Characterisation of two genetic loci involved in fetal haemoglobin production BCLIIA and HBSIL-MYB intergenic region

Jawaid, Kiran

Awarding institution:
King's College London

Title: Characterisation of two genetic loci involved in fetal haemoglobin production
BCLIIA and HBSIL-MYB intergenic region

Author: Kiran Jawaid

Characterisation of two genetic loci involved in fetal haemoglobin production:

BCL11A & *HBS1L-MYB* intergenic region

Kiran Jawaid

April 2013

Department of Molecular Haematology,
Division of Cancer Studies,
School of Medicine, King's College London

Submitted for degree of Doctor of Philosophy

ABSTRACT

The continuous production of fetal haemoglobin (HbF, $\alpha_2\gamma_2$) into adulthood is an ameliorating factor in sickle cell disease and β-thalassemia. We have previously mapped two quantitative trait loci (QTLs) controlling HbF levels, one in intron 2 of *BCL11A* gene, and the other, an intergenic region on chromosome 6 between the genes *HBS1L* and *MYB*, known as *HMIP*.

Histone modification and RNA polymerase II binding at these loci and at the globin genes themselves were analysed using microarray-based chromatin immunoprecipitation studies of primary human erythroid progenitor cells. In addition, we analysed binding of GATA-1 and KLF1, two major erythroid-specific transcription factors.

Strong GATA-1 binding at the *HMIP* region coinciding with strong histone acetylation and RNA polymerase II activity was seen, indicative of the presence of regulatory elements in the intergenic region. Moreover, differential GATA-1 binding was observed between individuals with low HbF and raised HbF levels at a site within the *HMIP* region most strongly associated with HbF variance.

BCL11A intron 2 also showed strong GATA-1 binding and histone H3 acetylation, and may therefore be responsible for the overall regulation of *BCL11A*. I have carried out extensive sequence analysis of individuals from the upper and lower extremes of BCL11A-associated HbF levels to fully characterize DNA variants. The nature of this sequence as a regulator of BCL11A expression remains to be determined and functional tests are on-going.

The role of BCL11A, in conjunction with KLF1, as a transcriptional regulator of the α and β globin loci, as well as the QTLs themselves, was also investigated. These results show that BCL11A, often alongside GATA-1, is intimately involved in the transcriptional regulation of the globin genes and that KLF1 also regulates BCL11A. The importance of different protein isoforms of BCL11A is also explored.

ACKNOWLEDGEMENTS

The completion of this thesis was made possible by the help and support of many individuals.

I thank my first supervisor, Dr. Steve Best, undoubtedly one of the most knowledgeable and hard-working academics I have had the pleasure of working with.

I sincerely thank my second supervisor, Professor Swee Lay Thein. I am extremely grateful for her advice and support and for her trust in my ability as a researcher.

Everyone at the Department of Molecular Haematology has been like a second family to me over the years. Their friendliness, interest in my work and valuable feedback has been the best part of my experience at King's. Thank you to Dr. Nisha Vasavda, Dr. Karin Wahlberg, Dr. Stephan Menzel, Dr. Emma Drasar, Helen Rooks, Annabelle Kelly and Claire Steward.

A special thanks to Suleyman Aketuna, my fellow PhD student, for the always helpful conversation and company in the lab.

I would like to thank my parents for their encouragement, my mother has been and continues to be my rock throughout everything and is the inspiration of my ambitions. I wish them both the best of health and imaan.

My three beloved brothers have supported my studies from the very beginning, thank you.

My husband Khurram has been a great supporter of my work and has been especially supportive of the writing of this thesis. I love you and I thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS AND PRESENTATIONS ARISING FROM THIS THESIS

PUBLICATIONS

Karin Wahlberg, Jie Jiang, Helen Rooks, Kiran Jawaid, Fumihiko Matsuda, Masao Yamaguchi, Mark Lathrop, Swee Lay Thein, Steve Best (2009), 'The *HBS1L-MYB* intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells', *Blood*, 114 (6), 1254-62.

Kiran Jawaid, Karin Wahlberg, Swee Lay Thein, Steve Best (2010), 'Binding patterns of BCL11A in the globin and GATA1 loci and characterization of the *BCL11A* fetal hemoglobin locus', *Blood Cells Mol Dis.,* 45 (2), 140-6

PRESENTATIONS

The 17th Haemoglobin Switching Meeting, St.Johns College, Oxford, UK, 2nd September 2010 to 6th September 2010

Poster presentation- "Binding patterns of BCL11A in the globin and *GATA1* loci and characterization of the *BCL11A* fetal haemoglobin locus"

King's College London, Division of Cell & Gene Based Therapy, Postgraduate Showcase Day 2009, 12th June 2009

Oral presentation- "Transcription factor binding studies to characterise the mechanisms of globin production"

King's College London, Division of Cell & Gene Based Therapy, Postgraduate Showcase Day 2010, 28th May 2010

Oral presentation- "Genetic regulation of fetal haemoglobin production"

King's College London, The James Black Centre, Research Seminar series, 8th October 2011

Oral presentation- "Fetal haemoglobin: An update of the genes regulating its production"

King's College London, Division of Cancer Studies, PhD Research Conference 2011, 30th June 2011

Oral presentation- "Genetic regulation of fetal haemoglobin production"

C H A P T E R 1

1 INTRODUCTION

1.1 Haemoglobin

In 1840, Liepzig, Germany, a crystalline substance was discovered in earthworms which was later identified as haemoglobin. Haemoglobin is the protein that carries oxygen in red blood cells (erythrocytes) and delivers it to all body tissues where it is exchanged with carbon dioxide and later expelled via the lungs.

The molecular structure of haemoglobin (Hb) was determined by Max Perutz in 1959 using x-ray crystallography. Haemoglobin consists of two α-like globin chains and two β-like globin chains (Figure 1.2) with each chain containing an oxygen carrying heme group (Perutz *et al.* 1960).

In keeping with the changing oxygen affinity demands during human development, different forms of the haemoglobin protein are synthesised to fulfil these requirements. A total of six types of haemoglobin exist, and the first three, Gower I ($\zeta_2\varepsilon_2$), Gower II ($\alpha_2\varepsilon_2$), and Portland ($\zeta_2\gamma_2$) are only produced in the embryo (Huehns *et al.* 1964). The first switch occurs from embryonic to fetal haemoglobin ($\alpha_2\gamma_2$) at about 6-8 weeks of gestation. Around birth the second switch is from fetal to adult haemoglobin (HbA_2; $\alpha_2\delta_2$ and HbA; $\alpha_2\beta_2$). HbA_2 is a minor form of haemoglobin which makes up 2-3% of total haemoglobin (Kunkel & Wallenius 1955). HbA represents more than 95% of total haemoglobin and remains the dominant form of haemoglobin throughout life (Stamatoyannopoulos 1972) (Figure 1.1).

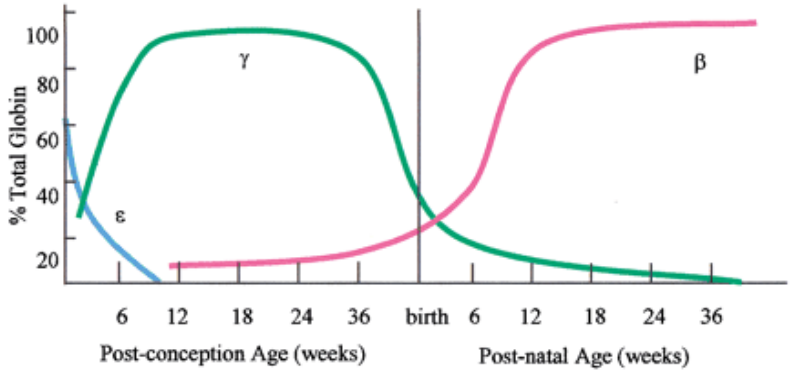


Figure 1.1 ß-like globin gene expression. The embryonic gene is expressed during the first six weeks of gestation. The first switch from ε- to γ-globin occurs shortly after conception, and the second switch from γ- to ß-globin occurs shortly after birth (Harju *et al.* 2002).

The genes responsible for the synthesis of Hb in humans are found in two gene clusters. The α-globin gene cluster at chromosome 16 contains the ζ-globin gene and the α-globin gene. The β-globin gene cluster is located on chromosome 11 and contains the ε, ^{G}γ, ^{A}γ,δ, and β genes (Stamatoyannopoulos 2005)(Fig. 1.5). In both clusters the globin genes are arranged in the order in which they are expressed during ontogeny (Collins & Weissman 1984).

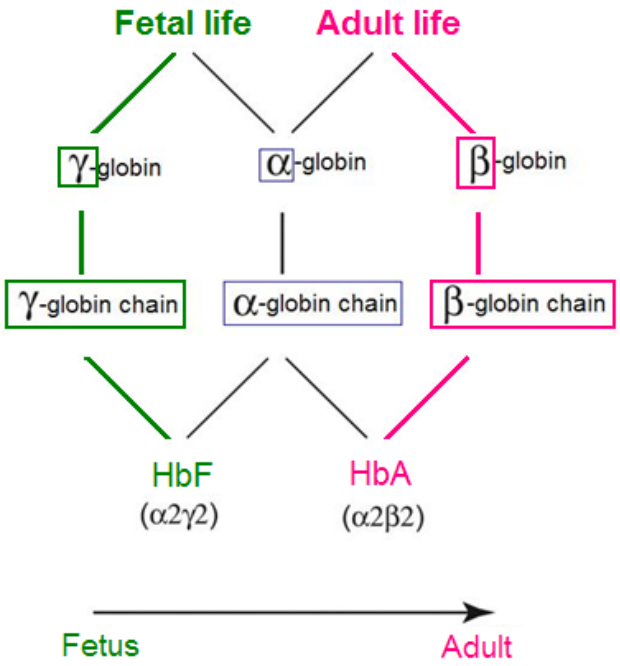


Figure 1.2 Schematic of the globin genes and gene products that constitute the fetal and adult haemoglobin molecules.

1.2 Haematopoiesis & Erythropoiesis

Haematopoiesis is defined as the formation of blood cells. This process takes place at multiple sites throughout development and can be divided into two stages: Primitive haematopoiesis occurs during the earliest weeks of gestation, in the embryonic yolk sac; definitive haematopoiesis occurs in several locations; the aorta-gonads-mesonephros (AGM) region is the initial site of definitive haematopoiesis followed by the liver and spleen. From around 2 months until 6-7 months of fetal life, the liver and spleen are the major haematopoietic organs, after which, the

bone marrow begins to take over the task and eventually becomes the exclusive site of haematopoiesis throughout childhood and adult life (Dzierzak 2005)(Fig. 1.3).

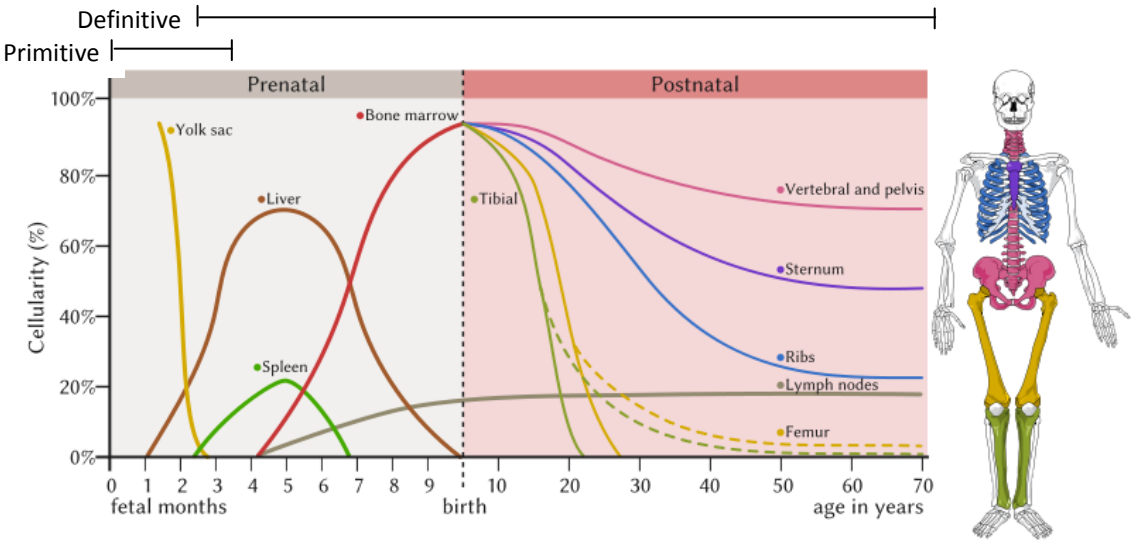


Figure 1.3 Sites of haematopoiesis during human development. (Image adapted from M.Komorniczak, 2011)

All blood cells are derived from a common pluripotent haematopoietic stem cell (HSC) and differentiation begins after they are committed to either the myeloid or lymphoid lineage, as a myeloid progenitor cell or lymphoid progenitor cell, respectively.

Erythrocytes (red blood cells) are derived from the myeloid progenitor cells. The earliest forms of myeloid progenitor cells are burst forming units-erythroid (BFU-e). BFU-e develop into colony forming units-erythroid (CFU-e). These progenitor cells develop into precursor cells and become erythropoietin responsive. These erythropoietin responsive cells (ERC) go through several stages, proliferating and maturing at each stage and are now morphologically recognisable (Philipsen & Wood, 2009).

CFU-e develop into pro-erythroblasts, basophilic erythroblasts, polychromatic erythroblasts and orthochromatic erythroblasts (Fig.1.4). In the next stage of development these cells lose the ability to divide, followed by extrusion of the nucleus and loss of organelles. The cells are now referred to as reticulocytes. Reticulocytes are released into the blood stream where they mature into erythrocytes (Philipsen & Wood, 2009).

Erythrocytes in circulation have a 120 day lifespan. Towards the end of this time they are cleared from the circulatory system as new reticulocytes are released and mature into erythrocytes. The rate by which red cell production is regulated is controlled by erythropoietin (EPO) levels. EPO stimulates erythropoiesis by increasing the rate of maturation of the ERCs into pro-erythroblasts (Philipsen & Wood, 2009).

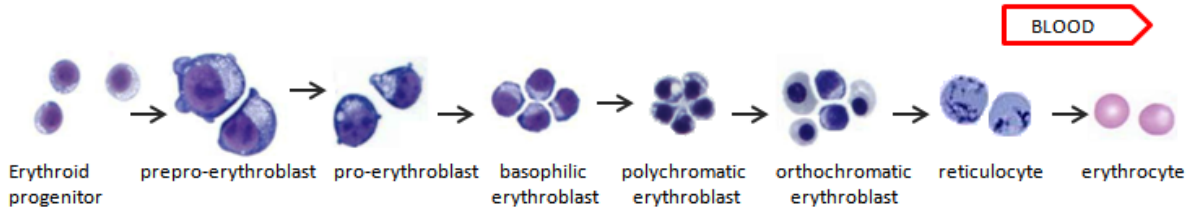


Figure 1.4 Human erythroid differentiation. Cells grown by the two-phase liquid culture method adapted from Fibach et al.

Mature erythrocytes contain haemoglobin which represents more than 95% of total cytoplasmic protein content of the cell.

The growth of erythroid progenitor cells is dependent on specific growth factors. Stem cell factor-SCF (also known as Steel factor or Kit ligand) is required for HSC to BFU-e stage, which express the SCF-R (Stem cell factor-Receptor) which binds SCF.

SCF is required to induce proliferation and differentiation of these cells (Broudy 1997). Interleukin-3 (IL-3) is a cytokine that promotes cell survival and proliferation by binding to IL-3R expressed on HSCs through to early progenitor cells. The granuolocyte-macrophage colony stimulating factor (GM-CSF) is required again from HSC to BFU-e stage and promotes proliferation (Migliaccio & Papayannopoulou 2001). Erythropoietin (EPO) is an essential growth factor in erythropoiesis which is expressed slightly later than the other growth factors but becomes paramount for cell development. EPO drives the proliferation and expansion of the developing erythroid progenitors from BFU-e stage through to terminal differentiation to haemoglobinized cells (Philipsen & Wood 2009). Erythroid progenitors can thus be identified by the cell-surface expression of receptors for these specific growth factors.

1.3 Control of the β-globin gene locus

Both cis (including promoters, enhancers, silencers and insulators) and trans-acting elements are involved in the expression of the β-globin like genes.

Globin gene promoters have sequences in common at specific distances from the transcriptional start site (TSS). The TATA sequence is found 30 bp upstream of the TSS, the CCAAT site is found 70-78 bp upstream of the TSS and the CACCC sequence is located 80-140 bp upstream of the TSS. The CACCC sequence is specific to erythroid specific gene promoters, whereas the other sequences are found in many eukaryotic promoters. These sequences are important in globin gene expression as

down regulation of the globin genes is seen when natural mutations occur at these sites (Forget, 2001).

1.3.1 The β-globin Locus Control Region (LCR)

The β-globin gene locus is controlled exclusively on a transcriptional level. Genes that have an 'open' chromatin structure (euchromatin) are transcriptionally active, whereas transcriptionally silent genes have a 'closed' chromatin structure (heterochromatin). Open chromatin is hypersensitive to nuclease digestion while closed chromatin is resistant to DNase I digestion. Thus, transcriptionally active sites can be identified by their hypersensitivity to DNase I.

6-20kb upstream of the β-globin locus is the *locus control region* or β-LCR, which contains five DNase I hypersensitive sites; HS1-5 (Fig.1.5). HS1-4 are hypersensitive to DNase I in erythroid cells, whereas HS5 is ubiquitous in many non-erythroid cells. The hypersensitive site cores are 200-300 bp which are highly conserved between species. The core sequences contain binding sites for ubiquitous and erythroid specific transcription factors (Hardison 2001).

It is proposed that this upstream region is responsible for activation of the β-globin gene locus by initiating and/or maintaining an open state of chromatin around the β-globin gene cluster and acting as an enhancer element of the globin genes (Forrester *et al.* 1986; Forrester *et al.* 1987).

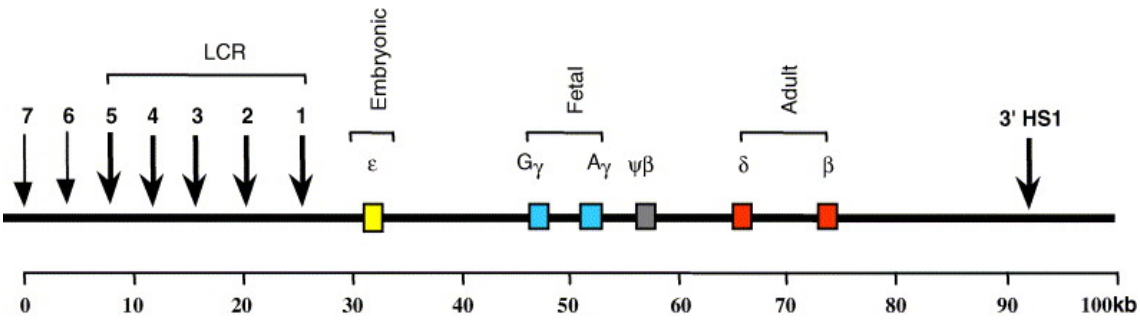


Figure 1.5 The human β-globin gene locus on chromosome 11. Arrows indicate DNase I hypersensitive sites. (G. Stamatoyannopoulos 2005).

The β-LCR is unique to other cis-acting regulatory elements as it can activate independent of position, as demonstrated in mice studies. Deletions of the β-LCR in transgenic mice results in a reduction of globin gene expression by 100-fold (Bender *et al.* 2000). However, when the β-LCR coupled with a β-globin gene is transfected into transgenic mice, this results in full expression of the gene, independent of the site of integration in the host genome (Grosveld *et al.* 1987). Moreover, deletion of individual HS sites prove that an intact LCR is necessary for full positional independent activation of the globin genes, therefore the HS sites act as a single unit (Milot *et al.* 1996). This highlights the importance of the LCR and globin gene co-operation in gene expression.

The HS sites themselves have individual functions, acting in cis to the β-globin gene locus; HS5 is thought to act as an insulator or 'boundary element' (Li & Stamatoyannopoulos 1994). HS5 protects the locus from the negative effects of neighbouring closed chromatin (Harju *et al.* 2002) and insulates a specific chromatin domain, thereby regulating a steady state between open and closed chromatin (Gribnau *et al.* 2000). HS2 functions as an enhancer element and HS3

acts as an activator of ε-globin and γ-globin gene expression (Navas *et al.* 1998), by exerting chromatin opening and remodelling activity (Ellis *et al.* 1996).

1.3.2 Globin gene activation by the β-LCR

Several models have been proposed to explain how the β-LCR functions to activate the downstream globin genes and of these, the looping model is the most widely accepted (Fig. 1.6). This model suggests that the HS sites of the β-LCR fold into a holocomplex, bringing the LCR closer to the appropriate gene and thereby the transcription factors bound to the LCR are delivered to the gene locus to interact with the basal transcriptional apparatus and activate globin gene expression (Ellis *et al.* 1996; Fraser *et al.* 1993; Stamatoyannopoulos 2005; Wijgerde *et al.* 1995) .

Studies have demonstrated the interaction of the LCR with downstream globin gene promoters via the looping model. Carter *et al.* (2002) used an RNA fluorescence *in situ* hybridization (FISH) method called RNA TRAP to tag and recover chromatin near the murine β-globin gene which recovered HS1, HS2 and HS3 together with the β-globin gene. This demonstrated that the HS sites and β-globin gene were in close proximity. A further study used chromatin conformation capture (3C) to measure the proximity of LCR sequences and active globin genes in erythroid cells. Clustering of active regulatory elements at these sequences was referred to as an active chromatin hub (ACH) (Tolhuis *et al.* 2002).

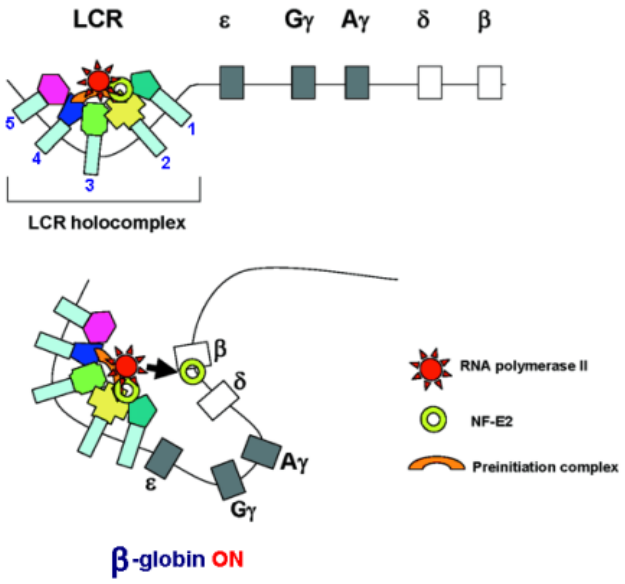


Figure 1.6 The looping model of gene activation at the β-globin locus. The LCR holocomplex is formed with transcription factors and other transcriptional apparatus. The holocomplex loops over the globin genes, bringing the transcriptional apparatus closer to the globin genes for transcription of the preferential gene (the β-globin gene in this example). Image adapted from Vieira *et al.* 2004.

1.4.3 Gene switching mechanisms

To explain how certain genes are switched off while others remain turned on, two mechanisms of gene switching are thought to be responsible for turning genes off; autonomous silencing and gene competition. Gene competition was first investigated in transgenic mice in which the β-LCR was linked to either the γ-globin (μLCR ^{A}γ) or the β-globin (μLCR β) gene, alone. In both cases genes failed to show developmental expression, i.e. the β-globin gene was expressed in cells throughout all developmental stages including the embryonic cells. In other words, β-globin gene expression throughout development was as it would be in adult cells. The γ-globin genes were also expressed in adult cells (Enver *et al.* 1990; Stamatoyannopoulos 2005).

When the LCR, γ and β genes were linked together in the same construct (μLCR ^{A}γψβδβ), developmental expression of the genes followed. The γ-globin gene was predominantly expressed in embryonic erythrocytes and the β-globin gene was highly expressed in adult mice erythrocytes. γ and β-globin genes were expressed similarly in fetal liver erythroblasts, similar to the human perinatal stage (Enver *et al.* 1990). The investigators concluded that the γ and β genes compete to interact with the LCR (Enver *et al.* 1990). During fetal life, proximity to the β-LCR leads to preferential expression of the γ-globin gene. During adult life the γ-globin gene is autonomously silenced and the LCR interacts with the next downstream gene i.e. β-globin.

Similar experiments in transgenic mice showed that mice carrying the embryonic, ε-globin gene do not express the gene at all, but when the ε-globin gene is linked to the LCR (μLCR ε) the gene is expressed at high levels, indicating the ε-gene expression is dependent on the LCR. At the same time, expression was nil in yolk sac erythropoiesis, suggesting all the components needed to turn off a gene are to be contained within the gene or adjacent sequence. This concept is called 'autonomous silencing' (Raich *et al.* 1990). Transcription factors bind on silencing sequences on proximal and distal ε-globin gene promoters to form a silencing complex which turns off ε-globin gene expression by disrupting interactions between the gene and LCR (Stamatoyannopoulos 2005). Autonomous silencing is thought to be the principal mechanism to turn off genes, though gene competition is also contributing.

1.5 Transcription factors at the β-globin locus

Transcription factors play an important part in gene regulation at the β-globin locus. The HS sites all share a core sequence ~250 nucleotides long, which is studded with motifs for transcription factors. The HS sites contain binding sites for erythroid specific transcription factors NF-E2, GATA-1 USF, Sp1 and KLF1 (Harju *et al.* 2002).

NF-E2 is an erythroid-specific DNA-binding protein that recognizes motifs (GCTGA(G/C)TCA) in the 5'HS2 of the β-globin LCR (Talbot & Grosveld 1991, Stamatoyannopoulos *et al*. 1995). Here, the NF-E2 binding sites are important for transcriptional activation and formation of HS sites in the LCR (Forsberg *et al.* 2000; Ney *et al.* 1990). Experimental evidence also shows NF-E2 binding sites are important for chromatin remodelling activity and necessary for ε-globin gene expression and formation of HS2 (Gong *et al.* 1996).

GATA-1 is also an erythroid specific transcription factor required for globin gene switching and erythroid cell maturation. It is a member of the GATA zinc finger family of transcription factors which bind to nucleic acid consensus sequence (T/A)GATA(A/G) (Evans & Felsenfeld 1989; Tsai *et al.* 1989). GATA-1 binds to sites in the globin gene promoters and HS site cores of HS1-5 which contain the GATA-1 recognition sequence (Orkin 1992). GATA-1 can also act as an activator when bound to the γ-globin gene promoter or HS1-5 (Jane & Cunningham 1996; J. A. Stamatoyannopoulos *et al.* 1995).

GATA-1 is also an activator of the ε-globin gene but acts as a repressor of ε-globin when bound to the ε-globin gene silencer with transcription factor YY1 (Kellum &

Schedl 1991; J. Li *et al.* 1998). GATA-1 is known to interact with itself and other transcription factors like SP1 and KLF1 (Merika & Orkin 1993).

Krüppel-like factor 1, KLF1 (previously known as erythroid Krüppel-like factor, EKLF) was identified by subtractive hybridization and found to be an erythroid cell-specific transcription factor, homologous to the Krüppel family of transcription factors which have roles in cell proliferation, differentiation and survival (Miller & BIeker, 1993). KLF1 contains three zinc fingers at the C-terminus which bind to a CACCC sequence. CACCC sequences are repeated in erythroid enhancers and promoters, including the β-globin gene promoter (Miller & Bieker, 1993). This sequence is noted as a site of point mutations that give rise to β-thalassaemia.

It is reported by Feng *et al.* (1994) that KLF1 is unable to transactivate in the presence of these point mutations due to a decrease in binding affinity for these target sites. The KLF1 protein binds to the β-globin promoter with 8 fold higher affinity than to the γ-globin promoter (Donze *et al.* 1995). These studies showed KLF1 as a stage-specific, β-globin-specific transcription factor and proposed that it is most likely an important factor in the fetal to adult (γ to β) haemoglobin switch.

Further studies in mice revealed that *KLF1*$^{-/-}$ mice die around day 14-15 of gestation due to anaemia caused by the failure to express β-globin. Moreover, the embryos showed features of β-globin deficiency as found in β-thalassaemia. *KLF1*$^{-/-}$ embryos appeared normal during embryonic yolk sac stage of haematopoiesis, but became fatally anaemic during early fetal life, at the precise time of the switch from embryonic to fetal-liver erythropoiesis (Nuez *et al.* 1995 & Perkins *et al.* 1995).

The activation of the different β-globin like genes is mediated in part through the β-globin LCR region (Chapter 1.3.2). KLF1 binding enhances the interaction between the β-globin promoter and the LCR and so the interaction between the γ-globin promoter and LCR decreases (Miller & Bieker 1993, Donze *et al.* 1995).

The human β-globin locus was studied in *KLF1* knockout/human β-locus transgenic mice (Wijgerde *et al.* 1996). The ε and γ globin genes were expressed normally in *KLF1*$^{-/-}$ foetuses, with a complete lack of β-globin expression. In *KLF1*$^{+/-}$ /β-globin transgenic mice, there was a shift in the γ to β ratio caused by an increase in the number of actively transcribed γ genes and a decrease in transcribed β genes. The authors proposed that the reduction in KLF1 in the *KLF1*$^{+/-}$ mice reduced the time the LCR was in complex with the β-globin gene. Because KLF1 does not directly bind γ-globin, an increase in γ-globin is due to the LCR being less occupied with β-globin, therefore less β-globin activation and more interaction is formed between the LCR and γ-globin (Wijgerde *et al.* 1996). This was accompanied by changes in chromatin structure at the β-globin promoter, and HS3 of the β-LCR.

Following this work, KLF1 was found to be required for the activity of 5'HS3 of the β-globin LCR by binding directly to the core fragment within HS3 (Tewari *et al.* 1998, Gilemans *et al.* 1998). Increasing levels of *KLF1* lead to changes in the balance from γ to β-globin gene expression, which results in an earlier switch of the globin genes and the amount of *KLF1* also influences the rate of the switching process (Tewari *et al.* 1998).

These studies, collectively demonstrate the importance of KLF1 in the γ to β globin switch and LCR-β-globin gene interactions.

Recently it has been proposed that KLF1 has a direct role in *BCL11A* activation. Borg *et al* (2010) mapped the cause of HbF persistence in a Maltese family to chromosome 19 and found a heterozygous truncated *KLF1* mutation in family members with high levels of HbF. When they looked at *BCL11A* expression, it was downregulated in these family members; moreover, γ-globin expression was upregulated. The same group demonstrated that *KLF1* knockdown in human erythroid progenitor cells resulted in an increase in γ-globin and a decrease in *BCL11A*. The opposite effect was observed with full length *KLF1* expression in individuals with high HbF, i.e. downregulation of γ-globin and upregulation of *BCL11A.* Furthermore, KLF1 binding was shown at the *BCL11A* promoter, demonstrating the direct role of KLF1 on *BCL11A.* These findings conclude that HbF persistence here is caused by *KLF1* haploinsufficiency. These findings were confirmed by others (Zhou *et al*. 2010), and now propose a new model for globin regulation (Fig. 1.7).

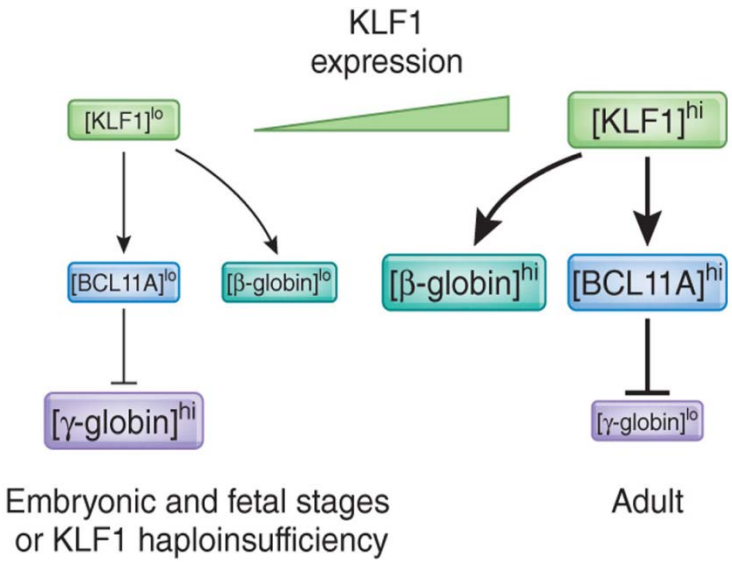


Figure 1.7 Latest model of globin gene regulation by KLF-1 via BCL11A (adapted from Bieker *et al.* 2010).

1.6 The Haemoglobinopathies

The two commonest inherited disorders of haemoglobin that impact global health are β-thalassaemia and sickle cell disease (SCD). Both disorders arise from mutations in the β-globin gene locus (Steinberg 2001).

An estimated 300,000 new-borns with these disorders are born worldwide, annually. These are prevalent in tropical and sub-tropical regions where carriers have a survival advantage to *falciparum* malaria, but these disorders have spread to most countries by population migration. β-thalassaemia is most common in Asia, the Mediterranean and Middle East, and SCD predominates in Africa (World Health Organization, fact sheet No.308, Jan 2011)

SCD is caused by a point mutation in the 6^{th} codon of the β-globin gene in which 'T' is substituted for 'A', changing a glutamic acid residue into a valine residue (Pauling. *et al*, 1949), and resulting in an altered β-globin chain, referred to as sickle globin (β^S) (Marotta 1977). Two β^S chains combine with two α-globin chains and make up a variant haemoglobin molecule called sickle haemoglobin (HbS- $\alpha_2\beta_2{}^S$). This HbS is less soluble in its deoxygenated state and polymerises, causing a distorted, sickle like shape of the red blood cells. These sickle cells become trapped in the circulation causing clinical manifestations such as ischaemic organ damage and episodic pain crises (Fig. 1.8). Sickled cells are also highly fragile with a shortened lifespan, leading to haemolytic anaemia (Stuart & Nagel, 2004).

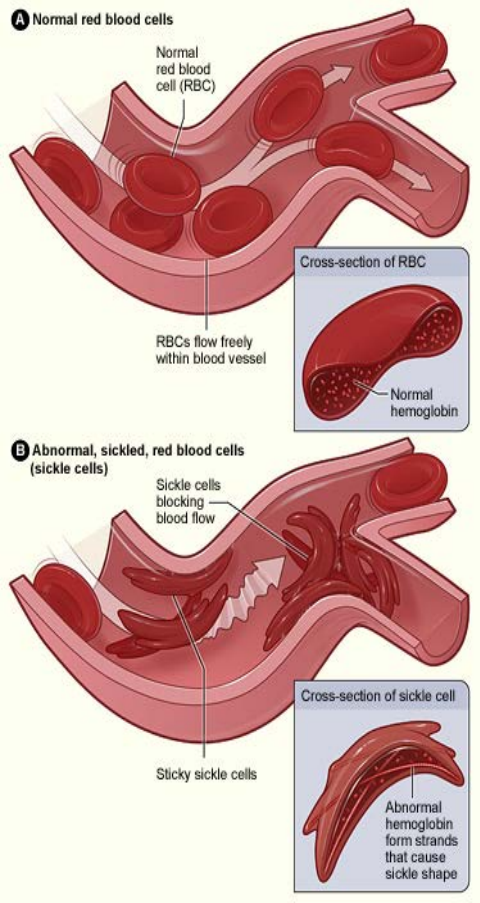


Figure 1.8 Appearance and passage of red blood cells through vessels in normal (A) and sickle cell disease (B) scenario. (Adapted from National Heart Lung and Blood Institute, 2011).

β-thalassaemia is characterized by a quantitative deficiency of the β-globin chain, caused by a wide array of mutations, mostly all of which affect the β-globin locus. The majority of mutations are point mutations within the gene and its immediate flanking sequences (Forget 2009*).*

Lack of its partner β-globin chains, leads to an accumulation of free α-globin chains which aggregate in red blood cell precursors, forming inclusion bodies and causing apoptosis, ineffective erythropoiesis, peripheral haemolysis and reduction in Hb synthesis (Thein 2002). Some red blood cells that make their way to circulation show microcytosis and other deformities and are eventually hemolysed.

Erythropoiesis is thus expanded to compensate for the lack of functional red blood cells and causes hyperplastic marrow; an increased proliferation of red blood cells in the bone marrow (Bank 2005, Stamatoyannopoulos 1992).

1.7 Fetal haemoglobin and the hereditary persistence of fetal haemoglobin

Fetal Hb (HbF; $\alpha_2\gamma_2$) has a much higher affinity for oxygen than the adult form of Hb (HbA; $\alpha_2\beta_2$), and is the dominant form of Hb expressed during fetal life. The switch from fetal to adult Hb is not complete; background levels of fetal Hb continue to be produced in adult life with the majority of the healthy adult population having HbF levels of ≤0.6% of total Hb (Thein & Menzel 2009).

Hereditary persistence of fetal haemoglobin (HPFH) is the term used to describe a condition in which high γ-globin expression persists into adult life, resulting in elevated HbF. Broadly speaking, there are two groups of HPFH; pancellular, referring to the even distribution of HbF in erythrocytes, with a clearly recognisable increase of 10-35% HbF in heterozygotes. These are caused by mutations (deletions or point mutations in the γ-globin promoter) affecting the β-globin gene cluster (Forget 1998).

The second group is referred to as heterocellular hereditary persistence of fetal haemoglobin (hHPFH) reflecting the uneven distribution of HbF among the red blood cells. Individuals with heterocellular HPFH represent the upper tail of the natural continuous distribution of HbF with HbF levels >0.8%. Heterocellular HPFH

is inherited as a complex trait. Recently three major quantitative trait loci (QTLs) have been identified including a QTL on the β globin cluster, represented by the Xmn1-^{G}γ site (Craig *et al.* 1996; Thein & Weatherall 1989; Thein *et al.* 2007; Menzel *et al.* 2007). Erythrocytes in which HbF is high enough to be detected are designated 'F' cells (Boyer *et.al* 1975).

1.7.1 Stress erythropoiesis

At times when the demand for oxygen is greater, such as anaemia and tissue hypoxia (during fetal development), erythroid progenitor cells, induced by an increase in EPO, have the ability to rapidly proliferate, releasing large numbers of erythrocytes into circulation. This process is called 'stress erythropoeisis'. A γ-globin to β-globin switch also occurs within early progenitor cells (Papayannopoulou *et al*. 1986). While we see a small number of F-cells in normal erythropoiesis, during stress erythropoiesis more cells terminate maturation prematurely, while still expressing γ-globin, and are released into circulation while still expressing predominantly HbF as F-cells, thus increasing HbF levels (Stamatoyannopoulos 2005).

Raised HbF is clinically beneficial for people affected with β-globin deficiencies such as β-thalassemia or SCD. In β-thalassemia, the γ-globin chains combine with redundant excess α-globin chains thus reducing the α/β chain imbalance that underlies the severity of anaemia in β-thalassaemia (Thein 2002). In the case of SCD, the rate of HbS polymerisation is dependent on the concentration of HbS in the red blood cells, so an increase in non-sickle Hb, for example, an increase in fetal

Hb, shifts the reaction dynamics away from polymerisation and ameliorates the disease symptoms (Noguchi 1984). At the same time γ-globin also has an inhibiting effect on HbS polymerisation.

For these reasons, higher levels of HbF have ameliorating effects in SCD and β-thalassaemia, and so, reactivation of HbF synthesis is an important approach for the treatment of these haemoglobinopathies.

1.7.2 Reactivation of HbF

Hydroxyurea is an agent used to induce HbF levels in patients by increasing F-cells in circulation which results in increased HbF. Hydroxyurea has been used to treat SCD patients since the early 1990s, but seems to have variable effects in increasing HbF in β-thalassaemia patients. Currently many drugs are under investigation for use as HbF inducers, but none are close to being used as therapeutics for haemoglobinopathies. This makes understanding the γ-globin to β-globin switch and control of HbF production ever more important. Understanding the mechanisms behind this can lead to targeted therapies for the treatment and even cure of β-haemoglobinopathies.

1.8 Quantitative trait loci influencing HbF levels

A quantitative trait locus (QTL) is a region of DNA that is associated with a particular phenotypic trait, which in this case is HbF. QTL can often be found at regions within a gene, thus identifying candidate genes contributing to the phenotype.

Three loci associated with HbF levels have recently been mapped by genetic studies; *Xmn1*-G*γ* polymorphism at the β-globin gene cluster, *HBS1L-MYB* intergenic polymorphism (*HMIP*) on chromosome 6q and *BCL11A* on chromosome 2p (Craig *et al.* 1996; Thein & Weatherall 1989; Thein *et al.* 2007; Menzel *et al.* 2007).

Multiple factors contribute to the variance in HbF levels in the population, including age, sex (2%) and environmental factors (9%), but an overwhelming 89% is attributed to genetic factors (Fig 1.9). The *Xmn1-Gγ* polymorphism at the β-globin locus accounts for up to 10% of the F cell variance in the general population (Craig *et al.* 1997; Thein 2002). However, more than 35% of the genetic variance in F cell levels is controlled by the other two loci; *HMIP* (19%) and *BCL11A*(15%) *,* which are unrelated to the β-globin locus (Thein 2002, Menzel *et al.* 2007).

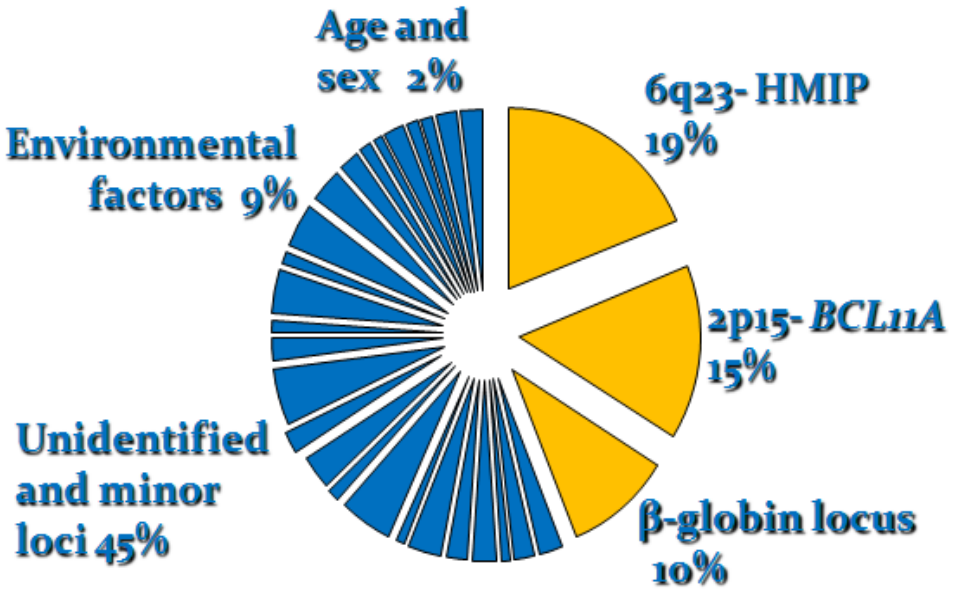


Figure 1.9 Factors contributing to the overall variance in HbF levels in the northern European population, highlighting the three main QTL identified to date (Figures based on Garner *et al.* 2000 & Menzel *et al.* 2007).

1.8.1 Chromosome 11 *Xmn1-Gγ* polymorphism

The *Xmn1*-Gγ polymorphism on chromosome 11 is caused by a C-T polymorphism at position -158 upstream of the Gγ-globin gene. *Xmn1*-Gγ has long been implicated as a QTL influencing HbF production through clinical and family studies. This polymorphism is present in about 30% of the population, and results in a slight increase in HbF levels in this group. However, the presence of this polymorphism in SCD and β-thalassaemia patients results in even higher HbF levels. (Thein *et al* 1987).

1.8.2 6q23 *HBS1L-MYB* intergenic region

The QTL for fetal haemoglobin on chromosome 6q23 was identified in a large consanguineous Asian-Indian family with β-thalassaemia and co-inherited hHPFH, by genome wide linkage analysis (Craig *et al.* 1996; Thein & Weatherall 1989). The family was first described in 1989 (Thein & Weatherall, 1989) through the propositus who was homozygous for β-thalassaemia and yet had an extremely mild clinical disease and HPFH (Thein *et al.* 1994). Several family members with or without β-thalassaemia had raised levels of HbF, up to 3.4% in normal members and up to as much as 24% in β-thalassaemia heterozygotes (Thein *et al.* 1994)

Fine mapping with 210 family members brought the associated area to a 1.5Mb region (Garner *et al.* 1998), in which lie 5 protein coding genes; *ALDH8L, HBS1L, MYB, AHI1 and PDE7B* (Close *et al.* 2004) (Fig.1.10). Three of these genes; *HBS1L, MYB* and *AHI1* are expressed in erythroid cells. None of these genes harboured any causative mutations and so it was proposed that the causative genetic variants may

be in the regulatory and non-coding regions and affect transcription of the genes rather than altering the gene function itself.

It may be that differences between the mRNAs of these genes in individuals with elevated HbF and individuals with normal levels of HbF may relate to sequence variants in the regulatory regions of these genes. This led to the investigation of the expression patterns of the genes in the 6q23 locus in the erythroid precursor cells of individuals with and without elevated HbF levels, by quantitative reverse transcription PCR (Jiang *et al.* 2006).

This study revealed both *MYB* and *HBS1L* were down-regulated in individuals with elevated HbF, making them candidate genes involved in the regulation of fetal haemoglobin.

K562 cells transfected with human cDNA of *MYB* and *HBS1L* showed that over-expression of *MYB* inhibits γ-globin gene expression, but *HBS1L* over-expression had no effect. This suggests that *MYB* may be the influencing factor at this QTL by acting as a negative regulator of the γ-globin gene (Jiang et al 2006).

1.8.2.1 MYB

The *MYB* gene (also known as *c-Myb*) was first recognized as the gene transformed by the avian myeloblastosis virus (AMV) which caused erythroblastic and myeloid leukaemia in quails and chickens (Oh & Reddy, 1999). The *MYB* gene is highly conserved through evolution (Lipsick, 1996); the major product of the human gene is a 75kDa nuclear protein of 636 amino acids. The protein is expressed in most

haematopoietic tissues and is highly expressed in immature haematopoietic cells (Gonda *et al.* 1982, Westin *et al.* 1982). Efforts were made to identify MYB target genes, by measuring differentially expressed genes between $MYB^{-/-}$ null mutants and wild-type cells. The GATA-1 transcription factor was found to be down-regulated in the $MYB^{-/-}$ cells, giving clues to the involvement of MYB in regulation of haematopoiesis (Lin *et al.* 1996).

Further evidence for this is seen when *MYB* knockout mice died *in utero* after day 15 of gestation due to severe anaemia (Mucenski *et al.* 1991). MYB is critical transcription factor in haematopoiesis, being involved in both cell proliferation and differentiation. It is required for maintenance of proliferation by exerting a direct role in cell cycle control (Oh *et al.* 1999). MYB is highly expressed in immature haematopoietic cells and is down-regulated as the cells differentiate (Westin *et al.* 1982), expression of MYB blocks differentiation in neuroblastoma cell lines, suggesting down-regulation of MYB is also required for terminal differentiation (Oh *et al.* 1999).

So far there has been no evidence for a direct role of MYB on the regulation of the γ-globin gene. It has been proposed that MYB influences HbF levels via altering erythroid kinetics. It is evident that MYB is involved in cell proliferation and differentiation; low *MYB* levels correlate with accelerated erythroid maturation (Jiang *et al.* 2006). *HMIP* block 2 SNPs (see chapter 1.8.3) that are highly associated with raised F cells are associated with reduced erythrocyte count, increased erythrocyte volume and increased platelet count (Menzel *et al.* 2007), suggesting that the HbF QTL has a pleiotropic effect on haematopoiesis.

Another mechanism whereby MYB can affect γ-globin production is via KLF1 activation of *BCL11A* (Bianchi *et al*. 2010) as discussed on page 14.

Raised HbF is a feature in some cases of Trisomy 13 (Huehns *et al.* 1964). Sankaran *et al.* (2011) mapped this elevation of HbF in Trisomy 13 cases to chromosomal band 13q14 which encodes two microRNAs: miR-15a and miR-16-1, which are expressed in erythroid precursor cells. Increasing the expression of these miRNAs in haematopoietic progenitor cells by 1.5 fold (as would be expected in a trisomy), resulted in an increase in γ-globin levels by 2.4 fold, which suggests that the elevated HbF levels seen in Trisomy 13 cases are likely due to increased expression of these miRNAs. Interestingly, when the authors investigated the targets of these miRNAs, the *MYB* gene was found to have two conserved 8mer miR-15a/16-1 targeting sites. *MYB* protein levels were reduced with even a slight increase in the expression of these miRNAs. This study further supports the role of MYB in the regulation of HbF, and its importance in erythropoiesis.

1.8.2.2 HBS1L

The *HBS1L* gene is mostly uncharacterised in humans. It is the human ortholog of the *Sacharomyces cerevisiae HBS1* gene, encoding a GTP-binding protein which shares structural features with elongation factor eEF-1A and release factor 3 (eRF3) families (Wallrapp *et al,* 1998). Its role, if any, and to what extent in the regulation of HbF is yet to be determined.

1.8.3 HBS1L-MYB intergenic polymorphisms (*HMIP*)

The 6q23 QTL story continued with further analysis of the region in Northern European twin pairs, by a high resolution association study (Thein *et al.* 2007). The study revealed multiple single nucleotide polymorphisms (SNPs) associated with HbF levels in the region between *HBS1L* and *MYB.* The SNPs reside in 3 linkage disequilibrium (LD) blocks (*HMIP* block 1-3) 79 kb within the intergenic region. *HMIP* blocks 1-3 account for all F cell variance at the 6q23 locus. *HMIP* blocks start 188bp upstream of the *HBS1L* gene and end 45 kb upstream of the *MYB* gene, both genes are transcribed in opposite directions (Figure 1.10). A 24 kb segment, starting 33 kb upstream of *HBS1L* represents *HMIP* block 2. Twelve markers in *HMIP* block 2 showed strongest associations with HbF (P values of 10^{-50}-10^{-75}) and all the SNPs in this block are in complete linkage disequilibrium.

These SNP associations with HbF levels were later confirmed in sickle cell disease patients (Lettre *et al.* 2008)

In further studies the SNPs in *HMIP* 2 were found to have pleiotropic effects on other haematological parameters in healthy Northern Europeans, including erythrocyte count, erythrocyte volume, platelet count and monocyte count (Menzel *et al.* 2007).

It is our thinking that *HMIP* block 2 in the *HBS1L-MYB* intergenic region contains regulatory elements which influence HbF levels, possibly by regulating the control of the flanking genes; *MYB* and/or *HBS1L*.

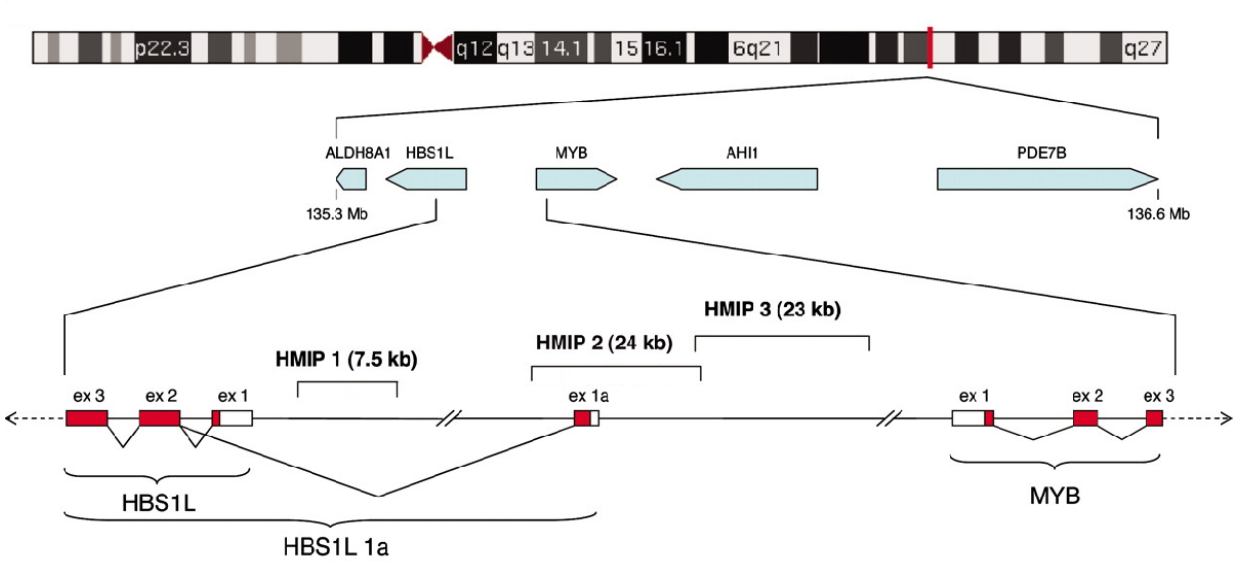


Figure 1.10 View of the QTL in chromosome 6q23 and *HMIP* blocks 1-3. Image adapted from (Thein et al. 2007).

1.8.4 2p15 *BCL11A* locus

To identify additional QTLs, our group conducted a genome wide association study (GWAS) in a Northern European cohort of 179 individuals with contrasting extreme F cell values (Menzel *et al.* 2007)

This study picked up the already known HbF associated loci at the *Xmn1-HBG2* site and *HMIP* on chromosome 6q23, and in addition identified a third, previously unknown QTL on chromosome 2p15, close to a gene encoding the zinc finger protein *BCL11A.* A further panel of 142 SNPs were genotyped which uncovered two clusters of association (Fig. 1.11a). The strongest association lay at a 14 kb region in intron 2 of *BCL11A*, with the second cluster spanning 67 kb in the 3' region of *BCL11A*, 8-7 kb downstream of exon 5 (Menzel *et al.* 2007).

The 2p15 QTL was estimated to contribute 15% of overall trait variance for HbF levels (Menzel *et al.* 2007). The peak of association within *BCL11A* was later confirmed by other groups (Lettre *et al.* 2008, Sedgewick *et al.* 2008, Uda *et al.* 2008). Uda *et al.* screened 4,000 Sardinians with varying HbF levels, finding association at the three major loci, with the strongest association for HbF levels with SNP rs11886868 (Figure 1.11a) in intron 2 of the *BCL11A* gene ($p<10^{-34}$).

To investigate if the influence of *BCL11A* on HbF modulates clinical phenotypes in β-thalassaemia, the same group genotyped SNP rs11886868 in 52 patients with β-thalassaemia intermedia and 74 with β-thalassaemia major. The 'C' allele of rs11886868, which is strongly associated with HbF was significantly more frequent in patients with β-thalassaemia intermedia, leading to the conclusion that the variant rs11886868 'C' allele influences a milder phenotype.

The authors further studied the effect of *BCL11A* in a cohort of 1,242 sickle cell disease (SCD) patients and found the rs11886868 'C' allele was also strongly associated with increased HbF in these patients ($p<10^{-20}$). Thus, BCL11A has important effects in haemoglobinopathies as well as non-anaemic individuals.

Sedgewick and colleagues (2008) through genetic association studies investigated the association of the known *BCL11A* SNPs in 3 different racial populations; 250 Chinese patients with β-thalassaemia major, 113 Thai patients with β-thalassaemia or HbE heterozygotes and 255 African Americans with SCD. Strong association of SNP rs766432 in intron 2 of *BCL11A* was found in all 3 groups.

In the Chinese β-thalassaemia group, individuals with the 'CC' genotype of rs766432 had twice as much HbF and F cells as those with an 'AA' genotype. Of the African American group with SCD, homozygotes for the 'C' allele of the same SNP had an average 7% HbF and homozygotes for the 'A' allele had an average 3% HbF. This suggests that the 'C' allele of rs766432 affects HbF levels in these populations.

SNP rs11886868 found by Uda *et al.* in the Sardinian and SCD patients is also in the same LD block in Africans and Europeans. This shows that the most important SNPs involved in HbF regulation at this locus are located within intron 2 which may be a regulatory motif.

In a third genome wide association study the *BCL11A* SNP rs11886868 was shown to have strong association with HbF levels in the 1,275 African Americans with SCD (Lettre *et al.* 2008). In addition SNPs rs4671393 and rs7557939 (in LD with rs11886868) were genotyped and showed stronger association with HbF levels than rs11886868 in both African Americans and Brazilians with SCD.

These studies collectively confirm that the genetic variants at the 2p15 QTL influence HbF levels in SCD and lie within intron 2 of the *BCL11A* gene.

1.8.4.1 *BCL11A*

In 2000, Nakamura and others made efforts to identify possible leukaemia disease genes in BXH2 mice, which highly develop leukaemia on retro-viral integration.

This was done using a method called retro-viral tagging. Retro-viruses integrate into the mouse haematopoietic cell DNA, disrupting the function of near-by genes,

resulting in the development of leukaemia. Therefore, by tagging the integration sites, isolating the DNA and sequencing, the genes targeted by the retro-virus could be identified. Identifying such retroviral integration sites by retroviral tagging meant that possible disease genes could be found.

One such common site of retroviral integration in BXH2 leukaemic mice cells is Ectopic viral integration site 9 (or Evi9). By chromosomal and physical mapping this site was located to chromosome 11 and because this was near the *c-rel* proto-oncogene this suggested that Evi9 may represent a new leukemia disease gene.

The same group demonstrated that a gene at Evi9 encodes a zinc-finger protein which binds to another zinc-finger protein called BCL6- which is a known human B-cell proto-oncogene product, so the new gene at Evi9 was coined *BCL11A*. Apart from the zinc finger domain it appeared BCL11A was distinct from anything in the current databases and two alternatively spliced *BCL11A* isoforms were identified by RT-PCR.

Saiki *et al.* (2000) investigated the possible role of *BCL11A* in the human haematopoetic system. This was done by screening a human fetal brain cDNA library using a mouse *BCL11A* cDNA fragment as a probe and a 3,063 nucleotide cDNA clone was produced. The sequence was analysed and revealed that human *BCL11A* encodes a 797 amino acid protein. Compared to the mouse cDNA sequence the human cDNA encodes 35 additional amino acids at the N-terminus but other than this, both mouse and human proteins share 99% homology.

Chromosomal localization by flourescence *in situ* hybridization (FISH) mapped the gene to chromosome 2p13 (Saiki *et al.* 2000). This site had previously been described in a chromosomal translocation with chromosome 14q32, in two children with chronic lymphocytic leukemia (Fell *et al.* 1986). The region of chromosome 2 was not characterised at the time.

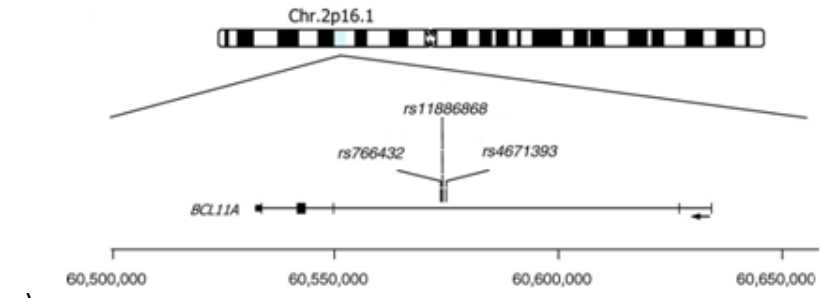
BCL11A is a highly conserved zinc finger gene and is expressed in haematopoietic progenitors, is down-regulated during myeloid differentiation (Saiki *et al.* 2000) and is essential for normal B and T cell development (Liu *et al.* 2003). BCL11A binds to GC-rich motifs and mediates transcriptional repression (Senawong *et al.* 2005, Sankaran *et al.* 2008). It is also involved in lymphoid malignancies through translocation or amplification events (Fell *et al.* 1986, Richardson *et al.* 1992, Satterwhite *et al.* 2001). *BCL11A* mutations have also been associated with B-cell malignancies (Liu *et al.* 2006).

It has been shown that BCL11A acts as a repressor of HbF production; using *BCL11A* knockdown, the levels of γ-globin and HbF were dramatically elevated in erythroid cells (Sankaran *et al.* 2008). However, this did not affect overall erythroid differentiation. BCL11A has also been shown to bind to discrete sites in the β-globin gene cluster, indicating a direct role of *BCL11A* in globin gene regulation (Sankaran *et al.* 2008).

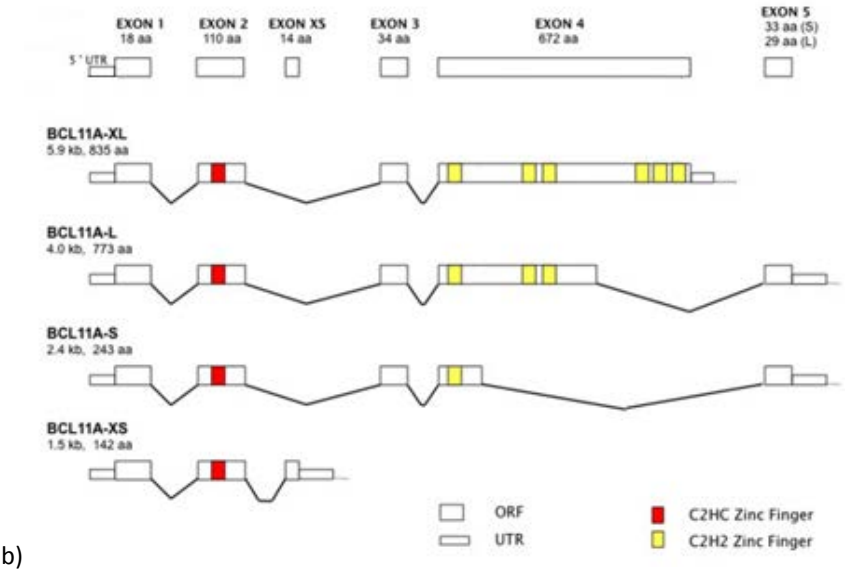
More recent studies using a human β-globin locus YAC transgene and a *BCL11A*$^{-/-}$ mouse, showed that in the absence of BCL11A, developmental silencing of the human γ-globin genes is impaired in the definitive erythroid lineage. Intermediate silencing in a *BCL11A*$^{+/-}$ mouse suggests that not only is BCL11A a developmental-

stage specific repressor but the effect is quantitative (Sankaran *et al.* 2009). Also, differences in *BCL11A* expression patterns between mouse and humans may be responsible for their divergent expression of β-like globin genes (Sankaran *et al.* 2009).

At least four *BCL11A* transcripts exist, giving rise to different protein isoforms; extra-long (XL; 5.9kb, 125kD), long (L; 3.8kb, 100kD), short (S; 2.4kb, 35kD) and extra short (XS; 1.5kb, 25kD). Exons 1& 2 are common to all four isoforms, exon 3 & 4 to isoforms XL, L and S. Isoforms L and S share exon 5 and isoform XS contains exon XS (Liu *et al.* 2006) (Fig. 1.11b).



a)



b)

Figure 1.11 a) QTL in chromosome 2- *BCL11A*. Highest scoring SNPs indicated at intron 2. **b)** Four major BCL11A isoforms, XL, L, S and XS (Images adapted from a) Thein *et al.* 2009, b) Liu *et al.* 2006)

The role of these different isoforms is largely unclear. Studies have shown that the shorter isoforms appear in primitive erythroid cells, whereas the larger forms appear in later developmental stages of erythroid maturation, when γ-globin expression is reduced (Sankaran *et al.* 2008).

BCL11A represents a possible therapeutic target for the treatment of β-haemoglobinopathies and studies are already showing its potential in reactivation of HbF and correcting the phenotypes of SCD in mouse models. Xu *et al.* (2011) employed SCD mice containing targeted deletions of murine α and β globins and the addition of human fetal and adult globin transgenes which were also null for *BCL11A* (SCD/ *BCL11A* $^{-}/^{-}$ mice). In these SCD/*BCL11A* $^{-}/^{-}$ mice, haematological parameters including RBC counts and Hb content were corrected and sickle cells were absent in comparison to SCD mice. SCD/ *BCL11A* $^{-}/^{-}$ mice showed increased expression of γ-globin, strong pancellular staining for HbF and F-cells accounted for 85.1% of total RBCs. These findings demonstrate that SCD phenotypes can be corrected by HbF reactivation via the inactivation of BCL11A.

1.9 HbF in MDS & Myeloid leukaemia

Raised HbF levels have also been observed in some acquired physiological conditions like pregnancy and acute blood loss and are associated with acquired malignancies of haematopoietic cells including myelodysplastic syndrome (MDS); a disorder of the myeloid cells which causes ineffective haematopoiesis. (Bourantes *et al.* 1991, Weatherall 1975).

A case study by Craig *et al*. (1996) examined F-cell production in a large group of MDS patients and found 62% of patients had raised levels of F-cells. 24.7% of MDS patients had higher HbF levels compared to 11% of normal individuals with raised HbF. Moreover, 61% of MDS patients with abnormal karyotypes have raised HbF (>5%). The group also reported that all four cases with monosomy 7 or 7q deletions had significantly higher F-cell levels.

Juvenile myelomonocytic leukaemia (JMML) is a rare mixed form of myelodysplastic/myeloproliferative disorder affecting young children (WHO classification, 2001). A key feature of JMML is elevated HbF. Chromosome 7 abnormalities and monosomy 7 is also common in JMML (25-30% of cases) (Emanuel 2008) and high levels of HbF in JMML is an indicator for poor prognosis, though why this is so, is unclear.

A recent study by Yin *et al*. (2009) measured gene expression in 285 patients with acute myeloid leukaemia (AML). A subset of AML patients with monosomy 7, associated with poor disease outcome showed high expression levels of *BCL11A.* Patients with JMML also showed high levels of *BCL11A* expression. Elevated HbF is

almost invariable in JMML, and yet *BCL11A* expression is seen to be elevated which is contrary to its role, as we know it, of acting as a repressor of HbF production.

Considering the current knowledge of myeloproliferative disorders and myeloid leukaemia, there seems to be a strong association with HbF levels and monosomy 7 in myeloid disorders, and more recently a possible role for BCL11A in these disorders.

BCL11A occurs as several protein isoforms (as discussed earlier). The shorter isoforms (L and S) appear to be restricted to primitive erythroblasts, and the full length (XL) isoforms to the adult-stage erythroblasts. All forms have been shown to acts as transcriptional repressors, but their functional overlap, interaction with each other and their individual roles remain to be discovered.

It is possible that different isoforms of *BCL11A* are expressed in different tissues and may have different roles in different cell types. This could possibly explain the high levels of *BCL11A* expression seen in JMML and myeloid disorders with monosomy 7, despite the very high levels of HbF, and may also shed more light on the mechanisms involved in HbF regulation.

1.10 Aims & Objectives

The overall aim of this research was to characterise the two trans-acting candidate regions at 6q23 and 2p15 involved in the control of HbF production.

1.10.1 6q 23, *HMIP* QTL

Transcription factors are critical regulators of gene expression. I aimed to investigate if there is differential transcription factor binding at the *HMIP* locus through chromatin immunoprecipitation (ChIP-chip) studies in primary human erythroid progenitors from individuals with different HbF phenotypes and genotypes.

1.10.2 2p15, *BCL11A* QTL

The majority of my research concentrates on this QTL. Evidence suggests BCL11A acts as a transcription factor (Sankaran *et al.* 2008), and so I explored the role of BCL11A in this capacity at different loci involved in haemoglobin production, including the α and β-globin loci. I also investigated other important erythroid transcription factors including KLF1 and GATA-1 at these loci.

The QTL at the *BCL11A* gene is mapped to a 14kb region within intron 2. I aimed to re-sequence this portion of the gene in individuals with high HbF and low HbF, to both confirm known, and identify any novel sequence variations between the extreme phenotypes.

Different protein isoforms of BCL11A exist. In order to better understand the function of these isoforms, I investigated their presence in different cell types.

CHAPTER 2

2 METHODS & MATERIALS

2.1 Cell culture

K562 cells (an erythroleukemia cell line originally isolated from a chronic myeloid leukemia patient in blast crisis) were maintained at 0.5-1 million/ml in RPMI medium (Sigma-Adrich, UK) supplemented with 10% FCS (PAA-laboratories, UK), 2mM L-glutamine (Sigma-Aldrich, UK), 0.1mg/ml streptomycin and 18units/ml penicillin (Sigma-Aldrich, UK) and incubated at 37°C in 5% CO_2.

2.1.1 Two phase erythroid culture

This liquid culture technique was adapted from the original method described by Eitan Fibach (Fibach et al. 1989), with modifications as detailed in Jiang et al. 2006.

50mls of peripheral blood was collected and centrifuged to separate the blood components. The yellow plasma interface is the component containing lymphocytes and progenitor cells and is called the buffy coat. This was collected and layered onto a density gradient centrifugation liquid- Histopaque-Ficoll (Sigma-Aldrich, UK) and centrifuged to separate red and white blood cells. The lymphocytes and other mononuclear cells remain at the plasma-histopaque interface from where they are collected and washed in PBS. Cells are counted and cultured in flasks at 1-2 million/ml in Phase I media (Stem Span Medium [Stem Cell Technologies, Canada], 10U/ml penicillin, 10µg/ml streptomycin, 1µg/ml cyclosporin A, 25ng/ml interleukin 3 [IL-3; Sigma, UK] and 50ng/ml human stem cell factor [SCF; Sigma, UK]) at 37° in 5% CO_2.

24 hours later the entire culture suspension is moved to a fresh flask, leaving behind the adherent monocytes in the first flask.

After 7 days in phase I, the culture is moved to a 50ml falcon tube, centrifuged and the cell pellet re-suspended in Phase II media (Stem Span Medium, 10U/ml penicillin, 10µg/ml streptomycin, 2U/ml human recombinant erythropoietin [EPO; Sigma, Poole UK] and 50ng/ml SCF) and maintained at a concentration of 1 million/ml for up to 20 days.

2.1.2 Cell morphology

Cytospins of erythroid cells were made by spinning $1x10^5$ cells in a 100-200µl volume onto glass slides. Slides were then stained using a Giemsa staining set (Hema "Gurr", VWR, UK.), and viewed with 40x objective under the microscope (Zeiss Axiolab).

2.1.3 Flow cytometry

Fluorescence activated cell sorting (FACS), a specialized type of flow cytometry, was used to test erythroid cell purity. Glycophorin A (GPA) and CD71 are human red blood cell membrane proteins (Furthmayr et al. 1975). GPA protein is highly expressed on the surface of mature erythroid cells, while CD71 is expressed in proliferating haematopoietic cells. Detection of GPA and CD71 indicates the purity of late erythroid progenitor cells in the culture.

2µl of fluorescent antibodies against human CD71 (fluorescein isothiocyanate [FITC] conjugated, BD Biosciences, Oxford, UK) and anti-human glycophorin A (GPA, R-phycoerythrin [RPE] conjugated, DAKO, Glostrup, Denmark) were added to $1x10^6$ erythroid cells that had been washed in PBS and then re-suspended in 100µl of PBS. Both antibodies were added for a double staining and one sample was prepared without antibody to serve as a negative control. Cells and antibody were incubated for 30 minutes in the dark, followed by two washes with PBS and cells are then re-suspended in a final volume of 1ml PBS containing 1% formaldehyde and 0.1% bovine serum albumin (BSA). FACS analysis of 10,000 cells was acquired using Becton Dickinson flow cytometer and CellQuest software (Becton Dickinson, UK).

2.2 Western blotting

2.2.1 Protein extraction

$5x10^6$ cells were collected, centrifuged and re-suspended in 250µl RIPA cell lysis buffer (20mM Tris-HCl, 150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 1mM EDTA 0.1% SDS) and 12.5µl 20X protease inhibitor (1 Complete EDTA-free Protease Inhibitor Cocktail tablet [Roche, UK] dissolved in 500µl H_2O).

Protein concentration was measured by spectrophotometry at absorbance 280 (Protein A280) and diluted in RIPA buffer and 20µl of 4X SDS loading buffer (10ml 1M Tris, 1ml 10% SDS, 10ml glycerol, 250µl 2-mercaptoethanol, 250µl 1% Bromophenol Blue, 4ml H_2O) to a final concentration 1µg/µl in 100 µl volume.

2.2.2 Protein electrophoresis and transfer

Samples were denatured for 10 mins in a heat block at 95°C. 25µg of each protein (25µl) and 10µl SeeBlue protein marker (Invitrogen) were loaded on a NuPAGE 10% Bis-Tris precast gel (Novex, Invitrogen) in MOPS SDS running buffer (Novex, Invitrogen). The gel was run overnight at 20V using the Novex Invitrogen Surelock X-Cell electrophoresis system, to separate proteins.

Proteins were transferred from gel to nitrocellulose membrane (Amersham Hybond ECL membrane, GE Healthcare) for 2 hours at 55V (X-Cell blot module, Novex Invitrogen) in transfer buffer (3.04g tris, 14.5g glycine, 200mls methanol 800mls H_20). Non-specific binding was blocked with 5% dry milk (Marvel) in PBS-Tween (5 PBS tablets dissolved in 5mls Tween-20 and 995mls H_20) for minimum 2 hours at 4°C with constant rolling.

2.2.3 Antibody probing

The membranes were then incubated overnight in 5% milk containing the primary antibody at 4°C with constant rolling.

The following morning after washing 3 times for 10 mins in PBS-Tween the membranes were incubated in a secondary antibody with 5% milk at room temperature for 1 hour 15mins with constant rolling. This was followed again by three 10 min washes in PBS-Tween.

2.2.4 Target protein detection

Target proteins were detected by an enhanced chemiluminescence kit (Amersham ECL Plus western blotting detection system, GE Healthcare). Following manufacturer's instructions, the membrane was incubated with ECL reagents for 5 mins and drained off. The membrane was wrapped in Saran wrap and placed in an exposure cassette. In a dark room the membrane was exposed to high performance chemiluminescence film (Amersham Hyperfilm ECL, GE Healthcare) for varying amounts of time (between 10 seconds to 3 mins) and the film was developed using the Compact X4 X-Ray film processor (Xograph Imaging Systems, UK)

To re-probe with another primary antibody, membranes were blocked again in 5% milk before incubating in primary antibody overnight as before.

2.3 Chromatin immunoprecipitation (ChIP)

ChIP experiments were carried out using the EZ-ChIP^{TM} kit (cat# 17-371) according to manufacturer's protocols with minor modifications (modifications kindly provided by Dr. David Garrick from the MRC molecular haematology unit, Oxford). All protease inhibitors used in the protocols were complete mini EDTA free tablets (Sigma-Aldrich, UK).

2.3.1 Cell-protein cross linking

The process of cross-linking links DNA to protein so that the DNA can be co-precipitated with the protein using a specific antibody. $5x10^7$ erythroid cells were taken for cross-linking in 10ml growth medium with 1% formaldehyde (Sigma-Aldrich, UK), and incubated for 10 min at room temperature with rolling. 125mM Glycine was added to quench the formaldehyde and then cells were washed in 5ml of cold PBS with protease inhibitors and centrifuged at 1000 rpm for 10 min at 4°C, twice. The pellets were divided into 5 aliquots of $1x10^7$ and stored at -80°C.

2.3.2 Cell lysis

$5x10^7$ cross-linked cells were lysed in 1ml SDS-lysis buffer (EZ-Magna ChIP™ Kit (cat.# 17-409) (Upstate-Millipore, UK) with protease inhibitor and incubated on ice for 15 min, vortexing every 5 mins. Cells were spun at 800 xg for 5 mins at 4°C and the pellet re-suspended in 1ml nucleus lysis buffer (EZ-Magna ChIP™ kit) with protease inhibitors.

2.3.3 Chromatin fragmentation by sonication

Cell/nuclei lysates ($5x10^7$ cells/ml equivalent) were sonicated with a Vibra Cell sonicator (Sonics, US) set at 40 % efficiency for 10x15 seconds using a small probe (2mm in diameter) to achieve DNA fragment lengths of 1000-200bp. The samples were kept on ice during sonication and left to rest on ice for 30secs in-between each round of sonication to avoid overheating with protein denaturation as a

consequence. To remove insoluble material, samples were spun at 10,000 xg for 10 min and the supernatant was moved to a new tube.

2.3.4 Gel analysis of sonication

For analysis of sonication efficiency, 5µl of sonicated material was collected and de-cross-linked over night at 65 °C after the addition of 90µl of H_2O and 4µl 5M NaCl. Following de-crosslinking, 1µl RNase A (10µg/µl) was added and the sample was incubated at 37 °C for 30-60 min to degrade RNA. Subsequently 2µl of 0.5M EDTA, 4µl Tris-HCl (1 M, pH 6.5) and 1µl Proteinase K (10µg/µl) was added followed by a 2 h incubation at 45 °C to degrade proteins. 10µl of chromatin samples were analysed on 1-1.5 % agarose gels.

2.3.5 Immunoprecipitation (IP)

Sonicated chromatin was diluted 10 times in ChIP dilution buffer supplied in the kit and provided with protease inhibitors. 2 x 10µl of sample was removed and stored at 4°C as input sample and kept until the following day.

10-15µg of antibody was added to the diluted chromatin (equivalent to $1x10^7$ cells) for each IP and 40µl of protein-G magnetic beads (EZ-Magna ChIP™ kit) were added and incubated overnight at 4°C on rotation. In the cold room (to keep proteases cold) beads were pelleted using a magnetic tube rack and the supernatant was discarded. Beads were then washed by re-suspending the bead pellet in 1ml of wash buffers supplied in the kit, incubated for 5 min at 4°C and pelleted again by

standing for 1 min on the magnet, the wash was discarded and pellet re-suspended in the subsequent wash buffer and incubated again. Washes were performed in the following order:

1. Low salt immune complex wash buffer (1ml)
2. High salt immune complex wash buffer (1ml)
3. LiCl immune complex wash buffer (1ml)
4. TE buffer (1ml)

The last TE wash was performed at RT.

To elute bound chromatin, 200µl of elution buffer (containing 10µg/ml Proteinase K) was added to re-suspend each sample and the two input samples from the previous day. All samples were incubated at 62°C for 3 hours with constant rotation. In this step the elution buffer separates the chromatin from the antibody and beads, the heat reverse-crosslinks the chromatin protein from the DNA and the Proteinase K treatment destroys the proteins.

Samples were then incubated at 95°C for 10 min and the beads were pelleted using the magnetic rack, the DNA is now in the supernatant which is transferred to a fresh tube.

2.3.6 Purification and ethanol precipitation of ChIP material

200µl of 1:1 phenol/chloroform-isoamylalcohol mixture was added to samples. Samples were vortexed and spun at 13000 rpm for 1 min. The aqueous top layer was transferred to a fresh tube containing 20µl of 3M sodium acetate and 20µg

glycogen (Roche, UK). DNA was precipitated at -20°Covernight and pelleted by centrifugation at 14 000 rpm for 40 min at 4°C. Pellets were washed with 150µl 70% EtOH followed by centrifugation at 14 000 rpm for 10 mins at 4°C. Dried pellets were re-suspended in 20µl of water.

2.3.7 Analysis of enrichment by real time PCR

Purified ChIP DNA was measured using the NanoDrop® Spectrophotometer ND-1000 (NanoDrop Technologies, USA) and analysed for fold enrichment with control primers (Appendix 1, Table 11.2) using real time SYBR-Green PCR which measures double stranded DNA generated. 0.4-0.5µl (~20ng) of ChIP sample and input sample was mixed with 2-4pmol each of forward and reverse primer, and 5µl of SYBR-Green PCR master mix (Applied Biosystems, UK) in a 10 µl reaction volume. All samples were analysed in duplicate using standard real time PCR conditions; 50°C for 2 min, 95°C for 10 min, 95°C for 15 sec and 60°C for 1 min repeated for 40 cycles.

Fold enrichment for an antibody (ab) at a specific primer target (amplicon) in ChIP samples was calculated relative to input DNA and normalised to a sequence in the *NEFM* gene according to the following formula:

$$2^{\overbrace{(C_T\,[\text{antibody}] - C_T\,[\text{input}])}^{\text{Amplicon}} - \overbrace{(C_T\,[\text{antibody}] - C_T\,[\text{input}])}^{\text{Reference}}}$$

2.3.8 Whole Genome Amplification of ChIP material

ChIP material was amplified using the GenomePlex® Whole Genome Amplification (WGA) kit (Sigma-Aldrich, UK) according to a protocol specifically developed for ChIP material (www.sigma-aldrich.com). 100ng of DNA was used for amplification which generated 1-5µg of amplified material (measured by NanoDrop®). Amplified material was purified using the QIAquick PCR Purification Kit (Qiagen, UK) according to the manufacturer's protocol with the exception that the buffer PBI, which is not recommended for microarray applications, was substituted for buffer PB.

2.4 Microarray analysis of ChIP material (ChIP-chip)

Sample labelling, hybridization, washing and microarray scan was performed using the NimbleGen system (Roche-NimbleGen, Madison, US, http://www.nimblegen.com) according to manufacturer's protocol (NimbleGen array's user guide for ChIP applications).

2.4.1 Labelling of ChIP material and microarray hybridisation

1µg each of input DNA and ChIP DNA was diluted to a 40µl volume with H_20. 40µl of Cy3 or Cy5 fluorescently labelled random primers were added to the input and ChIP samples, respectively. The Cy3 and Cy5 labelled random nonamers (TriLink Biotechnologies, USA) were previously diluted in random primer buffer (125mM Tris-HCl pH 7.4, 12.5mM $MgCl_2$ and 0.18 % β-mercaptoethanol) to a concentration of 1 O.D unit per 40µl. Samples and primer mixtures were denatured at 98°C for 10 mins.

Labelling reactions were set up in 100µl volumes. To each sample 10µl of 50X dNTP, 2µl (100 U) Klenow fragment (New England Biolabs, UK) and 8µl of Sigma water was added. Sample mixtures were incubated for 2 h at 37°C which labelled the ChIP DNA by primer extension.

Reactions were stopped by the addition of 10µl 0.5 M EDTA. 11.5µl NaCl (5M) and 110µl isopropanol was added to each sample to precipitate the DNA for 10 mins at room temperature. Samples were centrifuged at 12 000 rpm for 10 mins. DNA pellets were washed with 500µl 80% ice-cold ethanol. Labelled DNA samples were re-suspended in 4.5µl water. 0.5µl of labelled DNA was diluted 10 x and DNA concentrations were quantified using the Nanopdrop spectrophotometer. As Cy primers are light sensitive samples were protected from light during all steps.

2.4.2 Microarray Design

Custom built microarrays were generated, and were composed of 50mer oligonucleotide probes that tile non-repetitive genomic sequence of: 3.8 Mb of 6q23, including the intergenic region associated with HbF between co-ordinates chr6:133,500,000–137,300,000; 1.1 Mb of chr11p including the entire β globin cluster region from co-ordinates chr11:4,730,996–5,732,587 and 500 kb of the telomeric region of chromosome 16, including the α globin Encode region starting at co-ordinates chr16: 1–500,000 at 80 bp intervals.

2.4.3 ChIP-on-chip hybridisation

7.2µg of Cy3 input and Cy5 ChIP labelled DNA were mixed together and made up to 6µl with sigma water. The hybridisation reaction mixture contained labelled alignment oligos which hybridize to alignment features on the arrays needed for data extraction, 2x hybridisation buffer (11.8µl) and hybridisation component A (4.7µl), (NimbleGen Hybridisation kit). 15.6µl of this mixture was added to the 6µl sample and incubated at 95°C for 5 mins and then at 42°C until ready to load onto the array slide.

Hybridisations were done using NimbleChip x1 Mixers and the NimbleGen hybridisation system. The microarray slide has a hybridisation chamber over the array to which the hybridisation solution is loaded at a fill port on the mixer. Once the slide is in place and the sample loaded to cover the whole array (16-20µl of sample) the machine is switched on to mix the hybridisation solution over the array by a pumping mechanism which allows even hybridisation at 42 °C for 16–20 h.

Microarrays were washed using wash buffers, wash tanks and slide rack, all supplied in the NimbleGen Hybridisation kit. 25µl DTT (1M) was added to 25mls of each wash buffer. Microarrays and mixers were first disassembled in wash buffer 1 previously warmed to 42 °C. Following disassembly, the microarrays were washed by vigorous and constant agitation for 2 min and 15 s in wash buffer 1, 1 min in wash buffer 2 and 15sec in wash buffer 3. Immediately after the last wash, the microarrays were dried by centrifugation for 1 min using a microarray centrifuge.

2.4.4 Scanning of microarrays

The hybridised microarrays were placed in the GenePix 4000B scanner and scanned using GenePix software (Molecular Devices, Sunnyvale, USA) according to the NimbleGen Arrays protocol and images saved for analysis.

2.4.5 Microarray data analysis

Microarray data was extracted and analysed using the NimbleScan software according to the NimbleGen Arrays protocol. The NimbleScan software aligns the probes on the microarray to a design file describing probe location using the locations of alignment features on the array. The program then calculates a signal ratio of input and ChIP sample for each probe on the array and extracts statistically significant peaks from the ratio. These peaks indicate binding sites. Once this data is converted into track files, these can be uploaded and viewed on the UCSC genome browser (genome.ucsc.edu, assembly March 2006, NCBI36/hg18), thus displaying binding signals across the genome. In this way multiple tracks can be viewed in a single window at the desired loci.

2.5 Gene expression

All DNA, RNA and protein concentrations were measured on a Nanodrop® 2000 spectrophotometer (Thermo Scientific). Thermal cycler used throughout was the PTC-200 Peltier thermal cycler (MJ Research).

2.5.1 Total RNA extraction

Total RNA was extracted from cells using TRI Reagent (Sigma, Poole, UK) according to manufacturer's instructions. 1ml of TRI-reagent was used to lyse between 5-10 $x10^6$ cells, for less than $5x10^6$ cells a minimum of 0.5ml of TRI-reagent was used. RNA was resuspended in 100µl of DEPC-treated water (Applied Biosystems, Ambion) and stored at -80°C. RNA was quantified by Nanodrop. The absorption maximum of RNA is 260nm and the ratio of the absorbance at 260 and 280nm is used to assess the RNA purity of a sample. The Nanodrop measures the absorbance from the region of 200nm to 350nm to determine the RNA concentration and purity.

2.5.2 Reverse Transcription

First strand cDNA was generated from RNA by using Superscript™ III Reverse Transcriptase (Invitrogen, Paisley, UK). Reactions were set up with 500ng RNA, 1µl oligo dT (0.5µg/µl) and 1µl dNTPs (10mM) in a 13µl volume and incubated on a PCR block at 65˚C for 5 min. 4µl of 1st strand buffer (5x), 1µl DTT (0.1 M), 1µl Superscript™ III RT (200U/µl) and 1µl water were added to a final volume of 20µl. The reactions were incubated on a PCR block at 50˚C for 60 mins followed by 70˚C

for 15 mins and cooling to 4˚C. cDNA was diluted with water to a total volume of 50µl and stored at -20°C.

2.5.3 TaqMan Real Time PCR

Custom designed primers and probes for TaqMan real time PCR analysis were ordered from Applied Biosystems (UK) (Appendix 1, Table 11.1). Primer probe mixes for *BCL11A* (Hs00256254- Invitrogen, UK) and housekeeping genes HPRT1 (Hs99999909) and *GAPDH* (Hs99999905) were ordered as ready-made gene-expression assays from Applied Biosystems. 10µl reactions were set up with 2µl cDNA, 1.5pmol each of forward and reverse primer, 2pmol of fluorescent probe and 5µl of TaqMan PCR master mix (Applied Biosystems, UK) and run on an ABI Prism 7900HT Sequence Detection system (Applied Biosystems, UK) under the following conditions;

Step1) 50°C for 2 min,

Step 2) 95°C for 10 min,

Step 3) 95°C for 0.15 min and 60°C for 1 min repeated for 40 cycles.

Real time PCR measures the number of cycles to reach the threshold of fluorescence detection. The cycle at which the threshold is reached is called the threshold cycle or C_T, which provides a value for the starting copy number of a DNA target. A lower C_T value means fewer amplification cycles were required to reach the threshold of fluorescence detection and therefore the higher the starting amount of our specific DNA sequence.

Gene expression was calculated as fold change between levels of the target gene in the sample versus negative control and then normalized to the housekeeper genes.

2.5.4 DNase Treatment

RNA samples from overexpression plasmid transfected K562 cells were DNase treated for removal of plasmid DNA using Ambion® TURBO DNA-*free*™ Kit (Applied Biosystems, Ambion) prior to reverse transcription. 5µl of 10X TURBO DNase buffer and 1µl of TURBO DNase was added to 5µg of RNA in a 50µl reaction volume and incubated at 37°C for 30 mins. 5µl of DNase inactivation reagent was added to the reaction and incubated for 5 mins at room temperature and mixed at least 3 times over the 5 min incubation period. Finally sample was centrifuged at 10,000 xg for 1.5 min and transferred to a fresh tube.

2.6 DNA Sequencing

2.6.1 Standard PCR and gel purification

DNA was specifically amplified by standard PCR. 25µl reactions were set up with 5µl of DNA, 10µM each of forward and reverse primers (Appendix 1, Table 11.3, 11.4 & 11.7), 10mM dNTP, 2.5µl $MgCl_2$, 2.5µl 10X buffer and 0.4µl of AmpliTaq Gold. The reactions were run on a thermal cycler under the following conditions;

Step1) 94°C for 10 min

Step 2) 94°C for 0.30 min

Step 3) 55°C for 0.30 min

Step 4) 72°C for 0.30 min

Step 5) Repeat steps 1-4 for 30 cycles

Step 6) 72°C for 2 min

PCR product was initially visualized by running a small amount of the product on an agarose gel. Once PCR quality was confirmed, all the remaining PCR product was run on a fresh agarose gel. The gel was then placed on a UV imager, the product was visualized and the band(s) scored with a scalpel. Using the incisions as precise markers the product was cut out, placed in a 1.5ml eppendorf tube. The gel was dissolved and DNA purified as per the protocol outlined in the Promega Wizard SV Gel and PCR Clean-up System instruction booklet. DNA concentration was then measured by the Nanodrop®.

2.6.2 Cycle sequencing

Sequencing reactions were carried out using the ABI PRISM BigDye Terminator v3.0 Ready Reaction Cycle Sequencing Kit following protocols as outlined in the user's manual. 20µl reactions were set up with 200-300ng of DNA, 3.2pmol primer and 4µl of 5 X buffer and 2µl of Terminator Ready Reaction Mix. Reactions were run on a thermal cycler according to manufacturer's protocol (Applied Biosystems, Warrington, UK).

2.6.3 Ethanol precipitation

Reactions were transferred to 1.5ml eppendorf tubes containing 16µl deionized water and 64 µl of 95% ethanol. Samples were vortexed and incubated at room temperature for 15 mins, followed by a 20 minute spin at maximum speed.

Supernatant was discarded and sample washed with 400µl of 70% ethanol and spun again for 10 mins. Supernatant was again discarded and pellets dried at room temperature.

2.6.4 Electrophoresis and data collection

The pellets were re-suspended in 20µl of formamide, vortexed well and spun down. Samples were then transferred to a 96-well plate and pulse spun to 1000rpm. Samples on the plate were then denatured for 2 minutes at 95˚C on the thermal cycler. The plate was then set up and run on the sequencing machine 3130xl Genetic Analyser (ABI Prism by Applied Biosystems) as per the manufacturer's instructions. Data was analysed using the DNA sequence analysis programme Sequencher® version 4.6 (Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com).

2.7 Vector construct preparation

2.7.1 Restriction enzyme digests

Typically, restriction enzyme digests for 1µg of DNA were set up in a total volume of 25µl with 5 units of enzyme (New England Bioland, Ipswich, UK and Roche, Burgess Hill, UK) and the supplied buffer. Bovine serum albumin (BSA) was added according to manufacturer's protocol. Restriction digests were incubated in a water bath at 37 ºC (unless the manufacturer's protocol stated otherwise) for 2-4 hours.

2.7.2 A-tailing

30µl gel-purified RT-PCR products were A-tailed by adding 1U of AmpliTaq Gold, PCR buffer, 10mM dATP and incubating at 72˚C for 15 minutes.

2.7.3 Ligations

To avoid religation of vectors digested with single restriction enzymes, 2 units of alkaline phosphatase (Roche, Burgess Hill, UK) per 1µg of DNA was added to the restriction digest reaction.

Ligations were set up with an approximate vector to insert molar ratio of 1 to 3 in a total volume of 10µl. After the addition of 1 unit of T4 DNA ligase and T4 DNA ligase buffer (Roche, UK), the ligations were left in room temperature for 1 h followed by incubation over night at 4°C. Ligations were transformed the following day.

2.7.4 Heat shock transformations

Competent DH5α bacteria (Invitrogen, Paisley, UK) were used for transformations. Typically 5µl of ligation mixture was added to 50µl bacteria and incubated on ice for 30 min. Bacteria were heat-shocked at 42°C for 30 s and chilled on ice for 5 min. Following the addition of 250µl SOC medium (Invitrogen, Paisley, UK), the bacteria were incubated in 37°C with gentle shaking at 250 rpm. Transformed bacteria were grown on LB-agar plates with the addition of appropriate antibiotics and incubated at 37°C over night.

2.7.5 Colony PCR

A small amount of bacterial colony was picked with a sterile tip and resuspended in 10µl of water in PCR tubes. Following incubation at 96°C for 5 min, 15µl of PCR master mix (0.7pmol/µl forward primer, 0.7pmol/µl reverse primer, 4.1mM $MgCl_2$, 0.3mM dNTP, 2.5µl 10 x Taq gold buffer, 1.25 units Taq Gold) was added to each tube to obtain standard concentrations of reagents in a 25µl reaction volume. A standard PCR program was run and PCR products were analysed for positive clones on agarose gels.

2.7.6 Plasmid preparations

Bacteria for plasmid preparations were grown in LB medium in a shaking incubator at 37 °C. Plasmids were extracted from bacteria using a QIAprep® Spin Miniprep kit (Qiagen, Crawley, UK), or a HiSpeed Plasmid maxi kit (Qiagen, Crawley, UK) according to manufacturer's protocol.

All vector constructs were prepared in the group laboratory by Dr Steve Best.

2.8 Transfection of K562 cells by electroporation

4 x10^6 cells were centrifuged and re-suspended in 400µl of serum free RPMI culture medium per sample. Cells were transferred to electroporation cuvettes. 5µg of vector construct was added to the cells, the cuvette was capped and mixed gently. Transfection was achieved by electroporating cells using BioRad Gene Pulser® II following manufacturer's instructions at 960µF capacitance and 250V per pulse.

Following electroporation, cells were transferred to 10 ml of complete RPMI culture medium in small culture flasks and incubated at 37°C for 48hrs.

2.9 Luciferase reporter assays

Firefly and renilla luciferase levels were measured 24-36 h after electroporation using the Dual-Glo Luciferase Assay System (Promega, Southampton, UK). Cells were counted and the amount of cells used for assay was based on the lowest count. Cells were spun down at 1000 rpm for 5 min and resuspended in additive-free medium to obtain a volume of 225µl. Readings were initially taken for firefly luciferase and then for renilla luciferase. For each electroporation, triplicates of 75µl were loaded on a 96 well plate and 75µl firefly of luciferase reagent was subsequently added to each well. Reactions were left at room temperature for 10 min before luciferase levels were measured using a Lucy2 luminometer (Anthos, Engendorf, Austria) or a DTX 880 Multimode Detector (Beckman Coultier, High Wycombe, UK). Following Luciferase readings, 75µl of Renilla luciferase reagent was added to each well. Reactions were left for 10-15 min prior to renilla luciferase readings.

CHAPTER 3

3 Erythroid culture system

3.1 INTRODUCTION

In the past, *in vitro* haematopoiesis has been studied mostly by semi-solid culture systems in which colonies are formed from erythroid progenitor cells. Because the cells are immobilized in this culture system, it makes it very difficult to analyse the growth kinetics and characterise the developing cells. To study erythroid cell development, a more robust method to overcome the limitations of semisolid cultures was required and a new method was introduced by Fibach and colleagues in 1989. The group proposed a two-phase liquid culture system that supports the growth and maturation of human erythroid progenitor cells.

The conditions of culture systems are developed to mimic the natural environment of erythropoiesis. The cells yielded using these culture systems are very similar to their *in vivo* counterparts. This means morphological changes and gene expression changes accompanying erythroid differentiation can be studied in these cells. Due to their leukemic origin and extensive culturing, cell lines such as K562s have lost many features of the native erythroid cell and are therefore an unsuitable model to study erythroid gene regulation. The Fibach liquid culture system is therefore an ideal system for providing erythroid progenitor cells at different stages of erythroid differentiation. Primary human erythroid cells can be obtained from healthy individuals and patients for *in vitro* genomic studies such as transcription factor binding by chromatin immunoprecipitation.

Erythroid progenitor cells can be derived from bone marrow aspirates, fetal livers, cord blood and peripheral blood (PB). There are certain advantages to using peripheral blood over other sources including ease of access. More importantly, PB

contains a homogenous population of BFUe, providing a more uniform starting point for cultures (Fibach *et al.* 1989).

The first phase isolates erythroid progenitor cells by depleting all other cell types while expanding growth of the progenitor cells. The second phase promotes the proliferation and maturation of the erythroid progenitor cells. This selection is achieved by administration of key growth factors including erythropoietin (EPO) at the different stages of the culture.

For the work presented in this thesis I used the Fibach two-phase liquid culture system with some variations as previously established by our group (Jiang *et al.* 2006). The main difference is that I used a semi-synthetic Stem Span medium (Stem Cell Technologies) instead of the fetal calf serum originally described by Fibach (*et al.* 1989). I also substituted the 5637 cell conditioned medium originally described; with interleukin-3 (IL-3) and stem cell factor (SCF).

Firstly, density gradient centrifugation using Histopaque-Ficoll (Sigma-Aldrich, UK) is used to separate the blood components. Red blood cells are heavier and sink to the bottom and the mononuclear cells remain in a layer at the plasma-histopaque interface (buffy coat), from where they are collected. The buffy coat contains a small number of progenitor cells and mostly monocytes and lymphocytes. These cells are now suspended in Phase I medium containing IL-3 and SCF which bind to their respective receptors on the surface of the erythroid progenitor cells and stimulate proliferation of the cells. The Phase I medium also contains Cyclosporin A, an immunosuppressive agent which causes depletion of the lymphocytes. Monocytes adhere to the bottom of the culture flask and are removed by transferring the culture suspension gently into a fresh flask.

Over 7 days the progenitor cells have had a growth advantage and are now transferred into Phase II medium. Phase II medium contains EPO which stimulates the proliferation and differentiation of erythroid progenitor cells, and dexamethasone which increases the sensitivity of the cells to EPO (Golde *et al*. 1976). During incubation in Phase II medium, cells develop into mature erythroid precursor cells (Figure 1.4).

For each ChIP experiment I required 10^6 cells and this culture method was the best to yield large numbers of pure erythroid cells. I assessed cell development morphologically by analysing cell cytospins. The purity and development of cells was further analysed by flow cytometry. The full methods are outlined in chapter 2.1.1.

Erythroid cultures used for representative cytospins and FACS analysis in results section 3.2 were done in collaboration with fellow PhD students Suleyman Aktuna and Dr. Karin Wahlberg.

3.2 RESULTS

3.2.1 Analysis of cell morphology

In Phase II, proliferation of the erythroid progenitor cells were maintained at a cell count of 10^6/ml. Cell differentiation was monitored daily by cell cytospins (see chapter 2.1.2).

Small round erythroid progenitor cells were visible in the culture from day 1 of Phase II (Figure 3.1) with the presence of some prepro-erythroblasts which increased in number by day 4. The cell cytospin taken on day 7 showed the majority of the cells in the culture were now pro-erythroblasts. Basophilic erythroblasts began to appear from day 10 which were distinctly smaller than the pro-erythroblasts.

As the cells developed into polychromatic erythroblasts (day 12) they reduce in size and the nuclei become more condensed. By day 16 the cells are smaller yet, with a pale cytoplasm as the dark nucleus moves to one side of the cell which is now called an orthochromatic erythroblast. Soon the nucleus is expelled from the cell as seen on day 18 (Figure 3.1).

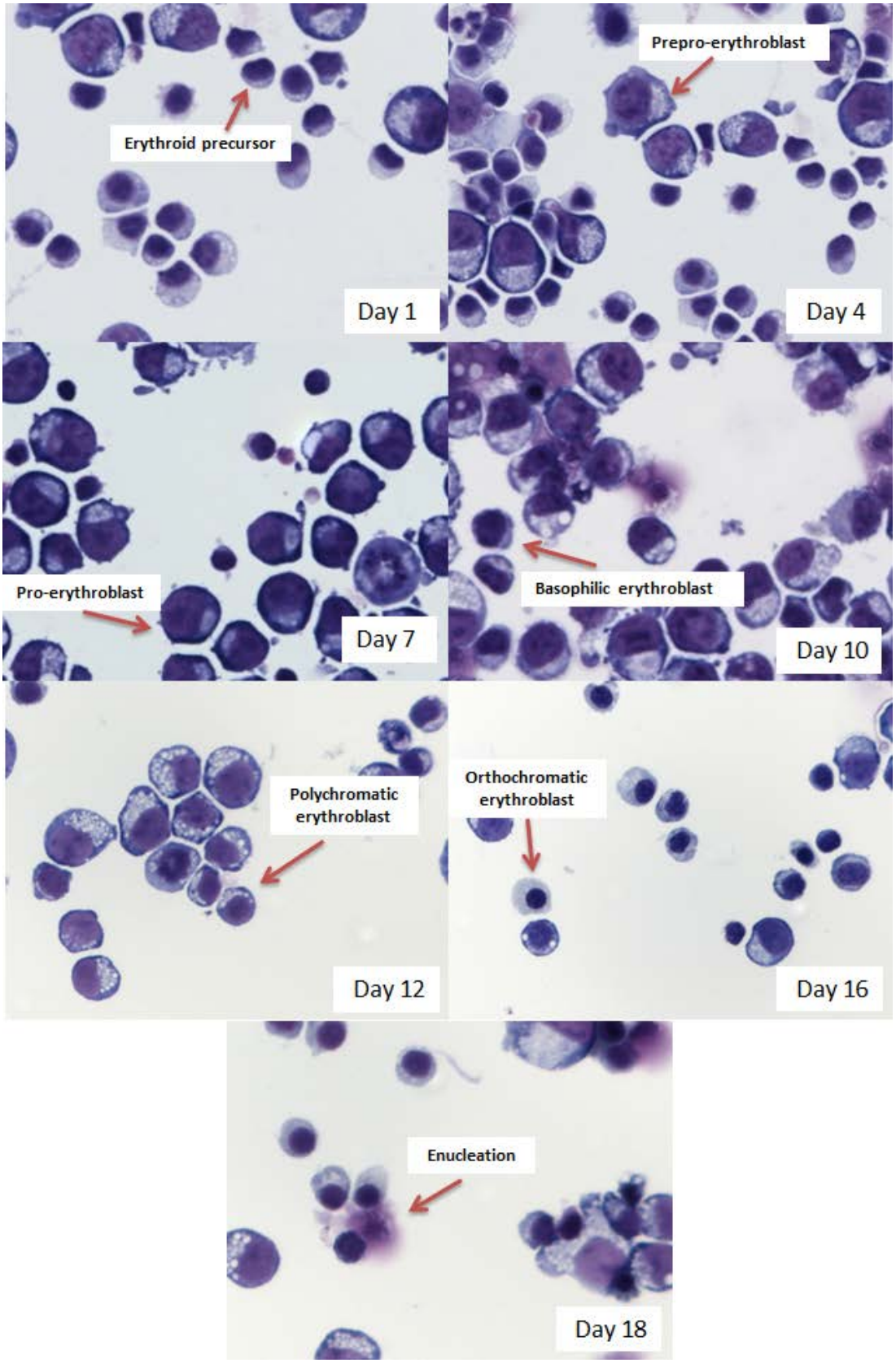


Figure 3.1 Differentiation of primary human erythroid progenitor cells. Cytospins of cells taken in Phase II day 1-18 of a two-phase liquid culture system adapted from Fibach *et al.* (1989). Morphological stage of cells are marked out on each day.

3.2.2 Purity and differentiation of erythroid cells in Phase II of culture

Cell differentiation and culture purity was further analysed by a specialised form of flow cytometry called fluorescence activated cell sorting (FACS). I used two human erythroid cell surface proteins as markers of cell development and purity; Glycophorin A (GPA) (DAKO, Glostrup, Denmark) and transferrin receptor protein 1 (CD71) (BD Biosciences, Oxford, UK) (Furthmayr *et al.* 1975). GPA is an erythroid specific protein and is up-regulated as the cells develop into erythrocytes. CD71 is highly expressed on the surface of proliferating haematopoietic cells, including erythroid progenitors during the early stages of development.

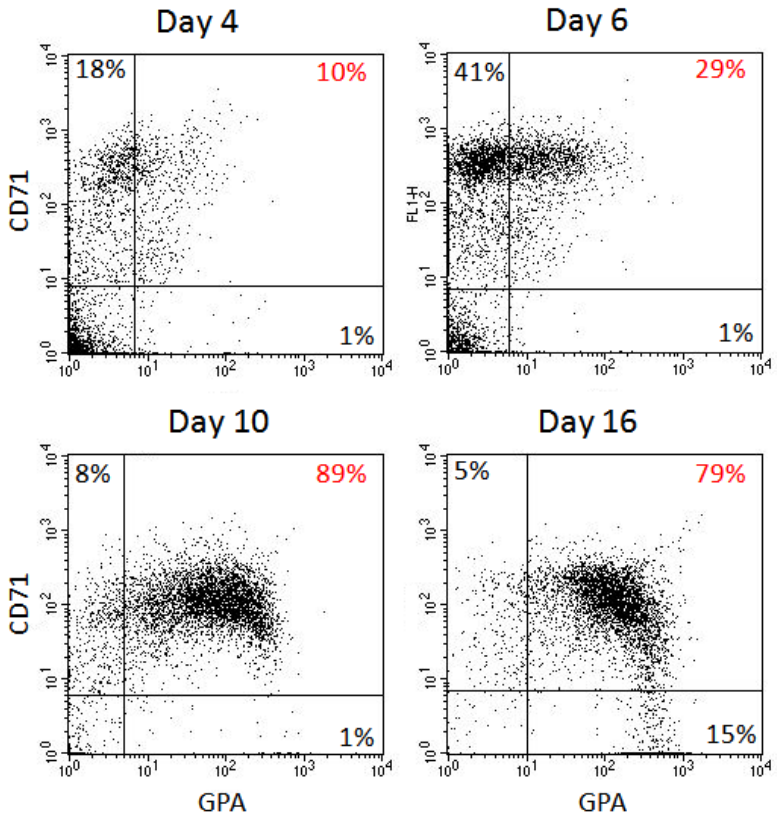


Figure 3.2 Differentiation and purity of primary human erythroid progenitor cells by FACS analysis. Cells in Phase II day 4, 6, 10 & 16 of modified Fibach culture are double stained with antibodies against CD71 and GPA. The y-axis shows CD71 expression and the x-axis shows GPA expression. Lower left box: double negative cells. Upper left box: CD71 positive, GPA negative cells. Upper right box: GPA and CD71 positive cells. Lower right box: GPA positive, CD71 negative cells. Markers (black lines) separating each field were set according to single stained and unstained control samples from the same culture at each day.

As cells proliferated in the early stages of culture, CD71 expression on cells increased from 28% on day 4 to 97% on day 10, the maximum that it would reach (Figure 3.2).

GPA is expressed later as the cells develop. We see a rapid increase in GPA expression between day 6, when 30% of cells analysed were positive for GPA and day 16 when we observed 90% of GPA positive cells. GPA positive, CD71 negative cells appeared after day 10 and by day 16 represented 15% of cells analysed (Figure 3.2). CD71 is a non-specific marker, in that it is a marker for all proliferating cells and not just erythroid cells. GPA however is an erythroid-specific marker, and therefore is the more important factor to monitor.

3.2.3 Gene expression profiles

For ChIP experiments, I needed to harvest cells at the peak expression stage of the transcription factor to be analysed. I constructed expression profiles of the different genes by qPCR of RNA extracted from erythroid cells at different days of Phase II and reverse transcribed to cDNA. Ct values were normalised to the house-keeping gene HPRT and fold change in expression was calculated relative to day 4 (*KLF1* and *BCL11A*) and day 6 (*GATA1)*.

KLF1 and *BCL11A* were up-regulated as cells developed, reaching the peak expression on day 10 (Figure 3.3). Day 10 also happens to be the peak expression of *GATA-1.*

This data shows that all three genes of interest (*KLF1, BCL11A* and *GATA1)* are at peak expression around day 10 of culture in phase II. This coincides with high levels

of GPA expressing cells (90%, Figure 3.2) at the basophilic erythroblast stage (Figure 3.1).

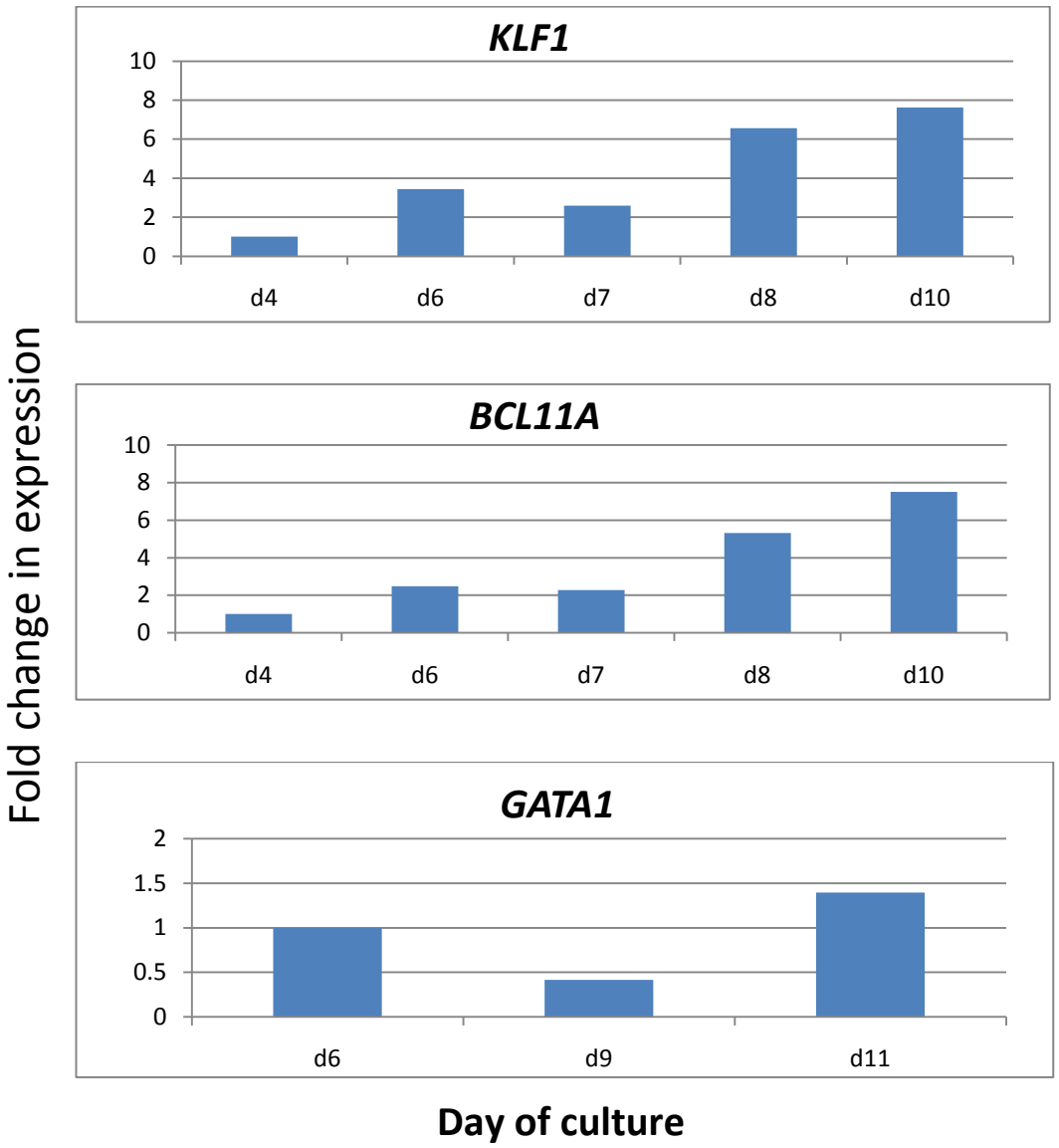


Figure 3.3 Expression profiles of *KLF1, BCL11A* and *GATA1* in primary human erythroid cells taken on different days of culture (Phase II). Analysed by qPCR in duplicate reactions using gene specific primers. Data was normalised to house-keeping gene HPRT and the first day on which cells were extracted for the expression profiles.

3.3 DISCUSSION

For the purpose of ChIP experiments I needed a large number of a homogenous population of primary human erythroid cells at synchronised stages of expression. Removal of macrophages in the early stages of culture (Phase I day 1) led to the generation of a purer culture of erythroid cells. However, small numbers of macrophages are beneficial to erythroid cell development *in vivo* and *in vitro* (Chasis & Mohandas 2008, Leimberg *et al.* 2005) and this was previously demonstrated by our group in the following experiments.

Monocytes in Phase I were depleted by use of a lysosomotropic agent, l-leucin methyl ester (LME), which causes selective death of monocytes in culture. LME treated cultures contained no macrophages on day 0 of phase II. A high amount of cell death was observed from day 6 of phase II with no viable cells visible on cytospins at day 10. In contrast, the control culture (no LME) had a small number of macrophages on day 0 of phase II and erythroid precursor cells had developed to a homogenous population of pro-erythroblasts at day 6.

Cell differentiation in the control culture was, however, slower than that of previous cultures containing higher numbers of macrophages. The technique by which the majority of monocytes are removed and a small number of macrophages remain in culture that is achieved in this method (i.e. the transfer of the culture to a fresh flask which removes many macrophages) is ideal for the generation of a pure culture.

Culture purity was confirmed by FACS, showing 90% of cells were positive for GPA at day 10, the majority of which were morphologically homogenous. At day 10, there was a high level of transcription factors KLF1, BCL11A and GATA-1 expression.

I conclude that Phase II day 10 of culture is the most appropriate time to harvest erythroid cells for ChIP analysis for KLF1, BCL11A and GATA-1.

CHAPTER 4

4 GATA-1 binding at the 6q23 locus

4.1 INTRODUCTION

The *HBS1L-MYB* intergenic region contributes 19% of the variation in HbF levels in healthy northern Europeans (Thein *et al.* 2007). SNPs that account for the effects of the 6q locus on HbF levels are distributed in 3 linkage disequilibrium blocks (*HMIP1*, *2* and *3*) which occupy a 79kb segment between the *HBS1L* and *MYB* genes. The SNPs showing the strongest association with HbF levels are found in block 2 (*HMIP2)*.

Wahlberg *et al. (*2009) identified DNase I hypersensitive sites in *HMIP* block 2 suggesting the presence of regulatory elements in the *HBS1L-MYB* intergenic region. I proceeded to analyse the region for transcription factor binding and chromatin modification by chromatin immunoprecipitation on microarray (ChIP-chip). The use of microarrays would allow us to see a more extensive area of the 6q23 locus.

Chromatin immunoprecipitation is a method to study the interaction of proteins of interest with their target DNA sequences *in vivo* (Figure 4.1). Material used for ChIP experiments was harvested from primary human erythroid cell cultures at a time when our protein of interest is expressed. Expression profiles for each target protein to be studied were constructed. Primary human erythroid cells were cultured from whole blood using the method adapted from Fibach *et al.* (1989) and detailed in chapter 2.1.2. RNA was extracted from cells at different days of culture and reverse transcribed into cDNA. Using primers specific for each target gene, expression was measured by qPCR and I used this information to estimate the optimal day to harvest cells for ChIP experiments.

On harvesting, cells were cross-linked using 1% formaldehyde, this step fixes the proteins to the target DNA sequence in their native state. The material is then sonicated to shear the DNA to fragments between 200-1000bp (NimbleGen array's user guide for ChIP applications). The material was then immunoprecipitated with an antibody to our specific protein of interest. Once the antibody is bound to its target protein, excess antibody is washed away and the antibody-protein-DNA material is reverse cross-linked, this time detaching the protein from its target DNA sequence.

At the end of this step, we are left with DNA that has been specifically targeted by our protein using a specific antibody. The amount of DNA resulting from a ChIP is usually very low and to analyse this on a microarray I required 1.5µg of DNA (NimbleGen array's user guide for ChIP applications). This ChIP DNA is amplified globally using a whole genome amplification kit (Sigma-Aldrich) before it is labelled with fluorescent dyes and hybridised to a custom micro-array. Our custom micro-arrays contain several different loci of interest including the globin gene loci and the 6q23 locus. The micro-array is scanned and analysed using computer software (SignalMap and NimbleScan); the fluorescence signals detected indicate the binding sites of our protein of interest. The full ChIP-chip protocol is outlined in Chapter 2.3.

GATA-1 is an erythroid specific transcription factor, required for globin gene regulation and erythroid cell maturation. The important regulatory role of GATA-1 in erythropoiesis makes it an ideal candidate to study transcriptional activity at the 6q23 locus. I decided to look at GATA-1 binding on the *HMIP* locus by ChIP-chip to further understand the function of this region.

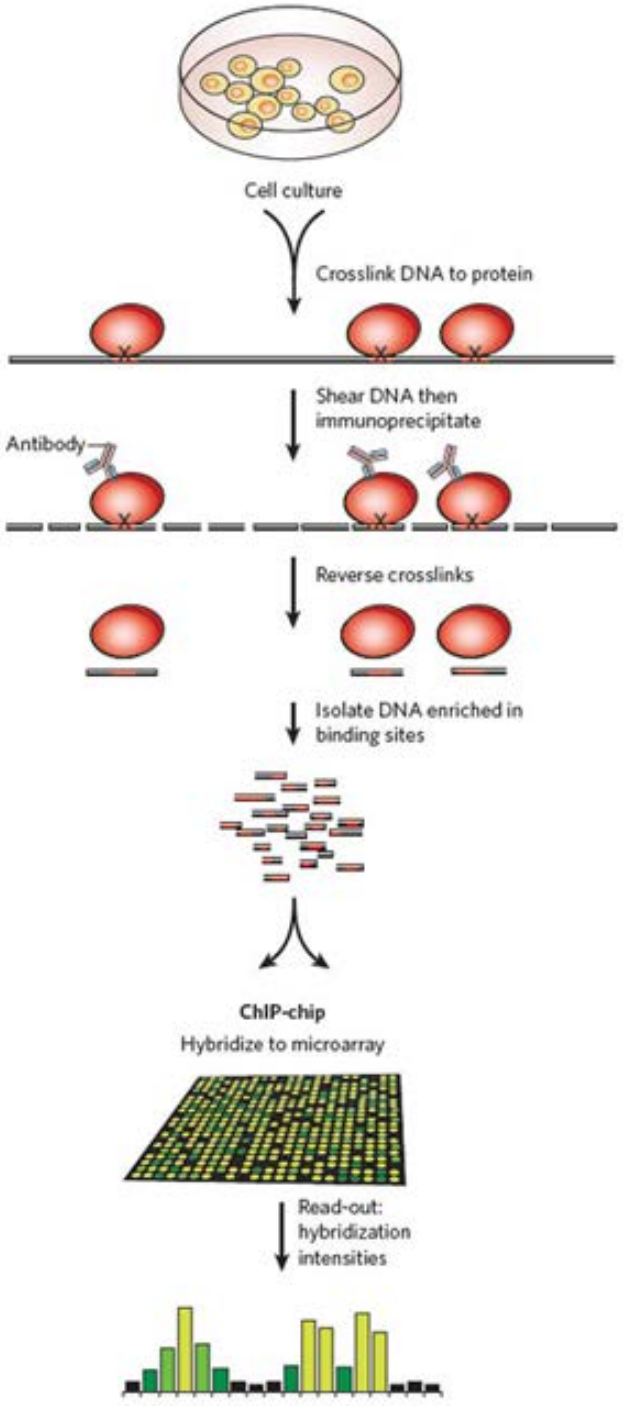


Figure 4.1 Illustration of ChIP-chip protocol. Image adapted from 'Genomic views of distant-acting enhancers' Axel Visel, Edward M. Rubin & Len A. Pennacchio. Nature 461, 199-205, 2009.

4.2 RESULTS

4.2.1 DNA sonication

Following crosslinking of primary human erythroid cells and cell lysis, the material was sonicated to shear the DNA. 1µl of the sonicated chromatin was run on a 1.5% agarose gel to confirm the size of the DNA fragments (Figure 4.2, lanes 2-3). Figure 4.2 shows chromatin was successfully sheared to 100-600bp.

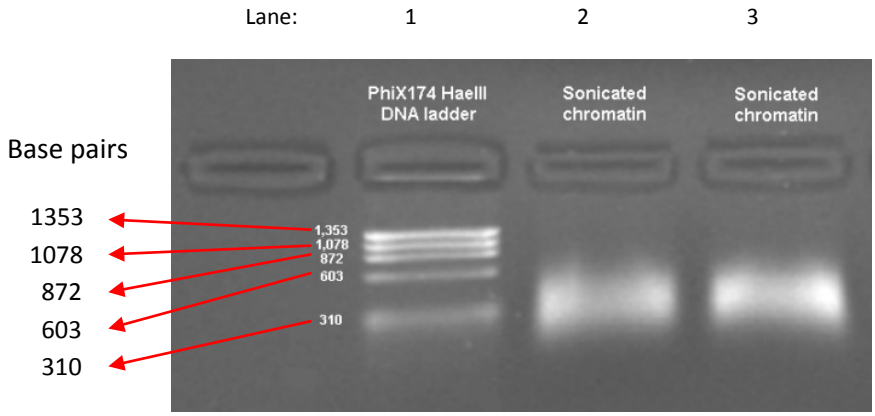


Figure 4.2 Sonication of chromatin. Chromatin sample was run in duplicate (lanes 2 and 3) along with PhiX DNA marker (lane 1), on a 1.5% agarose gel. Product size is as indicated in base pairs on the left of the image.

4.2.2 Control test for GATA-1 enrichment

Following confirmation of the chromatin size, I proceeded to immunoprecipitation using a ChIP grade anti-GATA1 antibody (M-20 sc1234, Santa Cruz). After reverse cross-linking and purification of the ChIP DNA, a small amount was tested for enrichment by SYBR-Green qPCR. Hypersensitive site 40 (HS-40) upstream of the α-globin locus on chromosome 16 (*α-HS40)* is a known GATA-1 binding site and was used as a positive control in GATA-1 ChIP experiments. This step is a check point to

confirm if ChIP with the target antibody has been successful. On average, nine-fold enrichment was seen after ChIP with anti-GATA-1 antibody, as demonstrated in Figure 4.3.

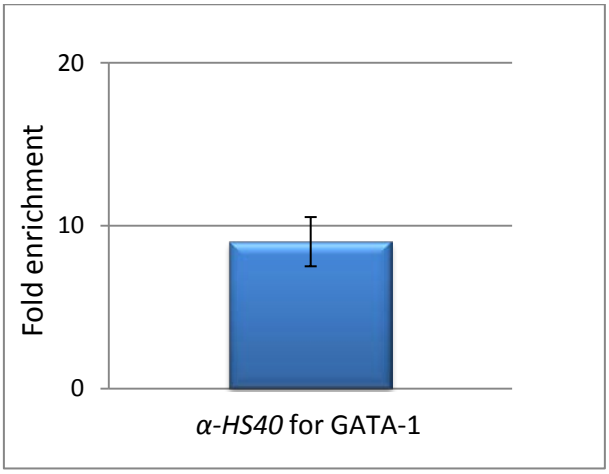


Figure 4.3 Enrichment of *α-HS40* to confirm GATA-1 ChIP (using antibody GATA-1 M-20 sc-1234, Santa Cruz) by SYBR-Green qPCR. Shown as fold enrichment, normalized to *NEFM* as a housekeeper gene and input sample (sample before ChIP with antibody). Datum is mean of at least 3 experiments, +/- SE.

After validating GATA-1 ChIP, I continued to process the remaining ChIP material for a global view of GATA-1 binding in the 6q23 intergenic locus (position chr6: 135,373,000-135,582,000). Initial analysis was done using SignalMap (Roche, NimbleGen) where data extracted from the microarray scan on NimbleScan was mapped onto tracks to view our specific genomic regions. I then constructed our own tracks of the GATA-1 peaks at the 6q23 locus and surrounding area which were uploaded and viewed on the online UCSC genome browser (genome.ucsc.edu) which provides a more detailed view of the region, with the flexibility of viewing SNP tracks, conserved GATA-1 sites, hypersensitive regions, DNA sequence, other transcription factor sites, as well as many other tracks available, all in the same view (Figure 4.4).

ChIP-chip data to investigate AcH3, AcH4 and RNA polymerase II binding in the 6q intergenic region in primary human erythroid cells was obtained in the same way in collaboration with fellow student Karin Wahlberg (Wahlberg *et al.* 2009).

4.2.3 GATA-1 binding at the 6q23 locus (*HMIP*)

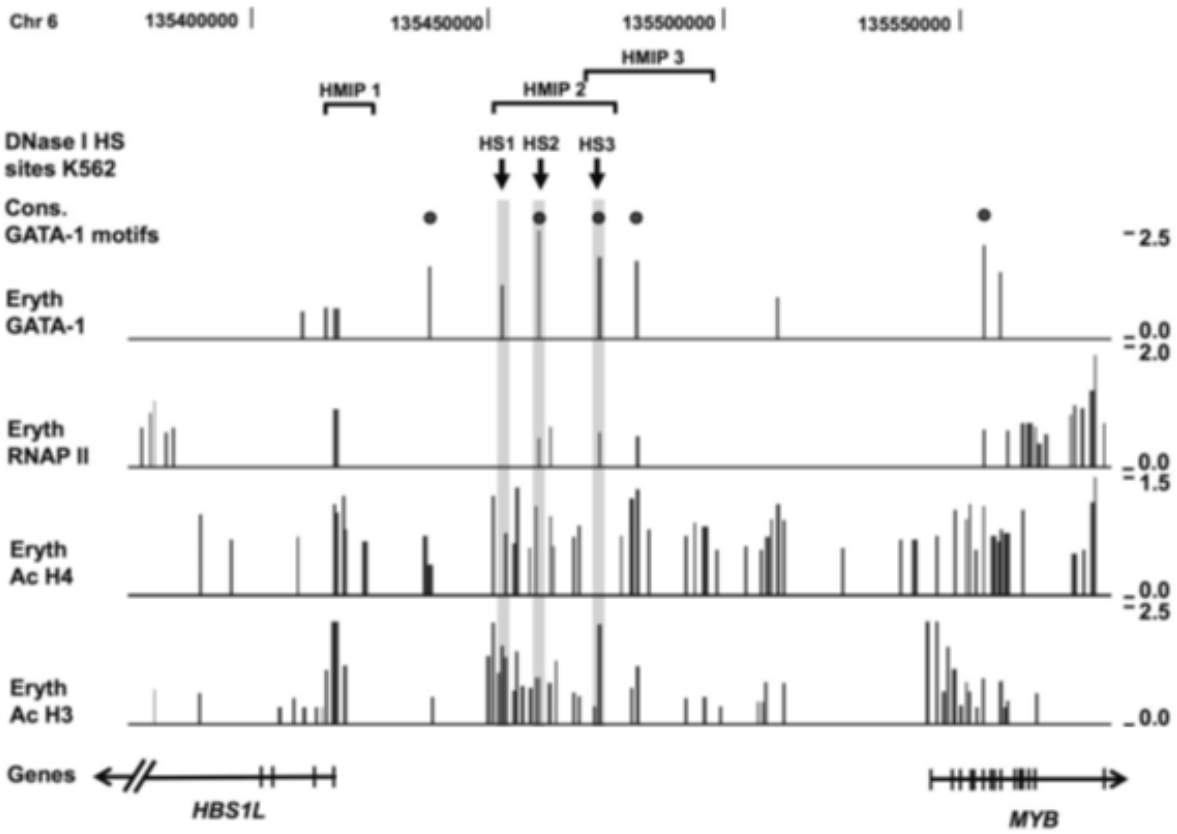


Figure 4.4 ChIP-chip data for the *HBS1L-MYB* intergenic region. Results from ChIP-chip experiments for GATA-1 (M20,sc-1234, Santa Cruz), RNA polymerase II (Upstate), AcH4 (Upstate), and AcH3 (Upstate) in primary human erythroid precursors. Covering 210-kb of 6q locus (position chr6: 135 373 000-135 582 000) including part of the *HBS1L* gene, the intergenic region, and the *MYB* gene. The *HMIP* block 1, 2, and 3 are shown above indicated by the horizontal brackets. Also shown is conserved GATA-1 motifs and DNase I hypersensitive sites (as identified in hemin-induced K562 cells by Wahlberg *et al.* 2009).

The intergenic region between the *HBS1L* and *MYB* genes showed strong signals for AcH3, AcH4, RNA polymerase II and GATA-1, indicating a high level of transcriptional activity in this area (Figure 4.4). AcH3 and AcH4, as well as RNA polymerase II signals were found around the promoters and coding region of the two flanking genes, which are both, highly expressed in human erythroid precursors cells. Generally there was strong AcH3 and AcH4 signal in the intergenic region, which was not seen anywhere else around the 6q segment. This indicates that the region is highly active in erythroid cells.

Strong GATA-1 signals were found around the *HBS1L* and *MYB* gene area, but concentrated in the intergenic region. Three of the six strongest GATA-1 peaks found in *HMIP2* coincide with DNase I hypersensitive sites previously identified in induced K562 cells, as a further indication of the functional regulatory role of this site (Wahlberg *et al.* 2009). Four of the strongest GATA-1 peaks in the intergenic region also coincide with conserved GATA-1 sites.

Our ChIP-chip data strongly supports the hypothesis that the intergenic region between *HBS1L* and *MYB* contains a regulatory element which is likely to control the flanking genes, which we propose, regulate HbF production.

4.2.4 GATA-1 binding at the 6q23 locus in extreme HbF phenotypes

To further investigate this region, I decided to analyse the region in individuals with high and low HbF levels to identify any potential differences in GATA-1 binding between the two genotypes which may help to explain the difference in HbF levels. One individual was homozygous for all the SNPs associated with high HbF in the 6q region (blocks 1, 2 and 3) and was designated +/+, while the other was homozygous for the absence of the high HbF SNPs and was designated -/-.

Primary erythroid cells were cultured from whole blood taken from the two individuals, both of whom were from the same family in which the QTL was first identified. ChIP-chip was carried out as previously described using the same anti-GATA-1 antibody (M-20 sc1234, Santa Cruz). I performed two microarray hybridisations and analysis, with the same material, as technical repeats.

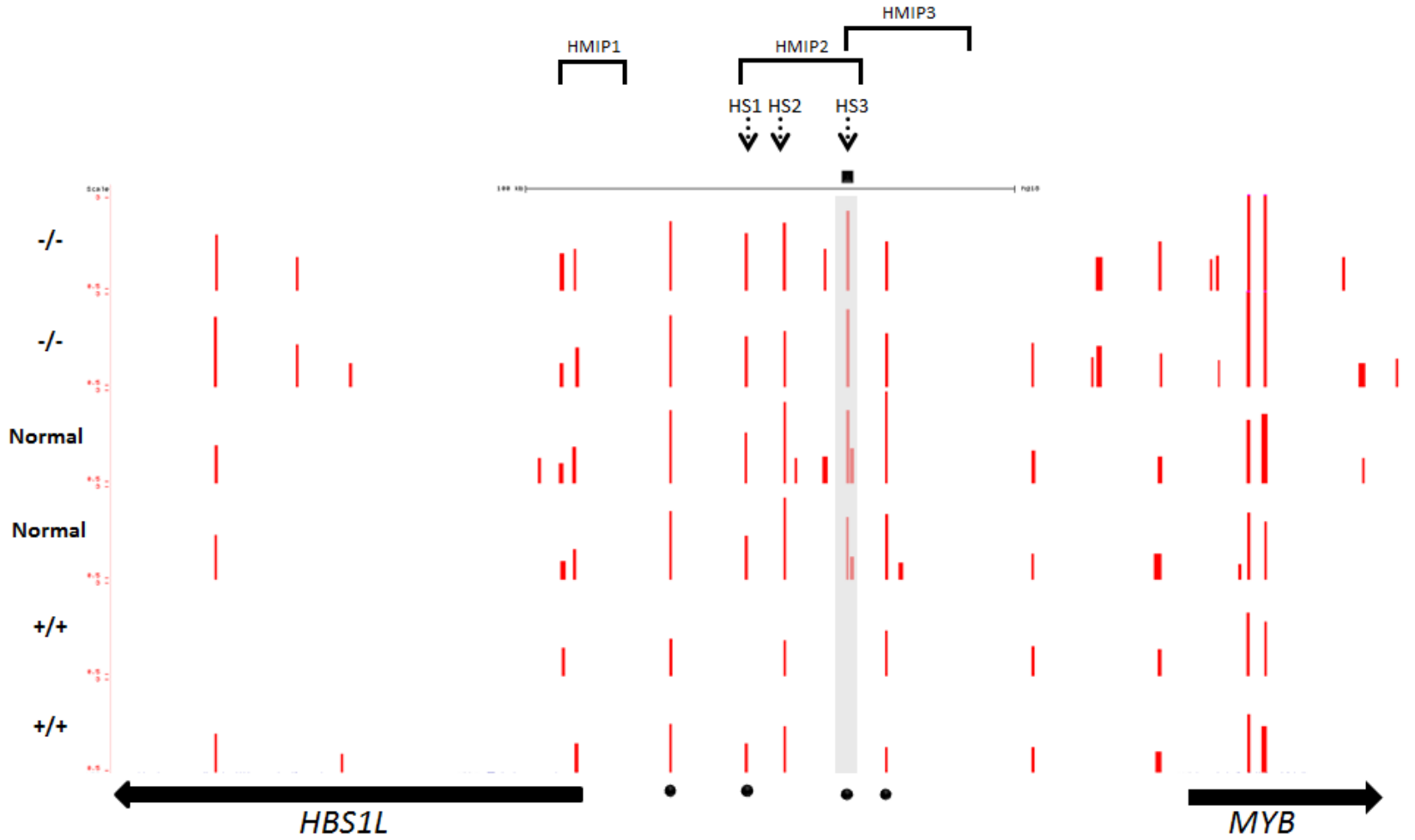


Figure 4.5 GATA-1 binding at the 6q23 QTL in erythroid precursor cells of individuals with high HbF, low HbF and normal levels of HbF, in duplicate (technical repeats on the same samples), by ChIP-on-chip analysis using NimbleGen custom microarrays. Red peaks represent GATA-1 binding, black arrows at the bottom indicate gene location; *HBS1L* and the *MYB* gene including the intergenic region between the genes. *HMIP* blocks 1-3 are shown by brackets above, DNase I hypersensitive sites are represented by broken arrows, and the blue dots at the bottom mark conserved GATA-1 binding sites at this region. Grey bar at the intergenic region indicates GATA-1 binding in the high scoring region *HMIP* block 2, the square below HS3 indicates SNP rs9494142.

GATA-1 signal was seen at the promoter region of the *HBS1L* gene in all individuals and not much signal anywhere else along the gene itself. In contrast, strong signal was found near the 5' end of *MYB* and within the gene itself, which again was consistent in all individuals.

In general, I found considerable GATA-1 binding in the *HMIP* region, particularly in the highly associated *HMIP2* block. The ChIP data in Figure 4.5 highlights an area within the *HBS1L-MYB* intergenic region (grey bar) that showed differential binding; GATA-1 binding is seen in erythroid precursor cells of both individuals with low levels of HbF and with normal levels of HbF, but is completely absent in the individual with high HbF. This site lies within *HMIP2* at HS3, one of the HbF-associated SNPs (rs9494142) also lies within this block. This area also coincides with a conserved GATA-1 binding site. Other than this binding site, all other GATA-1 binding sites in the critical intergenic region were consistent in all individuals.

Based on our earlier work (Figure 4.4) in collaboration with others (Wahlberg *et al.* 2009) and taken with the above data it is evident that this site may well contain a causative variant for raised HbF in these individuals.

4.3 DISCUSSION

Previously, genome wide association studies (GWAS) identified a QTL for raised HbF levels on chromosome 6. Further analysis revealed SNPs at a more precise location on chromosome 6q23, in the intergenic region between the *HBS1L* and *MYB* genes (*HMIP)*. These SNPs lie in 3 linkage disequilibrium blocks, *HMIP 1-3.*

HMIP2 accounts for the strongest effect on FC variance at this locus. Using chromatin immunoprecipitation experiments, I show that this area has significant histone acetylation and RNA polymerase II activity. I also show that several strong GATA-1 binding sites are concentrated within *HMIP2,* three of which coincide with DNase I hypersensitive sites. Such patterns of RNA polymerase II binding, histone acetylation and GATA-1 binding around DNase I hypersensitive sites is seen in known erythroid control regions, suggesting that this intergenic space may also be an active control region.

A profile of GATA-1 binding was performed at the 6q23 locus in two individuals from the same family with phenotypes at the two extreme ends of FC trait and associated with contrasting genotypes. One independent individual with normal HbF levels was used as a control.

This revealed similar GATA-1 binding patterns between the three phenotypes at the 6q intergenic region, except for one anomaly seen at *HMIP*2 which also coincides with a high scoring SNP (rs9494142) associated with HPFH. GATA-1 binds strongly at this site in the individuals with low HbF and normal levels of HbF, however there are no GATA-1 binding signals seen here in the individual

with raised HbF. This suggests that differential GATA-1 binding at the 6q QTL may contribute to the HbF phenotype.

Although these binding signals were confirmed by two technical repeats on one biological sample for each individual (Figure 4.5), these results are not quantitative and should be validated by qPCR. Assays designed to measure the enrichment of target sequences bound by GATA-1 (Figure 4.4 & 4.5) would both quantify and validate this data. Further analysis of this region with more biological samples is also required to confirm these results. This work is currently on-going.

Wahlberg *et al.* (2009) demonstrated that non-erythroid cells (HeLa cells) show substantially less H3 acetylation in the intergenic region, suggesting the transcriptional activity is indicative of a control element and is erythroid specific. HeLa cells do not express *MYB* and the lack of transcriptional activity in the region implies a link between the two, i.e. *MYB* expression and transcriptional activity in the intergenic region.

Following this work, other groups are also interested in characterising the dynamics of the *MYB* locus, since our data implies a distal regulatory region controls the expression of this gene.

Transcription factors including LDB1, GATA-1, FOG-1 and KLF1 are required to maintain chromatin looping to form an active chromatin hub (ACH) at the β-globin locus to initiate transcription (Tolhuis *et al.* 2002, Drissen *et al.* 2004, Vakoc *et al.* 2005, Splinter *et al.* 2006, Song *et al.* 2007). These LDB1 complexes

were found to bind at sites in the *HBS1L-MYB* intergenic region in MEL cells and primary mouse erythroid progenitors (Stadhouders *et al*. 2012). Further experiments showed characteristic features supporting enhancer activity (including the presence of histone acetyl transferase p300) and led to the conclusion that these binding sites suggest active regulatory elements in erythroid progenitors. The group also showed that the ACH destabilises and *MYB* expression decreases on erythroid differentiation in mouse erythroleukaemic (MEL) cells.

Previous studies have shown that *MYB* expression is regulated via transcription elongation through an attenuation site in intron 1 of the gene (Bender *et al.* 1987) and Stadhouders *et al.* showed this region interacts with the regulatory elements found in the intergenic region.

LDB1 and KLF1 were also shown to be essential for up-regulation of *MYB* expression as expression decreased by 50% when both factors were inhibited by shRNA interference in MEL cells (Stadhouders *et al*. 2012).

This study has provided further understanding about the function of the *HBS1L-MYB* intergenic region, although further links of this region with HbF regulation still remain to be found.

Earlier work showed *MYB* and *HBS1L* are down-regulated in individuals with elevated HbF and increased levels of *MYB* inhibit γ-globin gene expression while increased levels of *HBS1L* had no such effect on γ-globin expression. This

suggests *MYB* may be indirectly involved in the regulation of HbF levels (Jiang *et al.*2006).

In a later study we found that high HbF associated SNPs correlated with high *HBS1L* expression (in erythroid cells from a panel of healthy individuals). F-cell levels however, were not significantly correlated with high *HBS1L* expression. When the samples from both studies were combined and adjusted for genotype effects, F-cell levels were again negatively correlated with *HBS1L* expression. These findings revealed that multiple factors contribute to FC trait and *HBS1L* expression. These factors include the SNPs at the 6q intergenic region but are not limited to these (Thein *et al*. 2007).

SNPs in *HMIP* block 2 showing the strongest association with F-cell levels are associated with haematological traits (including reduced erythrocyte count) suggesting that the 6q intergenic region has a pleiotropic effect on haematopoiesis (Menzel *et al.* 2007).

MYB is a critical factor in haematopoiesis; the absence of *MYB,* in *MYB* knockout mice, leads to death *in utero* from anaemia and the down-regulation of erythroid transcription factor GATA-1 (Mucenski *et al.* 1991, Lin *et al.* 1996). Low levels of MYB in primary erythroid progenitor cells of high HbF individuals are associated with accelerated erythroid cell maturation. It was suggested that the relatively low levels of *MYB* may contribute to the HbF levels, indirectly via the regulation of F cells and erythroid kinetics (Jiang *et al.* 2006).

Collectively, these results lead us to propose that the *HBS1L-MYB* intergenic region acts as a distal control element and regulates HbF levels indirectly, perhaps via the flanking genes- *HBS1L* and *MYB.*

CHAPTER 5

5 BCL11A & GATA-1 binding at the globin genes and *BCL11A* loci

5.1 INTRODUCTION

A QTL for HbF at chromosome 2p15 was found by our group in 2007 and maps to a gene; *BCL11A* (Menzel *et al.* 2007). BCL11A is involved in lymphopoiesis and erythropoiesis and has been implicated in myeloid malignancies. More recently, BCL11A was demonstrated to act as a repressor of γ-globin expression (Sankaran *et al.,* 2008).

To investigate the role of BCL11A as a transcription factor, I aimed to study its interaction at the globin loci, *GATA1* locus and the *BCL11A* locus itself. I did this again by ChIP-chip analysis using the same method as described in chapter 3.

As GATA-1 is a key erythroid specific transcription factor, it was included in these ChIP-chip experiments in order to gain a better understanding of how both of these transcription factors influence globin regulation and to what extent they may overlap in function at these loci.

This was the first large-scale analysis of BCL11A carried out and reported.

5.2 RESULTS

5.2.1 BCL11A antibody optimisation

Three ChIP grade BCL11A antibodies were available commercially. In order to determine the most appropriate antibody to use in terms of efficacy and reproducibility of enrichment levels, I tested three antibodies by ChIP and SYBR-Green qPCR.

As a positive BCL11A ChIP enrichment control, I used primers designed to the β-globin LCR HS3 (Appendix 1, Table 11.2) a region which has previously been shown to be bound by BCL11A (Sankaran *et al.* 2008). The enrichment at this site for both BCL11A ChIP and normal IgG ChIP (negative control) was normalized to a reference sequence (*NEFM*) which acted as a housekeeper gene and also normalized to the input sample (DNA before ChIP with antibody).

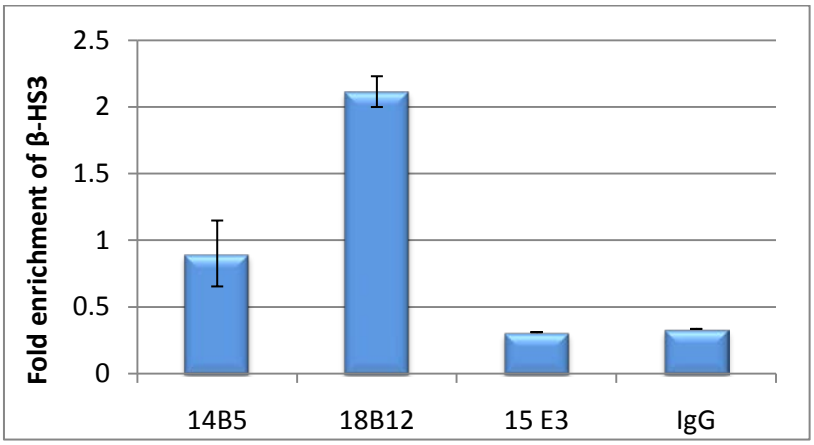


Figure 5.1: Fold enrichment of HS3 on the β-globin gene locus after ChIP by three different anti-*BCL11A* antibodies (14B5 Santa Cruz, 18B12 Santa Cruz, 15E3 Abcam) and anti-mouse IgG antibody (Upstate) as a negative control. Values for 14B5 and 18B12 are means of at least 3 experiments, +/- SE, except 15E3 and IgG data which are both from single experiments, but duplicate reactions.

Of the three antibodies tested in the BCL11A ChIP application (Figure 5.1), antibody 18B12 (Santa Cruz) gave the highest level of enrichment, so this antibody was used in all ChIP experiments for subsequent studies.

5.2.2 BCL11A & GATA-1 binding in the α and β-globin loci

I assessed BCL11A and GATA-1 binding throughout both the α and β globin gene loci. In order to assess the significance of our BCL11A and GATA-1 binding data and to place it in a useful context, we also analysed histone H3 acetylation (a mark associated with more open chromatin with enhancer/promoter activity) as well as RNAPII activity using an antibody which recognises both inactive and actively transcribing polymerase.

I have shown all loci in their orientation as given in the USCS genome browser March 2006 assembly. Within the β-globin locus, BCL11A binding was observed at several locations: at LCR-HS1, -HS2, -HS3 and -HS7, towards the 3' end of *HBE1*, ~3.4 kb downstream of *HBG1* and ~1.5 kb upstream of *HBD*, the latter two of which are in agreement with previous results (Sankaran *et al.* 2008). Binding within the β-LCR was coincident with GATA-1 binding (Figure 5.2) apart from LCR-HS4 where GATA-1 binding alone was found. I was unable to find BCL11A binding in the immediate promoters of any of the globin genes (Figures 5.2 and 5.3), which is in accord with a previous report using human erythroid cells (Sankaran *et al.* 2008).

Binding at the proximal γ-globin promoter has previously been observed in K562 cells using γ-globin promoter deletion constructs linked to a luciferase reporter

(Chen *et al.* 2009) implicating the presence of BCL11A protein. However, I found no expression of BCL11A in our K562 cells (Chapter 6, Fig. 6.1a). The absence of BCL11A in K562 cells would be consistent with the high level of γ-globin expression in K562 cells since *BCL11A* expression correlates inversely with γ-globin levels. I was therefore unable to assess the role of BCL11A in this cell line.

The human β-globin gene cluster is embedded in a larger cluster of olfactory receptor (OR) genes. Coincident BCL11A and GATA-1 binding was also seen at the OR52A1 gene, which is ~75 kb downstream of *HBB* (Fig. 5.2).

The RNAPII data suggests very strong transcription at the β-LCR-HS1, -3 and -4 with the strongest activity seen at the *HBB* gene, as would be expected. There were also pronounced peaks towards the 3' regions of *HBG2* and *HBG1* and *HBD* and around *HBBP1*. However, it is evident that there is a great deal of general intergenic transcription in agreement with previous studies (Gribnau *et al.* 2000, Miles *et al.* 2007).

Histone H3 acetylation in primary erythroid cells is extremely high throughout the LCR with a dramatic dip at the LCR between HS5 and HS6 which is coincident with an LTR (Figure 5.2 black square between HS5 and HS6 marks the location of the LTR promoter region). Acetylation levels at the globin genes are highest in a region peaking around *HBBP1* and then a second distinct region spanning from just 5' of *HBD* through to the end of *HBB* with another dramatic drop beginning at a LINE element 3' to HBB (Figure 5.2 black triangles mark the location of LINE repeats).

Results of GATA-1 and BCL11A binding in the α-globin locus are shown in Figure 5.3. BCL11A binding was entirely restricted to the multispecies conserved sequences (MCS) region of the α-globin locus, with three of the four peaks being completely coincident with GATA-1 signals at HS33, HS 40 and HS 48. These hypersensitive sites are involved in long range control of the α-globin genes in much the same way as the β-globin LCR controls the β-globin genes. The fourth BCL11A binding site, at HS46, showed no GATA-1 binding. The remaining GATA-1 signals occurred at sites in between the ζ-globin (HBZ) and the ψα2 pseudogene (HBM) and at the orthologous region corresponding to the mouse HS-12 element, hoHS-12 (De Gobbi *et al.* 2007).

Although there is substantial overlap between BCL11A and GATA-1 signals, we believe that the signals reflect genuine independent binding rather than cross-reactivity, resulting in immunoprecipitation of bound GATA-1 protein by BCL11A antibody. There is no evidence of cross-detection of GATA-1 using several commercially available BCL11A antibodies in Western blots using whole cell extracts. Further, several very strong GATA-1 binding sites, most notably at the BCL11A locus itself, showed no evidence of BCL11A signal.

Technically, it is not possible to say whether the signals found near the globin genes themselves reflect true initial binding sites for BCL11A/GATA-1 or if they bind at distal enhancers and are brought to these locations by chromatin looping.

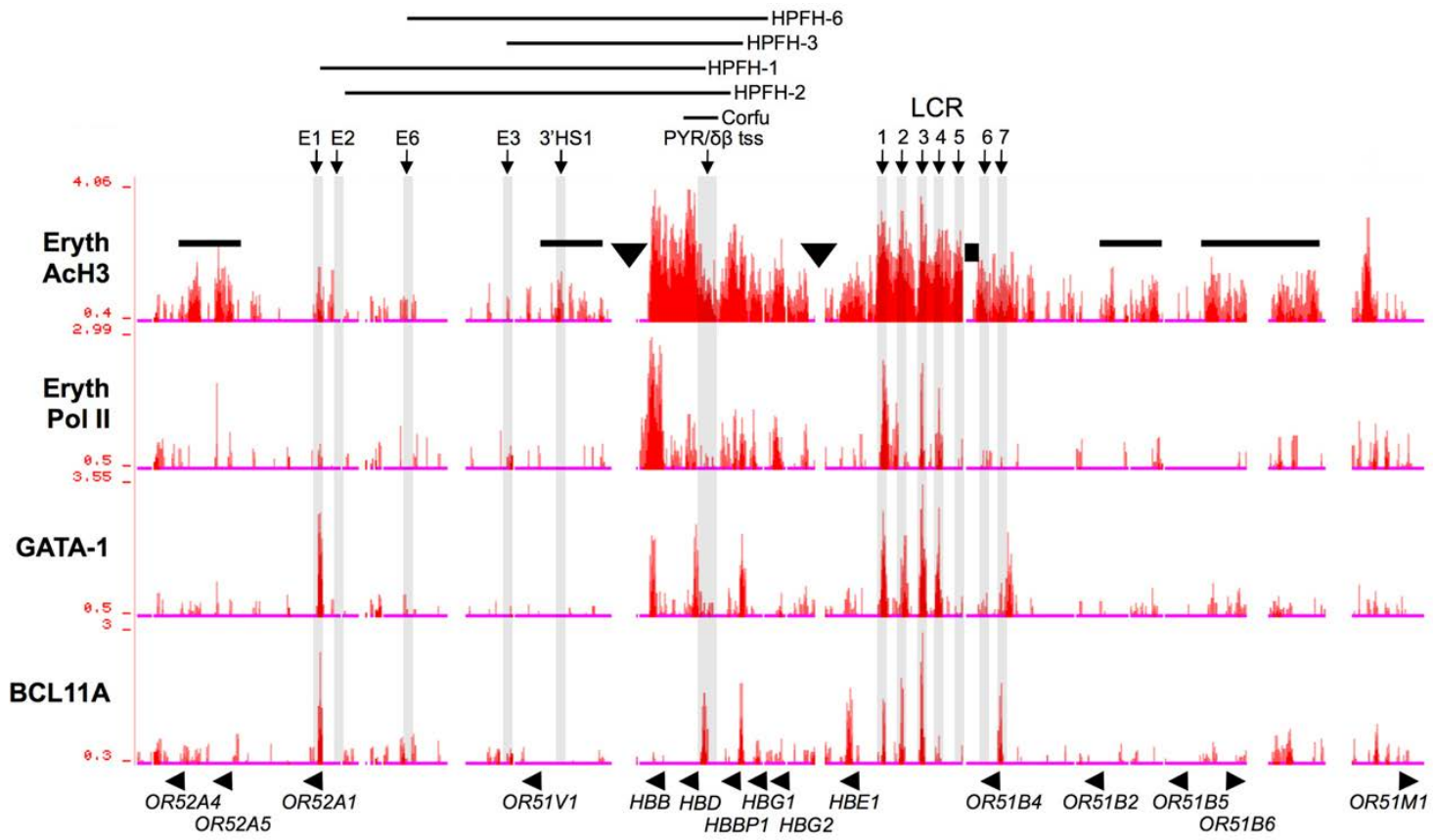


Figure 5.2 ChIP-chip data for the β-globin locus in primary human erythroid precursor cells. Panels show acetylated H3 (AcH3), RNA Polymerase II, GATA-1 and BCL11A binding. Peaks represent scaled log2 ratios of input and ChIP material. All peaks were verified as significant (FDR=0) as determined by the NimbleScan software. Gene locations are shown at the bottom of the figure. HPFH and Corfu deletions (see discussion pg. 14 for references) are shown as thin horizontal lines at the top. Vertical arrows and grey bars indicate the location of HPFH-associated enhancers (E1, E2, E6 and E3), the 3' hypersensitive site HS1, the Corfu PYR/δβ region and the hypersensitive sites (LCR 1–7). Short horizontal bars indicate H3 acetylation outside the immediate β-globin locus. Two triangles mark the location of large LINE repeats. A square between HS5 and HS6 marks the location of the LTR promoter region. AcH3 and RNAPol II ChIP-chip was done in collaboration with fellow PhD student Karin Wahlberg.

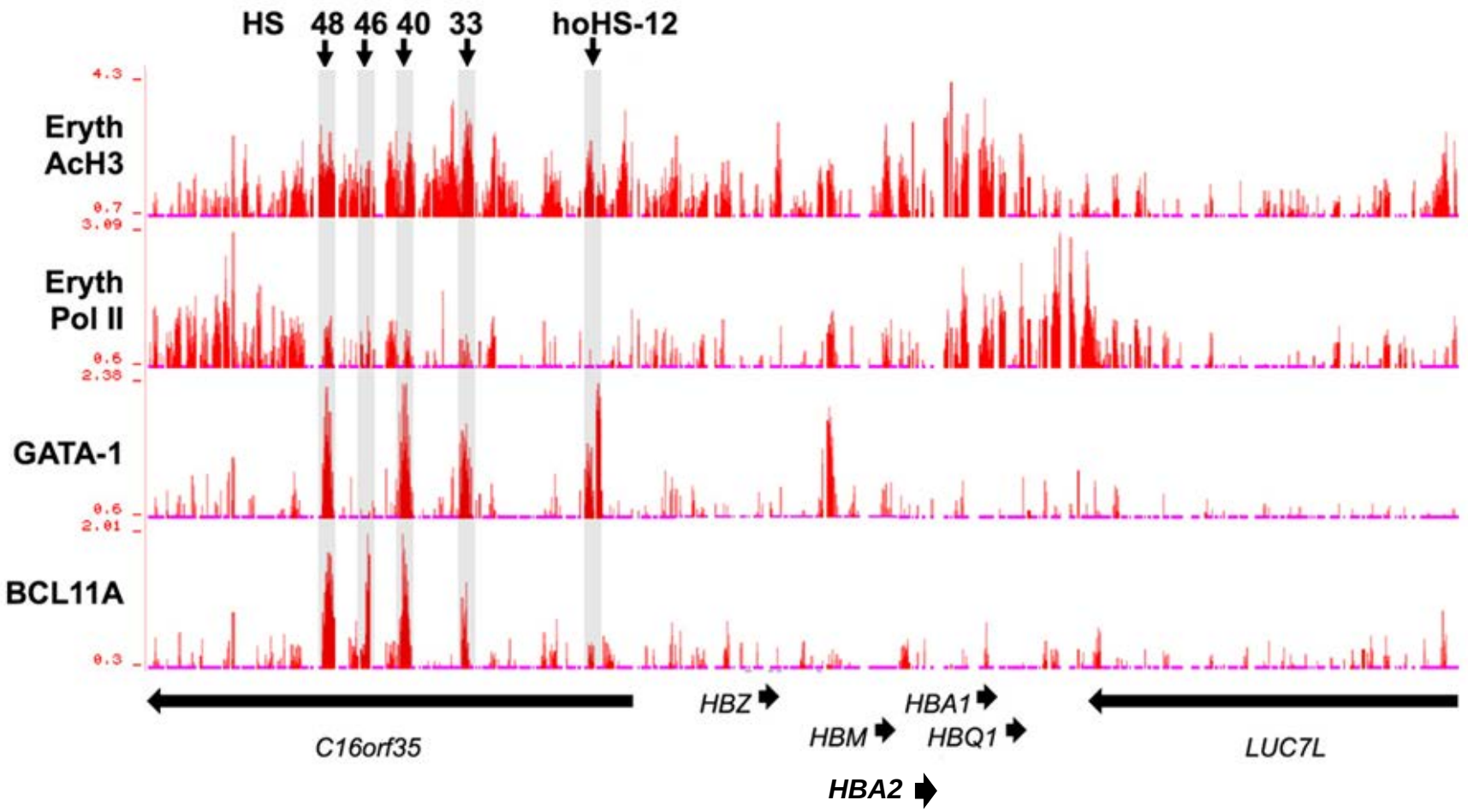


Figure 5.3 ChIP-chip data for the α-globin locus in primary human erythroid precursor cells. Vertical grey bars indicate Multiple Conserved Sites (MCS) upstream of the α-globin genes; HS-48, 46, 40 and 33. hoHS-12 indicates the orthologous region corresponding to the mouse HS-12 element (De Gobbi *et al.* 2007). Gene locations are indicated as black horizontal arrows at the bottom of the figure.

5.2.3 Differential binding of BCL11A and GATA-1 in *BCL11A*, *GATA1* and *HMIP*

Given the strong correlation between GATA-1 and BCL11A binding locations, I decided to investigate differential binding on these two genes themselves. I looked at the *BCL11A* gene itself for both GATA-1 and BCL11A binding activity, as well as for acetylated H3 and RNA polymerase II. We see an unusually large amount of H3 acetylation across the whole of *BCL11A*, in contrast to the neighbouring genes, where only restricted signals were seen (Figure 5.4 A).

Within *BCL11A*, it is clear that the acetylation pattern appears in two peaks; the first coincides with the promoter region and the second with intron 2. RNA polymerase II binding is also very pronounced (Figure 5.4 A), as would be expected from a highly expressed gene. The RNAPII antibody recognises both the active and stalled polymerase complex, but in highly expressed genes, we usually see an accumulation towards the 3' end of the gene as found in previous studies (Wahlberg *et al* 2009) and, once again, we see a peak of binding in intron 2.

GATA-1 binding is clearly observed not only at the promoter of *BCL11A*, but also in intron 2 (Figs. 5.4 A & B), which also coincides with conserved GATA-1 binding motifs (Fig 5.4 B, black dots mark location of conserved GATA-1 sites). ChIP analysis for BCL11A binding showed no activity in or around the *BCL11A* locus, which therefore makes auto-regulation unlikely.

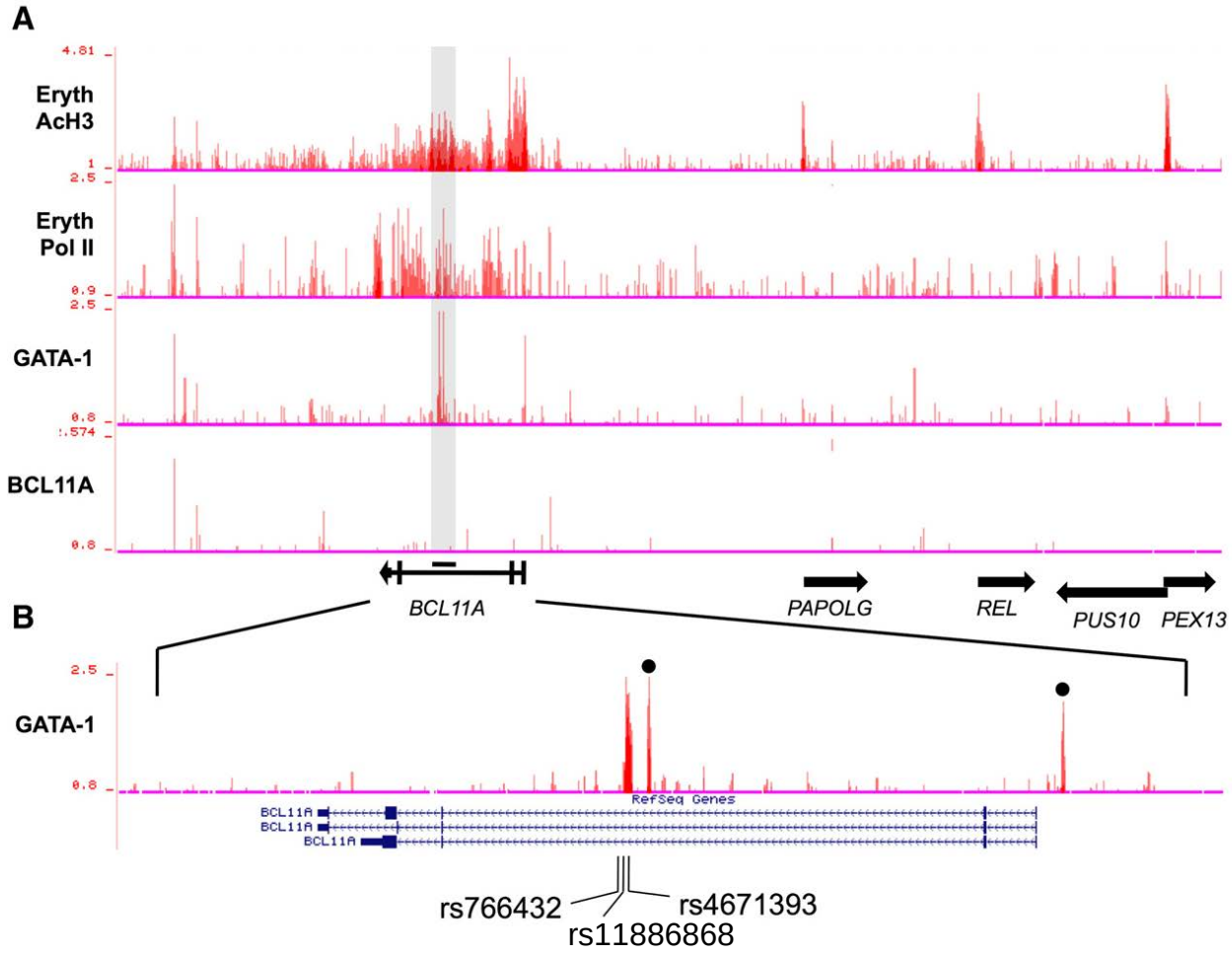


Figure 5.4 ChIP-chip data for the BCL11A region in primary human erythroid precursor cells. (A) Genes are marked with arrows at the bottom of the figure. The intron/exon structure of BCL11A is specifically indicated by vertical bars. The raised HbF-associated region in intron 2 of BCL11A is indicated by a horizontal bar above the gene. (B) A detailed view of the BCL11A gene. Conserved GATA-1 sites (•). Three SNPs are shown, which represent the strongest raised HbF-associated QTL signals found to date.

Several regions in and around the *GATA1* locus showed binding of GATA-1 including the two conserved GATA-1 sites in its promoter (marked by black dots in Figure 5.5). One site is the previously identified positive auto-regulatory promoter sequence which contains a double GATA motif ~700 bp upstream of the first exon (Nicolis *et al.* 1991, Tsai *et al.* 1991). Interestingly, the only binding of BCL11A in the entire region was at this same regulatory site.

BCL11A binding in the 6q23 locus was quite extensive throughout the genes and intergenic region (Figure 5.6). Binding sites were seen at the *MYB* and *HBS1L* promoters and coding regions of the genes. Strong binding signal was observed in the *HMIP* block 2 and most BCL11A binding sites were in coincidence with GATA-1 binding sites, particularly in the intergenic region.

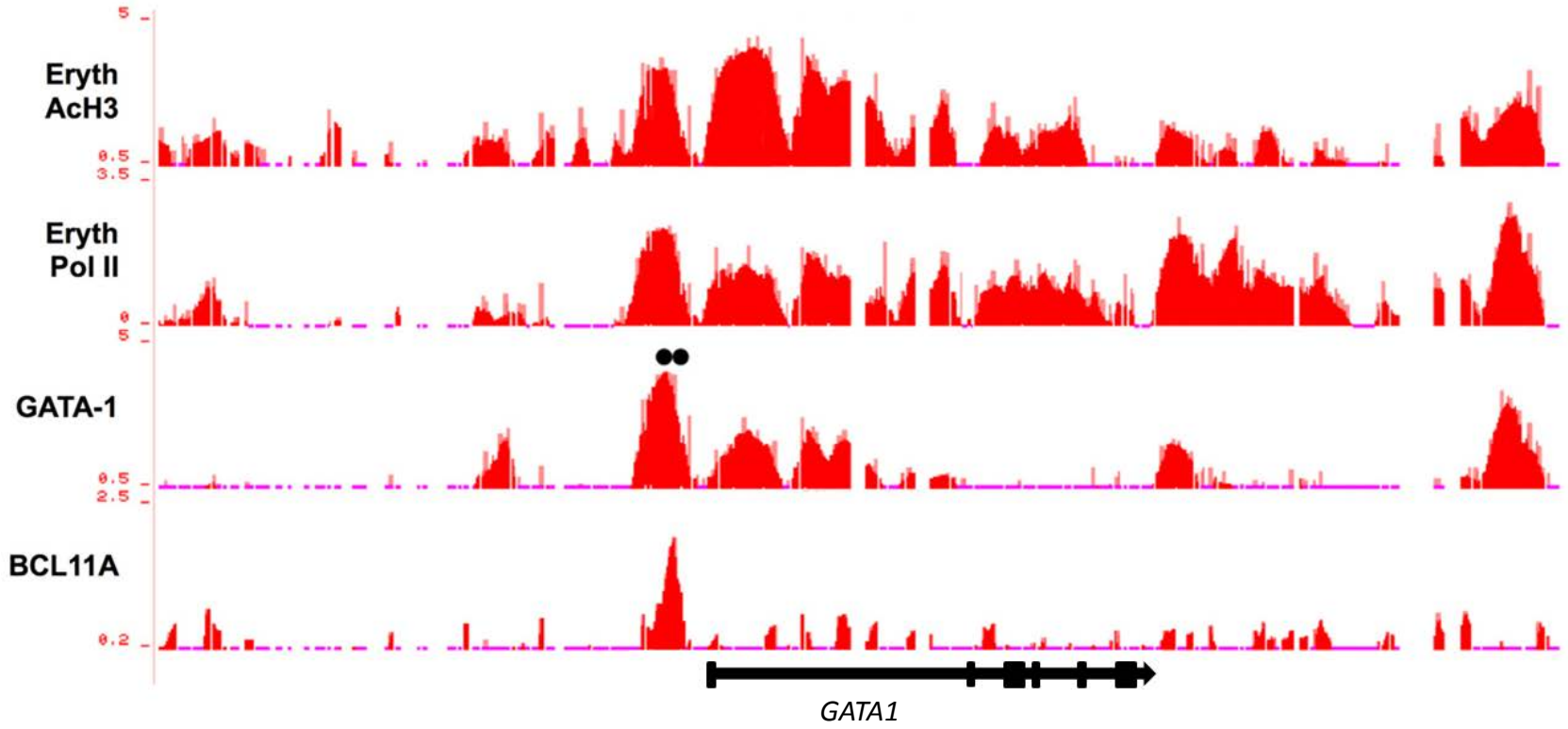


Figure 5.5 ChIP-chip analysis of BCL11A and GATA-1 binding at the *GATA1* locus in primary human erythroid precursor cells. The gene location is indicated below the figure and conserved GATA-1 binding sites are indicated on the GATA-1 panel (•).

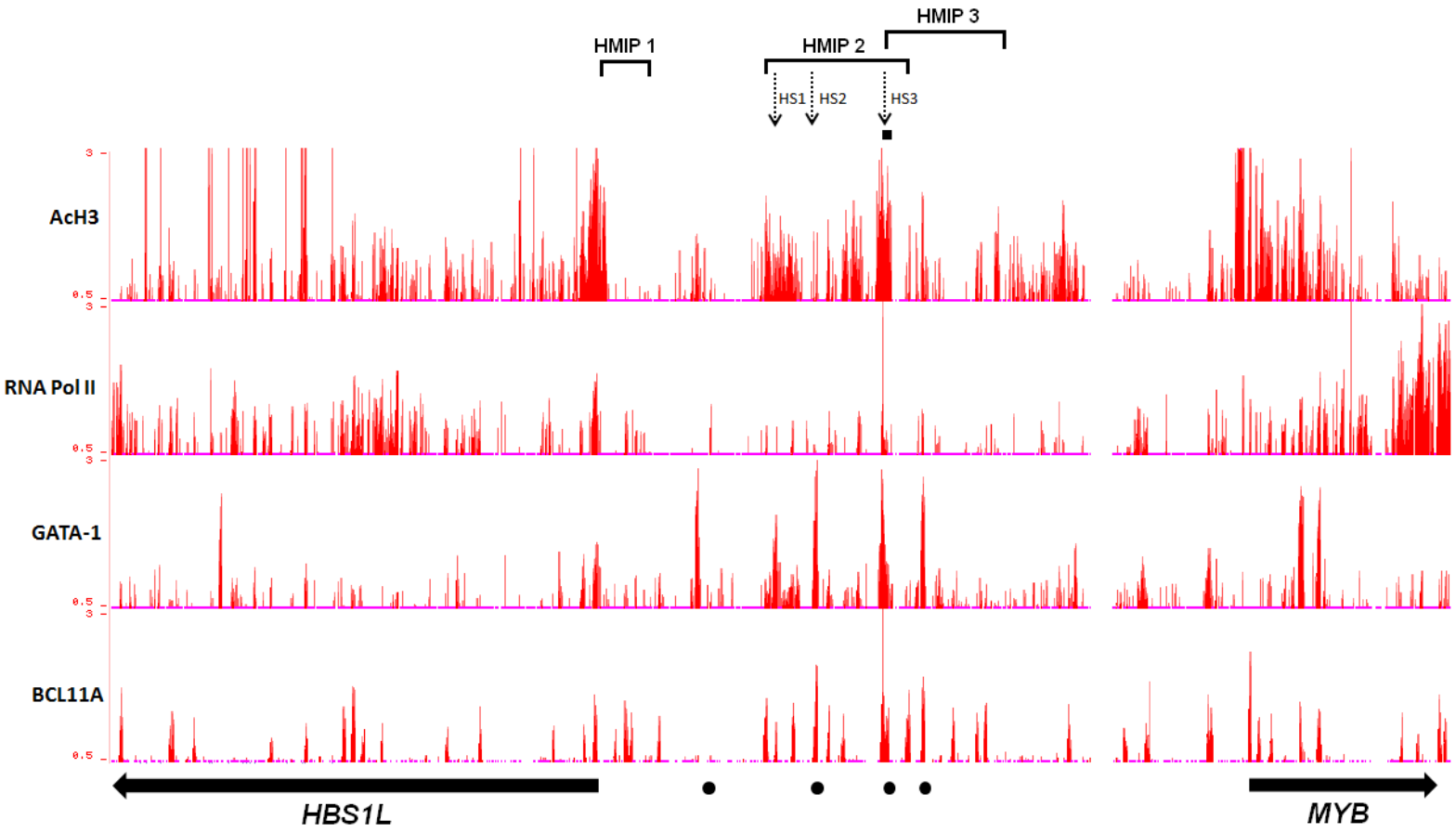


Figure 5.6 ChIP-chip data for the *HMIP* region on chromosome 6q23 in primary human erythroid precursor cells. Tracks show AcH3, RNA polymerase II, GATA-1 and BCL11A binding. Red peaks represent binding, black horizontal arrows at the bottom indicate gene location; *HBS1L* and the *MYB* gene including the intergenic region between the genes. *HMIP* blocks 1-3 are shown by brackets above the tracks, DNase I hypersensitive sites are represented by broken arrows, and the black dots at the bottom mark conserved GATA-1 binding sites at this region (•). The black square below HS3 indicates SNP rs9494142.

5.3 DISCUSSION

I provide here a survey of BCL11A binding in the α- and β-globin, *GATA1* and *BCL11A* loci. I observed strong BCL11A binding, in the β-globin locus particularly at sites previously associated with raised HbF conditions. Amongst the β-globin genes themselves, BCL11A/GATA1 binding was seen ~3.3 kb downstream of HBG1 where deletions are known to cause HPFH (W. G. Wood, 2001).

I also found binding of BCL11A ~2 kb upstream of *HBD*. This area, which is deleted in Corfu δβ thalassemia, is associated with high HbF levels, and is thought to be involved in disruption of γ-globin gene silencing (Wainscoat *et al.* 1985, Chakalova *et al*. 2005). It was suggested that this region, which contains an active transcription start site, marks the boundary of increasing transcription moving towards *HBB* (Gribnau *et al.* 2000, Miles *et al.* 2007). This is supported by our data which shows a pattern of increasing H3 acetylation moving towards *HBB*. Strong GATA-1 binding is also found in the promoter and 5'UTR regions of *HBD*. More recently the region 5' of the δ-globin gene, where we find BCL11A signal, has been elucidated as a region necessary for γ-globin gene silencing (Sankaran *et al.* 2011).

We also find increased levels of erythroid-specific H3 acetylation outside the β locus (black horizontal bars on Figure 5.2) centred around the olfactory receptor genes, indicative of regulatory activity. The *OR51B6* gene region has been shown, by association studies, to contain a possible novel HbF regulator region (Solovieff *et al.* 2010), although this finding has not been replicated. Interestingly, the BCL11A/GATA-1 binding seen at the *OR52A1* gene, ~75 kb 3' of *HBB*, is a known 3' enhancer region which was originally discovered via the HPFH-1 deletion

(Feingold & Forget 1989, Katsantoni *et al.* 2003). The HPFH deletions and their associated enhancer elements are shown in Figure 5.2. The HPFH-2, 3 and 6 enhancers showed no GATA-1 or BCL11A binding.

There were striking BCL11A ChIP signals at the α-globin locus (Figure 5.3), with strong BCL11A/GATA1 binding observed at the human multispecies conserved sequences (MCS) which correspond to four DNaseI hypersensitive sites (HS-48, HS-40, HS-33, HS-10) (Hughes *et al.* 2005) but again, there was no BCL11A binding in any immediate promoter regions. It has recently been shown that chromosome looping at the human α-globin locus is mediated via HS-40, confirming its role as the major enhancer of α-globin gene expression (Vernimmen *et al.* 2009). The importance of the other elements is not yet clear. However, they have all been shown to bind GATA-1 as part of a pentameric erythroid complex (De Gobbi *et al.* 2007). The fact that BCL11A binds at both α and β regulatory regions might be evidence of co-regulation of the α and β globins for maintenance of a balanced expression.

It would appear from my results that BCL11A is not always required for GATA-1 binding, as shown by some GATA-1 signals without BCL11A binding. To understand this will require experimental approaches similar to ones used to demonstrate that FOG-1 interaction with GATA-1 is required for GATA-1 occupancy at select sites, such as HS2, but is not required at others, for example at the β globin HS3 site and at the GATA-1 target gene *KLF1* (Letting *et al.* 2004.

Co-immunoprecipitation experiments in erythroid cells have shown that BCL11A is associated with known components of the repressive NuRD chromatin-remodelling complex, along with GATA-1 (Sankaran *et al.* 2008). Previously,

binding of a GATA-1/FOG-1/NuRD complex has been shown to silence hematopoietic genes (Rodriguez *et al.* 2005, Harju-Baker *et al.* 2008). I also observed BCL11A binding in the absence of GATA-1. Given that BCL11A is expressed widely, including cells which do not express GATA-1, it is assumed that such binding of BCL11A may be related to its other functions (Liu *et al.* 2006).

NuRD has also been shown to interact with other transcription factors with repressive activities, for example Ikaros (IKZF1) (Kim *et al.* 1999, Georgopoulos *et al.* 1994) which is required for normal T- and B-cell development, an interesting parallel with BCL11A, which is essential for normal lymphoid development (Liu *et al.* 2003). IKZF1-null mice carrying the human β-globin locus show delayed switching between human γ- and β-globin (Lopez *et al.* 2002). A pyrimidine-rich region, also known as the PYR element (Figure 5.2), upstream of δ globin, was identified as a target for an IKZF1-containing chromatin-remodelling complex and was found to be involved in γ- to β-globin switching. It contained both activator (SWI/SNF) and NuRD repressor components (O'Neill *et al.* 1999 & 2000).

This region is removed in the Corfu δβ thalassemia deletion and, if deleted in transgenic mice, results in a delayed γ to β switching. IKZF1 has recently been shown to interact with GATA-1 as part of a γ-globin gene silencing NuRD complex (Bottardi *et al.* 2009). My discovery of BCL11A binding at the Corfu deletion site, between the δβ transcription start site and the PYR element, could indicate that it may also be part of a repressive NuRD complex in adult erythroid cells.

Our results for AcH3 and RNAPII activity and GATA-1 binding in the *BCL11A* locus, all point towards a potentially important regulatory region in intron 2 coincident with previous genetic studies which identified the same region as highly

associated with HbF production (Menzel *et al.* 2007, Lettre *et al.* 2008, Uda *et al.* 2008, Sedgewick *et al.* 2008). The mechanism as to how this sequence regulates *BCL11A* expression remains to be determined, but this region could be a useful site of investigation for the manipulation of BCL11A levels for therapeutic reactivation of HbF.

BCL11A showed no binding to its own gene but strong signals were observed in the *GATA1* gene. *GATA1* has been shown to undergo positive auto-regulation via a paired GATA-1 motif in its promoter (Nicolis *et al.* 1991, Tsai *et al* 1991). I confirmed GATA-1 binding at this site but also show co-localised BCL11A binding. No BCL11A binding was found at any other point within the immediate region. Although BCL11A is thought to be part of a repressor complex, this GATA-1/BCL11A co-localisation might be evidence of a distinct activator complex. Alternatively, there might be competition between a BCL11A repressor complex and a GATA-1-containing activator complex which curbs *GATA1* positive auto-regulation. This activator-complex theory could also be true for the *HMIP* intergenic region where we see a lot of BCL11A and GATA-1 binding in unison.

I have been able to provide further evidence that BCL11A is involved in the transcriptional regulation of α- and β-globins and may also regulate and be regulated by GATA-1 as part of a distinct activator or repressor protein complex, in a way similar to that of FOG-1 and IKZF1. Moreover, I have shown for the first time that intron 2 of *BCL11A,* an area highly associated with HbF is particularly active transcriptionally and is likely to represent a regulatory region.

CHAPTER 6

6 KLF1 at the globin loci and *BCL11A* locus

6.1 Introduction

KLF1 was first discovered as a protein specific to erythroid cells. The CACCC sequence to which it binds is found in the enhancers and promoters of many erythroid specific genes, most notably the β-globin gene promoter (Miller & Bieker, 1993). Studies that followed its discovery revealed that KLF1 binds to this promoter with greater affinity than to any other promoter (Donze *et al.* 1995).

KLF1 also has a direct role in activating hypersensitive site 3 in the β-globin LCR (Tewari *et al.* 1998, Gilemans *et al.* 1998) and studies in KLF1$^{-/-}$ human β-locus transgenic mice showed a complete lack of β-globin gene expression (Wijgerde *et al.* 1996). These findings suggested that KLF1 has an important role in the globin switch.

More recently, KLF1 has been found to have a direct role in *BCL11A* activation, through variable HbF levels detected in families with *KLF1* mutations who were subsequently found to have altered *BCL11A* expression (Borg *et al.* 2010 Zhou *et al*. 2010).

To further understand the regulation of the *BCL11A* gene I wanted to investigate the role of KLF1 at this locus in a more extensive way. I performed ChIP-chip experiments using an anti-KLF1 antibody, in the same way as I investigated BCL11A and GATA-1 binding in the previous chapters.

6.2 RESULTS

6.2.1 KLF1 at the *BCL11A* locus

At the time of undertaking these experiments, there were no commercially available ChIP grade anti-KLF1 antibodies available. ChIP tested anti-KLF1 antibody (FRACT 5) was kindly donated by Dr. Sjaak Philipsen from the Erasmus Medical Centre, Netherlands. The same antibody had been used successfully in studies by the Philipsen group (Borg *et al.* 2010).

Chromatin from primary human erythroid cells was immunoprecipitated with anti-KLF1 (test) and anti-IgG (negative control) antibodies and the attached DNA extracted, purified and tested. The *HBB* promoter is a known KLF1 target and so primers specific to this site were used in the qPCR assays to assess the efficiency of the ChIP experiment by measuring enrichment of this positive control site.

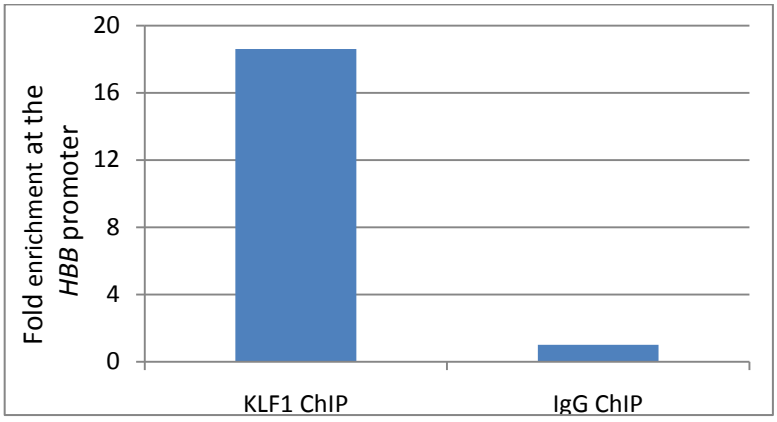


Figure 6.1 Fold enrichment of the *HBB* promoter, a positive control site for KLF1 binding. Expression was measured in material immunoprecipitated with KLF1 and normal IgG as a negative control for ChIP. Data is mean of duplicate PCR reactions.

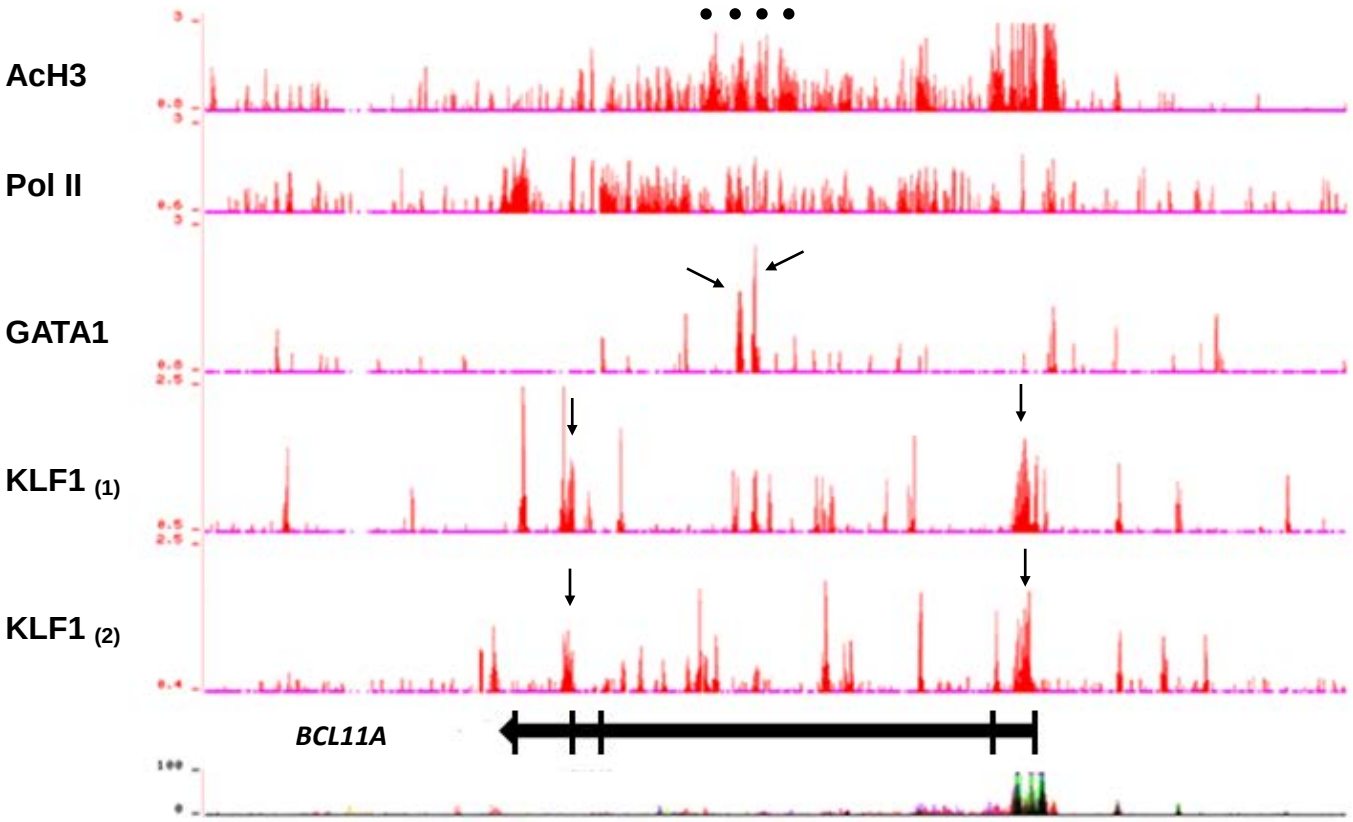


Figure 6.2 ChIP-chip data on the *BCL11A* locus in erythroid precursor cells. Showing GATA-1, acetylated-H3 (AcH3) and RNA polymerase II (Pol II), and KLF1 binding (duplicate ChIP-chip experiments). Red peaks represent binding; black horizontal arrow below peaks indicates gene location. The track below this indicates promoter associated signal in the area. The black vertical arrows indicate consistent and reproducible binding sites for GATA1 and KLF1. Black dots above the first track indicate a cluster of 4 strong AcH3 signals at the QTL. AcH3 and RNA PolII ChIP-chip experiments were done in collaboration with another PhD student at the time, Dr. Karin Wahlberg.

KLF1 binding was found at the *BCL11A* locus, mostly restricted to the gene itself. The strongest and most consistent peaks were at exon/intron 1 of the gene which coincides with the promoter associated signal from the ENCODE database, strengthening the accuracy of my ChIP-chip data. The second strongest peak was found at exon 4.

Some signal was found at the critical intron 2 interval, but this was not reproducible in the second independent ChIP-chip experiment and therefore we considered it less important than the other two stronger and more consistent signals (marked out by arrows on the KLF1 track in Figure 6.2).

In figure 6.2 I also included ChIP-chip data for AcH3, RNA Polymerase II and GATA-1 at the *BCL11A* locus (which was previously identified in chapter 4) in combination with the KLF1 data to put in context of general transcriptional activity at the locus. AcH3 and RNA Pol II was seen within the gene and abundantly in intron 2, close to the high scoring SNPs, and also where I found strong GATA-1 binding.

To validate the data obtained from the ChIP-chip experiments I repeated ChIP with the same anti-KLF1 antibody in human erythroid precursor cells and measured the enrichment of the most consistent sites by SYBR-Green qPCR using primers designed to the binding sites (peaks 1 & 2 in figure 6.3b) (Appendix 1, Table 11.6 for primers).

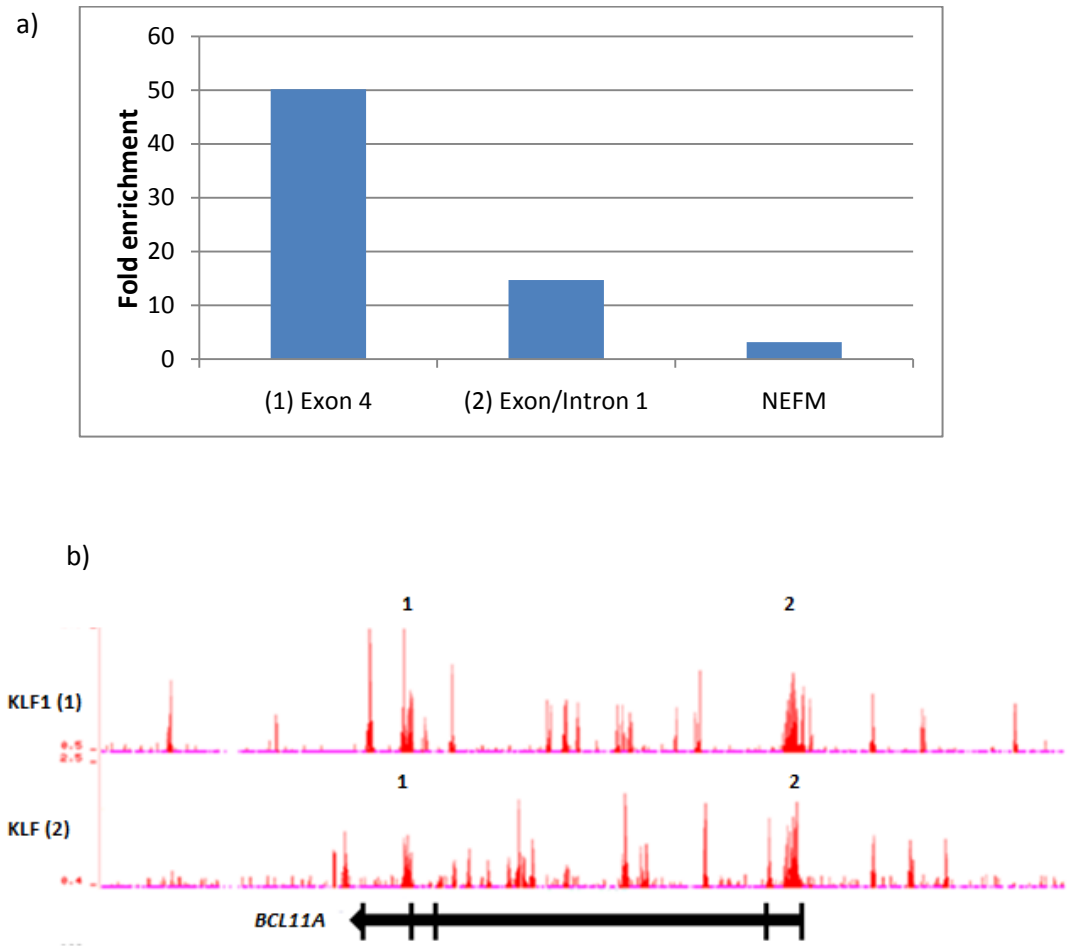


Figure 6.3 a) Fold enrichment of two of the consistent KLF1 sites on the *BCL11A* locus (labelled 1 and 2) after ChIP with anti-KLF1 antibody (FRACT 5, Dr. Sjaak Philipsen's laboratory, Erasmus, Netherlands) and *NEFM* housekeeper gene as a negative control and normalizer. Enrichment determined by SYBR-Green qPCR and analysed by normalizing to *NEFM* housekeeper gene. Values are mean fold enrichment of at least 3 PCR reactions. b) Corresponding KLF1 binding sites on the *BCL11A* locus found by ChIP-chip and measured for enrichment shown in a).

Both sites showed more than 10 fold enrichment when normalised to the housekeeping gene (*NEFM*), which also acted as the negative control. This result confirmed that the signals calculated from the ChIP-chip experiments are true signal for the binding of KLF1 at those specific sites.

6.2.2 KLF1 at the α- and β-globin locus

The β-globin gene (HBB) promoter contains a CACCC sequence which is a binding motif for KLF1. I decided to measure KLF1 activity at the β-globin locus because a) I wanted to measure KLF1 binding at a known binding site (the *HBB* promoter) to test the efficiency of the ChIP-chip experiments, and b) I wanted to see a more extensive view of KLF1 binding at the locus and nearby LCR.

So far there is very little literature on the role of KLF1 at other erythroid specific genes other than the *HBB* gene. Therefore, in addition I decided to assess KLF1 binding at the α-globin locus. I did this again using the same protocol and antibody previously used.

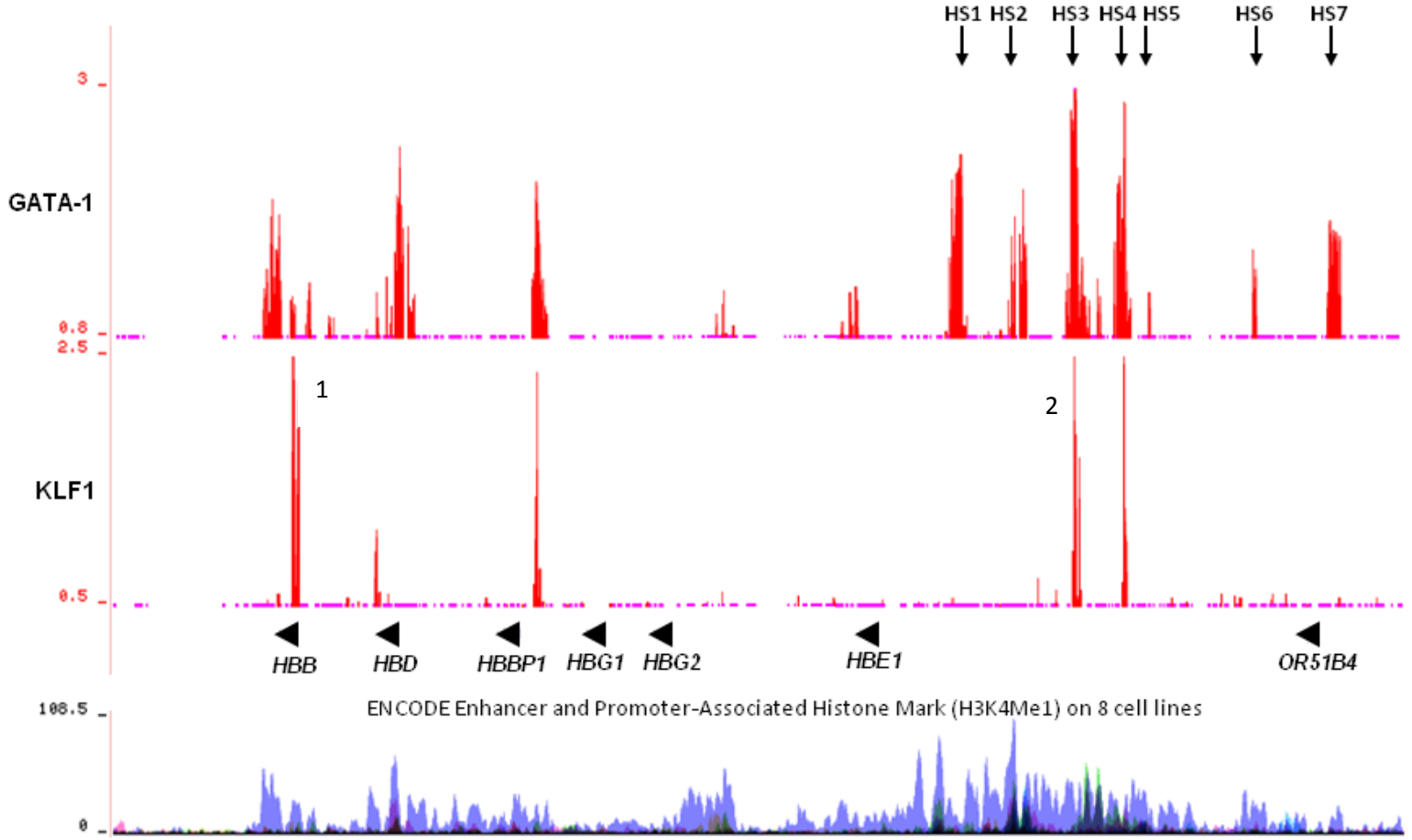


Figure 6.4 KLF1 and GATA-1 binding in the β-globin gene locus on chromosome 11 in primary human erythroid precursor cells by ChIP-on-chip analysis. Red peaks represent binding; black horizontal arrows below tracks indicate gene location. The hypersensitive sites of the β-LCR are marked by black vertical arrows above the tracks. The final track indicates enhancer/promoter associated signal in the area.

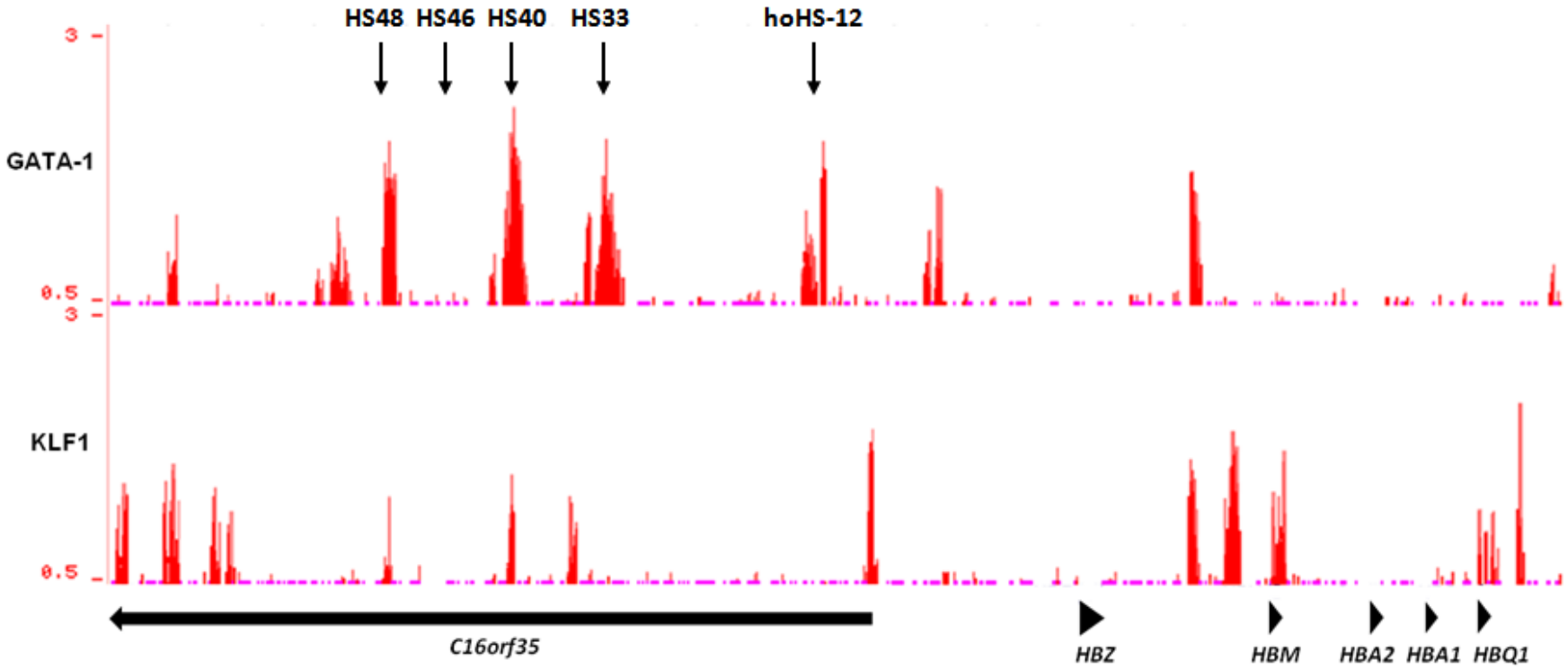
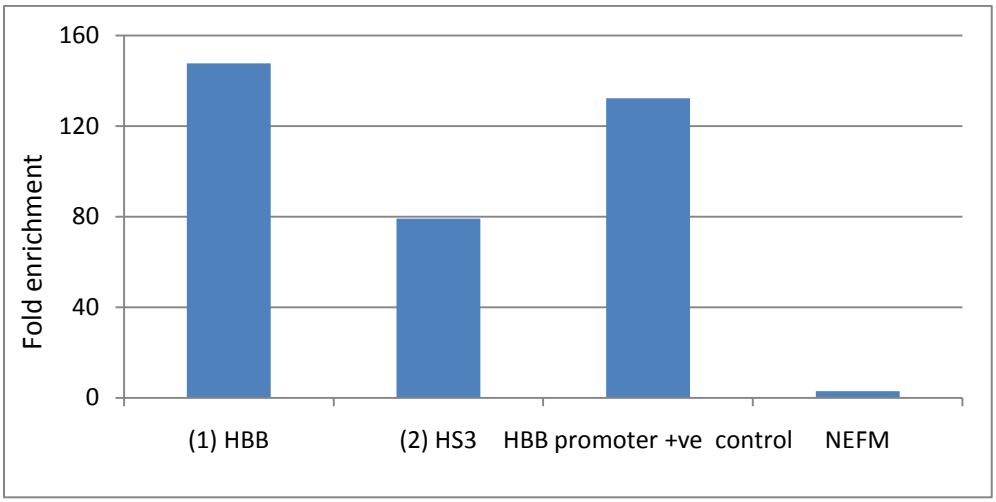


Figure 6.5 KLF1 and GATA-1 binding in the α-globin locus on chromosome 16 in primary human erythroid precursor cells by ChIP-on-chip analysis. Red peaks represent binding; black horizontal arrows below the tracks indicate gene location. The hypersensitive sites; HS-48, 46, 40 and 33, including the orthologous mouse hoHS-12 element are marked by black vertical arrows above the tracks.

a)



b)

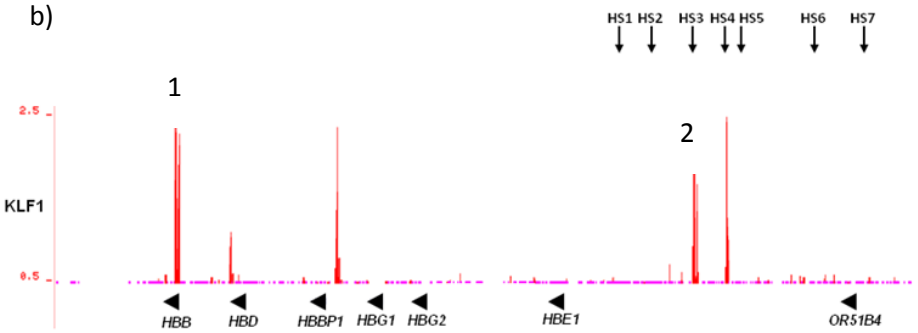


Figure 6.6 a) Fold enrichment of KLF1 binding sites 1 and 2 as indicated in figure 6.4. Enrichment was measured by SYBR-green qPCR using specific primers. The *HBB* promoter served as a positive control. All values were normalized to the *NEFM* housekeeping gene which was also used as a negative control for enrichment. Values are mean fold enrichment of at least 3 PCR reactions. b) Corresponding KLF1 binding sites on the β-globin locus found by ChIP-chip and measured for enrichment as shown in a).

I viewed GATA-1 with KLF1 binding signal at the β-globin locus (Figure 6.4). As expected, I found KLF1 binding at the *HBB* promoter, a known KLF-1 target (Miller & Bieker 1993). In addition, two strong KLF1 signals are seen at the β-globin LCR, one of which was at HS3. These KLF1 binding sites also coincide with GATA-1 binding sites at the LCR, where there is also considerable enhancer/promoter activity as indicated by the enhancer and promoter associated histone track in Figure 5.3.

Another KLF-1 binding signal was found between *HBBP* and *HBG1* where another GATA-1 binding site is also found.

I proceeded to confirm the findings by repeating ChIP of primary human erythroid cells with the same anti-KLF1 antibody followed by SYBR-green qPCR. To check enrichment as a positive control, primers were designed to peak 1 found at the *HBB* promoter (Figure 6.4) and peak 2 in the β-LCR (HS3). As a further check, I measured enrichment of the *HBB* promoter using previously published primers, designed specifically for the KLF1 CACCC sequence (Figure 6.6). This site is a direct target of KLF1, and had been used as a positive control for KLF1 in earlier ChIP experiments (Figure 6.1). *NEFM* was used as the normalizer and negative control.

Significant KLF1 binding sites were also found in the α-globin gene cluster and some discrete sites at the hypersensitive sites which are involved in long-range control of the α-globin genes (Figure 6.5). KLF1 signal between *HBZ* and *HBM* is in coincidence with GATA-1 signal as well as high levels of AcH3 and RNA Pol II at this region (Figure 5.3 for AcH3 and Pol II data). KLF1 also occupies sites at the *HBM* gene and promoter, with further signal detected at the *HBQ1* gene and downstream of this gene. The KLF1 binding sites found at the regulatory element of the locus marked by hypersensitive sites appears at HS48, 46 and around HS33, which are also occupied by GATA-1. AcH3 and RNA Pol II activity is also particularly high at the hypersensitive sites, as expected (Figure 5.3 for AcH3 and Pol II data).

6.2.3 KLF1 at the *HMIP* locus

Earlier I described GATA-1 binding in the *HMIP* region which is strongly associated with F-cell levels. I now measured KLF1 binding in this region to investigate the role of KLF1 in the intergenic region and on the *MYB* and *HBS1L* genes.

I found very little KLF1 binding in this region (Figure 6.7). Some signal was found at the *HBS1L* promoter which coincides with GATA-1 and BCL11A signal. Stronger signal was found at the *MYB* gene promoter which was in coincidence with BCL11A and a further signal was also seen within the *MYB* gene itself.

Apart from one binding signal near the 5' end of *HBS1L,* no KLF1 signal was found in the intergenic region.

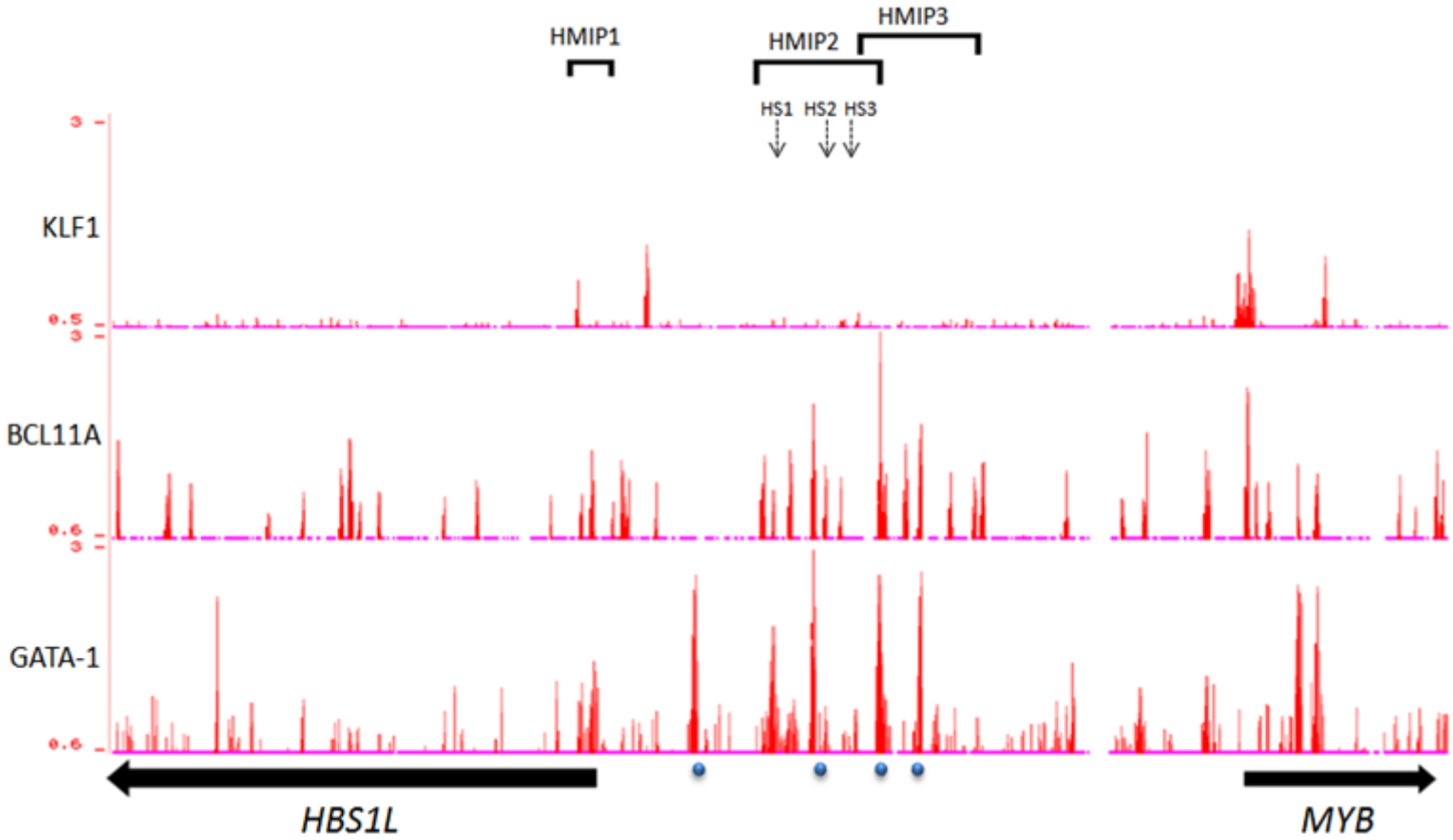


Figure 6.7 KLF1, BCL11A and GATA-1 binding at the 6q23 *HMIP* locus in primary human erythroid precursor cells by ChIP-on-chip analysis. Red peaks represent binding, black horizontal arrows at the bottom indicate gene location; *HBS1L* and the *MYB* gene including the intergenic region between the genes. *HMIP* blocks 1-3 are shown by brackets above the tracks, DNase I hypersensitive sites are represented by broken arrows, and the blue dots at the bottom mark conserved GATA-1 binding sites at this region.

6.3 DISCUSSION

KLF1 signal at *BCL11A* suggests that KLF1 has a regulatory role on the gene. Binding is found to be strongest at intron/exon 1 which is a possible enhancer of the gene as indicated by acetylated-histone H3 in this area (Figure 6.2). Strong binding is also seen at exon 4; which is the XL/L isoform specific exon and evidence suggests that the large isoforms are most abundantly expressed in primary adult human erythroid cells (Sankaran, 2008). KLF1 signal at this particular exon suggests that KLF1 may be involved in regulating transcription of specific *BCL11A* isoforms.

KLF1 binding is also seen at the *HBB* promoter consistent with other studies (Zhou *et al.* 2010). KLF1 signal seen in the β-LCR demonstrates the regulatory role of KLF1 on the β-globin gene. I found that KLF1 binds to the two regions of the β-globin locus as does BCL11A (see Chapter 5, Figure 5.2). In fact, BCL11A, GATA-1 and KLF1 binding coincides at these two positions; between the *HBG1* and *HBBP* genes, a site where deletions are known to cause HPFH (Bank 2006). Binding also coincides in the LCR at HS3 which is part of the regulatory element of the locus which is responsible for activation of the downstream β-globin genes.

It has already been suggested that BCL11A and GATA-1 may be working together in a NuRD complex (Chapter 5.3). Considering the KLF1 binding signals in coincidence with BCL11A and GATA-1, it is possible KLF1 may also be part of the same transcription repressing NuRD complex. This complements the latest model

of β-globin gene control suggesting KLF1 regulates BCL11A, resulting in the repression of γ-globin.

KLF-1 signal also coincides with GATA-1 signal at the α-globin locus within the gene cluster as well as the 5' regulatory elements of the locus, showing a more global role of KLF-1 within the globin genes (Figure 6.5).

KLF-1 may have an effect on the *MYB* gene at the 6q locus (Figure 6.7) as I found binding signals at the *MYB* gene promoter, but seems to show no interaction in the intergenic region.

Soon after our findings, others showed, through mouse studies, that KLF1 binds to a -81kb enhancer which contains a conserved CACCC-box motif, upstream of the MYB transcription start site on the *HBS1L-MYB* locus and is believed to activate *MYB* transcription (Stadhouders *et al.* 2011). However, I did not observe this binding site in my KLF1 ChIP-chip data in human erythroid precursor cells.

Other transcription factors have been shown to regulate *MYB* expression including GATA-1, Ets-1 and MYB itself (Bartunek *et al*. 2003, Sullivan *et al*. 1997, Nicolaides *et al*. 1991). It is possible that KLF-1 is another transcription factor that contributes to the overall expression of *MYB*.

Studies have shown that *MYB* introns contain regulatory sequences which are an important part of MYB control (Bender *et al.* 1987) which include enhancer elements in intron 1 and 4 (Dooley *et al.* 1996, Seib et al. 1994).

KLF1 is a critical regulator of globin gene expression and my research so far supports this role. The KLF1 binding profile presented in this chapter however, is

limited to primary human erythroid precursor cells and this binding activity may be erythroid specific. KLF1 ChIP experiments in other cell types or cell lines would help distinguish the specificity of KLF1.

CHAPTER 7

7 A study of *BCL11A* isoforms in erythroid cells and myeloid leukaemia

7.1 INTRODUCTION

A QTL on the *BCL11A* gene accounts for 15% of the variance in HbF levels in the healthy European population (Menzel 2007). Since the discovery of this QTL, much work has been undertaken to elucidate its role in HbF regulation and the globin switch. *BCL11A* was initially isolated by retro-viral tagging studies to identify leukaemia disease genes in mice (Nakamura *et al.* 1996). 10 years prior to these findings, two rare cases for chronic lymphocytic leukaemia were studied by Fell *et al.* (1986) and were found to have a chromosomal translocation involving part of what we now know as the *BCL11A* gene [t(2;14)(p13;q32)].

Until the identification of *BCL11A* as a QTL for HbF control, it was not even known to be expressed in erythroid cells. In a more recent study, raised expression levels of *BCL11A* were detected in some JMML cases and a subset of AML patients with monosomy 7 (MO7) with poor disease outcome. Monosomy 7 is also a common feature of JMML (Yin *et al.* 2008).

BCL11A exists as 3 major isoforms; XL, L and S. The distribution, function and purpose of these isoforms is as yet unclear. We wanted to identify which isoform(s) is present in erythroid cells and if the same or different isoform(s) is present in myeloid leukaemia cells, as this would give an indication as to the roles and function of the different isoforms in different cell types.

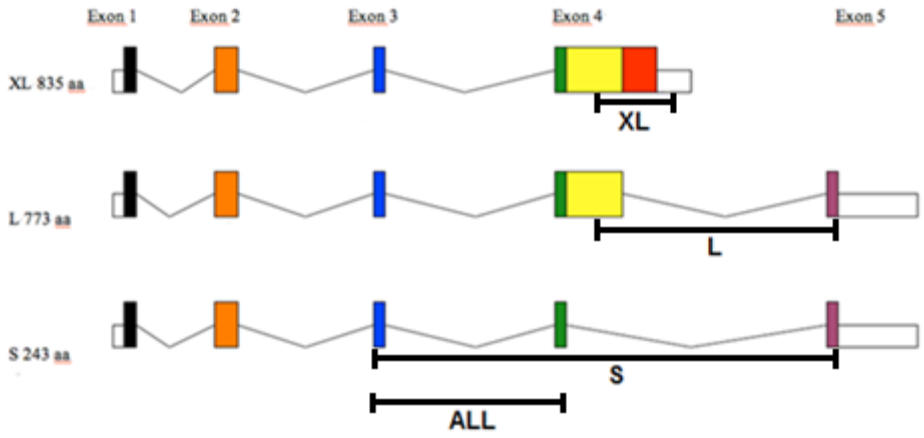


Figure 7.1 Representation of the *BCL11A* gene. Horizontal blue bars indicate the cross-exon design of amplicons to include exons specific to each isoform (XL, L, S) and all isoforms.

First, we over-expressed each isoform in K562 cells as positive controls for protein studies and qPCR studies in the different cell types.

7.2 RESULTS

7.2.1 Over-Expression of BCL11A isoforms in pEF6 vector

Erythroid cDNA was used to amplify cDNA corresponding to the three main isoforms of BCL11A. Even though there is very little of the L or S forms in erythroid cells, it was evident that sufficient mRNA was produced to make RT-PCR possible. Proof-reading polymerase (Velocity Taq, Bioline, UK) was used to reduce the possibility of polymerase-induced mutations.

The RT-PCR primers used contained restriction sites Bam*HI* and Not*I* (Appendix 2, Figure 11.2) to enable the in-frame cloning of the cDNA into the expression vector pEF6 (Invitrogen, UK) (Figure 7.2). This vector contains a highly active E1α

promoter, which would ensure expression in K562 cells, even though this cell type does not normally express BCL11A at any great level.

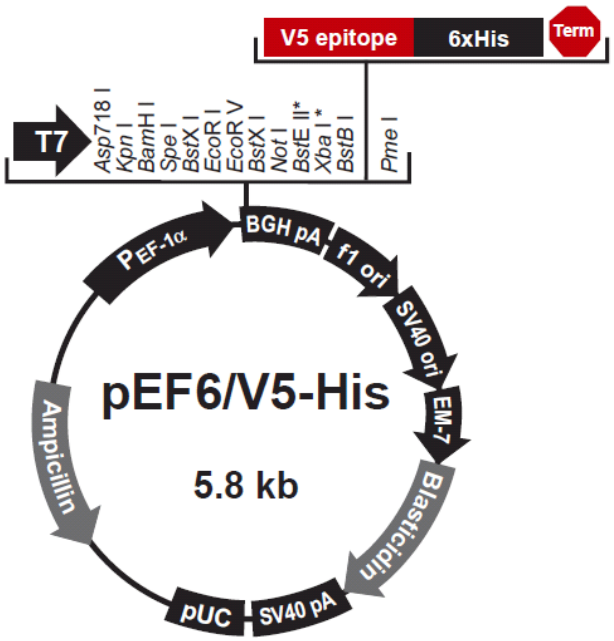


Figure 7.2. The pEF6/V5-His over-expression vector used to over-express the BCL11A isoforms. The V5 and His-tag linkers were not required and were therefore not designed to be added to the expressed BCL11A protein. This was achieved by leaving the naturally occurring BCL11A isoform stop codons intact.

Following digestion of the RT-PCR products with Bam*HI* and Not*I* to allow sticky end cloning, it proved impossible to successfully clone the amplicons directly into Bam*HI* - Not*I* digested pEF6. The reason for this was not clear, but was likely due to inefficient digestion of the PCR ends or degradation of the ends. It was therefore decided to clone the products into a TOPO vector first, sequence to verify products, digest with Bam*HI* and Not*I* and then clone into pEF6. The TOPO vector chosen was the TA-cloning vector TOPO 2.1 (Invitrogen, UK). RT-PCR products were gel-purified and A-tailed. The PCR products were then cloned into the TOPO vector according to the manufacturer's instructions. Several mini-

preps were prepared for each TOPO construct which were then sequenced to assess PCR fidelity.

Once sequence integrity was established, selected TOPO clones were digested with Bam*HI* and Not*I* to release their coding inserts, which were then isolated by gel purification. The pEF6 vector was opened with the same enzymes, dephosphorylated and the cDNA PCR fragment cloned in using the Rapid Ligation Kit (Roche, UK), according to the manufacturer's instructions. The resulting bacterial clones were screened by miniprep and inserts verified by restriction digest (Appendix 2, Figure 11.3).

Following electroporation into K562 cells, expression was assessed after 72 hours. RNA was extracted, reverse transcribed and TaqMan assays performed using a primer/probe set which would recognise all forms. Results show that high levels of expression were achieved for all isoforms (Figure 7.3). Protein for Western blotting analysis was also isolated at this time point.

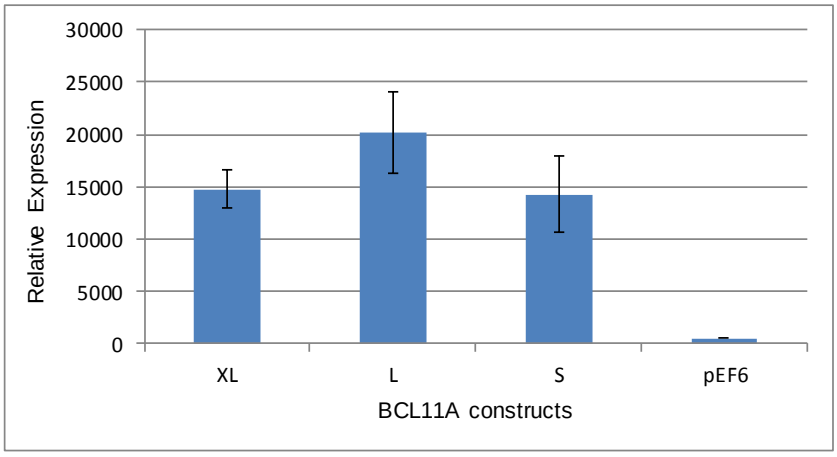


Figure 7.3 *BCL11A* isoform overexpression in K562 cells including negative control (K562 cells transfected with empty vector). Gene expression levels were determined by Taqman qPCR using primer probe mix specific to *BCL11A*. Values are means of at least 3 experiments +/- SE.

Figure 7.3 displays the results of the qPCR showing that all three isoforms were expressed in K562 cells in comparison to the negative control (an empty pEF6 plasmid transfected into K562 cells). This result confirmed that the transfection had been successful and that cDNA was being expressed from all constructs.

7.2.2 BCL11A isoforms in erythroid cells

I aimed first to identify the protein isoforms present in erythroid cells by western blotting, and then to measure by qPCR, the expression of the individual isoforms in different subsets of myeloid leukaemia. To differentiate the L from XL forms which are very similar in size, I isolated protein from K562 cells overexpressing each of the different isoforms as size markers. These were run on the same gel with protein extracted from erythroid cells for direct size comparison. K562 cells do not naturally express *BCL11A* which make them an appropriate cell model for overexpression studies.

To determine which isoform is translated in primary erythroid cells, a western blotting time course was performed (Figure 7.4a). Protein was isolated from primary human erythroid progenitor cells at different time points of the culture in phase II from day 0 to day 10. Native K562 cells were used as a negative control for BCL11A, protein from K562 cells overexpressing the isoforms was used as a positive control and β-actin was used as a loading control. I found that BCL11A was expressed in primary human erythroid cells from day 4 through to

day 8/9 of the culture. Only one isoform was present and appeared to be one of the large isoforms.

In Figure 7.4b I show both the XL and L protein isoforms are over-expressed in K562 cells (lanes 1&2), but I failed to see any presence of the S isoform (lane 8). In the same figure, a direct visual size comparison clearly shows that primary human erythroid cells express the XL isoform only.

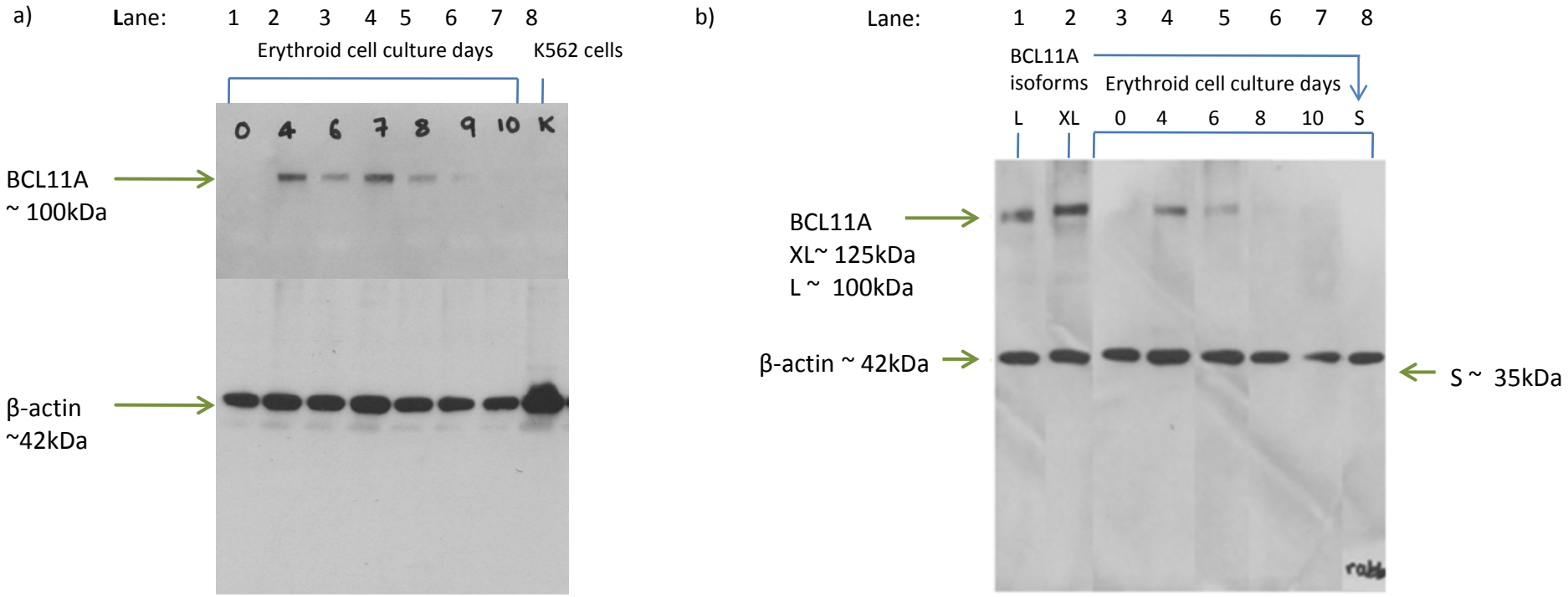


Figure 7.4 a) presence of BCL11A protein as detected by western blotting in primary human erythroid cells from Days 4, 6, 7, 8, 9 and 10 of phase II of culture (lanes 2-7). BCL11A antibody (14B5 sc-56011 from Santa Cruz Biotech) was used in a 1:200 dilution with 5% milk in PBST. Non-transfected K562 cells (lane 8) were used as a negative control for BCL11A protein expression. b) BCL11A western blot of K562 cells transfected with *BCL11A* L, XL and S overexpression plasmids (lanes 1,2 & 8, respectively). No S isoform was detected. BCL11A protein was detected in erythroid cells from days 4 and 6 of the culture (lanes 4 and 5, respectively) and appeared to be the XL form when compared to protein control (L and XL, lanes 1 and 2, respectively). For both blots β-actin (AC15 from Abcam) was used as a loading control in a 1:2000 dilution with 5% milk in PBST.

7.2.3 *BCL11A* isoforms in myeloid leukaemia

To investigate *BCL11A* isoform expression in myeloid leukaemia, cDNA samples were obtained from the Central England Haemato-oncology Research Biobank (CEHRB) with ethical approval. I received cDNA samples obtained from whole blood and bone marrow samples of patients with CMML, JMML, AML and AML with monosomy 7 and deletions within 7q (del 7q), as well as normal cDNA samples as the control group. *BCL11A* isoform expression was measured by SYBR-Green qPCR using primers specific to each isoform (Appendix 1, Table 11.5).

qPCR data displaying fold change in expression of isoforms in all patient samples across all disease groups is shown in Figure 7.5. All measurements were normalized to *HPRT* (housekeeper gene) and one individual from the CMML group. A control group of 10 individuals (no disease) was also included, as well as a group of patients with AML and normal karyotype (*n*=10). Full sample details including karyotype classification is listed in Appendix 4, Tables 11.8-11.12.

I also measured the isoforms in primary human erythroid progenitor cells, data shown in Figure 7.6. All isoforms were expressed but due to variation in primer efficacy and amplicon design, expression levels of different isoforms cannot be compared (see Discussion 7.3).

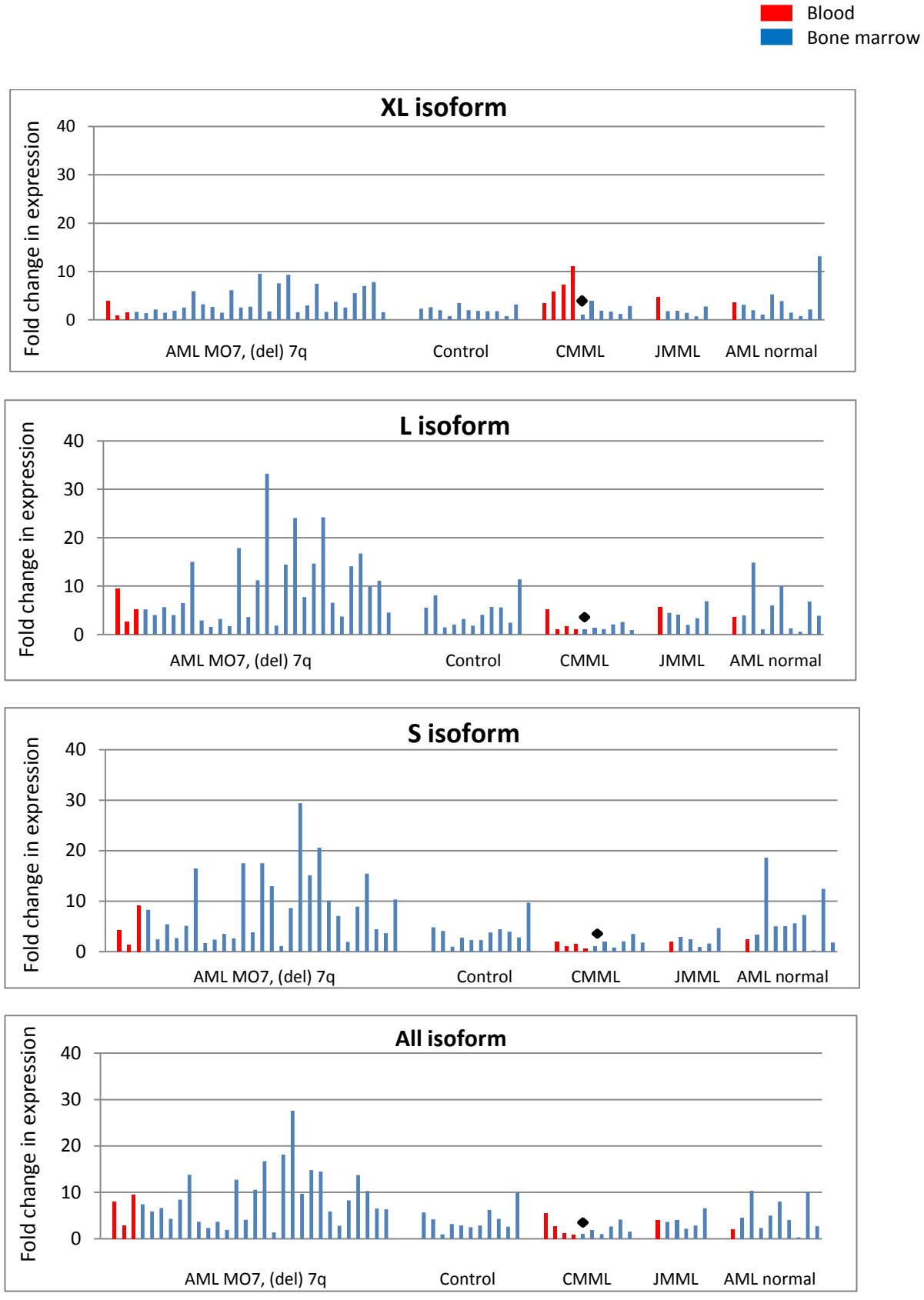


Figure 7.5 Fold change in expression of *BCL11A* isoform XL, L, S and all forms, in patients with AML M07/(del)7q (*n*=30), CMML (n=10), JMML (*n*=6), AML with normal karyotype (*n*=10) and a control group of normal individuals (*n*=10). Expression was measured by qPCR using primers specific to the individual isoforms. All CT values were normalized to the housekeeper gene *HPRT* and one randomly selected sample from the CMML group (◆) to allow ΔΔCt analysis.

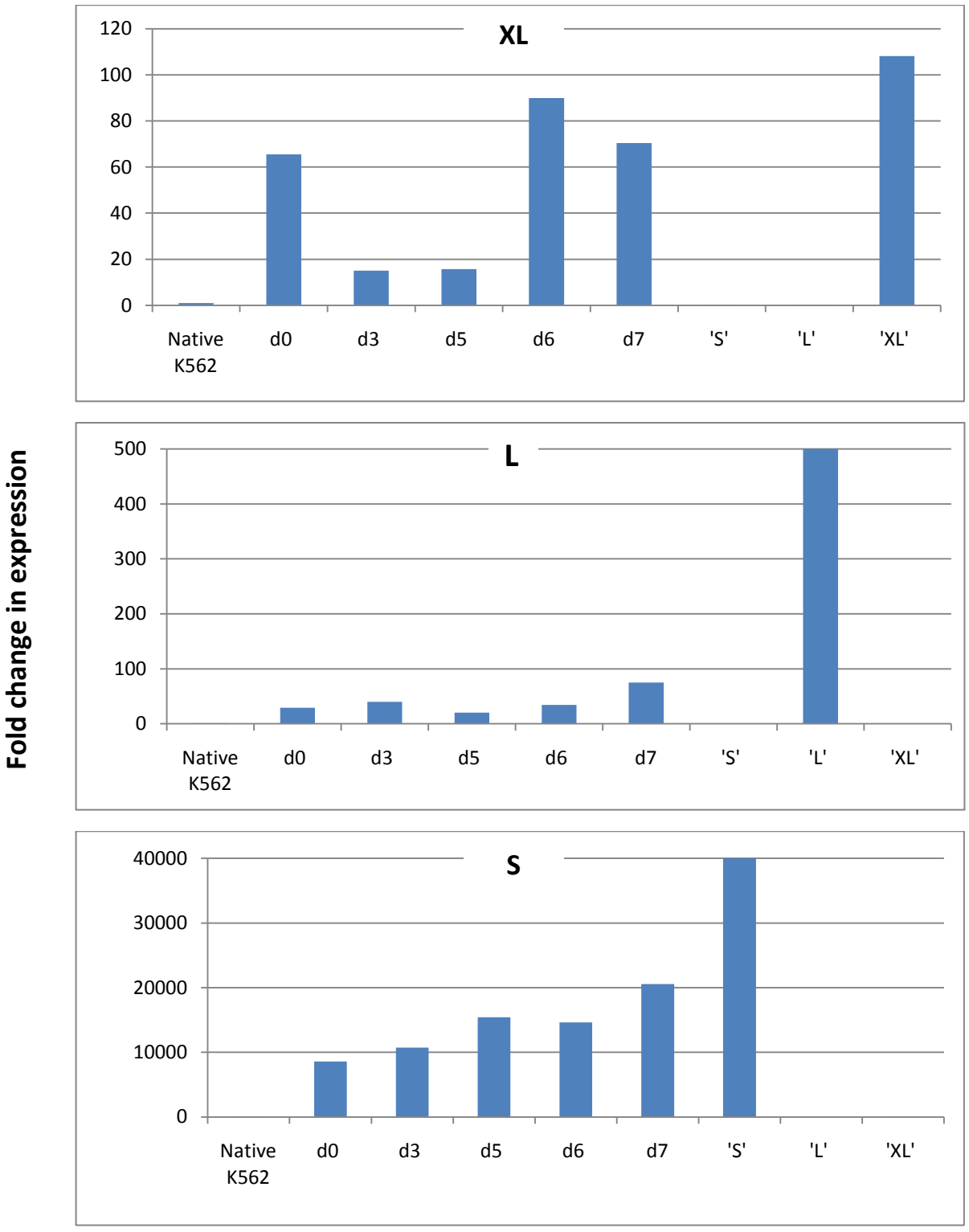


Figure 7.6 Fold change in expression of the *BCL11A* isoforms, 'XL', 'L' and 'S' in primary human erythroid progenitor cells from phase II day 0, 3,5,6 and 7 of culture. Native K562 cells (untransfected) were used as a negative control for *BCL11A* expression. cDNA from K562 cells transfected with over-expression plasmids for 'XL', 'L' and 'S' isoforms were measured as positive (and negative) controls for each isoform. All qPCR data (CT values) were normalized to *HPRT* and K562 cells.

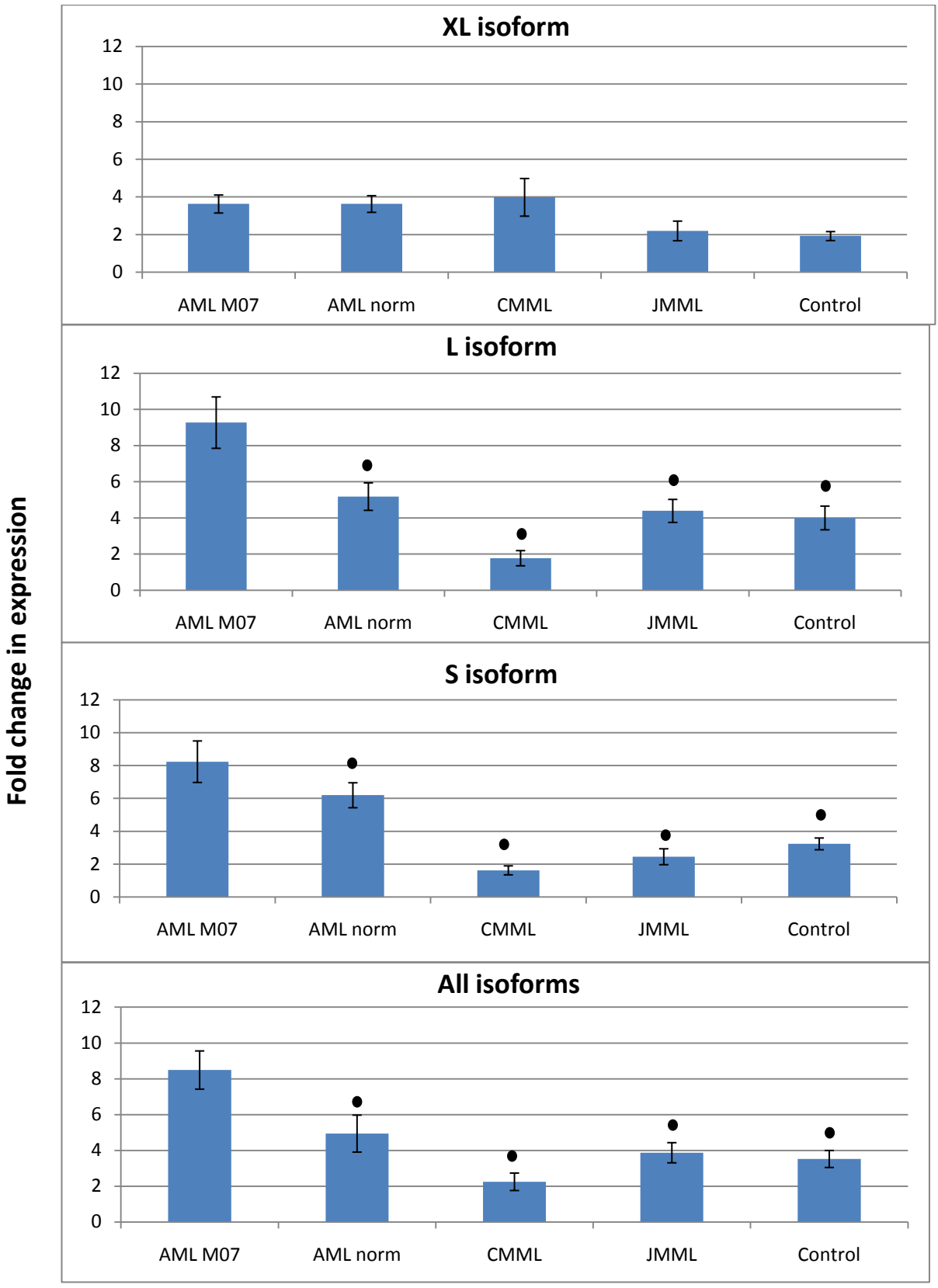


Figure 7.7 qPCR expression data (from Fig.7.5) was averaged for each group and displayed +/- SEM. Statistical significance between AML M07/(del) 7q and all other groups was measured by students t-test. (●) P= <0.05.

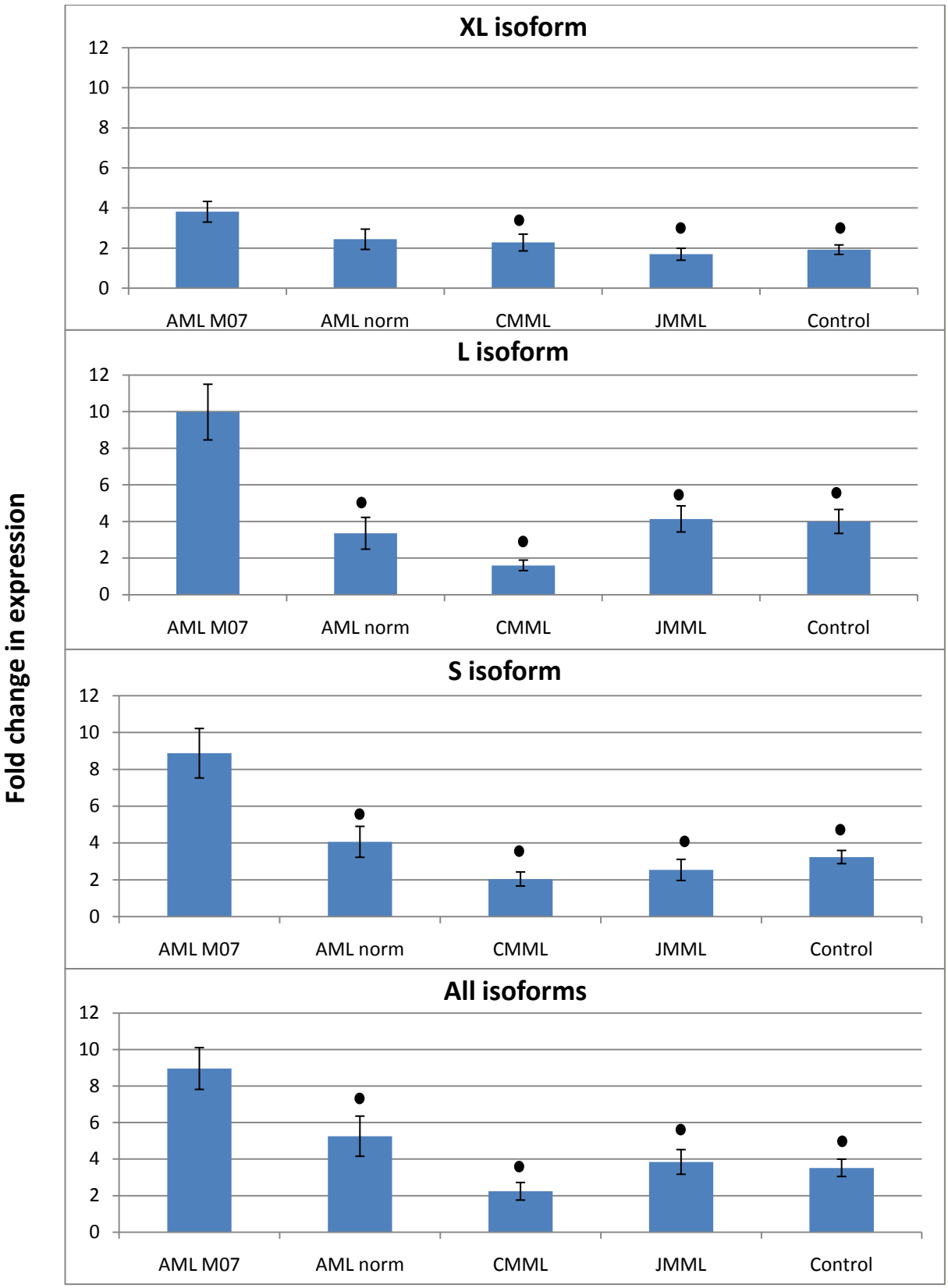


Figure 7.8 qPCR expression data (from Fig.7.5) was averaged for each group minus the blood samples and displayed +/- SEM. Statistical significance between AML M07/(del) 7q and all other groups was measured by students t-test. (●) P= <0.05.

7.2.4 *KLF1* expression in myeloid leukaemia

Since new studies have shown KLF1 regulates BCL11A expression, I decided to measure *KLF1* expression in the same myeloid leukaemia groups; AML, CMML and JMML to see if and how *KLF1* expression correlates with *BCL11A* expression in these cell types.

KLF1 expression was not significantly different between the three disease groups analysed except for a few individuals with higher expression, including one from the JMML group with MO7 and very high expression (Figure 7.9).

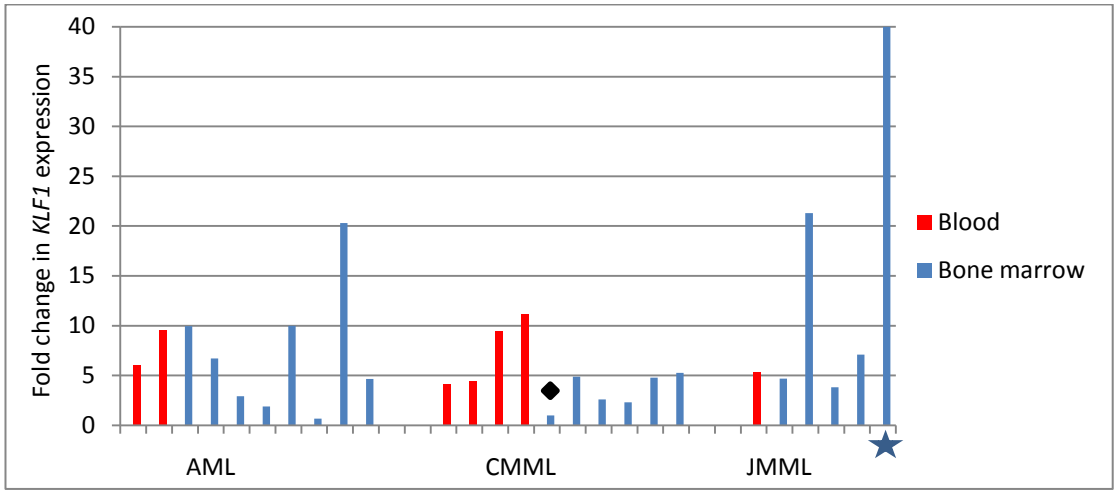


Figure 7.9 *KLF1* expression in AML, CMML and JMML groups. Y axis shows fold change in expression. Expression was measured by qPCR using primers specific to the individual isoforms. All CT values were normalized to the housekeeper gene *HPRT* and one randomly selected sample from the CMML group (◆). Individual with monosomy 7(★).

7.3 DISCUSSION

7.3.1 BCL11A in erythroid cells

Western blotting analysis revealed BCL11A protein is highly expressed in erythroid cells from day 4 of culture through to day 8, after which the protein is reduced. Figure 7.4a clearly shows a single band at each day of culture; therefore only one isoform is present in these erythroid cells. When this band is measured alongside the XL and L isoform in figure 7.4b), it is clear that this band corresponds with the XL isoform (control XL overexpressing K562 cells). Therefore I conclude that only the XL isoform is present in erythroid cells. These findings mirrored those previously reported by Sankaran *et al.* (2008) who also found the large isoforms expressed exclusively in adult erythroid cells and little or no expression of the short isoform in these cells.

Although we could over-express all three major isoforms of BCL11A in K562 cells (RNA levels detected in Figure 7.6), presence of the S isoform protein could not be detected in several experiments (Figure 7.4b). We have no clear explanation for this, but suggestions are that the S BCL11A isoform may be less stable or it could be transcribed but not translated in K562. Previous studies have shown that the S isoform is not present in definitive erythroid cells; however, the S isoform was detected in primitive erythroblasts, indicating that the BCL11A isoforms could be stage specific (Sanakaran *et al* 2008). It is therefore important to further understand the roles of the isoforms, as they may well lead to uncovering the role of BCL11A in the stage-specific globin switch.

7.3.2 BCL11A in myeloid leukaemia cells

Previous studies have shown a link between myeloid disorders, monosomy 7 and F-cell levels (Craig *et al.* 1996*)*. One study of MDS patients revealed raised F-cell levels in 62% of patients. Moreover, 4 MDS patients with monosomy 7 or 7q deletion had the highest number of F-cells (Craig *et al.* 1996).

JMML, a rare childhood form of myeloid disorder is often presented with raised HbF. 25-30% of JMML cases present with monosomy 7 or (del) 7q (Emanuel 2008). Reports have shown that elevated HbF is associated with JMML with normal karyotype, and in contrast, only normal to moderately elevated levels are reported in JMML patients with MO7 (Niemeyer *et al.* 1997).

BCL11A was found to be up-regulated in cases of AML with monosomy 7 and in some JMML cases (Yin *et al.* 2009). I obtained cDNA samples of patients with CMML, AML (with MO7 or -7q) and JMML. Of the groups studied, expression of *BCL11A* is highest in AML with monosomy 7 or (del) 7q, and the L and S isoforms of *BCL11A,* the most highly expressed.

After these initial results I expanded my sample size for the AML group with MO7 or 7q deletion to 30 samples. 10 AML samples with normal karyotypes were also included in the isoform expression study to see if the difference lay within chromosome 7 affected cases or if this was seen in normal AML cases too. A larger sample size confirmed the same pattern; patients with AML with MO7 or (del) 7q highly expressed the L and S isoforms compared to all other disease groups measured. This difference in expression was statistically significant ($p<0.05$, Figure 7.7 and 7.8). When measuring the bone marrow samples alone from each group,

statistically significant difference between AML with MO7 or (del) 7q and all other groups remained (Figure 7.8).

The amplification efficiency of the different isoforms is very different, and so, inter-isoform comparison is unreliable. For this reason, I could only compare expression of each isoform between the different disease and control groups but not their relative expression within each disease or control group.

My data shows that the XL isoform is predominant in erythroid cells and the L and S isoforms are highly expressed in acute myeloid leukaemia cells with M07 or other chromosome 7 abnormalities.

It may be that an element on chromosome 7 plays a role in the differential expression of *BCL11A* isoforms. Chromosome 7q has been widely implicated in myeloid malignancies and many studies have identified genes on chromosome 7 effected by MO7, (del) 7q or other chromosome 7 abnormalities. I carried out a literature search to find any common genes affected in chromosome 7 abnormalities that may be linked to haematopoiesis or *BCL11A* and may help to explain the high levels of *BCL11A*. The effected genes vary almost completely from study to study, probably because every patient included has a unique karyotype. A more precise region of chromosome 7 that could be involved with *BCL11A* cannot be pinpointed on the basis of existing data because deletions of 7q are different in every patient. Moreover, micro-deletion information is usually not provided in routine karyotyping as they are too small to detect by regular cytogenetic techniques and micro-deletions vary greatly. It is therefore very difficult to narrow down a causative region of chromosome 7.

In a more recent study, a SNP influencing platelet volume, count and function was identified on 7q (Soranzo *et al.* 2009), further highlighting the importance of this region in haematopoiesis. The significance of chromosome 7 in context of *BCL11A* regulation however is something which will need further investigation.

No HbF measurements were made for these samples by the CEHRB and as I only had cDNA samples, I was unable to measure HbF levels. As a surrogate, I measured the γ/β globin expression ratio, but this measurement was inconclusive. This may be because these myeloid leukaemia samples contain mostly blast cells and very few red cells, therefore making it difficult to measure globin expression. Information on the HbF in these samples would allow correlation, if any, with the levels of *BCL11A* in these samples.

In recent studies, KLF1 has been shown to regulate *BCL11A* (Borg *et al. 2010,* Zhou *et al*. 2010). I measured *KLF1* expression in the AML MO7/-7q, CMML and JMML groups. Interestingly, I found very high *KLF1* expression in the individual with JMML and MO7 (Figure 7.9), this individual also had slightly raised *BCL11A* in comparison to the rest of the group. However, this result is somewhat subtle and a real trend can only be confirmed with a larger sample size.

In a study of MDS patients with MO7, looking at gene expression profiles, genes related to haematopoetic progenitor cell proliferation and blood cell function were dysregulated. Genes down-regulated in MO7 were those controlling cell growth and differentiation (Chen *et al.* 2004). It is likely that disordered erythropoiesis in MDS and JMML with MO7/-7q contributes to HbF expression seen in MO7 cases (Papayannopoulou *et al.* 1991).

I observed in the CMML group that the XL isoform was higher in samples taken from whole blood versus samples taken from bone marrow (Figure 7.5). This may be because the whole blood samples are likely to contain more mature erythroid cells, which may contribute to higher XL expression observed in these samples. When all bone marrow samples were averaged (i.e. excluding the blood samples), I found a statistically significant difference in the expression of the XL isoform between the group AML with MO7 or (del) 7q and CMML, JMML and control groups (Figure 7.8).

In conclusion, the results presented in this chapter further our understanding of the *BCL11A* isoforms. I demonstrate here that the XL isoform is predominantly expressed in human primary erythroid progenitor cells and it is likely that an element linked to chromosome 7 has a regulatory effect on the expression of *BCL11A*, as seen through raised levels in AML samples with MO7/-7q.

A more in-depth study is warranted, which might reveal regulatory elements on chromosome 7 which control *BCL11A* expression.

CHAPTER 8

8 Analysis of SNPs in *BCL11A* intron 2 in twins with extreme HbF phenotypes.

8.1 INTRODUCTION

The strongest HbF-associated SNPs (key SNPs) are found in intron 2 of *BCL11A* (Figure 8.1). Data suggests that the region contains important regulatory sequences, as indicated by the GATA-1 erythroid transcription factor binding patterns at this region. It is not clear, however, how the region regulates the *BCL11A* gene, the expression of which influences γ-globin expression.

HbF-promoting variants (alleles) of these SNPs are in strong linkage disequilibrium ('high HbF haplotypes'). HbF-lowering alleles, on the other hand exist together in linkage disequilibrium as 'low HbF haplotypes'.

I decided to further analyse typical high HbF and low HbF haplotypes in this region to either confirm that all SNPs are known or to identify any novel sequence variants that might characterise these haplotypes. This would inform us if the known SNPs are in fact the only candidates for causative sequence variants effecting F-cell levels. Alternatively, there may be new, previously unknown SNPs which may individually or collectively contribute to the trait.

We re-sequenced 20kb of intron 2 (the critical region containing strongest-associated SNPs- Table 8.1) in five individuals with extreme F-cell levels who were homozygous for representative high HbF and low HbF haplotypes, from the St. Thomas' UK Adult Twin Registry (www.twinsuk.ac.uk). These subjects had previously been genotyped during the first GWAS that identified *BCL11A* as a QTL for HbF (Menzel *et al.* 2007). These individuals were homozygous for the four key SNPs associated with F-cell levels found through several GWAS; rs1123573, rs6732518, rs6545816, rs766432. Further SNP haplotypes could therefore be

inferred directly. Of the five individuals studied, three had very high F-cells and carried two copies of the high HbF haplotype each. Two subjects had very low HbF values and carried two copies of the low HbF haplotype each (Table 8.2).

Following re-sequencing work we wanted to know if the different SNP alleles caused differences in near-by transcription factor binding sites (TFBS). We carried out some basic bioinformatics analysis of transcription factor binding sites in the presence of the different SNP alleles for each of the high scoring SNPs.

We attempted to further analyse the nature of the intron 2 QTL region as a putative enhancer of the gene and designed reporter assays to explore this possibility in both high and low HbF individuals. From the re-sequencing results at the HbF *BCL11A* QTL, we found no new sequence variants in the region which could account for the raised HbF. Therefore we conclude that one or more of the known SNPs are responsible for the functional mechanism which could lead to the raised HbF phenotype in this group. One suggestion is that the QTL acts as an enhancer/regulator of *BCL11A* expression. In an attempt to understand how these SNPs, singly or in combination, might regulate the expression of *BCL11A*, and in turn, how they might affect HbF levels, we tested the function of the QTL SNPs using an enhancer assay based on a luciferase reporter system.

Primers were designed to amplify 2 large sections of the QTL containing the high-scoring SNPs and the novel GATA-1 binding sites (Figure 8.1). DNA from the same raised and low HbF genotype/phenotype individuals were used and compared in the luciferase assay.

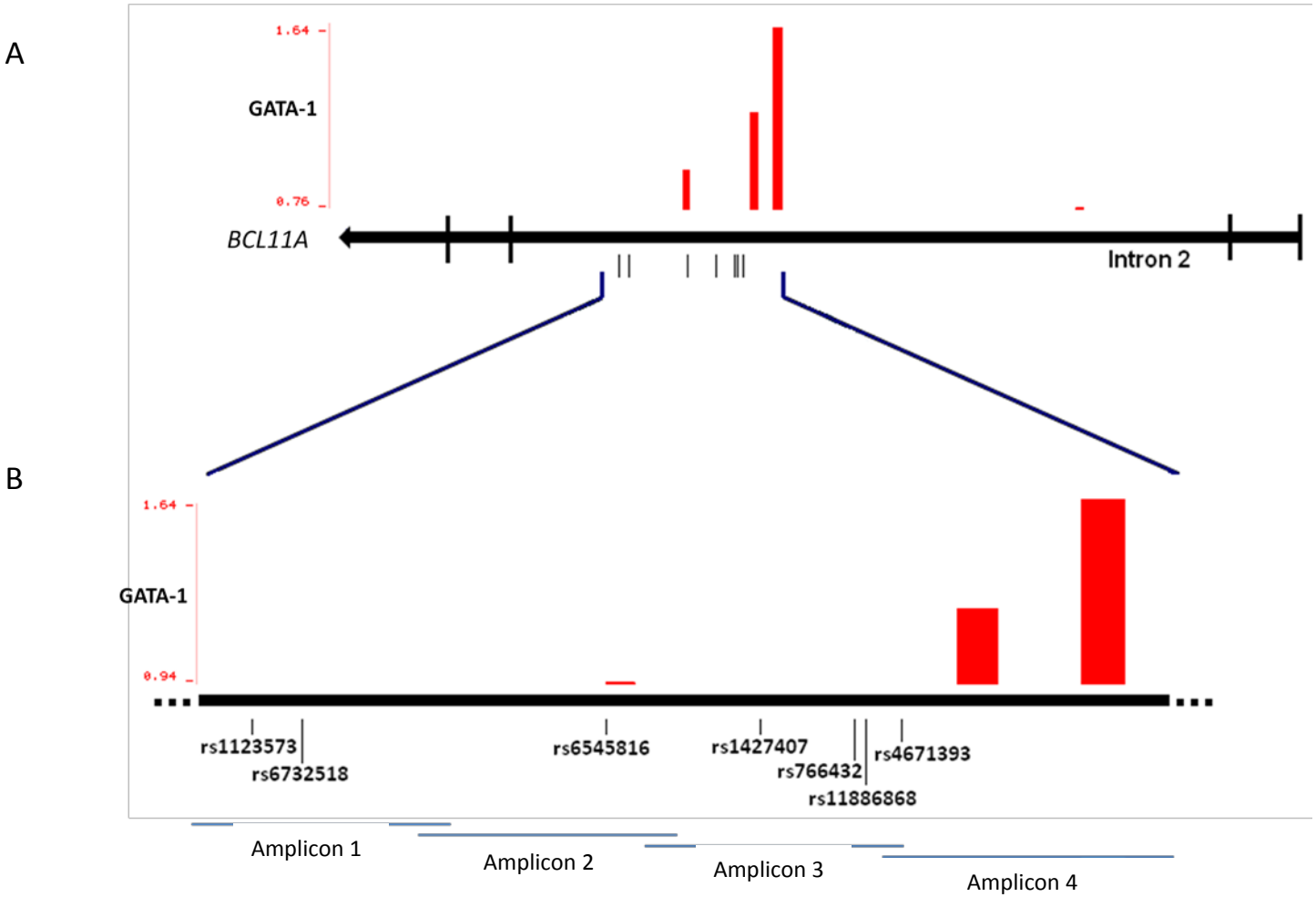


Figure 8.1 Schematic of *BCL11A* on chromosome 2. A) The first track shows GATA-1 binding signals in intron 2 (red peaks); black horizontal line represents gene, vertical lines mark exons. Vertical black lines below the gene mark location of key SNPs in the region. B) A detailed view of the intron 2 region; co-ordinates chr2:60,560,000-60,580,000 (20 kb). Black horizontal line indicates intron 2 of the gene, key SNPs are marked out with black vertical lines below this. The pale blue horizontal lines below the figure indicate the design of sequencing amplicons of c.5 kb each, spanning the 20kb region (Amplicon 1-4).

Table 8.1 Strongest HbF-associated SNPs found to date, including details of GWAS and populations in which they have been identified. * SNPs with highest association and located near the GATA-1 binding sites in intron 2 of *BCL11A*.

SNP	GWAS and populations
rs1123573	Menzel *et al.* 2007 (N.European)
rs6732518	Menzel *et al.* 2007 (N.European)
rs6545816	Menzel *et al.* 2007 (N.European) Sedgewick et al. 2008 (Thai)
rs1427407	Menzel *et al.* 2007 (N.European)
rs766432*	Menzel *et al.* 2007 (N.European) Sedgewick et al. 2008 (Chinese w. β-thal, Thai w. β-thal major, African American w. SCD)
rs11886868*	Menzel *et al.* 2007 (N.European) Uda et al. 2008 (Sardinian, w. β-thal, w. SCD)
rs4671393*	Menzel *et al.* 2007 (N.European) Lettre et al. 2008 (African American & Brazilian)

Table 8.2 SNP haplotypes that occur more frequently in either high F or low F phenotypes. Patient IDs are divided into high F and low F with IDs listed in the second column from the left.

		SNP haplotypes			
	Patient ID	rs1123573	rs6732518	rs6545816	rs766432
High F	1411	A	C	A	C
	1440	A	C	A	C
	3180	A	C	A	C
Low F	2109	G	T	C	A
	4041	G	T	C	A

8.2 RESULTS

8.2.1 Re-sequencing in extreme trait twins

The critical 20kb region of intron 2 was sequenced in 4 segments of around 5kb each (Figure 8.1). These segments were amplified using overlapping primers (Appendix 1, Table 11.4) and proof-reading polymerase, Velocity (Bioline, UK). The advantage of using a proof-reading polymerase is that it amplifies the exact same sequence as the template including mutations and SNPs, with few or no errors.

Once the large segments of the intron 2 region were amplified in all 5 individuals, the DNA was run on a 0.8% agarose gel and visualised. After confirmation of the amplified product, the bands were excised from the gel, column purified and sequenced directly (protocol in chapter 2.6.1).

The sequencing data was analysed using the Sequencher programme and assembled to the 20kb reference sequence (version, Mar.2006 (NCBI36/hg18)), accessed through the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

I designed sequencing primers tiling the length of each large amplicon with an overlap of about 100bp for complete coverage of the region.

The sequence was assembled for each individual against the template, 52 SNPs were identified and confirmed (Figure 8.2), of which 4 SNPs represented the “high” and “low” HbF haplotypes. However, no new sequence variants were identified in any of the individuals studied (Figure 8.2).

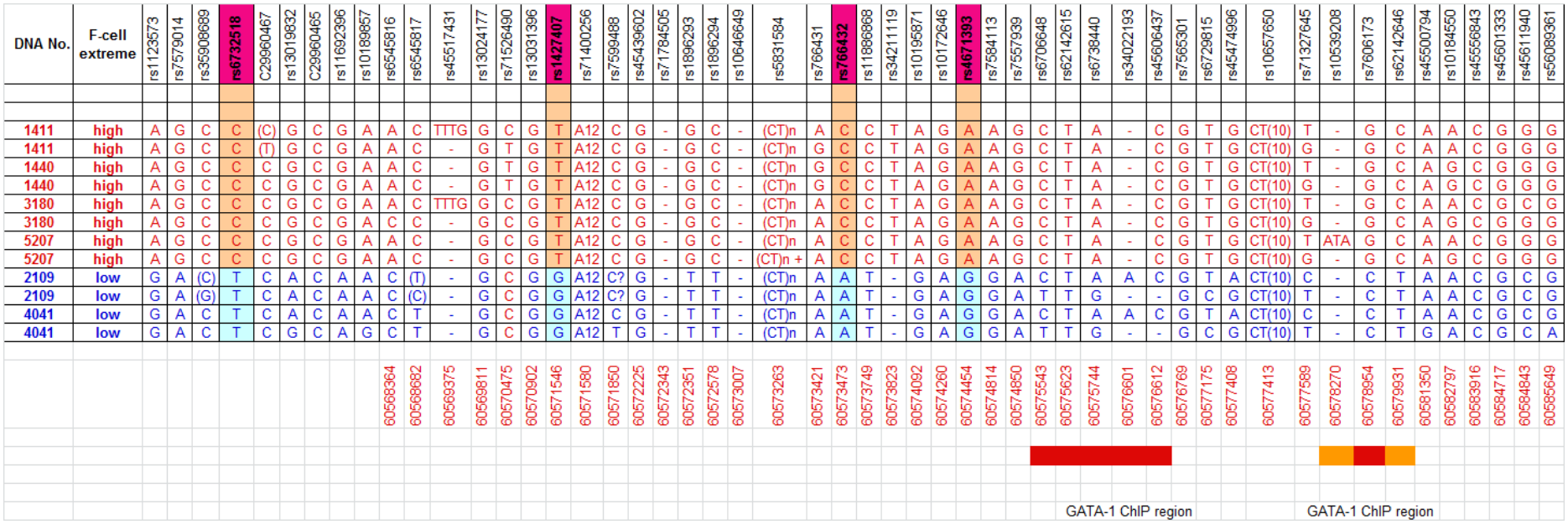


Figure 8.2 Sequencing data from extreme trait twins in the high scoring region of intron 2 (chr2:60,560,000-60,580,000). The figure displays key SNPs (pink box- high F samples are highlighted in peach and low F samples are highlighted in pale blue) as well as all other SNPs found in the region. Sample ID and HbF phenotype is listed in the first two columns to the left; the first row from the top displays SNP ID, SNP position co-ordinates in the genome (UCSC genome browsers, v. March 2006 NCBI36/hg18, http://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=299726899&clade=mammal&org=Human&db=hg18) are marked vertically in red below the SNPs. Red blocks below the SNP co-ordinates represent the location of GATA-1 binding sites, orange blocks represent weaker binding sites.

8.2.2 *In- silico* transcription factor binding analysis in extreme trait twins

Transcription factors (TF) have a key role in gene regulation (Cantor & Orkin, 2001) and erythroid development is largely mediated by many transcription factor complexes and cofactors, cis- or trans-acting to the expressed gene (Soler *et al.* 2010, Stadhouders *et al.* 2012). Any disruption or alteration in the binding sites of these important TFs can affect the expression of their target genes. It is possible that SNPs found in association with extreme HbF levels cause such an alteration in the binding of factors which contributes to the aberrant expression of HbF, indirectly via genes that influence HbF production. To investigate this we carried out a bioinformatics transcription factor binding analysis on the SNPs in both sets of individuals to assess if there were any differences in transcription factor binding as a result of the sequence variants.

We used an online programme called Transcription Element Search System (TESS) (http://www.cbil.upenn.edu/cgi-bin/tess) to predict TF binding sites at our specific sequences. TESS identifies binding sites using site or consensus sequences and positional weight matrices from the TRANSFAC database (database of eukaryotic transcription factors, their genomic binding sites and their DNA binding profiles).By inputting the different SNP alleles and including 25bp sequence either side of the SNP, we generated a transcription factor binding site map and compared this for each SNP between the high and low HbF samples (Figure 8.3-8.6).

Some differences in transcription factor binding were seen between the two alleles for each SNP. The high HbF associated 'T' allele of rs1427407 showed binding sites for YY1, a repressor and activator of promoters which is also involved in histone modification. The low HbF associated 'G' allele showed binding sites for LEF1, which is known to bind to the T-cell receptor- alpha enhancer, maximising enhancer activity. A c-Myb binding site is also located near this SNP, which is unaltered between the extreme traits. The 'A' allele of rs766432 is associated with low HbF and analysis revealed two transcription factors binding sites for TBP and TF11A, both of which are involved in positioning and assembly of RNA polymerase II.

A binding site for Nf1 (negative regulator of the RAS signalling pathway) was found near the high HbF associated 'C' allele of rs11886868. The high F 'A' allele of rs4671393 showed a TTF1 binding site, which acts as a transcription termination factor for RNA polymerase I. RAF binding site was found near the opposite allele 'G' (high HbF), which is also involved in transcription repression.

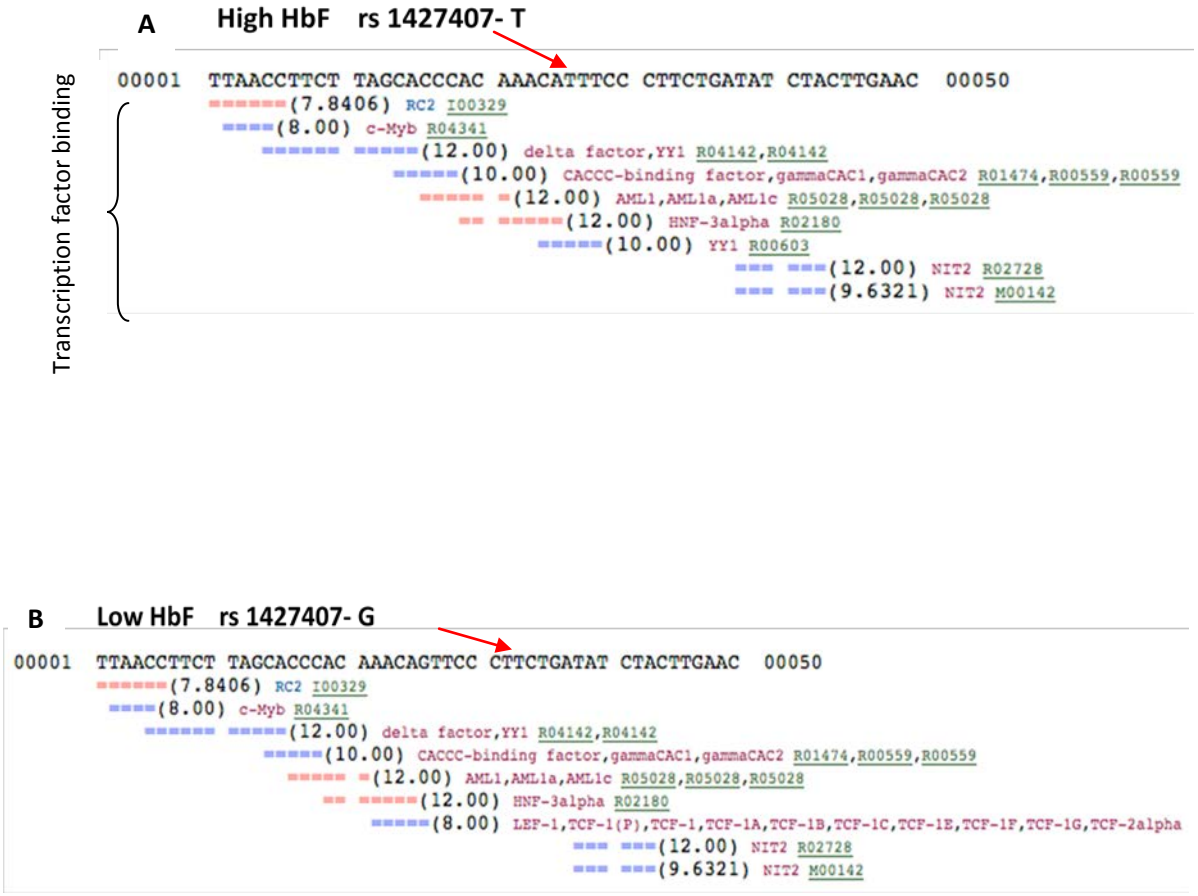
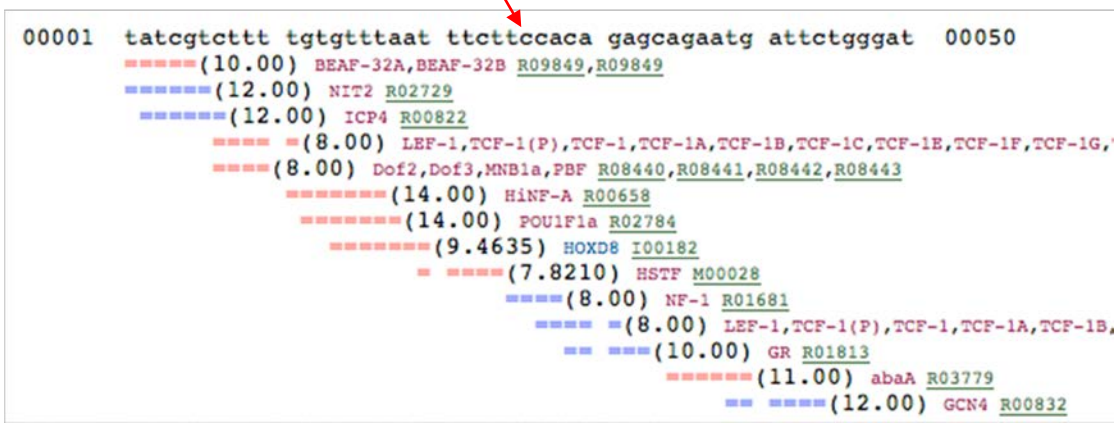


Figure 8.3 Transcription factor binding map for SNP rs1427407. A) 'T' nucleotide in high HbF and B) 'G' nucleotide in low HbF genotypes. Transcription factor binding sites are displayed below the 50bp sequence. The bracketed values show the log likelihood score of predicted transcription factor binding. The higher the score, the more likely the binding (maximum score is 14.00).

Figure 8.4 Transcription factor binding map for SNP rs766432. A) 'C' nucleotide in high HbF and B) 'A' nucleotide in low HbF genotypes. Transcription factor binding sites are displayed below the 50bp sequence. The bracketed values show the log likelihood score of predicted transcription factor binding. The higher the score, the more likely the binding (maximum score is 14.00).

A **High HbF rs11886868- C**



B **Low HbF rs11886868- T**

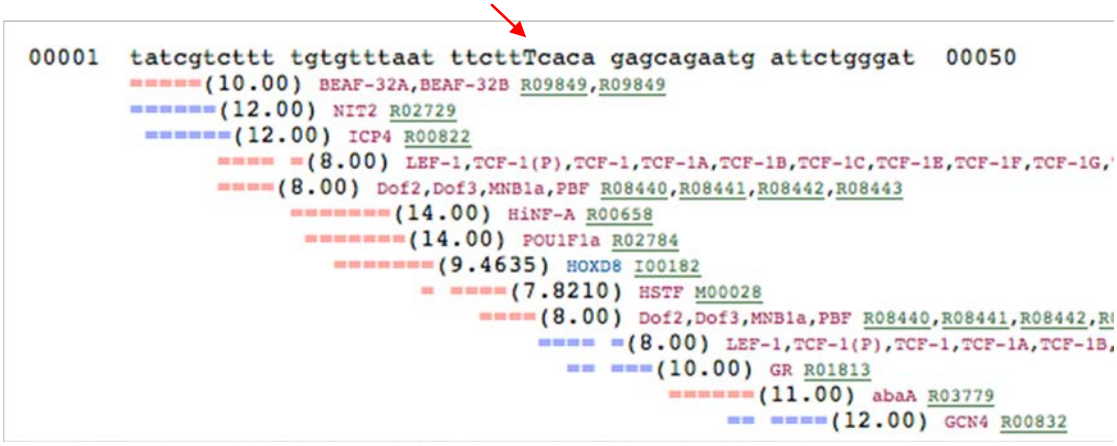


Figure 8.5 Transcription factor binding map for SNP rs11886868. A) 'C' nucleotide in high HbF and B) 'T' nucleotide in low HbF genotypes. Transcription factor binding sites are displayed below the 50bp sequence. The bracketed values show the log likelihood score of predicted transcription factor binding. The higher the score, the more likely the binding (maximum score is 14.00).

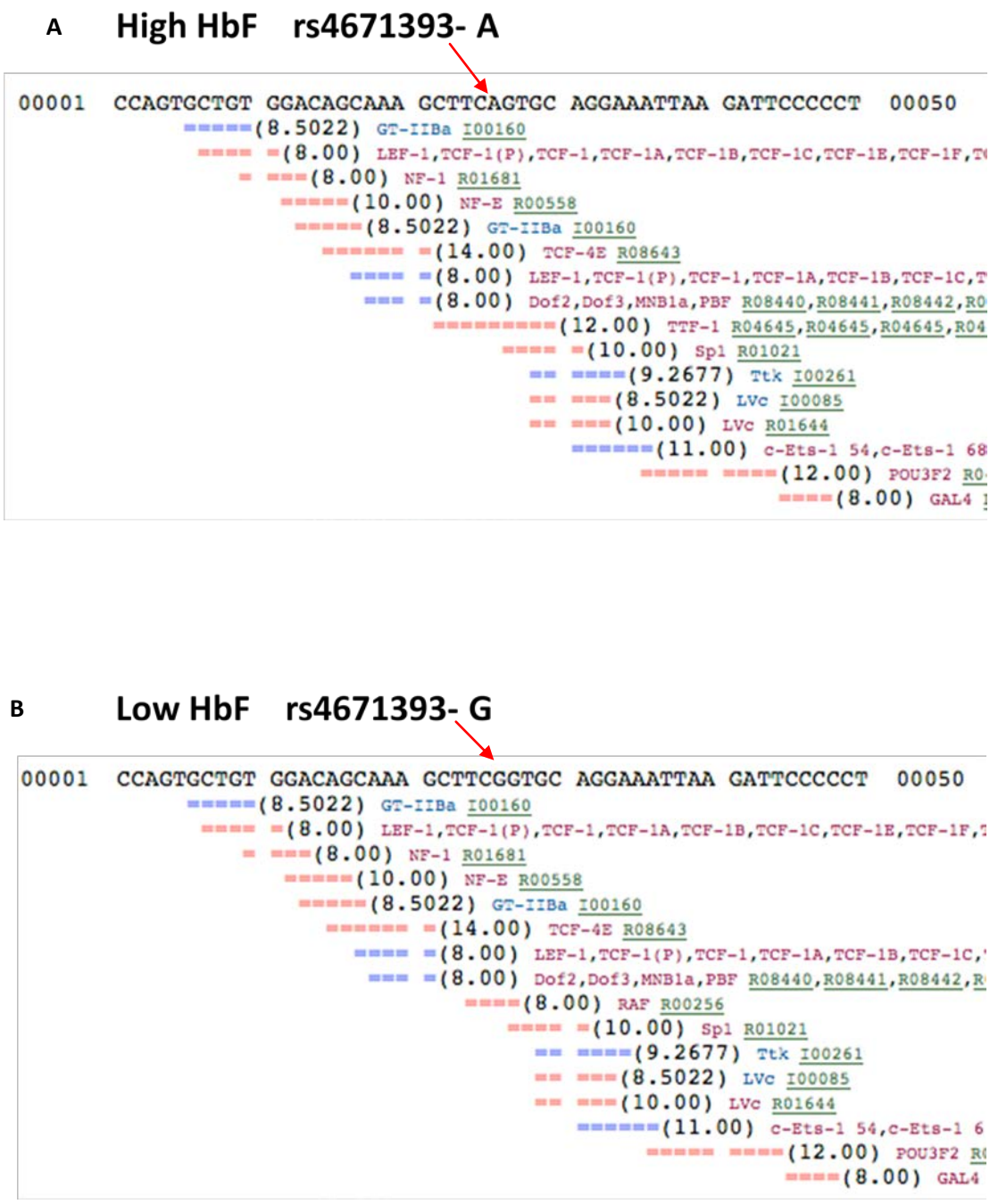


Figure 8.6 Transcription factor binding map for SNP rs4671393. A) 'A' nucleotide in high HbF and B) 'G' nucleotide in low HbF genotypes. Transcription factor binding sites are displayed below the 50bp sequence. The bracketed values show the log likelihood score of predicted transcription factor binding. The higher the score, the more likely the binding (maximum score is 14.00).

8.2.3 Functional analysis *BCL11A* intron 2 by reporter assay

The chosen method to test the function of the QTL SNPs was an enhancer assay based on a luciferase reporter system (Promega, UK). Primers were designed to amplify 2 large sections of the QTL containing the key SNPs and the novel GATA-1 binding sites (Figure 8.7). DNA from raised and low HbF genotype/phenotype individuals were used and compared in the luciferase assay. A third amplicon was designed from the β-globin HS3 LCR region (chr11: 5260110-5264249) which could act as a positive control in the enhancer assay.

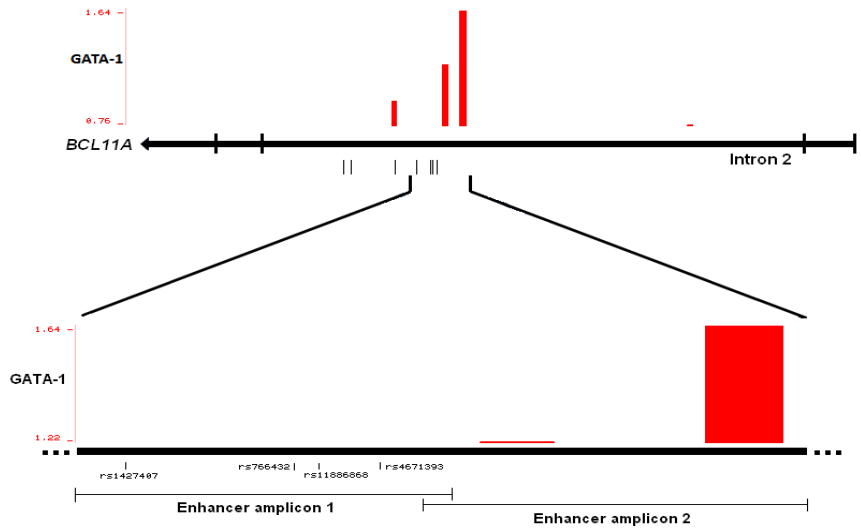


Figure 8.7 *BCL11A* gene locus on chromosome 2 showing the position of the putative enhancer regions being tested in intron 2, including the GATA-1 sites. Blue horizontal bars represent the gene and the track below this shows the key SNPs. Enhancer amplicon 1 (chr2 60570995-60575110) and Enhancer amplicon 2 (chr2 60575003-60579344) are indicated at the bottom by black horizontal lines.

The reported vector chosen was a minimal promoter variant of the pGL4 vector series from Promega (Figure 8.8). This vector is used for testing the enhancer activity of DNA. It carries a minimal mammalian promoter which drives the luciferase gene.

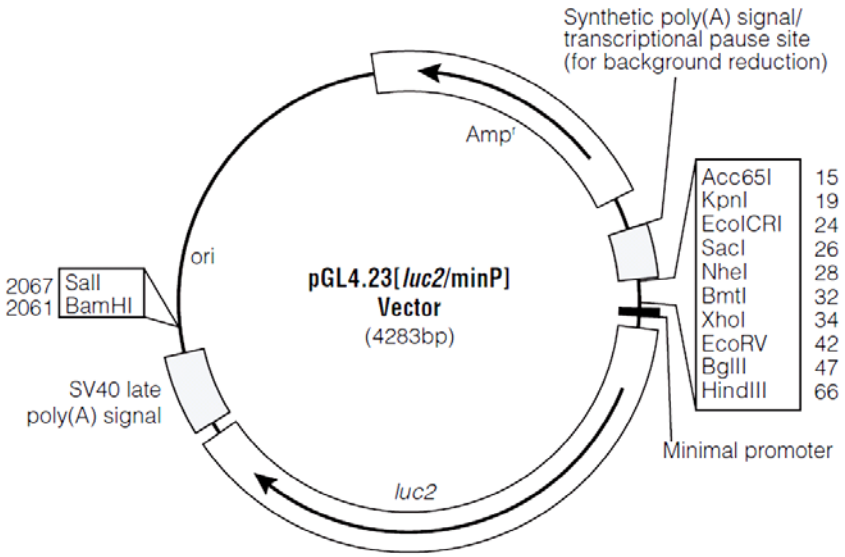


Figure 8.8. Map of the pGL4.23[luc2/minP] Vector (Promega, UK). For the enhancer 1 (chr2: 60570995-60575110) and HS3 (chr11;5260110-5264249) contructs, the KpnI and NheI sites were used. For the Enhancer 2 construct (chr2: 60575003-60579344), XhoI was used.

DNA was taken from two of the fully sequenced individuals (4041; low HbF-associated SNPs and 3180; high HbF-associated SNPs in BCL11A). Enhancer 1 (Enh1) and Enhancer 2 (Enh2) primer pairs were used to amplify their respective targets (Appendix 1, Table 11.3). The PCR products were gel purified and cloned into the TOPO-Blunt vector (Invitrogen, UK). Plasmids from several clones were prepared and subjected to restriction digest to verify that the insert contained insert of the correct genotype (Appendix 3, Figure 11.4 & 11.5 for enhancer 1 construct, and Figure 11.6 &11.7 for Enhancer 2 construct).

Following clone identity verification, the enhancers were transferred to the pGL4.23 vector. Plasmids were digested with Kpn*I*/Nhe*I* for the Enh1 construct and with Xho*I* for the Enh 2 construct. This was followed by gel purification of the released fragment, and ligation into the pEF vector (using the Roche Rapid Ligation Kit according to the manufacturer's protocol). Plasmid mini-preps were prepared from several colonies, and midi-preps before electroporation into K562 cells.

Each electroporation was carried out using 3 separate plasmid samples. After 48 hrs in culture, cells were harvested and assayed (using the Promego Dual Glo luciferase kit according to manufacturer's instructions). Luminometer reads were taken and each sample read 6 times (Figure 8.9). All methods used can be found in Chapter 2.

The data in Figure 8.9 show that there is some enhancer activity present in the Enh1 region but no significant difference between the "high" and "low" HbF individuals. The Enh2 constructs show a modest increase in enhancer activity, although again, there is no significant difference between the two genotypes. The HS3 enhancer shows the largest signal, which, given the nature of this element, is not unexpected.

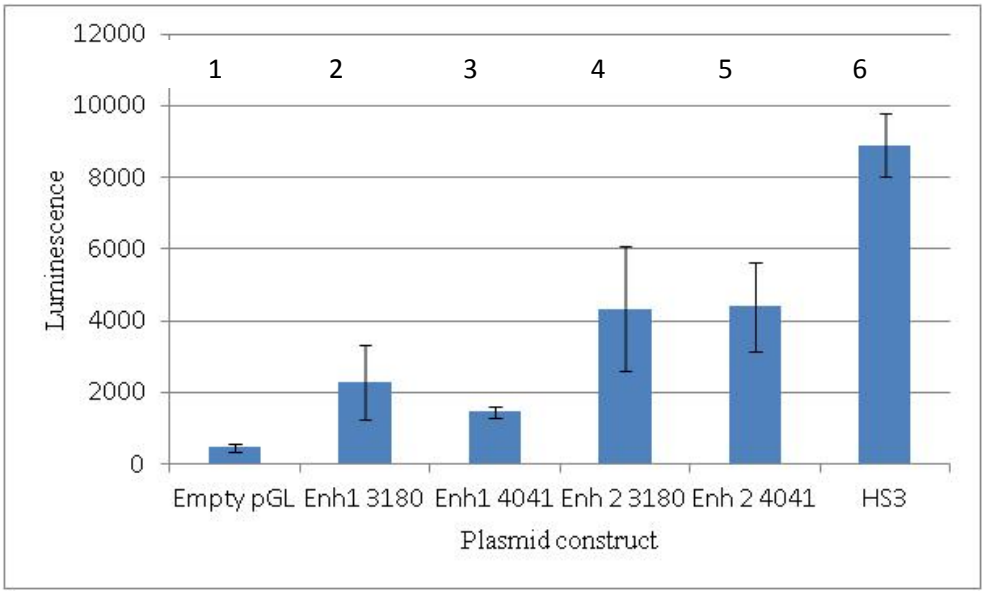


Figure 8.9 Results of the BCL11A enhancer reporter assays. The first lane is the empty pGL4.23 vector, used to assess background signal. Lanes 2 and 3 are results from the Enhancer 1 construct from the "high" (3180) and "low" (4041) HbF individuals, respectively. Lanes 4 and 5 are the corresponding Enhancer 2 constructs which contain known GATA-1 binding sites. Lane 6 shows the positive control β-globin HS3 region.

8.3 DISCUSSION

I re-sequenced the high scoring region of *BCL11A* intron 2 in individuals with high and low HbF phenotypes with corresponding genotypes. Sequence analysis revealed no gross or micro deletions or rearrangements in the sequence and no novel SNPs were found. The known SNPs were confirmed in these individuals suggesting that these SNPs are the causative sequence variants, and therefore the phenotype we see is most likely due to the effect of one or more of these polymorphisms.

A simple transcription factor binding site (TFBS) analysis revealed some TFBS were different between the high HbF and low HbF SNPs. Differences in score of TFBS indicates likelihood of a true binding site (higher score means it is more likely to be a true binding site and vice versa). Some of these factors are involved in repression or activation of promoters, histone modification and RNA polymerase II assembly. It is possible that the changes in the SNP alleles affect binding of these transcription factors which has a downstream effect on HbF production via altered *BCL11A* transcription. Moreover, if a TFBS near a SNP (such as the c-Myb binding site- Figure 8.3) is not altered, it is still possible that another factor binding to the c-MYb protein may be affected by this SNP. It is, however, noteworthy that this type of TFBS analysis is based on a prediction tool and *in vitro* studies are warranted to validate true differences in transcription factor binding.

To further distinguish how this region functions we studied part of the intron 2 sequence for enhancer activity. There was slightly higher activity in the enhancer 2 region which contains the GATA-1 sites; but we found no significant difference in enhancer like activity between the high and low F genotypes.

Although we did show evidence of general enhancer activity, we do, however, realize that the experiment design could be optimised. Amplicon design may not have been sufficient to incorporate all the important regions of the enhancer. Individual regions of an enhancer element may not show activity in isolation but only function in combination with other elements. An example is the β-globin LCR; individually only the β-globin LCR HS2 and 3 show enhancer activity on β-globin expression in stable reporter assays, however, the presence of HS1 and 4 are required for full induction of β-globin expression (Collis *et al*, 1990). Dividing the region in different constructs, as we have done, could disrupt the regulatory function of a potential enhancer.

The region may also need other elements for the enhancer to be fully functional, possibly trans-factors such as transcription factors which were not available in the K562 cell environment in which the constructs were transfected. This is possible as K562 cells do not naturally express high levels of *BCL11A,* which may be due to the loss of important factors required for *BCL11A* expression (Andersson *et al.* 1979).

Transfection efficiency is another factor which could have led to sub-optimal results. For these reasons other cell lines should be tested and even transgenic mouse models should be explored. The enhancer construct itself was prepared in one orientation which may not have been the correct one for the enhancer to function properly.

The findings of this experiment therefore do not disprove the hypothesis of an enhancer element in intron 2, but a better planned study is required to fully investigate this system in order to be able to distinguish one haplotype over another.

CHAPTER 9

9 Final Discussion

Genetic loci have been identified through family studies and GWAS as QTLs for HbF levels (Thein *et al.* 1987, Garner *et al.* 1998, Menzel *et al.* 2007). Two of these QTLs are trans-acting and un-related to the β-globin locus. The first was found in an intergenic region on chromosome 6q23 (*HMIP*) between two genes; *HBS1L* and *MYB,* which was known to be involved in erythropoiesis (Oh *et al.* 1999). The second and most recently identified QTL maps to intron 2 of a little studied gene called *BCL11A* on chromosome 2p15, which was previously implicated in B and T cell malignancies (Liu *et al.* 2003, Liu *et al.* 2006).

The aim of this research was to characterise these QTL to reveal the mechanism(s) by which they themselves are regulated and how they might affect HbF levels. HbF expression is stage specifically controlled in embryonic and fetal life. The continuous production of HbF into adult life has ameliorating effects in sickle cell disease (SCD) and β-thalassaemia, greatly reducing morbidity and mortality. The therapeutic targeted re-activation of HbF in the population suffering from these β-haemoglobinopathies could reduce disease severity and greatly increase quality of life and lessen the burden on the public health services.

The mechanism by which these QTLs influence the developmental control of the globin switch (γ-globin to β-globin), is poorly understood, but if comprehended, will provide better understanding, not only of globin gene regulation but also give insights in to the general mechanisms by which all genes are controlled developmentally, in humans.

This study has revealed substantial novel data regarding the regulatory nature of these QTLs which has further contributed towards our understanding of HbF control.

9.1 Discussion and further work

9.1.1 The *HMIP* QTL is indicative of a regulatory region

We have demonstrated in chapter 4.0 that the *HMIP* region has characteristics of a regulatory element. High levels of AcH3 and RNA Pol II binding reveal that the region is transcriptionally active. GATA-1 erythroid transcription factor binds at DNase I hypersensitive sites in the highest scoring block within the *HMIP* region- *HMIP2,* which is indicative of an erythroid-specific regulatory element.

The *MYB* gene has a role in erythroid progenitor cell proliferation and differentiation (Oh *et al.* 1999, Jiang *et al.* 2006) and altered levels can affect the rate of erythroid maturation. It is possible that the *HMIP* intergenic region regulates the *MYB* gene, which alters erythropoiesis (via altered erythroid kinetics) and indirectly affects HbF levels.

To further characterise this region I wanted to investigate any differences between individuals with high and low HbF phenotypes and corresponding genotypes. I assessed GATA-1 binding in the QTL region in high and low HbF individuals and found one GATA-1 signal absent in the high HbF individual.

This binding site was also in coincidence with DNase I hypersensitive site (HS3). This may be a causative variant which may interfere with the regulatory capacity of the *HMIP* region, causing altered *MYB* expression as seen in high HbF individuals (Jiang *et al.* 2006), and lead to the high HbF phenotype (as a result of more immature higher HbF expressing erythroid cells in circulation) (Jiang *et al.* 2006).

Assessment of KLF1 binding at the *HMIP* QTL revealed strong KLF1 binding in the *MYB* promoter and lesser binding in intron 1. Studies have shown that *MYB* introns contain

regulatory sequences involved in its control (Bender *et al.* 1987, Dooley *et al.* 1996, Seib *et al.* 1994). No binding was found in the intergenic *HMIP* region. Several factors have been suggested to regulate *MYB* expression including GATA-1 (Bartunek *et al.* 2003); it is possible KLF1 is also involved in *MYB* gene regulation.

Furthermore, taken together with other studies, our data suggests that the transcriptional activity in this region is erythroid-specific since it was not seen in other non-erythroid cells (Wahlberg *et al.* 2009).

The GATA-1 binding data is not quantitative and would need to be validated by qPCR. More functional studies of this site are being done to fully understand its potential in influencing HbF indirectly via *MYB* and/or *HBS1L.*

Partly on the basis of our findings, others have reported further studies in *MYB* regulation. Stadhouders *et al.* (2012) have shown, through ChIP-seq experiments to identify binding sites of key transcription factors, possible enhancers in *HMIP* and demonstrated that they form an active chromatin hub (ACH) incorporating the *MYB* promoter and intron 1. These findings are in agreement with transcription factor binding (KLF1 and BCL11A) that I have shown at intron 1 of *MYB* through ChIP-chip experiments (Figure 5.6).

Although we have documented and published evidence to support the regulatory nature of the *HBS1L-MYB* intergenic QTL, this needs to be validated and further characterised.

A possible way to assess how *HMIP* regulates *MYB/HBS1L* is by studying the physical interactions between the two. A useful method to do this is chromatin conformation capture (3C), a technique that allows the analysis of chromosome organisation in its

native state. Chromosomal regions can fold in order to bring enhancers and transcription factors closer to a gene to be transcribed, in this way we could determine if the *HMIP* region physically interacts with the flanking genes. Indeed, our collaborators have recently published such an analysis in the mouse locus (Stadhouders *et al.* 2012).

Targeted deletions of the region and the effect on the *HBS1L/MYB* genes can help determine the regulatory capacity of the region. Yeast artificial chromosomes (YAC) have been widely employed for functional analysis of the human β-globin locus in this way (Fedosyuk & Peterson, 2007, Navas *et al*. 2006, Navas *et al*. 2002, Bungert *et al*. 1999). YACs containing the intergenic QTL with deletions of specific sequences in transgenic mouse models or stably transfected cell lines would be used to measure the effect on the *HBS1L/MYB* genes. Alteration in the expression of these genes as a result of the QTL disruption could verify if the sequence is part of a regulatory element.

It is also important to further characterise the SNPs in *HMIP* block 2, which have the highest association with HbF levels and are located close to the GATA-1 sites. Electrophoretic mobility shift assays (EMSA) and differential allelic expression analysis would aid in identifying the causative SNP allele.

EMSAs are used to study protein–DNA or protein–RNA interactions and can determine if one or more proteins tend to bind to a specific DNA or RNA sequence. In our case we would be interested in determining any difference in transcription factor binding in the presence of the different SNP alleles.

Differential allelic expression measures the relative expression levels of two alleles of one gene. In a heterozygote for a functional cis-acting SNP that affects expression of a

gene, the mRNA from each copy of the gene will be expressed differently. An expressed SNP can be used as a marker to distinguish the mRNA transcripts of each allele. On-going work to elucidate the SNP functions is being undertaken by others in our group using such approaches.

9.1.2 *BCL11A* contains an auto-regulatory element in intron 2 and itself acts as a transcription factor at multiple loci.

BCL11A binding was frequently found to coincide with GATA-1 binding at erythroid-related loci including the α-globin and β-globin loci, thus demonstrating the regulatory role of BCL11A on globin expression (Sankaran *et al.* 2009). BCL11A is known to associate with the NuRD chromatin-remodelling complex and other transcription factors (Sankaran *et al.* 2008). I have provided, for the first time, an extensive multi-locus analysis of BCL11A, which has helped to demonstrate its role as an erythroid-related transcription factor.

As to the regulation of *BCL11A* itself, I have found very strong GATA-1 binding at *BCL11A* in intron 2, accompanied with high levels of transcriptional activity (ACH3 and RNA pol II). Three of the highest scoring SNPs are found near these GATA-1 sites.

Most research has been focused on the role of the gene in the regulation of γ-globin and HbF production (Sankaran *et al.* 2008 & 2009, Borg *et al.*2010, Zhou *et al*. 2010, Xu *et al.* 2012), and not on the regulation of the gene itself, which we believe may be controlled in some part by the intron 2 region. We have made efforts to understand this region and though our data indicates that the QTL is likely to be a control region of the gene, the mechanism by which this sequence exerts its regulatory role on *BCL11A*

expression is still to be determined, although it may regulate the levels of the various isoforms produced from this gene.

Through ChIP-on-chip experiments, I found KLF1 occupies sites on the *BCL11A* gene. Binding sites were identified at the promoter region of the gene and a second site on exon 4 which is specific to the larger *BCL11A* isoforms (XL & L), which we and others have found to be predominant in primary human erythroid progenitor cells (Sankaran *et al.* 2008).

SNPs in the *BCL11A* QTL are the causative variants in HbF levels.

Re-sequencing the *BCL11A* intron 2 QTL in extreme trait phenotype/genotypes showed no novel sequence variants, insertions or deletions, thereby confirming that the HbF phenotypes are due to one or more of the known SNPs. Work should now focus on elucidating how these SNPs cause the observed effect on HbF levels.

We further studied this region in high and low HbF individuals by screening the area in two fragments (Figure 8.7) for enhancer activity by luciferase reporter assay. We found slightly higher enhancer activity in the second fragment (Enhancer 2, Figure 8.9) which contained the GATA-1 sites, but found no significant difference between the high and low HbF individuals.

I found KLF1 activity on regulatory elements at both globin loci and binding of KLF1 was often found in parallel with GATA-1 binding. This is not unexpected as KLF1 participation in complexes with other erythroid transcription factors has been reported at other loci (Stadhouders *et al.* 2011).

Elements on chromosome 7 may affect *BCL11A* expression.

In an attempt to gain better understanding of the *BCL11A* isoforms, I measured levels of the isoforms in myeloid leukaemia. I found higher levels of L and S forms in cases with monosomy 7 or (del) 7q compared to all other myeloid leukaemia disease groups and the control group studied. Studies have linked sequence variants on chromosome 7 with haematological parameters (Soranzo *et al.* 2009) and genes related to haematopoetic progenitor cell proliferation and blood cell function are found to be dysregulated in M07 cases (Chen *et al.* 2004).

It is possible that an element on chromosome 7 may be involved in the regulation of the expression of the *BCL11A* isoforms, but narrowing this down to a region of 7q is difficult. There are very few erythroid related genes on chromosome 7, though there are more than 70 miRNAs. It is therefore a reasonable possibility that an miRNA encoded on chromosome 7q regulates translation of the isoforms. Individual miRNAs could be used to determine if there is any isoforms-specific effect on BCL11A. A rapid cell-based assay would be relatively simple to set up.

miRNA target prediction software would be a first step to identify candidate miRNAs that may bind BCL11A directly or indirectly. Once these are identified- miRNA knockdown studies would prove a suitable method to detect changes in expression levels of isoforms and verify the miRNA target.

HbF data for MO7 and (del)7q cases are very rarely available, but may provide a further link to BCL11A regulation. More work is necessary to uncover any associations between chromosome 7 and BCL11A.

The Intron 2 QTL on *BCL11A* is indicative of a regulatory region which may act as an enhancer of the gene. We made attempts to identify any enhancer-like activity in the area but the data produced was not substantial. It is worth using a better system to study the region for enhancer activity using a more reliable reporter assay or again, a large clone deletion/transgenic study.

Re-sequencing work in extreme phenotypes and genotypes confirmed that the known SNPs in the region, alone or collectively cause the variation in HbF levels in Northern Europeans. Which particular SNPs are causing the effect is unknown. One method to distinguish the causative SNP(s) is to measure differential transcription factor binding sites at important SNPs. We carried out some basic TFBS analysis using online TFBS prediction tools, but this is a fairly crude way to assess transcription factors at a given sequence. TFBS can only really be verified by *in vitro* or *in vivo* studies as is presently being carried out at the *HMIP* QTL.

Differential transcription factor binding results could be verified by ChIP analysis of human primary erythroid precursor cells cultured from individuals with the appropriate genotype.

BCL11A is a promising target for the therapy of SCD and β-thalassaemia as discussed in chapter 1.8.4.1, and studies have already provided evidence for this (Xu *et al.* 2012). These studies however are so far limited to mouse models and furthermore, the disruption of *BCL11A* is likely to have effects on other genes and pathways. For this reason, more work is needed to fully understand the function of *BCL11A*.

9.2 Concluding statement

The studies documented in this thesis strengthen our knowledge and understanding of the *HMIP* and *BCL11A* QTLs that influence HbF levels. Further studies are underway by our group and others to elucidate the mechanisms of function of these QTLs, which will hopefully lead to better understanding of the control of HbF production; something that should also prove useful in the therapeutic activation of HbF and treatment of β-haemoglobin disorders.

REFERENCES

Andersson L. C., Nilsson K., Gahmberg C. G., (1979) 'K562--a human erythroleukemic cell line', *Int J Cancer*, 23, 143-7.

Bank, A. (2005), 'Understanding globin regulation in beta-thalassemia: it's as simple as alpha, beta, gamma, delta', *J Clin Invest,* 115 (6), 1470-3.

Bank, A.(2006), 'Regulation of human fetal hemoglobin: new players, new complexities', *Blood,* 107 (2), 435-43.

Bartunek P., Kralova J., Blendinger G., Dvorak M., & Zenke M. (2003) 'GATA-1 and c-myb crosstalk during red blood cell differentiation through GATA-1 binding sites in the c-myb promoter' *Oncogene*, 22, 1927-1935.

Bender T.P., Thompson C.B., & Kuehl W.M. (1987) 'Differential expression of c-myb mRNA in murine B lymphomas by a block to transcription elongation', *Science*, 237, 1473-1476.

Bender, M. A., et al. (2000), 'Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region', *Mol Cell,* 5 (2), 387-93.

Bianchi E., Zini R., Salati S., Tenedini E., Norfo R., Tagliafico E., Manfredini R., Ferrari S., (2010), 'c-myb supports erythropoiesis through the transactivation of KLF1 and LMO2 expression', Blood, 116 (22), 99-110.

Bieker J. J., (2010) 'Putting a finger on the switch', *Nat Genet* 42 (9,) 733-4.

Bungert J., Tanimoto K., Patel S., Liu Q., Fear M., & Engel J.D. (1999) 'Hypersensitive site 2 specifies a unique function within the human beta-globin locus control region to stimulate globin gene transcription', *Mol Cell Biol* , 19, 3062-3072.

Borg J, et al., (2010) 'Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin' *Nat Genet,* 42 (9), 801-5

Bottardi S., Ross J., BourgoinV., et al., (2009) 'Ikaros and GATA-1 combinatorial effect is required for silencing of human gamma-globin genes', *Mol Cell Biol,* 29, 1526–1537.

Bourantas K. L., Georgiou I., Seferiadis K., (1991) 'Quantitation of HBF gamma-chain types by HPLC in patients with myelodysplastic syndrome,' *Haematologica*, 76 (4), 337-8.

Boyer S. H., Belding T.K., Margolet L., & Noyes A.N., (1975) 'Fetal hemoglobin restriction to a few erythrocytes (F cells) in normal human adults', *Science*, 188, 361-363.

Cantor A. B., Orkin S. H., (2001) 'Hematopoietic development: a balancing Act', *Curr Opin Genet Dev,* 11, 513–519

Carter D., Chakalova L., Osborne C.S., Dai Y.F., & Fraser P., (2002) 'Long-range chromatin regulatory interactions in vivo', *Nat Genet*, 32, 623-626.

Chakalova L., Osborne C. S., Dai Y. F., et al., (2005) 'The Corfu deltabeta thalassemia Deletion disrupts gamma-globin gene silencing and reveals post-transcriptional regulation of HbF expression', *Blood*, 105, 2154–2160.

Chasis J.A. & Mohandas N. (2008) 'Erythroblastic islands: niches for erythropoiesis', *Blood*, 112, 470-478.

Chen G., Zeng W., Miyazato A., Billings E., Maciejewski J. P., Kajigaya S., Sloand E. M., Young N. S., (2004) 'Distinctive gene expression profiles of CD34 cells from patients with myelodysplastic syndrome characterized by specific chromosomal abnormalities', *Blood,* 104 (13), 4210-8.

Chen Z., Luo H. Y., Steinberg M. H., Chui D. H., (2009) 'BCL11A represses HBG transcription in K562 cells', *Blood Cells Mol Dis,* 42, 144–149.

Close, J., et al. (2004), 'Genome annotation of a 1.5 Mb region of human chromosome 6q23 encompassing a quantitative trait locus for fetal hemoglobin expression in adults', *BMC Genomics,* 5 (1), 33.

Collins, F. S. and Weissman, S. M. (1984), 'The molecular genetics of human hemoglobin', *Prog Nucleic Acid Res Mol Biol,* 31, 315-462.

Craig, J. E., et al. (1996), 'Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach', *Nat Genet,* 12 (1), 58-64.

Craig, J. E., et al. (1997), 'Genetic heterogeneity in heterocellular hereditary persistence of fetal hemoglobin', *Blood,* 90 (1), 428-34.

De Gobbi, M., et al. (2007), 'Tissue-specific histone modification and transcription factor binding in alpha globin gene expression', *Blood,* 110 (13), 4503-10.

Donze, D., Townes, T. M., and Bieker, J. J. (1995), 'Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching', *J Biol Chem,* 270 (4), 1955-9.

Dooley S., Seib T., Welter C., & Blin N. (1996) 'c-myb intron I protein binding and association with transcriptional activity in leukemic cells', *Leuk Res*, 20, 429-439.

Drissen R., (2004) 'The active spatial organization of the beta-globin locus requires the transcription factor EKLF', *Genes Dev,* 18, 2485–2490.

Dzierzak E., (2005) 'The emergence of definitive hematopoietic stem cells in the mammal', *Curr Opin Hematol,* 12 (3), 197-202.

Ellis, J., et al. (1996), 'A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region', *EMBO J,* 15 (3), 562-8.

Emanuel P. D. (2008) 'RAS pathway mutations in juvenile myelomonocytic leukemia' *Acta Haematol,* 119 (4), 207-11.

Enver, T., et al. (1990), 'Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice', *Nature,* 344 (6264), 309-13.

Evans, T. and Felsenfeld, G. (1989), 'The erythroid-specific transcription factor Eryf1: a new finger protein', *Cell,* 58 (5), 877-85.

Fedosyuk H. & Peterson K.R. (2007) 'Deletion of the human beta-globin LCR 5'HS4 or 5'HS1 differentially affects beta-like globin gene expression in beta-YAC transgenic mice', *Blood Cells Mol.Dis.*, 39, 44-55.

Feingold, E. A. and Forget, B. G. (1989), 'The breakpoint of a large deletion causing hereditary persistence of fetal hemoglobin occurs within an erythroid DNA domain remote from the beta-globin gene cluster', *Blood,* 74 (6), 2178-86.

Fell H.P., Smith R.G., Tucker P.W., (1986) 'Molecular analysis of the t(2;14) translocation of childhood chronic lymphocytic leukemia', *Science*, 232, 491-4.

Feng.W. C., Southwood C. M., Bieker J. J., (1994) 'Analyses of β-thalassemia mutant DNA interactions with erythroid KrÜppel-like factor (EKLF), an erythroid cell-specific transcription factor', *J Biol Chem*, 269 (2), 1493-1500.

Fibach, E., et al. (1989), 'Proliferation and maturation of human erythroid progenitors in liquid culture', *Blood,* 73 (1), 100-3.

Forget B.G., (2001) 'Molecular Genetics of the Human Globin Genes' Disorders of Hemoglobin (ed. by M. H. Steinberg, B. G. Forget, D. R. Higgs, & R. L. Nagel), pp. 117-130. Cambridge.

Forget, B. G. (1998), 'Molecular basis of hereditary persistence of fetal hemoglobin', *Ann N Y Acad Sci,* 850, 38-44.

Forrester, W. C., et al. (1986), 'A developmentally stable chromatin structure in the human beta-globin gene cluster', *Proc Natl Acad Sci U S A,* 83 (5), 1359-63.

Forrester, W. C., et al. (1987), 'Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids', *Nucleic Acids Res,* 15 (24), 10159-77.

Forsberg, E. C., Downs, K. M., and Bresnick, E. H. (2000), 'Direct interaction of NF-E2 with hypersensitive site 2 of the beta-globin locus control region in living cells', *Blood,* 96 (1), 334-9.

Fraser, P., et al. (1993), 'Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes', *Genes Dev,* 7 (1), 106-13.

Furthmayr, H., Tomita, M., and Marchesi, V. T. (1975), 'Fractionation of the major sialoglycopeptides of the human red blood cell membrane', *Biochem Biophys Res Commun,* 65 (1), 113-21.

Garner C., Mitchell J., Hatzis T., Reittie J., Farrall M., & Thein S.L. (1998) 'Haplotype mapping of a major quantitative-trait locus for fetal hemoglobin production, on chromosome 6q23', *Am J Hum Gene*, 62, 1468-1474.

Georgopoulos K., Bigby M., Wang J. H., et al., (1994) 'The Ikaros gene is required for the development of all lymphoid lineages', *Cell,* 79, 143–156.

Gillemans N., Tewari R., Lindeboom F., Rottier R., de Wit T., Wijgerde M., Grosveld F., Philipsen S., (1998) 'Altered DNA-binding specificity mutants of EKLF and Sp1 show that EKLF is an activator of the beta-globin locus control region in vivo', *Genes Dev,* 12 (18), 2863-73

Gonda T. J., Sheiness D. K., Bishop J. M. (1982) 'Transcripts from the cellular homologs of retroviral oncogenes: distribution among chicken tissues', *Mol Cell Biol*, (6), 617-24.

Golde D.W., Bersch N., & Cline M.J., (1976) 'Potentiation of erythropoiesis in vitro by dexamethasone', *J Clin Invest*, 57, 57-62.

Gong, Q. H., McDowell, J. C., and Dean, A. (1996), 'Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene in vivo by 5' hypersensitive site 2 of the beta-globin locus control region', *Mol Cell Biol,* 16 (11), 6055-64.

Gribnau, J., et al. (2000), 'Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus', *Mol Cell,* 5 (2), 377-86.

Grosveld, F., et al. (1987), 'Position-independent, high-level expression of the human beta-globin gene in transgenic mice', *Cell,* 51 (6), 975-85.

Hardison R., (2001) 'Organisation, Evolution, and Regulation of the Globin Genes', Disorders of Hemoglobin (ed. by M. H. Steinberg, B. G. Forget, D. R. Higgs, & R. L. Nagel), pp. 95-116. Cambridge.

Harju-Baker, S., et al. (2008), 'Silencing of Agamma-globin gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the -566 GATA site', *Mol Cell Biol,* 28 (10), 3101-13.

Harju, S., McQueen, K. J., and Peterson, K. R. (2002), 'Chromatin structure and control of beta-like globin gene switching', *Exp Biol Med (Maywood),* 227 (9), 683-700.

Hong, W., et al. (2005), 'FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1', *EMBO J,* 24 (13), 2367-78.

Huehns, E. R., et al. (1964), 'Human Embryonic Hemoglobins', *Cold Spring Harbor Symposia on Quantitative Biology,* 29, 327-&.

Hughes J. R., Cheng J. F., Ventress N., et al., (2005) 'Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences', *Proc Natl Acad Sci,* 102, 9830–9835.

Jane, S. M. and Cunningham, J. M. (1996), 'Molecular mechanisms of hemoglobin switching', *Int J Biochem Cell Biol,* 28 (11), 1197-209.

Jiang, J., et al. (2006), 'cMYB is involved in the regulation of fetal hemoglobin production in adults', *Blood,* 108 (3), 1077-83.

Katsantoni, E. Z., et al. (2003), 'Persistent gamma-globin expression in adult transgenic mice is mediated by HPFH-2, HPFH-3, and HPFH-6 breakpoint sequences', *Blood,* 102 (9), 3412-9.

Kellum, R. and Schedl, P. (1991), 'A position-effect assay for boundaries of higher order chromosomal domains', *Cell,* 64 (5), 941-50.

Kim J., Sif S., Jones B., et al., (1999) 'Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes', *Immunity*, 10, 345–355.

Kunkel, H. G. and Wallenius, G. (1955), 'New hemoglobin in normal adult blood', *Science,* 122 (3163), 288.

Leimberg J.M., Konijn A.M., & Fibach E. (2005) 'Macrophages promote development of human erythroid precursors in transferrin-free culture medium', *Hematology*, 10, 73-76.

Letting D. L, Chen Y.Y. , Rakowski C., Reedy S., Blobel G. A., (2004) 'Context-dependent regulation of GATA-1 by friend of GATA-1', *Proc Natl Acad Sci*, 101, 476–481.

Lettre G., Sankaran V. G., Bezerra M. A., 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 Sc,*. 105, 11869–11874.

Li, J., et al. (1998), 'Multiple regulatory elements in the 5'-flanking sequence of the human epsilon-globin gene', *J Biol Chem,* 273 (17), 10202-9.

Li, Q. and Stamatoyannopoulos, G. (1994), 'Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator', *Blood,* 84 (5), 1399-401.

Lin H. H., Sternfeld D. C., Shinpock S. G., Popp R. A., Mucenski M. L., (1996) 'Functional analysis of the c-myb proto-oncogene', *Curr Top Microbiol Immunol*, 211, 79-87.

Lipsick J. S., (1996) 'One billion years of Myb', *Oncogene*, 13 (2) 223-35.

Liu P., Keller J. R., Ortiz M., et al., (2003) 'Bcl11a is essential for normal lymphoid Development', *Nat Immunol,* 4, 525–532.

Liu, H., et al. (2006), 'Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells', *Mol Cancer,* 5, 18.

Lopez R. A., Schoetz S., DeAngelis K., O'Neill D., Bank R., (2002) 'Multiple hematopoietic defects and delayed globin switching in Ikaros nullmice', *Proc Natl Acad Sci,* 99, 602–607.

Nakamura T., Yamazaki Y., Saiki Y., Moriyama M., Largaespada D. A., Jenkins N. A., (2000) 'Evi9 encodes a novel zinc finger protein that physically interacts with BCL6, a known human B-cell proto-oncogene product', *Mol Cell Biol,* 20 (9), 3178-86.

Nakamura T., Largaespada D. A., Shaughnessy Jr J. D., Jenkins N. A., Copeland N. G.*(1996) 'Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias', Nat. Genet. 12:149–153.*

Marotta, C. A., J. T. Wilson, et al. (1977). 'Human beta-globin messenger RNA. III. Nucleotide sequences derived from complementary DNA.' J Biol Chem 252(14): 5040-5053.

Menzel, S., et al. (2007), 'A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15', *Nat Genet,* 39 (10), 1197-9.

Merika, M. and Orkin, S. H. (1993), 'DNA-binding specificity of GATA family transcription factors', *Mol Cell Biol,* 13 (7), 3999-4010.

Migliaccio R. & Papayannopoulou T., (2001) 'Erythropoiesis'. Disorders of Hemoglobin (ed. by M. H. Steinberg, B. G. Forget, D. R. Higgs, & R. L. Nagel), pp. 52-71. Cambridge.

Miles J., Mitchell J.A., Chakalova L., Goyenechea B., Osborne C.S., O'Neill L., Tanimoto K., Engel J.D., & Fraser P. (2007) 'Intergenic transcription, cell-cycle and the developmentally regulated epigenetic profile of the human beta-globin locus', *PLoS.ONE.*, 2, e630.

Miller I.J. & Bieker J.J. (1993) 'A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins' *Mol Cell Biol*, 13, 2776-2786.

Milot,E., Strouboulis,J., Trimborn,T., Wijgerde,M., et al., (1996) 'Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription', *Cell*, 87, 105-114.

Mucenski M.L., McLain K., Kier A.B., Swerdlow S.H., Schreiner C.M., Miller T.A., Pietryga D.W., Scott W. J. Jr., & Potter S. S., (1991) 'A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis', *Cell*, 65, 677-689.

Navas, P. A., et al. (1998), 'Developmental specificity of the interaction between the locus control region and embryonic or fetal globin genes in transgenic mice with an HS3 core deletion', *Mol Cell Biol,* 18 (7), 4188-96.

Navas P.A., Li Q., Peterson K.R., & Stamatoyannopoulos G. (2006) 'Investigations of a human embryonic globin gene silencing element using YAC transgenic mice', *Exp Biol Med (Maywood)*, 231, 328-334.

Navas P.A., Li Q., Peterson K.R., Swank R.A., Rohde A., Roy J., & Stamatoyannopoulos G. (2002) 'Activation of the beta-like globin genes in transgenic mice is dependent on the presence of the beta-locus control region', *Hum Mo .Genet*, 11, 893-903.

Ney, P. A., et al. (1990), 'Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells', *Genes Dev,* 4 (6), 993-1006.

Nicolaides N.C , Gualdi R., Casadevall C., Manzella L., & Calabretta B. (1991) 'Positive autoregulation of c-myb expression via Myb binding sites in the 5' flanking region of the human c-myb gene' *Mol Cell Biol*, 11, 6166-6176.

Nicolis, S., et al. (1991), 'An erythroid specific enhancer upstream to the gene encoding the cell-type specific transcription factor GATA-1', *Nucleic Acids Res,* 19 (19), 5285-91.

Niemeyer C. M. (1997) 'Chronic myelomonocytic leukemia in childhood: a retrospective analysis of 110 cases. European Working Group on Myelodysplastic Syndromes in Childhood (EWOG-MDS)', *Blood,* 89 (10), 3534-43

Noguchi, C. T. (1984). 'Polymerization in erythrocytes containing S and non-S hemoglobins.' Biophys J 45(6): 1153-1158.

Nuez B., Michalovich D., Bygrave A., Ploemacher R., Grosveld F., (1995) 'Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene', *Nature,* 375, 316-8.

Oh I. H., Reddy E. P., (1999) 'The myb gene family in cell growth, differentiation and apoptosis', *Oncogene*, 18 (19) 3017-33.

O'Neill D. W., Yang J., Erdjument-Bromage H., et al., (1999) 'Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching', *Proc Natl Acad Scl,* 96, 349–354.

O'Neill D. W., Schoetz S. S., Lopez R. A., et al., (2000) 'An ikaros-containing chromatinremodeling complex in adult-type erythroid cells', *Mol Cell Biol,* 20, 7572–7582.

Orkin S.H., (1992) 'GATA-binding transcription factors in hematopoietic cells', *Blood*, 80, 575-581.

Papayannopoulou T., Brice M., & Stamatoyannopoulos G., (1986) 'Analysis of human hemoglobin switching in MEL x human fetal erythroid cell hybrids', *Cell*, 46, 469-476.

Papayannopoulou T., Nakamoto B., Anagnou N. P., Chui D., Dow L., Sanders J., (1991) 'Expression of embryonic globins by erythroid cells in juvenile chronic myelocytic leukemia', *Blood,* 77 (12), 2569-76.

Pauling., Itano H. A., & . (1949) 'Sickle cell anemia a molecular disease', *Science*, 110, 543-548.

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

Perutz, M. F., et al. (1960), 'Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis', *Nature,* 185 (4711), 416-22.

Philipsen S., Wood W. G., (2009) 'Erythropoiesi', Disorders of Hemoglobin (ed. by M. H. Steinberg, B. G. Forget, D. R. Higgs & D. J. Weatherall), 24-46.

Raich, N., et al. (1990), 'Autonomous developmental control of human embryonic globin gene switching in transgenic mice', *Science,* 250 (4984), 1147-9.

Richardson, A. L., Humphries, C. G., and Tucker, P. W. (1992), 'Molecular cloning and characterization of the t(2;14) translocation associated with childhood chronic lymphocytic leukemia', *Oncogene,* 7 (5), 961-70.

Rodriguez, P., et al. (2005), 'GATA-1 forms distinct activating and repressive complexes in erythroid cells', *EMBO J,* 24 (13), 2354-66.

Saiki Y., Yamazaki Y., Yoshida M., Katoh O., Nakamura T., (2000) 'Human EVI9, a homologue of the mouse myeloid leukemia gene, is expressed in the hematopoietic progenitors and down-regulated during myeloid differentiation of HL60 cells', *Genomics,* 70, 387–391.

Sankaran, V. G., et al. (2008), 'Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A', *Science,* 322 (5909), 1839-42.

Sankaran, V. G., et al. (2009), 'Developmental and species-divergent globin switching are driven by BCL11A', *Nature,* 460 (7259), 1093-7.

Sankaran VG, et al. (2011) 'MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13' *Proc Natl Acad Sci*, 108, 1519–1524

Satterwhite E., Sonoki T., Willis T. G., et al., (2001) 'The BCL11 gene family: involvement of

BCL11A in lymphoid malignancies', *Blood*, 98, 3413–3420.
Sedgewick A. E., Timofeev N., Sebastiani P., et al., (2008) 'BCL11A is a major HbF quantitative trait locus in three different populations with beta-hemoglobinopathies',*Blood Cells Mol Dis*
Seib T., Welter C., Engel M., Theisinger B., & Dooley S. (1994) 'Presence of regulatory sequences within intron 4 of human and murine c-myb genes' *Biochim Biophys Acta*, 1219, 285-292.
Senawong T., Peterson V. J., Leid M., (2005), 'BCL11A-dependent recruitment of SIRT1 to a promoter template in mammalian cells results in histone deacetylation and transcriptional repression', *Arch Biochem Biophys,* 434 (2), 316-25.
Soler E. et al*.,* (2010), 'The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation' *Genes Dev* 24, 277–289
Solovieff N., Milton J. N., Hartley S. W., et al., (2010) 'Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster', *Blood*, 115, 1815–1822.
Song S. H., Hou C., Dean A. (2007) 'A positive role for NLI/Ldb1 inlong-range beta-globin locus control region function*', Mol Cell* 28, 810–822
Soranzo N. et al (2009) 'Epub 2009 Feb 12.A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function', *Blood,* 113 (16) 3831-7.
Splinter E. et al. (2006) 'CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus' *Genes Dev* 20, 2349–2354.
Stadhouders R. et al. (2011) 'Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development', *EMBO J.* 31 (4), 986-99.
Stamatoyannopoulos, G. (1972), 'The molecular basis of hemoglobin disease', *Annu Rev Genet,* 6, 47-70.
Stamatoyannopoulos, G. (2005), 'Control of globin gene expression during development and erythroid differentiation', *Exp Hematol,* 33 (3), 259-71.
Stamatoyannopoulos, J. A. (1992), 'Future prospects for treatment of hemoglobinopathies', *West J Med,* 157 (6), 631-6.
Stamatoyannopoulos, J. A., et al. (1995), 'NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region', *EMBO J,* 14 (1), 106-16.
Steinberg, M. H. and G. P. Rodgers (2001). 'Pathophysiology of sickle cell disease: Role of cellular and genetic modifiers.' Seminars in Hematology 38(4): 299-306.
Stuart M.J. & Nagel R.L. (2004) 'Sickle-cell disease', *Lancet*, 364, 1343-1360.
Sullivan J., Feeley B., Guerra J., & Boxer L.M. (1997) 'Identification of the major positive regulators of c-myb expression in hematopoietic cells of different lineages' *J Biol Chem* , 272, 1943-1949.
Talbot D. & Grosveld F., (1991) 'The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites', *EMBO J*, 10, 1391-1398.
Tewari R., Gillemans N., et al., (1998) 'Erythroid Krüppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the beta-globin locus control region', *EMBO J*, 17 (8), 2334-2341.
Thein S.L., Sampietro M., Rohde K., Rochette J., Weatherall D.J., Lathrop G.M., & Demenais F., (1994) 'Detection of a major gene for heterocellular hereditary persistence of fetal hemoglobin after accounting for genetic modifiers' *Am J Hum Genet*, 54, 214-228.
Thein, S. L., et al. (1987), 'Association of thalassaemia intermedia with a beta-globin gene haplotype', *Br J Haematol,* 65 (3), 367-73.
Thein, S. L. and Weatherall, D. J. (1989), 'A non-deletion hereditary persistence of fetal hemoglobin (HPFH) determinant not linked to the beta-globin gene complex', *Prog Clin Biol Res,* 316B, 97-111.
Thein, S. L. (2002), 'Beta-thalassaemia prototype of a single gene disorder with multiple phenotypes', *Int J Hematol,* 76 Suppl 2, 96-104.

Thein, S. L., 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 (27), 11346-51.

Thein, S. L. and Menzel, S. (2009), 'Discovering the genetics underlying foetal haemoglobin production in adults', *Br J Haematol,* 145 (4), 455-67.

Tolhuis B., Palstra R.J., Splinter E., Grosveld F., & de L.W., (2002) 'Looping and interaction between hypersensitive sites in the active beta-globin locus', *Mol Cell*, 10, 1453-1465.

Tsai, S. F., et al. (1989), 'Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells', *Nature,* 339 (6224), 446-51.

Tsai, S. F., Strauss, E., and Orkin, S. H. (1991), 'Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter', *Genes Dev,* 5 (6), 919-31.

Uda M., Galanello R., 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*, 105, 1620-1625.

Vakoc C. R., (2005) 'Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1' *Mol Cell,* 17, 453–462.

Vernimmen, D., et al. (2009), 'Chromosome looping at the human {alpha} globin locus is mediated via the major upstream regulatory element (HS -40)', *Blood*.

Vernimmen D., Marques-Kranc F., Sharpe J. A., et al., (2009) 'Chromosome looping at the human {alpha}-globin locus is mediated via the major upstream regulatory element (HS-40)', *Blood,* 114, 4253–4260.

Vieira, K. F., et al. (2004), 'Recruitment of transcription complexes to the beta-globin gene locus in vivo and in vitro', *J Biol Chem,* 279 (48), 50350-7.

Visel A., Rubin E. M., Pennacchio L. A., (2009) 'Genomic views of distant-acting enhancers', *Nature*, 46, 199-205.

Wahlberg, K., et al. (2009), 'The HBS1L-MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells', *Blood,* 114 (6), 1254-62.

Wainscoat, J. S., et al. (1985), 'A novel deletion in the beta-globin gene complex', *Ann N Y Acad Sci,* 445, 20-7.

Wallrapp C., Verrier S.B., Zhouravleva G., Philippe H., Philippe M., Gress T.M., & Jean-Jean O. (1998) 'The product of the mammalian orthologue of the Saccharomyces cerevisiae HBS1 gene is phylogenetically related to eukaryotic release factor 3 (eRF3) but does not carry eRF3-like activity', *FEBS Lett.*, 440, 387-392.

Weatherall D. J., Clegg J. B., Wood W. G., Callender S. T., Sheridan B. L., Pritchard J., (1975) 'Foetal erythropoiesis in human leukaemia', *Nature,* 257, 710-2.

Westin E. H., Gallo R. C., Arya S. K., Eva A., Souza L. M., Baluda M. A., Aaronson S. A., Wong-Staal F. (1982) 'Differential expression of the amv gene in human hematopoietic cells', *Proc Natl Acad Sci*, 79 (7), 2194-8.

Wijgerde, M., Grosveld, F., and Fraser, P. (1995), 'Transcription complex stability and chromatin dynamics in vivo', *Nature,* 377 (6546), 209-13.

Wijgerde M., Gribnau J., Trimborn T., Nuez,B., Philipsen S., Grosveld F., & Fraser P. (1996) 'The role of EKLF in human beta-globin gene competition' *Genes Dev*, 10, 2894-2902.

Wood W.G., (2001) 'Hereditary Persistence of Fetal Haemoglobin and äâ Thalassemia', Disorders of Hemoglobin (ed. by M. H. Steinberg, B. G. Forget, D. R. Higgs, & R. L. Nagel), pp. 356-388.

Xu J. et al. (2011) 'Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing' *Science*, 334 (6058), 993-996.

Yin, B., et al. (2009), 'A retroviral mutagenesis screen reveals strong cooperation between Bcl11a overexpression and loss of the Nf1 tumor suppressor gene', *Blood,* 113 (5), 1075-85.

Zhou D, Liu K, Sun CW, Pawlik KM, Townes TM. (2010) 'KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching', *Nat Genet,* 42 (9), 742-4

A P P E N D I C E S

Appendix 1- Primer sequences

Table 11.1 Primer probe sets used for real time TaqMan® PCR

Gene	Forward primer sequence and exon targeted	Reverse primer sequence and exon targeted	Probe sequence and exon targeted
α-globin	TCTGGTCCCCACAGACTCAGA (exon 1)	GAAGTGCGGGAAGTAGGTCTTG (exon 2)	CCTGGAGAGGATGTTC (exon 1 and 2)
β-globin	CTGTGACAAGCTGCACGTGG (exon 2)	GCCAGCACACAGACCAGCA (exon 3)	CCTGAGAACTTCAGGCTCCTGGCCAAC (exon 2 and 3)
γ-globin	CGTCACTGTGACAAGCTGCAT (exon 2)	CGAAATGGATTGCCAAAACG (exon 3)	CCAGCACATTTCCCAGGAGCTTGAAGT (exon 2 and 3)
MYB	ATGATGAAGACCATGAGAAGGAAA (exon 7)	AACAGGTGCACTGTCTCCATGA (exon 9)	TGCTACCAACACAGAACCACACATGCA (exon 8 and 9)
HBS1L	ACAAGAATGAGGCAACAGTATCTACAG (exon 4)	TCAGATTCACTTCGCGATGTCT (exon 5)	CAAAAGGAAAACCAGTAGATTCC (exon 5)
GATA1	CAATGCCTGCGGCCTCTA (exon 5)	CATCCTTCCGCATGGTCAGT (exon 6)	CTACAAGCTACACCAGGTGA (exon 5 and 6)

Table 11.2 Primers for real time SYBR-Green PCR

Primer target	Forward primer sequence	Reverse primer sequence
α-HS40 Positive control for GATA1 binding (ChIP)	CGACCCTCTGGAACCTATCAG	TTGGCCTCCAGAAGCACTG
NEFM Housekeeping gene, negative control and normaliser	AGCTGGACAAGAAGGTGCAG	CCGTCGAGATGTCTGTCTTC
GAPDH Housekeeping gene, negative control and normaliser	ACTAGCGGTTTTACGGGCG	TCGAACAGGAGGAGCAGAGAGCGA
HBB promoter (β-globin) Positive control for KLF1 binding (ChIP)	CAATAGATGGCTCTGCCCTGAC	CAGGTACGGCTGTCATCACTTA
β-*HS3* Positive control for BCL11A binding (ChIP)	CAGATGGATGGGGCAATGAAG	GGGGGTATAGGGGAGCAG
KLF1	CCCAGTCACTAGGAGAGTCCAA	GCTACACCAAGAGCTCCCACCT
BCL11A	Ready-made TaqMan primer mix- Hs00256254.m1 (Invitrogen, UK)	

Table 11.3 Primers for *BCL11A* intron 2 enhancer study

Target	Forward primer	Reverse primer
Enhancer amplicon 1	TCATCGGTACCAACTCCTTAAGCCATCATTCCA	TCATCGCTAGCTTTCCACTCTTCTCTCCCCATA
Enhancer amplicon 2	TCATCCTCGAGCCCTCTGTAAACAAGGAGGTTG	TCATCCTCGAGCTGCCTGCTTCCTTCTCTTTAC
β-LCR HS3	TCA TCG GTA CCT TTG CTC TCT GTT CCT GTG AAA	TCA TCG CTA GCT TTG ATT GCT GCT GTC TTC AGT

Table 11.4 *BCL11A* intron 2 sequencing primers

Primer target	Forward primer sequence	Reverse primer sequence
Amplicon 1	TTCCCTATCCACAGACCACAGCCC	ACAGCCATAACAGGGTTTCCAGTG
Amplicon 2	ACCGTCTCTTTGGTGCAGTGGGAA	AGCCGTCTTTCAGGGTAGCACCAA
Amplicon 3	TGCGCTTCTCAGACCCAAATGCTC	TCCATGGGGAAATGGGTCCTGCTT
Amplicon 4	GCCCCTCTGCCTCTCTACATTCCA	CATCACATGCCTTACCAGTCAGGG

Table 11.5 SYBR-Green qPCR primers specific for *BCL11A* isoforms

Isoform	Forward primer sequence and exon targeted	Reverse primer sequence and exon targeted
XL	AGCTCAAAAGAGGGCAGACGCA (exon 4)	CTT CTC CTG TGG ACA GTG AGA TTG (exon 4)
L	AGCTCAAAAGAGGGCAGACGCA (exon 4)	AGA ACT TAA GGG CTC TCG AGC TTC (exon 5)
S	CGT TCT GCA CAT GGA GCT CTA ATC (exon 3)	AGA ACT TAA GGG CTC TCG AGC TTC (exon 5)
ALL isoforms	CGT TCT GCA CAT GGA GCT CTA ATC (exon 3)	AGT TGT ACA TGT GTA GCT GCT GGG (exon 4)

Table 11.6 qPCR primers to validate KLF1 ChIP-chip data

Primer target	Forward primer sequence	Reverse primer sequence
B-HS3 (peak 2)	GGGAGCAGTCCCATGTAGTAGT	TAGGTGGTTAGGTCAGGTTGGT
***HBB* promoter (peak 1)**	GGCTCTGCCCTGACTTTTATG	AAGTCCAACTCCTAAGCCAGTG
***BCL11A* exon 4 (peak 1)**	GCCTCTCTCGATACTGATCCTG	CTTAGAAAGCGAACACGGAAGT
***BCL11A* exon/intron 1 (promoter)**	CACCACCACCAGCTCTTATACA	GGAGAAGTGTGTTTGAGTGTGC
***HBB* promoter (positive control for KLF1 enrichment)**	CAATAGATGGCTCTGCCCTGAC	CAGGTACGGCTGTCATCACTTA

Appendix 2- Over-expression plasmid construction

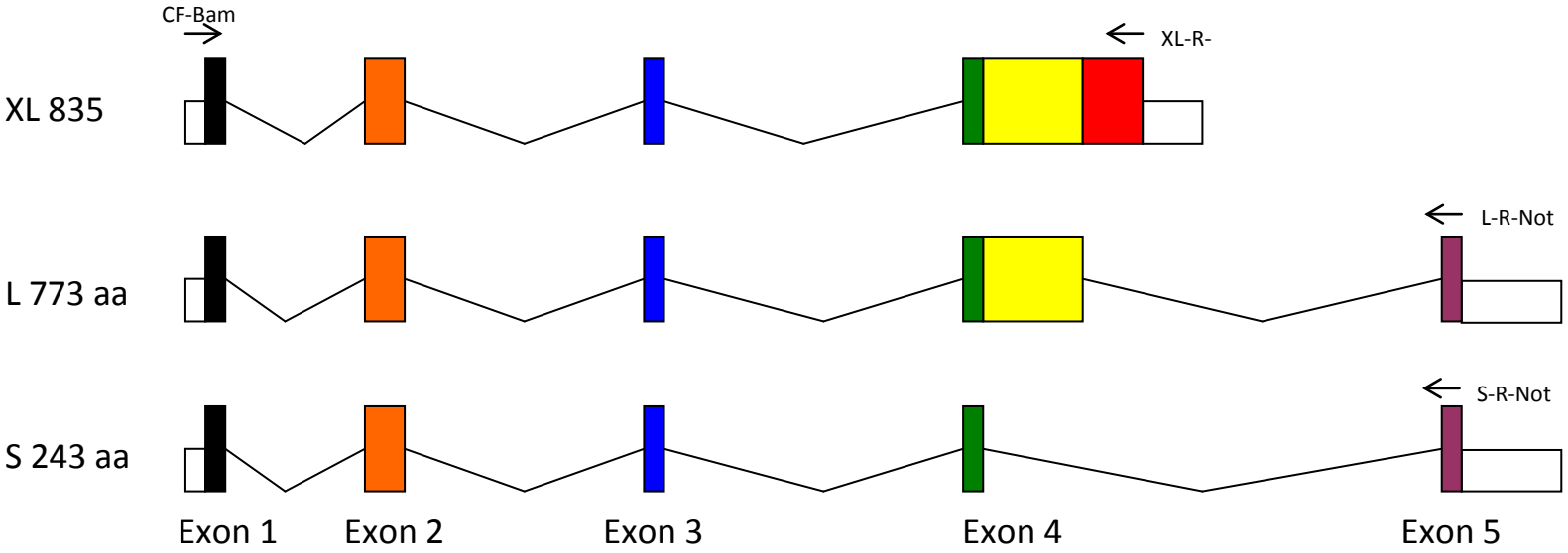


Figure 11.1 Design and RT-PCR primer position for the *BCL11A* over-expression constructs

Table 11.7 Amplification primers for *BCL11A* isoforms

Primer ID and target isoform	Primer sequence
CF-Bam: (common forward primer for all forms)	ttattggatccaccatgtctcgccgcaagcaaggcaaac
S-R-NotI: (reverse primer for S form)	ttattgcggccgcatcaaattttctcagaacttaagggc
L-R-NotI: (reverse primer for L form)	ttattgcggccgcatcagaacttaagggctctcgagctt
XL-R-NotI: (reverse primer for XL form)	ttattgcggccgcactattcagtttttatatcattattc

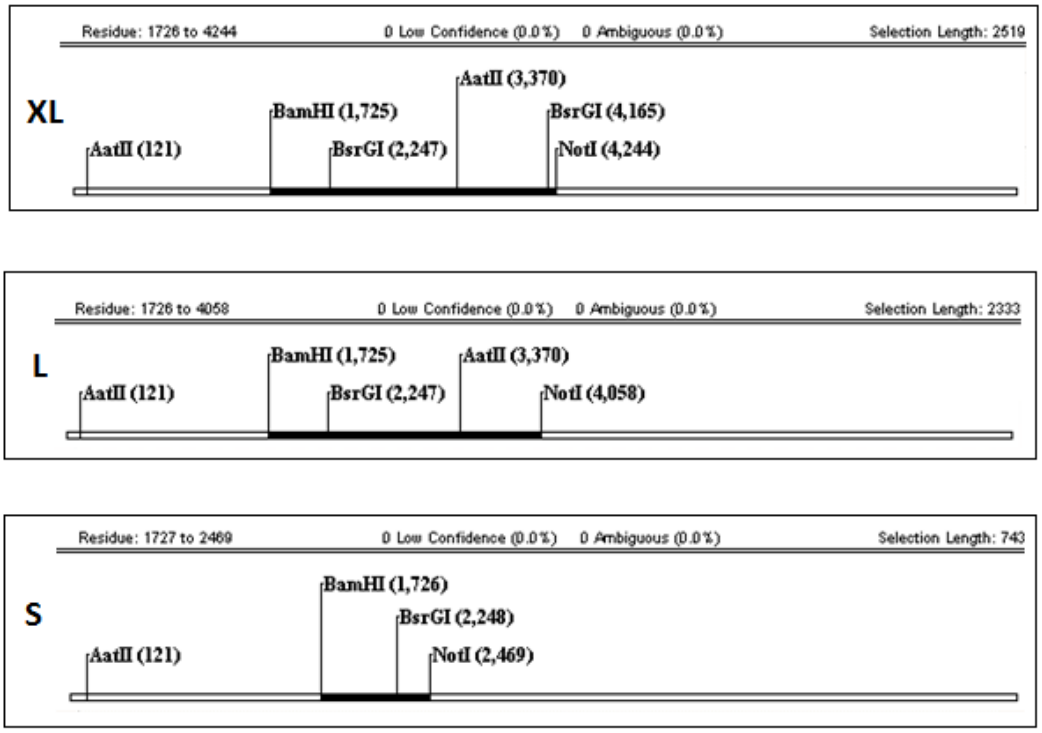


Figure 11.2 Schematics of the pEF6 constructs with BCL11A inserts are shown for each of the isoforms. The dark regions show the BCL11A cDNAs, flanked by the Bam*HI* and Not*I* cloning sites. The clear regions are pEF6 vector backbone. As part of the test digests to check for vector integrity, vector-specific enzymes were also used to assess proper insert orientation (Figure 11.3)

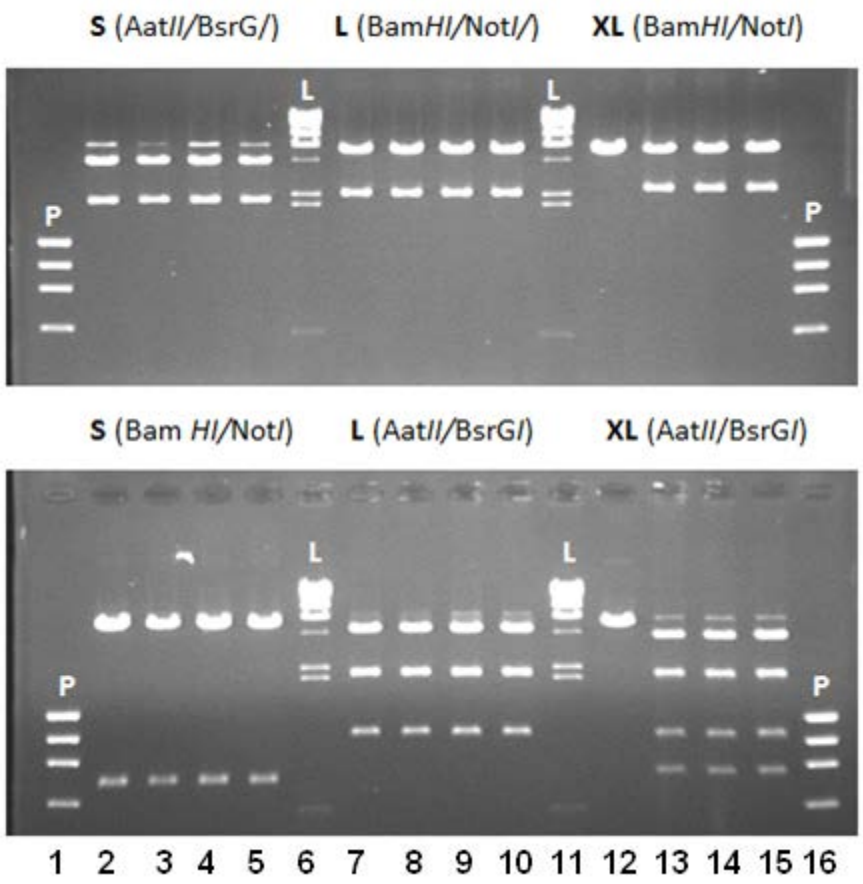


Figure 11.3 Restriction digests of 4 plasmid preps from each of the pEF6-cloned BCL11A isoforms. The Bam*HI*/Not*I* digests release the full-length cDNA (XL 2519 bp; L 2333 bp; S 743 bp), leaving the larger pEF backbone (5.8 kb). The AatII/BsrGi digests check for insert orientation: (expected bands if correct: XL: 4243, 2126, 1123 bp; L: 4852, 2126, 1123 bp; S: 4391, 2126 bp). Having no insert, XL clone number 1 is the only one to be incorrect. PhiX/Hae*III* DNA ladders (P) are on the outside lanes and Lambda/Hind*III* (L) are on lanes 6 and 11 on each panel.

Appendix 3- Lucifrase reporter constructs

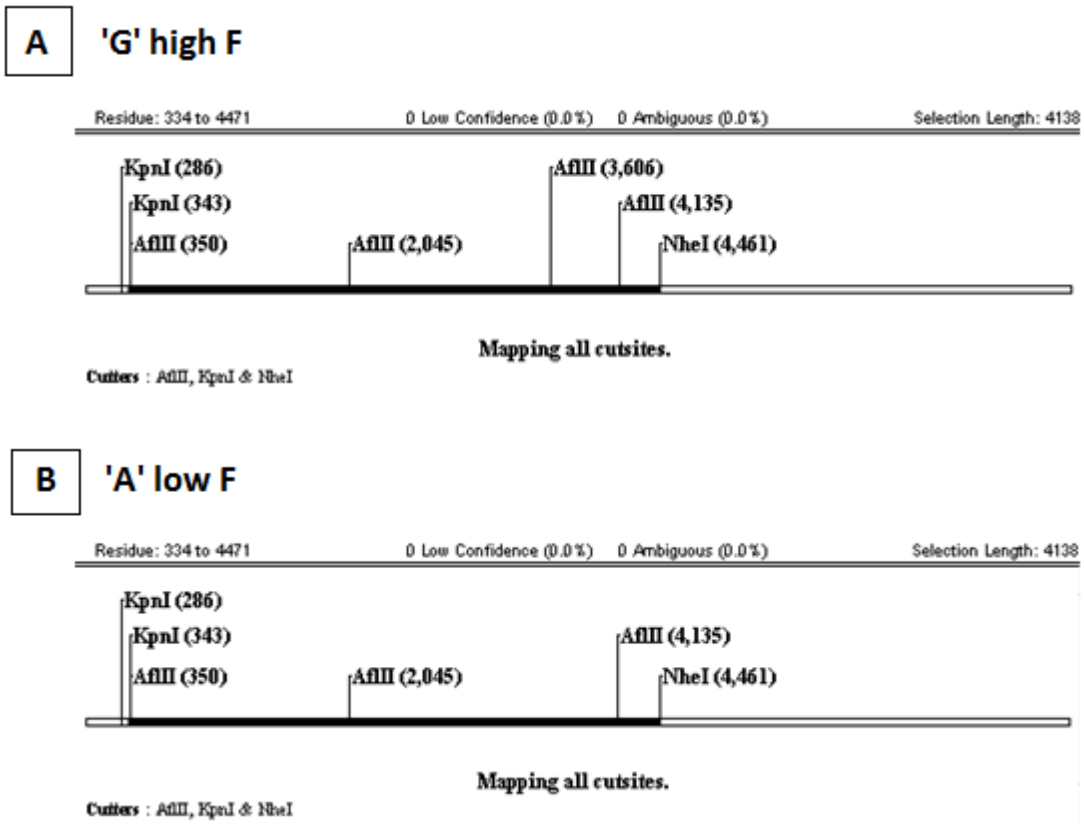


Figure 11.4 Verification of the Enhancer 1 construct. The "high" HbF (A) and the "low" HbF (B) individual can be distinguished by the presence or absence of an Afl*II* restriction site within the SNP rs10172646 (here at position 3,606 of the cloned PCR fragment, which is shown by the black bar between the cloning sites Kpn*I*/Nhe*I*). The cutter denotes the "G" genotype of the "high" HbF individual 3180 (upper panel-A) and the non-cutter denotes the "A" genotype of the "low" HbF individual 4041 (lower panel-B).

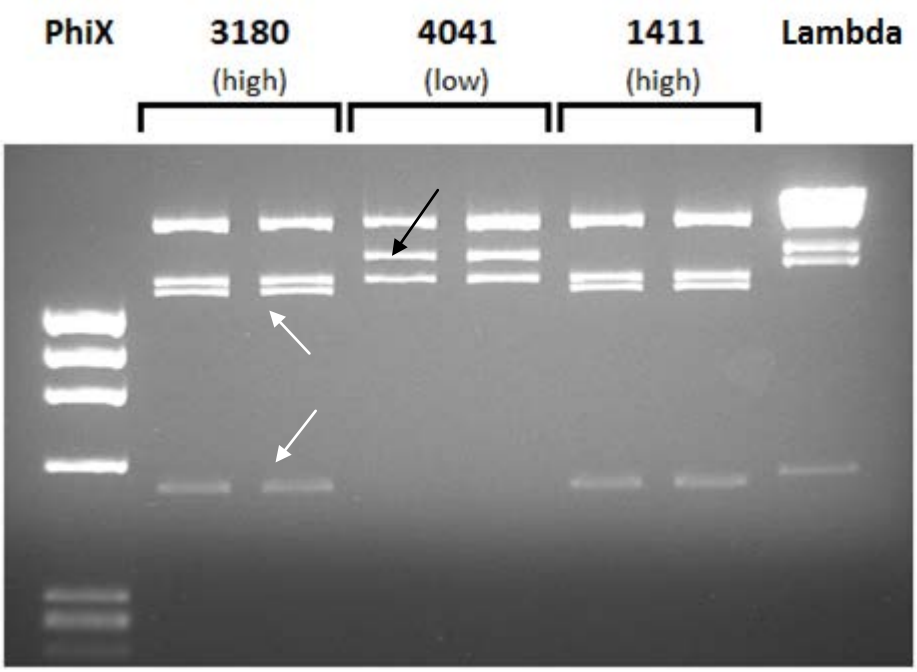


Figure 11.5 Enhancer 1 genotype verification at SNP rs10172646. A 2089 bp AflI fragment (indicated by a black arrow) is cleaved by AflI if there is a G nucleotide (the "high" genotype) into a 1561 and 529 bp fragments (indicated by white arrows). The "low" genotype, with an A nucleotide, does not cut. Samples 3180 and 4041 were used in subsequent experiments.

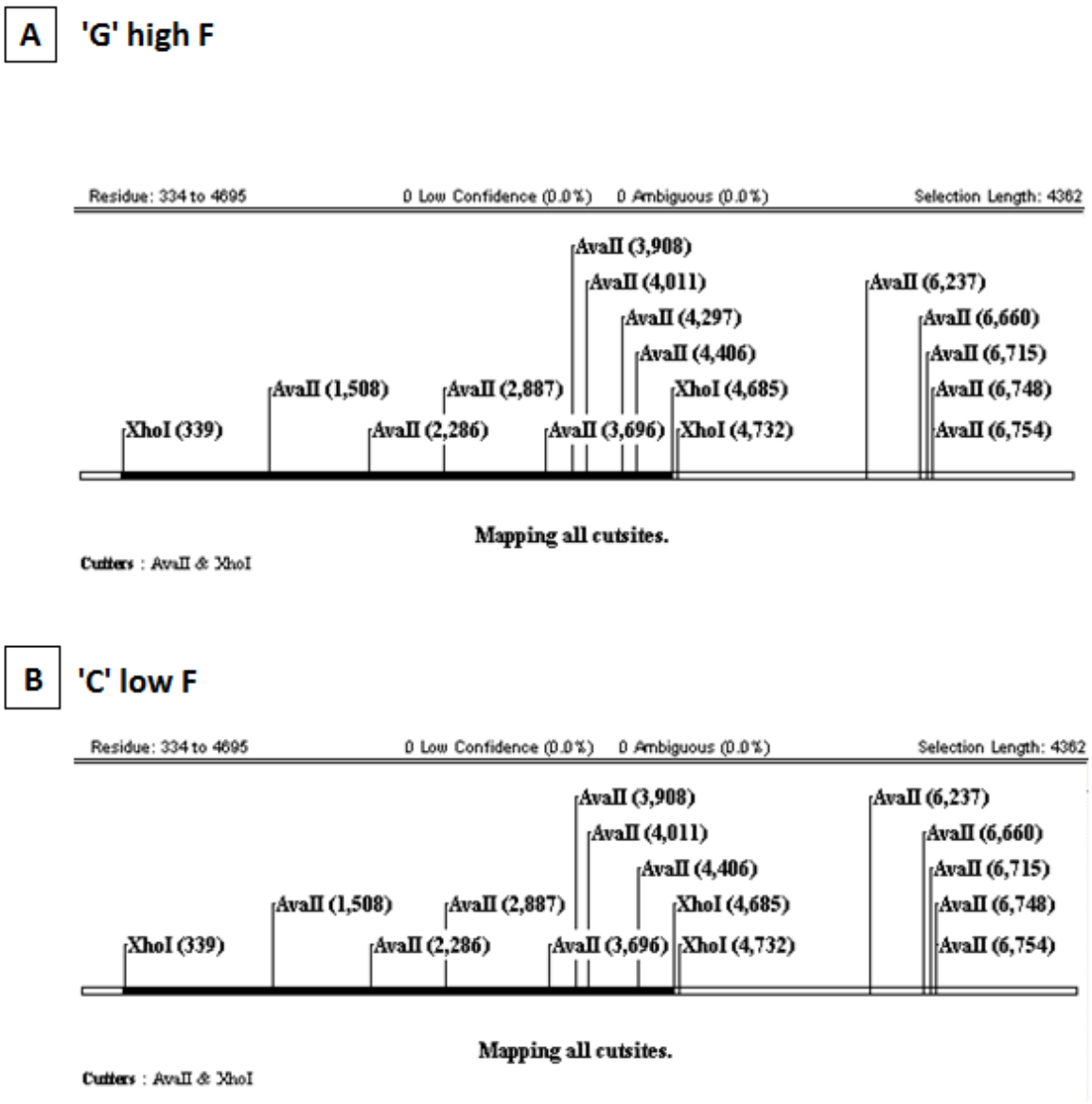


Figure 11.6. Verification of the Enhancer 2 construct, the "high" HbF (A) and the "low" HbF (B) individual can be distinguished by the presence or absence of an Ava*II* restriction site within the SNP rs7606173 (here at position 4,297 of the cloned PCR fragment, which is shown by the black bar between the Xho*I* cloning site). The cutter denotes the "G" genotype of the "high" HbF individual 3180 (upper panel-A) and the non-cutter denotes the "C" genotype of the "low" HbF individual 4041 (lower panel-B).

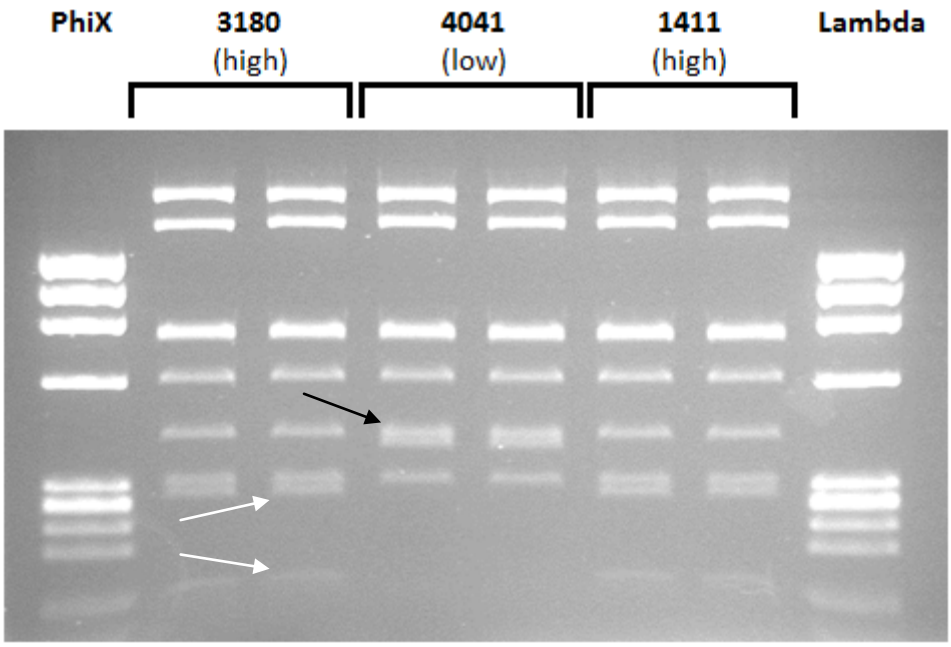


Figure 11.7 Enhancer 2 genotype verification at SNP rs7606173. A 395 bp Ava*II* fragment (indicated by a black arrow) is cleaved by Ava*II* if there is a G nucleotide (the "high" genotype) into a 286 and 109 bp fragments (indicated by white arrows). The "low" genotype, with a C nucleotide, does not cut. Samples 3180 and 4041 were used in subsequent experiments.

Appendix 4- Patient karyotypes for qPCR studies in Chapter 7

Table 11.8 AML with Monosomy 7 (-7) or del (7q) (*n*=30)

Patient	DOB	Sample date	Date received	Sample type	RNA condition	Karyotype
FP	21-Dec-78	07/06/2007	07/06/2007	Marrow	Good	45,XX,-7[6]/ 46,XX[4]
KGC	21-Jan-91	16/08/2007	16/08/2007	Marrow	High	45,XX,-7[12]/ 46,XX[1]
RP	06-Aug-35	28/08/2007	29/08/2007	Marrow	Good	45,XY,-7[10]/ 46,XY[2]
SH	30-Sep-53	05/02/2008	05/02/2008	Marrow	High	46,XY,del(7)(q21q36)[6]/ 46,XY,del(7)(q22q34)[4]/ 46,XY,del(7)(q22q36)[3]/ 46,XY,del(5)(q13q33),del(7)(q22q36)[9]/ 46,XY[2]
CH	03-Sep-38	03/11/2008	03/11/2008	Blood	High	45,XX,t(3;6)(q26;q25),-7[10] Metaphase FISH: EVI1 gene rearrangement positive[5]
BR	08-Jan-52	29/01/2009	29/01/2009	Marrow	V.High	46,XY,del(7)(q3q3)[12]/ 46,XY[1]
NJ	26-Feb-42	25/11/2009	16/11/2009	Marrow	V.High	45,XY,del(6)(q1q2),-7[9]/ 46,XY[1] No evidence of a MLL gene rearrangement [106]
KK	13-May-36	14/10/2010	15/10/2010	Marrow	Good	45,XY,add(1)(p3),-7[7]/ 46,XY,add(1)(p3),-7,+mar[2]/ 46,XY,del(20)(q11q13)[2]/ 46,XY[5]
LW	10-Feb-28	04/12/2007	04/12/2007	Marrow	Good	45,XY,-7[9]/ 46,XY[1]
JVR	13-May-35	18/08/2008	18/08/2008	Marrow	Good	47,XX,del(7)(q22q3),+8[10]

…Continued

SLC	08-Apr-94	28/11/2008	28/11/2008	Marrow	V.High	46,XX,del(7)(q2q3),t(16;16)(p13;q22)[15] CBFB/MYH11 gene rearrangement [5] Metaphase FISH analysis
RG	08-Apr-77	29/09/2009	29/09/2009	Marrow	V.High	46,XX,del(7)(q2q3),t(16;16)(p13;q22)[15] CBFB/MYH11 gene rearrangement [5] Metaphase FISH analysis
PW	08-Jul-38	25/01/2010	26/01/2010	Marrow	High	46,XY,del(7)(q22q32),inv(16)(p13q22)[9]/ 46,XY[1]
JAL	29-Nov-00	28/05/2010	29/05/2010	Marrow	V.High	48,XY,del(7)(q31q36),der(16)del(16)(p13p13)inv(16)(p13q22) del(16)(q22q22),+22,+r[10] CBFB/MYH11 positive Metaphase FISH analysis
DD	06-Oct-41	09/09/2010	10/09/2010	Marrow	High	46,XY,del(7)(q2)[5]/ 46,XY[64] Loss of one copy of D7S522 [4/99] Interphase FISH analysis
CCS	06-Mar-96	08/03/2007	08/03/2007	Marrow	V.High	45,X,-Y,t(8;21)(q22;q22)[3]/ 45,X,-Y,del(7)(q31q34),t(8;21)(q22;q22)[6]/ 46,XY[1]
DD	27-Jun-24	22/10/2007	22/10/2007	Marrow	Good	45,XX,-7,+14,i(14)(q10)[10]
CM	02-Apr-42	20/03/2008	20/03/2008	Marrow	Good	46,XY,del(7)(q2q3)[9]/ 46,XY[1]
RMN	19-Sep-42	24/03/2008	25/03/2008	Marrow	High	46,XY,t(15;17)(q22;q21)[2]/ 45,XY,del(7)(q1),-14,t(15;17)(q22;q21),der(22)t(14;22)(q1;p1)[5]/46,XY[3] Interphase FISH: PML/RARA gene rearrangement positive [21/105] 13q14 deleted x1 [19/110]
HH	28-Jul-32	23/06/2009	24/06/2009	Marrow	High	46,XY,-3,add(3)(q1),-7,dic(13;18)(p1;p1),+mar1,+mar2,+r[8]
SS	31-Aug-68	08/01/2008	08/01/2008	Blood	Good	45,XY,-5,add(6q),-7,add(12)(p1),-22,+2mar[3]/ 45~46,XY,-3,-5,add(6)(q1),-7,add(12p),-15,-21,add(22)(q13), +mar1x2~3,+2mar[cp5]/ 46,XY[3]
RG	13-Apr-46	05/09/2008	05/09/2008	Marrow	Good	45,XY,-7[13]/46,XY[2]
KS	13-Mar-43	11/09/2008	12/09/2008	Marrow	High	45,XY,-7[10]

…Continued

SRC	29-Oct-33	13/02/2008	13/02/2008	Marrow	Good	45,XX,-7[13]/46,XX[2]
KJ	15-Feb-92	06/11/2008	07/11/2008	Marrow	Good	5,XX,-7[8]/46,XX[2]
MA	28-Jan-46	23/12/2008	23/12/2008	Marrow	V.High	45,XY,t(3;3)(q21;q26),-7[9]/46,XY[1] EVI1 rearrangement positive [5/5], Metaphase FISH
GF	12-Aug-45	03/06/2009	04/06/2009	Blood	High	45,XY,add(5)(q1), add(7)(q2), add(12)(p1),-20[8]/ 45,XY,add(5)(q1), -7,add(12)(p1),-20,+mar[3]/ 45,XY,add(3)(q1),add(5)(q1),add(7)(q2) ,add(12)(p1),-20[3]/ 46,XY[1]
EAJ	11-Mar-05	04/09/2009	05/09/2009	Marrow	High	45,XY,-7[20]
KM	19-Mar-40	19/02/2010	19/02/2010	Marrow	V.High	45,XY,-7[10]
JA	24-Sep-40	11/06/2008	12/06/2008	Marrow	Good	45,XX,-7[4]/ 46,XX[6]

Table 11.9 JMML samples (*n*=6)

Patient ID	DOB	Sample date	Date of receipt	Sample type	RNA condition	Karyotype
MI	10-Feb-05	19/07/2007	19/07/2007	Blood	Good	46,XY[20]
NJEC	11-Mar-06	22/11/2007	23/11/2007	Marrow	Good	46,XX[20]
RAK	31-Mar-09	30/07/2010	31/07/2010	Marrow	Good	46,XX[20]
SL	28-Apr-03	15/09/2005	15/09/2005	Marrow	Very high	46,XX[20]
IH	13-May-03	03/01/2006	03/01/2006	Marrow	High	46,XY,-6,+r[3]/ 46,XY[37]
TJH	22-Jan-04	05/10/2004	06/10/2004	Marrow	Good	45,XY,-7[18]/ 46,XY[2]

Table 11.10 CMML samples (*n*=10)

Patient ID	DOB	Sample date	Date of receipt	Sample type	RNA condition	Karyotype
MP	11-Mar-36	06/04/2009	07/04/2009	Marrow	V.High	46,XX[20]
TN	02-Jun-45	22/04/2010	23/04/2010	Marrow	V.High	46,XY[20]
KS	08-Oct-40	16/10/2009	16/10/2009	Blood	High	PDGFRA gene rearrangement negative [110] PDGFRB gene rearrangement negative [96] Interphase FISH analysis
LA	22-May-27	10/08/2009	12/08/2009	Marrow	High	46,XY[20]
BS	10-Jun-37	08/09/2009	08/09/2009	Marrow	High	46,XY[20]
VN	04-Jun-23	01/02/2010	01/02/2010	Marrow	V.High	46,XX[20]
NJB	20-Aug-43	02/02/2010	03/02/2010	Blood	High	46,XY[20]
BK	05-Sep-30	09/04/2010	10/04/2010	Blood	Good	46,XX[20]
IM	18-Jun-31	05/05/2010	06/05/2010	Blood	Good	46,XX[20]
MP	08-May-38	12/05/2010	13/05/2010	Marrow	V.High	46,XY[20]

Table 11.11 AML with normal karyotype

Patient ID	DOB	Sample date	Date received	Sample type	RNA Condition	Karyotype
DD	24-Dec-55	01/06/09	01/06/09	Marrow	High	46,XY[20]
HA	12-Jul-53	28/08/09	28/08/09	Marrow	V.high	46,XX[20]
FD	11-Jun-39	15/06/09	16/06/09	Marrow	V.High	46,XX[20]
LG	23-Dec-64	18/06/09	18/06/09	Marrow	High	46,XX[20]
HB	15-Apr-45	06/07/09	07/07/09	Marrow	Good	46,XX[20]
RG	02-Apr-36	08/07/09	09/07/09	Marrow	V.High	46,XX[20]
FN	09-Aug-45	13/07/09	14/07/09	Marrow	V.High	46,XY[20]
GH	13-Oct-44	14/07/09	15/07/09	Blood	High	46,XY[20]
LG	25-Nov-34	27/07/09	28/07/09	Marrow	V.High	46,XY[20]
MD	26-Sep-49	17/08/09	17/08/09	Marrow	Good	46,XY[20]

Table 11.12 Control group

Patient ID	DOB	Sample date	Date received	Sample type	RNA condition
CB	16-Sep-51	13/03/08	13/03/08	Marrow	good
CBH	06-Oct-43	27/06/08	27/06/08	Marrow	high
ES	06-Sep-40	12/08/08	13/08/08	Marrow	good
AW	26-Mar-01	04/06/09	04/06/09	Marrow	v.high
BW	10-Feb-36	16/09/09	16/09/09	Marrow	high
CB	07-Sep-41	25/09/09	25/09/09	Marrow	high
EM	09/01/1997	09/05/08	09/05/08	Marrow	high
WC	27/11/1934	17/03/08	17/03/08	Marrow	high
EM	27-Mar-44	23/12/08	23/12/08	Marrow	good
DH	22-Aug-27	21/09/10	21/09/10	Marrow	v.high

Appendix 5- List of abbreviations

ACH	Active chromatin hub
AcH3	Acetylated histone H3
AcH4	Acetylated histone H4
αHS40	α-globin locus DNase I hypersensitive site 40
AML	Acute myeloid leukaemia
βHS3	β-globin LCR DNase I hypersensitive site 3
BFU-e	Burst forming unit erythroid
Bp	Base pairs
CFU-e	Colony forming unit erythroid
ChIP-chip	Chromatin immunoprecipitation on microarray
CMML	Chronic myelomonocytic leukaemia
Cy3	Cyanine 3
Cy5	Cyanine 5
Enh	Enhancer
EPO	Erythropoietin
ERC	Erythropoietin responsive cells
F	Fetal haemoglobin
FACS	Fluorescence activated cell sorting
FC	F-cells
FCS	Foetal calf serum
FOG-1	Friend of GATA-1
GM-CSF	Granulocyte-macrophage colony stimulating factor
GPA	Glycophorin A
HbA	Adult haemoglobin ($\alpha_2\beta_2$)
HBA	α-globin

HbA2	Haemoglobin minor ($\alpha_2\delta_2$)
HbAS	Sickle trait, heterozygous for HbS
HBB	β-globin
HBBP1	βψ-pseudogene
HBD	δ-globin
HBE	ε-globin
HbF	Foetal haemoglobin ($\alpha_2\gamma_2$)
HBG	γ-globin
HBM	ψα2-pseudogene
HBQ1	θ-globin
HbS	Sickle haemoglobin, $\alpha_2\beta^s{}_2$
HBZ	ζ-globin
HMIP	*HBS1L-MYB* intergenic polymorphism
Ho	Human orthologous
HPC	Haematopoietic progenitor cell
HPFH	Hereditary persistence of foetal haemoglobin
HS	Hypersensitive site (to Dnase I)
HSC	Haematopoietic stem cell
IL-3	Interleukin 3
JMML	Juvenile myelomonocytic leukeamia
Kb	Kilo base pairs
KLF	Krüpple like factor-1
LCR	Locus control region
LD	Likage disequilibrium
LINE	Long interspersed nuclear element
LME	L-leucine methyl ester
LTR	Long terminal repeat
Mb	Mega base pair

MDS	Myelodysplastic syndrome
MO7	Monosomy 7
MYB	Myeloblastosis oncogene
NEFM	Neurofilament medium gene
PB	Peripheral blood
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline-Tween
PCR	Polymerase chain reaction
Q-PCR	Quantitative PCR
QTL	Quantitative trait locus
RNAP II	RNA polymerase II
RT-PCR	Reverse transcription PCR
SCD	Sickle cell disease
SCF	Stem cell factor
SNP	Single nucleotide polymorphism
TFBS	Transcription factor binding site
TF	Transcription factor
TSS	Transcriptional start site
UTR	Untranslated region
YAC	Yeast artificial chromosome
(del) 7q	Deletions within the q arm of chromosome 7