age of >2-5 years, 3 were documented as having "normal" development (60%), while the other 2 cases have only "mild" developmental delay (40%). Among those survivors who were in the infant-toddler age group at the time of report, 57% (4/7 cases) were noted to have "normal" development and the remaining 43% (3/7 cases) experienced only "mild" neurodevelopmental delay (Figure 3.2, Table 3.3). None of the affected individuals who had received intrauterine treatment suffers from severely delayed neuropsychological development.

This finding shows that the vast majority (78%) of the survivors who had received intrauterine interventions have favorable long-term neurodevelopmental outcomes (either "normal", 48%, or "mild" neurodevelopmental delay, 30%). The proportions of affected individuals with normal long-term neurodevelopmental outcomes are not significantly different between the patients who had survived naturally until birth and those who had been maintained with intrauterine intervention (59% *vs* 48%). This suggests that, once patients survive gestation and the postnatal period, *in utero* treatment is not associated with improved neurodevelopmental outcome and these two groups of survivors are likely to have a similar chance to develop normally in the long-term.

To further investigate other factors associated with neurodevelopmental outcomes in the survivor group with intrauterine treatment, several parameters were assessed. Median GA at 1st intrauterine intervention as well as median GA at birth is not significantly different between those survivors

who had “normal or near normal development” and those who suffer from “significant developmental delay” (26 weeks *vs* 24 weeks and 36 weeks *vs* 34 weeks). Requirement for mechanical or assisted ventilation during the neonatal period was reported in 4 out of 5 survivors suffering from long-term neurodevelopmental delay, whereas 64% (9/14 cases) of the survivors who have normal or near normal neurological development were noted to have a similar requirement (p=non-significant). Considering the survivors who had received intrauterine treatment, it is still unclear whether critical neonatal events and early infant illness are predictive of poor long-term neuropsychological outcomes. Only 3 out of the 5 survivors with significant neurodevelopmental delay seem to suffer from more severe neonatal events or prolonged hospitalization in early life (cases 38, 50, 60, Table 3.3), while some affected infants with similar events eventually develop normally or have only mild neurodevelopmental delay (cases 40, 52, 53, Table 3.3). One report of a survivor who underwent prolonged hospitalization during the first 1-2 years of life and who was transplanted at the age of 21 months suffered continuing significant neurodevelopmental delay (case 50, Table 3.3). This suggests that neurodevelopmental delay may not always be recoverable by early curative therapy.

Learning difficulties and below-average intellectual levels are the most frequently reported aspects of developmental delay, accounting for at least 48% of reported neurodevelopmental delay in the survivors. Other types of less frequently reported delay include gross motor delay, speech, hearing and language problems.

In conclusion, the majority of the BHFS survivors, both in the intrauterine treated and the non-treated groups demonstrate surprisingly favorable long-term neurodevelopmental outcomes. Once an individual has survived gestation, intrauterine treatment is not predictive of favorable long-term neuropsychological development. The survivors who experience prolonged and more severe critical neonatal periods are at risk of poor developmental outcomes. However, a number of those affected infants eventually develop normally in the long-term.

3.2.8 Current treatments

Although many of the patients have survived because of prenatal and immediate postnatal blood transfusion, the vast majority of these patients have the burden of life-long transfusion dependence. Curative HSC transplantation is mainly restricted by availability in each center, availability of donors and limitations of funding, especially in less developed countries.

In the current review, only 17% (10/60 cases) of the survivors underwent postnatal HSC transplantation and have gone on to become transfusion-independent. In most cases, transplants were performed early in life, ranging from 18-44 months of age. Consequently, these patients had a lower risk of chronic transfusion-related complications, such as iron overload. HSCs obtained for transplantation were derived from various sources including Human Leukocyte Antigen (HLA) -matched related bone marrow (cases 5, 13 - Table 3.2 and 39 -Table 3.3), HLA-mismatched related cord blood (case 9 – Table 3.2), HLA-matched unrelated bone marrow (case 37 – Table 3.3), HLA-matched unrelated peripheral blood (cases 43 and 50 – Table 3.3), and HLA-mismatched (sharing 5 out of 6 markers tested) unrelated cord blood (case 12

– Table 3.2). Acute skin (case 12, 21 – Table 3.2) and gut (case 21 – Table 3.2) graft-versus-host disease (GVHD) was noted in two cases that had been transplanted with HLA-matched (case 21 – Table 3.2) and mismatched (case 12 – Table 3.2) unrelated stem cells. Hearing loss in one ear (resulting from chemotherapeutic agents used for the conditioning regimen) and vocal cord paralysis (resulting from intubation required during the stem cell transplantation process) requiring speech therapy were reported as chronic consequences of HSC transplantation (case 12 – Table 3.2). HSC transplantations had been offered, but were unfortunately unsuccessful because of graft rejection in an additional 3 survivors. These patients are currently transfusion-dependent.

A total of 83% (50/60 cases) of the BHFS survivors have survived to date in a transfusion-dependent manner and all except one case (case 17 – Table 3.2, see details in Chapter 6) have been regularly transfused starting from a few days to a few weeks after birth. Frequency of transfusions given to the patients varies from 2 to 5 weeks depending on the judgment of local hematologists and availability of blood components. The adequacy of the transfusion given is often monitored using pre-transfusional total hemoglobin level. Similar to transfusion therapy for individuals with β-thalassemia major, the BHFS survivors are frequently transfused to maintain a pre-transfusional hemoglobin level of 8.5-10 g/dl. In some cases, functional hemoglobin of 9-10 g/dl, calculated by (HbA% ($\alpha_2\beta_2$) + HbF% ($\alpha_2\gamma_2$) x total hemoglobin level), is used as a threshold for transfusion instead of total hemoglobin level alone (V. Viprakasit and A. Amid, personal communication). As a result, the BHFS survivors whose transfusions are monitored using functional hemoglobin

levels tend to require more frequent transfusion (up to every 2 weeks) compared to those whose transfusions are monitored using total hemoglobin levels. However, the BHFS survivors whose transfusions were determined by functional hemoglobin were reported as having improved growth (V. Viprakasit, personal communication). Other laboratory parameters employed for blood transfusion monitoring of the survivors are reticulocyte counts and HbH (β_4) levels. It has been observed in a single institute, where survivors were transfused according to total hemoglobin level, that the patients had increased reticulocyte counts as well as hepatomegaly and splenomegaly over the course of their treatment (A. Amid, personal communication). This suggests that hemolytic anemia resulting in extramedullary hematopoiesis is the more prominent pathophysiology in survivors rather than ineffective erythropoiesis, reflecting the situation in HbH disease, and that these survivors require hypertransfusion to suppress both stress erythropoiesis from bone marrow and extramedullary erythropoiesis from other organs. In addition, the BHFS survivors who have been treated in this institute demonstrate very high pre-transfusional HbH levels (25 to 60% of total Hb component) as well as extremely high EPO levels, which are significantly higher than those observed in β-thalassemia major patients with a lower pre-trasfusional hemoglobin level. This suggests that inadequate transfusion of functional-hemoglobin-containing red blood cells over many years may give rise to accumulation of non-functional HbH, resulting in tissue hypoxia and an extremely high EPO level. In an attempt to solve these problems, occasional exchange transfusions with an aim of removing non-functional HbH, followed

by hypertransfusions have been employed to treat transfusion-dependent BHFS survivors in this institute.

To prevent detrimental iron overload, termed hemochromatosis, chelation therapy has been provided to the regularly transfused BHFS survivors. Iron overload is typically monitored using serum ferritin levels and when the level reaches ~1,000 ng/mL chelation therapy is administered. Severe iron overload has been reported in some of the older patients, however, observations in the Hospital for Sick Children, Toronto, Canada, which is currently taking care of 7 BHFS survivors, suggest that individuals with the BHFS have a significantly lower ferritin/liver iron concentration (LIC) ratio, suggesting that serum ferritin may not be an accurate measure of iron overload in these patients. This may be because of differences in the pathophysiology between the BHFS and β-thalassemia major, in fact serum ferritin in the BHFS may be more similar to the pattern observed in hemolytic disease. This indicates that serum ferritin alone might under estimate the degree of total body iron accumulation in the BHFS survivors.

Having considered these observations, transfusion and chelation management of the BHFS may be significantly different than that of β-thalassemia major. These observations suggest that functional hemoglobin should be monitored in the BHFS survivors and chelation should be tailored to each patient using both routinely performed serum ferritin levels and more accurate body iron assessments, such as LIC or MRI T2*.

3.2.9 Other Comorbidities of the long-term BHFS survivors

As in other cases of hemolytic anemia, splenomegaly of varying degrees resulting from extramedullary erythropoiesis may be expected in BHFS

survivors. The size of the spleen in these patients is also dependent on adequacy of blood transfusion to suppress this process. In this review, significant splenic enlargement of up to 19 cm below the left costal margin (case 53) was observed. Significant splenomegaly was reported to cause secondary thrombocytopenia in at least one case (case 38). There is currently no evidence regarding whether splenectomy might improve this condition in the BHFS survivors.

Almost all of the congenital defects reported involving the genitourinary system (such as hypospadias and undescended testes), the gastrointestinal system (such as jejunal atresia) and the cardiovascular system (such as ASD and PDA) had been surgically repaired during the neonatal period, early or late childhood. Only congenital limb defects remain permanent co-morbidities of the patients reported here. Hypothyroidism and hyperthyroidism (Graves' disease, status post thyroid ablation) are reported in 2 (cases 40 and 43, Table 3.3) and 1 case (case 2, Table 3.2) respectively. Whether these thyroid conditions are directly associated with the BHFS is unknown. Other co-morbidities found in the BHFS survivors show no consistent association and include asthma, tricuspid valve dysplasia, premature ventricular contraction, migraine, and idiopathic urticaria.

3.3 Discussion

To date, this is the largest review describing the natural history and long-term clinical outcomes of the BHFS survivors and includes total of 60 published and unpublished survivors worldwide. This study aimed to answer whether the BHFS should ever be considered a salvageable disease and to provide guides for treatment of these survivors and on counseling of couples

expecting affected infants. That at least 60 of the affected individuals with the BHFS have survived in long-term indicates that improvement of prenatal diagnosis, intrauterine therapy and neonatal intensive care over the past 2-3 decades are capable of maintaining the patients' lives and that this condition may no longer be considered a universally fatal disorder.

Although embryos with homozygosity for deletions that span ζ-globin and α-globin die at a very early stage of gestation it is interesting that a single ζ-globin gene can confer viability. The registry contains a single individual who is a compound heterozygote for the SEA and Filipino deletions (haplotypes $--^{SEA}/--^{Fil}$) who became a long-term BHFS survivor (case 58 – Table 3.3).

Less invasive methods for early detection of hydropic fetuses, such as accurate ultrasound measurement of placental thickness (reflecting accumulation of fluid in the placenta), which can detect hydropic features as early as the 1st trimester have been developed (Ghosh et al., 1994; Ko et al., 1995). The findings in this work suggest that couples at risk of having affected fetuses with the BHFS should participate in antenatal care as early in gestation as possible. This will provide time for couples to decide on whether to continue pregnancy after counseling and also allow obstetricians the opportunity for early intervention to treat developing fetuses. In this registry, only the BHFS patients who have survived through the neonatal period are included. Therefore, it is important to keep in mind that the majority of affected fetuses without any intrauterine therapy wouldn't have survived gestation at all or would have died soon after birth and the proportion of these patients varies widely on capability of obstetric and neonatal intensive care of each center. I would like to recommend that intrauterine treatment be initiated in every case

where the couple concerned has decided to continue pregnancy, to make sure the maximum GA at birth is achieved. Intrauterine therapy can be performed using various methods, including red cell transfusion, exchange transfusion with or without subsequent red cell transfusion and intrauterine HSC transplantation. This review finds that intrauterine transfusion alone is the most commonly employed intrauterine therapy for affected fetuses with the BHFS. Intrauterine exchange transfusion seems to be of benefit for affected fetuses who have fetal ascites as a predominant symptom (Bizzarro et al., 2003; Carr et al., 1995; Fung et al., 1998) or have demonstrated poor cardiac function (Fung et al., 1998), suggesting fetal volume overload should be strictly avoided. However, as there is an increased risk of intrauterine infection with exchange transfusion when compared to intrauterine transfusion alone, this risk should be considered prior to the procedure. *In utero* HSC transplantation for the BHFS, so far has not been successful, this may be because the most appropriate sources of HSCs to provide *in utero* engraftment have not yet been identified and also the optimum timing to conduct the procedure during gestation is yet to be extensively studied.

When considering whether to continue a BHFS pregnancy, maternal complications should also be taken into account. Mothers carrying hydropic fetuses have increased risks of serious obstetric complications, including preterm delivery, poly/oligohydramnios, pre-eclampsia and malpresentation, all of which can occur regardless of whether intrauterine therapies are given to the fetus and often lead to caesarean section or assisted vaginal delivery. Although the frequencies of most of these maternal complications are much lower in this review compared to those reported in the previous literature

(reviewed in (Higgs, 2009b)), suggesting significant improvement of obstetrical care, it remains important for the mothers to have frequent follow-up visits at well-equipped antenatal care clinics and that early signs and symptoms of these complications are communicated to the mothers throughout pregnancy. However, providing affected fetuses with intrauterine treatments prolongs the course of pregnancy, thus decreasing the frequency of preterm delivery, which is the most common obstetrical complication found in mothers carrying fetuses affected with the BHFS.

Given that affected fetuses with the BHFS who had not received intrauterine therapies are expected to suffer from a more severe degree of *in utero* anemia, hypoxia and congestive heart failure, it is interesting that a considerable number of affected fetuses with the BHFS who were not transfused *in utero* (48%) were born without hydropic features. The factors underlying the development of the range of hydropic abnormalities are still unknown, however, one possible hypothesis may be that the affected infants without hydrops at birth might have a degree of persistent expression of embryonic ζ-globin chains. However, to understand whether this is the case further detailed study is required. Because affected infants with the BHFS who died soon after birth are not routinely reported, it is not possible to suggest the proportion of survivors who completed gestation without intrauterine intervention and who later survived because of neonatal resuscitation and intensive care. The proportion of this group of the patients who have survived is also expected to vary from place to place, depending on the neonatal intensive care capability of each center. Nevertheless, this review suggests that at least 27 of such cases, some of which had very low Apgar scores or

even required CPR at birth, have made it through the neonatal period with modern ventilatory support in neonatal intensive care and survived in the long-term. Intrauterine therapies have benefitted the BHFS survivors during their perinatal and neonatal courses. Firstly, the vast majority (92%) of those affected infants who received transfusion *in utero* were born with a favorable BW of ≥ 10th centile. Secondly, these affected infants had significantly better Apgar scores at 1 and 5 minutes after birth as compared to those of the BHFS survivors without intrauterine treatment, suggesting that they perform better at birth and are more likely to respond to initial resuscitation. Thirdly, although the proportion of the survivors who needed mechanical or assisted ventilation is not significantly different between the intrauterine treatment and the no intrauterine treatment groups, the former group tends to have less stormy neonatal periods. This is demonstrated by significantly less median time required for mechanical or assisted ventilation (3 days *vs* 19 days) in neonatal intensive care.

One of the concerns for the BHFS survivors is a high incidence of associated congenital abnormalities. In this review, congenital defects were found to be present in 50% of the survivors and this prevalence is considerably higher than the 17% previously reported (reviewed in (Higgs, 2009b)). The two most common congenital defects are urogenital abnormalities and limb defects of varying degrees. Such a high frequency of urogenital organ formation defects (hypospadias being the most common), occurring up to 71% of the male survivors, is one of the most interesting findings in this study. This supports the hypothesis that the co-occurrence of homozygous α^0-thalassemia and urogenital defects are likely to be linked (Dame et al., 1999a; Fung et al.,

1999) and that these defects are unlikely to occur solely as a result of chronic hypoxia *in utero*. Formation of the normal male urethra is typically completed by ~12 weeks gestation, the period by which the remaining Hb Gower I ($\zeta_2\varepsilon_2$) and Hb Portland I ($\zeta_2\gamma_2$) are likely to be adequate for oxygen delivery to regions of uncompromised blood flow. Therefore, there may be a hitherto unknown developmentally important determinant present within the α-globin locus and further studies are required to understand the mechanism underlying this condition. The vast majority of these frequently found urogenital defects in the BHFS survivors, however, are operable in a large number of general hospitals and have seldom contributed to the long-term co-morbidities of the survivors.

Unlike the associated urogenital abnormalities, limb defects identified in the BHFS survivors are more likely to result from *in utero* hypoxia. Although limb formation occurs around the same period (~6 to 8 weeks of gestation, when some ζ-globin chains are likely to remain) as urogenital organ formation, blood supply is often compromised to distal limb regions leading to higher chance of developmental disruption. This hypothesis is supported by the fact that the limb defects found in the BHFS survivors are mostly unilateral, which is suggestive of secondary defects resulting from severe fetal hypoxia. These distal limb defects are often not correctable and do contribute to long-term co-morbidities of the BHFS survivors. Because limb development occurs at ~6-8 weeks of gestation (Abuelo et al., 1997; Chitayat et al., 1997; Lam et al., 1997), when currently available technologies are incapable of diagnosis of the BHFS and the affected fetuses are too small for intervention, it is almost impossible that these limb defects can be prevented. However, with advances

in prenatal ultrasonographic assessment, limb defects can be detected in routine ultrasound and should be taken in to account for counseling couples expecting affected babies.

This review indicates that growth retardation is a major adverse long-term outcome for the BHFS survivors as up to 50% (19/38 cases) suffer from severe growth retardation (Wt and/or Ht ≤ 3rd centile) in any given age group. Having received intrauterine therapies does not seem to provide additional benefits to the BHFS survivors in terms of their long-term growth, as compared to those survivors who had not been transfused *in utero*. Among the survivors without intrauterine treatment, those who were born more prematurely, allowing earlier extrauterine transfusion and intensive care, are more likely to have better growth in long-term. Long-term growth of the BHFS survivors depends on multiple factors. Prolonged critical neonatal periods and chronic illness during early life can put affected infants with the BHFS, as well as infants with other disorders, at risk of poor growth. For transfusion-dependent BHFS survivors, optimal blood transfusion is extremely likely to play an important role in improvement of growth. Thus, optimal transfusion regimens should be considered and I would highly recommend growth monitoring in all BHFS survivors.

Surprisingly, favorable long-term neurodevelopmental outcomes among the BHFS survivors are the most encouraging findings in this registry. Significant long-term neuropsychological developmental delay (including both the severely delayed and the 6-12 months' delayed groups) has been reported in only 18% of all the BHFS survivors at any given age. This is an unexpected finding, given that affected individuals with the BHFS have suffered from fetal

anemia and hypoxia from early gestation and brain development is likely to have been affected. The most common aspects of developmental delay are learning difficulty and below-average intellectual levels. The current review also demonstrates that the proportion of affected individuals with normal long-term neurodevelopmental outcomes is not significantly different between the BHFS survivors with and without prior intrauterine treatment. This suggests that once affected infants with the BHFS have survived gestation and postnatal periods, they all have a similar chance to develop normally regardless of prenatal history of intrauterine treatment. This finding is inconsistent with a previous study of 11 BHFS survivors without intrauterine treatment, which suggested that almost half of the patients (5/11) suffered from neurological deficit (Lee et al., 2007).

Although BHFS survivors often have favorable neurodevelopmental outcomes, the majority of the patients (83%) have the burden of life-long transfusion dependence. This also leads to transfusion-related complications, especially iron overload requiring chelation, similar to that found in transfusion-dependent β-thalassemia major. However, this review suggests that transfusion management and iron overload monitoring in the BHFS survivors differs from management of β-thalassemia major. Monitoring of pre-transfusional functional hemoglobin, ((HbA + Hb F) x total Hb level), and hypertransfusion should be considered, as they tend to promote optimal growth in long-term survivors of the BHFS. Both routine serum ferritin level and more accurate body iron assessments, such as LIC and MRI T2*, are likely to be of benefit to the survivors.

In summary, maintaining affected fetuses with the BHFS should not be routinely recommended to couples. Instead, extensive discussion with the family on the long-term outcomes of the affected babies and maternal risks should be undertaken. However, if a couple decides to continue pregnancy, intrauterine therapy is recommended at least for viability of the affected fetus. Although affected individuals with the BHFS, who have been kept alive with intrauterine transfusion and/or immediate postnatal transfusion, are mostly transfusion-dependent, the majority of them develop normally in long-term (at least up to their 20's), thus in my opinion the BHFS should be considered a salvageable condition. Increasing numbers of reports of long-term BHFS survivors with favorable outcomes also draw attention to the area of developing novel therapies. Since almost all of the BHFS survivors are homozygous for the SEA deletion, which deletes both α-globin paralogs but leaves the embryonically expressed ζ-globin gene intact, it is theoretically possible to reactivate the ζ-globin gene in definitive erythroid cells. This is likely to ameliorate the clinical symptoms of patients with the BHFS, as well as other severe α-thalassemias. However, the regulation of ζ-globin expression is poorly understood. A major aim of this thesis is to identify regulation of the ζ-globin gene both in *cis*- and *trans*- and results are discussed in the following chapters.

Chapter 4 : *Cis*-acting regulation of ζ-globin expression

4.1 Introduction

The *cis*-acting regulatory landscape of the α-globin locus has been extensively studied in HSC-derived definitive erythroblasts and the critical sites together with their associated epigenetic marks have been well defined (De Gobbi et al., 2007; Higgs and Wood, 2008b). However, the conformation of this locus in primitive erythroblasts, when ζ-globin is actively transcribed, has not previously been studied. In humans, the expression of the α-globin genes during definitive erythropoiesis is regulated in *cis* by four conserved, non-coding regulatory sequences, known as Multispecies Conserved Sequences (MCSs) (Hughes et al., 2005). These MCSs correspond to DNaseI hypersensitive sites (DHSs) and they are identified according to their location in kilobases (kb) relative to the ζ-globin mRNA cap site: HS-48 (MCS-R1), HS-40 (MCS-R2), HS-33 (MCS-R3) and HS-10 (MCS-R4). In the mouse, the orthologous elements are termed HS-31 (MCS-R1), HS-26 (MCS-R2), HS-21 (MCS-R3) and HS-8 (MCS-R4). In addition there is a non-conserved mouse-specific *cis*-regulatory element termed HS-12 (Hughes et al., 2005; Kielman et al., 1996)(Figure 1.3, Chapter 1). Three of the regulatory elements, MCS-R1-R3, are in introns of an adjacent gene *Nprl3*. Previous studies (see details in Chapter 1, Section 1.2.1) have suggested that MCS-R2 is the critical element contributing to α-globin transcription in definitive erythroblasts. In addition, a number of mouse lines harboring deletions of individual

hypersensitive sites have recently been generated in my host department. The contribution of each upstream *cis*-regulatory element to α-globin expression has been extensively studied and only combined homozygous deletion of both HS-31 and HS-26 in tandem was found to cause severe anemia leading to embryonic lethality at E13.5. This suggests that α-globin expression is predominantly co-regulated by HS-31 (MCS-R1) and HS-26 (MCS-R2) in HSC-derived definitive erythroid cells while the other elements make smaller positive contributions to α-globin transcription (Hay D, *et al*., manuscript in review).

The presence of open chromatin at the upstream *cis*-regulatory elements and the contribution of each individual element on ζ- and α-globin gene expression in primitive erythroid cells have not previously been assayed. However, several lines of evidence suggest that the ζ-globin gene might be less reliant on distal *cis*-acting regulatory elements than the α-globin genes. Firstly, the ζ-globin gene promoter is larger and more complex than those of the α-globin genes (see details in Chapter 1, Section 1.5) suggesting that a greater number of *trans*-acting factors bind this region. Secondly, a mouse mutant homozygous for a targeted deletion of three critical α-globin regulatory elements (MCSR1-R3) and the intervening genetic material (termed the smallest natural deletion or SND allele) die during gestation at ~E14.5, shortly after the ζ- to α-globin switch normally occurs (D. Higgs, personal communication). The survival of these mice to this stage suggests that they are capable both of expression and repression of ζ-globin in the absence of the α-globin regulatory elements. Thirdly, in transgenic mice, the human ζ-globin gene, when attached to the HS-40 (MCS-R2) element shows

developmental specificity as it is expressed almost exclusively at the embryonic stage (Vyas et al., 1992). This suggests that the ζ-globin promoter and the HS-40 (MCS-R2) element together contain all the *cis*-acting regulatory elements required to achieve developmental stage-specific expression and silencing. Lastly, induced pluripotent stem cells (iPSCs) derived from patients with complete loss of the α-globin, but not ζ-globin genes, and trans-differentiated into erythroblasts suggests that ζ-globin is transcribed and repressed in a similar fashion to control cells indicating an independent (non-competitive) means of repression (Chang and Bouhassira, 2012).

In this Chapter, I have characterized the *cis*-regulatory landscape of the ζ-globin gene in murine primitive erythroid cells using DNase I hypersensitivity. In addition, I have investigated the interaction of the previously identified distal *cis*-regulatory elements with the ζ-globin promoter, as well as the contribution of the individual *cis*-regulatory elements to ζ-globin expression during primitive erythropoiesis.

4.2 Results

4.2.1 Validation of ζ-globin expressing primitive erythroid cells

To obtain an appropriate source of primitive erythroid cells for the study of *cis*- and *trans*-regulation of ζ-globin expression, two factors have been considered: 1) Primitive erythroid cells should be at the stage when ζ-globin transcripts are most abundant. 2) Sufficient numbers of primitive erythroid cells should be available to permit analysis of DNaseI hypersensitivity (~30 million cells).

It has been demonstrated, in a previous study in mouse, that all three α-like genes are expressed in primitive erythroblasts from their emergence at E7.5 in the yolk-sac blood islands. By E8.5, ζ-globin transcripts account for ~40% of accumulated total α-like (α+ζ) globin transcripts and this proportion remains approximately the same at E10.5 (Kingsley et al., 2006). Silencing of ζ-globin occurs between E10.5 to E12.5, when ζ-globin transcripts account for ~10% of total α-like (α+ζ) globin transcripts (Kingsley et al., 2006). I assayed ζ-globin expression, using quantitative PCR (qPCR), in primary primitive erythroblasts derived from the embryonic yolk sac at E10.5, E11.5 and E12.5. I found primitive erythroblasts derived from E10.5 and E11.5 yolk sacs to contain the maximum proportion of ζ-globin accumulated transcripts (accounting for ~ 40% of total α-like (α+ζ) globin transcripts) (Figure 4.1A). However, only primitive erythroblasts derived from yolk sacs at E10.5 were selected for downstream experiments and analyses in this study, as they are more likely to be at the stage when the ζ-globin gene is transcriptionally active. A total of ~ 1 x 10^6 cells was obtained from each individual E10.5 yolk sac. I also performed Wright's stain of primitive erythroblasts at this stage to assess morphology. Because primitive erythroblasts mature in semi-synchronous manner, the morphology of the cells is highly homogeneous at E10.5 with the vast majority of cells being at the intermediate erythroblast stage of differentiation (mostly polychromatic erythroblasts, Figure 4.1B). In addition, differential counts of E10.5 yolk sac derived cells demonstrated that ~95% are primitive erythroid cells, while the remainder is composed mainly of macrophages, which are unlikely to confound downstream analysis of chromatin accessibility and conformation.

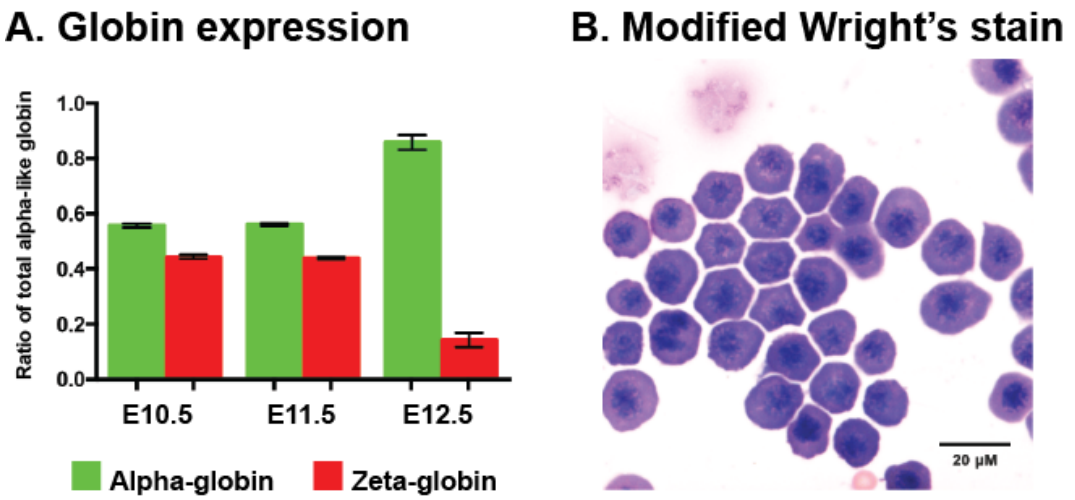


Figure 4.1 Alpha-like globin expression of primitive erythroblasts and morphology of primitive erythroblasts at embryonic day 10.5 (E10.5)

A. qPCR analysis shows α-like globin expression of primitive erythroblasts obtained from murine yolk sacs at different developmental stages. The green bars represent α-globin expression and the red bars represent ζ-globin expression. Expression levels were normalized to *Pabpc1* and are shown as ratios of total α-like globin expression (Y-axis). Error bars represent standard deviations between two biological replicates each repeated 3 times. B. Modified Wright's stain shows a homogenous population of intermediate (polychromatic) erythroblasts obtained from E10.5 yolk sacs.

4.2.2 Identification of the cis-regulatory landscape of the ζ-globin gene in primitive erythroid cells using DNaseI/ATAC-seq

Regions in the genome specifically bound by transcription factors, as identified by ChIP analysis, often correlate with active gene promoters and enhancers and areas of open chromatin. These open chromatin regions are hypersensitive to enzymes, such as DNaseI and Tn5 Transposase (Buenrostro et al., 2013). This sensitivity can be exploited to enable the large-scale isolation of fragments that are then subjected to massively parallel sequencing and alignment with the relevant genome assembly.

I initially characterized areas of open chromatin in the α-globin locus in primitive erythroblasts derived from the yolk sacs of E10.5 mouse embryos using a DNaseI hypersensitivity assay coupled with high-throughput sequencing. I used ~30 million cells per assay and the chromatin was exposed to 8 different concentrations of DNaseI (see methods in Chapter 2).

Digested chromatin was analysed using 1% agarose gels (Figure 4.2). To determine the amount of enzyme required, a key determinant of success in this protocol, each digested chromatin sample was digested with a range of amounts of DNaseI enzyme including 0, 5, 10, 20, 40, 60, 80, and 100 units. DNaseI digestion with 20, 40, 60, and 80, units represented the most appropriate digestion of the chromatin, just before the entire chromatin was extensively digested (Figure 4.2). These digestions were selected for library preparation and sequencing. The reads from the library generated using chromatin from the 80 units digestion obtained from massively parallel sequencing (Illumina Miseq) contained the most informative reads. Bioinformatic analysis was performed using custom scripts generated by M. Suciu.

At this stage, the ζ-globin gene is marked by a region of increased DNaseI hypersensitivity, which taken together with the expression data suggests that it is actively transcribed at this stage (Figure 4.3). All five previously characterized *cis*-regulatory elements are DNaseI sensitive and no novel sites gained sensitivity in this region. This indicates the activity of the same enhancers upstream of the α-globin locus in primitive and definitive erythroid cells (Figure 4.3). Provocatively, the DNaseI peak co-incident with ζ-globin extends ~300 bp 3′ of the ζ-globin 3′UTR, and this is investigated further in Section 4.2.5.

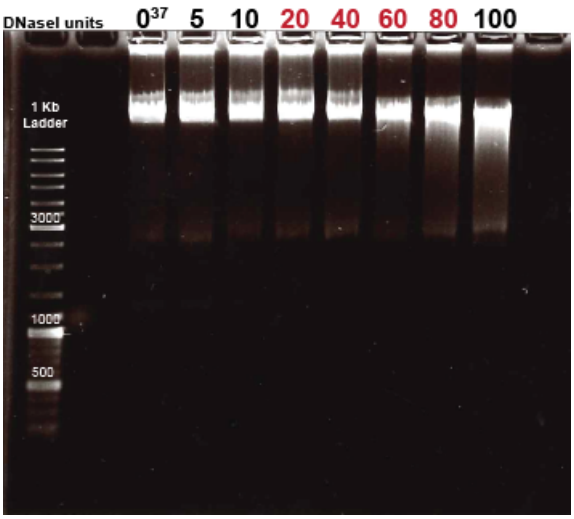


Figure 4.2 Selection of DNaseI digestions

The figure shows chromatin obtained from primitive erythroblasts (n=1) digested with different concentrations of DNaseI visualized on a 1% agarose gel. The left lane contains a 1 kb ladder. The numbers in white indicate the molecular weight of bands in base pairs. The numbers above each lane indicate the amount of DNaseI. "0^{37}" represents the negative control, to which DNaseI was not added. Increasing amounts of DNaseI enzyme were added to each chromatin sample (see methods in Chapter 2). The level of degradation seen in Lanes 20, 40, 60 and 80 has previously been empirically determined to yield optimum results. Therefore the chromatin shown in these lanes were selected for library preparation and Hi-throughput sequencing.

A novel technique for the identifying open chromatin sites has recently been developed, this is termed the Assay for Transposase-Accessible Chromatin (ATAC)(Buenrostro et al., 2013). This technique is based on direct *in vitro* transposition of sequencing adaptors into native chromatin (see methods in Chapter 2). One advantage of this assay is that it can be performed using only 50,000 cells per assay. Thus, this method is much more efficient both in terms of experimental effort and input materials than the DNaseI hypersensitivity assay. I performed ATAC-seq using three biological replicates of E10.5 yolk sac-derived primitive erythroid cells and sequenced adaptor ligated DNA fragments using the Illumina NextSeq platform. The data were aligned to the reference genome in the same way as the DNaseI HS data detailed above. The HSs identified are reproducible within the three biological replicates and

the data derived from ATAC-seq are highly comparable to those derived from DNaseI-seq (Figure 4.4).

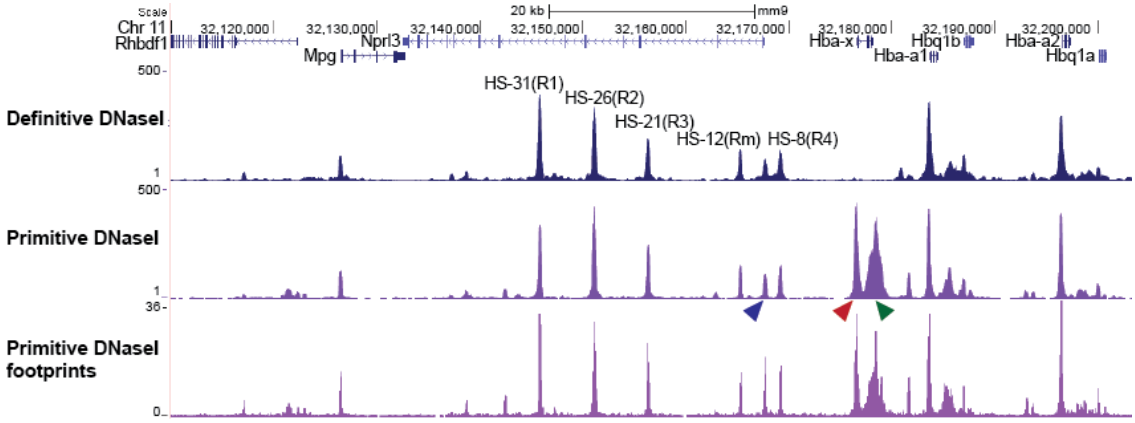


Figure 4.3 DNaseI tracks

The figure shows region of conserved synteny of the mouse α-globin cluster. The upper track shows analysis of DNaseI hypersensitive sites (HS) in phenylhydrazine-treated-spleen-derived definitive erythroid cells (n=1) given by Maria Suciu, coinciding with active promoters and the well-characterized distal enhancers present in the α-globin locus (R1-4 and the mouse-specific enhancer Rm). The distal enhancers are numbered according to their distance in kb 5′ of the ζ-globin (Hba-x) promoter. The middle track shows DNaseI HSs in primitive erythroid cells at embryonic day 10.5 (E10.5) (n=1). The lower track shows primitive erythroid DNaseI footprints identified using the Illumina HiSeq platform (n=1, see expanded view of footprints at the ζ-globin gene in Figure 4.9). The red arrow indicates the active ζ-globin promoter. The green arrow indicates a significant DNaseI peak identified at the 3′ end of the ζ-globin gene. The HS corresponding to the *Nprl3* gene promoter is noted with the blue arrow.

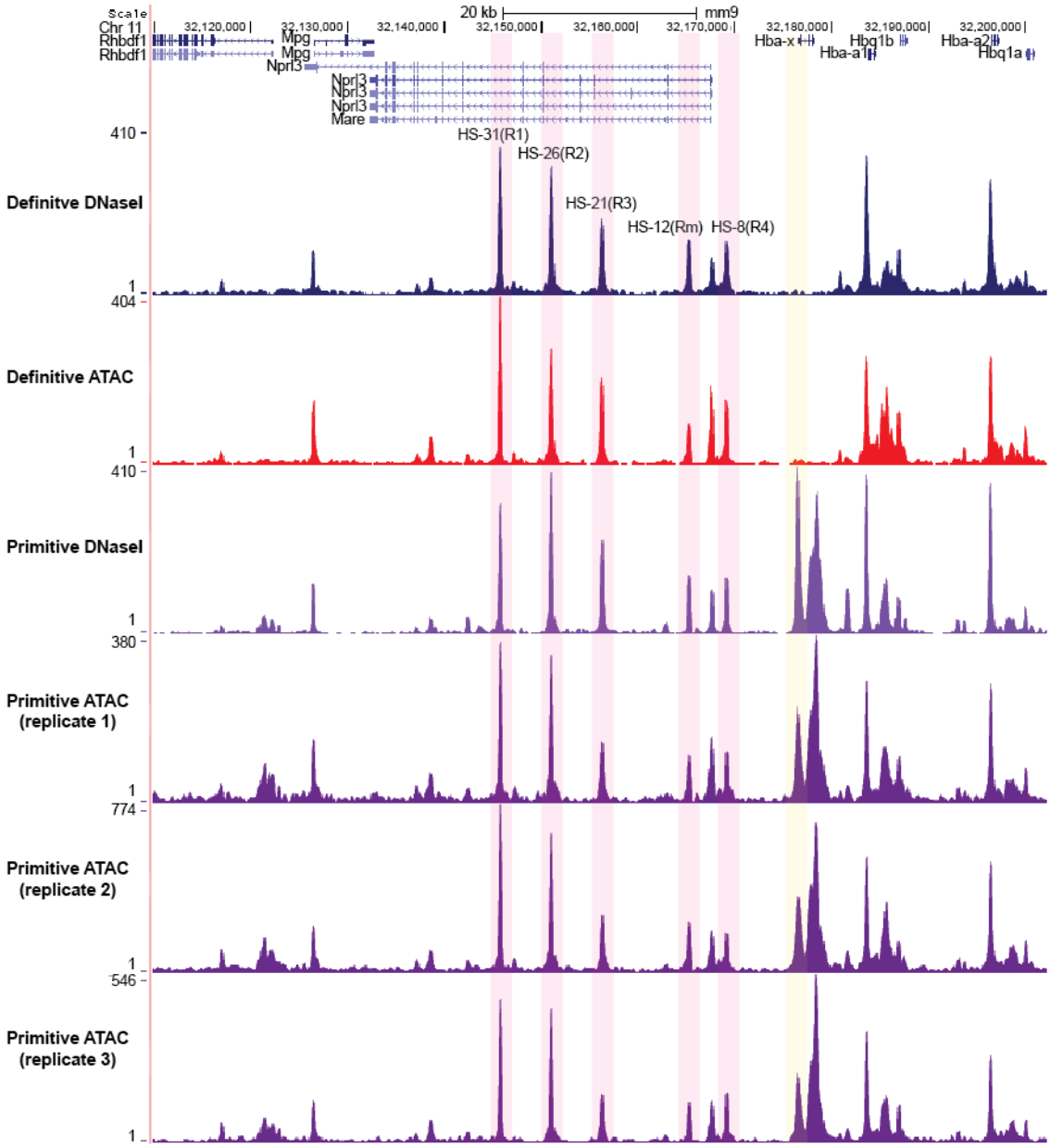


Figure 4.4 DNaseI *vs* ATAC tracks

The figure shows syntenic region of the mouse α-globin cluster. The top two tracks show the *cis*-regulatory HSs of the α-globin cluster in phenylhydrazine-treated-spleen-derived definitive erythroid cells as determined by DNaseI hypersensitivity assay (DHS, shown in blue, n=1) and ATAC (shown in orange, n=1). These two tracks were obtained and given by Maria Suciu for comparison with corresponding data obtained from primitive erythroid cells. The third track (light purple) shows DNaseI hypersensitive sites (HS) in primitive erythroid cells (n=1), coinciding with active promoters and the well-characterized distal enhancers (highlighted in pink) present in the α-globin locus at this developmental stage. The lower three tracks (dark purple) show ATAC tracks from three biological replicates of E10.5 primitive erythroid cells and show that the *cis*-regulatory landscape identified using this novel technique is highly reproducible and comparable to that identified using DHS assays. The yellow shading indicates the active ζ-globin (Hba-x) promoter in primitive erythroid cells. Other actively transcribed genes in the region, including the *Rhbdf1*, *Mpg* and *Nprl3* genes, are marked by DNaseI and ATAC peaks over their promoters. Note that the *Rhbdf1* promoter is hypersensitive in primitive, but not in phenylhydrazine-treated-spleen-derived definitive erythroid cells. Number of reads in all tracks has been normalized for visual inspection.

4.2.3 Interaction of the upstream cis-regulatory elements and the ζ- and α-globin promoters as identified by Capture-C

It is now possible to analyze in detail the complex interaction between *cis*-regulatory elements and promoters at individual loci using chromosome conformation capture. Recently, a high-throughput approach interrogating hundreds of specific *cis*-interactions at high resolution in a single experiment, termed “Capture-C”, has been developed (Hughes et al., 2014). In this technique, the high resolution is made possible by capturing *cis*-interactions at specific viewpoints of interest, using biotinylated sequence-specific oligonucleotides prior to library generation and high-throughput sequencing. The Capture-C technique has additionally been modified to provide significantly higher resolution by further enrichment by performing a second round of capture (Davies et al., 2016).

To investigate developmental stage-specific interactions between the upstream *cis*-regulatory elements and the ζ- and α-globin promoters, I have performed the Capture-C assay using murine E10.5 yolk sac-derived primitive erythroblasts. Oligonucleotides targeting DNA sequences within single *Dpn*II fragments in the ζ- and α-globin promoters were designed. Interactions were also captured from the promoter of Mitoferrin (Slc25A37), an erythroid-expressed gene encoding an iron transport protein required for heme biosynthesis, as an internal positive control for each experiment (oligonucleotide sequences are given in Chapter 2). I also performed the Capture-C assay using E14 embryonic stem cells (ES Cells), derived from 129SvEv mice, where the globin locus is inactive. The data obtained were

used to identify background levels of interaction to compare with the data derived from globin-expressing erythroid cells. All sequencing of Capture-C libraries was performed using the Illumina HiSeq platform and bioinformatic analysis was performed using pre-existing custom scripts and pipelines (J. Davies and J. Telenius). To visualize and compare interactions between the upstream *cis*-acting regulatory elements and the ζ- and α-globin promoters in different cell types, read counts within individual captured regions were normalized to 100,000 reads obtained from each experiment.

Capture-C showed that during murine primitive erythropoiesis the five *cis*-regulatory elements, corresponding to HS-31 (R1), HS-26 (R2), HS-21 (R3), mouse-specific HS-12 (Rm), and HS-8 (R4), interact with the ζ-globin promoter as well as the α-globin promoters (Figure 4.5). This analysis shows that during primitive erythropoiesis the α-globin promoters also interact with the enhancers known to function in definitive erythropoiesis. Additionally, the ζ-globin promoter interacts with the same group of enhancers. To determine positive interactions above background, normalized reads identifying interactions from ES cells were subtracted from the interactions between the individual distal enhancers and the ζ- and α-globin promoters in primitive erythroblasts. The data suggests that, from the viewpoints of the α-globin promoters, *cis*- interactions with all of the known MCS-Rs exist in primitive erythroblasts and that the overall direction, extent and conformation of the interactions appear similar to that seen in definitive erythroblasts (Figure 4.5). The level of *cis*-interactions from the viewpoint of the ζ-globin promoter is similar to that of the α-globin promoters. This suggests that ζ-globin is regulated in a similar fashion to the α-globin genes during primitive

erythropoiesis. To determine any differential contribution of the enhancers, expression analysis of globin genes in knockout mutant mouse models is required (see Section 4.2.4).

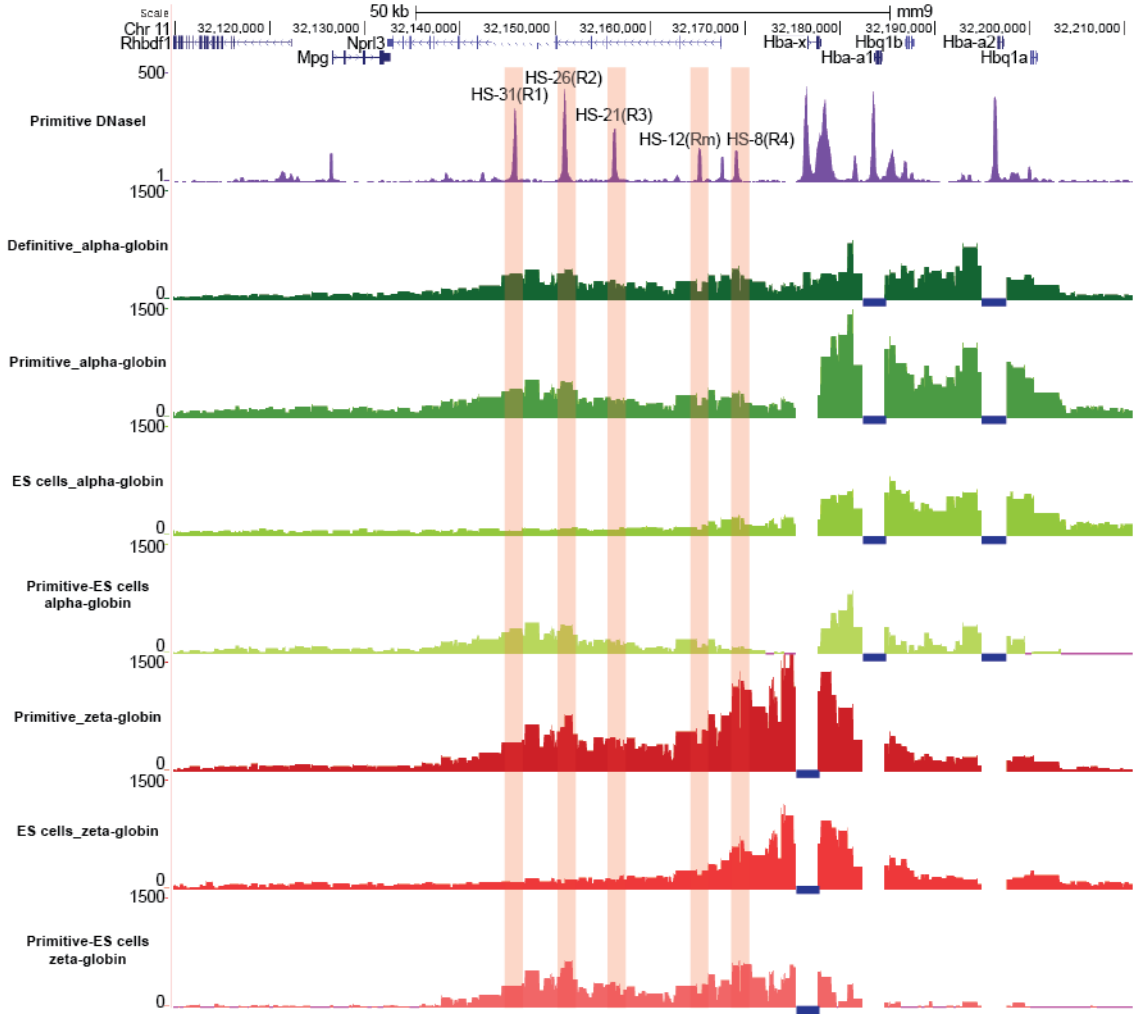


Figure 4.5 Capture-C tracks

The top track shows DNaseI HSs in primitive erythroid cells (n=1) and the second track shows interactions of the distal enhancers with α-globin promoters in phenylhydrazine-treated-spleen-derived definitive erythroid cells from previous studies (n=1, data provided by James Davies for comparison). The third track shows interactions between enhancers and the α-globin promoters in primitive erythroid cells (n=1). Using interactions identified in ES cells (the fourth track, n=2) as a baseline, enhancer-α-globin-promoter interaction differences in primitive erythroid cells were calculated and displayed in the fifth track. The sixth and seventh tracks show interactions between enhancers and the ζ-globin promoter in primitive erythroid cells (n=1) and ES cells (n=2) respectively. The bottom track shows interactions between enhancers and the ζ-globin promoter above the baseline in primitive erythroid cells. Blue rectangles indicate capture viewpoints. Orange transparent shading shows interactions at the individual *cis*-regulatory elements. Note that sequences underlying the hypersensitive site at the *Rhbdf1* promoter in primitive erythroid cells (see Figure 4.4) do not interact with the α- and ζ-globin promoter at this stage. The number of reads in all Capture-C tracks has been normalized for visual inspection.

4.2.4 Contribution of the individual upstream cis-acting regulatory elements on ζ- and α-globin expression in primitive erythroid cells

To investigate the contribution of the individual upstream *cis*-acting regulatory elements on expression of the ζ- and α-globin genes during primitive erythropoiesis, I have measured the levels of accumulated transcripts, using NanoString® technology in E10.5 erythroblasts derived from a number of mouse mutants. A series of mice harboring deletions of individual HSs (HS-31, HS-26, HS-21, HS-12 and HS-8), and double deletions of HSs (HS-31/26 and HS-26/21) have recently been generated in my host department (D. Hay *et al.*, manuscript in review) (Figure 4.6). Since homozygosity for deletions of both HS-31/-26 leads to embryonic lethality at E13.5, this mouse mutant line has been maintained in a heterozygous manner. Homozygotes of the remaining HS-knockouts survive gestation and develop normally, therefore these mutants have been maintained in a homozygous manner.

Gene Name	WT BL6	WT (mixed background)	HS-31(R1)	HS-26(R2)	HS-21(R3)	HS-21(R3) tech rep.	HS-12(Rm)	HS-12(Rm) tech rep.	HS-31/26 (R1R2)_Het	HS-31/26 (R1R2)_Hom	HS-26/21(R2R3)
Hba-a1 (α)	217117	222474	114327	138712	204243	199002	199507	202642	114491	20131	83969
Hba-x (ζ)	150331	136125	137880	143305	121060	118357	110979	107738	84955	59561	132997
Hbb-b1 (β)	7457	6814	6302	10366	11613	11551	8939	8863	6237	11151	4871
Hbb-bh1 (βh1)	71852	97862	149472	89884	77662	72908	102050	97207	125644	140248	166774
Hbb-bh2 (βh2)	111	64	97	83	52	50	75	76	39	53	58
Hbb-y (εγ)	339451	298167	328287	369091	287507	279264	250814	227504	264654	312575	325842
Btf3	1298	1042	942	1104	864	870	574	538	869	1042	1248
Car2	2796	2703	2813	3198	2736	2637	2446	2433	2216	3339	3210
Eef2	3834	2846	2725	3380	2099	2151	1497	1468	2406	2624	3742
Pabpc1	5152	4579	4327	4858	4832	4653	3376	3428	4019	5236	5025
Rpl38	2267	1946	1852	2281	1666	1695	1301	1276	1624	1879	2367
Rps18	7849	7197	7804	8307	5272	5120	4338	4438	6244	8065	8817
Alas2	2420	2640	2466	2787	2232	2051	2153	2154	2979	3368	2929
Fth1	15076	16711	16961	15431	14727	14556	13848	14365	16064	19190	18555
NEG_A	10	7	8	12	4	7	6	4	9	8	8
NEG_B	417	430	485	436	396	400	350	334	406	476	437
NEG_C	30	26	24	19	26	23	17	28	10	7	8
NEG_D	46	46	35	42	49	45	47	28	25	15	27
NEG_E	8	12	6	15	10	11	5	8	6	13	13
NEG_F	166	173	175	172	173	163	160	110	148	162	198
NEG_G	37	49	62	61	24	33	32	30	42	52	65
NEG_H	58	77	106	69	61	67	66	83	106	116	121
POS_A	7163	5784	5497	6189	4857	4692	3962	3923	4027	4799	6116
POS_B	1880	1429	1452	1756	1379	1227	1079	1017	1100	1316	1662
POS_C	746	594	541	698	516	464	456	402	440	411	617
POS_D	212	217	193	195	173	173	179	153	185	204	234
POS_E	185	209	207	206	167	165	164	138	149	194	183
POS_F	23	25	17	19	19	8	12	10	9	24	19

Table 4.1 Example of raw counts of transcripts obtained from a single run of the nCounter® Gene expression assay

The top row indicates genotypes of specific *cis*-element knockouts. Raw transcript counts of globin genes are shown in blue. Counts of transcripts obtained

Gene Name	WT BL6	WT (mixed background)	HS-31(R1)	HS-26(R2)	HS-21(R3)	HS-21(R3) tech rep	HS-12 (Rm)	HS-12(Rm) tech rep	HS-8(R4)	HS-31/26(R1R2)_Het	HS-31/26(R1R2)_Hom	HS-26/21(R2R3)
Hba-a1 (α)	197948	2283670	124348	127759	247727	246431	316507	319127	184756	134081	18694	71190
Hba-x (ζ)	136953	139578	150063	132001	146656	146384	175835	169450	143313	99380	56101	112958
Hbb-b1 (β)	6469	6608	6406	9213	13672	13883	13693	13509	5066	6897	10175	3804
Hbb-bh1 (βh1)	65280	100232	162719	82659	93925	90001	161647	152841	126926	147184	132649	141733
Hbb-bh2 (βh2)	1	1	1	1	1	1	2	2	1	1	1	1
Hbb-y (εγ)	309673	306204	357949	340547	348897	346003	398034	358339	350876	310502	296136	277248
Btf3	844	673	554	658	612	632	401	379	704	590	584	717
Car2	2212	2381	2597	2592	2886	2824	3376	3368	2522	2173	2763	2389
Eef2	3160	2528	2501	2760	2113	2221	1868	1846	2495	2396	2085	2842
Pabpc1	4364	4310	4250	4125	5433	5325	4853	4937	4150	4291	4563	3935
Rpl38	1729	1603	1548	1745	1586	1655	1556	1543	1664	1477	1378	1670
Rps18	6827	7002	8046	7311	5968	5904	6382	6530	7590	6905	7247	7165
Alas2	1868	2316	2218	2213	2274	2097	2910	2928	2138	3069	2791	2149
Fth1	13427	16785	18044	13891	17456	17611	21493	22186	15335	18443	17801	15461
NEG_A	1	1	1	1	1	1	1	1	1	1	1	1
NEG_B	43	43	50	45	36	40	28	36	41	40	50	31
NEG_C	1	1	1	1	1	1	1	1	1	1	1	1
NEG_D	1	1	1	1	1	1	1	1	1	1	1	1
NEG_E	1	1	1	1	1	1	1	1	1	1	1	1
NEG_F	1	1	1	1	1	1	1	1	1	1	1	1
NEG_G	1	1	1	1	1	1	1	1	1	1	1	1
NEG_H	1	1	1	1	1	1	1	1	1	1	1	1
POS_A	6789	5397	5062	5798	4497	4331	3640	3625	6748	3660	4373	5710
POS_B	1506	1041	1017	1364	1019	866	757	719	1488	733	890	1256
POS_C	372	207	106	307	157	103	134	104	410	74	1	210.8
POS_D	1	1	1	1	1	1	1	1	1	1	1	1
POS_E	1	1	1	1	1	1	1	1	1	1	1	1
POS_F	1	1	1	1	1	1	1	1	1	1	1	1

Table 4.2 Example of normalized counts of transcripts obtained from a single run of the nCounter® Gene Expression Assay (from Table 4.1)

The top row indicates the genotypes of specific *cis*-element knockouts. All data shown were normalized using internal positive controls, 6 generally expressed genes and 2 erythroid-specific genes (*Alas2* and *Fth1*). Background counts, assessed using 8 negative controls, were subtracted from normalized transcript counts. Globin transcripts are shown in blue. The generally expressed and erythroid-specific transcripts are shown in green. The negative and positive controls included in the CodeSet are shown in orange and purple respectively.

Globin transcripts account for the vast majority of total transcripts in erythoid cells, however they can be accurately measured by the nCounter® expression assay. This assay is based on digital detection and direct molecular barcoding of target molecules through the use of a color-coded probe pair, thus allowing direct count of a diversity of target transcripts within a single well of hybridized RNA samples (see methods Chapter 2). By this means, this assay is likely to be more accurate than the qPCR technique that relies on amplification. In this study, I have performed the nCounter® expression assay using three biological replicates of total RNA from each of the seven types of homozygous HS-knockouts and also heterozygotes of double knockouts of HS-31/26. Genotyping of embryos obtained from heterozygous crosses of mice harboring double-deletions of HS-31/26 (R1R2) was performed using standard PCR and custom primers (see methods Chapter 2) and embryos with homozygous and heterozygous genotypes selected for this study are shown in Figure 4.7.

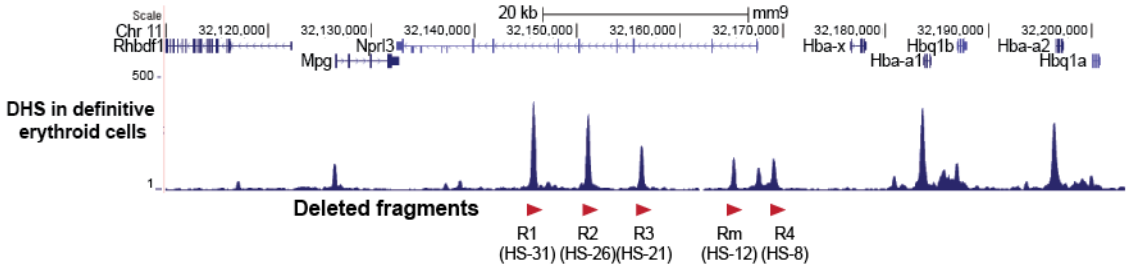


Figure 4.6 Position of elements removed by homologous recombination

The diagram shows the α-globin locus in the mouse. The genes in the region are shown in the upper panel. The center panel shows DNaseI hypersensitive sites (DHS) in phenylhydrazine-treated-spleen-derived definitive erythroid cells obtained as shown in Figure 4.3 and red arrows (lower) indicate the position of predicted *cis*-acting regulatory sequences and the extent of deleted DNA fragments in the mice lines analyzed in this study.

In this experiment, I have carried out nCounter® expression assays in three separate runs. Each run contained at least one biological replicate of total RNA derived from primitive erythroblasts of individual HS-knockouts, the WT

littermates of the HS-31/26 (R1R2) double knockout (mixed background) and WT C57BL6. Technical replicates were included in each cartridge to establish the degree of reproducibility between samples. An example of the raw counts of a run obtained from this experiment is shown in Table 4.1. Since the nCounter® expression assay is a fairly new technique to measure gene expression, careful data analysis is crucial and is described below in detail.

As the first step of data analysis, I have normalized all platform-associated sources of variation across samples, including slight differences in hybridization efficiency, automated purification and binding efficiency of the Capture and Reporter probes by using the internal positive controls. For each positive control, *in vitro* transcribed RNA targets (6 targets per lane), the sequences of which are not homologous to any known organism, are pre-mixed with the Reporter CodeSet during manufacturing. A lane-specific value representative of positive control counts was calculated for each lane using the geometric mean $((x^1x^2\ldots x^n)^{(1/n)}$, where x= raw count of each positive control target, n= total number of positive controls). The average of these calculated values across all lanes was used as the reference against which each lane is normalized. A scaling factor is then calculated for each of the lanes, based on the value of the positive controls in each individual lane relative to the average of this value across all lanes. This normalization factor may then be used to adjust the counts for each gene target and negative control in the associated lane.

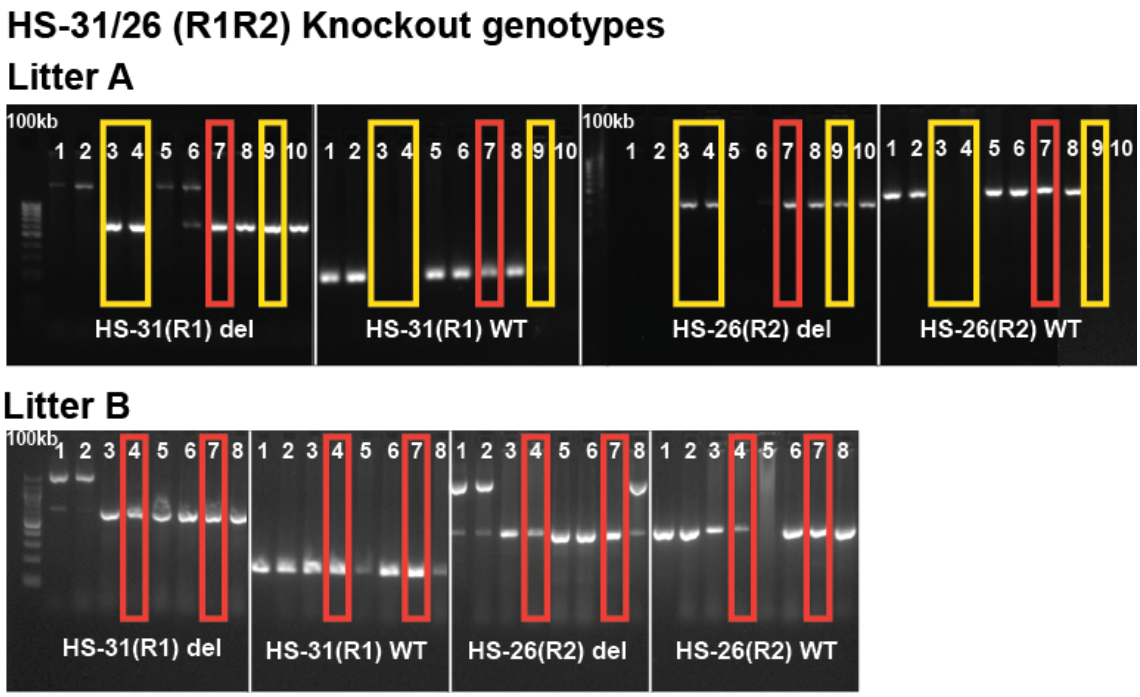


Figure 4.7 Genotypes of mice harbouring HS-31/26 (R1R2) double deletions selected for the nCounter® Gene Expression Assay

The upper panel of each 2% agarose gel indicates embryo numbers from each litter. The lower panels indicate genotypes identified using custom primers (see Chapter 2). The three homozygotes for double knockout of HS-31/26 (R1R2) selected for the experiment are embryos 3, 4 and 9 from litter A. Genotyping of these embryos (yellow boxes) demonstrates the presence of the HS-31 (R1) and HS-26 (R2) deletion product, but not the WT product. The three heterozygotes for the double knockout HS-31/26 (R1R2) selected are embryo 7 from litter A, embryos 4 and 7 from litter B. Genotyping of the 3 embryos (orange boxes) demonstrates the presence of both HS-31 (R1), HS-26 (R2) deletion products and the presence of the corresponding WT products.

To correct for differences in sample input between assays, I have performed CodeSet content normalization. I selected a total of 6 generally expressed genes, including *Btf3, Car2, Eef2, Pabpc1, Rpl38, Rps18,* and 2 erythroid-specific genes with invariant expression: *Alas2* and *Fth1*. These genes were used to determine the accuracy of the normalization: the geometric mean (see previous page) of the reference genes was calculated for each lane. The average of these calculated values across all lanes was used as the reference against which each lane is normalized. A lane-specific normalization factor is then calculated based on the geometric mean in each lane relative to the average of the geometric mean across all lanes. The normalized endogenous

target counts in each lane were derived by multiplying the counts by that lane's normalization factor.

Accurate estimation of probe background is also essential for interpreting expression data. Each CodeSet includes eight reporters for which no corresponding transcript is present. These negative controls can therefore be used to estimate systematic background counts within any single hybridization reaction. Therefore, the average count of the 8 negative controls (plus 2 standard deviations) was subtracted from the normalized endogenous target counts prior to downstream expression data analysis. An example of counts normalized from Table 4.1 is shown in Table 4.2.

The average ζ- to total α-like globin expression ratio in wild type (WT) at this developmental stage is ~36%. The average of total α-like (α, ζ) to total β-like globin transcript counts (βh1, βh2, εγ, β) in WT are close to 1, and the ratio is highly consistent in primitive erythroid cells derived from E10.5 yolk sacs of either C57BL6 or mice from mixed genetic backgrounds including 129SvEv, C57BL6/J and BalbC (0.96 in WT of mixed background and 1.1 in WT C57BL6). This suggests that strain differences do not contribute to a significant difference in the relative level of α- and β-like globin expression.

Assuming that β-like globin expression is not altered when any of the α-upstream *cis*-regulatory elements are removed, the level of α-like globin expression in primitive erythroid cells derived from yolk sacs of mice harboring specific *cis*-element deletions was calculated and shown relative to the proportion of β-like globin expression within the same RNA sample. The mean of the ratios of α-like globin expression across three biological replicates for each of the knockout genotypes is shown in Figure 4.8.

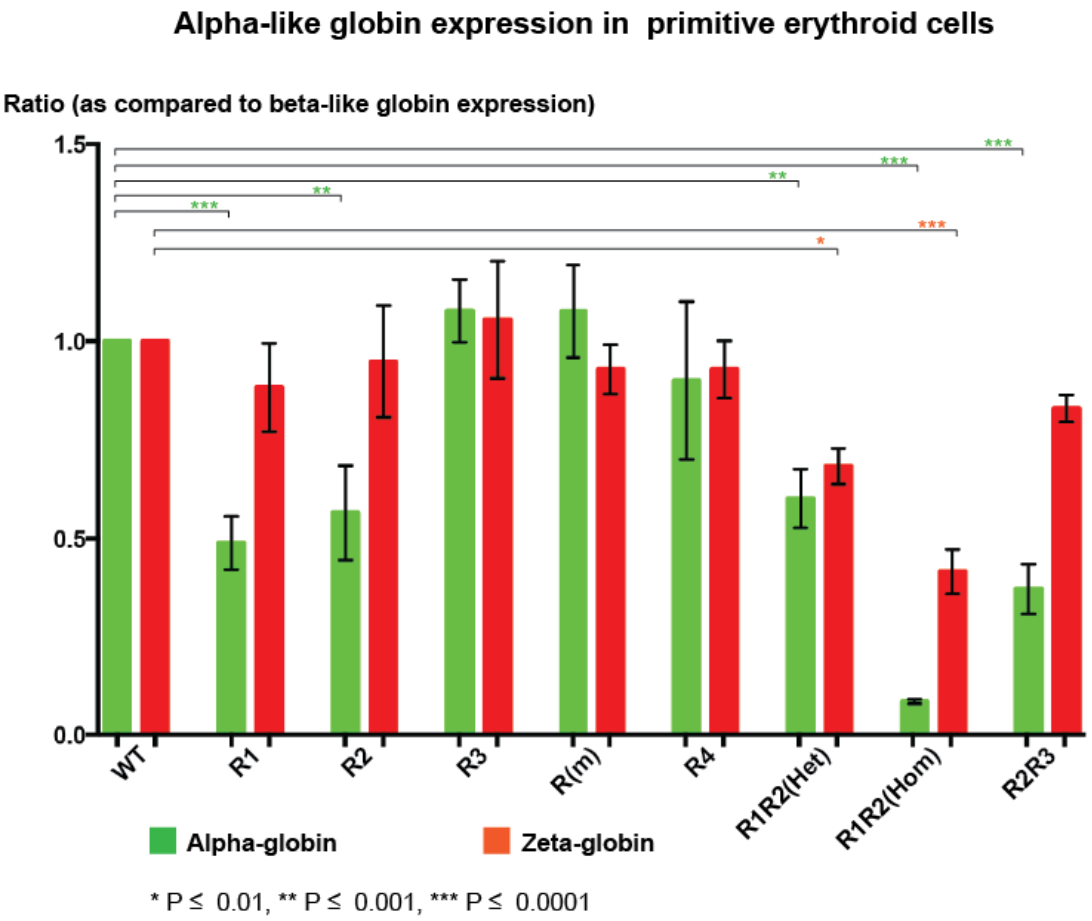


Figure 4.8 Contribution of individual upstream *cis*-acting regulatory elements to α-like globin expression during primitive erythropoiesis

The bar graph shows α-like globin expression, performed using the nCounter® gene expression assay, in primitive erythroblasts derived from murine yolk sacs at embryonic day 10.5 (E10.5). The X-axis indicates the different genotypes. The Y-axis shows the ratio of α-like globin expression in comparison with total β-like globin transcripts. The green bars represent α-globin expression and the red bars represent ζ-globin expression. Error bars represent standard deviation (SD) across three biological replicates. Asterisks (*) indicate the level of statistical significance.

Of the five distal *cis*-acting regulatory elements, R1 (HS-31) and R2 (HS-26) contribute most to α-globin expression in primitive erythroid cells, as removal of the individual elements resulted in significant reduction of α-globin transcripts to approximately 50% and 60% respectively as compared to WT levels. Homozygotes for the double knockouts of R1R2 are shown to have an additive effect since α-globin transcripts are found at ~8% of the WT level. A significant reduction in α-globin transcripts, to ~40% of that seen in WT primitive erythroid cells, was observed in cells derived from homozygotes for

the R2R3 double knockout. Reflecting previous findings in definitive erythropoiesis, deletion of R3, R(m) and R4 does not seem to adversely affect α-globin expression.

Interestingly, unlike α-globin, deletions of R1 or R2 alone do not significantly affect the accumulation of ζ-globin transcripts (present at 88% and 95% of the WT level respectively). ζ-globin expression is reduced to approximately 40% of WT in homozygotes harboring R1R2 double deletions suggesting R1 and R2 might have synergistic effect on ζ-globin expression in primitive erythroid cells. More importantly, this finding supports the hypothesis that ζ-globin is not as reliant on the upstream *cis*-regulatory elements as α-globin, because its expression level is not reduced as far as that of α-globin by deletion of these elements. There are no novel distal HSs in the α-globin locus during primitive erythropoiesis and the Capture-C experiment suggests that the ζ-globin promoter interacts within the previously defined domain. Taken together these data suggest that the developmental stage-specificity of the ζ-globin gene is regulated to a large extent by elements that lie within the gene itself or its promoter and the associated key *trans*-acting factors. The remaining *cis*-acting regulatory elements, including R3, R(m) and R4 have little effect on ζ-globin expression and it is clearly shown in this study that these three elements are unlikely to be responsible for the expression of ζ-globin at this stage. Therefore, the apparent interaction between MCS-R4 and the ζ-globin promoter, detected by Capture-C, is likely to result from the proximity of these two elements. Significant reduction of both α- and ζ-globin transcripts in homozygotes harboring double deletions of R1R2 results in reduction of total α-like globin expression to ~20% (Table 4.2) and this result is highly similar in

all three biological replicates of the same genotype. However, this amount is adequate for embryonic survival, since homozygotes for the R1R2 deletions die at E13.5 after the ζ-globin gene is repressed (D. Hay *et al.,* manuscript in review).

4.2.5 DNaseI footprints and predicted transcription factor binding at the ζ-globin promoter and the 3' ζ-globin DNaseI peak

The data from the DNaseI-seq experiment is genome-wide, therefore this is an excellent resource to allow differences in transcription factor binding patterns between primitive and definitive erythropoiesis to be determined. However, a comparison of over- and under-represented sequences underlying the footprints present in this data, as well as footprints within the two major distal *cis*-acting regulatory elements (R1 and R2) of the α-like globin genes requires a large amount of further bioinformatic work, which is beyond the scope of this thesis. Therefore this work has been undertaken as a part of another D.Phil project by Maria Suciu in the Molecular Haematology Unit. Since the developmental stage-specificity of ζ-globin expression may be encoded within *cis*-acting sequences in the ζ-globin gene itself or the flanking sequences, it is relevant to consider the footprints present within this region. Within promoters, enhancers and unclassified DNaseI peaks, sequences bound by transcription factors are protected from the DNaseI enzyme and therefore can be visualized as crypts within each entire DNaseI peak and the sequences underlying such crypts are known as “footprints”. Sequence footprints can be used to search in available databases for potential binding of transcription factors. Detailed analysis of the DNaseI peaks over this region suggests that there are > 50 sequence footprints present, each of which is

likely to represent binding of a specific protein (Figure 4.9). Analysis of the sequence footprints using predictive software suggests several hundred potential binding proteins. To determine which of these factors are responsible for embryonic expression of ζ-globin, a candidate list would need to be generated and refined using cell-specific expression data (see Chapter 5) and ChIP analysis performed. In summary, the work presented in this chapter is an essential starting point for the identification of the transcription factors responsible for embryonic ζ-globin expression, however, further study will be required to identify the factors responsible.

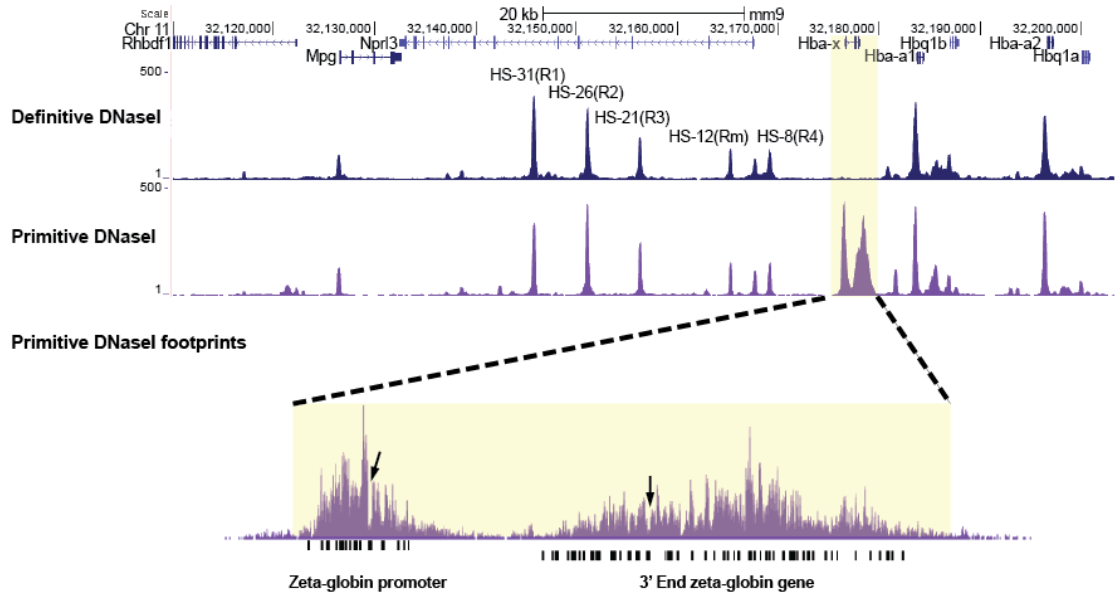


Figure 4.9 Footprints within the ζ-globin DNaseI peaks

The upper and middle tracks show DHSs in phenylhydrazine-treated-spleen-derived definitive and primitive erythroid cells taken from Figure 4.3. The ζ-globin gene is highlighted in yellow. The lower track shows an expanded view of the ζ-globin DHSs and examples of "crypts" representing footprints are shown with black arrows. Significant footprints and the position of associated underlying motifs (black squares) are shown. Footprint analysis was performed by Maria Suciu.

4.3 Discussion

Although previous work has suggested that ζ-globin expression may be reliant on *cis*-acting regulatory elements, all of this work was conducted in transgenic mice containing the human α-globin locus and also in the K562 cell line (De

Gobbi et al., 2007; Gourdon et al., 1994; Huang et al., 1998; Pondel et al., 1992; Ren et al., 1993; Rombel et al., 1995; Zhang et al., 1993; Zhang et al., 1995). In this thesis I have integrated the pre-existing technique of assaying for DNaseI hypersensitivity with recently developed techniques, including ATAC-seq, Capture-C and nCounter® Gene expression assays to identify the *cis*-regulatory landscape of α-like globin genes in primary primitive erythroid cells. Erythroblasts derived from yolk sacs at E10.5 showed activity of the ζ-globin promoter, as demonstrated by DNaseI-seq and ATAC-seq coupled with expression analysis (Figure 4.3, 4.4) and are therefore suitable for study of the ζ-globin *cis*-acting regulatory elements.

The finding from this work demonstrates for the first time that the previously characterized five *cis*-acting regulatory elements 5′ of the α-like globin genes are shared between primitive and definitive erythropoiesis and there is no developmental stage-specific distal *cis*-acting element in the locus in primitive erythroid cells. These five *cis*-regulatory elements, corresponding to HS-31 (R1), HS-26 (R2), HS-21 (R3), HS-12 (Rm) and HS-8 (R4), interact with the α- and ζ-globin promoters in primitive erythroid cells, as demonstrated by Capture-C.

This work has demonstrated the contribution of the individual *cis*-acting regulatory elements on α-like globin expression during primitive erythropoiesis. Previous studies in definitive erythroid cells have shown that among the five *cis*-acting regulatory elements, only HS-31 (R1) and HS-26 (R2) predominantly regulate α-globin transcription and the function of HS-21 (R3), HS-12 (Rm) and HS-8 (R4) remain unclear. Interestingly, these three *cis*-acting elements of undefined function are not shown to play a significant

role in regulating α-like globin expression in primitive erythroid cells. This finding suggests that HS-21 (R3), HS-12 (Rm) and HS-8 (R4) are not developmental stage-specific *cis*-regulatory elements of the α-globin locus and their functions are yet to be identified.

Accumulation of ζ-globin transcripts are reduced to ~40% of the WT level in primitive erythroblasts derived from mice homozygous for double deletions of HS-31/26 (R1R2), suggesting ζ-globin is, at least, partly regulated in *cis*-predominantly by these two elements. On the other hand, this finding supports the hypothesis that ζ-globin is less reliant on the upstream enhancers than the α-globin genes, as ζ-globin expression is not decreased as much as α-globin expression in R1R2 knockouts. Furthermore, developmental stage-specificity of the ζ-globin gene is likely to be defined within the ζ-globin promoter or the flanking regions on both sides of the ζ-globin gene. Previous studies performed in K562 cells (a human ζ-globin expressing erythroid cell line) demonstrated using DNaseI and gel mobility shift assays that there are binding sites of GATA1 (previously named NF-E1), Sp1 and CCAAT box-binding factors located in the ζ-globin promoter (Yu et al., 1990). Binding of relevant *trans*-acting factors to these sequences may be involved in developmental stage-specific expression of the ζ-globin gene. The work presented in this chapter is a critical first step in identification of further *trans*-acting factors responsible for stage-specific ζ-globin expression.

Chapter 5 : Identification of *trans*-acting factors regulating ζ-globin expression

5.1 Introduction

Transcription factors are key regulators of lineage fate as they underlie differential transcriptional programs. Recently great effort has been focused on the elucidation of hemoglobin switching at the β-globin locus, in an attempt to identify the underlying mechanisms and transcription factors responsible. This is because reactivation of functional fetal or embryonic globin in adult erythropoiesis may provide a potential long-term cure for individuals affected with hemoglobinophathies, including β-thalassemia and sickle cell disease.

Recent studies of globin switching at the β-globin locus have identified several *trans*-acting factors which, when mutated or depleted, cause increased expression of the embryonic or fetal globins. These include KLF1, BCL11a, KLF3, KLF8, TR2, TR4, DNMT1, MYB and the NuRD, CoREST and DRED complexes (reviewed in Chapter 1.4). However, little is known about the effect of these proteins on the α-globin locus. One of the first key erythroid transcription factors to be identified was *KLF1* and more recently mutations in this gene have been found to underlie hereditary persistent expression of fetal hemoglobin (HPFH). KLF1 is a zinc finger protein that binds differentially to the embryonic, fetal and adult globin gene promoters during development. Decreased expression of *KLF1* underlies increased transcription of embryonic and fetal globin at the β-globin locus (Zhou et al., 2010). KLF1 is also thought to regulate expression of a second zinc finger protein BCL11A which is likely

to mediate silencing of the embryonic and fetal β-globin genes through both long-range interaction within the β-globin cluster and local interactions with chromatin-associated SOX6 at the proximal promoters of the γ-globin genes (Xu et al., 2010). Klf1 has been shown to bind specific sites in the α-globin locus in mouse (Tallack et al., 2010) and BCL11A has also shown to bind the α-globin locus in human erythroid cells (Jawaid et al., 2010), however, their effect on switching the α-globin genes is unclear. Evidence that KLF1 may also affect regulation of the α-globin genes comes from a recent cohort study of 8 patients with chronic hemolytic anemia or chronic non-spherocytic hemolytic anemia who were found to be compound heterozygotes for loss of function *KLF1* mutations. In addition to raised levels of fetal haemoglobin, these patients were also found to have increased levels of the embryonic ζ- and ε-globins (Viprakasit et al., 2014). Further evidence that KLF1 may affect expression of ζ-globin comes from patients with congenital dyserythropoetic anemia harbouring a p.E325K gain of function change in *KLF1* who were also found to express embryonic ζ- and ε-globin (Arnaud et al., 2010).

Recent studies have compared the transcriptomes of primitive and definitive erythroid cells using microarrays and have identified core erythroid genes and differentially expressed (DE) genes (Kingsley et al., 2013). However, the genes encoding transcription factors underlying the differential globin expression profile between the two lineages have yet to be elucidated.

The work in this thesis aims to identify novel *trans*-acting factors, which regulate ζ-globin expression, and also gain further insight into the mechanism of action of known *trans*-acting factors such as Klf1 and Bcl11a. As a first step this part of the project aims to identify transcriptional differences between

primitive and definitive erythroid cells by RNA sequencing and gene expression arrays using the mouse as a model system. Differentially expressed genes will then be systematically prioritized for further analysis and functional testing using a Cas9/CRISPR library screen. By understanding the mechanisms involved in ζ-globin gene expression it may be possible to design novel therapeutic approaches to achieve a better outcome for patients with α-hemoglobinopathies.

5.2 Results

5.2.1 Selecting primitive and definitive erythroid cells for transcriptome comparison

I have compared the transcriptomes of primitive and definitive erythroblasts with the aim of investigating the *trans*-regulatory network underlying the different globin expression profiles between primitive and definitive erythropoiesis.

Primitive erythroblasts were initially obtained from two sources: 1) Primary primitive erythroblasts; 2) *In vitro* cultured primitive erythroblasts. Primary primitive erythroblasts were derived from yolk sacs at embryonic day 10.5 (E10.5). A total of ~1x10^6 erythroblasts, giving rise to ~1 μg of total RNA, were obtained from a single embryo. 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, also termed polychromatic erythroblasts (Figure 5.1A). At this stage ζ-globin accumulated transcripts accounted for ~40% of total α-like globin transcripts present at this stage as demonstrated by qPCR (Figure 5.1B). *In vitro* cultured

primitive erythroblasts were also obtained for comparison with the primary cells in this experiment, using a protocol adapted from Greenfest-Allen and colleagues (Greenfest-Allen et al., 2013) (see methods in Chapter 2). In the *in vitro* culture system primary primitive erythroblasts were obtained from yolk sacs at E8.5 and were cultured for 48 hours. In our hands, a total of ~$1x10^6$ erythroblasts per litter of embryos were obtained after 2 days of culture. During the culture period, primitive erythroblasts had semi-synchronously matured *ex vivo* and become reasonably homogenous at the morphologically equivalent stage to that found in E10.5 yolk sac (Figure 5.1A). However, accumulation of ζ-globin transcripts in *in vitro* cultured primitive erythroid cells at this stage varied from ~30% to ~55% of total α-like globin transcripts as measured by qPCR (Figure 5.1B).

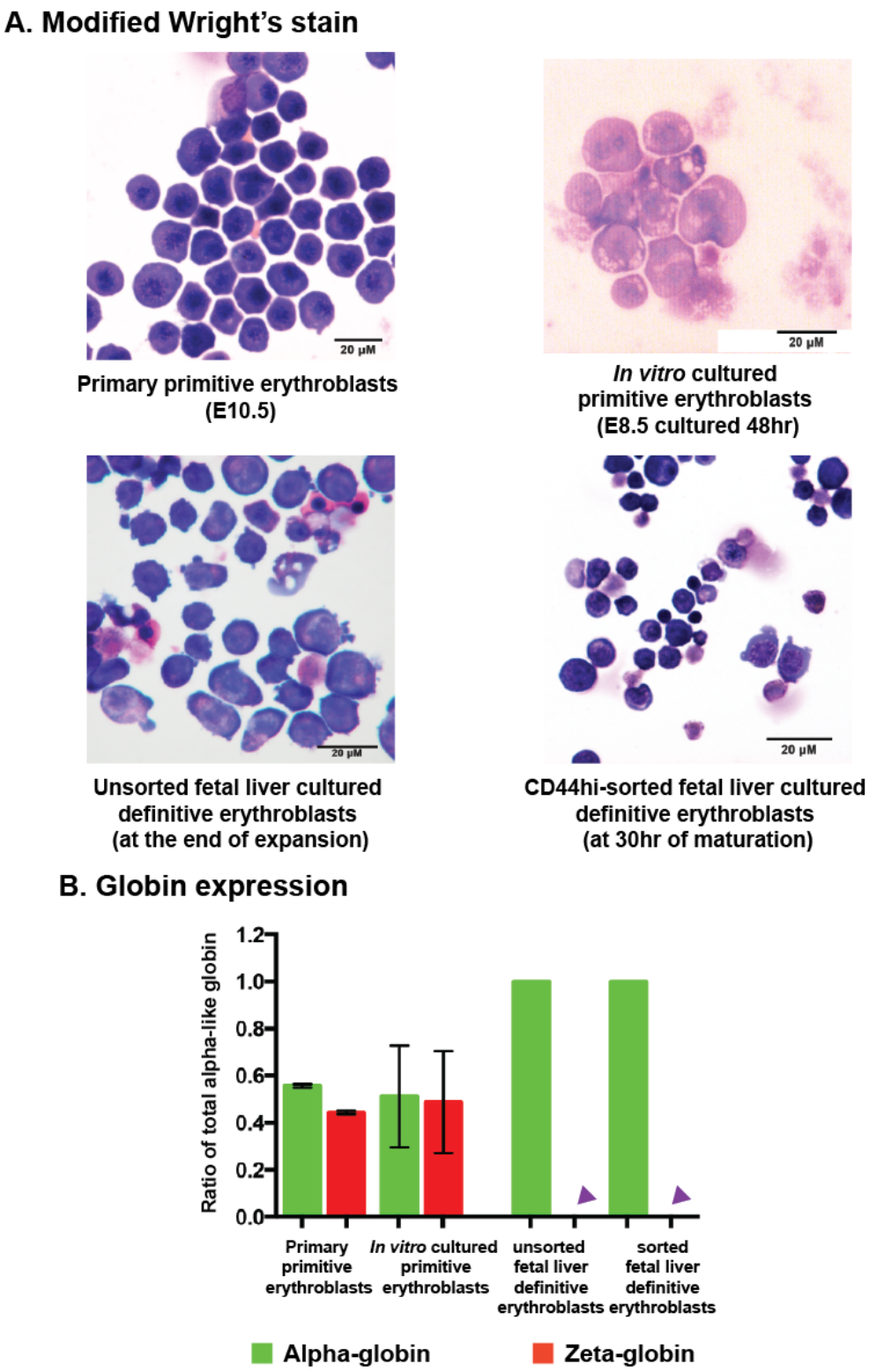


Figure 5.1 Modified Wright's stain of all erythroid samples used in the RNA-seq experiment and their α-like globin expression

A. Different sources of primitive and definitive erythroblasts employed in the RNA-seq experiment. Primitive erythroblasts obtained from E10.5 yolk sacs (left upper) are highly homogeneous at the polychromatic stage. E8.5 primitive erythroblasts matured semi-synchronously *in vitro* during 48 hours of culture and reached maturational stage equivalent to E10.5 primary primitive erythroblasts (right upper). Morphology of definitive erythroblasts derived from unsorted E12.5 fetal liver culture is heterogeneous ranging from proerythroblasts

to nucleated red blood cells (left lower). The majority of definitive erythroblasts derived from 30-hour culture of CD44hi-sorted E12.5 fetal liver cells matured up to the polychromatic stage as shown in the right lower panel. B. qPCR data showing ratios of α-globin (green) and ζ-globin (red) transcripts in each RNA sample (globin expression assays performed using 2 biological replicates of each RNA sample). Error bars represent standard deviation (SD) between two biological replicates. α- and ζ-globin expression in E10.5 primary primitive erythroblasts was obtained from Figure 4.1. ζ-globin is repressed (transcripts undetectable by qPCR) in both populations of definitive erythroid cells (purple arrows).

As reviewed in Section 1.3.3 and 1.3.4, EMPs migrate to and rapidly differentiate in the mouse fetal liver between E10.5-E11.5 before mature EMP-derived definitive erythroblasts eventually enter the circulation. HSCs seed the fetal liver between E11.5-E12.5, where maturation of HSC-derived definitive erythroblasts occur. Therefore, at E12.5 the fetal liver most likely consists mainly of HSC-derived, rather than EMP-derived definitive erythroblasts. HSC-derived definitive erythroid cells for transcriptome comparison were obtained from *in vitro* mouse fetal liver culture at two different stages (see methods in Chapter 2). Initially, definitive erythroblasts were derived from E12.5 unsorted fetal liver cells that had been cultured to the end of expansion phase (with low EPO level). The numbers of early erythroblasts (Figure 5.1A) obtained at this stage range from 5 x 10^7 to 1 x 10^8 cells from a single mouse fetal liver. ζ-globin is repressed, while α-globin transcripts are observed in the unsorted fetal liver cultured-derived erythroid cells as determined by qPCR (Figure 5.1B). Subsequently, I have employed a protocol recently modified from the original version given by Marieke Von Lindern (see methods in Chapter 2) to improve the purity and morphological stage synchronization. This protocol involves expansion of cells derived from E12.5 fetal liver for up to 6-7 days followed by Ter119 depletion to remove mature red cells. Subsequently, CD44-hi early erythroid progenitors (Chen et al., 2009) were sorted to plate on erythroid differentiation media followed by *in*

vitro erythroid maturation for up to 30 hours. I used this protocol with the aim of obtaining erythroblasts at the intermediate stage, when α-globin is more prominently expressed (Figure 5.1A). Using this approach, the *in vitro* fetal liver culture-derived definitive erythroid cells are more morphologically comparable with primitive erythroid cells obtained at E10.5. The immunophenotype and morphology of definitive erythroid cells obtained at different stages of the protocol are shown in Figure 5.2. I have identified the globin expression profile in sorted fetal liver cell culture-derived erythroblasts using qPCR and confirm that ζ-globin is also repressed at the selected maturational stage.

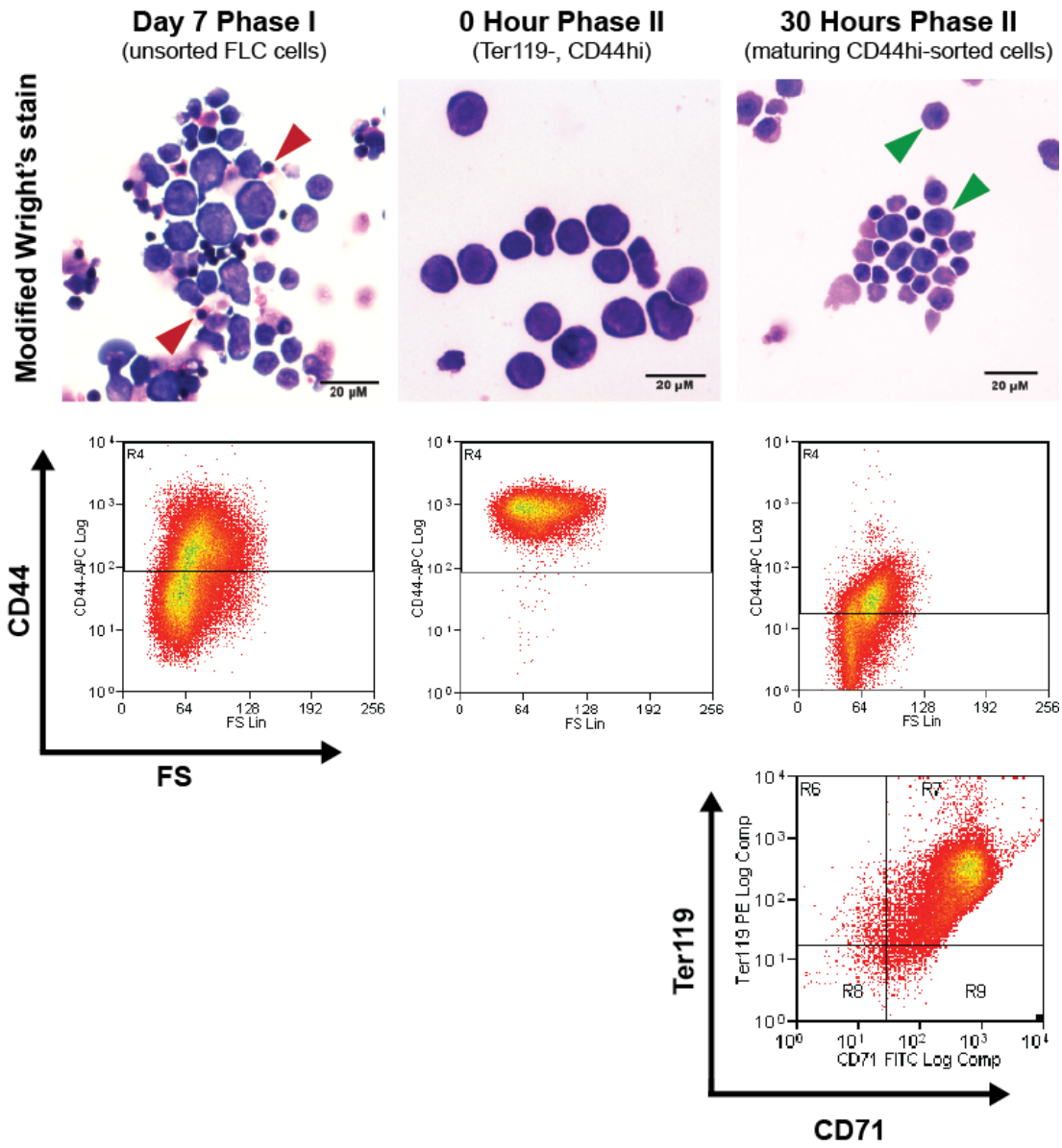


Figure 5.2 Morphology and immunophenotype of definitive erythroid cells derived from unsorted and sorted fetal liver culture

The upper panel shows morphology of cells during different stages of E12.5 fetal liver culture. There is a heterogeneous mix of differently staged erythroid cells at the end of the expansion phase (Day 7 Phase I), some of which are at the terminal stage of differentiation (red arrows). After Ter119 depletion to remove mature erythroid cells and CD44hi cell sorting, a homogenous population of early erythroblasts was obtained for the initiation of the maturation phase (0 Hour Phase II). The culture was maintained for up to 30 hours of maturation, the stage when the majority of erythroid cells are intermediate erythroblasts (green arrows). The middle panel shows CD44 expression of erythroid cells as assessed by flow cytometric analysis. There is a range of CD44 expression at the end of expansion phase, consistent with a mix of different maturational-stage erythroid cells (left). CD44hi sorting resulted in > 95% purity of CD44hi erythroblasts (middle). During terminal maturation of erythropoiesis, CD44 is gradually reduced as reported in previous studies (Chen et al., 2009) (right). The lower panel shows that the vast majority of the cells at 30 hours of maturation are double positive for Ter119 and CD71, indicating they are intermediate erythroblasts.

5.2.2 Experimental design and analysis of confounding factors

The experimental design to compare the transcriptomes between primitive and definitive erythroblasts using RNA-seq is summarized in Figure 5.3. Total RNA derived from each sample was DNase treated to remove genomic DNA prior to downstream analysis. To obtain maximum informative reads, the total RNA samples were globin- and ribo-depleted to remove globin (accounting for ~90% of polyadenylated transcripts in erythroid cells) and ribosomal transcripts (accounting for ~80% of total RNA) before sequencing. The globin- and ribo-depletion processes are achieved using biotinylated oligonucleotides followed by streptavidin-mediated removal. Initially, total RNA derived from definitive erythroblasts was globin- and ribo-depleted using the commercially available Globin Zero Gold® kit (Epicentre), which depletes adult globin mRNA and ribosomal RNA. The initial evaluation of RNA-seq data demonstrated that total globin transcripts were reduced to < 0.1 % of total reads and the data set obtained from the RNA-seq analysis contains the majority of the 1% of the lowest expressed genes present in the recent study by Kingsley and co-workers (Kingsley et al., 2013). In subsequent RNA-seq sample preparation, total RNA obtained from primitive erythroblast samples was globin- and ribo-depleted with a different depletion kit from NuGen®, which depletes ribosomal RNA and adult globin mRNA and which was also modified to deplete embryonic globin mRNA using additional custom oligonucleotides. Target transcripts of the two different globin depletion kits are shown in Table 5.1. The depletion process in the NuGen® kit was performed after reverse transcription of total RNA to cDNA. The majority of all targeted globin was depleted compared to pre-depletion levels, as demonstrated using qPCR

(Figure 5.4). Globin reads obtained from sequencing accounted for < 0.5% of total reads, suggesting successful globin depletion using the NuGen® kit with custom oligonucleotides.

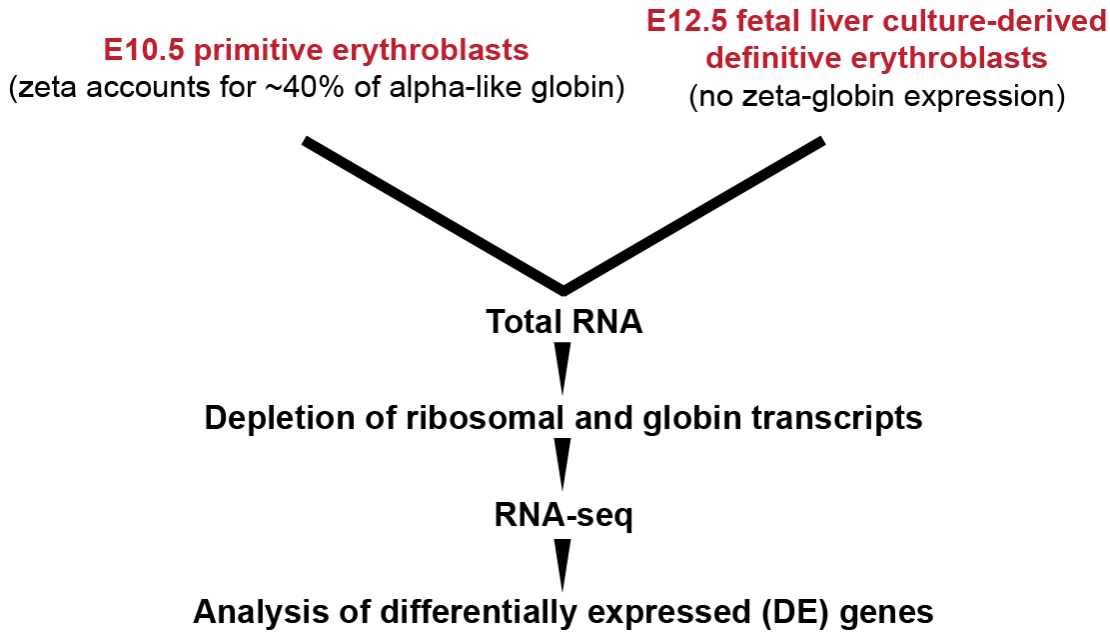


Figure 5.3 Experimental design of the RNA-seq experiment

	Globin Zero Gold Kit®	NuGen Globin Depletion Kit® with custom oligonucleotides
Globin mRNA	Hba-a1, Hba-a2, Hbb-b1, Hbb-b2	Hba-a1, Hba-a2, Hbb-b1, Hba-x, Hbb-bh1, Hbb-y
Mitochondrial rRNAN	18S, 12S	12S, 16S
Nuclear rRNA	28S, 18S, 5.8S, 5S	18S, 28S, 5.8S

Table 5.1 Target transcripts of the globin depletion kits employed before RNA-seq.
Targets for custom oligonucleotides added to the pre-existing commercially available kit are highlighted in red.

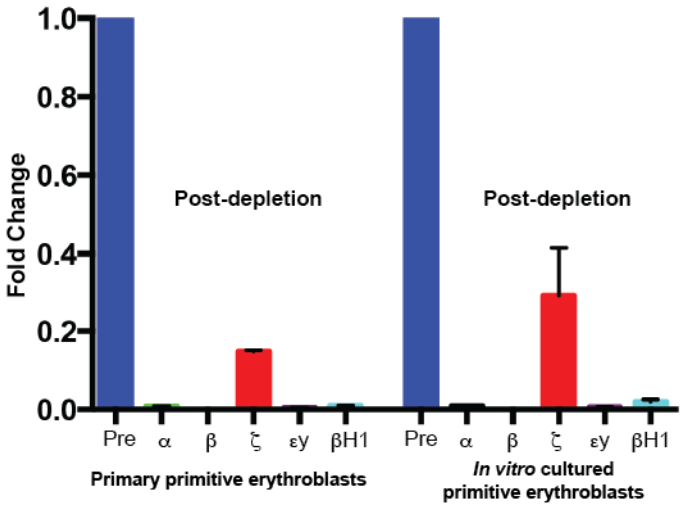


Figure 5.4 Globin expression analysis of pre- and post-globin depletion RNA samples

Globin depletion of E10.5 primary (n=2) and *in vitro* cultured primitive erythroid cells (n=2) was performed using the NuGen® depletion kit with custom oligonucleotides. Globin expression assays were performed using qPCR, post-depletion levels of each globin transcript were compared to their pre-depletion levels. Error bars represent SD from 2 biological replicates. Although ζ-globin transcripts were less effectively depleted compared to other globins, post-depletion levels of all globin transcripts were significantly reduced from their pre-depletion levels with p-value ≤ 0.0001.

The effects of the different globin- and ribo-depletion kits on expression profiles were evaluated using a single sample of total RNA derived from unsorted fetal liver cell culture. Surprisingly, as many as 878 genes were differentially expressed at log fold change (logFC) of > 2 within the same biological sample undergoing different globin- and ribo-depletion processes. The difference may be visualized using the Gene Expression Dynamic Inspector (GEDI, (Eichler et al., 2003), Figure 5.5A). To investigate what differences the two different globin- and ribo-depletion kits had contributed, 878 DE genes were functionally annotated using the DAVID bioinformatics database (Huang da et al., 2009) (Figure 5.5B). This demonstrates that the majority of the DE genes are annotated as encoding various types of metal ion binding proteins, suggesting different non-specific off-target effects of the two kits. However, a number of genes encoding transcription factors are also

affected by differences between the kits, suggesting that the same globin- and ribo-depletion process should be employed with every RNA sample analyzed.

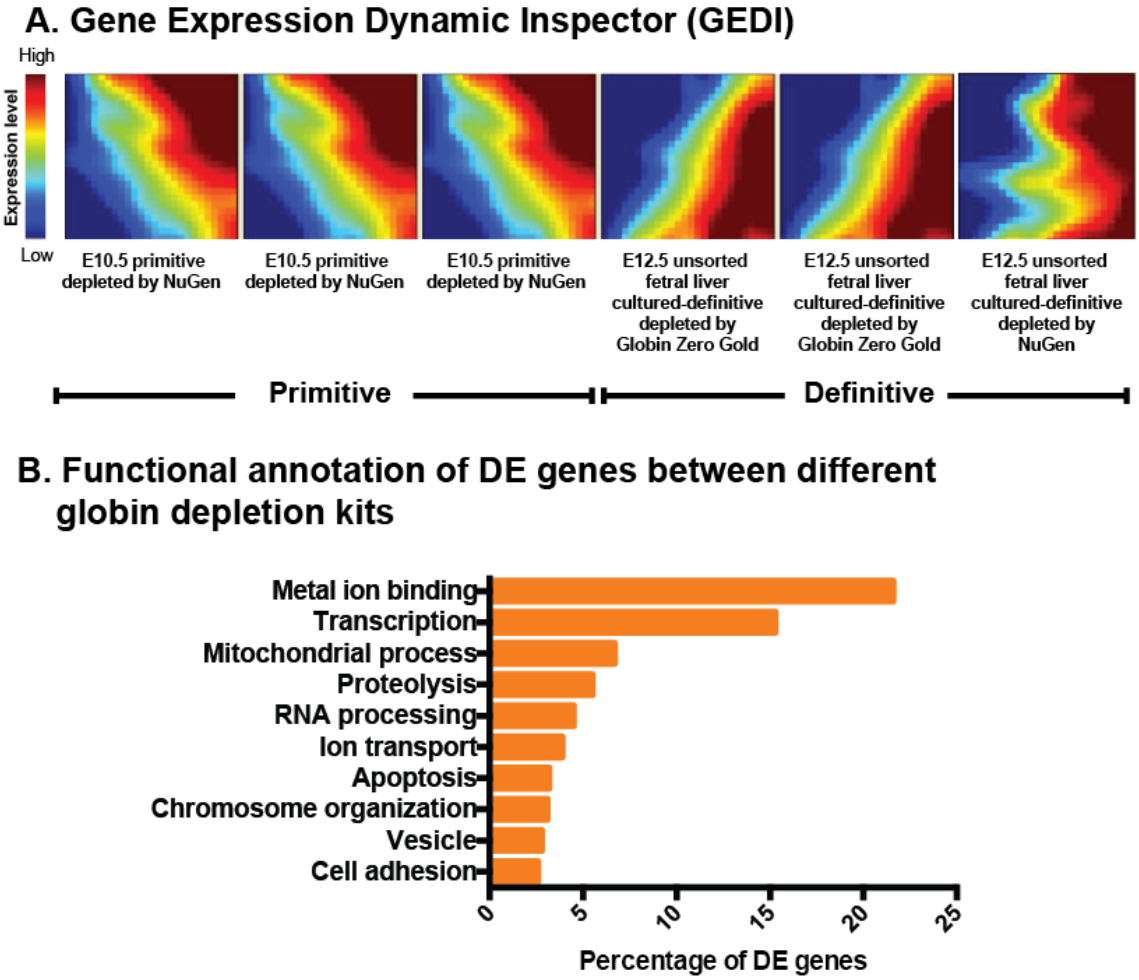


Figure 5.5 Effect of different globin- and ribosomal depletion kits on RNA-seq data

A. GEDI plots of total RNA samples from E10.5 yolk sac derived-primitive and E12.5 unsorted fetal liver cultured-definitive erythroblasts, which had been globin- and ribo-depleted with two different depletion kits. Mosaics (individual pictures) visualizing gene expression profiles of each replicate of RNA from E10.5 primitive erythroblasts (left, n=3), and E12.5 unsorted fetal liver definitive erythroblasts from E12.5 unsorted fetal liver cell culture (right, n=3). The tiles represent individual “miniclusters” of genes. The color of each tile of the mosaic is determined by the average expression level of that respective minicluster evaluated in each sample. All total RNA samples obtained from primitive erythroblasts were globin- and ribo-depleted using the NuGen® kit with custom oligonucleotides, resulting in similar gene expression patterns. Two total RNA samples obtained from fetal liver cultured-definitive erythroblasts were depleted using the Globin Zero Gold® kit (n=2). The gene expression patterns of these samples are visually similar and they are distinct from the expression pattern of the other fetal liver cultured-definitive erythroblast RNA sample (right, n=1), which was selected for a pilot experiment using the NuGen® depletion kit. B. The figure shows functional annotation using the DAVID bioinformatics database of DE genes between 2 technical replicates of E12.5 unsorted fetal liver cultured-definitive erythroblasts that underwent different depletion processes.

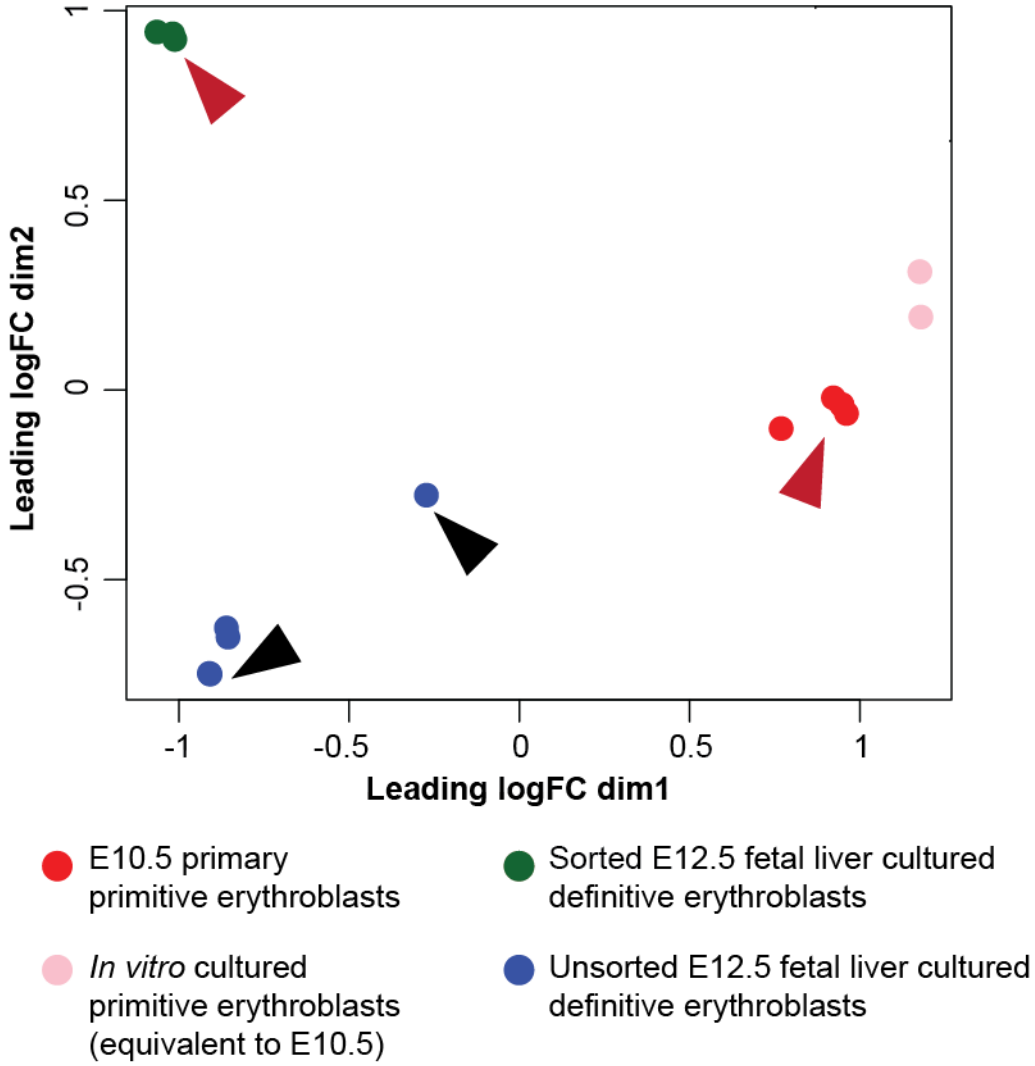


Figure 5.6 Principal component analysis from the RNA-seq experiment

Replicates of E10.5 primary (3 biological replicates plus 1 technical replicate for correction of batch effect) and *in vitro* cultured primitive erythroid samples (2 biological replicates) were colored in red and pink respectively. Definitive erythroid samples derived from CD44hi sorted E12.5 fetal liver cell culture (3 biological replicates) are colored in green and definitive erythroid samples obtained from unsorted E12.5 fetal liver cells culture are colored in blue (3 biological replicates plus 2 additional technical replicates for correction of batch effect and identification of DE genes occurred as a result of different globin- and ribosomal depletion kits). Two technical replicates of unsorted fetal liver-derived definitive erythroblasts that underwent the two different globin- and ribo-depletion processes are indicated with black arrows. The two groups of samples, E10.5 primary primitive erythroblasts and sorted E12.5 fetal liver cultured definitive erythroblasts, selected for final transcriptome comparison are indicated with red arrows.

To determine whether *in vitro* cultured primitive erythroid cells resemble primary primitive erythroblasts at the equivalent stage to that obtained from murine yolk sacs, the difference in expression profiles between these two

erythroid populations was analyzed. The two biological replicates of *in vitro* cultured primitive erythroblasts were not well clustered, as demonstrated in the Principal Component Analysis (PCA) plot (Figure 5.6, pink dots). There were a total of 1,514 significantly DE genes between the *in vitro* cultured primitive erythroblasts and the primary primitive erythroblasts (1,310 genes upregulated in E10.5 yolk sac-derived primitive erythroblasts and 204 genes upregulated in *in vitro* cultured primitive erythroblasts at the equivalent stage). I have functionally annotated the 1,514 genes using the DAVID database and demonstrated that cell signaling molecules and transcriptional networks contribute to the major differences between these two erythroid cell populations (Figure 5.7). In addition, genes encoding cell cycle regulation are upregulated in primitive erythroblasts derived from culture system, relative to primary cells. Because the culture system did not yield reproducible expression data in my hands, I selected primary cells as the source of primitive material for this experiment.

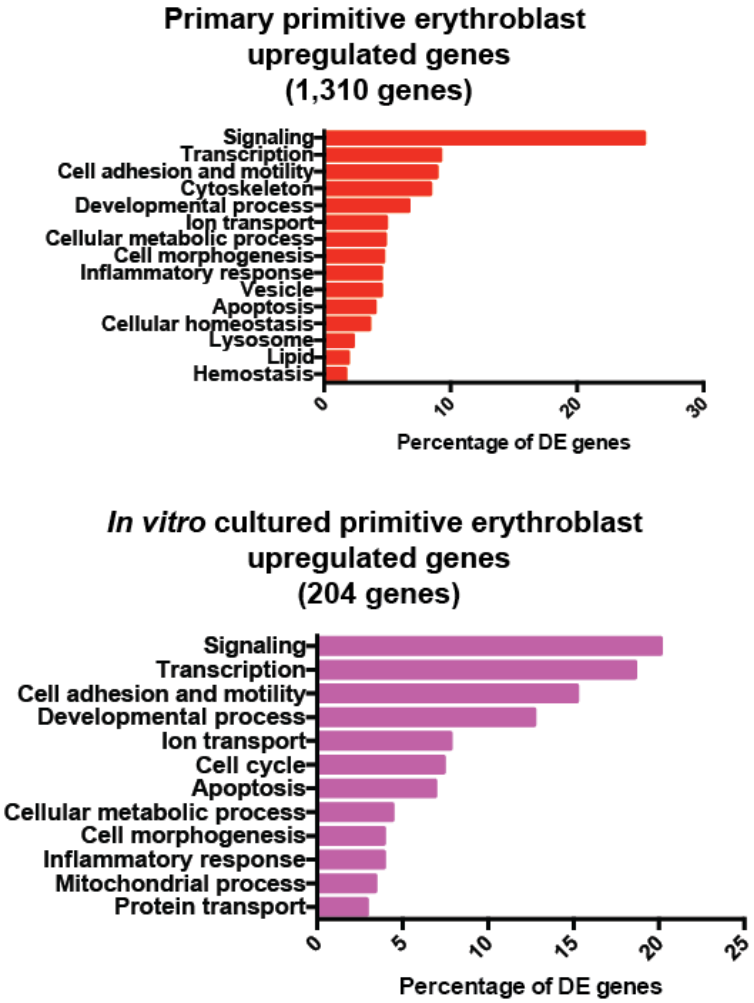


Figure 5.7 Functional annotation of DE genes between primary primitive erythroblasts and *in vitro* cultured primitive erythroblasts

Red bars demonstrate functional annotation (DAVID bioinformatics database) of genes upregulated in E10.5 yolk sac-derived erythroid cells. Pink bars show functional annotation of genes upregulated in primitive erythroblasts at the equivalent stage derived from the *in vitro* culture system.

Transcriptional profile differences between definitive erythroid cells obtained from unsorted fetal liver cell culture and CD44hi fetal liver cells that had been cultured to the intermediate erythroblast stage were investigated. Expression profiles of the two erythroid populations are drastically different (a total of 3,426 DE genes with LogFC > 2, of which 1,950 are upregulated in sorted fetal liver cell culture erythroblasts and 1,476 are upregulated in erythroblasts derived from mixed fetal liver cell culture), this is reflected in the PCA analysis (Figure 5.6). When functionally annotated using the DAVID database, RNA

samples derived from unsorted fetal liver cell culture were predominantly enriched for genes involved in mitochondrial processes (Figure 5.8). In addition, a significant proportion of genes encoding proteins with transcription factor activity are differentially expressed between the two fetal liver cell culture populations and genes involved in hematopoiesis (including erythropoiesis) and heme/iron metabolism are enriched in CD44hi sorted fetal liver cells. These findings suggest that definitive erythroblasts derived from unsorted fetal liver cell culture are too heterogeneous and may be contaminated with fetal hepatocytes and therefore should not be used for the final transcriptome comparison analysis between primitive and definitive erythroid cells.

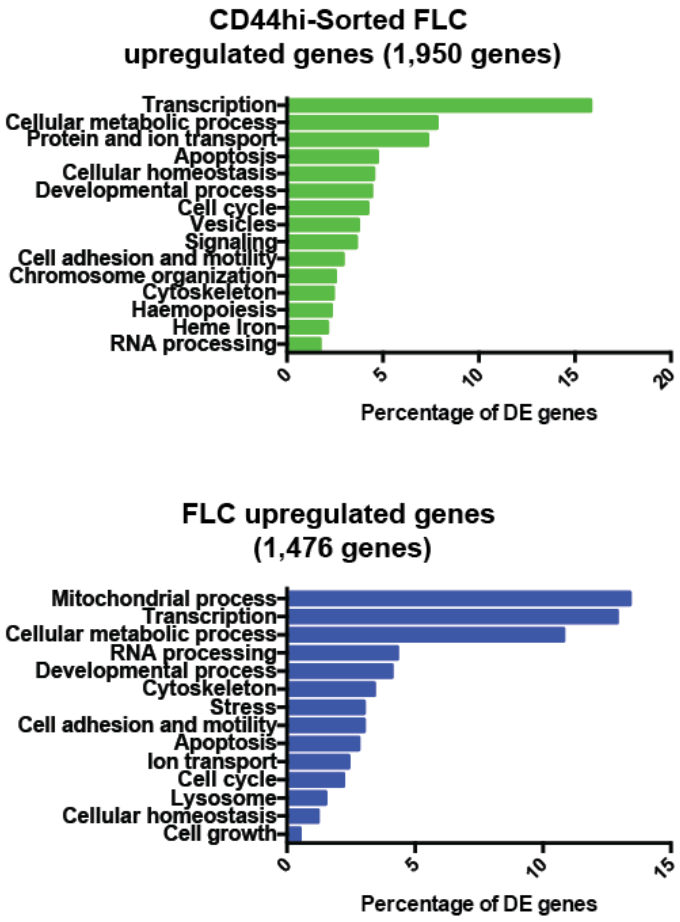


Figure 5.8 Functional annotation of DE genes between CD44hi-sorted fetal liver cultured (FLC) erythroblasts and unsorted FLC erythroblasts

Green bars represent functional annotation (the DAVID bioinformatics database) of genes upregulated in definitive erythroblasts derived from CD44hi-sorted fetal liver cell culture. Blue bars show functional annotation of genes upregulated in definitive erythroblasts derived from mixed fetal liver cell culture.

5.2.3 Global analysis of DE genes between primitive and definitive erythroblasts as determined by RNA sequencing

Comparison of the transcriptional networks between primitive and definitive erythropoiesis was performed using expression data from primary primitive erythroblasts and definitive erythroblasts derived from CD44hi fetal liver cell culture. The three biological replicates of total RNA from each erythroid lineage had undergone the same globin- and ribo-depletion process using the NuGen® Globin Depletion kit with custom oligonucleotides prior to

sequencing. The samples selected for the final analysis are shown in Figure 5.6.

Principal component analysis demonstrates that transcriptome data from the biological replicates of primitive (Figure 5.6, red dots) and CD44hi-sorted fetal liver cultured definitive (Figure 5.6, green dots) erythroid cells are each closely clustered. All DE genes with a false discovery rate (FDR, adjusted p-value) of < 0.05 are plotted in Figure 5.9A. Given the manifold differences between primitive and definitive erythroblasts including their structure and biochemistry, a large number of DE genes was expected. Greater than 4,000 DE genes (LogFC ≥ 2) were identified between these two erythroid lineages, of which 2,675 genes are upregulated in primitive erythroblasts and 1,560 genes are upregulated in definitive erythroblasts (Figure 5.9A, B). This large difference is consistent with previous studies showing that they are distinct cell types (see Section 1.3). However, comparing transcriptomes of “primary” primitive erythroblasts with “*in vitro* cultured” definitive erythroblasts from fetal liver cells may also contribute to significant number of DE genes. This is most likely owing to manifold transcriptional differences between *in vitro* cultured-derived and primary cells, including genes involve in signaling pathways, cell cycle processes and genes encoding transcription factors, similar to that identified when primary primitive erythroblasts and *in vitro* cultured primitive erythroblasts were compared (Figure 5.7). In addition, a large proportion (> 1,000) of genes were upregulated in primitive erythroblasts at logFC > 5 suggesting they may be a more complex cell type than definitive cells (Figure 5.9B). Some key genes known to be involved in globin switching in the β-globin locus such as *Bcl11a* and *Sox6* (known γ-globin repressors) show

transcriptional differences in our data as they have much higher level of expression in definitive compared to primitive erythroid cells (LogFC 3.8 and 7.3 respectively). A subset of DE genes reported in a previous study (Kingsley et al., 2013), including *Arid3a*, *Cited2*, *Cebpa*, and *Irf9* show a similar direction of differential expression; *Arid3a* and *Cited2* are enriched in primitive whereas *Cebpa* and *Irf9* are enriched in definitive erythroid cells.

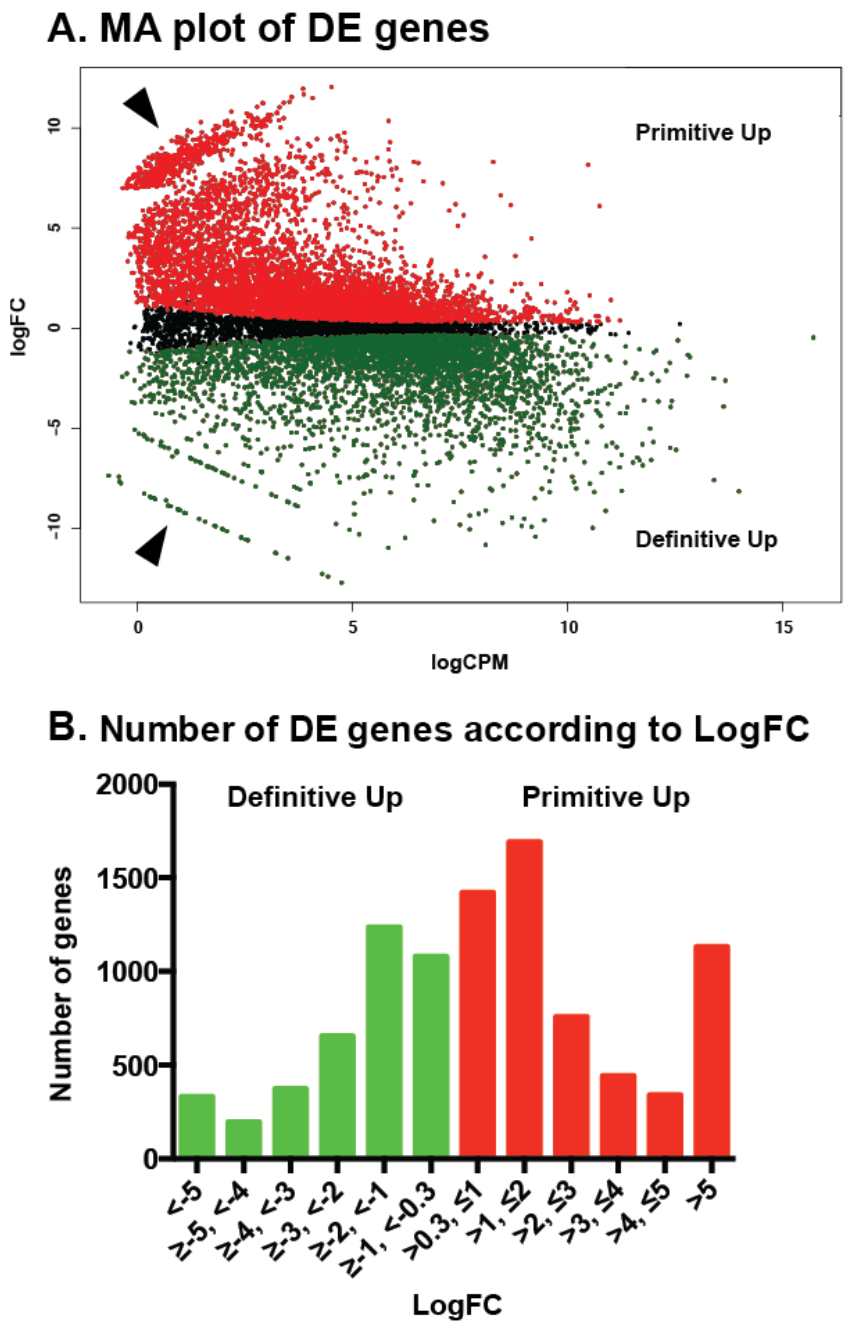


Figure 5.9A. Analysis of differentially expressed (DE) genes from the RNA-seq experiment)

A. MA plot (plot of the average (A) of the log intensities *vs.* the differences in average intensities (minus, M)) of DE genes between E10.5 primary primitive erythroblasts (n=3) and CD44hi-sorted E12.5 fetal liver cultured-definitive erythroblasts (n=3). Genes upregulated in primitive erythroblasts are shown in red and genes upregulated in definitive erythroblasts are highlighted in green. Genes lie in the left upper and left lower corner of the plot (black arrows) are those exclusively expressed in either primitive or definitive erythroid cells. B. Number of DE genes according to Log Fold Change (LogFC). Red bars represent genes upregulated in primitive and green bars represent genes upregulated in definitive erythroid cells.

5.2.4 Identification of DE genes between primitive and definitive erythropoiesis using gene expression arrays

To refine the large DE gene list from the RNA-seq analysis and identify consistently differentially expressed genes between primitive and definitive erythropoiesis for prioritization, I have performed a parallel expression analysis using Affymetrix® gene expression arrays (see methods in Chapter 2). This initial prioritization strategy is based on the hypothesis that expression of the ζ-globin gene is distinct between these two cell types, therefore genes coding for *trans*-acting factors involved in expression and silencing of ζ-globin, if transcriptionally regulated, should be identified as DE genes regardless of experimental methodology. Aliquots of the same RNA samples of primary primitive erythroblasts, CD44hi sorted fetal liver cultured definitive erythroblasts and unsorted fetal liver culture definitive erythroblasts used for the RNA-seq experiment were used in the microarray experiment. However, only expression data from the former two erythroid populations were employed for the downstream analysis. Total RNA samples were DNase-treated and RNA quality assessed using a bioanalyzer (Agilent Technologies). Depletion of globin- and ribosomal-transcripts was not performed for the microarray experiments since array methodology is not suitable for quantification of highly expressed transcripts as probes become saturated during the hybridization processes and depletion may confound accurate measurement of target transcripts. Furthermore, abundant transcripts are unlikely to affect hybridization, and therefore quantification, of rarer transcripts as occurs during RNA-sequencing where the number of informative reads is diminished by the presence of abundant transcripts. Quality control and gene

expression analysis were performed using Expression Console® and Transcriptome Analysis Console® Software (Affymetrix®).

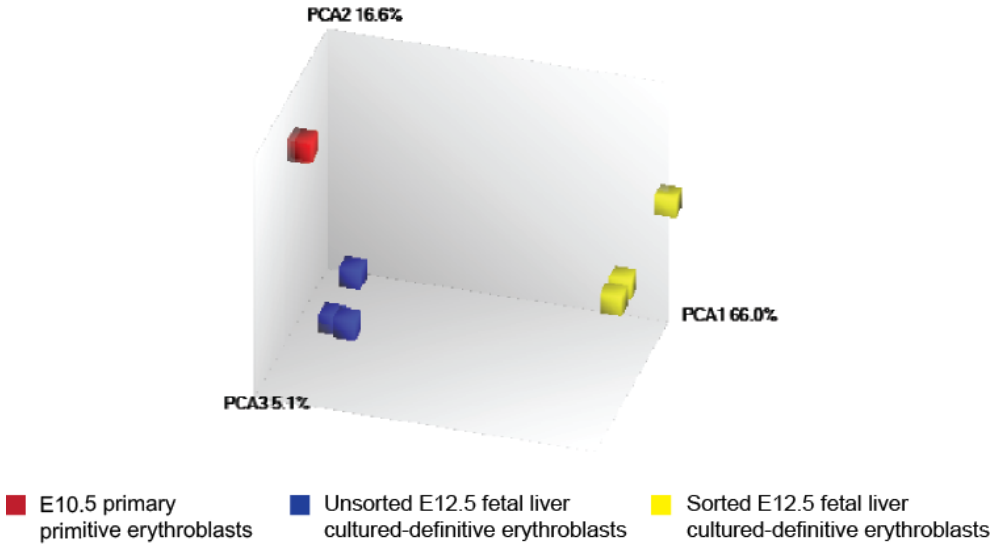


Figure 5.10 3D representation of principal component analysis from the gene expression array experiment

Expression of E10.5 primary primitive (n=3) and two groups of E12.5 fetal liver culture-derived definitive erythroblasts (n=3 each) was assayed using mouse transcriptome arrays (MTA1.0, Affymetrix®). Biological replicates of each erythroid population are closely clustered and the three types of samples are distinct. However, only expression data from primitive erythroblasts and sorted fetal liver cultured-definitive erythroblasts were employed for downstream analysis.

Similar to the RNA-seq analysis results, microarray principal component analysis demonstrates that transcriptome data from biological replicates of primitive (Figure 5.10, red rectangles) and definitive sorted fetal liver cultured (Figure 5.10, yellow rectangles) erythroid cells are each closely clustered and the two clusters are distinct. Moreover, the analysis also confirms that unsorted fetal liver cultured definitive erythroblasts (Figure 5.10, dark blue rectangles) are significantly different from sorted fetal liver cultured erythroblasts (Figure 5.10, yellow rectangles). Only transcriptome data of primary primitive and sorted fetal liver culture-derived definitive erythroid cells

obtained from the microarrays experiment were used to overlap with relevant data sets obtained from RNA-seq for further prioritization in this thesis.

Affymetrix® gene expression array analysis resulted in 3,965 probes covering 2,988 DE genes showing LogFC ≥ 2, of which 1,794 genes were upregulated in primitive and 1,194 genes were upregulated in definitive erythroid cells (Figure 5.11A, B). *Bcl11a* and *Sox6*, which are key genes known to be involved in β-globin switching, were also shown to have a large expression change in this microarray data, as they were differentially upregulated in definitive erythroblasts at LogFC of 2.9 and 8.8 respectively.

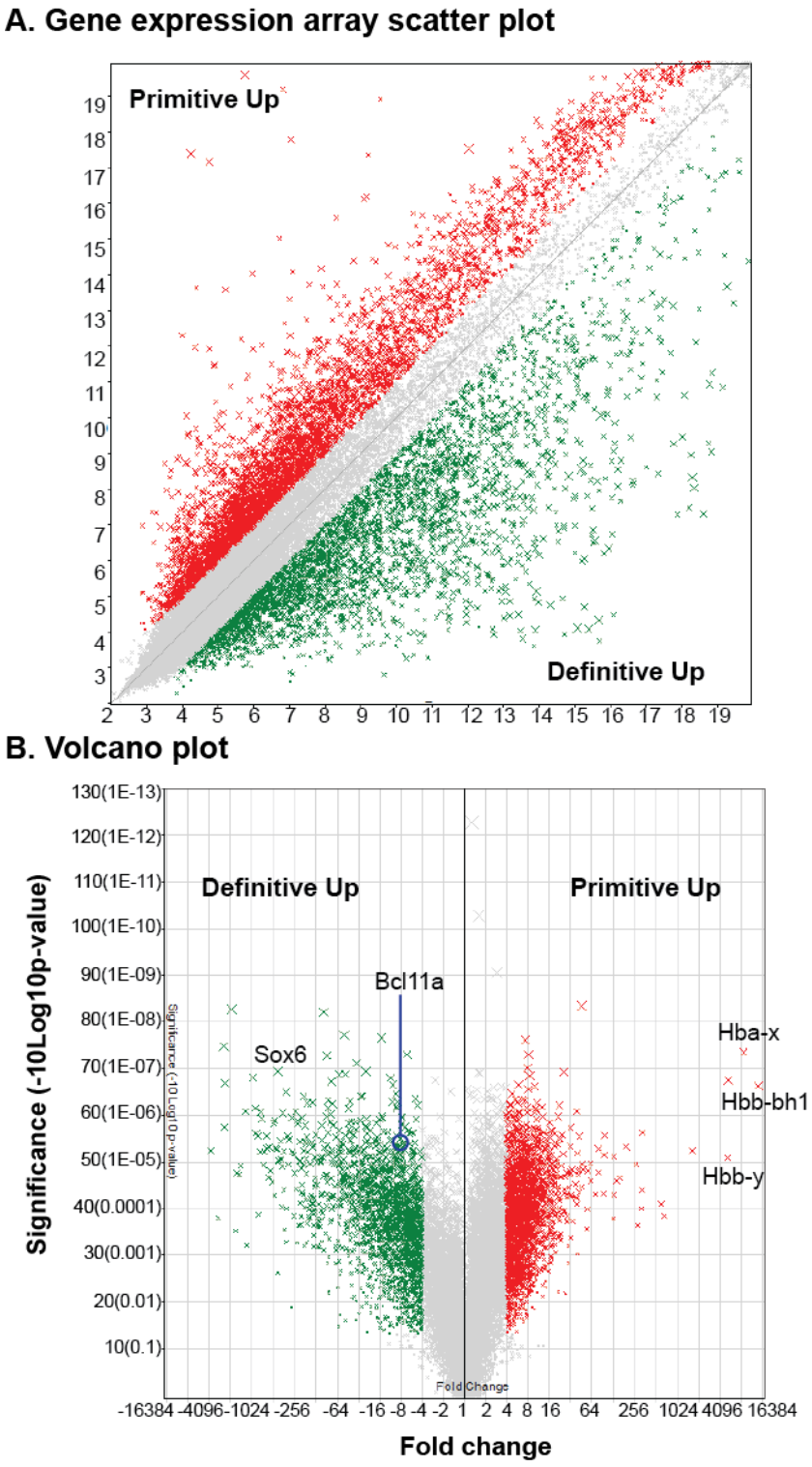


Figure 5.11 Analysis of DE genes from the gene expression array experiment

A. Scatter plot showing DE genes (Fold change > 2) between E10.5 primary primitive (n=3) and CD44hi-sorted fetal liver cultured-definitive erythroblasts (n=3). Genes upregulated in primitive erythroblasts are shown in red and genes upregulated in definitive erythroblasts are plotted in green. B. Volcano plot of DE genes showing LogFC ≥ 2 (linear fold change ≥4) in either erythroid lineage. These DE genes were used for downstream analysis. Embryonic α- (*Hba-x*) and β-like (*Hbb-y, Hbb-bh1*) globin genes are the top DE genes in primitive erythroid cells. Key genes (*Sox6* and *Bcl11a*) known to be involved in β-globin switching are also present in this dataset.

5.2.5 Prioritization of candidate genes

The aim of the prioritization scheme is to obtain a refined list of candidate genes underlying the different globin expression profiles between primitive and definitive erythropoiesis for future functional testing. To achieve this, expression data from RNA-seq and Affymetrix® expression arrays were combined. I have initially considered genes with a differential expression level of LogFC ≥ 2 (linear fold change ≥ 4), with a FDR (adjusted p-value) of < 0.05 as the cut-off point for significance. A total of 1,168 genes (355 genes upregulated in E10.5 yolk sac-derived primitive erythroblasts and 813 genes upregulated in sorted E12.5 fetal liver culture-derived definitive erythroblasts) were consistently shown to have significant level of differential expression in both datasets (Figure 5.12A). These genes are functionally annotated using the DAVID bioinformatics database as shown in Figure 5.12B. Genes encoding transcription factors are ranked the top among DE genes in definitive erythroblasts, whereas genes encoding proteins responsible for cellular metabolic processes, signaling pathways and developmental processes are more enriched in primitive erythroid cells. This finding suggests that, in the transition from primitive to definitive erythropoiesis, a number of transcription factors and signaling molecules underlie the phenotypic differences between the two lineages. Many of these are likely to be required to repress embryonically expressed genes, including the ζ-globin genes. I have found that pathway analysis using web-based tools, such as Metacore®, are not of a great benefit for this transcriptome comparison experiment, since pathways underlying globin switching are poorly defined because those governing ζ-globin expression have not previously been identified.

I have also considered whether the list of 1,168 overlapping DE genes may contain contaminant genes. Sources of contamination identified in previous studies include genes expressed at high levels in the yolk sac, hepatocytes, myeloid and lymphoid cells. These likely contaminants include 59 sentinel contaminants and 205 probable contaminants and may indicate contamination during dissection or *in vitro* culture (Kingsley et al., 2013). A total of 21 likely contaminant genes were identified in my overlapping DE gene list (Table 5.2), none of which were annotated as having transcription factor activity.

To further filter the DE genes for functional validation, I have prioritized genes, which have been annotated as having transcription factor activity. A total of 2,935 genes annotated as encoding mouse transcription factors were obtained from the Riken Transcription Factor Database (http://genome.gsc.riken.jp/TFdb/), the Animal Transcription Factor Database (AnimalTFDB, (Zhang et al., 2012)), and the Molecular Signatures Database (MSigDB, (Liberzon, 2014)), 146 of which were identified in the list of significantly DE genes in the overlapping data set.

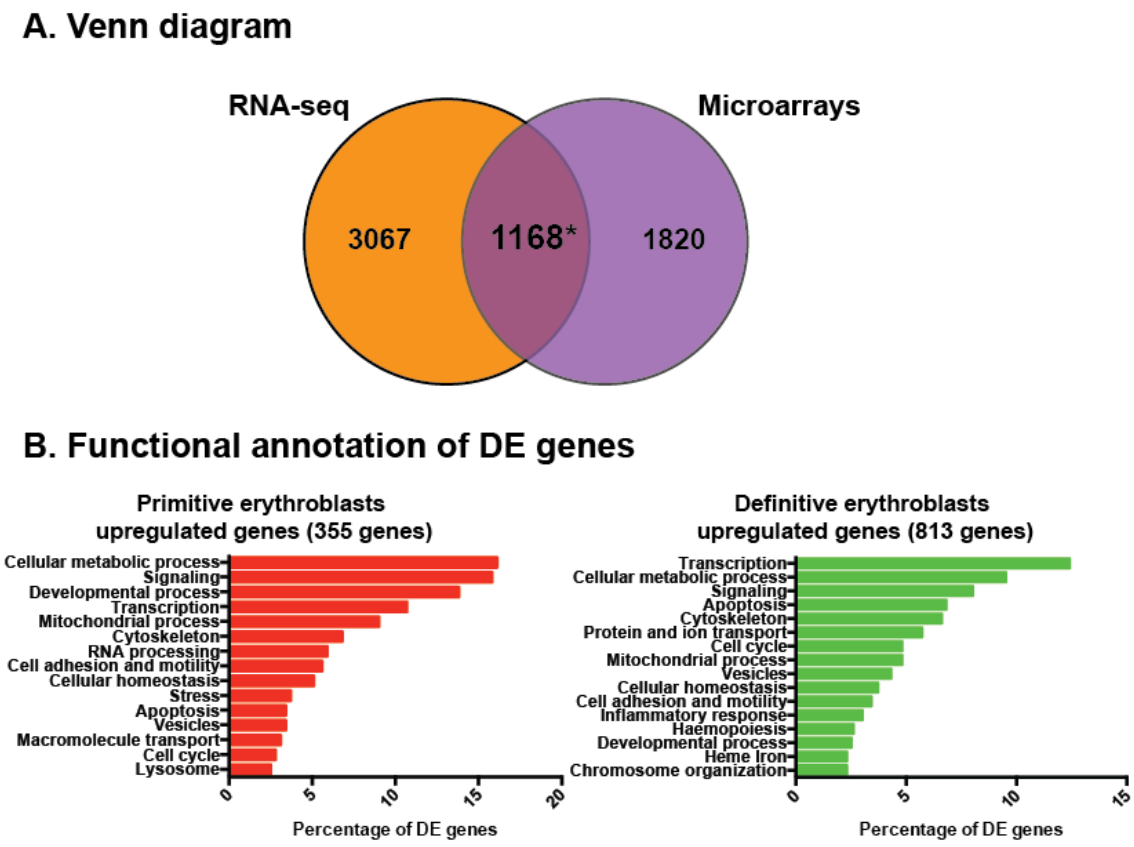


Figure 5.12 Robustly differentially expressed gene identified from the RNA-seq and gene expression array experiments

A. Venn diagram showing the number of significantly DE genes (LogFC ≥ 2) between E10.5 yolk sac-derived primary primitive (n=3) and CD44hi-sorted E12.5 fetal liver cultured-definitive erythroblasts (n=3) from RNA-seq (orange) and microarray (purple) datasets. The number of DE genes identified from both datasets is indicated with an asterisk. B. Functional annotation of overlapping DE genes. Genes upregulated in primitive erythroblasts are shown with red bar graphs and genes upregulated in definitive erythroblasts are shown with green bar graphs. Functional annotation was performed using the DAVID bioinformatics database.

Gene symbol	Gene name	FDR	logFC
Afp	Alpha fetoprotein	9.27E-300	8.256739026
Apoa1	Apolipoprotein A-I	1.85E-126	7.917459433
Apoa2	Apolipoprotein A-II	8.51E-48	7.502389401
Apob	Apolipoprotein B	1.93E-221	7.928810745
Cubn	Cubilin (intrinsic factor-cobalamin receptor)	3.1E-185	7.56614369
Dock2	Dedicator of cyto-kinesis 2	1.1E-70	-4.640990245
Fgg	Fibrinogen gamma chain	1.28E-31	7.022089736
Glipr1	GLI pathogenesis-related 1 (glioma)	1.66E-09	-3.495571847
Gpc3	Glypican 3	2.37E-105	6.88189599
Il7r	Interleukin 7 receptor	3.22E-11	-5.510677516
Lcn2	Lipocalin 2	1.5E-16	-12.27181409
Lgals2	Lectin, galactose-binding, soluble 2	2.98E-38	7.877973009
Ngp	Neutrophilic granule protein	8.16E-14	-10.29751544
Nrk	Nik related kinase	7.48E-95	7.65648573
Pcbd1	Pterin 4 alpha carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor 1 alpha (TCF1) 1	1.84E-30	7.657935736
Rbp4	Retinol binding protein 4, plasma	1.7E-135	8.627000237
Rian	RNA imprinted and accumulated in nucleus	1.92E-111	8.880399559
S100a9	S100 calcium binding protein A9 (calgranulin B)	3.49E-13	-7.8429168
S100g	S100 calcium binding protein G	1E-70	9.384121111
Tnfrsf13c	Tumor necrosis factor receptor superfamily, member 13c	9.86E-38	-3.971016524
Ttr	Transthyretin	6.24E-137	7.785289312

Table 5.2 Overlapping DE genes likely to be contaminants

The table shows overlapping DE genes from this study that are likely to be contaminants as identified in the previous study by Kingsley and colleagues (Kingsley et al., 2013). These genes are shown in alphabetical order and were excluded from the current study. Positive values of LogFC indicate the level of enrichment in primitive erythropoiesis samples and negative values denote enrichment in definitive erythropoiesis samples.

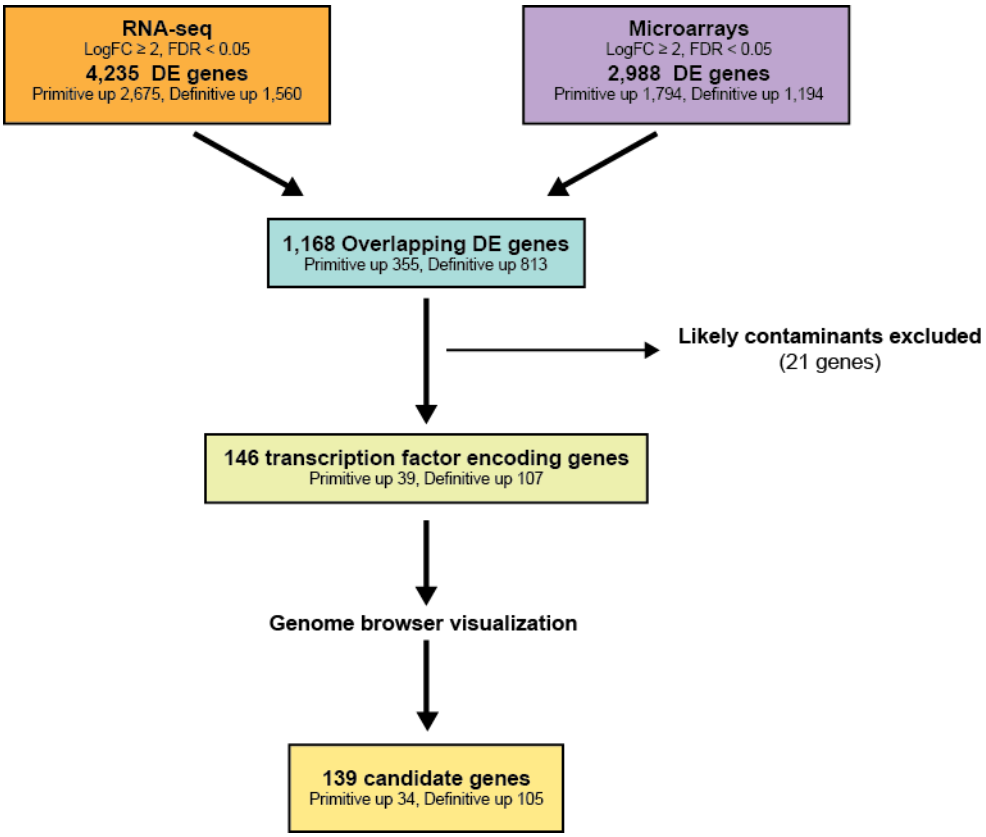


Figure 5.13 Candidate genes prioritization pipeline

Of the 146 transcription factor-encoding genes, 39 are upregulated in primitive and 107 are upregulated in definitive erythroid cells. When a gene is not at all expressed or expressed at a very low level in one dataset, it is often interpreted *in silico* that it's expression level is significantly different. Therefore, I have confirmed differential expression of the individual candidate genes by visualizing their expression and their active promoters individually on a UCSC genome browser session created from the RNA-seq and DNaseI/ATAC experiments. Five of the 39 "Primitive upregulated" genes were filtered out, resulting in 34 remaining genes in this list (Table 5.3). Two genes were filtered out from the "Definitive upregulated" list, resulting in 105 remaining genes (Table 5.4). The candidate gene prioritization pipeline is summarized in Figure 5.13.

Gene symbol	Gene name	FDR	logFC
Aatf	Apoptosis antagonizing transcription factor	5.95E-36	2.548882802
Arid3a	AT rich interactive domain 3A (BRIGHT-like); AT rich interactive domain 3A (Bright like); AT rich interactive domain 3A (BRIGHT-like) (Arid3a), mRNA.	1.98E-213	5.106160542
Atm	Ataxia telangiectasia mutated; ataxia telangiectasia mutated homolog (human)	1.03E-47	2.498985423
Bcor	BCL6 interacting corepressor; RIKEN cDNA 2900008C10 gene; Bcl6 interacting corepressor; Novel transcript	4.51E-39	2.081959365
Cdk4	Cyclin-dependent kinase 4	4.46E-53	2.387910858
Cited2	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2	2.78E-21	2.053865548
Cited4	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 4	4.61E-58	2.567909092
Csrp2	Cysteine and glycine-rich protein 2	1.72E-36	4.336329675
Gtf2i	General transcription factor II I	4.86E-57	2.49523149
Hdac11	Histone deacetylase 11	9.46E-41	3.25435444
Hesx1	Homeobox gene expressed in ES cells	9.30E-67	8.398983188
Hif1a	Hypoxia inducible factor 1, alpha subunit	1.52E-38	2.365047927
Hmga2	High mobility group AT-hook 2	6.51E-113	6.053440927
Id2	Inhibitor of DNA binding 2	6.25E-09	2.542969297
Kat2a	K (lysine) acetyltransferase 2A	9.32E-34	2.529723642
Maf	Avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog	8.92E-40	5.78188484
Maged1	Melanoma antigen, family D, 1	5.04E-22	2.394016115
Mybbp1a	MYB binding protein (P160) 1a	1.93E-77	2.626691685
Myc	Myelocytomatosis oncogene	2.80E-50	2.634335472
Ncoa3	Nuclear receptor coactivator 3	1.52E-42	2.248450134
Ndn	Necdin	1.17E-48	10.73211632
Peg3	Paternally expressed 3	7.51E-63	2.854351741
Snai2	Snail family zinc finger 2	8.17E-26	7.071097973
Sox11	SRY (sex determining region Y)-box 11; SRY-box containing gene 11	3.84E-167	9.300699345
Sox4	SRY (sex determining region Y)-box 4	3.14E-43	5.930073884
Tada2a	Transcriptional adaptor 2A; transcriptional adaptor 2 (ADA2 homolog, yeast)-like	9.89E-18	2.324321693
Tcf7l2	Transcription factor 7-like 2, T-cell specific, HMG-box	8.95E-87	4.733141923
Thra	Thyroid hormone receptor alpha	4.02E-63	3.114467139
Ttll4	Tubulin tyrosine ligase-like family, member 4	2.40E-30	2.158910764
Ybx1	Y box protein 1	1.64E-53	2.222051414
Zdhhc21	Zinc finger, DHHC domain containing 21	1.26E-63	3.520293241
Zfp568	Zinc finger protein 568	4.69E-93	3.97507505
Zmiz1	Zinc finger, MIZ-type containing 1	2.15E-73	2.669776454
Znhit6	Zinc finger, HIT type 6	4.60E-28	2.416206584

Table 5.3 Candidate transcription factor encoding genes upregulated in primitive erythroblasts

Total of 34 candidate genes upregulated in primitive erythroblasts are shown in alphabetical order.

Gene Symbol	Description	FDR	logFC
Abtb1	Ankyrin repeat and BTB (POZ) domain containing 1	2.48E-150	-5.849892799
Aff1	AF4/FMR2 family, member 1	1.72E-63	-3.256750132
Arid4b	AT rich interactive domain 4B (RBP1-like)	6.14E-20	-2.74638359
Arih2	Ariadne homolog 2 (Drosophila); predicted gene, 26298	4.03E-38	-2.445586957
Asb1	Ankyrin repeat and SOCS box-containing 1	1.79E-58	-3.191336891
Atf3	Activating transcription factor 3	6.16E-100	-7.843435793
Atf5	Activating transcription factor 5	1.51E-41	-3.682017553
Atf6	Activating transcription factor 6	3.52E-43	-2.743124384
Bach1	BTB and CNC homology 1	6.29E-64	-3.535951688
Baz2a	Bromodomain adjacent to zinc finger domain, 2A	2.47E-47	-2.512978883
Bbx	Bobby sox homolog (Drosophila)	8.48E-13	-2.553987902
Bcl11a	B cell CLL/lymphoma 11A (zinc finger protein); B-cell CLL/lymphoma 11A (zinc finger protein)	9.69E-51	-3.821972534
Carhsp1	Calcium regulated heat stable protein 1	8.16E-93	-4.368435625
Cbfa2t3	Core-binding factor, runt domain, alpha subunit 2, Translocated to, 3 (human)	5.38E-59	-3.770273116
Cdkn2c	Cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	6.97E-40	-3.036986528
Cdr2	Cerebellar degeneration-related 2	1.81E-117	-4.498120174
Cebpa	CCAAT/enhancer binding protein (C/EBP), alpha	1.78E-21	-4.694324845
Cebpg	CCAAT/enhancer binding protein (C/EBP), gamma	4.23E-43	-2.925056483
Chd2	Chromodomain helicase DNA binding protein 2; RIKEN cDNA 1810026B05 gene; novel transcript overlapping Chd2	9.86E-57	-2.993329889
Cir1	Corepressor interacting with RBPJ, 1	2.92E-44	-3.582876831
Crebbp	CREB binding protein	6.62E-53	-2.640898521
Creg1	Cellular repressor of E1A-stimulated genes 1	2.55E-76	-3.267322237
Crem	cAMP responsive element modulator; cAMP responsive element modulator (Crem), transcript variant 9, mRNA.	8.32E-32	-7.074566934
Crip2	Cysteine rich protein 2	4.84E-52	-3.685656029
Ddit3	DNA-damage inducible transcript 3	1.44E-56	-5.601050856
Dedd2	Death effector domain-containing DNA binding protein 2	4.59E-102	-4.532482756
Dennd4a	DENN/MADD domain containing 4A	1.50E-215	-6.848009549
E2f2	E2F transcription factor 2	2.09E-88	-4.212382957
Egr1	Early growth response 1	2.42E-39	-3.693178729
Ell	Elongation factor RNA polymerase II	1.31E-40	-2.734362579
Ep300	E1A binding protein p300	4.52E-39	-2.274117614
Ezh1	Enhancer of zeste homolog 1 (Drosophila)	1.85E-66	-3.666616481
Fosl1	Fos-like antigen 1	5.34E-42	-6.083104841
Foxo3	Forkhead box O3	1.31E-114	-4.447060032
Hbp1	High mobility group box transcription factor 1	1.51E-137	-5.342550671
Htatip2	HIV-1 tat interactive protein 2, homolog (human)	7.83E-136	-5.10752788
Ifnar2	Interferon (alpha and beta) receptor 2	1.29E-72	-3.43668534
Irf9	Interferon regulatory factor 9	8.80E-44	-3.283015561
Jun	Jun proto-oncogene; Jun oncogene	1.63E-93	-5.246292758
Jund	Jun D proto-oncogene	3.67E-90	-4.213612176
Kdm5b	Lysine (K)-specific demethylase 5B	6.77E-58	-3.753735001
Klf3	Kruppel-like factor 3 (basic)	1.91E-76	-3.31116586
Klf6	Kruppel-like factor 6	2.34E-114	-5.912430357
Klf7	Kruppel-like factor 7 (ubiquitous)	6.10E-42	-4.583774204
Klf9	Kruppel-like factor 9	2.83E-55	-5.645880203
Lcor	Ligand dependent nuclear receptor corepressor	1.64E-45	-3.224021857
Lmo2	LIM domain only 2	1.26E-57	-3.235800188
Mafg	V-maf musculoaponeurotic fibrosarcoma oncogene family, protein G (avian)	8.22E-65	-3.613648021
Mafk	V-maf musculoaponeurotic fibrosarcoma oncogene family, protein K (avian)	2.49E-133	-4.994197297
Mbnl2	Muscleblind-like 2	3.48E-25	-3.120898616
Med7	Mediator complex subunit 7	2.62E-40	-2.51263941
Mef2d	Myocyte enhancer factor 2D	1.41E-57	-3.659751411

Table 5.4 Candidate transcription factor encoding genes upregulated in definitive erythroblasts

Total of 105 candidate genes upregulated in definitive erythroblasts are shown in alphabetical order. The level of enrichment is shown in negative value of LogFC.

Gene symbol	Gene name	FDR	LogFC
Mier1	Mesoderm induction early response 1 homolog (Xenopus laevis; mesoderm induction early response 1 homolog (Xenopus laevis (Mier1), transcript variant 1, mRNA.	5.15E-37	-2.704677492
Mier3	Mesoderm induction early response 1, family member 3	9.19E-32	-2.251240484
Mtf1	Metal response element binding transcription factor 1	2.62E-39	-2.396529754
Mxd1	MAX dimerization protein 1	4.69E-279	-8.12362846
Mxi1	Max interacting protein 1	7.43E-84	-4.303247879
Nfe2	Nuclear factor, erythroid derived 2	7.64E-90	-3.889753673
Nfkbia	Nuclear factor of kappa light polypeptide gene enhancer in B cells inhibitor, alpha	1.03E-24	-5.22433942
Nr3c1	Nuclear receptor subfamily 3, group C, member 1	4.84E-33	-3.942434668
Nrip1	Nuclear receptor interacting protein 1	3.33E-33	-2.759831983
Optn	Optineurin	1.64E-133	-5.530137358
Ostf1	Osteoclast stimulating factor 1; Osteoclast Stimulating Factor 1	1.07E-84	-3.848350236
Otud7b	OTU domain containing 7B	7.50E-59	-3.11226803
Per1	Period circadian clock 1; period homolog 1 (Drosophila)	8.97E-153	-5.851303791
Phf8	PHD finger protein 8	6.80E-46	-2.776660419
Ppard	Peroxisome proliferator activator receptor delta; RIKEN cDNA 1810013A23 gene	7.37E-69	-3.845322356
Ppp1r10	Protein phosphatase 1, regulatory subunit 10; microRNA 1894	2.09E-32	-2.226784579
Ppp1r12c	Protein phosphatase 1, regulatory (inhibitor) subunit 12C	5.83E-73	-3.553264043
Psmd10	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 10; proteosome (prosome, macropain) 26S subunit, non-ATPase, 10	9.70E-36	-2.363449244
Rab11a	RAB11A, member RAS oncogene family; RAB11a, member RAS oncogene family	1.49E-39	-2.393188755
Rab8b	RAB8B, member RAS oncogene family	4.22E-42	-2.424416503
Rb1	Retinoblastoma 1	4.15E-77	-4.084293247
Rfx2	Regulatory factor X, 2 (influences HLA class II expression)	4.97E-61	-3.637835568
Rnf10	Ring finger protein 10	2.51E-68	-3.233614604
Rnf114	Ring finger protein 114	9.67E-63	-3.60556702
Rnf115	Ring finger protein 115	7.16E-63	-3.755724357
Runx1	Runt related transcription factor 1; runt related transcription factor 1 (Runx1), transcript variant 2, mRNA.	6.52E-105	-4.398135185
Sox6	SRY (sex determining region Y)-box 6; SRY-box containing gene 6 (Sox6), transcript variant 2, mRNA.; SRY-box containing gene 6	6.22E-145	-7.259825003
Sqstm1	Sequestosome 1	9.64E-63	-4.611188134
Ssbp2	Single-stranded DNA binding protein 2	4.40E-57	-3.55410519
Tceb3	Transcription elongation factor B (SIII), polypeptide 3	5.80E-43	-2.71256881
Tfb2m	Transcription factor B2, mitochondrial	6.94E-35	-2.579103108
Tfdp2	Transcription factor Dp 2	9.34E-88	-3.820365265
Tnfaip3	Tumor necrosis factor, alpha-induced protein 3	8.67E-12	-5.087568661
Trib3	Tribbles homolog 3 (Drosophila)	5.77E-25	-3.032348271
Trim10	Tripartite motif-containing 10	1.30E-62	-3.061642835
Trim25	Tripartite motif-containing 25; tripartite motif protein 25; tripartite motif-containing 25 (Trim25), mRNA.	3.71E-51	-3.20409407
Tsc22d2	TSC22 domain family, member 2	4.46E-47	-2.704631488
Tsc22d3	TSC22 domain family, member 3; TSC22 domain family 3	2.17E-82	-4.921323354
Ubr4	Ubiquitin protein ligase E3 component n-recognin 4	4.89E-33	-2.134646832
Xbp1	X-box binding protein 1; X-box binding protein 1 (Xbp1), mRNA.	8.40E-44	-2.779693691
Zbtb46	Zinc finger and BTB domain containing 46	3.74E-37	-4.2339543
Zfand5	Zinc finger, AN1-type domain 5	3.39E-108	-4.102718083
Zfand6	Zinc finger, AN1-type domain 6	1.15E-73	-3.518827392
Zfp217	Zinc finger protein 217	2.93E-79	-4.668578109
Zfp36	Zinc finger protein 36	4.88E-43	-4.791822739
Zfp691	Zinc finger protein 691	3.97E-72	-4.157026751
Zfp869	Zinc finger protein 869	7.43E-30	-2.327103057
Zfp874b	Zinc finger protein 874b	7.60E-44	-3.575211419
Zfp945	Zinc finger protein 945; Novel protein (A630033E08Rik)	2.28E-07	-3.135718533
Zfp949	Zinc finger protein 949; novel protein (4930422I07Rik)	1.24E-80	-4.605282173
Zkscan14	Zinc finger with KRAB and SCAN domains 14	4.67E-58	-4.371389293
Zmym5	Zinc finger, MYM-type 5	1.29E-33	-3.151535235
Zswim4	Zinc finger SWIM-type containing 4	1.34E-83	-5.273872138

Table 5.4 Candidate transcription factor encoding genes upregulated in definitive erythroblasts (continued)

For further prioritization, the 139 candidate transcription factor-encoding genes (Table 5.3, 5.4) were ranked by FDR (adjusted p-value) and subsequently by the level of differential expression. The top 20% of candidate genes showing greatest expression difference in either primitive or definitive erythroid cells are described in detail in Table 5.5 and 5.6.

Gene symbol	Gene name	FDR	logFC	Known functions	Homozygous knockout phenotypes	Related human disorders
Arid3a	AT rich interactive domain 3A (BRIGHT-like)	1.98E-213	5.106160542	Embryo patterning, control of cell cycle progression	Embryonic lethality between E11.5 andE13.5 due to impaired erythropoiesis	Unknown
Sox11	SRY (sex determining region Y)-box 11	3.84E-167	9.300699345	Embryonic neurogenesis, tissue modelling during development	Neonatal lethality with impaired ossification, absent spleen, abnormal neural system development	Mental retardation (autosomal dominant), Coffin-Siris syndrome
Hmga2	High mobility group AT-hook 2	6.51E-113	6.053440927	Cell cycle regulation, chromosome condensation during meiosis of spermatocytes, adipogenesis	Proportionate dwarfing with a significant reduction in body weight, reduced fat tissue	Leiomyoma, 12q14 Microdeletion syndrome
Zfp568	Zinc finger protein 568	4.69E-93	3.97507505	Unknown	Unknown	Unknown
Tcf7l2	Transcription factor 7-like 2, T-cell specific, HMG-box	8.95E-87	4.733141923	Participates in the Wnt signalling pathway and modulates MYC expression and blood glucose homeostasis	Die shortly after birth due to intestinal epithelia abnormalities	Pitt-Hopkins syndrome, tropical calcific pancreatitis
Mybbp1a	MYB binding protein (P160) 1a	1.93E-77	2.626691685	Binds to the Myb proto-oncogene protein, nucleolar stress, tumor suppression, synthesis of ribosomal DNA	Embryonic lethality before blastocyst formation	Brain glioma
Zmiz1	Zinc finger, MIZ-type containing 1	2.15E-73	2.669776454	Regulates the activity of various transcription factors, including the androgen receptor, Smad3/4, and p53, may play a role in sumoylation	Embryonic lethality during organogenesis with failure of yolk sac vascular remodelling and abnormal embryonic vascular development	Lymphoblastic leukemia

Table 5.5 Details of top 20% of candidate transcription factor encoding genes upregulated in primitive erythroblasts

Candidate genes are ranked by FDR, and subsequently by level of differential expression (LogFC). Summary of gene functions and related human disorders were derived from GeneCards® Human gene database (http://www.genecards.org/). Descriptions of homozygous knockout phenotypes were obtained from the Mouse Genome Informatics (MGI) database (http://www.informatics.jax.org/).

Gene symbol	Gene name	FDR	logFC	Known functions	Homozygous knockout phenotypes	Related human disorders
Mxd1	MAX dimerization protein 1	4.69E-279	-8.12362846	Mediates cellular proliferation, differentiation, and apoptosis. Antagonizes MYC-mediated transcriptional activation of target genes	Altered myelopoiesis, increased proliferative potential of bone marrow granulocytic precursors	Acute leukemia
Dennd4a	DENN/MADD domain containing 4A	1.50E-215	-6.848009549	Promotes the exchange of GDP to GTP, converting inactive GDP-bound Rab proteins into their active GTP-bound form	Unknown	Unknown
Per1	Period circadian clock 1; period homolog 1 (Drosophila)	8.97E-153	-5.851303791	Transcriptional repressor that forms a core component of the circadian clock. The circadian clock, may relate to HIF1Alpha-pathway	Persistence of circadian rhythm, but with a shorter period and impaired ability to maintain the precision and the stability of the period	Advanced and delayed sleep phase syndrome
Abtb1	Ankyrin repeat and BTB (POZ) domain containing 1	2.48E-150	-5.849892799	May act as a mediator of the PTEN growth-suppressive signalling pathway. May play a role in developmental processes.	Unknown	Unknown
Sox6	SRY (sex determining region Y)-box 6	6.22E-145	-7.259825003	A transcriptional activator that is required for normal development of the central nervous system, chondrogenesis and maintenance of cardiac and skeletal muscle cells, play a role in gamma globin suppression in definitive erythropoiesis	Cardioskeletal myopathy, cardiac blockage, delayed growth, and early postnatal lethality.	Unknown
Hbp1	High mobility group box transcription factor 1	1.51E-137	-5.342550671	Plays a role in the regulation of the cell cycle and of the Wnt pathway	Heterozygotes for a gene trapped allele have behavioural abnormalities and show alterations in leukocyte, platelet and natural killer (NK) cell number, blood urea nitrogen levels, and circulating amylase and calcium levels	Ureter cancer
Htatip2	HIV-1 tat interactive protein 2, homolog (human)	7.83E-136	-5.10752788	Oxidoreductase required for tumor suppression, may act as a redox sensor linked to transcription through regulation of nuclear import	Inactivation of this gene increases susceptibility to tumorigenesis	Gnathodiaphyseal dysplasia, HIV-1
Optn	Optineurin	1.64E-133	-5.530137358	Maintenance of the golgi complex, membrane trafficking, exocytosis, neuroprotective role in the eye and optic nerve, cell cycle	Unknown	Amyotropic lateral sclerosis, glaucoma
Mafk	V-maf musculoaponeurotic fibrosarcoma oncogene family, protein K (avian)	2.49E-133	-4.994197297	Maf protein (MafF, MafG, or MafK) and p45 are subunits of NFE2 DNA-binding activity. Maf homodimers suppress transcription at NFE2 sites	Homozygotes for a knock-out allele are viable, fertile, healthy and phenotypically normal with no detectable erythroid deficiencies	Fibrosarcoma
Cdr2	Cerebellar degeneration-related 2	1.81E-117	-4.498120174	Not clearly understood	Viable, fertile, overtly normal	Gait apraxia, paraneoplastic neurologic disorders

Table 5.6 Details of top 20% of candidate transcription factor encoding genes upregulated in definitive erythroblasts

Candidate genes are ranked by FDR and subsequently by the level of differential expression (LogFC). Summary of gene functions and related human disorders were derived from the GeneCards® Human gene database (http://www.genecards.org/). Descriptions of homozygous knockout phenotypes were obtained from the Mouse Genome Informatics (MGI) database (http://www.informatics.jax.org/).

Gene symbol	Gene name	FDR	logFC	Known functions	Homozygous knockout phenotypes	Related human disorders
Foxo3	Forkhead box O3	1.31E-114	-4.447060032	A trigger of apoptosis in the absence of survival factors, including neuronal cell death upon oxidative stress.	Ovarian defects leading to female sterility	Acute leukemia, rhabdomyosarcoma
Klf6	Kruppel-like factor 6	2.34E-114	-5.912430357	Tumor suppressor activity, could play a role in B-cell growth and development	Embryonic lethality during organogenesis, small size, pallor, decreased cellular proliferation and delayed liver development	Gastric cancer, global amnesia
Zfand5	Zinc finger, AN1-type domain 5	3.39E-108	-4.102718083	Involved in protein degradation via the ubiquitin-proteasome system, plays a role in the regulation of NF-kappa-B activation and apoptosis	Resistance to age related obesity and resistance to induced muscular atrophy	Unknown
Runx1	Runt related transcription factor 1; runt related transcription factor 1 (Runx1), transcript variant 2, mRNA.	6.52E-105	-4.398135185	The development of normal haematopoiesis	Mutations affect hematopoiesis, and in some cases result in defective angiogenesis and intraventricular hemorrhage. Null homozygotes die by embryonic day 12.5; heterozygotes have reduced erythroid and myeloid progenitor numbers	Familial platelet disorder with associated myeloid malignancy
Dedd2	Death effector domain-containing DNA binding protein 2	4.59E-102	-4.532482756	May play a critical role in death receptor-induced apoptosis and may target CASP8 and CASP10 to the nucleus	Unknown	Unknown
Atf3	Activating transcription factor 3	6.16E-100	-7.843435793	Involved in the complex process of cellular stress response	Enhanced allergen-induced airway hyperresponsiveness, pulmonary eosinophilia, and chemokine and Th2 cytokine responses in lung tissue and lung-derived CD4+ lymphocytes	Hypospadias
Jun	Jun proto-oncogene; Jun oncogene	1.63E-93	-5.246292758	Involves PI-3K cascade, MAPK signaling pathway, involves steroidogenic gene expression upon cAMP signaling pathway stimulation	Die at midgestation with impaired hepatogenesis, altered fetal liver erythropoiesis and edema	Hemorrhagic cystitis, sarcoma
Carhsp1	Calcium regulated heat stable protein 1	8.16E-93	-4.368435625	Binds mRNA and regulates the stability of target mRNA	Unknown	Unknown
Jund	Jun D proto-oncogene	3.67E-90	-4.213612176	A functional component of the AP1 transcription factor complex, protect cells from p53-dependent senescence and apoptosis	Reduced growth, sensitivity to LPS-induced hepatitis, and male infertility due to hormonal imbalance and impaired spermatogenesis	Human T-cell leukemia virus type 1, breast cancer
Nfe2	Nuclear factor, erythroid derived 2	7.64E-90	-3.889753673	Essential for regulating erythroid and megakaryocytic maturation and differentiation, may play a role in all aspects of hemoglobin production from globin and heme synthesis to procurement of iron	Homozygotes for a targeted null mutation lack platelets and most die as neonates from internal bleeding. Survivors exhibit hypochromia, reticulocytosis, and splenomegaly.	Essential thrombocythemia
E2f2	E2F transcription factor 2	2.09E-88	-4.212382957	Plays a crucial role in the control of cell cycle and action of tumor suppressor proteins, binds specifically to RB1 in a cell-cycle dependent manner	Premature death with signs of inflammatory and autoimmune disorders such as increased memory T cells, enlarged spleen, glomerulonephritis, inflamed liver, inflamed lung, hair loss, and erythema	Unknown

Table 5.6 Details of top 20% of candidate transcription factor encoding genes upregulated in definitive erythroblasts (continued)

All candidates listed in Tables 5.5 and 5.6 are genes encoding potential *trans*-acting factors responsible for ζ-globin expression in primitive erythropoiesis and repression in definitive erythropoiesis. Functional validation of these genes will be done systematically using a Cas9/CRISPR library system in a separate future project (see Section 7.7).

5.3 Discussion

An initial key step to identify novel *trans*-acting factors regulating the ζ-globin gene is to select the most appropriate erythroid populations, which best represent primitive erythropoiesis with ζ-globin expression and definitive erythropoiesis repressing ζ-globin. I have demonstrated here that ζ-globin is expressed at a significant level (~40% of total α-like globin transcripts) in primary erythroblasts derived from E10.5 yolk sacs and the proportion of ζ-globin transcripts is relatively consistent across the biological replicates. In addition, an adequate number of primitive erythroblasts (~1 x 10^6 cells per embryo) was obtained giving rise to a sufficient amount of total RNA (~1μg per embryo) at this developmental stage. We had initially performed primitive erythroid culture up to the equivalent maturational stage of E10.5 with two main aims: 1) to expand the number of primitive erythroblasts and 2) to establish a model system for future functional validation of candidate genes. In our hands, the number of primitive erythroblasts obtained from 2-day culture of E8.5 erythroblasts was fewer than those obtained from yolk sacs at the equivalent stage (E10.5). The maturational stage of primitive erythroblasts from the culture system was not as synchronous as that found in primary primitive erythroblasts. The level of ζ-globin transcripts also varied across biological replicates of cultured erythroblasts. Consistent with these findings,

expression data of *in vitro* cultured primitive erythroblasts are not reproducible across biological replicates and are significantly different from those of primary primitive erythroblasts in the RNA-seq datasets. This suggests that the culture system does not accurately recapitulate *in vivo* developmental processes. However, the established primitive erythroid culture system adapted from previous studies (Greenfest-Allen et al., 2013) may be useful as a model system to functionally test candidate genes, at least to assess changes in level of ζ-globin expression when genes encoding candidate ζ-globin activators are deleted.

An alternative source to obtain primary definitive erythroblasts may be cells isolated from murine bone marrow. However, pilot experiments involving sorting for $CD44^+$, $CD71^+$ and $Ter119^+$ cells from bone marrow of two mice resulted in fewer than 0.5 $x10^6$ intermediate erythroblasts (data not shown). In our experimental design, total RNA obtained from erythroid cells required globin- and ribo-depletion, resulting in significant loss of initial material for RNA-seq. To avoid amplification bias, I aimed to use at least 1-2 µg of total RNA as an input before depletion. For these reasons I decided to use erythroblasts derived from the fetal liver culture system as representative of definitive erythroblasts in this work. However, it is important to keep in mind that comparing "primary" primitive and "*in vitro* cultured" definitive erythroid cells can affect the RNA-seq results since this is likely to contribute to identification of excessive number of DE genes for further prioritization. A significant proportion of these DE genes may have been identified as a result of culture system effects, such as genes involving in cell cycle regulation, rather than being involved in normal expression and repression of the ζ-globin

gene. For this reason, further prioritization strategies for these DE genes focusing on transcription factor encoding genes is likely to yield plausible candidate genes responsible for ζ-globin expression. Nevertheless, I am currently collecting primary intermediate erythroblasts derived from bone marrow of a number of mice for a subsequent expression analysis to eliminate the culture system effect on differential expression results.

In the current RNA-seq and microarray experiments, fetal liver culture had been employed for the derivation of definitive erythroid cells. I found that sorting for early erythroblasts using $CD44^{+}$ resulted in improvement of maturational synchronization and 30 hours of differentiation yielded intermediate-stage definitive erythroblasts, which were morphologically comparable with the primitive erythroblasts derived from E10.5 yolk sacs. Definitive erythroblasts at this stage, although representing fetal-stage erythropoiesis, are suitable for study of ζ-globin regulation since no ζ-globin transcripts are detectable.

In the RNA-seq experiment, the globin- and ribo- depletion strategies were of great benefit, since globin transcripts were successfully reduced to such a level that rare transcripts could be routinely detected. However, a single type of globin- and ribo-depletion kit was employed for the final analysis, since variability of off-target effects between kits resulted in a significant difference in expression data.

Given the diversity in cellular structure and biochemistry of primitive and definitive erythroblasts (see Chapter 1, Sections 1.3.5, 1.3.6), a large number of DE genes may be expected. Genes previously identified as having transcriptional differences between primitive and definitive erythroid cells were

also identified as DE genes in the current study. Differential expression of the aquaporin genes (*Aqp1,3,8,9*) is thought to modulate a differential oxidative state between primitive and definitive erythroid cells as demonstrated in a previous study (Kingsley et al., 2013) and these were identified in the current dataset. *Aqp 3* and *8* are upregulated in primitive erythroblasts at LogFC 2.9 and 3.4 respectively, and *Aqp 1* and *9* are upregulated in definitive erythroblasts at LogFC of 9.3 and 7.4 respectively. The RNA-seq performed in this study demonstrates that there are as many as > 4,000 DE genes (LogFC ≥ 2) in total between these two erythroid populations supporting that they are distinct cell types (see Section 1.3). To confirm this finding, I have performed gene expression arrays using the same RNA samples and found a large number of DE genes (~ 3,000 genes) between primitive and definitive erythroblasts, of which ~ 1,200 genes were consistently identified as significantly DE genes in both datasets. This number of overlapping DE genes is large enough to support the observation that differences between primitive and definitive erythropoiesis go far beyond differences in globin expression profiles. Functional annotation of the overlapping DE genes demonstrate that, besides major differences in cellular metabolic processes and signaling pathways between primitive and definitive erythroid cells, primitive erythroblasts may also play a role in the development of embryos as genes involved in developmental processes were ranked third among genes enriched in primitive erythroblasts (Figure 5.12B, red bar graphs). However, further work will be required to establish whether this is the case.

It is of note that there are ~1,000 more DE genes identified from RNA-seq compared to expression arrays, showing the benefit of RNA-seq in the

detection of previously uncharacterized transcripts. However, since a large number of DE genes between primitive and definitive erythroid cells were identified using RNA-seq alone, it was difficult to prioritize and specifically identify candidate genes underlying ζ-globin expression. I reasoned that genes encoding *trans*-acting factors regulating ζ-globin transcription, if transcriptionally regulated, should be consistently differentially expressed at a significant level that can be identified from both the RNA-seq and the microarray datasets. This is supported by identification of some key genes known to be involved in β-globin switching, such as *Bcl11a* and *Sox6*, in the overlapping DE dataset.

From this dataset, I am particularly interested in transcription factor encoding genes since genes upregulated in primitive erythroblasts are likely to be responsible for normal ζ-globin expression and those upregulated in definitive erythroid cells are likely to be involved in its repression. Previous studies demonstrated that compound heterozygous loss-of-function mutations of *KLF1* in humans lead to increased levels of ζ-globin in definitive erythroid cells, up to 8% of total α-like globin transcripts, suggesting its role in the normal repression of the embryonic ζ-globin gene in definitive erythropoiesis (Viprakasit et al., 2014). However, Klf1 is one of the core erythroid transcription factors and is critical for the development of primitive and definitive red cells (Hodge et al., 2006), thus *Klf1* was not differentially expressed between the two erythroid lineages in my transcriptome datasets. Two transcription factors transcriptionally regulated downstream of Klf1 previously characterized as being part of the β-globin switching circuit, Bcl11a and Sox6, were identified as candidate genes upregulated in definitive

erythroblasts in this work. However, currently, there is no direct evidence that these genes play roles in ζ-globin repression. An interesting candidate as a potential repressor of the ζ-globin gene identified in the transcriptome experiment is a gene encoding another KLF family member, *Klf3*, which demonstrated ~10 fold enrichment in definitive erythroid cells, as identified by RNA-seq. Homozygous knockout of *Klf3* alone and double-knockout of *Klf3/Klf8* have been shown to result in elevated level of embryonic globin, including *Hba-x (encoding ζ-globin)* expression in a recent study (Funnell et al., 2013). In that study chromatin immunoprecipitation (ChIP) using Klf3 antibodies was further performed in differentiated MEL cells and the result demonstrated that, although Klf3 does not directly bind to the ζ-globin promoter, it binds to the α-globin promoters, supporting a potential role of Klf3 as a repressor of ζ-globin in definitive erythroid cells.

Among the top ranked candidate transcription-factor encoding genes in this study, another core erythroid gene, *Nfe2*, was identified. This gene is enriched in definitive erythroblasts at ~15 fold in the current study. NFE2 DNA-binding activity consists of a heterodimer containing an 18-kD Maf protein (MafF, MafG, or MafK) and p45 (Andrews, 1998). MafG and MafK are functionally redundant and able to form dimers with themselves, which suppresses transcription at NFE2 sites. NFE2 has previously been shown to bind the major α-globin enhancers, MCS-R1 and R2, in definitive erythroid cells in both human and mouse (De Gobbi et al., 2007). Although, to my knowledge, there is no evidence that NFE2 directly binds the ζ-globin promoter, previous studies in transgenic mice showed that when an NFE2/AP1 binding motif within the HS-40 element was mutated by 1 bp, there was de-repression of ζ-

globin in adult erythroid cells (Huang et al., 1998). Subsequent studies from the same group using gel mobility shift and transient co-transfection assays demonstrated that this 1-bp mutation completely abolishes the binding of the small MafK homodimer, rather than the binding of NFE2 itself (Wen et al., 2000). Interestingly, *MafK* is also one of the top ranked genes upregulated in definitive erythroid cells in this study (fold enrichment 32), suggesting its potential role, together with *Nfe2*, as developmental stage-specific ζ-globin repressor. In addition, MafK null mutant mice are viable and fertile, however, it is doubtful that ζ-globin de-repression has been assessed in adults.

Because the functions of many of the short listed genes are not known or incompletely annotated, further systemic prioritization and functional validation of these genes is required. This will include confirmation of differential expression using qPCR and functional validation using a targeted Cas9/CRISPR library system (see details in Section 7.7). Reduction of ζ-globin transcripts will be assessed in cultured primitive erythroblasts, in which genes encoding candidate ζ-globin activators (genes upregulated in primitive erythroblasts) are deleted, whereas de-repression of ζ-globin will be assessed in CD44hi-sorted E12.5 fetal liver culture-derived definitive erythroblasts, in which candidate ζ-globin repressors (genes upregulated in definitive erythroblasts) are deleted (also see Section 7.7).

Chapter 6 : Study of an exceptional case of the BHFS

6.1 Introduction

The BHFS survivor registry (Chapter 3) shows that all except one case inherited the South East Asian (SEA) deletion on both alleles, consistent with the high allele frequency of this deletion. This deletion removes both adult α-globin genes leaving the embryonically expressed ζ-globin gene intact. The case study presented in this Chapter is one of the BHFS survivors in the registry, who inherited homozygous SEA deletion and remarkably survived the first year of life with minimal transfusion, most likely because of persistent expression of ζ-globin.

The aim of this part of my project is to use the information on the *cis*- and *trans*-regulation of ζ-globin obtained using the mouse model in an attempt to gain insight into the molecular mechanism underlying the persistent ζ-globin expression. Additionally, genetic variants identified in the patient may offer insight into factors underlying ζ-globin expression.

6.2 Results

6.2.1 Case History

This Thai boy (KD) was the first pregnancy of a 24-year-old Thai woman and her 28-year-old Thai-Chinese husband (non-consanguineous marriage). He was born *via* vaginal delivery at 39 weeks of gestation with a birth weight of 2.3 kg (small for gestational age). He appeared to be inactive and suffered

from apnea with cyanosis (O_2 saturation 88-89%) at birth, thus was referred from a regional hospital. Maternal antenatal care history is not available. Physical examination revealed no pallor, no jaundice, however, he did present with dyspnea and cyanosis with clear breath sounds. There was a systolic murmur (grade III/VI) detected at the left lower sternal border and an enlarged liver and spleen (~2 cm below right and left costal margin) were noted. Short penile length, hypospadias with inguinal testes were also observed.

On the first day of life his blood count showed low hemoglobin level (Hb) for gestational age at 11.2 g/dl (range 14-22 g/dl, mean 18 g/dl) (Jopling et al., 2009) with a hematocrit of 44% (range 43-63%, mean 53%). The mean corpuscular volume (MCV) and mean corpuscular hemoglobin concentration (MCHC) were also low at 78.9 fL (normal 106 ± 4 fL) and 25.5 g/dl (normal 34 ± 1 g/dl) (Christensen et al., 2012) respectively, while his reticulocyte count was in the normal range at 6.6% (normal < 7% on day 1). His white blood cell count demonstrated leukocytosis with elevated numbers of myeloid progenitors and mild thrombocytopenia was present, suggesting the possibility of neonatal infection. His peripheral blood smear showed a mild degree of hypochromia, a moderate degree of anisocytosis and poikilocytosis with numerous nucleated red cells and polychromasia, indicating a degree of hemolysis with compensated erythropoiesis. However, his abnormal hemoglobin level and other red cell indicies were not sufficient to arouse suspicion of a severe anemic condition such as the BHFS, rather a diagnosis of neonatal sepsis was made. The first blood transfusion was given a few days after birth primarily for the treatment of idiopathic hemolytic jaundice. During the neonatal period, KD was treated for neonatal sepsis, persistent

pulmonary hypertension of the newborn (PPHN) and patent ductus arteriosus (PDA) with signs and symptoms of congestive heart failure; however, his PDA underwent spontaneous closure in subsequent months.

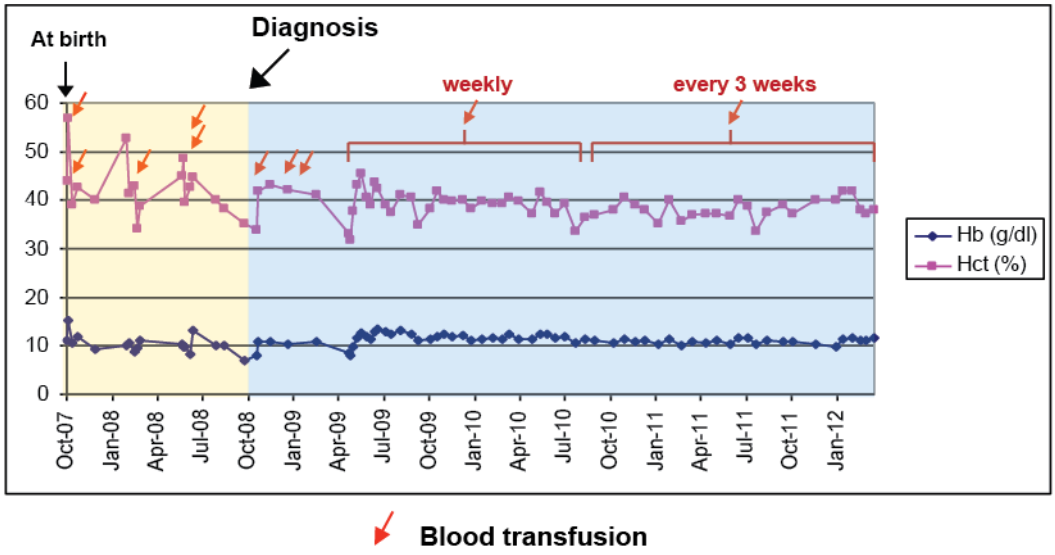


Figure 6.1 The clinical course of an exceptional case of the BHFS (KD)

KD was given blood transfusions twice during the neonatal period for the treatment of idiopathic hemolytic jaundice. From 2-9 months of age, he received 3 additional transfusions when he was admitted with pneumonia and hypoxia (yellow shading of chart). However, his Hb and Hct levels were always above 10 g/dl and 35% pre-transfusion during the first year of life. After the diagnosis of the BHFS at 1 year of age, KD has been regularly transfused (blue area).

After he was discharged from the hospital, he was readmitted multiple times with a diagnosis of recurrent pneumonia during his first year of life. His liver and spleen became gradually enlarged and his growth was severely retarded: his body weight was only ~3.5 kg at 1 year of age, similar to that of a newborn baby. Glycogen storage disease was suspected but diagnostic investigations did not support that diagnosis. Despite KD's Hb level remaining at ~10 g/dl, a few more blood transfusions were given during admissions for various reasons (including hypoxia and dyspnea). The clinical course of KD and his blood transfusion history are summarized in Figure 6.1.

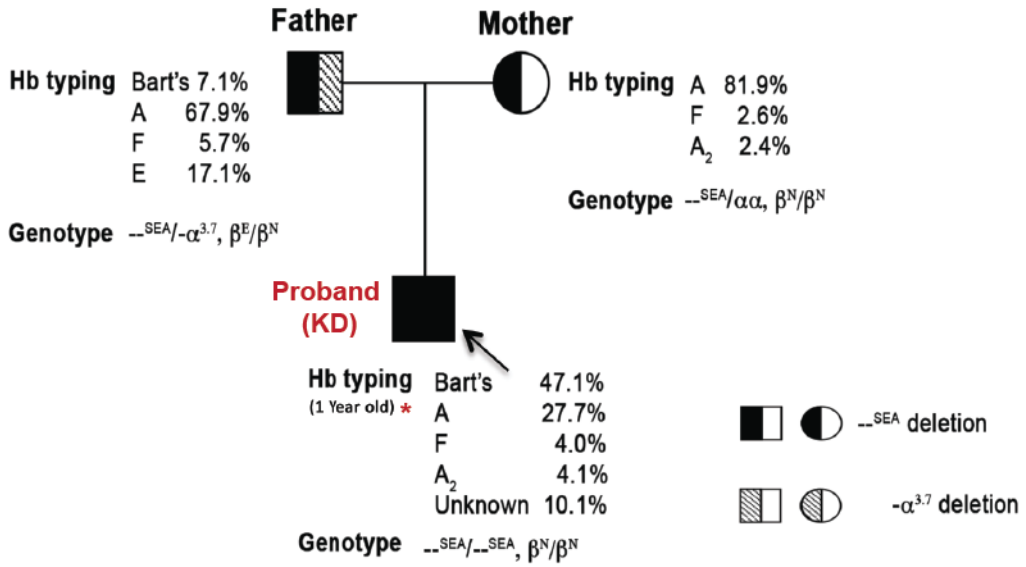


Figure 6.2 KD Pedigree

Hb typing and α- and β-globin loci genotypes of KD and his parents. Hemoglobin electrophoresis (Hb typing) was performed at the local hospital using low-pressure liquid chromatography (LPLC). * Indicates Hb typing of the patient at the time of first diagnosis (10 weeks after blood transfusion).

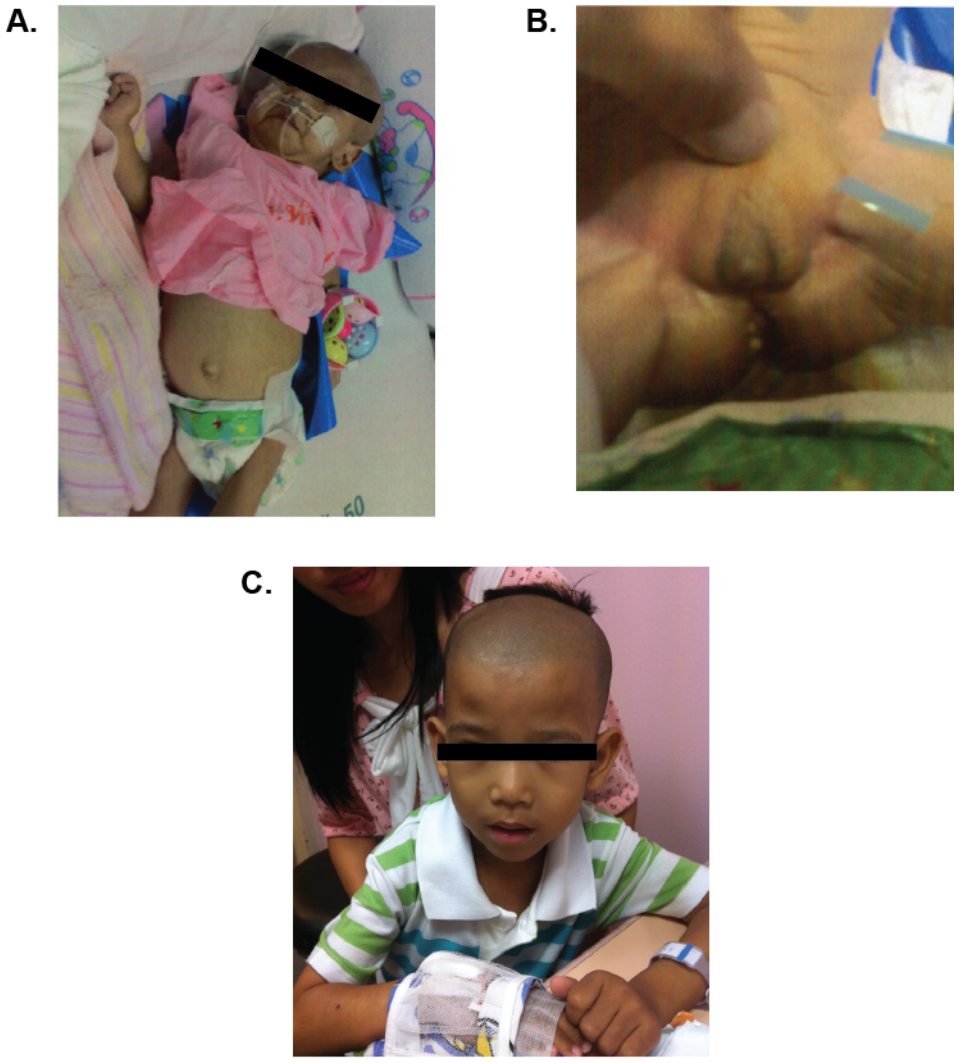


Figure 6.3 General appearance of KD

A. General appearance of KD at the time of diagnosis (1 year old), when he suffered from severe growth retardation and developmental delay. B. Physical examination revealed hypospadias. C. General appearance of KD at 4 years of age. After initiation of hypertransfusions, his growth and neurodevelopment have been improving.

At one year of age, KD received more complete hematologic analysis, because he suffered from mild anemia with Hb level decreased to ~ 9 g/dl. A presumptive diagnosis of the BHFS was made by Hb electrophoresis (low performance liquid chromatography, LPLC): Hb Bart's (γ_4) 47.1%, Hb A ($\alpha_2\beta_2$) 27.7%, Hb A_2 ($\alpha_2\delta_2$) 4.1%, Hb F ($\alpha 2\gamma 2$) 8.4% and an unknown Hb 10.1%. However, Hb A and Hb A_2 were detected because the electrophoresis was performed ~10 weeks after blood transfusion. The diagnosis of BHFS was

subsequently confirmed by gap PCR performed using DNA extracted from his peripheral blood, which demonstrated homozygosity for the SEA deletion. The α- and β-globin genotypes and globin analysis of each family member is shown in Figure 6.2. The general appearance of the patient is shown in Figure 6.3.

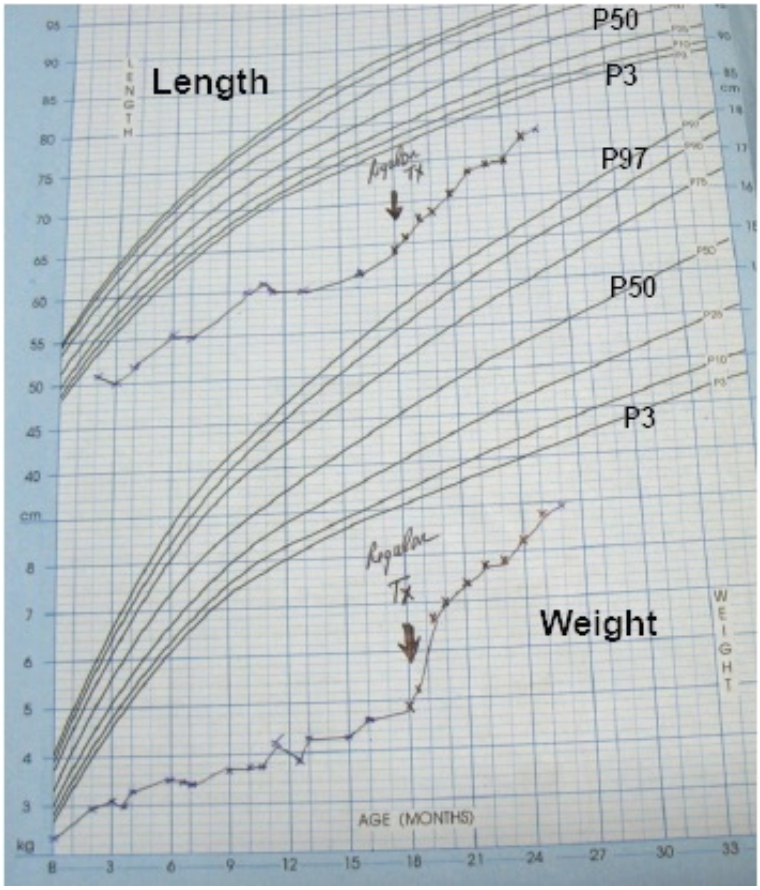


Figure 6.4 Growth curve

The figure shows KD's growth during his first 2 years of life, plotted on a growth chart for Thai male children. He had suffered from poor growth (weight and height) until regular blood transfusion was initiated (black arrows), after which time improved growth has been observed. P3, P50 and P97 indicate weight and height at 3rd, 50th and 97th percentile among the population of Thai male children.

After diagnosis the patient received hypertransfusion, initially on a weekly basis and subsequently every 2 - 3 weeks. The aim of his transfusion regimen is to maintain his functional hemoglobin (HbA + HbF) level at ~ 10 g/dl. As a consequence of transfusion therapy he has iron overload and is currently on

an iron chelation regimen. His growth (Figure 6.4) and neurodevelopment have dramatically improved after initiation of hypertransfusion. The mother subsequently carried another affected fetus, which was aborted after family counseling. The patient is 6.5 years old at the time of writing, he is going to normal school, however, he has a mild to moderate degree of learning disability. His hypospadias has been surgically corrected.

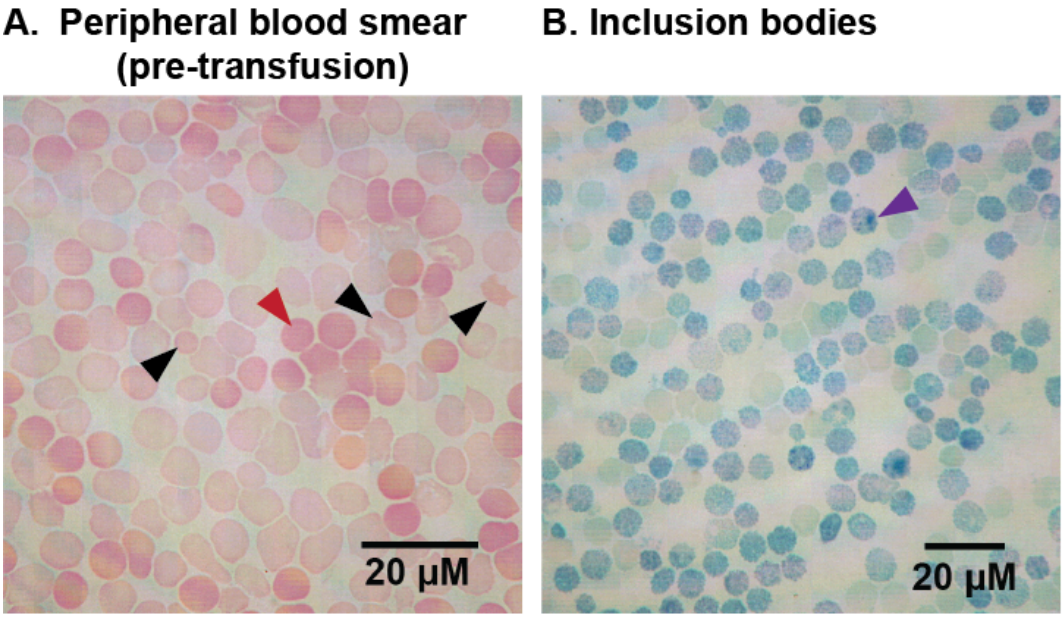


Figure 6.5 KD's blood smears performed when first referred to our center (at 4 years of age)

A. Modified Wright stain of KD's pre-transfusional peripheral blood smear. The majority of red blood cells seen are normochromic normocytic transfused cells (red arrow), however, the patient's own red cells showing mild to moderate hypochromia, microcytosis, anisopoikilocytosis and fragmented red cells were also observed (black arrows). B. Brilliant cresyl blue stain of KD's pre-transfusional peripheral blood shows inclusion bodies (purple) arrow), which are aggregates of HbH (β_4) formed in red blood cells. Pictures were taken by Dr. Chris Fisher.

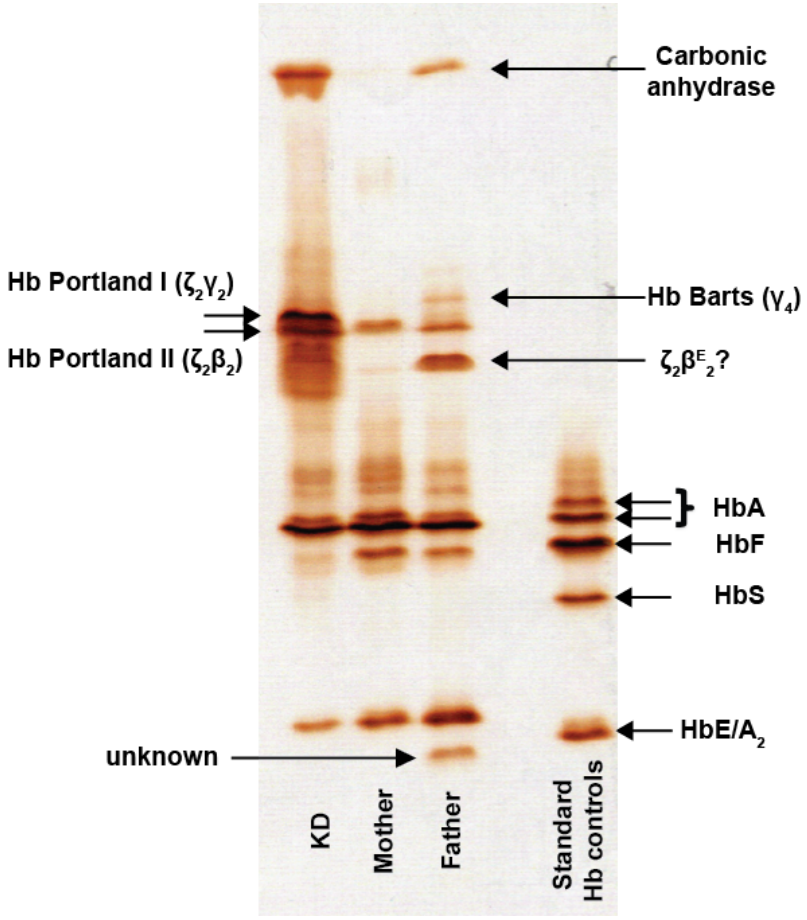


Figure 6.6 Isoelectric focusing (IEF) of embryonic globins identified in KD and his parents (at 4 years of age)

Comparing hemoglobin profiles from KD with the standard hemoglobin controls reveals 2 distinct abnormal hemoglobin protein bands separated at a more cathodic position than HbA and HbF. These hemoglobin bands were subsequently identified by mass spectrometry to be Hb Portland I ($\zeta_2\gamma_2$) and Hb Portland II ($\zeta_2\beta_2$). Hemoglobin profiles of both parents also demonstrate the presence of Hb Portland II. This finding is consistent with previous reports of Hb Portland being detected in heterozygous carriers of the SEA deletion (Chui et al., 1986; Chung et al., 1984; Ireland et al., 1993). The presence of HbF in the mother is most likely owing to a 7 bp insertion in *KLF1*, leading to a loss of function frameshift mutation. Presence of thick HbE, HbF, and Hb Bart's (γ_4) bands in the father, in addition to HbA, is consistent with his AE Bart's disease genotype ($--^{SEA}/-\alpha^{3.7}$, β^E/β^N). The IEF was performed by Dr. Chris Fisher.

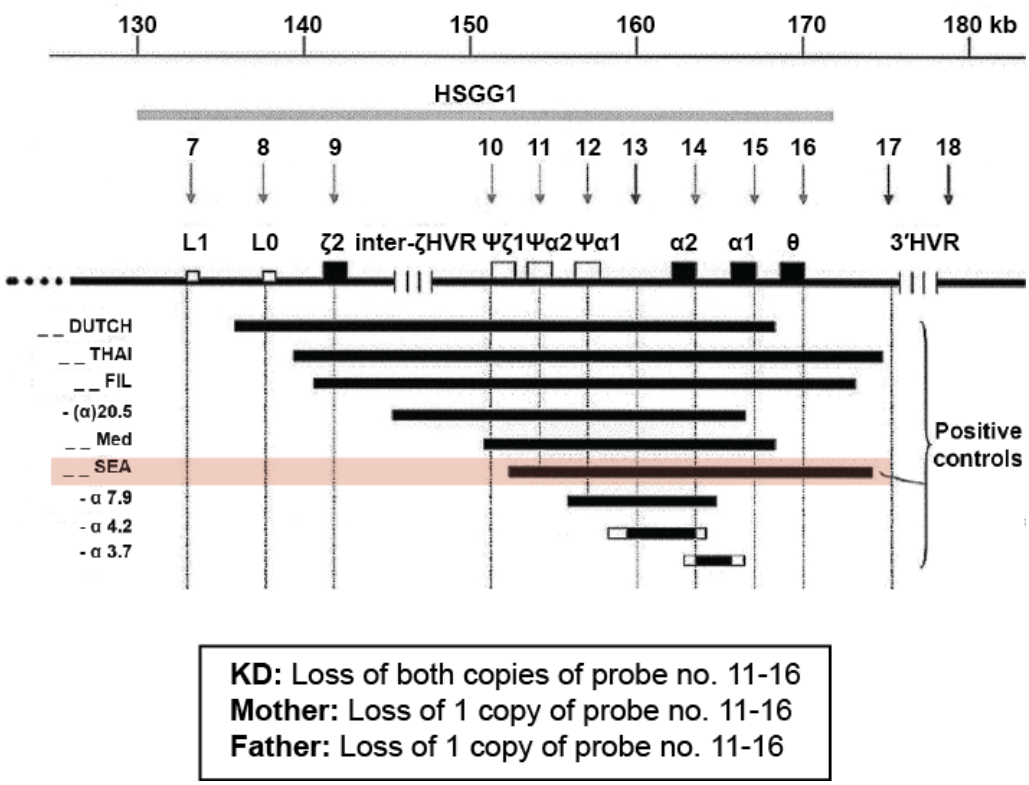


Figure 6.7 Multiplex ligation-dependent probe amplification (MLPA) study of the SEA deletion in KD

Losses of both copies of probes 11-16 in KD and one copy of the same probes in the parents confirm homozygous and heterozygous carriers of the SEA deletion (red highlight) respectively. MLPA was performed by Dr. Chris Fisher.

KD was referred to our department as a 4 year old as an unusual case of the BHFS. Given that he does not produce any α-globin, it is very surprising that he had survived his first year of life at all with only minimal transfusion. The patient's pre-transfusion peripheral blood smear and brilliant cresyl blue stain showing presence of inclusion bodies (HbH, β_4) were performed and are shown in Figure 6.5A,B. Multiplex ligation-dependent amplification (MLPA) confirming the SEA deletion and isoelectric focusing (IEF) showing different types of globin, including Hb Portland I ($\zeta_2\gamma_2$) and II ($\zeta_2\beta_2$), identified from KD and his parents' peripheral blood were performed (Figure 6.6, 6.7).

6.2.2 Analysis of globin transcripts in the patient's erythroid cells

KD's survival during the first year of life in a transfusion independent manner together with initial identification of Hb Portland using IEF lead to the speculation that he may have persistent expression of the functional embryonic ζ-globin gene. To confirm this, I performed expression analysis of RNA isolated from peripheral reticulocytes, which revealed that embryonic ζ-globin transcript accumulation in KD is increased to a level far beyond that found in either of his parents (Figure 6.8). This suggests persistent expression of ζ-globin may be substituting for α-globin to some extent and ameliorating his condition. The γ-globin gene was upregulated in both parents (Figure 6.8D), however, this is likely to be explained by their known genotypes. The father has AE Bart's disease ($--^{SEA}/-\alpha^{3.7}$, β^{E}/β^{N}, Figure 6.2), in which increased γ-globin expression is commonly found. Sanger sequencing shows the mother to harbor a previously described seven base pair insertion in *KLF1,* leading to a loss of function frameshift mutation (G176RfsX179) prior to the zinc finger domains (Viprakasit et al., 2014). Loss of function mutations (including this frameshift allele) in *KLF1* cause de-repression of fetal and embryonic globins in adults (Viprakasit et al., 2014). Interestingly, γ-globin expression is also upregulated in the patient who has not inherited the abnormal *KLF1* allele from his mother nor AE Bart's from his father. This observation suggests that the patient may harbor a variant in a *trans*-activating factor which de-represses both ζ- and γ-globin expression.

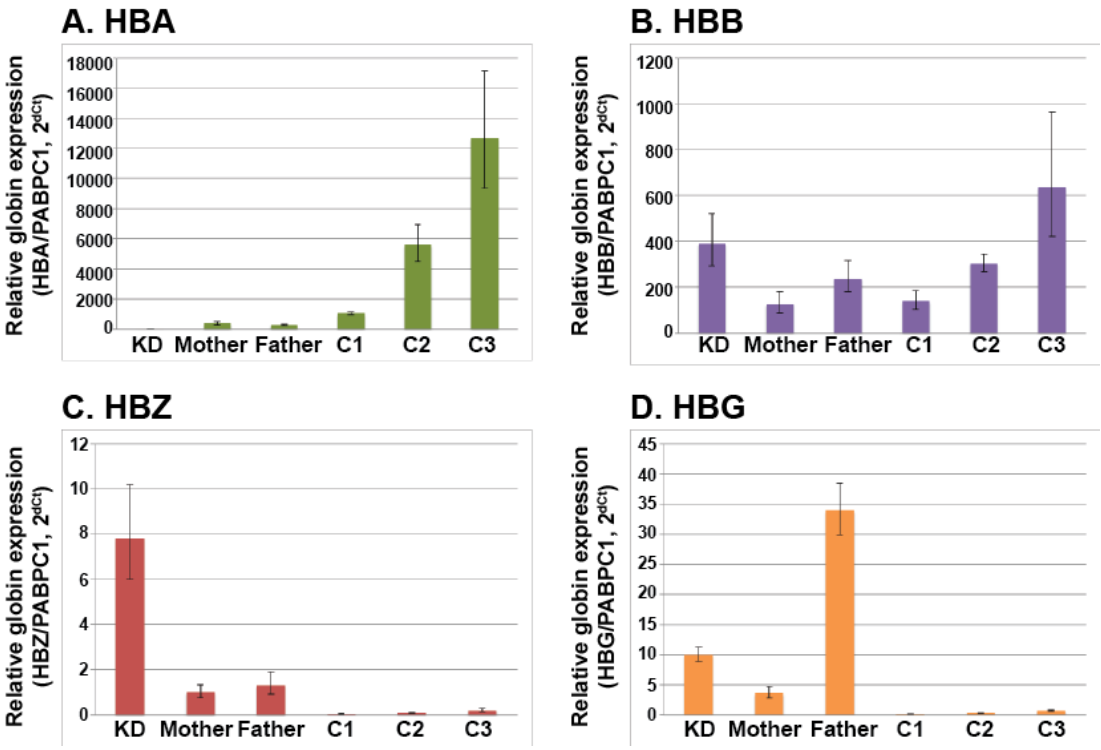


Figure 6.8 Globin expression analysis of KD and his parents

Quantification of globin transcript levels in RNA isolated from peripheral blood of KD and his mother and father compared with that of 3 healthy individuals of the same ethnic background (Thai) (C1, C2, C3). The α-globin expression (A) of KD was absent consistent with the loss of all 4 α-globin genes. Instead, his embryonic ζ-globin expression (C) was greatly increased. γ-globin expression (D) was increased not only in the father, who has AE Bart's disease ($--^{SEA}/-\alpha^{3.7}$, β^{E}/β^{N}), but also in KD and his mother. β-globin expression (B) of each member of the family was comparable to that of healthy individuals. Data was normalized to *PABPC1* and shown in 2^{dCt}, when dCt is the qPCR threshold cycle difference. Error bars represent SD across three technical replicates.

To further investigate whether KD's unusual phenotype of persistent ζ-globin expression can be recapitulated *in vitro*, as well as to establish a model system in which the mechanism underlying this de-repression may be investigated, patient-derived induced pluripotent stem cells (iPSCs) were generated from mononuclear cells (through collaboration with Dr. Lee Carpenter of the Blood Transfusion Service in Oxford). I initially differentiated these iPSCs together with iPSCs generated from a normal individual into erythroblasts using a published protocol ((Chang and Bouhassira, 2012), see methods in Chapter 2). Previous studies suggest that when ES cells/iPSCs are differentiated in culture, the first cell population emerging has a gene

expression profile consistent with primitive hematopoiesis and this wave is followed by the emergence of distinct populations of EMP-definitive hematopoiesis (McGrath et al., 2015). I found that erythroblasts derived from both iPSC lines matured toward terminal differentiation over a 24-day period, as shown by morphology and immunophenotype (Figure 6.9A,B). However, enucleation was not observed using this differentiation protocol. Globin expression analysis of iPSC-derived erythroblasts shows that α-globin transcripts were present only in WT, but not KD's cells (Figure 6.10A), consistent with the patient's α^0-thalassemia phenotype. Only low levels of β-globin transcripts were detected in erythroblasts derived from both KD and WT iPSC lines at D17, D21 and D24 of differentiation (Figure 6.10B). This finding is consistent with the observation in previous studies of erythroid trans-differentiation from iPSCs using an identical culture protocol (Chang and Bouhassira, 2012). I found that KD's erythroblasts persistently express *HBZ* (ζ-globin) throughout differentiation in this culture system and ζ-globin transcripts remained detectable at day 24 of differentiation (most likely in EMP-like definitive erythroblasts at this stage) as demonstrated using qPCR, when wild type (WT) erythroblasts demonstrated ζ-globin repression (Figure 6.10C). This small amount of ζ-globin transcript was sufficient to form globin tetramers with the β-like globins, resulting in well-hemoglobinized terminally matured erythroid cells, evidenced by dark pink colored-cytoplasm of nucleated red blood cells (Figure 6.9A, black arrow). This patient's mature erythroid cytoplasmic color is also comparable to that of WT cells at the relevant stage (figure 6.9A, WT cells D24). Fetal β-like globin (Figure 6.10D, *HBG* encoding γ-globin) and embryonic β-like globin expression (Figure

6.10E, *HBE* encoding ε-globin) in patient erythroblasts were increased along the course of differentiation, as compared to those of WT erythroblasts supporting the hypothesis that the mechanism underlying ζ-globin de-repression in this patient might also cause de-repression of embryonic and fetal β-like globin.

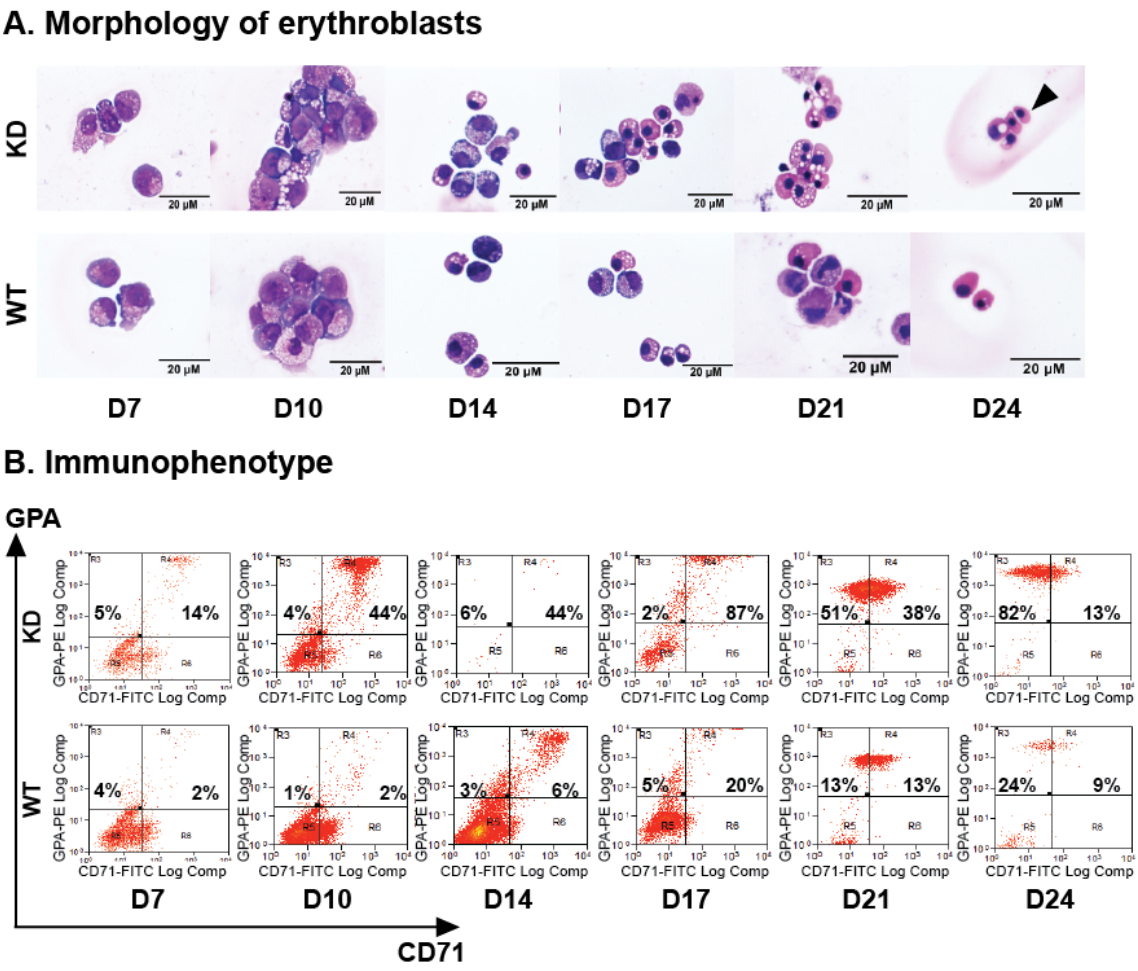


Figure 6.9 *In vitro* erythroid differentiation from iPSCs derived from KD and a normal individual (WT)

A. Modified Wright stain showing morphology of iPSC derived erythroblasts. Both KD and WT iPSC lines (n=1 each) were differentiated using the 2-stage protocol of Chang and Bouhassira, 2012 (Chang and Bouhassira, 2012). The morphology gradually changed from early (D7, D10) to intermediate (D14, D17) and to late erythroblasts by D21-24. At each time point, maturational stages of differentiating erythroblasts are comparable between KD and WT iPSCs. At the end of differentiation (D24), nucleated red blood cells derived from KD iPSCs (black arrow) are as well hemoglobinized as those derived from WT iPSCs. B. Flow cytometric analysis (n=1 for each iPSC line) shows the immunophenotype of iPSC-derived erythroblasts at each stage. CD71 and glycophorin A (GPA) were highly expressed on the surface of differentiated erythroid cells at D7-17, consistent with morphological staging of early and intermediate erythroblasts. During the late phase of *in vitro* differentiation (D21-24), CD71 surface expression gradually decreased, while high levels of GPA persisted, consistent with terminally matured erythroblasts. Proportion of $CD71^{+}GPA^{+}$ and $CD71^{-}GPA^{+}$ cells as compared to total cells harvested at each time point are shown in each plot.

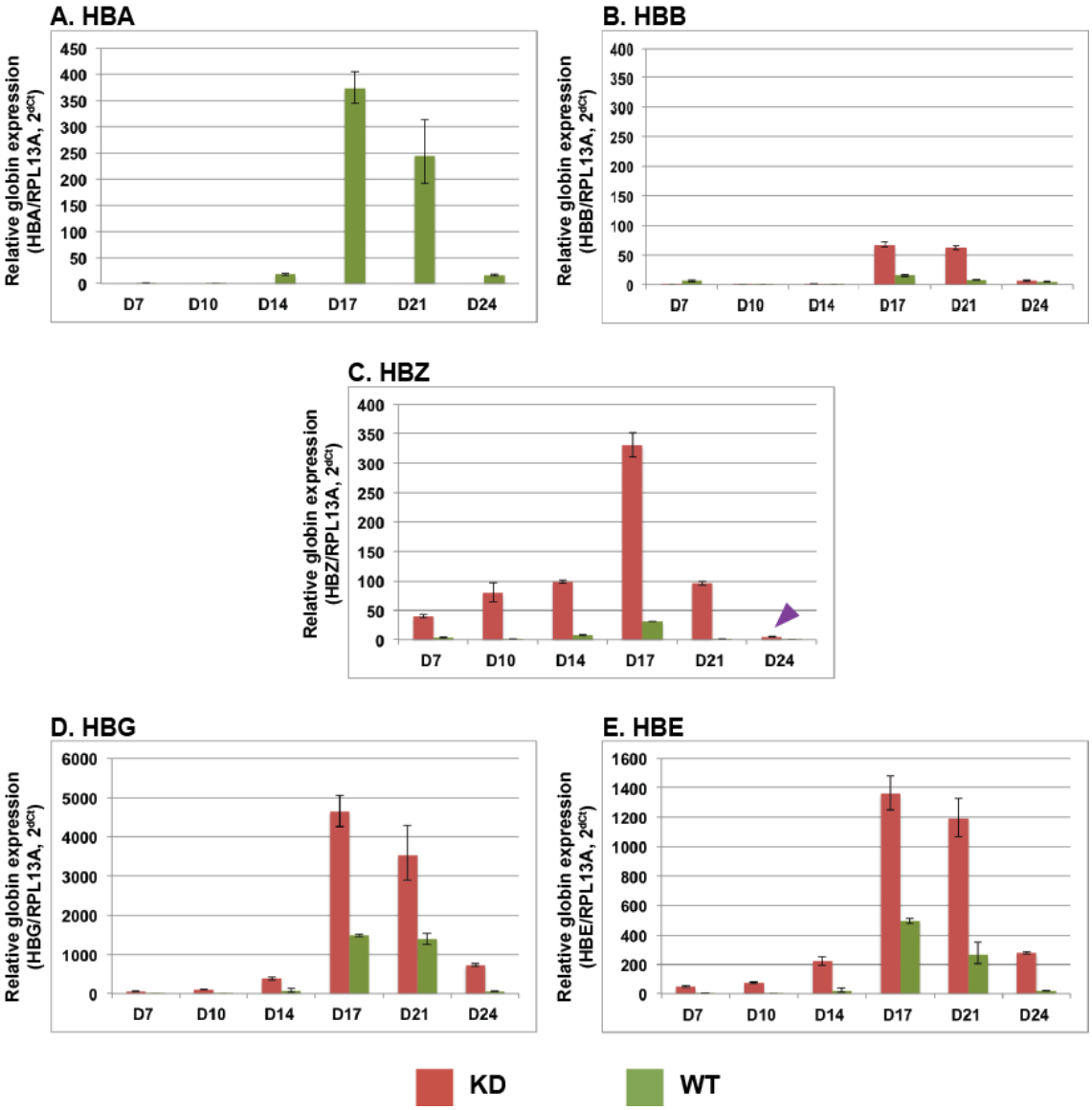


Figure 6.10 Globin expression of erythroblasts derived from KD and WT iPSCs

Globin expression analysis of iPSC derived erythroblasts during 24-day differentiation (n=1 in each iPSC line). α-globin transcripts were undetectable by qPCR in patient-derived differentiated iPSCs, thus only α-globin expression in control-differentiated iPSCs is seen (A). Only a low level of β-globin transcript was detected in erythroblasts derived from both iPSC lines in D17, D21 and D24 of differentiation (B). Patient erythroblasts demonstrate persistently high levels of ζ-globin transcripts when compared with that of WT erythroblasts at all stages (C). Towards the end of differentiation, ζ-globin transcripts remained detectable at a small amount in the KD cell line (purple arrow) and were completely repressed in the WT line. In fact, this amount of ζ-globin comprised all of α-like globin in KD's erythroid cells and was sufficient to form globin tetramers with β-like globin, resulting in well-hemoglobinized terminally matured erythroblasts (Figure 6.9A, black arrow). Note that KD's erythroblasts also demonstrate upregulation of both fetal (D) and embryonic (E) β-like globin genes. qPCR data was normalized to *RPL13A* and shown in 2^{dCt}, where dCt is threshold cycle difference. Error bars represent SD across three technical replicates.

6.2.3 Investigation of cis- variants in the unusual case of the BHFS

DNA isolated from a lymphoblastoid cell line derived from the patient was subjected to whole genome sequencing (see Methods in Chapter 2). To first exclude the possibility that a *cis*-acting regulatory change caused de-repression of ζ-globin, I initially assayed all variants in the terminal 2 Mb of chromosome 16 and found those in the terminal 345 kb to be uniformly homozygous (Figure 6.11), indicating that KD inherited identical α-globin alleles from both parents. I identified the haplotype carried by the patient in the HapMap project *(http://hapmap.ncbi.nlm.nih.gov/)* in the related Japanese and Chinese populations (Figure 6.12 A,B). Because the patient was found to be homozygous for a common haplotype and because no heterozygous changes were found to have arisen on either allele I consider it to be unlikely that a *cis*-acting regulatory change underlies the abnormal expression of ζ-globin in this patient. To further ensure no variant had been overlooked I visually inspected the DNA sequence of the known *cis*-acting regulatory sequences of the α-globin genes using a customized GBrowse database (Stein et al., 2002) and identified no heterozygous changes.

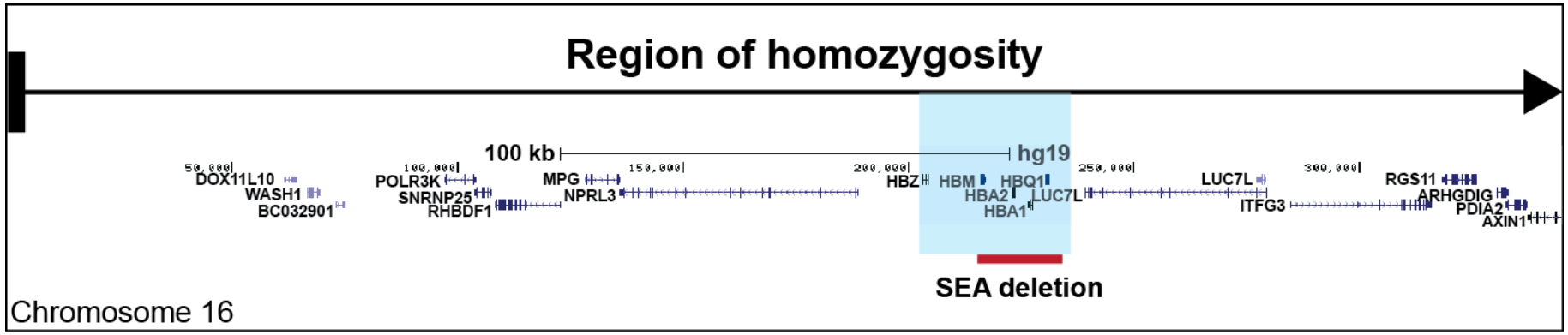


Figure 6.11 Region of homozygosity in KD

The terminal region of the p arm of human chromosome 16 showing the full extent of the homozygous region identified in KD. The genes in the region are shown. The α-globin locus is shaded in blue. The patient's SEA deletion between 215,395 bp and 234,700 bp (hg19), is shown by a red line.

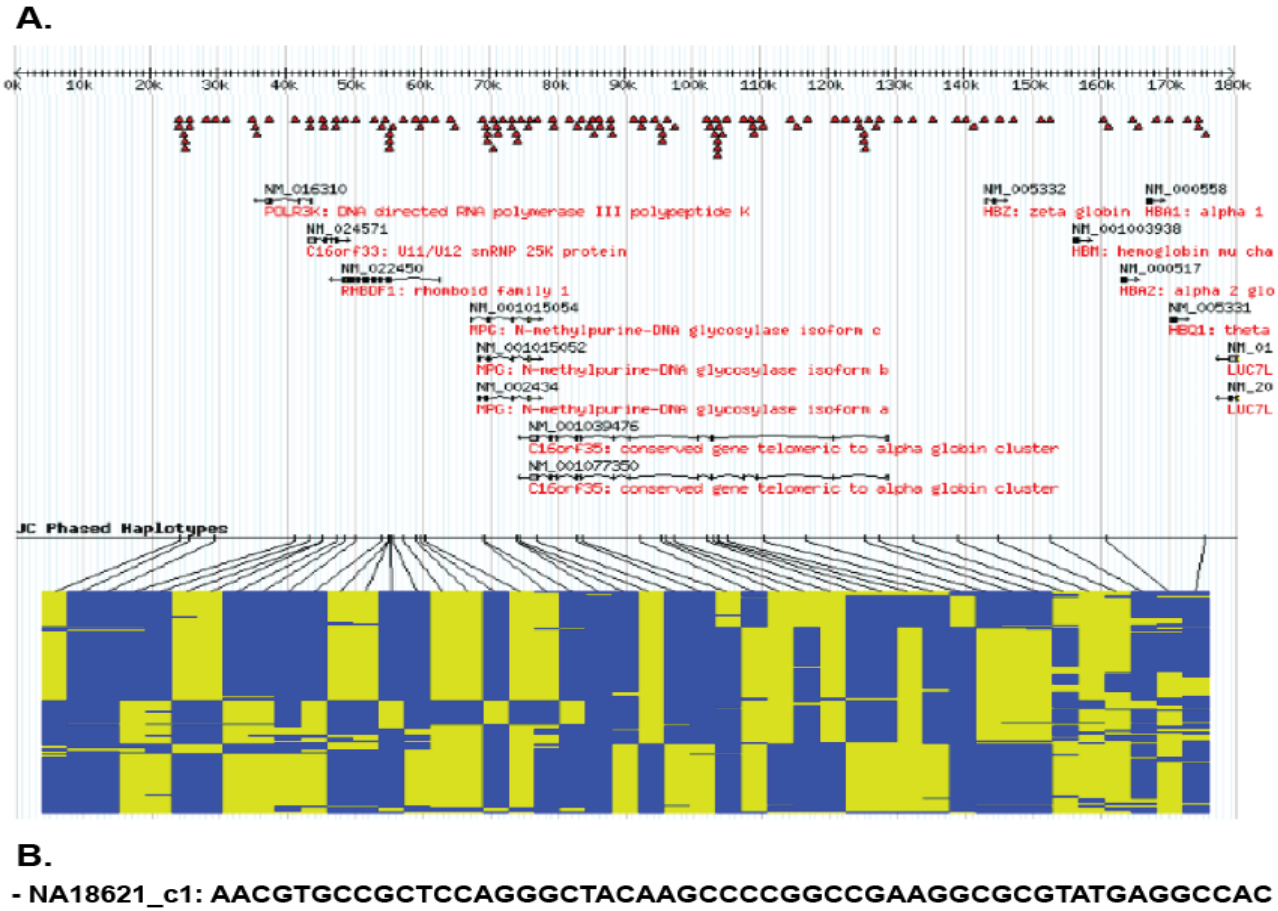


Figure 6.12 Phased haplotypes along the terminal 2 Mb of chromosome 16 in the Japanese and Chinese populations and the unique BHFS survivor's haplotype

Genes included in the terminal region of chromosome 16 and region-specific phased haplotypes in Japanese (JPT) and Chinese (CHB) population derived from the HapMap project (*http://hapmap.ncbi.nlm.nih.gov/*) A. The region analyzed included the *POLR3K* gene 5′ and the *LUC7L* gene 3′ of the α-globin cluster (upper panel). Blue and yellow colors represent clusters of SNPs showing linkage disequilibrium. Along the vertical axis 189 haplotypes are represented. Details of the unique BHFS survivor's haplotype, present in this population is also shown B.

6.2.4 Identification of candidate copy number variations

In the absence of plausible *cis*-acting regulatory changes and because of the observed γ-globin expression in the patient, I have investigated copy number variation, which may affect *trans*-acting factors. To achieve this, the patient's DNA has been analyzed by array comparative genomic hybridization (aCGH) using the Human OmniExpressExome array platform (Illumina®, see methods in Chapter 2), which includes over 273,000 exonic markers genome-wide, to detect large insertions and deletions. Optimal calling for gain and loss, in comparison with a reference sample, was performed using Nexus copy number analysis software (Illumina®). Calls were visually inspected and curated prior to reporting. Using a minimum of 3 probes to detect variations, a total of 81 copy number (CN) losses (4 homozygous, 77 heterozygous) spanning 178 genes and 1 CN gain covering 2 genes were detected (Table 6.1). Consistent with homozygosity for the SEA deletion, one of the homozygous CN losses was confined within the α-globin cluster, spanning the *HBA1*, *HBA2*, *HBM* and *HBQ1* genes. Given that ζ-globin de-repression leading to survival of the BHFS is a very rare condition, it is unlikely that causative variants are commonly found in healthy populations. For that reason, regions of CN loss and gain identified in this study were compared with common variants reported in the Database of Genomic Variants (DGV, (MacDonald et al., 2014)). Removal of variants present in the DGV resulted in a total of 5 CN losses that included 6 genes (Table 6.2). Of the six genes that are haploinsufficient in this patient and absent from the DGV, three are annotated as signaling proteins *(LRP3, PGLYRP1, IGFL4),* one is a structural

protein (*MACF1*), one is a developmentally important growth factor *(FGF2)* (Huang da et al., 2009) and the remaining variant *(C15ORF40)* is of unclassified function. *Lrp3* is upregulated in primitive erythroblasts at LogFC 9 and *Pglyrp1* is upregulated in definitive erythroblasts at LogFC 7. However, all these genes (Table 6.2) remain formal candidates for KD's unusual phenotype.

Chromosomal Region	Event	Length	Cytoband	Gene Symbols
chr7:99,696,312-99,704,474	CN Gain	8163	q22.1	MCM7, AP4M1
chr1:86,393,133-86,408,189	Homozygous Copy Loss	15057	p22.3	COL24A1
chr4:123,747,412-123,750,051	Homozygous Copy Loss	2640	q27	FGF2
chr12:70,680,128-70,683,938	Homozygous Copy Loss	3811	q15	CNOT2
chr16:214,014-236,637	**Homozygous Copy Loss**	**22624**	**p13.3**	**HBM, HBA2, HBA1, HBQ1**
chr1:1,126,260-1,153,009	CN Loss	26750	p36.33	TTLL10, TNFRSF18, TNFRSF4, SDF4
chr1:1,505,049-1,563,885	CN Loss	58837	p36.33	SSU72, C1orf233, MIB2
chr1:2,329,999-2,407,472	CN Loss	77474	p36.32	RER1, PEX10
chr1:39,878,583-39,880,016	CN Loss	1434	p34.3	MACF1
chr1:223,564,339-223,567,805	CN Loss	3467	q41	C1orf65
chr1:228,289,570-228,412,361	CN Loss	122792	q42.13	C1orf35, MRPL55, GUK1, GJC2, IBA57, C1orf145, OBSCN
chr1:228,503,635-228,505,230	CN Loss	1596	q42.13	OBSCN
chr1:232,941,072-232,941,760	CN Loss	689	q42.2	KIAA1383
chr1:928,057-950,367	CN Loss	22311	p36.33	HES4, ISG15
chr2:241,369,224-241,418,989	CN Loss	49766	q37.3	GPC1, PP14571, MIR149, ANKMY1
chr4:674,623-677,095	CN Loss	2473	p16.3	MYL5, MFSD7
chr4:960,571-985,084	CN Loss	24514	p16.3	DGKQ, SLC26A1, IDUA
chr5:132,146,660-132,159,449	CN Loss	12790	q31.1	SOWAHA, SHROOM1
chr6:53,516,666-53,517,088	CN loss	423	p12.1	KLHL31
chr7:973,085-1,119,312	CN Loss	146228	p22.3	ADAP1, COX19, CYP2W1, MIR339, GPR146, C7orf50
chr7:2,578,134-2,582,892	CN Loss	4759	p22.3	BRAT1
chr7:53,102,808-53,103,961	CN Loss	1154	p12.1	POM121L12
chr7:73,015,500-73,079,942	CN Loss	64443	q11.23	MLXIPL
chr7:73,182,511-73,198,844	CN Loss	16334	q11.23	CLDN3
chr7:124,581,882-124,707,132	CN Loss	125251	q31.33	no gene in this region
chr7:149,488,847-149,490,521	CN Loss	1675	q36.1	SSPO
chr8:10,466,661-10,467,227	CN Loss	567	p23.1	RP1L1
chr8:15,401,755-15,426,013	CN Loss	24259	p22	TUSC3
chr8:144,873,505-144,894,995	CN Loss	21491	q24.3	SCRIB
chr8:144,943,209-144,990,859	CN Loss	47651	q24.3	EPPK1, PLEC
chr8:144,997,281-145,000,428	CN Loss	3148	q24.3	PLEC
chr9:0-216,463	CN Loss	216464	p24.3	DDX11L5, WASH1, FAM138C, FOXD4, CBWD1, C9orf66, DOCK8
chr9:2,718,264-2,719,088	CN Loss	825	p24.2	KCNV2
chr9:140,093,643-140,109,229	CN Loss	15587	q34.3	TPRN, TMEM203, NDOR1
chr9:140,147,870-140,245,684	CN Loss	97815	q34.3	LOC100129722, C9orf173, COBRA1, TOR4A, NRARP, EXD3
chr9:140,274,209-140,329,948	CN Loss	55740	q34.3	EXD3, NOXA1, ENTPD8
chr9:140,365,974-140,392,536	CN Loss	26563	q34.3	PNPLA7
chr10:134,971,899-134,999,723	CN Loss	27825	q26.3	KNDC1
chr11:284,298-291,478	CN Loss	7181	p15.5	NLRP6, ATHL1
chr11:551,216-556,930	CN Loss	5715	p15.5	LRRC56, C11orf35
chr11:610,288-614,309	CN Loss	4022	p15.5	PHRF1, IRF7
chr11:1,782,856-1,857,142	CN Loss	74287	p15.5	MOB2, CTSD, SYT8
chr11:65,351,089-65,367,602	CN Loss	16514	q13.1	EHBP1L1, KCNK7, MAP3K11
chr14:38,724,273-38,725,323	CN Loss	1051	q21.1	CLEC14A
chr14:103,391,274-103,397,833	CN Loss	6560	q32.32	AMN
chr14:103,568,541-103,571,722	CN Loss	3182	q32.32	EXOC3L4
chr15:83,679,992-83,682,070	CN Loss	2079	q25.2	C15orf40
chr16:578,309-602,810	CN Loss	24502	p13.3	MIR5587, SOLH, MIR3176
chr16:639,910-683,918	CN Loss	44009	p13.3	RAB40C, WFIKKN1
chr16:703,635-711,989	CN Loss	8355	p13.3	WDR90

Chromosomal Region	Event	Length	Cytoband	Gene Symbols
chr16:818,693-840,228	CN Loss	21536	p13.3	MSLN, MIR662, RPUSD1, CHTF18
chr16:1,129,869-1,143,689	CN Loss	13821	p13.3	SSTR5, C1QTNF8
chr16:1,838,315-1,846,216	CN Loss	7902	p13.3	NUBP2, IGFALS
chr16:2,138,181-2,142,549	CN Loss	4369	p13.3	TSC2, MIR1225, PKD1
chr16:4,798,319-4,802,848	CN Loss	4530	p13.3	C16orf71, ZNF500
chr16:67,203,730-67,208,408	CN Loss	4679	q22.1	HSF4, NOL3
chr16:68,056,448-68,082,838	CN Loss	26391	q22.1	DDX28, DUS2L
chr16:83,999,211-84,008,827	CN Loss	9617	q23.3	OSGIN1, NECAB2
chr16:88,594,592-88,630,537	CN Loss	35946	q24.2	ZFPM1
chr17:40,557,297-40,575,374	CN Loss	18078	q21.2	PTRF
chr17:42,250,885-42,255,635	CN Loss	4751	q21.31	ASB16, ASB16-AS1
chr17:80,914,635-80,953,083	CN Loss	38449	q25.3	B3GNTL1
chr18:77,170,700-77,173,092	CN Loss	2393	q23	NFATC1
chr19:548,957-599,174	CN Loss	50218	p13.3	GZMM, BSG, HCN2
chr19:1,987,682-2,067,281	CN Loss	79600	p13.3	BTBD2, MKNK2
chr19:3,528,835-3,544,767	CN Loss	15933	p13.3	FZR1, C19orf71, MFSD12
chr19:3,589,810-3,607,662	CN Loss	17853	p13.3	GIPC3, TBXA2R, CACTIN-AS1
chr19:4,571,192-4,638,620	CN Loss	67429	p13.3	no gene in this region
chr19:8,201,006-8,201,835	CN Loss	830	p13.2	FBN3
chr19:8,564,518-8,587,678	CN Loss	23161	p13.2	PRAM1, ZNF414, MYO1F
chr19:13,869,790-13,876,303	CN Loss	6514	p13.2	CCDC130, MRI1
chr19:18,368,203-18,368,878	CN Loss	676	p13.11	KIAA1683
chr19:33,696,835-33,699,184	CN Loss	2350	q13.11	LRP3
chr19:45,682,608-45,729,582	CN Loss	46975	q13.32	BLOC1S3, EXOC3L2
chr19:46,524,370-46,550,612	CN Loss	26243	q13.32	PGLYRP1, IGFL4
chr19:55,871,868-55,888,325	CN Loss	16458	q13.42	FAM71E2, IL11, TMEM190
chr19:59,070,134-59,074,743	CN Loss	4610	q13.43	UBE2M, LOC100131691, MZF1
chr20:62,114,545-62,169,661	CN Loss	55117	q13.33	EEF1A2, PPDPF, PTK6
chr20:62,178,472-62,185,648	CN Loss	7177	q13.33	SRMS, C20orf195
chr21:47,551,984-47,552,390	CN Loss	407	q22.3	COL6A2
chr22:38,220,375-38,223,930	CN Loss	3556	q13.1	GALR3
chr22:50,928,899-50,943,464	CN Loss	14566	q13.33	LMF2

Table 6.1 Genomic structural variants in KD as identified using aCGH

Only CN gains and losses are shown. Purple highlighting indicates the region of the SEA deletion. Genes within the structural variations, which do not overlap with common variants in the DGV are shown in red.

Gene symbol	Gene Name	Event	Probes	Known Functions
FGF2*	Fibroblast growth factor 2 (basic)	Homozygous Copy Loss	3	Plays an important role in the regulation of cell survival, cell division, angiogenesis, cell differentiation and cell migration
MACF1	Microtubule-actin crosslinking factor 1	CN Loss	10	Facilitates actin-microtubule interactions at the cell periphery and couples the microtubule network to cellular junctions
C15orf40	Chromosome 15 open reading frame 40	CN Loss	7	Unknown
LRP3	Low Density Lipoprotein Receptor-Related Protein 3	CN Loss	10	May be involved in the internalization of lipophilic molecules and/or signal transduction
PGLYRP1	IGF-like family member 4	CN Loss	21	Plays a role in innate immunity, may bind and kill gram-positive bacteria
IGFL4	IGF-like family member 4	CN Loss	21	Play critical roles in cellular energy metabolism and in growth and development

Table 6.2 Genes covered by CNVs and absent in the DGV

* When checked against the patient's WGS data, a read depth of > 25 is present at this region, suggesting heterozygous loss rather than homozygous loss. Summary of gene functions were derived from GeneCards® Human gene database (http://www.genecards.org/).

6.2.5 Identification and prioritization of trans-candidate variants from the patient's whole genome sequencing

Initially, I obtained DNA extracted from the patient's lymphoblastoid cell line and sequenced key genes known to be involved in β-globin switching, including *KLF1*, *KLF3*, *KLF8* and *BCL11A* using Sanger sequencing. Although a 7 bp insertion in KLF1, leading to a loss of function frameshift mutation prior to zinc finger domains, was identified in his mother, KD did not inherit this allele. In addition, no abnormal sequences in *KLF3*, *KLF8* and *BCL11A* were found.

To identify *trans*-acting variants that may be responsible for ζ-globin de-repression in this patient, I first examined his whole genome sequence (WGS) data and prioritized coding variants for further analysis. The DNA derived from the proband was subjected to massively parallel sequencing using an Illumina HiSeq machine (see Methods in Chapter 2). Coverage was measured in terms of read depth over exons, a read depth of > 10 reads was present in 168,577 exons out of a total of 227,110 exons annotated in the human genome (demonstrating coverage of >74% at this read depth). A total of 23,742 predicted variants present in the coding and intronic regions were initially identified.

To further narrow our search, we selected only variants predicted by ANNOVAR (Wang et al., 2010) to alter protein coding sequences; specifically, insertions or deletions predicted to alter the reading frame, non-synonymous amino acid changes or loss or gain of a stop codon. This filter left 3,588 predicted coding changes. Visual inspection of these sequence calls, using a

customized GBrowse database (Stein et al., 2002) revealed ~80% to be invalid, their presence most likely owing to low coverage or to result from misalignment of a highly similar sequence. Those present in dbSNP (build 137) (http://www.ncbi.nlm.nih.gov/projects/SNP/) with an allele frequency of > 5% or present in similar datasets obtained from three other Thai individuals were removed. Subsequently, only those non-synonymous amino acid changes in evolutionarily conserved DNA sequences were selected, together with frameshift changes and gain or loss of stop codons. This resulted in 74 variants remaining (Table 6.3).

In the absence of appropriate family members to perform segregation analysis (to associate variants with phenotypic abnormalities, in this case is significant upregulation of ζ-globin), I have integrated the DE genes identified from the mouse transcriptome comparison between primitive and definitive erythroid cells to KD's WGS data and only 3 DE genes contain variants likely to be deleterious. However, only one of these genes is a plausible candidate (*CALCOCO1*) as it has transcription factor activity, while the other two genes play roles in cellular transport and in the cytoskeleton and therefore unlikely to be involved in ζ-globin regulation (Table 6.4). To ensure that the causative variant has not been overlooked, further work needs to be conducted to establish a full list of likely deleterious variants that may be tested in a high-throughput functional screen (see details in Section 7.7).

Gene symbol	Gene name
ACIN1	Apoptotic chromatin condensation inducer1
AP4B1	Adaptor-related protein complex AP-4, beta 1
ARID3A	AT rich interactive domain 3A (BRIGHT-like)
ASB3	Ankyrin repeat and SOCS box containing 3
BAZ2B	Bromodomain adjacent to zinc finger domain, 2B
BRAT1	BRCA1-associated ATM activator 1
BRD8	Bromodomain containing 8
C11orf75	Single-pass membrane protein with coiled-coil domains 4
C15orf39	Chromosome 15 open reading frame 39
C16orf46	Chromosome 16 open reading frame 46
C5orf42	Chromosome 5 open reading frame 42

Gene symbol	Gene name
CALCOCO1	Calcium binding and coiled-coil domain 1
CBFA2T2	Core-binding factor, runt domain, alpha subunit 2; translocated to, 2
CCDC142	Coiled-coil domain containing 142
CEBPA	CCAAT/Enhancer Binding Protein (C/EBP), Alpha
CTBP2	C-terminal binding protein 2
CXorf66	Chromosome X open reading frame 66
DBX2	Develping brain homeobox 2
ELP2	Elongator acetyltransferase complex subunit 2
ERCC3	Excision repair cross-complementing rodent repair deficiency, complementation group 3
FAM149A	Family with sequence similarity 149, member A
FAM217A	Family with sequence similarity 217, member A
FANCC	Fanconi anemia, complementation group C
FANCM	Fanconi anemia, complementation group M
FARP1	FERM, RhoGEF (Arhgef) and pleckstrin domain protein 1 (chondrocyte-derived)
FEN1	Flap structure-specific endonuclease 1
FKBP4	FK506 binding protein 4, 59kDa
FOXP3	Forkhead box P3
FXR1	Fragile X mental retardation, autosomal homolog 1
GLIS3	GLIS Family Zinc Finger 3, both activator and repressor, involve development of pancreatic beta-cell, liver, kidney, thyroid, eye
GTF2E1	General transcription factor IIE, polypeptide 1, alpha 56 kDa
HEATR4	HEAT repeat containing 4
ING3	Inhibitor of growth family, member 3
LAMTOR4	Late endosomal/lysosomal adaptor, MAPK and MTOR activator 4
LFNG	*LFNG* O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase
MCM4	Minichromosome maintenance complex component 4
MDN1	*MDN1*, midasin homolog (yeast)
MFSD5	Major facilitator superfamily domain containing 5
MGA	MAX dimerization protein
NAA30	N (alpha)-acetyltransferase 30, NatC catalytic subunit
NCOR1	Nuclear receptor corepressor 1
NDUFAF5	NADH dehydrogenase (ubiquinone) complex I, assembly factor 5
NOTCH1	Notch 1
NR1H4	Nuclear receptor subfamily 1, group H, member 4
PARP4	Poly (ADP-ribose) polymerase family, member 4
PCDH15	Protocadherin-related 15
PIGO	Phosphatidylinositol glycan anchor biosynthesis, class O
POTEA	POTE ankyrin domain family, member A
POU3F1	POU class 3 homeobox 1
PTCRA	Pre T-cell antigen receptor alpha
PTPRD	Protein tyrosine phosphatase, receptor type, D
RASSF5	Ras association (RalGDS/AF-6) domain family member 5
SIPA1L2	Signal-induced proliferation-associated 1 like 2
SRCAP	Snf2-related CREBBP activator protein
TACC2	Transforming, acidic coiled-coil containing protein 2
TDRD3	Tudor domain containing 3
TFAP2A	Transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)
TLX2	T-cell leukemia homeobox 2
TMEM91	Transmembrane protein 91
TP53RK	TP53 regulating kinase
TRPS1	Zinc Finger Transcription Factor Trps12
TSHZ2	Teashirt zinc finger homeobox 2
TTI1	TELO2 Interacting Protein 1
TULP4	Tubby like protein 4
VGLL2	Vestigial like 2 (Drosophila)
VSX1	Visual system homeobox 1
WNT10A	Wingless-type MMTV integration site family, member 10A
WNT10B	Wingless-type MMTV integration site family, member 10B
XRCC6BP1	XRCC6 binding protein 1
ZBTB17	Zinc finger and BTB domain containing 17
ZBTB2	Zinc finger and BTB domain containing 2
ZEB1	Zinc Finger E-Box Binding Homeobox 1, transcriptional repressor of IL-2
ZFHX3	Zinc finger homeobox 3
ZNF23	Zinc finger protein 23

Table 6.3 List of genes containing candidate variants identified from KD's WGS

Gene symbol	Gene name	Enrichment in mouse transcriptome comparison	Level of DE (LogFC)	Known functions	Related human disorders
CALCOCO1	Calcium binding and coiled coil domain 1	Definitive erythroblats	5.492301703	Functions as a coactivator for aryl hydrocarbon and nuclear receptors (NR). Involved in the transcriptional activation of target genes in the Wnt/CTNNB1 pathway.	Unknown
AP4B1	Adaptor-related protein complex AP-4, beta 1	Definitive erythroblats	3.256924377	Targeting proteins from the trans-Golgi network to the endosomal-lysosomal system.	Spastic paraplegia
FARP1	FERM, RhoGEF (Arhgef) and pleckstrin domain protein 1 (chondrocyte-derived)	Primitive erythroblats	6.604225011	May play a role in semaphorin signaling. Plays a role in the assembly and disassembly of dendritic filopodia, the formation of dendritic spines, regulation of dendrite length and ultimately the formation of synapses (By similarity).	Unknown

Table 6.4 List of genes that contain likely deleterious variants identified in KD's WGS and which were also identified as DE genes in the mouse transcriptome comparison datasets

Summary of gene functions and related human disorders were derived from GeneCards® Human gene database (http://www.genecards.org/).

6.3 Discussion

Survival of an individual with the BHFS with minimal transfusion during the first year of life is an exceptional condition. Although, the patient received a few blood transfusions, this level of transfusion is normally not sufficient to maintain an individual affected with the BHFS. In fact, this transfusion therapy was not given for the treatment of anemia, since he was slightly anemic at birth and had maintained only a slightly low level of Hb (~10 g/dl) during his first year. My initial globin expression study of the unique BHFS survivor suggests he survived with the α^0-thalassemia genotype through abnormally high expression of ζ-globin, consistent with presence of embryonic globins identified in his peripheral blood using IEF. It is of note that both assay of ζ-

globin transcripts and identification of Hb Portland I and II were performed from pre-transfusional blood of KD after he was 4 years of age, confirming his persistent ζ-globin expression.

Although KD had survived through his first year of life, he suffered from congestive heart failure, hepatosplenomegaly and poor growth. This was mainly due to the presence of non-functional hemoglobin, Hb Bart's (γ_4), leading to ineffective erythropoiesis, hemolysis (with compensated extramedullary hematopoiesis) and poor oxygen release to body organs. After the diagnosis of the BHFS the patient was regularly transfused. Nevertheless, it has been shown that hypertransfusion strategies with monitoring of functional Hb level can lead to significant growth improvement, as shown by his growth chart and fairly good neurodevelopmental outcomes. He has now entered a mainstream school with only 6-12 months of developmental delay. KD also had hypospadias, however, this does not contribute to his current co-morbidity, since he received a surgical correction without complication.

I have obtained whole genome sequence data for this individual and my analysis of the haplotypes of his α-globin loci shows that he is homozygous for the common SEA allele and neither allele contains a novel variant within 345 kb of the ζ-globin gene. This makes it very unlikely that a *cis*-acting change is responsible for the de-repression of ζ-globin in this individual.

To identify causative *trans*-acting variants, I initially looked for genomic structural variations using aCGH and identified a list of 6 prioritized candidate copy number variations (Table 6.2). In a second approach, I have prioritized *trans*-acting candidate variants from KD's WGS data. This is challenging, since there is no appropriate family member to allow segregation analysis to

be performed. To generate a list of candidates, I performed a systematic prioritization by considering only coding variants affecting evolutionarily conserved protein domains and missense and loss or gain of stop codons. This led to a list of 74 variants, which are all formal candidates that may underlie ectopic expression of ζ-globin. In an attempt to further refine this list, I identified DE genes from the mouse primitive/definitive screen. However, only one is a plausible candidate because it has transcription factor activity, although there is no evidence implicating it in globin switching. Therefore, to continue this work, further sequencing needs to be undertaken to improve the level of genome coverage and the genes identified in the resulting list of deleterious variants tested using a high throughput loss of function screen (see Section 7.7).

In this study, I have successfully differentiated iPSCs generated from the patient into the erythroid lineage and demonstrated that this *in vitro* culture system can be a model system for functional validation of the variants found in this study as it recapitulated the patient's ζ-globin de-repression phenotype. In addition, future study of gene expression and epigenetic changes at each stage of the iPSC- derived erythroid differentiation may be employed to gain insight into the underlying molecular nature of his elevated ζ-globin expression.

Chapter 7 : Final discussion and conclusions

7.1 Introduction

Although the BHFS has long been considered a universally fatal disorder, an increasing number of patients have survived because of prenatal and immediate postnatal blood transfusion. This work has documented the natural history and clinical outcomes of long-term survivors with the BHFS to gain insight into whether this disease should now be considered manageable and to assess whether the burden of treatment is too great.

The majority of the common causes of α-thalassemia, including the South East Asian (SEA) deletion, remove both adult α-globin paralogs leaving the embryonically expressed ζ-globin gene intact. Previous work in a transgenic mouse model demonstrated that ζ-globin can functionally substitute for α-globin in adult erythroid cells (Russell and Liebhaber, 1998). Additionally, evidence that upregulation of ζ-globin significantly ameliorated disease severity in the unusual case of BHFS (KD) reported here, suggests that ζ-globin can also functionally substitute for α-globin in humans. Therefore, it is theoretically possible to re-activate or prevent silencing of the ζ-globin gene in definitive erythroid cells, in the absence of both adult α-globin genes, to provide an alternative strategy to stem cell transplantation for long-term therapy for individuals with the BHFS.

In this thesis, I have used the mouse as a model system to investigate *cis*- and *trans*-regulation of ζ-globin. There are two lines of evidence suggesting that the mouse is a good model system for study of the α-globin switch. First,

in both human and mouse, there are only two developmental expression stages for the α-like genes, the embryonic stage when all α-like globin genes are expressed and the adult stage (including fetal stage beyond 6-8 weeks of gestation through adult life) when only α-globin genes continue to be expressed. Second, evidence from previous studies demonstrates that the human ζ-globin gene switches off normally in a developmental stage-specific manner in transgenic mice (Higgs et al., 1998).

In the last part of this work, I have integrated information on the *cis*- and *trans*-regulation of the ζ-globin gene obtained using the mouse model to gain insight into an exceptional case of a BHFS survivor, who remarkably survived the first year of life with minimal transfusion most likely because of persistent expression of ζ-globin.

7.2 Natural history and long-term clinical outcomes of survivors with the BHFS

Over the past two decades, great effort has been made to implement strategies for prenatal screening of severe thalassemia including the BHFS in many countries. In some areas of the world such as the Northern Thailand, where a high frequency of α^0-thalassemia carriers are present, prenatal prevention of homozygous α^0-thalassemia births is found to be very effective (Tongsong et al., 2013). Nevertheless, with ever improving perinatal care, survivors are occurring with increasing frequency, most commonly due to unexpected, undiagnosed cases rather than a planned delivery. To my knowledge, the database of 60 cases presented in this thesis is the largest review describing the natural history and long-term clinical outcomes of

survivors with the BHFS to date. These data are useful for assessing whether the BHFS should now be considered manageable.

Unsurprisingly, the BHFS survivor registry reveals that the vast majority of the survivors are of East Asian origin and they inherited the most common α^0-thalassemia deletion (the SEA deletion) on both alleles. A total of 45% of the survivors were diagnosed after birth and have survived because of immediate postnatal transfusion, while the majority of the remainder were diagnosed in mid to late gestation and have survived through intrauterine treatment. Intrauterine therapies including blood transfusion and exchange transfusion were most commonly initiated between 20-29 weeks of gestation. Although *in utero* HSC transplantations have been reported in two cases at gestational ages of 13 and 15 weeks, the treatments were unsuccessful. These findings suggest that any associated congenital abnormalities occurring in the first trimester of pregnancy cannot be reversed by currently available intrauterine treatments.

Intrauterine therapies may afford the chance for affected individuals to survive gestation and also provide advantages for perinatal and neonatal care. These advantages include prolongation of the course of pregnancy, improvement of birth weight and Apgar scores and shortening of the course of mechanical/assisted ventilation during the neonatal period. This review suggests that growth retardation is one of the adverse long-term outcomes found among survivors with the BHFS, as up to 50% of the patients suffer from severe growth retardation. This occurs regardless of whether affected individuals received transfusion *in utero*, since growth of these patients was more likely to result from multiple factors, including illness in the neonatal

period and early life and the treatment regimen employed (transfusion and transplantation). In contrast, this study suggests that the BHFS survivors, surprisingly, have encouraging long-term neurodevelopmental outcomes. Significant neurodevelopmental delay (severely delayed and 6-12 months' delayed) was found in 18% of all survivors in this registry. The remainder have developed normally or experienced only mild developmental delay. Comparable numbers of survivors who received either intrauterine treatment and non-intrauterine treatment were shown to have normal long-term neurodevelopmental outcomes, suggesting once affected individuals have survived beyond the neonatal period, they have a similar chance to develop normally. This finding is in contrast to a review of 11 BHFS survivors that suggested that up to ~50% of survivors without intrauterine treatment suffered from a neurological deficit (Lee et al., 2007).

My analysis also shows a high frequency (up to 50%) of associated congenital abnormalities are present in the survivors. The two types of abnormalities most commonly associated with this condition are urogenital defects and varying degrees of limb defects ranging from asymmetrical hand size to absence of distal elements of the foot. Both types of defects are most likely to have arisen by 12 weeks of gestation or even earlier, suggesting that the currently available intrauterine therapies given in second or third trimester of pregnancy are unlikely to correct those defects. A very interesting finding from this work is that the urogenital abnormalities (hypospadias being the most common) were found in as high a frequency as 71% of the male survivors. This supports the hypothesis that the urogenital defects present in α^0-thalassemia are unlikely to occur solely as a result of chronic hypoxia *in utero*

(Dame et al., 1999a; Fung et al., 1999), although the pathogenic mechanism has yet to be discovered. Two studies have tried to identify a genetic determinant within the α-globin cluster (Chr16p13.3), which may be linked to urogenital organ formation (Dame et al., 1999a; Utsch et al., 2001). However, these approaches were not successful and the link between the α-thalassemia and urogenital abnormalities remains an open question. The majority of the associated congenital abnormalities, found in individuals affected with the BHFS, including hypospadias, ASD and jejunal atresia are surgically correctable conditions, however, more severe defects such as the absence of some parts of limbs remain co-morbidities lifelong.

Although the neuropsychological development of long-term BHFS survivors is encouraging, the majority of these patients (83%) have the burden of life-long transfusion. This review suggests that the transfusion regimen given to the BHFS survivors should not be the same as that given to individuals with β-thalassemia major. Increased frequency of transfusion, timed according to optimal levels of pre-transfusional functional hemoglobin, may be beneficial for improvement of growth, reduction of HbH (β_4) level and spleen size.

In conclusion, the current review suggests that maintaining affected fetuses with the BHFS should not be routinely recommended in clinical practice. Efforts should be made to counsel couples at risk regarding the natural history, long-term outcomes and possible associated congenital anomalies of the BHFS. In addition, extensive discussion regarding increased risk of maternal complications, such as pre-term delivery, pre-eclampsia, poly/oligohydramnios and malpresentation, should be undertaken. However, upon a parental decision to continue the BHFS pregnancy, this condition

should be considered manageable with reasonable long-term outcomes and intrauterine intervention should be initiated as early in gestation as possible. The only currently available long-term cure for this group of patients is stem cell transplantation which is, however, not feasible for most cases. Therefore, reactivation of ζ-globin in adult erythroid cells to functionally substitute for α-globin may serve as an alternative treatment to ameliorate clinical symptoms of patients with severe α-thalassemia including the BHFS. Little is known about how ζ-globin is normally activated in primitive and silenced in definitive erythroid cells. As the first step towards developing a new treatment, I have identified the *cis*-acting regulatory network of ζ-globin and identified candidates that may effect *trans*-regulation of ζ-globin expression.

7.3 Summary and context of *cis*- regulation of ζ-globin

Although previous work has demonstrated that ζ-globin expression may be reliant on the α-*cis*-acting regulatory elements, all of these studies were conducted in transgenic mice containing the human α-globin locus and in the quasi-erythroid aneuploid cell line called K562 (De Gobbi et al., 2007; Gourdon et al., 1994; Huang et al., 1998; Pondel et al., 1992; Ren et al., 1993; Rombel et al., 1995; Zhang et al., 1993; Zhang et al., 1995). It is now possible to analyze in detail the *cis*-regulatory landscape, the complex interaction between *cis*-elements and promoters and chromatin accessibility in an unprecedentedly small number of cells. This work has integrated these approaches to illucidate *cis*-regulatory control of ζ-globin in mouse primary primitive erythroid cells for the first time.

The ζ-globin gene is actively transcribed in primitive erythroblasts derived from the yolk sac at embryonic day 10.5 (E10.5), as shown by the presence of accumulated transcripts and the presence of hypersensitivity at the ζ-globin promoter. At this developmental stage, all five previously characterized α-*cis*-regulatory elements (MCS-R1-R4 and Rm) are hypersensitive sites and no extra site is present in this 5′ region, indicating that the α-globin locus employs the same set of distal enhancers in both primitive and definitive erythroid cells. In primitive erythroblasts, these α-*cis*-regulatory elements physically interact with both the α- and ζ-globin gene promoters as demonstrated by Capture-C. I have further demonstrated the contribution of each upstream *cis*-acting element to α- and ζ-globin expression in primitive erythroblasts using a series of mice harboring deletions of the individual MCS-Rs. Homozygotes for double deletion of R1 and R2 showed the most significant reduction of α-globin transcripts (~8% of WT) in E10.5 yolk sac-derived primitive erythroid cells. Homozygous deletion of R1 or R2 alone led to a reduction of α-globin transcripts to the lesser extent of ~50% and ~60% of that found in WT respectively. These data suggest that R1 and R2 are the major enhancers of α-globin expression in primitive erythroid cells. This finding is similar to *cis*-acting control of the α-globin genes in definitive erythroid cells reported in (Hay D, et al. manuscript in review). Interestingly, R1 and R2 do not contribute to ζ-globin expression to the same extent as α-globin in primitive erythroid cells, as only a small reduction of ζ-globin transcripts was observed in homozygotes for deletion of the individual R1 and R2 elements (transcript level ~88% and 95% of the WT level respectively) and ζ-globin expression was ~40% of WT when R1 and R2 were deleted together homozygously.

Similarly to α-globin, ζ-globin expression is also not dependent on the other *cis*-acting regulatory elements (R3, R(m) and R4), as none of homozygous deletions of these elements resulted in significant reduction of ζ-globin transcripts in primitive erythroid cells. These findings support the hypothesis that ζ-globin is less reliant on the upstream *cis*-regulatory elements than α-globin and suggests that elements responsible for ζ-globin expression and repression may lie within the ζ-globin promoter and *trans*-acting factors. In addition, this study demonstrates that R3, R(m) and R4, which were found to have little contribution to α-globin expression in definitive erythroid cells (Hay D, et al. manuscript in review), do not contribute significantly to α- or ζ-globin expression in primitive erythroid cells. Therefore, functions of R3, R(m) and R4 are yet to be clarified.

In this work, I have identified a DNaseI HS at the ζ-globin promoter indicating its active state and DNaseI hypersensitivity extending ~300 bp downstream of the ζ-globin 3'UTR in primitive erythroblasts. DNaseI footprint analysis suggests that DNA sequences underlying the promoter sequences may be functionally important for regulation of the ζ-globin gene as they contain a number of putative transcription factor binding sites, including 6 potential GATA binding sites, supporting previous studies in K562 cells using DNaseI and gel mobility shift assays that suggests GATA1 binds the ζ-globin promoter (Yu et al., 1990). The significance of the sequences underlying the ζ-globin 3'UTR DNaseI peak is much less clear. This peak may be an artifact resulting from open chromatin as assessed by DNaseI assay at the 3′ end of highly expressed genes, such as the ζ-globin gene in these cells. Additionally, less than 50% of the nucleotides underlying this 3′UTR DNaseI peak is

evolutionarily conserved. However, there are several predicted transcription factor binding sites in this region, including four GATA1 binding sites and further analysis of these sequences may be warranted. Further footprinting analysis of these regions may assist prioritization of candidate genes identified from the mouse transcriptome comparison. However, to validate predicted binding of these transcription factors, chromatin immunoprecipitation (ChIP) experiments are required.

7.4 Summary and context of *trans*-regulation *of* ζ-globin

To identify novel *trans*-acting regulators of ζ-globin expression, I have compared the transcriptomes of primitive and HSC-derived definitive erythroid cells. Validation of globin expression profiles using qPCR and principal component analysis (PCA) of the initial RNA-seq experiment suggested that primitive erythroblasts derived from murine yolk sacs at E10.5 (ζ-globin accounting for ~40% of total α-like globin transcripts) and stage-matched definitive erythroblasts derived from E12.5 fetal liver cultures (no ζ-globin transcript detected) were the most suitable for the comparison. I employed both RNA-seq and microarray techniques as each offer specific advantages. RNA-seq offers unbiased detection of novel transcripts, a broader dynamic range and increased specificity and sensitivity. When using expression arrays there is no requirement for depletion of abundant transcripts, which may lead to off-target effects, additionally, data analysis is much more robust.

Given the differences in cell structure, biochemistry and cellular metabolic processes (reviewed in Chapter 1.3) in addition to the different globin

expression profiles between primitive and definitive erythroid cells, it is unsurprising that there are a large number of DE genes identified between these two cell populations as previous studies have demonstrated that they are different cell types (see Section 1.3). More than 4,000 DE genes (LogFC ≥ 2) were identified using RNA-seq, of which 2,675 are enriched in primitive erythroblasts and 1,560 are enriched in definitive erythroblasts. However, this large number of DE genes may partly result from the fact that the transcriptome of "primary" primitive erythroblasts was compared to that of "*in vitro* cultured" definitive erythroblasts. Therefore, a significant number of these DE genes are likely to be identified because of culture system effects (eg. genes involve in cell cycle regulation) rather than being involved in normal expression and repression of the ζ-globin gene and careful prioritization of these DE genes is warranted. I have also employed expression microarrays for transcriptome comparison analysis between primary E10.5 yolk sac derived-primitive and CD44hi-sorted E12.5 fetal liver culture-derived definitive erythroid cells. To determine the genes that are robustly differentially expressed between the two cell populations, I selected only those genes that were shown to be differentially expressed in both datasets. Reassuringly, this approach identified both *Bcl11a* and *Sox6*, key genes known to be involved in the β-globin switch.

A total of 1,168 DE genes between primitive and definitive erythroid cells were consistently identified using RNA-seq and microarrays. Of these, 355 were upregulated in primitive and 813 were upregulated in definitive erythroid cells. Functional annotation using the DAVID bioinformatics database (Huang et al., 2007a; Huang et al., 2007b), suggests that a large proportion of genes (up to

~17%) enriched in primitive erythroblasts contribute to cellular metabolic processes, signaling or developmental processes. The most common functional category (~13%) of genes enriched in definitive erythroblasts is that of transcription and co-transcription factors. This may imply that as erythropoiesis changes from the primitive to the definitive stage, developmental stage-specific gene expression, including globin genes, occurs as different transcriptional networks emerge.

Systematic prioritization of the DE genes derived from RNA-seq and microarray experiments, primarily focusing on those that encode for transcription factors, resulted in a list of 139 candidate genes, of which 34 are upregulated in primitive and 105 are upregulated in definitive erythroid cells (see Chapter 5, Table 5.3,5.4). A proportion of those enriched in primitive erythroblasts are likely to be activators of the ζ-globin gene, while some of those enriched in definitive erythroblasts are likely to be developmental stage-specific repressors of ζ-globin. Although KLF1 is a potential repressor of ζ-globin, as shown by naturally occurring mutations of this gene that result in an increased level of ζ-globin (Arnaud et al., 2010; Viprakasit et al., 2014), it was not identified as one of DE genes in this work. This finding is consistent with a previous study comparing transcriptomes from primitive and definitive erythroid cells (Kingsley et al., 2013) and is not surprising since Klf1 is a core erythroid transcription factor critical for development of both primitive and definitive red cells (Hodge et al., 2006). However, the list of *trans*-acting candidate genes identified in this study includes several potential ζ-globin repressors such as Klf3, Nfe2 and its partner for DNA binding activity MafK. A previous study has shown that mice homozygous for *Klf3* deletion either alone

or together with *Klf8 showed* elevation of ζ-globin transcripts (Funnell et al., 2013). The possibility of Nfe2 and MafK acting as repressors of ζ-globin is raised following identification of NFE2 binding in the human and mouse α-globin locus in previous studies (De Gobbi et al., 2007; Huang et al., 1998). Previous reports of transgenic mice showed that when an NFE2/AP1 binding motif within the HS-40 element was mutated by 1 bp, there was de-repression of ζ-globin in adult erythroid cells (Huang et al., 1998).

However, systematic functional validation of these 139 candidate DE genes is required and this will be performed using a Cas9/CRISPR library screen, which is beyond the scope of this project (see Section 7.7). Final candidate genes obtained from the Cas9/CRISPR screen will be considered in conjunction with the footprint analysis of the ζ-globin promoter and DNA sequences underlying DNaseI peak adjacent to the 3' UTR of the ζ-globin gene and a refined list subjected to further functional studies.

7.5 Identification of candidate genes that may cause persistent ζ-globin expression in an exceptional case of BHFS

This work reports the clinical course of an exceptional case of the BHFS who survived the first year of life with minimal transfusion. Although detectable levels (up to ~0.42%) of ζ-globin transcripts are observed among heterozygous carriers of the SEA deletion (Chui et al., 1986; Tang et al., 1992), my analysis suggests that this individual survived with the α^0-thalassemia genotype through an abnormally high level of ζ-globin expression, far beyond that found in either of his parents carrying this

deletion. This finding is further supported by de-repression of the ζ-globin gene at the terminal stage of erythroid differentiation performed using patient-derived iPSCs as compared to WT iPSCs.

Analysis of whole genome sequencing data from this patient indicates that his persistent expression of ζ-globin is unlikely to result from a *cis*-acting variant. The terminal 345 kb of his chromosome 16 was found to be uniformly homozygous at the nucleotide level, indicating that KD inherited the same SEA allele from both parents. Because the SEA haplotype in this family is a common haplotype in the related Japanese and Chinese populations (*http://hapmap.ncbi.nlm.nih.gov/*) and because no heterozygous changes were found to have arisen on either allele I consider it to be unlikely that a *cis*-acting regulatory change underlies the abnormal expression of ζ-globin in this patient.

To investigate whether *trans*-acting variants may underlie his unusual phenotype, I performed WGS and aCGH analysis. This resulted in a list of 80 candidate genes, because of the lack of appropriate family members in which to perform segregation analysis, I attempted to refine the list by overlapping variants with DE genes from the mouse primitive and definitive screen. However, no likely variants resulted from this analysis, therefore a high throughput screen of all 80 candidate genes would be required to identify the causative gene (see Section 7.7). Comparative gene expression studies of erythroblasts differentiated from the patient-derived iPSCs and WT iPSCs may lead to better insight into mechanisms underlying ζ-globin de-repression. In addition, the patient-derived iPSCs established in this study could also serve as a model for functional validation of these variants.

7.6 Implications

The BHFS survivor registry suggests that the major burden of this disorder is life-long transfusion dependence. This could be ameliorated if a novel therapy reactivating ζ-globin could be identified. The remainder of the work in this thesis represents an important first step towards this goal. This study has refined our understanding of the *cis*-regulation of ζ-globin and identified a number of genes likely to regulate ζ-globin expression and repression. Further work is required to determine which of these genes are responsible for the regulation of ζ-globin and thereby suggest novel therapies.

7.7 Future directions

The key aim of the planned future project is to determine which of the candidate *trans*-acting factors identified from the mouse transcriptome comparison between primitive and definitive erythroid cells and WGS of the exceptional case of the BHFS contribute to the normal expression and silencing of the ζ-globin gene. To achieve this a mouse line in which the coding sequence of the ζ-globin gene (*Hba-x*) has been replaced with sequence encoding green fluorescent protein (GFP) will be generated. We will culture primitive and definitive erythroblasts derived from this mouse and perform a targeted loss-of-function screen using CRISPR/Cas9 gene editing methodology. We will perform the screen using a library of sgRNAs designed against all 1,168 robustly differentially expressed genes identified from RNA-seq and microarray experiments (Figure 5.12A) and 80 candidate genes identified in the unique BHFS patient (Table 6.2, 6.3). With inclusion of 5 guides per gene, this equates to ~6,300 guide RNAs. We will introduce these

guide RNAs into primary cells from *Hba-x*–GFP mice by transduction of a lentiviral library containing equimolar amounts of each guide RNA. To ensure unbiased expression of Cas9 protein we will use cells derived from mice constitutively expressing Cas9 from an endogenous locus. We will also test a number of candidate genes using a plasmid based strategy, whereby plasmids encoding specific guide short guide RNAs and wild type *Streptococcus pyogenes* Cas9 protein are introduced into immortalized CD34+ve cells by nucleofection. We will then screen for *HBZ* de-repression in these cells using qPCR methodology. By these means, we aim to identify a short list of depleted genes that are likely to be ζ-globin repressors in GFP-positive E12.5 fetal liver culture-derived definitive erythroid cells and immortalized human CD34+ve cells and depleted genes that are likely to be ζ-globin activators in GFP-negative *in vitro* cultured primitive erythroid cells. In addition, we will also use cultures from the *Hba-x*-GFP mouse line to conduct an unbiased small molecule screen to identify small molecule inhibitors able to induce ζ-globin expression. We will validate putative hits from the CRISPR/Cas9 library screen and the small molecule screen in human erythroid cultures and the immortalized CD34+ve cells. We will also breed the *Hba-x*-GFP mouse line to homozygosity to determine whether ζ-globin function is redundant. To test for binding of the refined list of potential *trans*-acting factors, we will perform Chromatin Immunoprecipitation (ChIP) experiments and we may also attempt to modify specific binding sites of these proteins.

References

Abuelo, D.N., Forman, E.N., and Rubin, L.P. (1997). Limb defects and congenital anomalies of the genitalia in an infant with homozygous alpha-thalassemia. Am J Med Genet *68*, 158-161.

Adam, M.P., Chueh, J., El-Sayed, Y.Y., Stenzel, A., Vogel, H., Weaver, D.D., and Hoyme, H.E. (2005). Vascular-type disruptive defects in fetuses with homozygous alpha-thalassemia: report of two cases and review of the literature. Prenat Diagn *25*, 1088-1096.

Andrews, N.C. (1998). The NF-E2 transcription factor. Int J Biochem Cell Biol *30*, 429-432.

Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., and Higgs, D.R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. Embo j *23*, 2841-2852.

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.

Baron, M.H. (2013). Concise Review: early embryonic erythropoiesis: not so primitive after all. Stem Cells *31*, 849-856.

Basu, P., Lung, T.K., Lemsaddek, W., Sargent, T.G., Williams, D.C., Jr., Basu, M., Redmond, L.C., Lingrel, J.B., Haar, J.L., and Lloyd, J.A. (2007). EKLF and KLF2 have compensatory roles in embryonic beta-globin gene expression and primitive erythropoiesis. Blood *110*, 3417-3425.

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.

Benesch, R.E., Ranney, H.M., Benesch, R., and Smith, G.M. (1961). The chemistry of the Bohr effect. II. Some properties of hemoglobin H. J Biol Chem *236*, 2926-2929.

Bertrand, J.Y., Jalil, A., Klaine, M., Jung, S., Cumano, A., and Godin, I. (2005). Three pathways to mature macrophages in the early mouse yolk sac. Blood *106*, 3004-3011.

Bianchi, D.W., Beyer, E.C., Stark, A.R., Saffan, D., Sachs, B.P., and Wolfe, L. (1986). Normal long-term survival with alpha-thalassemia. J Pediatr *108*, 716-718.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E*., et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Bizzarro, M.J., Copel, J.A., Pearson, H.A., Pober, B., and Bhandari, V. (2003). Pulmonary hypoplasia and persistent pulmonary hypertension in the newborn with homozygous alpha-thalassemia: a case report and review of the literature. J Matern Fetal Neonatal Med *14*, 411-416.

Borg, J., Papadopoulos, P., Georgitsi, M., Gutierrez, L., Grech, G., Fanis, P., Phylactides, M., Verkerk, A.J., van der Spek, P.J., Scerri, C.A*., et al.* (2010). Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. Nat Genet *42*, 801-805.

Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J., and Buckle, V.J. (2006). Coregulated human globin genes are frequently in spatial proximity when active. The Journal of cell biology *172*, 177-187.

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. Nature methods *10*, 1213-1218.

Carey, B.W., Markoulaki, S., Hanna, J.H., Faddah, D.A., Buganim, Y., Kim, J., Ganz, K., Steine, E.J., Cassady, J.P., Creyghton, M.P.*, et al.* (2011). Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. Cell stem cell *9*, 588-598.

Carr, S., Rubin, L., Dixon, D., Star, J., and Dailey, J. (1995). Intrauterine therapy for homozygous alpha-thalassemia. Obstet Gynecol *85*, 876-879.

Chang, C.J., and Bouhassira, E.E. (2012). Zinc-finger nuclease-mediated correction of alpha-thalassemia in iPS cells. Blood *120*, 3906-3914.

Chang, C.J., Mitra, K., Koya, M., Velho, M., Desprat, R., Lenz, J., and Bouhassira, E.E. (2011). Production of embryonic and fetal-like red blood cells from human induced pluripotent stem cells. PLoS One *6*, e25761.

Chang, K.H., Huang, A., Hirata, R.K., Wang, P.R., Russell, D.W., and Papayannopoulou, T. (2010). Globin phenotype of erythroid cells derived from human induced pluripotent stem cells. Blood *115*, 2553-2554.

Chen, K., Liu, J., Heck, S., Chasis, J.A., An, X., and Mohandas, N. (2009). Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis. Proc Natl Acad Sci U S A *106*, 17413-17418.

Chen, M.J., Li, Y., De Obaldia, M.E., Yang, Q., Yzaguirre, A.D., Yamada-Inagawa, T., Vink, C.S., Bhandoola, A., Dzierzak, E., and Speck, N.A. (2011). Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. Cell stem cell *9*, 541-552.

Chik, K.W., Shing, M.M., Li, C.K., Leung, T.F., Tsang, K.S., Yuen, H.L., Cheng, S.B., and Yuen, P.M. (1998). Treatment of hemoglobin Bart's hydrops with bone marrow transplantation. J Pediatr *132*, 1039-1042.

Chitayat, D., Silver, M.M., O'Brien, K., Wyatt, P., Waye, J.S., Chiu, D.H., Babul, R., and Thomas, M. (1997). Limb defects in homozygous alpha-thalassemia: report of three cases. Am J Med Genet *68*, 162-167.

Chmait, R.H., Baskin, J.L., Carson, S., Randolph, L.M., and Hamilton, A. (2014). Treatment of alpha(0)-thalassemia (--/--) via serial fetal and post-natal transfusions: Can early fetal intervention improve outcomes? Hematology.

Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J.C., and Keller, G. (1998). A common precursor for hematopoietic and endothelial cells. Development *125*, 725-732.

Christensen, R.D., Del Vecchio, A., and Henry, E. (2012). Expected erythrocyte, platelet and neutrophil values for term and preterm neonates. J Matern Fetal Neonatal Med *25*, 77-79.

Chui, D.H., Mentzer, W.C., Patterson, M., Iarocci, T.A., Embury, S.H., Perrine, S.P., Mibashan, R.S., and Higgs, D.R. (1989). Human embryonic zeta-globin chains in fetal and newborn blood. Blood *74*, 1409-1414.

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

Chui, D.H., Wong, S.C., Chung, S.W., Patterson, M., Bhargava, S., and Poon, M.C. (1986). Embryonic zeta-globin chains in adults: a marker for alpha-thalassemia-1 haplotype due to a greater than 17.5-kb deletion. N Engl J Med *314*, 76-79.

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. Proc Natl Acad Sci U S A *81*, 6188-6191.

Coelho, A., Picanco, 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.

Cole, R.J., and Paul, J. (1966). The effects of erythropoietin on haem synthesis in mouse yolk sac and cultured foetal liver cells. Journal of embryology and experimental morphology *15*, 245-260.

Cui, S., Kolodziej, K.E., Obara, N., Amaral-Psarris, A., Demmers, J., Shi, L., Engel, J.D., Grosveld, F., Strouboulis, J., and Tanabe, O. (2011). Nuclear receptors TR2 and TR4 recruit multiple epigenetic transcriptional corepressors that associate specifically with the embryonic beta-type globin promoters in differentiated adult erythroid cells. Mol Cell Biol *31*, 3298-3311.

Dame, C., Albers, N., and Bartmann, P. (1999a). Candidate gene involved in genital ontogeny in chromosome region 16p13.3. Am J Med Genet *86*, 302-303.

Dame, C., Albers, N., Hasan, C., Bode, U., Eigel, A., Hansmann, M., Brenner, R., and Bartmann, P. (1999b). Homozygous alpha-thalassaemia and hypospadias--common aetiology or incidental association? Long-term survival of Hb Bart's hydrops syndrome leads to new aspects for counselling of alpha-thalassaemic traits. Eur J Pediatr *158*, 217-220.

Davies, J.O., 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. Nature methods *13*, 74-80.

de Bruijn, M.F., Speck, N.A., Peeters, M.C., and Dzierzak, E. (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. Embo j *19*, 2465-2474.

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 alpha globin gene expression. Blood *110*, 4503-4510.

DeSimone, J., Heller, P., Hall, L., and Zwiers, D. (1982). 5-Azacytidine stimulates fetal hemoglobin synthesis in anemic baboons. Proc Natl Acad Sci U S A *79*, 4428-4431.

Diamond, M.P., Cotgrove, I., and Parker, A. (1965). Case of intrauterine death due to alpha-thalassaemia. British medical journal *2*, 278-279.

Drissen, R., von Lindern, 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.

Dwinnell, S.J., Coad, S., Butler, B., Albersheim, S., Wadsworth, L.D., Wu, J.K., and Delisle, M.F. (2011). In Utero diagnosis and management of a fetus with homozygous alpha-Thalassemia in the second trimester: a case report and literature review. J Pediatr Hematol Oncol *33*, e358-360.

Eichler, G.S., Huang, S., and Ingber, D.E. (2003). Gene Expression Dynamics Inspector (GEDI): for integrative analysis of expression profiles. Bioinformatics *19*, 2321-2322.

Ema, M., Yokomizo, T., Wakamatsu, A., Terunuma, T., Yamamoto, M., and Takahashi, S. (2006). Primitive erythropoiesis from mesodermal precursors expressing VE-cadherin, PECAM-1, Tie2, endoglin, and CD34 in the mouse embryo. Blood *108*, 4018-4024.

England, S.J., McGrath, K.E., Frame, J.M., and Palis, J. (2011). Immature erythroblasts with extensive ex vivo self-renewal capacity emerge from the early mammalian fetus. Blood *117*, 2708-2717.

Fathallah, H., Weinberg, R.S., Galperin, Y., Sutton, M., and Atweh, G.F. (2007). Role of epigenetic modifications in normal globin gene regulation and butyrate-mediated induction of fetal hemoglobin. Blood *110*, 3391-3397.

Fei, Y.J., Kutlar, F., Harris, H.F., 2nd, Wilson, M.M., Milana, A., Sciacca, P., Schiliro, G., Masala, B., Manca, L., Altay, C.*, et al.* (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.

Ferkowicz, M.J., and Yoder, M.C. (2005). Blood island formation: longstanding observations and modern interpretations. Exp Hematol *33*, 1041-1047.

Filipe, A., Li, Q., Deveaux, S., Godin, I., Romeo, P.H., Stamatoyannopoulos, G., and Mignotte, V. (1999). Regulation of embryonic/fetal globin genes by nuclear hormone receptors: a novel perspective on hemoglobin switching. Embo j *18*, 687-697.

Fischel-Ghodsian, N., Higgs, D.R., and Beyer, E.C. (1987). Function of a new globin gene. Nature *329*, 397.

Fischel-Ghodsian, N., Vickers, M.A., Seip, M., Winichagoon, P., and Higgs, D.R. (1988). Characterization of two deletions that remove the entire human zeta-alpha globin gene complex (- -THAI and - -FIL). Br J Haematol *70*, 233-238.

Forget, B.G., and Hardison, R. (2009). The Normal Structure and Regulation of Human Globin Gene Clusters. In Disorders of Hemoglobin, M.H. Steinberg, B.G. Forget, D.R. Higgs, and D.J. Weatherall, eds. (New York: Cambridge University Press).

Frame, J.M., McGrath, K.E., and Palis, J. (2013). Erythro-myeloid progenitors: "definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells. Blood Cells Mol Dis *51*, 220-225.

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.

Fucharoen, S., and Winichagoon, P. (1992). Thalassemia in SouthEast Asia: problems and strategy for prevention and control. The Southeast Asian journal of tropical medicine and public health *23*, 647-655.

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. Proc Natl Acad Sci U S A *93*, 12355-12358.

Fung, T.Y., Kin, L.T., Kong, L.C., and Keung, L.C. (1999). Homozygous alpha-thalassemia associated with hypospadias in three survivors. Am J Med Genet *82*, 225-227.

Fung, T.Y., Lau, T.K., Tam, W.H., and Li, C.K. (1998). In utero exchange transfusion in homozygous alpha-thalassaemia: a case report. Prenat Diagn *18*, 838-841.

Funnell, A.P., 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.

Galanello, R., Sanna, M.A., Maccioni, L., Gasperini, D., Melis, M.A., Rosatelli, C., Monni, G., and Cao, A. (1990). Fetal hydrops in Sardinia: implications for genetic counselling. Clinical genetics *38*, 327-331.

Galanello, R., Sanna, S., Perseu, L., Sollaino, M.C., Satta, S., Lai, M.E., Barella, S., Uda, M., Usala, G., Abecasis, G.R.*, et al.* (2009). Amelioration of Sardinian beta0 thalassemia by genetic modifiers. Blood *114*, 3935-3937.

Galarneau, G., Palmer, C.D., Sankaran, V.G., Orkin, S.H., Hirschhorn, J.N., and Lettre, G. (2010). Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. Nat Genet *42*, 1049-1051.

Gale, R.E., Clegg, J.B., and Huehns, E.R. (1979). Human embryonic haemoglobins Gower 1 and Gower 2. Nature *280*, 162-164.

Gallienne, A.E., Dreau, H.M., Schuh, A., Old, J.M., and Henderson, S. (2012). Ten novel mutations in the erythroid transcription factor KLF1 gene associated with increased fetal hemoglobin levels in adults. Haematologica *97*, 340-343.

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.

Ghosh, A., Tang, M.H., Lam, Y.H., Fung, E., and Chan, V. (1994). Ultrasound measurement of placental thickness to detect pregnancies affected by homozygous alpha-thalassaemia-1. Lancet *344*, 988-989.

Giardine, B., van Baal, S., Kaimakis, P., Riemer, C., Miller, W., Samara, M., Kollia, P., Anagnou, N.P., Chui, D.H., Wajcman, H.*, et al.* (2007). HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. Human mutation *28*, 206.

Gourdon, G., Sharpe, J.A., Wells, D., Wood, W.G., and Higgs, D.R. (1994). Analysis of a 70 kb segment of DNA containing the human zeta and alpha-globin genes linked to their regulatory element (HS-40) in transgenic mice. Nucleic Acids Res *22*, 4139-4147.

Greenfest-Allen, E., Malik, J., Palis, J., and Stoeckert, C.J., Jr. (2013). Stat and interferon genes identified by network analysis differentially regulate primitive and definitive erythropoiesis. BMC Syst Biol *7*, 38.

Gumuscu, B., Thompson, E.I., Grovas, A.C., Zach, T.L., Warkentin, P.I., and Coccia, P.F. (2013). Successful Unrelated Cord Blood Transplantation For Homozygous alpha-Thalassemia. J Pediatr Hematol Oncol.

Guy, G., Coady, D.J., Jansen, V., Snyder, J., and Zinberg, S. (1985). alpha-Thalassemia hydrops fetalis: clinical and ultrasonographic considerations. Am J Obstet Gynecol *153*, 500-504.

Harmon, J.V., Jr., Osathanondh, R., and Holmes, L.B. (1995). Symmetrical terminal transverse limb defects: report of a twenty-week fetus. Teratology *51*, 237-242.

Harteveld, C.L., and Higgs, D.R. (2010). Alpha-thalassaemia. Orphanet J Rare Dis *5*, 13.

Hayward, A., Ambruso, D., Battaglia, F., Donlon, T., Eddelman, K., Giller, R., Hobbins, J., Hsia, Y.E., Quinones, R., Shpall, E.*, et al.* (1998). Microchimerism and tolerance following intrauterine transplantation and transfusion for alpha-thalassemia-1. Fetal Diagn Ther *13*, 8-14.

He, Z., Lian, L., Asakura, T., and Russell, J.E. (2000). Functional effects of replacing human alpha- and beta-globins with their embryonic globin homologues in defined haemoglobin heterotetramers. Br J Haematol *109*, 882-890.

He, Z., and Russell, J.E. (2001). Expression, purification, and characterization of human hemoglobins Gower-1 (zeta(2)epsilon(2)), Gower-2 (alpha(2)epsilon(2)), and Portland-2 (zeta(2)beta(2)) assembled in complex transgenic-knockout mice. Blood *97*, 1099-1105.

He, Z., Song, D., van Zalen, S., and Russell, J.E. (2014). Structural determinants of human zeta-globin mRNA stability. J Hematol Oncol *7*, 35.

Hecht, F., Motulsky, A.G., Lemire, R.J., and Shepard, T.E. (1966). Predominance of hemoglobin Gower 1 in early human embryonic development. Science *152*, 91-92.

Higgs, D.R. (2009a). The Molecular Basis of alpha-thalassemia. In Disorders of Hemoglobin, M.H. Steinberg, B.G. Forget, D.R. Higgs, and D.J. Weatherall, eds. (New York: Cambridge University Press), pp. 241-265.

Higgs, D.R. (2009b). The Pathophysiology and Clinical Features of alpha-thalassemia. In Disorders of Hemoglobin, M.H. Steinberg, B.G. Forget, D.R. Higgs, and D.J. Weatherall, eds. (United State of America: Cambridge University Press), p. 266.

Higgs, D.R., Garrick, D., Anguita, E., De Gobbi, M., Hughes, J., Muers, M., Vernimmen, D., Lower, K., Law, M., Argentaro, A.*, et al.* (2005). Understanding alpha-globin gene regulation: Aiming to improve the management of thalassemia. Ann N Y Acad Sci *1054*, 92-102.

Higgs, D.R., Sharpe, J.A., and Wood, W.G. (1998). Understanding alpha globin gene expression: a step towards effective gene therapy. Semin Hematol *35*, 93-104.

Higgs, D.R., and Wood, W.G. (2008a). Genetic complexity in sickle cell disease. Proc Natl Acad Sci U S A *105*, 11595-11596.

Higgs, D.R., and Wood, W.G. (2008b). Long-range regulation of alpha globin gene expression during erythropoiesis. Curr Opin Hematol *15*, 176-183.

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.

Hill, A.V., Nicholls, R.D., Thein, S.L., and Higgs, D.R. (1985). Recombination within the human embryonic xi-globin locus: a common xi-xi chromosome produced by gene conversion of the psi xi gene. Cell *42*, 809-819.

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

Horton, B.F., Thompson, R.B., Dozy, A.M., Nechtman, C.M., Nichols, E., and Huisman, T.H. (1962). Inhomogeneity of hemoglobin. VI. The minor hemoglobin components of cord blood. Blood *20*, 302-314.

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. Proc Natl Acad Sci U S A *95*, 14669-14674.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007a). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome biology *8*, R183.

Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C*., et al.* (2007b). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res *35*, W169-175.

Huber, T.L., Kouskoff, V., Fehling, H.J., Palis, J., and Keller, G. (2004). Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. Nature *432*, 625-630.

Huehns, E.R., Dance, N., Beaven, G.H., Hecht, F., and Motulsky, A.G. (1964). Human embryonic hemoglobins. Cold Spring Harbor symposia on quantitative biology *29*, 327-331.

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.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet.

Ireland, J.H., Luo, H.Y., Chui, D.H., Chu, B., Yuen, J., and Hsia, Y.E. (1993). Detection of the (--SEA) double alpha-globin gene deletion by a simple immunologic assay for embryonic zeta-globin chains. Am J Hematol *44*, 22-28.

Isern, J., Fraser, S.T., He, Z., and Baron, M.H. (2008). The fetal liver is a niche for maturation of primitive erythroid cells. Proc Natl Acad Sci U S A *105*, 6662-6667.

Isern, J., Fraser, S.T., He, Z., Zhang, H., and Baron, M.H. (2010). Dose-dependent regulation of primitive erythroid maturation and identity by the transcription factor Eklf. Blood *116*, 3972-3980.

Isern, J., He, Z., Fraser, S.T., Nowotschin, S., Ferrer-Vaquer, A., Moore, R., Hadjantonakis, A.K., Schulz, V., Tuck, D., Gallagher, P.G*., et al.* (2011). Single-lineage transcriptome analysis reveals key regulatory pathways in primitive erythroid progenitors in the mouse embryo. Blood *117*, 4924-4934.

Jackson D.N., Strauss A.A., Groncy P.K., Bianchi D.W., and Akabutus J. (1990). Outcome of neonatal survivors with homozygous alpha-thalassemia. Pediatric Research *27*, 266A.

Jane, S.M., Nienhuis, A.W., and Cunningham, J.M. (1995). Hemoglobin switching in man and chicken is mediated by a heteromeric complex between the ubiquitous transcription factor CP2 and a developmentally specific protein. Embo j *14*, 97-105.

Jawaid, K., Wahlberg, K., Thein, S.L., and Best, S. (2010). Binding patterns of BCL11A in the globin and GATA1 loci and characterization of the BCL11A fetal hemoglobin locus. Blood Cells Mol Dis *45*, 140-146.

Jiang, J., Best, S., Menzel, S., Silver, N., Lai, M.I., Surdulescu, G.L., Spector, T.D., and Thein, S.L. (2006). cMYB is involved in the regulation of fetal hemoglobin production in adults. Blood *108*, 1077-1083.

Jopling, J., Henry, E., Wiedmeier, S.E., and Christensen, R.D. (2009). Reference ranges for hematocrit and blood hemoglobin concentration during the neonatal period: data from a multihospital health care system. Pediatrics *123*, e333-337.

Joshi, D.D., Nickerson, H.J., and McManus, M.J. (2004). Hydrops fetalis caused by homozygous alpha-thalassemia and Rh antigen alloimmunization: report of a survivor and literature review. Clin Med Res *2*, 228-232.

Kattamis, C., Metaxotou-Mavromati, A., Tsiarta, E., Metaxatou, C., Wasi, P., Wood, W.G., Pressley, L., Higgs, D.R., Clegg, J.B., and Weatherall, D.J. (1980). Haemoglobin Bart's hydrops syndrome in Greece. British medical journal *281*, 268-270.

Kennedy, M., D'Souza, S.L., Lynch-Kattman, M., Schwantz, S., and Keller, G. (2007). Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. Blood *109*, 2679-2687.

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 Res *12*, 996-1006.

Kielman, M.F., Smits, R., Hof, I., and Bernini, L.F. (1996). Characterization and comparison of the human and mouse Dist1/alpha-globin complex reveals a tightly packed multiple gene cluster containing differentially expressed transcription units. Genomics *32*, 341-351.

Kiledjian, M., Wang, X., and Liebhaber, S.A. (1995). Identification of two KH domain proteins in the alpha-globin mRNP stability complex. Embo j *14*, 4357-4364.

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.

Ko, T.M., Hsieh, F.J., Hsu, P.M., and Lee, T.Y. (1991). Molecular characterization of severe alpha-thalassemias causing hydrops fetalis in Taiwan. Am J Med Genet *39*, 317-320.

Ko, T.M., Tseng, L.H., Hsu, P.M., Hwa, H.L., Lee, T.Y., and Chuang, S.M. (1995). Ultrasonographic scanning of placental thickness and the prenatal diagnosis of homozygous alpha-thalassaemia 1 in the second trimester. Prenat Diagn *15*, 7-10.

Kou, K.O., Lee, H., Lau, B., Wong, W.S., Kan, A., Tang, M., Lau, E.T., Poon, C.F., and Leung, K.Y. (2013). Two Unusual Cases of Haemoglobin Bart's Hydrops Fetalis due to Uniparental Disomy or Non-Paternity. Fetal Diagn Ther.

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.* (2012). Nprl3 is required for normal development of the cardiovascular system. Mammalian genome : official journal of the International Mammalian Genome Society *23*, 404-415.

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.

Lam, T.K., Chan, V., Fok, T.F., Li, C.K., and Feng, C.S. (1992). Long-term survival of a baby with homozygous alpha-thalassemia-1. Acta Haematol *88*, 198-200.

Lam, Y.H., Tang, M.H., Sin, S.Y., Ghosh, A., and Lee, C.P. (1997). Limb reduction defects in fetuses with homozygous alpha-thalassaemia-1. Prenat Diagn *17*, 1143-1146.

Lapillonne, H., Kobari, L., Mazurier, C., Tropel, P., Giarratana, M.C., Zanella-Cleon, I., Kiger, L., Wattenhofer-Donze, 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.

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, S.U., Maeda, M., Ishikawa, Y., Li, S.M., Wilson, A., Jubb, A.M., Sakurai, N., Weng, L., Fiorini, E., Radtke, F.*, et al.* (2013). LRF-mediated Dll4 repression in erythroblasts is necessary for hematopoietic stem cell maintenance. Blood *121*, 918-929.

Lee, S.Y., Chow, C.B., Li, C.K., and Chiu, M.C. (2007). Outcome of intensive care of homozygous alpha-thalassaemia without prior intra-uterine therapy. J Paediatr Child Health *43*, 546-550.

Lee, S.Y., Li, C.K., Ling, S.C., and Shiu, Y.K. (2009). Survival of homozygous alpha-thalassemia with aplasia/hypoplasia of phalanges and jejunal atresia. J Matern Fetal Neonatal Med *22*, 711-713.

Lemischka, I.R. (1991). Clonal, in vivo behavior of the totipotent hematopoietic stem cell. Seminars in immunology *3*, 349-355.

Lemmens-Zygulska, M., Eigel, A., Helbig, B., Sanguansermsri, T., Horst, J., and Flatz, G. (1996). Prevalence of alpha-thalassemias in northern Thailand. Hum Genet *98*, 345-347.

Lettre, G., Sankaran, V.G., Bezerra, M.A., Araujo, 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, W.C., Oepkes, D., Seaward, G., and Ryan, G. (2002). Serial sonographic findings of four fetuses with homozygous alpha-thalassemia-1 from 21 weeks onwards. Ultrasound Obstet Gynecol *19*, 56-59.

Ley, T.J., DeSimone, J., Anagnou, N.P., Keller, G.H., Humphries, R.K., Turner, P.H., Young, N.S., Keller, P., and Nienhuis, A.W. (1982). 5-azacytidine selectively increases gamma-globin synthesis in a patient with beta+ thalassemia. N Engl J Med *307*, 1469-1475.

Ley, T.J., DeSimone, J., Noguchi, C.T., Turner, P.H., Schechter, A.N., Heller, P., and Nienhuis, A.W. (1983). 5-Azacytidine increases gamma-globin synthesis and reduces the proportion of dense cells in patients with sickle cell anemia. Blood *62*, 370-380.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, L., Lee, J.Y., Gross, J., Song, S.H., Dean, A., and Love, P.E. (2010). A requirement for Lim domain binding protein 1 in erythropoiesis. The Journal of experimental medicine *207*, 2543-2550.

Liang, S.T., Wong, V.C., So, W.W., Ma, H.K., Chan, V., and Todd, D. (1985). Homozygous alpha-thalassaemia: clinical presentation, diagnosis and management. A review of 46 cases. Br J Obstet Gynaecol *92*, 680-684.

Liao, C., Wei, J., Li, Q., Li, J., Li, L., and Li, D. (2007). Nonimmune hydrops fetalis diagnosed during the second half of pregnancy in Southern China. Fetal Diagn Ther *22*, 302-305.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Liberzon, A. (2014). A description of the Molecular Signatures Database (MSigDB) Web site. Methods Mol Biol *1150*, 153-160.

Lie-Injo Luan, E.N.G. (1959). Haemoglobin of new-born infants in Indonesia. Nature *183*, 1125-1126.

Lin, C.S., Lim, S.K., D'Agati, V., and Costantini, F. (1996). Differential effects of an erythropoietin receptor gene disruption on primitive and definitive erythropoiesis. Genes Dev *10*, 154-164.

Liu, C.A., Huang, H.C., and Chou, Y.Y. (2002). Retrospective analysis of 17 liveborn neonates with hydrops fetalis. Chang Gung Med J *25*, 826-831.

Lucitti, J.L., Jones, E.A., Huang, C., Chen, J., Fraser, S.E., and Dickinson, M.E. (2007). Vascular remodeling of the mouse yolk sac requires hemodynamic force. Development *134*, 3317-3326.

Lucke, T., Pfister, S., and Durken, M. (2005). Neurodevelopmental outcome and haematological course of a long-time survivor with homozygous alpha-thalassaemia: case report and review of the literature. Acta Paediatr *94*, 1330-1333.

Luckett, W.P. (1978). Origin and differentiation of the yolk sac and extraembryonic mesoderm in presomite human and rhesus monkey embryos. The American journal of anatomy *152*, 59-97.

Lung, H.Y., Meeus, I.S., Weinberg, R.S., and Atweh, G.F. (2000). In vivo silencing of the human gamma-globin gene in murine erythroid cells following retroviral transduction. Blood Cells Mol Dis *26*, 613-619.

Lunter, G., and Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res *21*, 936-939.

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.

MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res *42*, D986-992.

Maherali, N., and Hochedlinger, K. (2008). Guidelines and techniques for the generation of induced pluripotent stem cells. Cell stem cell *3*, 595-605.

Malik, J., Kim, A.R., Tyre, K.A., Cherukuri, A.R., and Palis, J. (2013). Erythropoietin critically regulates the terminal maturation of murine and human primitive erythroblasts. Haematologica *98*, 1778-1787.

McGrath, K.E., Frame, J.M., Fegan, K.H., Bowen, J.R., Conway, S.J., Catherman, S.C., Kingsley, P.D., Koniski, A.D., and Palis, J. (2015). Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. Cell reports *11*, 1892-1904.

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 beta-globin locus in the mammalian embryo. Blood *117*, 4600-4608.

McGrath, K.E., Kingsley, P.D., Koniski, A.D., Porter, R.L., Bushnell, T.P., and Palis, J. (2008). Enucleation of primitive erythroid cells generates a transient population of "pyrenocytes" in the mammalian fetus. Blood *111*, 2409-2417.

McGrath, K.E., Koniski, A.D., Malik, J., and Palis, J. (2003). Circulation is established in a stepwise pattern in the mammalian embryo. Blood *101*, 1669-1676.

McGrath, K.E., and Palis, J. (2005). Hematopoiesis in the yolk sac: more than meets the eye. Exp Hematol *33*, 1021-1028.

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. Nat Genet *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*, e96-108.

Michlitsch, J.G., and Walters, M.C. (2008). Recent advances in bone marrow transplantation in hemoglobinopathies. Current molecular medicine *8*, 675-689.

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. The Journal of clinical investigation *78*, 51-60.

Miller, J.L. (2004). A genome-based approach for the study of erythroid biology and disease. Blood Cells Mol Dis *32*, 341-343.

Mohandas, N., and Gallagher, P.G. (2008). Red cell membrane: past, present, and future. Blood *112*, 3939-3948.

Moreau-Gaudry, F., Xia, P., Jiang, G., Perelman, N.P., Bauer, G., Ellis, J., Surinya, K.H., Mavilio, F., Shen, C.K., and Malik, P. (2001). High-level erythroid-specific gene expression in primary human and murine hematopoietic cells with self-inactivating lentiviral vectors. Blood *98*, 2664-2672.

Mucenski, M.L., McLain, K., Kier, A.B., Swerdlow, S.H., Schreiner, C.M., Miller, T.A., Pietryga, D.W., Scott, W.J., Jr., and Potter, S.S. (1991). A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. Cell *65*, 677-689.

Mukai, H.Y., Motohashi, H., Ohneda, O., Suzuki, N., Nagano, M., and Yamamoto, M. (2006). Transgene insertion in proximity to the c-myb gene disrupts erythroid-megakaryocytic lineage bifurcation. Mol Cell Biol *26*, 7953-7965.

Muller, A.M., Medvinsky, A., Strouboulis, J., Grosveld, F., and Dzierzak, E. (1994). Development of hematopoietic stem cell activity in the mouse embryo. Immunity *1*, 291-301.

Nakayama, R., Yamada, D., Steinmiller, V., Hsia, E., and Hale, R.W. (1986). Hydrops fetalis secondary to Bart hemoglobinopathy. Obstet Gynecol *67*, 176-180.

Naqvi A, Waye J. S., Morrow E, and Nisbet-Brown Olivieri N. F. (1997). Normal development of an infant with homozygous alpha-thalassemia. Blood *90*, 132a.

Ng, P.C., Fok, T.F., Lee, C.H., Cheung, K.L., Li, C.K., So, K.W., Wong, W., and Yuen, P.M. (1998). Is homozygous alpha-thalassaemia a lethal condition in the 1990s? Acta Paediatr *87*, 1197-1199.

Nilson, D.G., Sabatino, D.E., Bodine, D.M., and Gallagher, P.G. (2006). Major erythrocyte membrane protein genes in EKLF-deficient mice. Exp Hematol *34*, 705-712.

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

Nuinoon, M., Makarasara, W., Mushiroda, T., Setianingsih, I., Wahidiyat, P.A., Sripichai, O., Kumasaka, N., Takahashi, A., Svasti, S., Munkongdee, T.*, et al.* (2010). A genome-wide association identified the common genetic variants influence disease severity in beta0-thalassemia/hemoglobin E. Hum Genet *127*, 303-314.

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.

Palis, J. (2014). Primitive and definitive erythropoiesis in mammals. Front Physiol *5*, 3.

Palis, J., Malik, J., McGrath, K.E., and Kingsley, P.D. (2010). Primitive erythropoiesis in the mammalian embryo. Int J Dev Biol *54*, 1011-1018.

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.

Pang, C.J., Lemsaddek, W., Alhashem, Y.N., Bondzi, C., Redmond, L.C., Ah-Son, N., Dumur, C.I., Archer, K.J., Haar, J.L., Lloyd, J.A*., et al.* (2012). Kruppel-like factor 1 (KLF1), KLF2, and Myc control a regulatory network essential for embryonic erythropoiesis. Mol Cell Biol *32*, 2628-2644.

Pera, M.F. (2008). Stem cells. A new year and a new era. Nature *451*, 135-136.

Perkins, A.C., Sharpe, A.H., and Orkin, S.H. (1995). Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature *375*, 318-322.

Perrine, S.P., Ginder, G.D., Faller, D.V., Dover, G.H., Ikuta, T., Witkowska, H.E., Cai, S.P., Vichinsky, E.P., and Olivieri, N.F. (1993). A short-term trial of butyrate to stimulate fetal-globin-gene expression in the beta-globin disorders. N Engl J Med *328*, 81-86.

Perrine, S.P., Olivieri, N.F., Faller, D.V., Vichinsky, E.P., Dover, G.J., and Ginder, G.D. (1994). Butyrate derivatives. New agents for stimulating fetal globin production in the beta-globin disorders. The American journal of pediatric hematology/oncology *16*, 67-71.

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 epsilon----gamma-globin switches in primitive and definite erythropoietic lineage. Nature *313*, 235-238.

Peters, L.L., White, R.A., Birkenmeier, C.S., Bloom, M.L., Lux, S.E., and Barker, J.E. (1992). Changing patterns in cytoskeletal mRNA expression and protein synthesis during murine erythropoiesis in vivo. Proc Natl Acad Sci U S A *89*, 5749-5753.

Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H., and Costantini, F. (1995). Development of hematopoietic cells lacking transcription factor GATA-1. Development *121*, 163-172.

Platt, O.S. (2008). Hydroxyurea for the treatment of sickle cell anemia. N Engl J Med *358*, 1362-1369.

Platt, O.S., Orkin, S.H., Dover, G., Beardsley, G.P., Miller, B., and Nathan, D.G. (1984). Hydroxyurea enhances fetal hemoglobin production in sickle cell anemia. The Journal of clinical investigation *74*, 652-656.

Pondel, M.D., Proudfoot, N.J., Whitelaw, C., and Whitelaw, E. (1992). The developmental regulation of the human zeta-globin gene in transgenic mice employing beta-galactosidase as a reporter gene. Nucleic Acids Res *20*, 5655-5660.

Pongtanakul, B., Sanpakit, K., Chongkolwatana, V., and Viprakasit, V. (2013). Normal cognitive functioning in a patient with Hb Bart's hydrops successfully cured by hematopoietic SCT. Bone Marrow Transplant.

Premawardhena, A., Fisher, C.A., Olivieri, N.F., de Silva, S., Arambepola, M., Perera, W., O'Donnell, A., Peto, T.E., Viprakasit, V., Merson, L*., et al.* (2005). Haemoglobin E beta thalassaemia in Sri Lanka. Lancet *366*, 1467-1470.

Pressley, L., Higgs, D.R., Clegg, J.B., and Weatherall, D.J. (1980). Gene deletions in alpha thalassemia prove that the 5' zeta locus is functional. Proc Natl Acad Sci U S A *77*, 3586-3589.

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, C., Olivier, E.N., Velho, M., and Bouhassira, E.E. (2008). Globin switches in yolk sac-like primitive and fetal-like definitive red blood cells produced from human embryonic stem cells. Blood *111*, 2400-2408.

Redmond, L.C., Haar, J.L., Giebel, M.L., Dumur, C.I., Basu, P., Ware, J.L., and Lloyd, J.A. (2006). Isolation of erythroid cells from the mouse embryonic yolk sac by laser capture microdissection and subsequent microarray hybridization. Blood Cells Mol Dis *37*, 27-32.

Ren, S., Luo, X.N., and Atweh, G.F. (1993). The major regulatory element upstream of the alpha-globin gene has classical and inducible enhancer activity. Blood *81*, 1058-1066.

Robb, L., Lyons, I., Li, R., Hartley, L., Kontgen, F., Harvey, R.P., Metcalf, D., and Begley, C.G. (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. Proc Natl Acad Sci U S A *92*, 7075-7079.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Rombel, I., Hu, K.Y., Zhang, Q., Papayannopoulou, T., Stamatoyannopoulos, G., and Shen, C.K. (1995). Transcriptional activation of human adult alpha-globin genes by hypersensitive site-40 enhancer: function of nuclear factor-binding motifs occupied in erythroid cells. Proc Natl Acad Sci U S A *92*, 6454-6458.

Rossant, J. (2011). The impact of developmental biology on pluripotent stem cell research: successes and challenges. Dev Cell *21*, 20-23.

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

Russell, J.E., Morales, J., Makeyev, A.V., and Liebhaber, S.A. (1998). Sequence divergence in the 3' untranslated regions of human zeta- and alpha-globin mRNAs mediates a difference in their stabilities and contributes to efficient alpha-to-zeta gene development switching. Mol Cell Biol *18*, 2173-2183.

Sabath, D.E., Koehler, K.M., and Yang, W.Q. (1996). Structure and function of the zeta-globin upstream regulatory element. Nucleic Acids Res *24*, 4978-4986.

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.

Sankaran, V.G., Menne, T.F., Scepanovic, D., Vergilio, J.A., Ji, P., Kim, J., Thiru, P., Orkin, S.H., Lander, E.S., and Lodish, H.F. (2011a). 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., 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., and Nathan, D.G. (2010). Thalassemia: an overview of 50 years of clinical research. Hematol Oncol Clin North Am *24*, 1005-1020.

Sankaran, V.G., and Orkin, S.H. (2013). The switch from fetal to adult hemoglobin. Cold Spring Harb Perspect Med *3*, a011643.

Sankaran, V.G., Xu, J., Byron, R., Greisman, H.A., Fisher, C., Weatherall, D.J., Sabath, D.E., Groudine, M., Orkin, S.H., Premawardhena, A.*, et al.* (2011b). A functional element necessary for fetal hemoglobin silencing. N Engl J Med *365*, 807-814.

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.

Seldon, W.A., Rubinstein, C., and Fraser, A.A. (1962). The incidence of atrial septal defect in adults. British heart journal *24*, 557-560.

Sharma, R.S., Yu, V., and Walters, W.A. (1979). Haemoglobin Bart's hydrops fetalis syndrome in an infant of Greek origin and prenatal diagnosis of alpha-thalassaemia. The Medical journal of Australia *2*, 404, 433-404.

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.

Sharpe, J.A., Summerhill, R.J., Vyas, P., Gourdon, G., Higgs, D.R., and Wood, W.G. (1993a). Role of upstream DNase I hypersensitive sites in the regulation of human alpha globin gene expression. Blood *82*, 1666-1671.

Sharpe, J.A., Wells, D.J., Whitelaw, E., Vyas, P., Higgs, D.R., and Wood, W.G. (1993b). Analysis of the human alpha-globin gene cluster in transgenic mice. Proc Natl Acad Sci U S A *90*, 11262-11266.

Sher, G.D., Ginder, G.D., Little, J., Yang, S., Dover, G.J., and Olivieri, N.F. (1995). Extended therapy with intravenous arginine butyrate in patients with beta-hemoglobinopathies. N Engl J Med *332*, 1606-1610.

Shimizu, R., Takahashi, S., Ohneda, K., Engel, J.D., and Yamamoto, M. (2001). In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. Embo j *20*, 5250-5260.

Shivdasani, R.A., Mayer, E.L., and Orkin, S.H. (1995). Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature *373*, 432-434.

Singer, S.T., Styles, L., Bojanowski, J., Quirolo, K., Foote, D., and Vichinsky, E.P. (2000). Changing outcome of homozygous alpha-thalassemia: cautious optimism. J Pediatr Hematol Oncol *22*, 539-542.

Sohan, K., Billington, M., Pamphilon, D., Goulden, N., and Kyle, P. (2002). Normal growth and development following in utero diagnosis and treatment of homozygous alpha-thalassaemia. BJOG *109*, 1308-1310.

Spangrude, G.J., Smith, L., Uchida, N., Ikuta, K., Heimfeld, S., Friedman, J., and Weissman, I.L. (1991). Mouse hematopoietic stem cells. Blood *78*, 1395-1402.

Stadhouders, R., Thongjuea, S., Andrieu-Soler, C., Palstra, R.J., Bryne, J.C., van den Heuvel, A., Stevens, M., de Boer, E., Kockx, C., van der Sloot, A.*, et al.* (2012). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. Embo j *31*, 986-999.

Stamatoyannopoulos, G., Navas, P.A., and Li, Q. (2009). Molecular and Cellular Basis of Hemoglobin switching. In Disorders of Hemoglobin, M.H. Steinberg, B.G. Forget, D.R. Higgs, and D.J. Weatherall, eds. (New York: Cambridge University Press).

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A.*, et al.* (2002). The generic genome browser: a building block for a model organism system database. Genome Res *12*, 1599-1610.

Steinberg, M.H., and Nagel, R.L. (2009). Hemoglobins of the Embryos, Fetus, and Adult. In Disorders of Hemoglobin, M.H. Steinberg, B.G. Forget, D.R. Higgs, and D.J. Weatherall, eds. (New York: Cambridge University Press).

Steiner, R., and Vogel, H. (1973). On the kinetics of erythroid cell differentiation in fetal mice. I. Microspectrophotometric determination of the hemoglobin content in erythroid cells during gestation. Journal of cellular physiology *81*, 323-338.

Suwanrath-Kengpol, C., Kor-anantakul, O., Suntharasaj, T., and Leetanaporn, R. (2005). Etiology and outcome of non-immune hydrops fetalis in southern Thailand. Gynecol Obstet Invest *59*, 134-137.

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 Res *20*, 1052-1063.

Tan, S.L., Tseng, A.M., and Thong, P.W. (1989). Bart's hydrops fetalis--clinical presentation and management--an analysis of 25 cases. Aust N Z J Obstet Gynaecol *29*, 233-237.

Tanabe, O., Katsuoka, F., Campbell, A.D., Song, W., Yamamoto, M., Tanimoto, K., and Engel, J.D. (2002). An embryonic/fetal beta-type globin gene repressor contains a nuclear receptor TR2/TR4 heterodimer. Embo j *21*, 3434-3442.

Tanabe, O., McPhee, D., Kobayashi, S., Shen, Y., Brandt, W., Jiang, X., Campbell, A.D., Chen, Y.T., Chang, C., Yamamoto, M.*, et al.* (2007). Embryonic and fetal beta-globin gene repression by the orphan nuclear receptors, TR2 and TR4. Embo j *26*, 2295-2306.

Tang, W., Luo, H.Y., Albitar, M., Patterson, M., Eng, B., Waye, J.S., Liebhaber, S.A., Higgs, D.R., and Chui, D.H. (1992). Human embryonic zeta-globin chain expression in deletional alpha-thalassemias. Blood *80*, 517-522.

Tang, X.B., Feng, D.X., Di, L.J., Huang, Y., Fu, X.H., Liu, G., Tang, Y., Liu, D.P., and Liang, C.C. (2007). HS-48 alone has no enhancement role on the expression of human alpha-globin gene cluster. Blood Cells Mol Dis *38*, 32-36.

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

Thein, S.L., Menzel, S., Peng, X., Best, S., Jiang, J., Close, J., Silver, N., Gerovasilli, A., Ping, C., Yamaguchi, M.*, et al.* (2007). Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. Proc Natl Acad Sci U S A *104*, 11346-11351.

Thornley, I., Lehmann, L., Ferguson, W.S., Davis, I., Forman, E.N., and Guinan, E.C. (2003). Homozygous alpha-thalassemia treated with intrauterine transfusions and postnatal hematopoietic stem cell transplantation. Bone Marrow Transplant *32*, 341-342.

Thumasathit, B., Nondasuta, A., Silpisornkosol, S., Lousuebsakul, B., Unchalipongse, P., and Mangkornkanok, M. (1968). Hydrops fetalis associated with Bart's hemoglobin in northern Thailand. J Pediatr *73*, 132-138.

Tober, J., Koniski, A., McGrath, K.E., Vemishetti, R., Emerson, R., de Mesy-Bentley, K.K., Waugh, R., and Palis, J. (2007). The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. Blood *109*, 1433-1441.

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.

Tongsong, T., Charoenkwan, P., Sirivatanapa, P., Wanapirak, C., Piyamongkol, W., Sirichotiyakul, S., Srisupundit, K., Tongprasert, F., Luewan, S., Ratanasiri, T.*, et al.* (2013). Effectiveness of the model for prenatal control of severe thalassemia. Prenat Diagn *33*, 477-483.

Touraine, J.L., Raudrant, D., Royo, C., Rebaud, A., Roncarolo, M.G., Souillet, G., Philippe, N., Touraine, F., and Betuel, H. (1989). In-utero transplantation of stem cells in bare lymphocyte syndrome. Lancet *1*, 1382.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trimborn, T., Gribnau, J., Grosveld, F., and Fraser, P. (1999). Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci. Genes Dev *13*, 112-124.

Tsang, A.P., Fujiwara, Y., Hom, D.B., and Orkin, S.H. (1998). Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. Genes Dev *12*, 1176-1188.

Tuchinda, S., Nagai, K., and Lehmann, H. (1975). Oxygen dissociation curve of haemoglobin Portland. FEBS letters *49*, 390-391.

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.

Utsch, B., Albers, N., Dame, C., Bartmann, P., Lentze, M.J., and Ludwig, M. (2001). Homozygous alpha-thalassemia associated with hypospadias: SEA-type deletion does not affect expression of the -14 gene and loss of the straight theta1-globin gene on 16p13.3 is compensated by its duplicate straight theta2 on chromosome 10. Am J Med Genet *101*, 286-287.

Vaeusorn, O., Fucharoen, S., and Ruangpiroj, T. (1985). Fetal pathology and maternal morbidity in hemoglobin Bart's hydrops fetalis: an analysis of 65 cases. In International Conference on Thalassemia (Bangkok, Thailand).

van Dijk, T.B., Gillemans, N., Pourfarzad, F., van Lom, K., von Lindern, M., Grosveld, F., and Philipsen, S. (2010). Fetal globin expression is regulated by Friend of Prmt1. Blood *116*, 4349-4352.

Van Handel, B., Prashad, S.L., Hassanzadeh-Kiabi, N., Huang, A., Magnusson, M., Atanassova, B., Chen, A., Hamalainen, E.I., and Mikkola, H.K. (2010). The first trimester human placenta is a site for terminal maturation of primitive erythroid cells. Blood *116*, 3321-3330.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J *26*, 2041-2051.

Vernimmen, D., Lynch, M.D., De Gobbi, M., Garrick, D., Sharpe, J.A., Sloane-Stanley, J.A., Smith, A.J., 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., Smith, A.J., 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.

Viprakasit, V., Ekwattanakit, S., Riolueang, S., Chalaow, N., Fisher, C., Lower, K., Kanno, H., Tachavanich, K., Bejrachandra, S., Saipin, J*., et al.* (2014). Mutations in Kruppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression. Blood *123*, 1586-1595.

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.

Waggoner, S.A., and Liebhaber, S.A. (2003). Regulation of alpha-globin mRNA stability. Exp Biol Med (Maywood) *228*, 387-395.

Wahlberg, K., Jiang, J., Rooks, H., Jawaid, K., Matsuda, F., Yamaguchi, M., Lathrop, M., Thein, S.L., and Best, S. (2009). The HBS1L-MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells. Blood *114*, 1254-1262.

Wallace, H.A., Marques-Kranc, F., Richardson, M., Luna-Crespo, F., Sharpe, J.A., Hughes, J., Wood, W.G., Higgs, D.R., and Smith, A.J. (2007). Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. Cell *128*, 197-209.

Wang, C., and Ryan, G. (2009). Transfusion medicine illustrated: Intrauterine transfusion for homozygous alpha(0) thalassemia reverses hydrops fetalis. Transfusion *49*, 1043-1044.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38*, e164.

Warren, A.J., Colledge, W.H., Carlton, M.B., Evans, M.J., Smith, A.J., and Rabbitts, T.H. (1994). The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. Cell *78*, 45-57.

Wasi, P., Na-Nakorn, S., and Pootrakul, P. (1974). The alpha thalassemias. Clinical Haematology *3*, 383-410.

Waugh, R.E., Mantalaris, A., Bauserman, R.G., Hwang, W.C., and Wu, J.H. (2001). Membrane instability in late-stage erythropoiesis. Blood *97*, 1869-1875.

Waye, J.S., Eng, B., and Chui, D.H. (1992). Identification of an extensive zeta-alpha globin gene deletion in a Chinese individual. Br J Haematol *80*, 378-380.

Weatherall, D.J. (2001). Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. Nature reviews Genetics *2*, 245-255.

Weatherall, D.J., and Clegg, J.B. (2001). The thalassemia syndromes, Fourth edn (Malden, MA: Blackwell Science).

Weatherall, D.J., and Williams, T.N. (2009). Population Genetics and Global Health Burden. In Disorders of Hemoglobin, M.H. Steinberg, B.G. Forget, D.R. Higgs, and D.J. Weatherall, eds. (New York: Cambridge University Press).

Wen, S.C., Roder, K., Hu, K.Y., Rombel, I., Gavva, N.R., Daftari, P., Kuo, Y.Y., Wang, C., and Shen, C.K. (2000). Loading of DNA-binding factors to an erythroid enhancer. Mol Cell Biol *20*, 1993-2003.

Westgren, M., Ringden, O., Eik-Nes, S., Ek, S., Anvret, M., Brubakk, A.M., Bui, T.H., Giambona, A., Kiserud, T., Kjaeldgaard, A.*, et al.* (1996). Lack of evidence of permanent engraftment after in utero fetal stem cell transplantation in congenital hemoglobinopathies. Transplantation *61*, 1176-1179.

Winichagoon, P., Fucharoen, S., and Wasi, P. (1992). The molecular basis of alpha-thalassemia in Thailand. The Southeast Asian journal of tropical medicine and public health *23 Suppl 2*, 7-13.

Wong, P.M., Chung, S.W., Reicheld, S.M., and Chui, D.H. (1986). Hemoglobin switching during murine embryonic development: evidence for two populations of embryonic erythropoietic progenitor cells. Blood *67*, 716-721.

Wu, H., Liu, X., Jaenisch, R., and Lodish, H.F. (1995). Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. Cell *83*, 59-67.

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.

Yi, J.S., Moertel, C.L., and Baker, K.S. (2009). Homozygous alpha-thalassemia treated with intrauterine transfusions and unrelated donor hematopoietic cell transplantation. J Pediatr *154*, 766-768.

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.

Yoshida, H., Kawane, K., Koike, M., Mori, Y., Uchiyama, Y., and Nagata, S. (2005). Phosphatidylserine-dependent engulfment by macrophages of nuclei from erythroid precursor cells. Nature *437*, 754-758.

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.

Zambidis, E.T., Peault, B., Park, T.S., Bunz, F., and Civin, C.I. (2005). Hematopoietic differentiation of human embryonic stem cells progresses through sequential hematoendothelial, primitive, and definitive stages resembling human yolk sac development. Blood *106*, 860-870.

Zhang, H.M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H., and Guo, A.Y. (2012). AnimalTFDB: a comprehensive animal transcription factor database. Nucleic Acids Res *40*, D144-149.

Zhang, J., Socolovsky, M., Gross, A.W., and Lodish, H.F. (2003). Role of Ras signaling in erythroid differentiation of mouse fetal liver cells: functional analysis by a flow cytometry-based novel culture system. Blood *102*, 3938-3946.

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, Q., Rombel, I., Reddy, G.N., Gang, J.B., and Shen, C.K. (1995). Functional roles of in vivo footprinted DNA motifs within an alpha-globin enhancer. Erythroid lineage and developmental stage specificities. J Biol Chem *270*, 8501-8505.

Zhang, Z., and Gerstein, M. (2004). Large-scale analysis of pseudogenes in the human genome. Curr Opin Genet Dev *14*, 328-335.

Zhao, Q., Zhou, W., Rank, G., Sutton, R., Wang, X., Cumming, H., Cerruti, L., Cunningham, J.M., and Jane, S.M. (2006). Repression of human gamma-globin gene expression by a short isoform of the NF-E4 protein is associated with loss of NF-E2 and RNA polymerase II recruitment to the promoter. Blood *107*, 2138-2145.

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. Nat Genet *42*, 742-744.

Zhou, X., Ha, S.Y., Chan, G.C., Luk, C.W., Chan, V., Hawkins, B., Lam, Y.H., Liang, R.H., and Lau, Y.L. (2001). Successful mismatched sibling cord blood transplant in Hb Bart's disease. Bone Marrow Transplant *28*, 105-107.