Genetic basis of hereditary persistence of fetal haemoglobin

Alice E Gallienne

A thesis submitted in partial fulfilment of the requirements of Oxford Brookes University for the degree of Doctor of Philosophy

This research was carried out in collaboration with the Oxford Radcliffe NHS Trust

October 2013

ABSTRACT

This thesis investigated the extent to which genetic factors underlie the variations observed in fetal haemoglobin (HbF) levels. This is important as it is known that an elevated HbF level can ameliorate the symptoms of many of the haemoglobinopathies.

The frequency and range of hereditary persistence of fetal haemoglobin (HPFH) in the UK population was determined which was otherwise unknown. All 4 categories of deletion mutations were identified and demonstrated the high frequency of deletional HPFH in UK patients. Four potentially novel deletions were identified and 2 fully characterised. One novel deletion was the first reported case of a large β^0-thalassaemia deletion mutation in the Afghan population. Mutations in the γ-globin gene promoters were identified as a frequent (21% of patients) cause of non-deletion HPFH in the UK. The majority of mutations being in white British individuals with elevated HbF levels only, probably arisen independently through genetic drift. The strongest association with the three polymorphisms and HbF expression was seen in β-thalassaemia trait subjects with the *XmnI-HBG2* polymorphism. The SNPs in *BCL11A* and *HBS1L-MYB* failed to show statistical correlations with HbF. Heterozygosity for ten novel mutations in the *KLF1* gene were indentified in patients with a high HbF indicating that a single altered *KLF1* allele can elevate HbF. The identification of a *KLF1* mutation in an individual with a particularly mild form of sickle cell disease could provide *in-vivo* evidence that controlled reduction of *KLF1* expression could be an effective treatment for sickle cell disease. Finally, the finding that Asian Indian newborns undergo haemoglobin switching earlier than other ethnic groups was investigated. Birth weight, gestation and chromosomal abnormalities were not responsible and the frequency of the 4.9kb γ-thalassaemia deletion mutation was determined to be low.

ACKNOWLEDGEMENTS

Firstly I would like to thank my supervisors. I am very grateful to Dr John Old for giving me the opportunity to carry out this research in his laboratory and count it a great privilege to have been able to work with him. Thanks to Dr Shirley Henderson for guiding me through the process of scientific writing and for her enthusiasm for my work. Special thanks to Dr Tony Bradshaw who has been behind me all the way and been a constant source of encouragement. His help and kindness I will not forget.

Thanks to all my colleagues in the Molecular Haematology Department who have been very supportive and to the staff in the Life Sciences Department at Oxford Brookes University.

Finally and most importantly I want to thank Olly, Zoë and Sam who have tirelessly encouraged me and got me through some of the most difficult times. To my family and friends I must say a huge thank you for your understanding and putting up with me never being around.

LIST OF ABBREVIATIONS

ANOVA	analysis of variance
ARMS	amplification refractory mutation system
BCL11A	B-cell lymphoma/leukemia 11A
bp	base pair
CAP	catabolite activator protein
^{o}C	degrees centigrade
CDA	congenital dyserythropoietic anaemia
CP1	Cysteine proteinase-1
CDP	CCAAT displacement protein
cnLOH	copy neutral loss of heterozygosity
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTPs	deoxynucleotide triphosphates
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EKLF	erythroid kruppel-like Factor
kb	kilobase
KCL	potassium chloride
$KHCO_3$	potassium hydrogen carbonate
KLF	kruppel-like factor
FBC	full blood count
FC	f cell
fl	femtolitre
FOG	friend of GATA
g	gram
g/dl	gram per decilitre
g/L	gram per litre
GATA	GATA binding-factor
Glu	glutamic acid
GWAS	genome wide association study
Hb	haemoglobin
HbA	adult haemoglobin
HbF	fetal haemoglobin
HbS	sickle haemoglobin
HBS1L	HBS1-like (*S. cerevisiae*)
Hct	hematocrit
HGVS	human genome variation society
HMIP	HBS1L-MYB intergenic polymorphism
HPFH	hereditary persistence of fetal haemoglobin
HPLC	high performance liquid chromatography
HS	Dnase I hypersensitive site
HS 1-5	Dnase I hypersensitive site 1-5
HS-40	Dnase I hypersensitive site -40
HVR	hypervariable region
IVS	intervening sequence
IEF	isoelectric focusing
LCR	locus control region
Lys	lysine

M	molar
Mb	megabase
MCV	mean cell volume
MCH	mean cell haemoglobin
mg	milligram
µg	microgram
min	minute
µl	microlitre
mM	millimolar
µM	micromolar
ml	millilitre
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulphate
MLPA	multiplex ligation-dependent probe amplification
mRNA	messenger RNA
MYB	myeloblastosis gene
NaCl	sodium chloride
NaOH	sodium hydroxide
NF-E2	nuclear factor erythroid 2
NF-E3	nuclear factor erythroid 3
ng	nanogram
NH_4Cl	ammonium chloride
OD	optical density
OCT-1	octamer transcription factor
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pg	picogram
pI	isoelectric point
pmole	picomole
QTL	quantitative trait locus
RBC	red blood cell
RE	reticuloendothelial system
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
SD	standard deviation
SDS	sodium dodecyl sulphate
sec	second
SIFT	software implemented fault tolerance
SNP	single nucleotide polymorphism
SOX6	SRY (sex determining region Y)-box 6
Sp1	specificity protein 1
SS	sickle cell disease
SSP	stage selector protein
TBE	tris-borate-EDTA
TE	tris-EDTA
TF	transcription factor
TRIM	tripartite motif gene
TRIS	2-amino-2-(hydroxyl)-1, 3-propanedol

U	unit
UCSC	university of California, Santa Cruz
UK	United Kingdom
UPD	uniparental disomy
UV	ultra violet
V	volt
Val	valine
WHO	world health organistion
Zn	zinc

LIST OF FIGURES

LIST OF TABLES

CONTENTS

CHAPTER ONE: INTRODUCTION

1.1 Introduction

Fetal haemoglobin (HbF, $\alpha_2\gamma_2$) is the major haemoglobin (Hb) produced during intrauterine life. As both fetal and adult Hb (HbA, $\alpha_2\beta_2$) contain α-globin chains, the switch from fetal to adult Hb production is essentially that of the replacement of γ-globin with β-globin gene expression. After birth the percentage of HbF progressively falls over several months to less than 1% of the total haemoglobin (97% HbA, 3% HbA_2). By six months of age mutations which affect the β-globin gene such as β-thalassaemia and sickle cell disease become clinically apparent. However, in some individuals the HbF levels do not fall to <1% and remain elevated. These conditions are known as hereditary persistence of fetal haemoglobin (HPFH) and $\delta\beta$-thalassaemia. In a proportion of cases the genetic reasons for HPFH disorders are understood; however, in many cases the molecular basis of the HPFH disorder remains obscure.

The degree of HbF persistence varies greatly between individuals and this variability is largely genetically controlled (Garner *et al*., 2000). Whilst persistence of high levels of HbF production have no clinical consequences in healthy individuals, high HbF levels have immense clinical benefits associated with milder disease progression and fewer complication in patients with sickle cell disease and β-thalassaemia (Thein and Menzel, 2009). Even a relatively modest increase in production of HbF is known to significantly benefit haemoglobinopathy patients. Major advances have been made in our understanding of the molecular mechanisms and pathophysiology of these haemoglobinopathies. However, problems still arise in genetic counselling and

clinical management as predicting the disease severity remains difficult. The ability to be able to identify the different genetic factors involved could provide more precise estimates of disease severity and increasing our knowledge of the molecular mechanisms could reveal new targets for therapeutic intervention (Thein, 2008). With this in mind, this project was aimed at trying to establish the genetic reasons responsible for the variation in HbF levels seen in the UK population and provide further understanding of the molecular mechanisms involved.

1.2 Haemoglobin

Haemoglobin is a vehicle for transporting oxygen. It picks up oxygen as the red cell passes through the lungs and releases oxygen when the red cell reaches the capillaries of the tissues.

1.2.1 Structure and genetic control of normal human haemoglobin Human haemoglobin was one of the first proteins to have its amino acid structure and DNA sequence identified and has been intensely studied partly due to its abundance in blood. All normal haemoglobin molecules have a tetrameric structure consisting of two different pairs of identical peptide chains, either α-like or β-like globin chains. The α-like peptide chains (α and ζ) are made up of 141 amino acids whereas the β-like chains (ε, γ, δ and β) each consist of 146 amino acids. The structure of haemoglobin changes throughout development (Figure 1.1) giving rise to embryonic, fetal and adult haemoglobins. In the embryo there are three types of haemoglobin, HbPortland ($\zeta_2\gamma_2$), HbGower 1 ($\zeta_2\varepsilon_2$) and HbGower 2 ($\alpha_2\varepsilon_2$). Fetal haemoglobin ($\alpha_2\gamma_2$) replaces the embryonic

haemoglobins by 10 weeks post-conception (Figure 1.1). HbF is comprised of two different types of γ-chains, $^{G}\gamma$ and $^{A}\gamma$, differing only at amino acid position 136 which is either glycine or alanine respectively. Adult haemoglobin (HbA) has two α-chains combined with two β-chains ($\alpha_2\beta_2$). The β-globin gene is expressed at low levels during early fetal life with the main switch to adult haemoglobin occurring 3-6 months after birth (Figure 1.1). In addition to HbA, adults also have a minor haemoglobin component HbA_2 ($\alpha_2\delta_2$). HbA_2 is a haemoglobin without an obvious physiological function and it has been suggested that the δ-gene will eventually end up as a pseudogene (Steinberg & Nagel, 2009) Normal adult haemoglobin consists of HbA (~96%), HbA_2 (~3%) and a small amount of HbF (~1%).

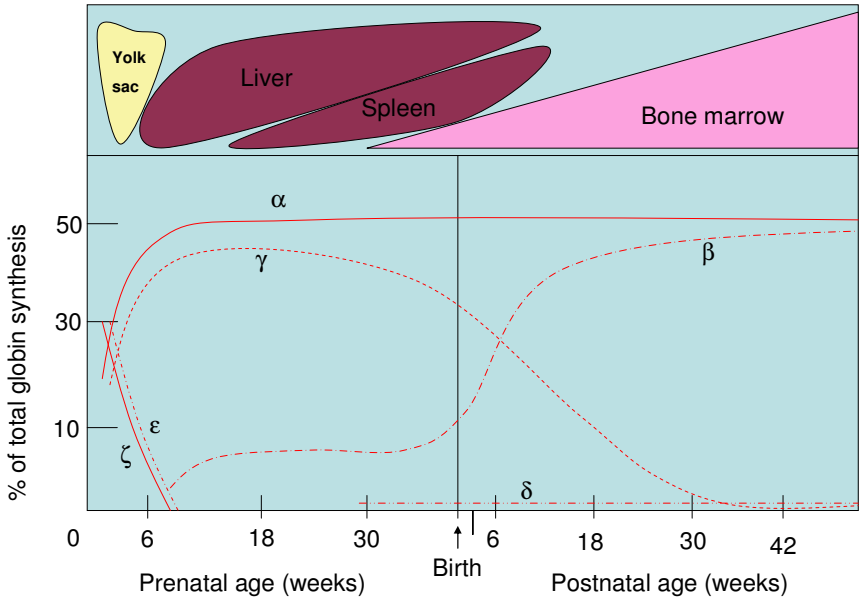


Figure 1.1 Synthesis of individual globin chains in prenatal and postnatal life. The sites of haemopoiesis at different stages of development and the levels of expression of the different globin chains at various gestational ages are shown (adapted from Hoffbrand *et al.*, 1993).

Genetic analysis has established that the genes responsible for globin synthesis lie in two developmentally regulated multigene clusters on separate chromosomes known as the α-globin and the β-globin clusters.

1.2.2 α-globin cluster

The α-globin cluster is over 70 kilobases long and is on the short arm of chromosome 16 (16p13.3). The cluster has three functional genes ζ2, α1 and α2, two non-transcribed pseudogenes ψζ1, ψα1, the θ and recently designated α^D gene. Both θ and α^D are transcribed but no protein or haemoglobin products have been detected (Albitar *et al*., 1992). The pseudogenes are highly homologous to their corresponding genes but they do not direct protein synthesis due to changes within their sequences acquired during evolution. The genes are arranged on the cluster in the order in which they are expressed during development 5'- ζ2,- ψζ1- α^D- ψα1- α2 - α1- θ -3' (Figure 1.2). There are normally two functional copies of the α-globin gene per chromosome in humans and therefore normal individuals will inherit four α-globin genes in total, written as αα/αα. Around 40kb upstream of the ζ2 gene lies the HS-40 site which is an important binding site for several transcription factors.The HS-40 site was found to be an enhancer and critical for the expression of the downstream α-globin genes.

1.2.3 β-globin cluster

The β-globin cluster spans a region of 50kb and is located on the short arm of chromosome 11 (11p15.5). It has five functional genes ε, $^G\gamma$, $^A\gamma$, δ, β and one non-functional gene ψβ. Like the α-globin cluster the genes are arranged in the order in which they are expressed during development 5'-ε-$^G\gamma$-$^A\gamma$-ψβ-δ- β-3'

(Figure 1.2). Upstream of the β-globin cluster is the β-locus control region (β-LCR) which consists of five DNase I hypersensitive sites (HS1-5) (Tuan *et al.*, 1985). The β-LCR is important for the expression of the β-like globin genes as it maintains an open chromatin domain and acts as an enhancer element; in its absence the level of gene expression is low.

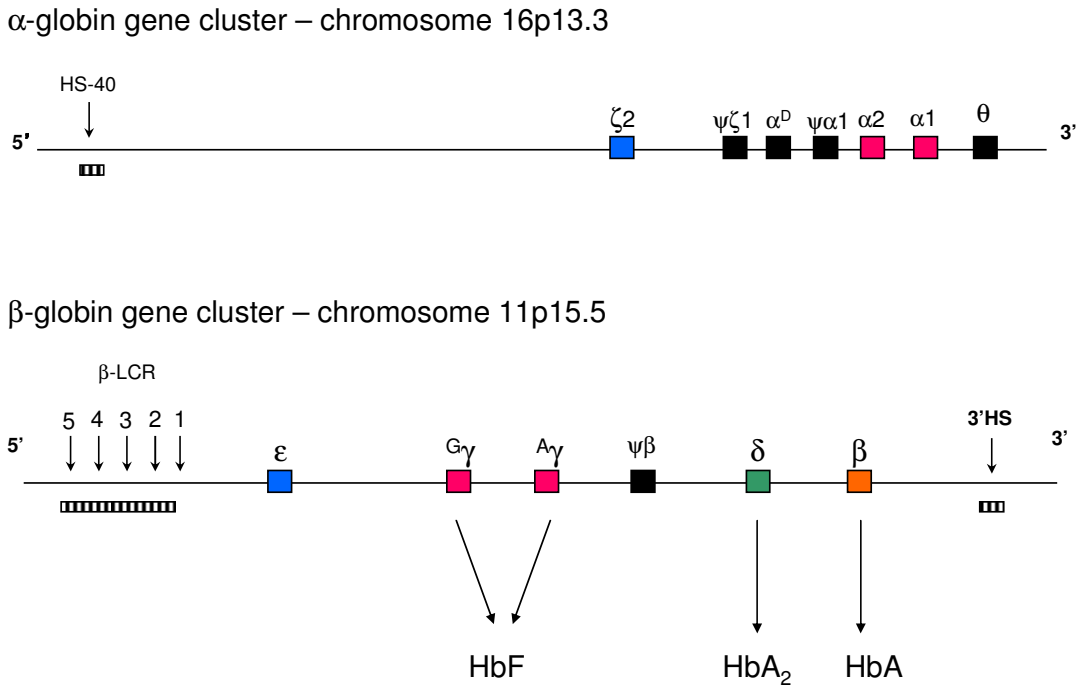


Figure 1.2 The α- and β-globin clusters. HS-40 = DNase1-hypersensitive site at -40kb relative to the ζ-globin gene. β-LCR= β-locus control region.

1.2.4 Haemoglobin switching

An important and extensively studied event that occurs in the β-globin gene cluster is the so-called haemoglobin switching process, the suppression of the γ-globin genes accompanied by the complementary rise in the expression of the previously silent β-globin gene (Stamatoyannopoulos *et al.*, 2001). While the α-like globin genes have a single switch (embryonic ζ → fetal/adult α), the β-globin genes undergo two switches (embryonic ε → fetal γ → adult β). In the

β-globin gene cluster the expression of the globin genes is regulated by complex interactions between the β-LCR and the promoters of the globin genes themselves. The β-LCR and the promoter regions of the globin genes are bound by either repressing or activating proteins called transcription factors (Stamatoyannopoulos, 2005).

Each HS of the β-LCR contains genetically conserved domains for the transcription factors to bind. Binding sites for transcription factors have also been identified in the promoter regions immediately upstream of each of the globin genes (Cao *et al*., 2002). These sequences include a TATA box at position -30, CCAAT box at position -70, CACCC box at position -90 and an AGATAT box at position -345 relative to the CAP site. As the β-LCR is bound by activating transcription factors (such as *GATA1* and *KLF1*) the associated chromatin then becomes activated (Fraser *et al*., 1998). Individual globin gene promoters and their associated chromatin might also become activated or repressed by the binding of different transcription factor complexes. In the later stages of erythropoiesis the β-LCR and the globin gene promoters appear to interact via a looping mechanism (Figure 1.3). This interaction causes the recruitment and activation of RNA polymerase II, which is needed to transcribe the gene into mRNA. This nuclear RNA is then processed into messenger RNA (mRNA) transported into the cytoplasm of erythroblasts and translated into the globin chains (Higgs *et al*., 2012).

It is suggested that the switch between embryonic, fetal and adult globins relies on the competition between the globin gene promoters for access to their

activating upstream regulatory elements in the β-LCR. The ability of the promoters to compete might in turn rely on changes in the activating or repressing transcription factors that they bind (Sankaran & Orkin, 2013). For example in fetal life the γ-promoter binds an activating complex that preferentially interacts with the β-LCR. Yet in adult life the γ-globin genes are bound by repressive factors (*BCL11A*) whereas the β-genes are bound with activating factors (*KLF1*).

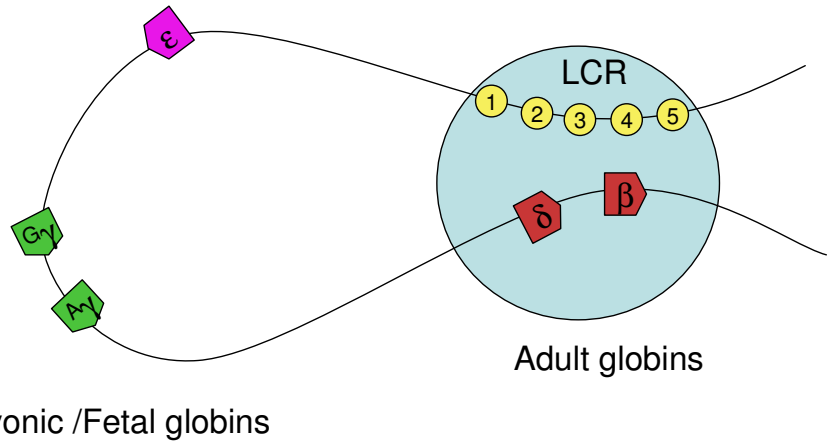


Figure 1.3 Long-range chromatin looping in the β-globin locus.
The β-LCR and adult β-globin gene come into close proximity in adult stage erythroid cells in which the gene is actively transcribed. The embryonic and fetal expressed genes that are silent at this stage of development are excluded from the close contacts. (Adapted and modified from Kiefer *et al.*, 2008)

Understanding the control of this process has obvious therapeutic implications for the therapy of β-thalassaemia and sickle cell disease. The γ-globin gene can functionally substitute for the defective β-globin gene in these diseases; therefore any advancement in the comprehension of the Hb switching process could benefit therapies based on the reactivation of the γ-globin gene (Weatherall and Clegg, 2001).

1.3 Inherited disorders of haemoglobin

The haemoglobinopathies are the most common single gene disorders in humans (Weatherall and Clegg, 2001). It is believed that their high prevalence in certain parts of the world reflects the heterozygote advantage of many of these disorders against severe malaria. The World Health Organisation (WHO) has estimated that close to 7% of the world's population i.e. 270 million are carriers for these conditions (especially α- and β-thalassaemia, HbS, HbC and HbE). There are approximately 300-400,000 affected babies born each year with severe forms of these diseases (WHO 1994). There is currently no definitive cure for any haemoglobinopathy apart from bone marrow transplantation which is expensive and available only to patients with compatible donors. Patients with these disorders often require lifelong treatment. Therefore the haemoglobinopathies pose a major health and economic burden to countries where they are common. These conditions are indigenous to tropical and subtropical regions where most of the developing countries are located however with the increasing frequency of population migration into the Western world, continents such as Australia, Europe and North America are identifying these disorders (Weatherall, 2000). Although improvements have been seen in overall infant mortality worldwide, these inherited disorders will still be an ever increasing haematological problem as these infants will require life long treatment and care.

The haemoglobinopathies are a diverse group of autosomal recessive disorders. They can be divided into two common types, the structural haemoglobin variants and the thalassaemias.

1.3.1. Structural haemoglobin variants

The structural haemoglobin variants are caused by single amino acid substitutions in either the α- or β- globin chains and the thalassaemias are a result of reduced production of one or more of the globin chains which leads to imbalanced globin chain synthesis. Over 900 structural variants have been reported so far with the majority being harmless and rare *(*Giardine *et al*., 2007) (Table 1.1). Clinically, the most important are HbS (sickle haemoglobin, β-globin codon 6, A→T, Glu→Val), HbE (β-globin codon 26, G→A, Glu→Lys) and HbC (β-globin codon 6, G→A, Glu→Lys) because they have reached high frequencies in some populations. HbS has reached high frequencies in parts of Africa, Saudi Arabia and India whereas HbC is mainly restricted to West Africa. Homozygosity for HbS (HbSS) and compound heterozygosity for HbS and HbC (HbSC) cause sickle diseases that pose significant health problems for the countries in which they are endemic. Yet, HbE which is the most common variant globally is not harmful in the homozygous state but when it interacts with β-thalassaemia it produces a thalassaemia phenotype which can range in severity from thalassaemia intermedia to transfusion dependant thalassaemia major (as explained in section 1.3.2). HbE is extremely common in Asia with frequencies reaching 50% in some areas and therefore in regions where β-thalassaemia is also prevalent this disorder has caused a significant health burden.

Table 1.1 Classification of genetic disorders of haemoglobin*

I. Structural variants – > 900 described. Mainly due to single amino-acid substitutions
II. **Thalassaemias** – Disorders due to defective and imbalanced globin production.
(1) α- thalassaemias – result from > 80 different deletions or point mutations in the α-globin genes.
1. α^0-thalassaemia (--/)
2. α^+-thalassaemia (-α/)
- Deletional (-α/)
- Non-deletional ($\alpha\alpha^T$)
(2) β-thalassaemias – result from over 200 different mutations in the β-globin genes.
1. β^0- thalassaemia
2. β^+-thalassaemia
- variants with unusually high level of HbF or HbA_2
- normal HbA_2
- 'silent'
- dominant
- unlinked to β-gene cluster
(3) $\delta\beta$-thalassaemias
1. $(\delta\beta)^+$ - thalassaemia
2. $(\delta\beta)^0$ - thalassaemia
3. $(\alpha\gamma\delta\beta)^0$ - thalassaemia
(4) γ-thalassaemia
(5) δ-thalassaemia
1. δ^0 - thalassaemia
2. δ^+ - thalassaemia
(6) $\varepsilon\gamma\delta\beta$-thalassaemia
III. Structural variants that result in a thalassaemic phenotype
IV. Hereditary persistence of fetal haemoglobin (HPFH)
(1) Deletional – $(\delta\beta)^0$
(2) Non-deletional – linked to β-globin gene cluster (γ-gene rearrangements and promoter mutations)
- unlinked to the β-globin gene cluster (*H-MYB*, *BCL11A*)

*Modified from Weatherall 2001

1.3.2 The thalassaemias

The thalassaemias are classified depending on which globin gene is defective (α, β, δβ and εγδβ) (Table 1.1). The α- and β-thalassaemias are found at high frequencies in some populations with the δβ and εγδβ being much rarer. The α- and β-thalassaemias can be further subdivided according to the amount of globin chain production. In α^+- and β^+-thalassaemia, decreased quantities of α- or β-globin chains are produced, whereas in α^0- and β^0- thalassaemia α- or β-globin production is absent. The α-thalassaemias can be divided into three clinically important types: α^+ -thalassaemia, a mild form in which only one of the linked α-globin genes is inactive (-α/); α^0-thalassaemia, a severe form in which both α-globin genes are inactive (--/); and HbH disease, compound heterozygosity of α^+- and α^0-thalassaemia in which there is only one functional α-globin gene (-α/--). The clinical picture of HbH disease is that of a chronic haemolytic anaemia of variable severity often requiring transfusion. The loss of all α-globin genes (--/--) is lethal and is referred to as HbBart's hydrops fetalis syndrome. Approximately 25% of the fetuses die *in utero* and the rest at delivery or soon after, although there are cases of hydropic infants who have been kept alive with regular transfusions (Joshi *et al*., 2004, Singer *et al*., 2000). Clinically β-thalassaemias are classified into thalassaemia major, thalassaemia intermedia and thalassaemia trait/minor according to their severity. Thalassaemia major results in a severe anaemia requiring regular blood transfusions and usually results from the interaction of two β^0 mutations or a β^0 mutation and a severe β^+ mutation. The intermedia phenotype is less severe and usually results from the interaction of two β^+ mutations or a β^+ mutation and

a β^0 mutation. The carrier state (β-thalassaemia trait) is usually asymptomatic. Clinically the most important forms of thalassaemia worldwide are β-thalassaemia major HbE/β-thalassaemia, HbH disease and HbBarts hydrops fetalis. Cases of thalassaemia have been found in most parts of the world but principally they are found in South China, Southeast Asia, the Mediterranean, Indian sub-continent, Africa and the Pacific Islands.

1.4 Fetal Haemoglobin (HbF)

The switch from γ- to β-globin gene expression at birth is also reflected by the change in the relative ratio of the two γ-globin chains. During intra-uterine life and at birth the $^{G}\gamma/^{A}\gamma$ ratio is about 70:30 but during the first months of life the ratio changes gradually to 40:60. Although HbF is replaced by adult haemoglobin by 1 year after birth, small amounts of HbF continue to be synthesized throughout adult life. This small amount of HbF is not evenly distributed but concentrated to a subset of erythrocytes termed F cells (FCs) (Boyer *et al*., 1975). FCs are not similar to fetal cells either in the amount of HbF per cell or in other characteristics which distinguish fetal and adult red cells (Wood, 1993). Studies have shown that FCs and non-FCs are derived from the same stem cell population and that FCs represent the progeny of erythroid progenitor cells that have undergone an 'accelerated' pathway of erythroid differentiation. Under normal circumstances very few cells follow this pathway but in conditions in which erythroid demand is increased, a higher proportion of erythroid cells follow such a pathway and hence a transitory increase in FCs occurs (Stamatoyannopoulos *et al*., 1987) (Figure 1.4). Various acquired and genetic disorders can increase HbF and FC levels in adult life.

1.5 Acquired increases in HbF levels in adults

In most acquired disorders the increased HbF production is due to perturbation of erythropoiesis; and increased demand for red cells leads to 'stress' erythropoiesis with preferential production of FCs. Increased HbF production is often seen in situations of acute erythroid expansion as in severe iron deficiency anaemia following treatment, in normal individuals following acute blood loss, in bone marrow regeneration after bone marrow transplantation, in cytotoxic treatment and in pregnancy, particularly in early second trimester when the blood volume increases sharply (Rochette *et al*., 1994). Two hypotheses for the increased HbF in pregnancy have been put forward; 1) a possible response of a placental or fetal inducer of HbF entering the maternal circulation and 2) the secretion of hormones which effect HbF production. Patients with haemopoietic malignancies such as juvenile chronic myeloid leukaemia and myelodysplastic syndromes often show a striking increase in HbF levels caused by a gross distortion in the regulation of erythropoiesis. Increased HbF levels have also been reported sporadically in patients with choriocarcinoma, testicular cancer, bronchogenic carcinoma, hepatoma and thyrotoxicosis (Rochette *et al*., 1994). The exact mechanism for the increased HbF seen in patients with solid tumours and cancers is being widely studied. In adult life the inactivated γ-genes in erythroid cells are methylated and the theory is that there are factors able to induce DNA demethylation and therefore activation of the γ-globin genes (Wolk *et al*., 2006). It is also thought that the expression of HbF in cancer prior to treatment may be a strategy of the cancer to gain increased access to oxygen, since HbF has a higher affinity for oxygen than adult haemoglobin and therefore HbF can be used as a potential marker for malignancy (Wolk *et al*, 2012).

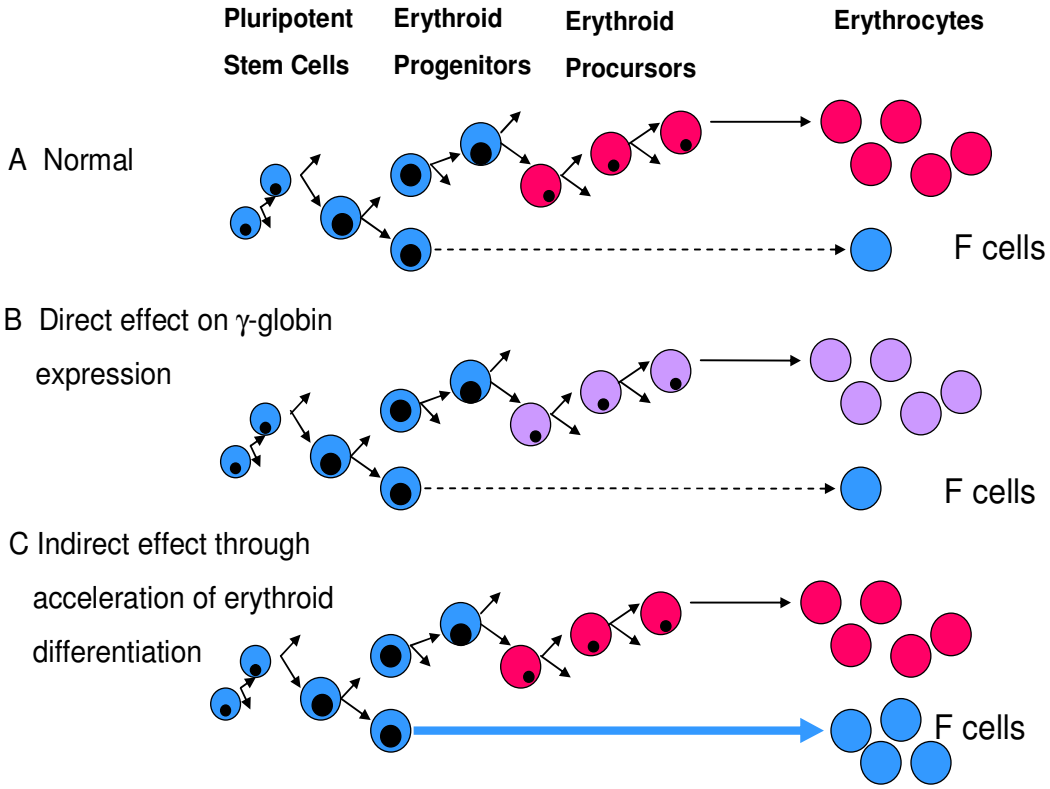


Figure 1.4 Proposed mechanisms of HbF production.

A) Normal fetal and adult haemoglobin are produced by cells from the same lineage i.e. haemoglobin switching occurs through a change in transcription program within the same cell. A change in transcription program occurs with erythropoietic differentiation, from one that predominantly supports γ-globin expression to one that supports only β-globin expression. FC represents the progeny of erythroid progenitor cells that terminate early before the change in transcription occurs. Two plausible mechanisms of HbF modulation have emerged from biological studies of *trans*-acting factors:

B) Direct effect on γ-globin gene expression. Where there is activation of γ- globin transcription or inhibition of γ-globin gene repression, thereby increasing the amount of HbF in each normal erythrocyte without producing more F cells.

C) Indirect via perturbation of kinetics of erythropoiesis such as in severe anaemia and the acquired disorders which increase HbF production by increasing the number of F cells produced. (Adapted and modified from Stamatoyannopoulos, 2005)

1.6 Genetic disorders increasing HbF levels in adults

Increased HbF production may occur as a primary genetic disorder or secondary to other inherited disorders.

1.6.1 Secondary increases in HbF levels

1.6.1.1 Increased HbF levels in sickle cell anaemia

Carriers for the sickle cell gene have normal HbF levels but in sickle cell homozygotes HbF levels can range from 1-20%. Factors that account for this striking variation include the preferential survival of HbF cells and the variability in the intrinsic capacity for HbF synthesis. Polymerisation of HbS leads to irreversibly sickled cells, which will be removed from the circulation by the reticuloendothelial (RE) system. HbF can effectively reduce the intracellular concentration of HbS so that the formation of HbS polymers becomes less likely. High HbF containing red cells gain this advantage and thus have a higher chance of survival.

1.6.1.2 Increased HbF levels in β-thalassaemia

Slightly increased HbF levels are often seen in patients heterozygous for β-thalassaemia as a result of the selective survival of HbF containing red cells. In homozygous β-thalassaemia HbF levels range from 10% to 100% depending on mutations. Despite this increase in γ-globin gene transcription, which is a response to erythropoietic stress, and the selective survival of HbF containing red cells, the majority of patients homozygous for β-thalassaemia (either β^+ or β^0) have very low Hb levels (3-5g/dl) and require regular blood transfusions for their survival. This is because the increase in HbF production is insufficient to

compensate for the reduced HbA production. However some individuals who are homozygous for β^0-thalassaemia produce more HbF and are able to maintain Hb levels of 8-10g/dl without a blood transfusion and therefore have a β-thalassaemia intermedia phenotype. Family studies have shown that in some cases the high HbF determinant segregates independently of the β-globin complex, indicating that *trans*-acting factors are involved (Wood, 1993).

1.6.2 Primary increases in HbF levels

The inherited disorders in which the underlying mutation is a direct cause of increased HbF production have been traditionally classified into two groups; δβ-thalassaemia and HPFH (deletion and non-deletion). Both δβ-thalassaemia and deletional HPFH are caused by deletions of the δ- and β-globin genes and it is apparent that the size and positions of the deletions do not correlate to the overall phenotype and that both disorders are closely related. They are classified depending on the level of HbF and whether the phenotype in the homozygous state is clinically affected or not. Therefore as there is overlap between the two disorders it has been suggested that they should be classified as one group. However, this reclassification has not happened as the precise mechanisms responsible for the increased HbF levels in these conditions are not fully understood and it is hoped that with increased knowledge a new classification might emerge naturally.

1.6.2.1 Increased HbF levels in δβ-thalassaemia

Heterozygotes for δβ-thalassaemia deletions ($^{G}\gamma^{A}\gamma(\delta\beta)^0$ and $^{G}\gamma(^{A}\gamma\delta\beta)^0$) are characterised by elevated levels of HbF ranging from 5-20% with a normal HbA_2 value (Table 1.2). The red cell indices are similar to those of β-thalassaemia

trait although the MCH and MCV are slightly higher. This is because in δβ-thalassaemia the defect in β-chain production is partially compensated by an increase in the output of γ-chains from the chromosome carrying the thalassaemia determinant (Shiokawa *et al*., 1988). Homozygotes have a clinical course of thalassaemia intermedia (Galanello *et al*., 2002).

Table 1.2 Reported δβ-thalassaemia mutations*.

Phenotype	Mutation	HGVS nomenclature	Size (kb)	%F hetero-zygote	Ethnic Origin	Reference
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	East European $(\delta\beta)^{0}$	-	9.12	13-24	East European	(Palena *et al.*, 1994)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Laotian $(\delta\beta)^{0}$	-	12.58	11.5	Laotian, Vietnamese	(Zhang *et al.*, 1988)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Thai $(\delta\beta)^{0}$	-	30	10	Thai	(Trent *et al.*, 1988)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Macedonian $^{G}\gamma^{A}\gamma(\delta\beta)^{00}$/Turkish inv- del $(\delta\beta)^{0}$	-	11.465+1.593del7.6inv	7-14	Macedonian, Turkish	(Efremov *et al.*, 1986b, Kulozik *et al.*, 1992)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Turkish $(\delta\beta)^{0}$	-	30	15	Turkish	(Oner *et al.*, 1996)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Leiden 7.4 kb $(\delta\beta)^{0}$	-	7.4	18.5	Turkish	(Phylipsen *et al.*, 2009)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Thai 11.3 $(\delta\beta)^{0}$	-	11.3	23.2	Thai	(Svasti *et al.*, 2007)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Indian $(\delta\beta)^{0}$	NG_000007.3:g.50509_83170del32662	32.62	5-15	India	(Mishima *et al.*, 1989)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Japanese $(\delta\beta)^{0}$	NG_000007.3:g.51483_165148del113666	113.63	5-7	Japan	(Matsunaga *et al.*, 1985)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Spanish $(\delta\beta)^{0}$	NG_000007.3:g.60375_153285del92911	115	5-15	Spanish	(Feingold and Forget, 1989)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Black $(\delta\beta)^{0}$	NG_000007.3:g.60530_60730del11822	11.77	24.8	African	(Anagnou *et al.*, 1985)

Phenotype	Mutation	HGVS nomenclature	Size (kb)	%F hetero-zygote	Ethnic Origin	Reference
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Sicilian $(\delta\beta)^{0}$	NG_000007.3:g.64336_77738del13403	13.38	4-19	Mediterranean	(Henthorn *et al.*, 1990)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Japanese 2 $(\delta\beta)^{0}$	-	27	-	Japan	(Yamashiro *et al.*, 2005)
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$	Senegalese $(\delta\beta)^{0}$	NG_000007.3:g.(63154_63209)_(70570_70625)del7417	74	14.2	Mediterranean	(Zertal-Zidani *et al.*, 2001, Saller *et al.* 2012)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Italian $(^{A}\gamma\delta\beta)^{0}$	-	52	12-17	Italian	(De Angioletti *et al.*, 1997)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Canadian $(^{A}\gamma\delta\beta)^{0}$	-	55.1	9.7-17.3	Canadian	(Voruganti *et al.*, 2009)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Leiden 69.5 $(^{A}\gamma\delta\beta)^{0}$	-	69.5	10.6	Italian	(Phylipsen *et al.*, 2009)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Asian $(^{A}\gamma\delta\beta)^{0}$	-	-	-	Asian	Ithanet id: ithalD 1533
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Turkish $(^{A}\gamma\delta\beta)^{0}$	NG_000007.3:g.45410_81665del36256	36.21	10-14	Turkish	(Tuan *et al.*, 1983, Henthorn *et al.*, 1990)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	German $(^{A}\gamma\delta\beta)^{0}$	NG_000007.3:g.(45922_46319)_(98640_99640)del	52	9.9-12.5	German	(Anagnou *et al.*, 1988)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Malaysian-2 $(^{A}\gamma\delta\beta)^{0}$	NG_000007.3:g.(47376_47553)_(89149_90149)del	42	-	Malay	(George *et al.*, 1986)

Phenotype	Mutation	HGVS nomenclature	Size (kb)	%F hetero-zygote	Ethnic Origin	Reference
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Chinese $(^{A}\gamma\delta\beta)^{0}$	NG_000007.3:g.48795_127698del78904	78.85	9-20	Chinese	(Jones *et al*., 1981b)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Black $(^{A}\gamma\delta\beta)^{0}$	NG_000007.3:g.49040_84889del35850	35.81	6-21	African	(Henthorn *et al*., 1985)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Indian $(^{A}\gamma\delta\beta)^{0}$ inv	-	0.834+7.460del15.5inv	10-18	Indian, Bangladeshi, Kuwaiti, Iranian	(Jones *et al*., 1981a)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Cantonese $(^{A}\gamma\delta\beta)^{0}$	-	-	19-20	Cantonese	(Zeng *et al*., 1985)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Malyasian-1 $(^{A}\gamma\delta\beta)^{0}$	-	-	25	Malay	(Trent *et al*., 1984)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Belgian $(^{A}\gamma\delta\beta)^{0}$	-	50	14-15	Belgian	(Losekoot *et al*., 1991)
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$	Yunnanese $(^{A}\gamma\delta\beta)^{0}$	-	88	9-17	Chinese	(Zhang *et al*., 1993)

*Adapted from ITHANET database (Lederer *et al*,. 2009).

1.6.2.2 Increased HbF levels in hereditary persistence of fetal haemoglobin (HPFH)

HPFH is the name given to a heterogenous group of disorders where there is a genetically determined persistence of fetal haemoglobin production into adult life, in the absence of any other haematological disorder (Bollekens and Forget, 1991). Two distinct types of HPFH have been classified, non-deletion and deletional. These are discussed in greater detail in section 1.7 and 1.8. The mechanisms by which deletions such as δβ-thalassaemia and HPFH lead to

increased levels of HbF production has been the subject of much interest. Three mechanisms have been proposed:

1. The deletion removes regulatory sequences within the β-globin gene cluster resulting in increased HbF levels due to the balance of the regulatory sequences which remain in the cluster. The identification of a 3.5kb ^{A}γ-δ-globin intergenic region which is a binding site for *BCL11A* has been shown to be an important site in haemoglobin silencing (Sankaran *et al*,. 2011). Removal of this region, for example in the HPFH deletions, results in higher HbF levels than in deletions which leave this region intact.
2. The deletion enables the 3' enhancer elements downstream of the β-globin gene to be in closer proximity with the γ-globin genes thereby increasing their expression.
3. Deletion of the β-globin gene region allows the β-LCR to continue interacting with the normally silenced adult γ-globin genes.

The elevation in γ-chain production in deletional HPFH is sufficient to compensate for the loss of β-chain production from the deleted β-globin gene; however this is not the case in δβ-thalassaemia. It may be that there is no simple explanation for the persistent γ-gene expression in deletional HPFH. Although it seems likely that the mechanisms listed above play an important role in this process with the degree of importance varying depending on the deletion. Given that there are a large number of elements around the β-globin gene cluster that affect regulation of the globin genes, it is entirely possible that other elements outside of the gene cluster are also involved in gene regulation.

1.7 Deletional HPFH

Deletional HPFH is relatively rare and is caused by extensive deletions of the β-globin gene cluster (including the δ- and β-globin genes). This form of HPFH demonstrates clear Mendelian inheritance and heterozygotes have higher HbF levels of 10-40% with normal red cell indices (Table 1.3). Heterozygotes and homozygotes for these deletions are unaffected clinically.

Table 1.3 Reported deletional HPFH mutations*.

Mutation	HGVS nomenclature	Size (kb)	%F heterozygote	Ethnic Origin	Reference
HPFH-1; Black	NG_000007.3:g.59478_144395del84918	84.92	20-30	African	(Tuan *et al*., 1983, Kutlar *et al*., 1984)
HPFH-2; Ghanaian	NG_000007.3:g.54867_139178del84312	83.68	20-30	African	(Tuan *et al*., 1983, Kutlar *et al*., 1984)
HPFH-3; Indian	NG_000007.3:g.50509_83170del32662	47.73	22-23	Indian	(Kutlar *et al*., 1984, Schroeder *et al*., 1973)
HPFH-4; Italian 1	NG_000007.3:g.	40	14-30	Italian	(Saglio *et al*., 1986)
HPFH-5; Italian 2	NG_000007.3:g.	12.91	16-20	Sicilian	(Camaschella *et al*., 1990)
HPFH-6; SE Asian	NG_000007.3:g.45595_124872del79278	79.28	18-27	Thai	(Kosteas *et al*., 1997)
HPFH-7; Vietnamese	NG_000007.3:g.	30	14-27	Vietnamese, Cambodian, Chinese	(Motum *et al*., 1993)
Algerian HPFH	NG_000007.3:g.48747_72606del23860	23.9	41	Algerian	(Joly *et al*., 2009)
French West-Indies HPFH	NG_000007.3:g.48762_72489del23728	23.7	36.5	French West-Indies	(Joly *et al*., 2009)
French HPFH	NG_000007.3:g.53013_72746del19734	19.7	35.7	French	(Joly *et al*., 2009)

*Adapted from ITHANET database (Lederer *et al*,. 2009).

1.8 Non-deletion HPFH

This type of HPFH has normal β-globin gene expression and all the genes within the β-globin cluster remain intact. The increase in HbF production is extremely heterogeneous ranging from 1-30%. Non-deletion forms of HPFH can be due to point mutations in either the $^{G}\gamma$- or $^{A}\gamma$-gene promoters, γ-globin gene rearrangements or 'Swiss type of HPFH' which does not follow Mendelian inheritance. As in the deletion type of HPFH, individuals with non-deletion HPFH have normal red cell morphology and counts with no clinical abnormalities.

1.8.1 γ-globin gene promoter mutations

Over the last fifty years several different types of non-deletion HPFH due to γ-globin gene promoter mutations have been reported. Sequencing of both the γ-genes revealed single nucleotide substitutions in the promoters of the γ-genes in a region that could affect their transcription. Studies on transgenic mice comparing the expression of normal and mutated genes have shown that these nucleotide substitutions in the promoters of the γ-genes are responsible for the increased levels of HbF and were not simply associated polymorphisms (Peterson *et al.*, 1995). To date, 25 such mutations have been described with 12 $^{G}\gamma$ HPFHs and 13 $^{A}\gamma$ HPFHs (Table 1.4). HbF levels in heterozygotes range from modest elevations to increases similar to those observed in deletion HPFH conditions. Homozygotes have been observed in both the British and Greek types of $^{A}\gamma$ HPFH presenting with normal red cell indices and HbF levels twice the level seen in heterozygotes (Camaschella *et al.*, 1989, Weatherall *et al.*, 1975)

Table 1.4 Reported non-deletion HPFH point mutations*.

Gene	Mutation	HGVS nomenclature	%F heterozygote	Ethnic origin	Reference
$^{G}\gamma$	-37 (A-T)	HBG2:c.-90 A>T	2.3	Dutch	(Bouva *et al.*, 2006)
$^{G}\gamma$	-109 (G-T)	HBG2:c.-162 A>C	4.1	Greek	(Chassanidis *et al.*, 2009)
$^{G}\gamma$	-110 (A-C)	HBG2:c.-163 A>C	3.1	Czech	(Indrak *et al.*, 1991)
$^{G}\gamma$	-114 (C-T)	HBG2:c.-167 C>T	11-14	Japanese	(Fucharoen *et al.*, 1990)
$^{G}\gamma$	-114 (C-A)	HBG2:c.-167 C>A	0.6-3.5	Algerian	(Zertal-Zidani *et al.*, 1999)
$^{G}\gamma$	-114 (C-G)	HBG2:c.-167 C>G	8.6	Australian	(Motum *et al.*, 1994)
$^{G}\gamma$	-161 (G-A)	HBG2:c.-215 G>A	4	African	(Leonova *et al.*, 1996)
$^{G}\gamma$	-175 (T-C)	HBG2:c.-228 T>C	17-30	African	(Friedman and Schwartz, 1976)
$^{G}\gamma$	-196 (C-T)	HBG2:c.-249 C>T	8.6	Greek	(Tasiopoulou *et al.*, 2008)
$^{G}\gamma$	-200 (+C)	HBG2:c.-253.254insC	18-28	Tunisian	(Pissard *et al.*, 1996)
$^{G}\gamma$	-202 (C-G)	HBG2:c.-255 C>G	14-21	African	(Collins *et al.*, 1984)
$^{G}\gamma$	-567 (T-G)	HBG2:c.-610 T>G	5.9-10.2	Iranian	(Chen *et al.*, 2008)
$^{A}\gamma$	-114(C-T)	HBG1:c.-167 C>T	3-6.5	African	(Oner *et al.*, 1991)
$^{A}\gamma$	-114 to -102	HBG1:c.-167-155del	30-32	African	(Gilman *et al.*, 1988b)
$^{A}\gamma$	-117 (G-A)	HBG1:c.-170 G>A	7.1-19	Greek, Italian, African	(Collins *et al.*, 1985, Gelinas *et al.*, 1985, Ottolenghi *et al.*, 1988, Huang *et al.*, 1987)
$^{A}\gamma$	-158 (C-T)	HBG1:c.-211 C>T	2.9-5.1	Greek	(Patrinos *et al.*, 1998)
$^{A}\gamma$	-175 (T-C)	HBG1:c.-228 T>C	36-41	African	(Stoming *et al.,* 1989, Coleman *et al.*, 1993)
$^{A}\gamma$	-195 (C-G)	HBG1:c.-248 C>G	4.5-7.0	Brazilian	(Costa *et al.*, 1990)
$^{A}\gamma$	-196 (C-T)	HBG1:c.-249 C>T	14-21	Italian, Chinese	(Giglioni *et al.*, 1984, Gelinas *et al.*, 1986)
$^{A}\gamma$	-198 (T-C)	HBG1:c.-251 T>C	3.5-12	British	(Tate *et al.*, 1986)
$^{A}\gamma$	-201 (C-T)	HBG1:c.-254 C>T	10.2	Greek	(Tasiopoulou *et al.*, 2008)
$^{A}\gamma$	-202 (C-T)	HBG1:c.-255 C>T	1.6-3.4	African	(Gilman *et al.*, 1988a)
$^{A}\gamma$	-211 (C-T)	HBG1:c.-264 C>T	2.7-6.3	African	(Arends *et al.*, 2004)

Gene	Mutation	HGVS nomenclature	%F heterozygote	Ethnic origin	Reference
$^{A}\gamma$	-225 to -222 (delAGCA)	HBG1:c-279-276del	6.7	African	(Gilman et al., 1988)
$^{A}\gamma$	-226 to -223 (delAAGC)	HBG1:c.-280-277del	3.2-5.4	Chinese	(Huang *et al*., 2000)

*Adapted from ITHANET databases (Lederer *et al*., 2009).

Analysis of protein binding to the promoters of the γ-genes has demonstrated potential binding sites for ubiquitous and erythroid-specific *trans*-acting transcription factors (Wood, 1993). Many of the nucleotide substitutions in non-deletion HPFH lie within or are close to the binding sites for these transcription factors and it seems likely that they exert their effect by altering the pattern of protein binding (Figure 1.5). However the precise mechanisms by which each mutation results in increased HbF production may differ. (Forget, 1998). Mutations could result in increased affinity for transcription factors which activate the γ-genes in adult life, decreased affinity for repressor molecules or a combination of both mechanisms.

1.8.2 γ-globin gene rearrangements

Several gene rearrangements have been reported in the β-globin gene locus and some have been identified which are responsible for different $^{G}\gamma$- and $^{A}\gamma$-ratios. The two most frequently reported are point mutations or gene conversions resulting in $^{G}\gamma$-$^{G}\gamma$- or $^{A}\gamma$- $^{A}\gamma$- which replaces the normal $^{G}\gamma$-$^{A}\gamma$- and the presence of multiple genes with three, four and even five γ-globin genes having been discovered (Huisman *et al*., 1991).

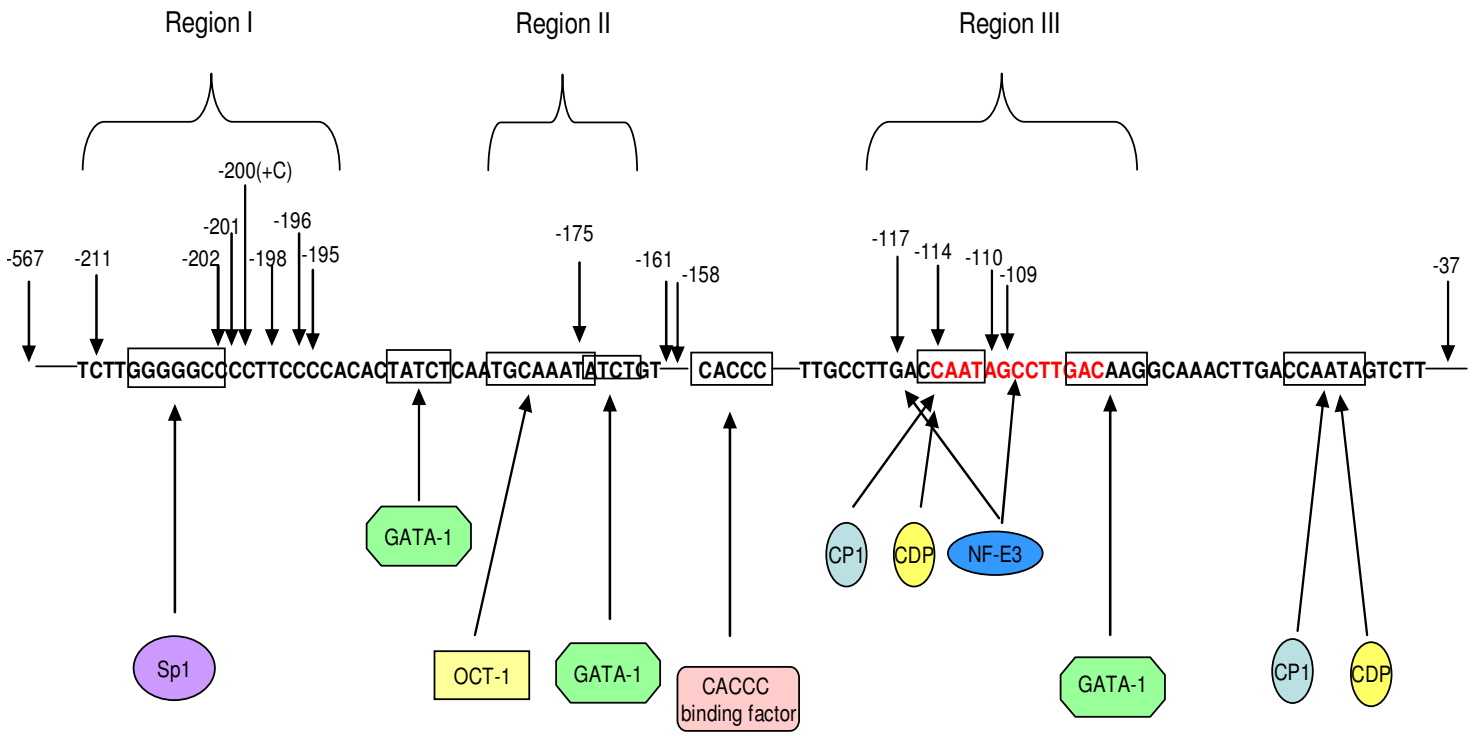


Figure 1.5 Transcription factor binding regions and positions of known mutations in the γ-globin gene promoters. Boxed areas show conserved sequences which are binding sites for the transcription factors indicated below the sequence. Arrows above the sequence show the positions of the known non-deletion HPFH point mutations with the sequence in red indicating the 13bp deletion (-114 to -102).

The first gene arrangement to be described in adults was the $^{G}\gamma$- $^{G}\gamma$- found in several African families. This rearrangement in also known as the Atlanta type of non-deletion HPFH and is characterised by HbF levels ranging from 1.3-9.8% with $^{G}\gamma$- values of 98% (Huisman *et al.*, 1985a). Further studies went on to show that the C→T polymorphism at position -158 of the $^{G}\gamma$-globin gene occur on both genes in the Atlanta non-deletion HPFH and mRNA data implied that it is actually the -158 polymorphism in both $^{G}\gamma$-globin genes which exerts its effects on HbF (Efremov *et al.*, 1994). Multiple $^{G}\gamma$-globin genes also result in increased adult HbF as in the case of the $^{G}\gamma$-globin triplication where HbF levels vary from 3.4-6.4% in heterozygotes but again only provided the C→T polymorphism at position -158 of the $^{G}\gamma$-globin gene occurs on the $^{G}\gamma$-globin genes (Efremov *et al.*, 1986a). Individuals with this triplication but without the C→T polymorphism at -158 were shown to have lower levels of HbF and also lower $^{G}\gamma$-values (Yang *et al.*, 1986). Yang *et al* (1986) also identified members of a Turkish family with γ-globin gene quadruplication and only slightly increased HbF levels of 0.7-3.8%.

1.8.3 Swiss type of non-deletion HPFH

This group of non-deletion HPFH disorders appear to have a genetically determined increase in HbF production in adult life but the inheritance is not well defined. In most cases they are characterised by only a very small increase in HbF (rarely above 5%). However there is evidence that they may play an important role in modifying the phenotype of β-thalassaemia or sickle cell anaemia and therefore efforts have been made to target therapies which

activate HbF. These conditions are referred to as 'Swiss HPFH' after the discovery in the 1960s of HbF persistence in Swiss army recruits (Marti and Buetler, 1961) or heterocellular HPFH since the distribution of HbF is uneven among the red cells (Boyer *et al*., 1975). The study by Marti and Buetler (1961) showed heritability, Swiss HPFH is quite different from classical HPFH as the inheritance does not follow Mendelian patterns and the HbF elevation is modest. Further studies showed that the amount of HbF varies considerably and that the distribution among the normal population is continuous and positively skewed (Thein and Craig, 1998). Heterocellular HPFH is found in approximately 10% of the population with HbF levels between 0.8 and 5% (Thein *et al*., 2009). Studies in twins went on to show that genetic factors in HbF levels account for 89% of the variability in F cell levels (Garner *et al*., 2000) with the remaining 11% of variance being accounted for by age and sex (2%) and unknown environmental factors.

In the past 10 years, several *trans*-acting mutations (mutations that are not actually physically linked to the locus) have been identified that alter the pattern of globin gene expression (Higgs *et al*., 2012). Insights into the regulation of globin genes that are obtained from these sporadic mutations have been greatly enhanced by extensive observations from family studies, twin studies, and genome-wide association studies (GWAS). Findings from these GWAS studies have identified other *trans*-acting factors (*BCL11A* and *HBS1L-MYB*) that can normally regulate the patterns of globin gene expression.

1.8.3.1 *XmnI-HBG2* polymorphism at the β-globin gene locus

In 1985 a single nucleotide polymorphism (SNP) C→T at position -158 of the $^{G}\gamma$-globin gene, later termed the *XmnI-HBG2* polymorphism was identified and shown to promote expression of the $^{G}\gamma$-globin gene (Gilman and Huisman, 1985). This polymorphism is common in all population groups with a frequency of about 0.35 (Garner *et al.*, 2000) and has been shown to increase HbF levels under conditions of erythroid expansion such as sickle cell anaemia and β-thalassaemia major leading to a milder disease (Thein *et al.*, 1987, Labie *et al.*, 1985). Certain haplotypes with this polymorphism have been shown to have higher HbF levels in sickle cell disease individuals from India, eastern Saudi Arabia and the Senegal (Labie and Elion, 1996) (Figure 1.6). The C→T change at position -158 was also shown to be associated with the 'Swiss-type' of HPFH in non-anaemic Europeans (Efremov *et al.*, 1987) and in individuals with high, normal or slightly raised HbF levels (Sampietro *et al.*, 1992). However this is not a consistent finding as this polymorphism is not present in all affected individuals and has been found in individuals with normal HbF levels. In one large pedigree with normal individuals having HbF levels between 2-6% the high HbF levels segregated with a β-globin gene haplotype which lacked the -158 C→T polymorphism (Donald *et al.*, 1988). It seems likely that in order to produce a full high HbF phenotype the *XmnI-HBG2* polymorphism must exist on a genetic background requiring the presence of additional factors (Thein and Menzel, 2009).

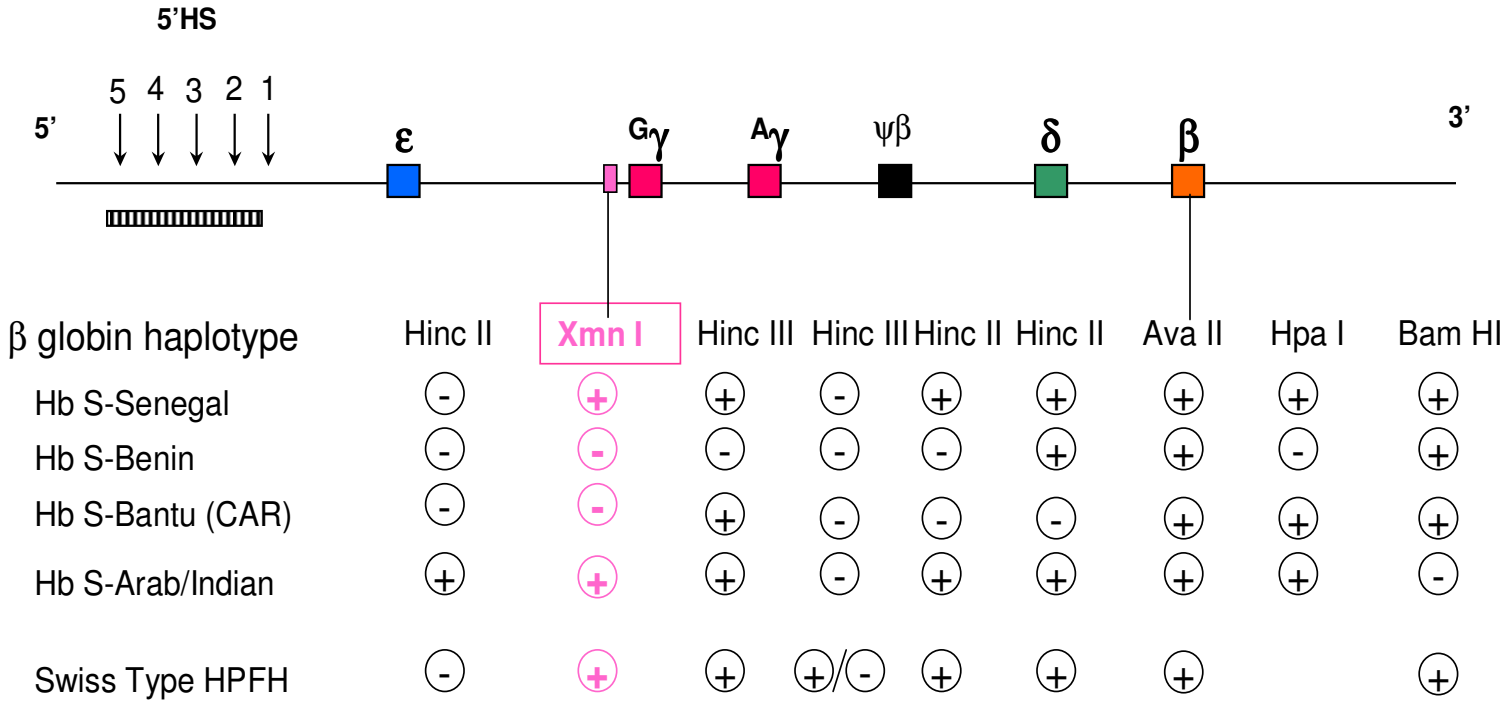


Figure 1.6 Genomic organisation of the β-globin gene cluster with the different polymorphic restriction enzyme sites. β-globin haplotypes are generated by linkage patterns of the RFLPs. The haplotypes associated with the sickle gene from various regions are shown. The Senegal and Arab/Indian and Swiss Type HPFH haplotypes carry the Xmn variant. The Senegal and Arab/Indian haplotypes have the highest HbF levels and the mildest clinical course. The Bantu haplotype has the lowest HbF levels and the severest clinical course (adapted and modified from Steinberg & Nagel, 2009).

1.8.3.2 X-linked on chromosome Xp22.2

Several studies in the 1980s showed that females had higher HbF levels than males. Rutland *et al* (1983) showed that females had higher HbF levels throughout life, whilst Mason *et al* (1982) observed higher HbF levels in normal females aged between 1-6 years than in males. They also noted that this finding was not evident in patients with sickle cell anaemia. Another study undertaken in Japanese adults showed that the high HbF phenotype might be an X-linked trait with 11% of males and 21% of females being carriers for heterocellular HPFH (Miyoshi *et al.,* 1988). An extensive series of studies was then carried out by Dover *et al* (1992) looking at HbF levels in individuals with sickle cell anaemia. They noticed that females had much higher levels of HbF than the males and undertook segregation analysis based on the phenotypes they had found, discovering a putative locus at Xp22.2 which he called the FC locus. This group then went on to use multiple linear regression analysis to determine the importance of this locus along with other known variables which increase HbF such as age, sex, α-globin gene number and β-globin gene haplotype. They concluded that all the variables in the study accounted for 50% of the variation in HbF levels with the FC locus being the most important factor accounting for 35-41% of the total HbF variation seen in homozygous sickle cell subjects (Chang *et al*., 1995). When they removed this locus from their regression model they found that the second most important predictor of HbF levels were the genetic factors associated with the common β-globin gene haplotypes. Therefore 50% of the variation in HbF levels in homozygous sickle cell disease patients remained unexplained. However others went on to

conclude that they could not replicate the same finding and that further *trans*-acting factors must be involved in the increased HbF phenotype.

1.8.3.3 *HBS1L-MYB* intergenic region on chromosome 6q23

For over 20 years scientists at the Weatherall Institute of Molecular Medicine (WIMM) in Oxford have been studying a large Asian-Indian pedigree with β-thalassaemia and heterocellular HPFH in order to try and explain some of the genetic variance seen in HbF levels (Wood *et al.*, 1977, Thein *et al.*, 1994, Craig *et al.*, 1996). These and other early studies suggested that the high HbF determinate segregates independently of the β-globin gene in some families with β-thalassaemia and sickle cell anaemia and families were discovered through probands with unexpectedly mild disease (Cappellini *et al.*, 1981, Old *et al.*, 1982). In the case of the Asian-Indian family individuals without β-thalassaemia had HbF levels ranging from 0-3%, while heterozygotes for β-thalassaemia having HbF levels from 0.8-24% (Thein *et al.*, 1994). Therefore the inheritance of the high HbF in this family was unclear. Genome-wide linkage analysis using 210 microsatellite markers was undertaken and this led to the discovery of linkage to a small segment of chromosome 6, 6q22.3-q24 (Craig *et al.*, 1996). Interestingly they found no evidence of any contribution from the FC locus on Xp22.2 and further fine mapping refined the area to a 1.5 Mb region containing 5 known genes (*ALDH8A1, HBS1L*, *MYB, AHI1* and *PDE7B*) none of which had a mutation and three (*HBS1L, MYB* and *AHI1*) that are expressed in erythroid cells (Garner *et al.*, 1998, Close *et al.*, 2004). Further high resolution association mapping was carried out on individuals of northern European ancestry and showed linkage to a set of 7 common SNPs

spanning a region of 79kb. The SNPs were distributed into three linkage blocks referred to as *HBS1L-MYB* intergenic polymorphisms (HMIP) blocks 1, 2 and 3 (Thein *et al*., 2007). These three haplotype blocks completely account for the FC variance due to the 6q23 locus but the second block had the strongest effect and further studies in African individuals also found the same strong association with FC levels and *HMIP-2* (Creary *et al*., 2009). The *HMIP-2* block has been shown to contain a regulatory locus as evidenced by several *GATA-1* signals that coincided with DNase1 hypersensitive sites in erythroid precursor cells (Wahlberg *et al*., 2009). It has been suggested that this regulatory locus controls *MYB* expression which in turn influences erythroid differentiation and indirectly the control of HbF levels (Thein *et al*., 2009).

1.8.3.4 *BCL11A* (B-cell lymphoma/leukemia 11A) on chromosome 2p16

By 2006 genome-wide association studies (GWAS) were being undertaken to look at common polymorphisms and relate them to phenotypes and variations at chosen loci. GWAS studies when applied to the variation in HbF yielded interesting findings showing that the variation can be accounted for by just three loci (Menzel *et al*., 2007, Uda *et al*., 2008, Lettre *et al*., 2008, Sedgewick *et al*., 2008, Galanello *et al*., 2009, Nuinoon *et al*., 2010). These loci include the β-globin cluster itself, the intergenic region between *HBS1L* and *MYB* genes and also a new locus in intron 2 of the oncogene *BCL11A* on chromosome 2p16. *BCL11A* is a zinc-finger transcriptional repressor found to be active in B lymphoid cells and expressed in erythroid cells and had not been previously implicated in haemoglobin switching (Bauer and Orkin., 2011). Several studies went on to show that *BCL11A* silences the γ-genes (Sankaran *et al*., 2008,

2009). Sankaran's studies showed that knockdown of *BCL11A* increases HbF synthesis in human erythroid progenitors and complete knock-out of the gene in mice causes profoundly delayed switching from embryonic to adult globins as well as silencing of the γ-globin genes. *BCL11A* has a binding site not in the γ-promoter but in the β-LCR as well as intergenic regions on the β-globin gene cluster which are associated with γ-globin gene repression (Xu *et al*., 2010, Jawaid *et al*., 2010). *BCL11A* associates with other transcription factors such as *GATA-1* and *SOX-6* (Xu *et al*., 2010, Jawaid *et al*., 2010) and unlike *BCL11A, SOX6* directly occupies the γ-globin promoter (Figure 1.7). Xu *et al* (2010) showed that double knockdown of *BCL11A* and *SOX6* has an additive effect on HbF derepression, suggesting that *BCL11A* may exert part of its γ-globin repression with *SOX6*. SNPs within the 14kb intron 2 of *BCL11A* correlate most strongly with HbF expression. These findings are very exciting and put *BCL11A* in the running as a potential therapeutic target for the reactivation of HbF.

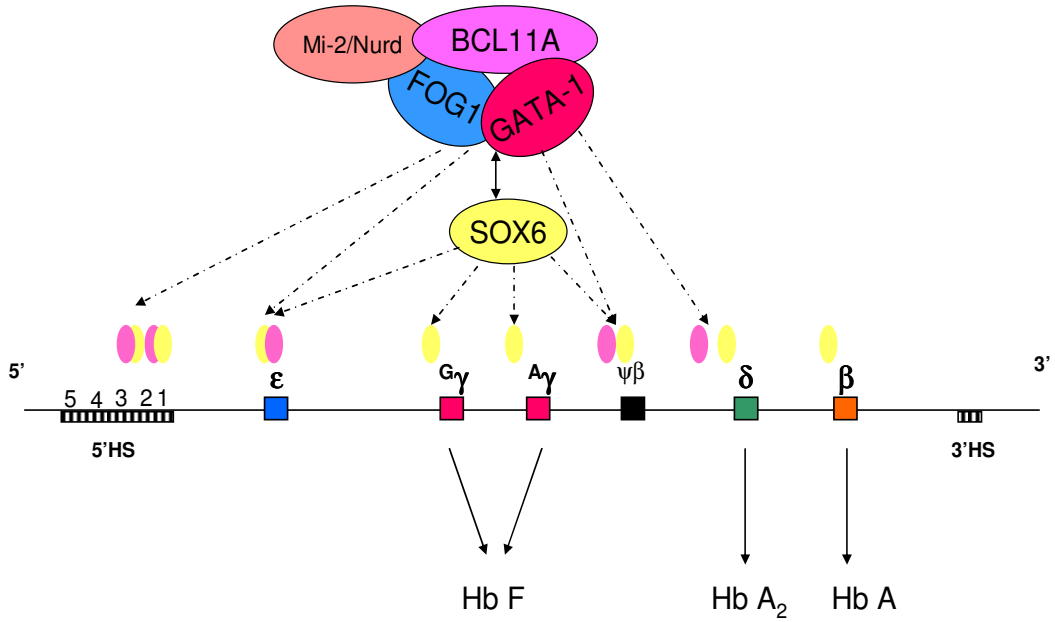


Figure 1.7 Model of *BCL11A* silencing of the γ-globin genes. The diagram illustrates the physical interaction between *BCL11A* and the *Mi-2/NuRD* complexes, erythroid transcription factors *GATA1* and *FOG1*, and the HMG-box protein *SOX6*. Rather than binding to the promoters of the γ-globin or β-globin genes as these latter factors do, *BCL11A* protein occupies the upstream LCR and γδ-intergenic regions of the β-globin cluster in adult human erythroid progenitors. Transcriptional silencing of the γ-globin expression by *BCL11A* involves long-range interaction within the β-globin cluster and local interactions with the chromatin associated *SOX6* proteins at the proximal promoters of the γ-globin genes. (Adapted and modified from Xu *et al.*, 2010)

1.8.3.5 *KLF1* (Erythroid Krüppel-like Factor 1) on chromosome 19p13.2

Recently an additional potential locus was identified when point mutations in the *KLF1* (also known as *EKLF*) gene were found to be associated with HPFH in a Maltese family (Borg *et al., 2010*) and in a family from Sardinia (Satta *et al.*, 2011). *KLF1* is an essential erythroid transcription factor first identified in 1993 that binds to an important DNA binding site, the CACCC motif in the β-globin gene (Miller and Bieker, 1993). Mutations in the CACCC region of the β-globin gene have been shown to be a cause of β-thalassaemia as *KLF1* binding is

prevented (Feng *et al*., 1994). Experiments in mice showed that ablation of *KLF1* caused profound β-thalassaemia and lethality (Nuez *et al*., 1995). The gene found on chromosome 19 comprises of a proline rich N-terminal region containing a transactivation domain and a C-terminal region containing three zinc finger domains essential for DNA binding.

KLF1 mutations have been associated with several erythroid phenotypes. In the two family studies exhibiting the HPFH phenotype, HbF levels ranged from 0.9-30.9%. Variability in HbF was found throughout both families with different *KLF1* missense mutations. In 2008 Singleton *et al* (2008) reported the first mutations in human *KLF1*. They identified 9 different loss-of-function mutations being responsible for the blood group phenotype In(Lu). However no clinical features or haematological parameters were reported. Arnaud *et al* (2010) found *KLF1* mutations in 2 unrelated patients with congenital dyserythropoietic anaemia (CDA). Their patients had a heterozygous missense mutation (p.E325K) in the second zinc finger of the *KLF1* gene which caused a dominant effect resulting in a severe haemolytic anaemia. Very high levels of HbF were seen (31.6% and 44%) and they also had expression of embryonic haemoglobins ζ-and ε-globin. A further study in Sardinia went on to show that *KLF1* mutations are linked to borderline HbA_2 levels and normal to moderately increased HbF levels (Perseu *et al*., 2011).

Studies started to link together *KLF1* and *BCL11A* with the phenotype of HPFH (Borg *et al*., 2010) In the Maltese family with HPFH it was found that samples from the affected individuals had *KLF1* levels directly proportional to *BCL11A*

levels and inversely proportional to γ-globin levels (Bieker, 2010). Importantly it was also shown that human *KLF1* binds to the *BCL11A* gene. Zhou *et al* (2010) then went on to confirm these findings showing that *BCL11A* levels were dramatically downregulated in *KLF1* knockdown mice and in human cells and that a decrease in *BCL11A* levels is seen with an increase in γ-globin levels. This suggests that *KLF1* plays a critical role in regulating the switch between fetal and adult haemoglobin expression both by direct activation of β-globin and indirect repression of γ-globin gene expression in adult erythroid progenitors via regulation of *BCL11A* (Fig 1.8) (Siatecka and Bieker, 2011).

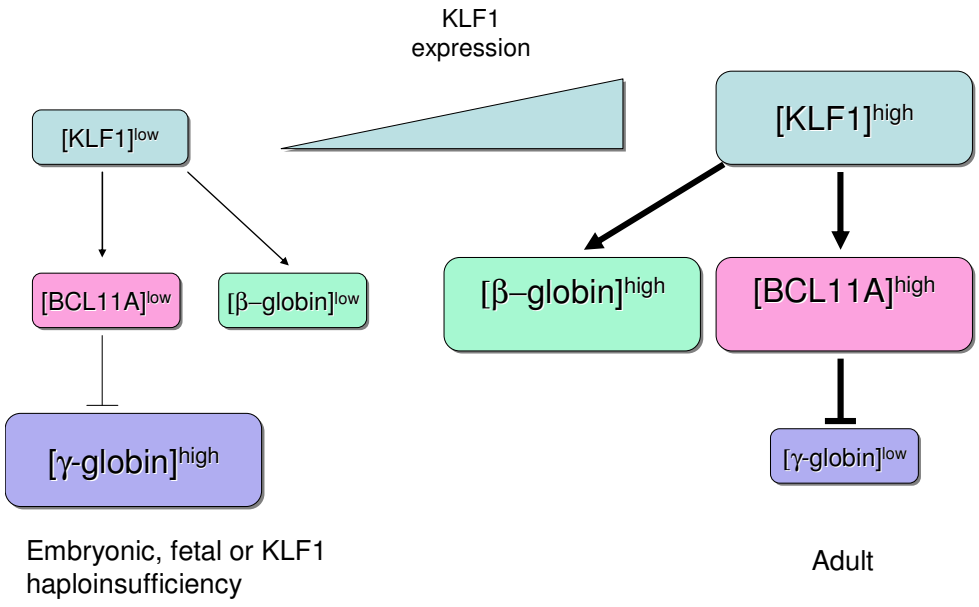


Figure 1.8 *KLF1* regulates globin switching. During embryonic and fetal development or in *KLF1*-haploinsufficient adults (left), *KLF1* levels are low, resulting in low levels of adult β-globin and *BCL11A* and high levels of γ-globin. In adults with two functional copies of *KLF1* (right), increased expression of *KLF1* in definitive red blood cells promotes high levels of adult β-globin and *BCL11A* expression, which in turn represses γ-globin expression (adapted from Bieker, 2010).

Although major *trans*-acting factors have been located the mechanisms through which they operate to increase HbF production are still not identified fully. The molecular basis for persistent HbF expression in adults is extremely heterogenous. It is possible to identify the deletional HPFH mutations both genotypically and phenotypically and it is indicated that the γ-globin genes can be reactivated to high levels of expression. However there are still no answers as to what the mechanisms are for the increased HbF synthesis. The *trans*-acting factors are believed to be responsible for the inheritance of non-deletional HPFH accounting for 20-50% of the HbF variability (Thein and Menzel, 2009). Therefore much work is still to be done in order to obtain all the factors that contribute to HbF production and to define the molecular network through which these operate.

1.9 Therapeutics for HbF induction

The ameliorating effect of HbF has led to the development of pharmacological agents that reactivate HbF synthesis; however current therapeutic agents do have toxic side effects (Perrine, 2005). Currently the only potentially curative therapies that are available for sickle cell disease and β-thalassaemia major are bone marrow transplantation and gene therapy but both have significant limitations for widespread use, particularly in the developing world (Sankaran and Orkin, 2013). Gene therapy remains largely experimental although advances in this approach are being made (Cavazzana-Calvo *et al.,* 2010). The major treatment for these disorders involves symptomatic care and red blood cell transfusion as is clinically necessary, yet transfusions can lead to major clinical problems such as iron overload. The recent identification of transcription factors that play a role in globin gene regulation and in particular HbF expression has lead to the promise of potential therapeutic targets which could lead to better, less toxic and more effective strategies for HbF induction (Sankaran, 2011). Further work will be necessary before these molecular targets can be translated into therapeutic advances but the fact that so many targets have been recently identified lends promise in this field. A better understanding of the mechanisms involved in the silencing of HbF will undoubtedly yield additional insights into how this process can best be modulated.

1.10 Thesis aims

The work presented in this thesis aims to identify the main genetic factors known to be involved in HbF expression in a cohort of patients with elevated HbF levels. This objective will be achieved by:

1) Identifying β-globin gene cluster deletion mutations in a high HbF cohort and determining the range and frequency of these deletions which is unknown.

2) Fine mapping of any novel mutations with a view to understanding the molecular mechanisms involved.

2) Ascertaining the frequency and range of γ-globin gene promoter mutations in the UK population.

3) Determining whether polymorphisms in the three major loci (*XmnI-HBG2, BCL11A* and *HBS1L-MYB*) are associated with increased HbF in a high HbF cohort.

4) Performing sequence analysis of the *KLF1* gene to see whether mutations in this essential transcription factor play a significant role in increasing HbF levels in adults.

5) Investigating the finding that Asian Indian newborns appear to have less HbF at birth than newborns from other ethnic groups. Could they be switching earlier?

CHAPTER TWO: MATERIALS AND METHODS

2.1 Introduction

This chapter outlines the standard protocols used in this project. Other methods used and developed are described in detail in the appropriate sections of the thesis (Chapters 3, 4, 5, 6, 7 and 8). Unless otherwise stated, all chemicals were of analytical grade or higher and were obtained from Sigma-Aldrich or BDH.

2.2 Haematological analysis

2.2.1 Full blood count (FBC)

Venous blood was collected in Vacutainer™ tubes using ethylenediamine tetraacetic acid (EDTA) as an anticoagulant. Hb level, red blood cell (RBC) count, mean cell volume (MCV) and mean cell haemoglobin (MCH) were determined using a Sysmex XE2100 haematology analyser.

2.2.2 High performance liquid chromatography (HPLC)

The measurement of haemoglobins in blood samples can be determined by passing solutions of haemoglobin through a weak cation-exchange high performance liquid chromatography (HPLC) system. Adult blood samples collected in EDTA were analysed using the Bio-Rad Variant™ II Hemoglobin Testing System (Bio-Rad Laboratories Inc) according to the manufacturer's instructions. Analysis of the haemoglobin fractions were provided by the Bio-Rad™ CDM software package and all buffers and columns were provided in the β-thalassaemia short™ kits supplied by Bio-Rad™. Newborn blood spots were analysed on a Vnbs (variant newborn system) Bio-Rad™ Sickle Cell Program with kits supplied by Bio-Rad™.

2.2.3 Isoelectric focusing (IEF)

Isoelectric focusing is a technique used for separating haemoglobin molecules by differences in their isoelectric point (pI). The pI is the pH at which a protein has no net charge and therefore does not migrate further in an electric field. The Multiphor II IEF tank was supplied by Amersham Biosciences linked to a Caron 2050W Circulating water bath and Thermo Electron Corporation 2060P power pack. Reagents and pre-cast agarose gels were supplied in kit form (Resolve® Neonatal Hemoglobin Test Kit) by Perkin Elmer.

2.3 Extraction of genomic DNA

2.3.1 Manual phenol-chloroform DNA extraction from whole blood

DNA from samples with less than 1ml of whole blood, as seen in paediatric samples, were manually extracted using phenol chloroform. DNA was prepared from peripheral blood leukocytes. Venous blood in Vacutainer™ tubes containing EDTA were washed with 2 volumes of reticulocyte saline (0.13M NaCL, 5mM KCL, and 7.4mM $MgCl_2$). The sample was inverted several times to ensure mixing and then centrifuged at 3500rpm for 5 mins at 4^0C. Afterwards, the supernatant was discarded and the cell pellet resuspended in 1ml of 2x lysis solution (0.77M NH_4Cl, 0.046M $KHCO_3$). The solution was mixed by invertion and rotated for 10 mins to lyse the red blood cells. The centrifugation step was repeated and after discarding the supernatant the buffy coat was resuspended in 150µl of lysing solution (100mM NaCl, 25mM EDTA) with 0.5ML 10% SDS and 20ml 20µg/ml Proteinase K (Roche). The sample was then incubated overnight at 37^0C. Proteins were then removed by

extracting twice with an equal volume of phenol. This was followed by repeating the extraction with an equal volume of chloroform. Following each extraction the sample was centrifuged at 3500rpm for 5 minutes at room temperature. The aqueous phase was then transferred to a clean tube and DNA precipitated by adding two volumes of 100% ethanol and 1/10 volume of 7.5M ammonium acetate. The mixture was centrifuged again and then the pellet washed in 70% ethanol. The DNA pellet was air-dried and then resuspended in DNase free water (Sigma).

2.3.2 Automated chemagen DNA extraction from whole blood

Samples with volumes of >1ml were placed on an automated DNA extractor. The extractor used was the Chemagen-Perkin Elmer Integrated Automated DNA Extractor. It consists of two integrated systems: the Chemagen MSM I extraction robot and the Perkin Elmer MPII liquid handler. Before placing onto the extractor, 1ml of EDTA whole blood had 10µl of protease and 1.5mls of lysis buffer1 added into a 50ml Falcon tube. The sample was then incubated for 10 mins and 120µl of paramagnetic bead solution was added to the lysed blood sample together with 4ml of binding buffer. This buffer should enhance the binding of the DNA to the magnetic beads. The Falcon tube was then loaded onto the automated DNA extraction platform. The DNA released from the lysed white blood cells gets bound to the magnetic beads and then subsequently transferred from one washing buffer to another by way of the magnetic rods on the extractor. Each washing step includes a homogenisation. There were four serial washing buffers with different ethanol concentrations. In the final step the DNA gets released from the beads in the elution buffer (DNase free water,

Sigma®) and transferred into a labeled tube. All reagents and chemagen kits are supplied by Perkin Elmer.

2.3.3 Extraction of DNA from hair follicles

DNA was extracted from hair follicles using the following protocol: 200µl of Buffer X1 (10mM TrisCl pH 8.0, 10mM EDTA, 100mM NaCl, 40mM DTT, 2% SDS and 250µg/ml proteinase K) was added to five hair follicles placed in a 1.5ml microcentrifuge tube and incubated overnight at 55^0C until dissolved. Standard phenol cholorform extraction of the DNA was then performed on the lysate (as described in section 2.3.1).

2.3.4 Extraction of DNA from dried blood spots

Two 3mm dried blood spots were added to a 1.5ml Eppendorf tube. 1ml of 0.5% saponin in PBS was added and incubated overnight at 4°C. The spots were transferred to a new 1.5ml Eppendorf containing 1ml of PBS and incubated for 1hr at 4°C. Then 75µl of 20% Chelex in distilled water was added to a 0.5ml Eppendorf tube. The blood spots were transferred into this tube. The tube was vortexed for 30 secs and incubated at 99 °C for 15 mins in a thermocycler with a heated lid. Further vortexing for 10-15 seconds was performed and then the Eppendorf tube was spun in a centrifuge at 10,000rpm for 5 mins. The supernatant was transferred into a final 0.5ml Eppendorf tube for storage.

2.4 Measurement of DNA concentration by nanodrop

The quality and concentration of the DNA extracted using the above methods was determined using a Nanodrop 2000c UV-Vis spectrophotometer (Nanodrop Technologies). The equipment utilises a sample retention technology which

allows direct measurement of 1µl of double-stranded DNA samples without dilution. With the sampling arm open, 1.5µl of DNA was pipetted directly onto the lower measurement pedestal after blanking. By closing the sampling arm, a column was automatically drawn between the upper and lower measurement pedestals through two corresponding optical fibers and a spectral measurement established. After measurement the sampling arm was opened and the upper and lower pedestals wiped clean with distilled water. The measurements were recorded and stored on a database and on the side of the DNA sample tube.

2.5 Polymerase chain reaction (PCR)

Many of the experiments mentioned throughout this thesis will employ the PCR technique. The conditions of each PCR amplification were optimised for each specific set of primers, the thermostable polymerase used and the length of the target sequence. Unless otherwise stated, a standard 25µl reaction contained 100-250ng of genomic DNA, 2 units of *Taq* polymerase (Invitrogen), 10pmol of each primer, 0.2mM dNTPs (GE Healthcare Life Sciences), 1.5mM $MgCl_2$ (Abgene) and 1 X Buffer IV (Abgene). All primers were designed and the melting temperatures calculated manually using the formula: Tm= [2(A+T) +4(C+G)] where A, T, C and G are the number of each of these nucleotides in the primer sequence. PCRs were carried out in a thermocycler (T3 Biometra) in 0.5ml PCR tubes with no heated lid and mineral oil in most cases. When using AmpliTaq DNA polymerase, the initial melting cycle was for 5 minutes.

The standard cycling conditions were:

1 cycle of:	95^0C for 3 min
30-40 cycles of:	95^0C for 1 min
	Tm^{0}C for 1 min
	72^0C for 1 min
1 cycle of:	72^0C for 10 min

When QIAGEN Multiplex PCR Master Mix (QIAGEN) was used instead of Buffer IV then the following conditions and reaction mix was used. A 25µl reaction contained 1µl of 100-250ng of genomic DNA, 10pmol of each primer, 12.5µl Multiplex PCR Master Mix (QIAGEN) and 9.5µl DNase-free water (Sigma). The conditions were as per the standard PCR cycling above except an initial activation of the HotStart Taq polymerase is required:

1 cycle of:	95^0C for 15 min
30-40 cycles of:	94^0C for 1 min
	Tm ^{0}C for 1 min
	72^0C for 1 min
1 cycle of:	72^0C for 10 min

2.5.1 Long range PCR

SeqTarget LongRange PCR kit from QIAGEN was used. This kit has a blend of enzyme and thermostable DNA polymerases with enhanced proofreading ability. This kit can amplify long-range PCR fragments 2kb to 20kb, G-C rich

regions and other difficult templates. The kit also includes Q-solution which facilitates amplification of difficult templates by modifying the melting behavior of DNA and a control primer pair which allows checking of the suitability of the human genomic DNA for use as a template. The reaction mix was prepared as per the kit instructions and the cycling protocol chosen depending on whether the amplification product would be 2-8kb or greater.

2.5.2 Gap-PCR for the common β-globin gene cluster deletion mutations

The technique of Gap-PCR is used routinely to detect many common deletion mutations. The technique is based upon the inability of PCR primers that are far apart to direct amplification unless a deletion brings them closer together. PCR primer pairs are designed to flank a known deletion, generating a unique amplicon that will be smaller in the mutant sequence compared with the wild type. Seven common β-globin gene deletion mutations were screened for by Gap-PCR. These consisted of HPFH-1, HPFH-2, HPFH-3, Sicilian $(\delta\beta)^0$ thalassaemia, Chinese $^{G}\gamma(^{A}\gamma\delta\beta)^0$ thalassaemia and two complex rearrangements: the Asian-Indian inversion-deletion $^{G}\gamma(^{A}\gamma\delta\beta)^0$ thalassaemia and the Turkish form of inversion-deletion $(\delta\beta)^0$ thalassaemia. The Gap-PCR primers and amplification conditions used for all these deletions were as previously described (Craig *et al.*, 1994).

2.6 Agarose gel electrophoresis of DNA

DNA fragments were separated by horizontal agarose gel electrophoresis. 1.5 to 3% (w/v) agarose gels were prepared in 1X TBE with the gel concentration depending on the size of the DNA fragments to be resolved. The sample was mixed with 1/10 volume of 10X TBE gel loading buffer and ran with appropriate

DNA markers. A potential difference of 2-5V/cm was applied across the gel until a satisfactory resolution was achieved. The gels were stained in 0.5µg/ml ethidium bromide solution and then destained in distilled water for 10 mins. The gels were placed on a UV transilluminator (Bio-Rad Gel-Doc™) to illuminate the DNA under ultraviolet light and a permanent record captured onto thermal print paper.

2.7 DNA sequence analysis

Sequencing was based on Sanger's chain termination method (Sanger *et al.*, 1977). Sequencing was carried out using an ABI PRISM® 3100 or a Beckman Coulter CEQ 8000 genetic analyser. The initial PCR amplification was performed as per the conditions in section 2.5 above.

2.7.1 PCR product purification

After PCR amplification 10µl of the PCR product was mixed with 10µl of Micro Clean (Microzone Ltd). The plate wells were pipette mixed and left for 5 minutes at room temperature. The plate was spun at 4700rpm for 40 minutes, turned upside down and spun for 10 seconds at very low speed (10rpm) then air dried. The pellet was resuspended in 12µls of DNase-free water (Sigma) and left for 10 minutes to rehydrate.

2.7.2 Cycle sequencing

2.7.2.1 Beckman coulter CEQ 8000

For each reaction 4µl of QuickStart Mix (GenomeLab™), 0.4µl of sequencing primers, 4.6µl of distilled water and 1.5µl of purified PCR products were added

to the well of a PCR plate, mixed thoroughly and sealed. The plate was placed into a thermal cycler (T3 Biometra) with the following conditions: 30 cycles of 96^0C for 20secs, 50^0C for 20secs and 60^0C for 4 mins and pause at 15^0C. Each sample was then purified by ethanol precipitation.

2.7.2.2 ABI PRISM® 3100

Sequencing reactions contained 1µl of PCR product and 9µl of reaction mix containing 2pmol of sequencing primer, 2µl ready reaction premix (ABI PRISM® Big Dye™ Primer Cycle Sequencing Kit, Invitrogen), 1µl Big Dye sequencing buffer and 3µl DNase-free water. Cycling conditions were an initial activation of 96^0C for 3 minutes followed by 25 cycles of: 96^0C for 15 secs, 50^0C for 10 secs, 60^0C for 2 mins and a final extension of 15^0C for 5 mins. Each sample was then purified by ethanol precipitation.

2.7.3 Ethanol precipitation

For each reaction 15µl of stop solution (2µl of 100mM EDTA, 2µl CH_3COONa (3M, pH 5.2), 1µl of glycogen and 10µl of distilled water) was added to the PCR products. 60µl of ice cold 100% ethanol was added to each well and the plate mixed and sealed. The plate was centrifuged at 4700rpm for 30mins at 4^0C. The plate was turned upside down to remove the supernatant and a further 200µl of ice cold 70% ethanol added to the wells. Further centrifugation at 4700rpm for 5mins at 4^0C and the supernatant removed. The PCR product was washed with 70% ethanol again by repeating the last step. Products were left to air dry for 30mins before 40µl of sample loading solution (SLS, Beckman

Coulter) or 8.5µl HiDi (ABI PRISM®) was added followed by a drop of mineral oil.

2.7.3 Analysis of sequence data by mutation surveyor

Mutation surveyor® (SoftGenetics, USA) was used to analyse the data obtained from the sequencers. This software is designed to find mutations and SNPs using a patented technology that allows base-to-base comparison of single and bi-directional DNA sequence traces generated by a standard automated sequencer. The software has a capacity to perform mutational analysis on up to 400 lanes of data simultaneously. The manufacturer claims that the software is capable of detecting >99% of mutations and can process up to 1 billion base pairs of sequence data per day under fully unattended operation (see http://www.softgenetics.com for further details).

2.8 Pyrosequencing assays for the SNPs in BCL11A and HBS1L-MYB

Pyrosequencing is a sequencing-by-synthesis method that enables rapid real-time sequencing of short DNA sequences, which can be used for the analysis of single nucleotide polymorphisms (Timbs *et al.*, 2012). This method was adopted to identify the two most common SNPs found in northern European and African populations in the gene *BCL11A* (rs 11886868) and the intergenic region *HBS1L-MYB* (rs 9399137). Using the Pyromark Assay Design Software 2.0 (QIAGEN) several sets of primers were designed and the following pairs worked optimally: rs 11886868 forward

TGCCCCTTTGCTGTCAAT, reverse ACCATGGATGAATCCCAGAAT (biotinylated) and sequencing primer CGTCTTTTGTGTTTAATTC.

Rs 9399137 forward CAACATCACCTTAAAAGGCGGTAT (biotinylated), reverse GCAGGGTTGCTTGTGAAAAAA and sequencing primer TGCTTGTGAAAAAACTGT. The standard QIAGEN PCR Multiplex Master Mix method was used (see section 2.5 above) with the following conditions: 95^0C for 10 mins followed by 30 cycles of 95^0C for 1 min, 63^0C for 1 min, 72^0C for 1.5 mins with a final extension of 72^0C for 10 mins. The pyrosequencing assays were carried out on a Pyromark Q24 as per the manufactures instructions and all reagents were supplied by QIAGEN. See Appendix 3 and 4 for examples of pyrosequencing traces and assay design.

2.9 Multiplex ligation-dependent probe amplification (MLPA)

The MLPA technique was used to detect deletions and copy number variations in the β-globin gene cluster. DNA probe sets are designed to span the area of interest and after hybridisation are amplified using only one pair of PCR primers (Appendix 1). The quantity of amplification product from each probe will be proportional to the copy number of the target sequence. The amplification products are then separated by capillary electrophoresis. A deletion on one chromosome can be detected by reduced relative peak heights in comparison to controls. The extent and position of the deletion were determined by mapping the positions of the probes with reduced peak height. Several different β-globin gene cluster probe sets were adopted but all developed and designed around the probe sets by Harteveld *et al* (2005). The method used was as per

the manufacturers' instructions and the β-globin gene cluster probe sets and reagent kits were supplied by ServiceXS (Leiden, Netherlands) and MRC-Holland (Amsterdam, Netherlands). The ServiceXS kit comprised of a total of three probe sets consisting of 50 probe pairs spanning a region of ~500kb and an average distance of ~10kb (β-globin^{xs} MLPA Kit, ServiceXS). The MRC-Holland P102-B2 HBB MLPA probemix contained 28 probes within the β-globin gene cluster region, including several probes for the upstream regulatory sequences and one probe that was specific for the haemoglobin S mutation (rs334) which is frequently found in Africans and African-Americans. All patient samples were tested with known controls and run at least twice. Deletions were detected visually by reduced probe signal strength of one or more probes.

2.10 Genome-wide SNP array

Since 2005, whole-genome SNP arrays have become an important tool for discovering variants that contribute to human diseases and phenotypes. Dr Ruth Clifford (The Oxford Biomedical Research Centre, Oxford) performed whole-genome SNP array analysis on the novel/rare deletions discovered in chapter three. Analysis was carried out on an iScan (Illumina®) and hybridisation to illumina SNP chips (HumanOmni1-Quad) was performed according to the manufacturer's protocols (www.illumina.com/products). The HumanOmni1-Quad featured over one million strategically selected markers that deliver dense genome-wide coverage and include data from the 1000 genomes project. The data was processed using GenomeStudioV2009.2 (Illumina, Inc., San Diego, California, USA) and then analysed using Nexus 6.1 Discovery Edition (BioDiscovery, Inc., El Segundo, California, USA).

2.11 Next-generation whole exome sequencing

Dr Lorna Gregory at The Wellcome Trust Centre for Human Genetics, Oxford performed next-generation whole exome sequencing on case 12 from chapter seven. Analysis was performed on a HiSeq 2500 system run in rapid-run mode (Illumina®) according to the manufacturer's protocols (http://www.illumina.com/products). The library was a Truseq exome library (62Mb capture kit) with 2x100bp paired-end reads. The analysis was run through Platypus (Rimmer *et al*, 2012) and STAMPY (Lunter and Goodson, 2011) software and aligned to the Genome Reference Consortium, GRCh37 sequence (www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/). The remaining data was annotated in ANNOVAR and the gene list checked visually for mutations (Wang *et al*., 2010).

CHAPTER THREE: BETA GLOBIN GENE CLUSTER DELETIONS ASSOCIATED WITH INCREASED HBF LEVELS IN ADULTS

3.1 Introduction

As discussed in chapter 1, the human β-like globin genes are arranged in a single cluster on the short arm of chromosome 11 in their order of expression during development; 5' ε–^{G}γ–^{A}γ–δ–β 3' (Weatherall *et al.*, 1985). Deletions and rearrangements in the β-globin gene cluster can produce 4 distinct phenotypes. The phenotypes are classified according to the particular genes which are deleted and consist of the clinically significant phenotypes of β-thalassaemia, δβ-thalassaemia, εγδβ-thalassaemia and the less clinically significant phenotype of HPFH. At least 40 different deletions larger than 1kb involving the β-globin gene cluster have been described to date in the HbVar database (Giardine *et al.*, 2007, Patrinos *et al.*, 2004).

It is well known that large deletions of the β-globin gene cluster are a significant cause of elevated HbF levels in adults. What is not well known however is the frequency and range of these mutations in the UK as they have traditionally been problematic to diagnose. The work described in this chapter aimed to evaluate the efficacy of the recently available technique of multiplex ligation-dependent probe amplification (MLPA) to determine the frequency and range of these mutations in the UK population and therefore aid in the identification of these deletions within a diagnostic setting.

3.2 Detecting β-globin gene deletions

Deletion mutations in the β-globin gene cluster have traditionally been detected by Southern blot analysis or Gap-PCR. However the technique of Southern blotting is not best suited to a routine diagnostic laboratory as it requires large amounts of DNA and is time consuming and technically demanding. Gap-PCR analysis is a simple and quick technique but can only be used to detect deletions that have previously characterised breakpoint sequences. Consequently Gap-PCR assays have been designed for the 9 deletions/rearrangements which are known to occur frequently (Craig *et al*., 1994). Due to these technical difficulties at the time of carrying out this study the frequency and range of β-globin gene cluster deletion mutations was unknown in most populations.

The technique of MLPA first described by Schouten *et al* (2002) has been shown to be relatively simple and able to detect copy number variations and large deletions of target DNA sequences. This method has been applied successfully to a number of genes in which deletions and duplications are common. In 2005 Harteveld *et al* designed MLPA probe sets for the β-globin gene and α-globin gene clusters and reported the technique to be robust, simple and specific for the detection and characterisation of deletions and duplications. The β-globin gene cluster probe set consisted of three probe sets containing 50 probe pairs spanning a region of 500kb and an average distance of 10kb (Appendix 2). Probes within each set were designed to produce PCR products differing by 2bp in length which allowed separation in the size range 80-125bp.

3.3 Study subjects

Over a 3 year period (December 2005-December 2008) 4,041 patient blood samples were referred to the National Haemoglobinopathy Reference Laboratory in Oxford for haematological and DNA analysis. During this time 316 samples were identified as potentially having a β-globin gene cluster deletion. They had a raised HbF or thalassaemic red cell indices or a combination of both. Point mutations responsible for β-thalassaemia and α-thalassaemia (both deletion and non-deletion mutations) had been excluded in all samples. Individuals with an elevated HbF level due to the hybrid haemoglobins Lepore and Kenya were diagnosed by haematological methods and confirmed by Gap-PCR, and therefore not included in this study.

3.4 Laboratory procedures

Peripheral blood counts and erythrocyte indices were determined using an automated cell counter. Haemoglobin quantifications were carried out using a cation exchange high performance liquid chromatography (HPLC) system (VARIANT^{TM}, Bio-Rad Laboratories, Hercules, CA USA). DNA was extracted from peripheral blood leukocytes by conventional phenol chloroform extraction or on an automated DNA extractor (Chemagen). Deletions of the β-globin gene cluster were determined using the MLPA method described previously (Harteveld *et al*., 2005). The β-globin gene cluster probe sets and reagent kits were supplied (β-globin^{xs} MLPA Kit; ServiceXS, Leiden, The Netherlands). Deletions were detected by reduced signal strength of one or more probes. Deletions with a position and size suggestive of one of the common deletion mutations were confirmed by Gap-PCR (Figure 3.1).

3.5 Results

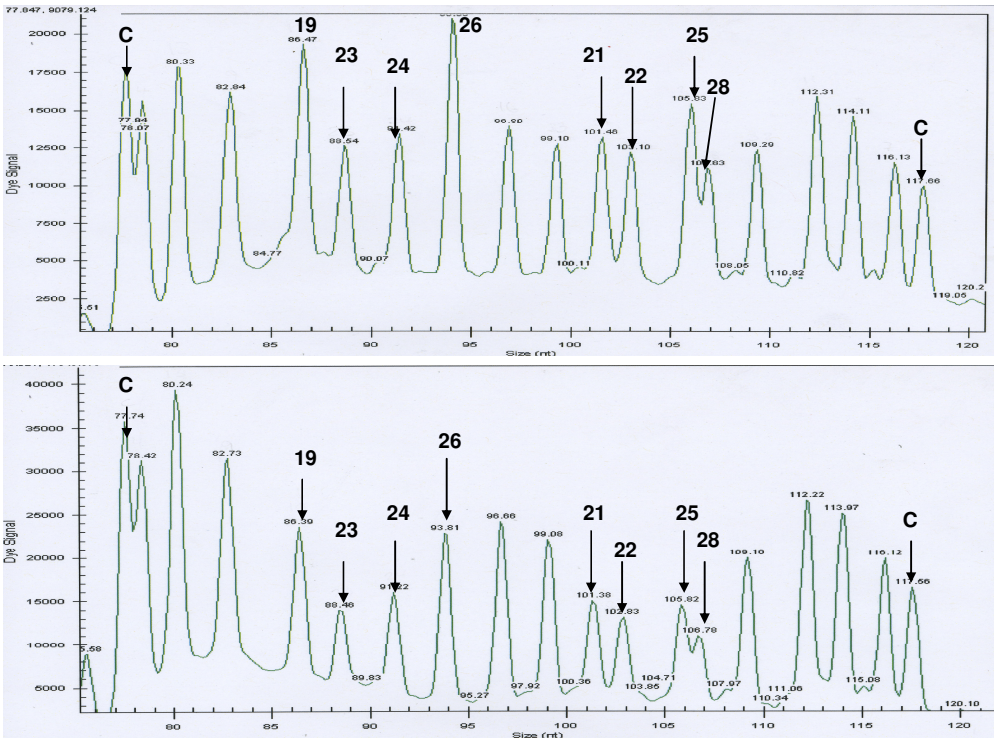
A total of 316 selected samples were analysed by MLPA and where appropriate Gap-PCR for a β-globin gene cluster deletion performed. Seventy five (24%) were found to have a mutation (Figure 3.2). Of these, 52 samples (16.5%) were confirmed by Gap-PCR to have one of the 7 common HPFH or δβ-thalassaemia mutations (Table 3.1). In all cases the genotype matched the phenotype in that patients with a HPFH deletion had the higher levels of HbF (up to 35%) and normal red cell indices. Patients with δβ-thalassaemia had more modest elevations in HbF (4-14%) with hypochromic, microcytic red cell indices. Patients with HPFH and therefore the highest HbF levels were found to have deletions which removed the 3.5kb intergenic region between the $^{A}\gamma$- and δ- globin genes known to be an important region for HbF silencing (Figure 3.2). The lowest HbF levels were seen in deletions which did not remove the 3.5kb intergenic region as in the Turkish and Sicilian $(\delta\beta)^0$-thalassaemia.

The remaining 23 samples (7.3%) had one of 10 different deletions (Figure 3.2). Analysis of the location and size of 6 of these deletions and their associated haematological parameters, together with the ethnicity of the patient, suggested they may be previously reported rare deletions (Table 3.2). However, MLPA on its own does not identify the precise breakpoints of a deletion so without further characterisation it is not possible to be certain as to their identification. The other 4 deletions identified do not match any previously reported deletions and appear to be novel (Table 3.2). These mutations consisted of a 2-8kb β^0-thalassaemia deletion identified in an Afro-Caribbean family, a 0.2-1kb β^0-thalassaemia deletion in an Afghanistan family, a 30-40kb $(\delta\beta)^0$-thalassaemia

deletion in an African patient, and a >160kb $(\varepsilon^{G}\gamma^{A}\gamma\delta\beta)^{0}$ thalassaemia deletion found in an Asian Indian family.

In summary a total of 17 different β-globin gene cluster deletions were identified, comprising 13 previously reported mutations and four new mutations illustrating the heterogeneity of the molecular basis of the β-thalassaemia disorders in the UK. Mutations were found in all major UK population ethnic groups (Asian Indian, African, Mediterranean, Southeast Asian and white British).

a)



b)

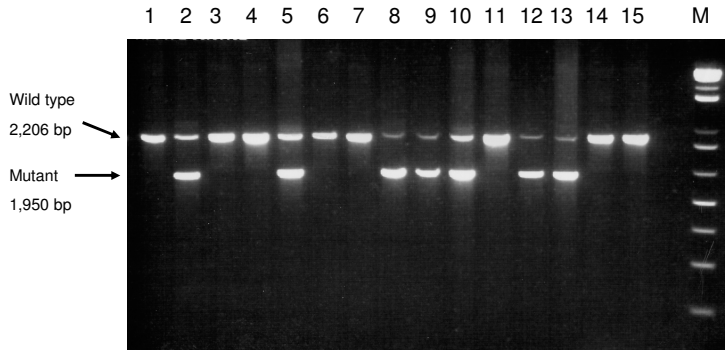


Figure 3.1 HPFH-2 deletion mutation by MLPA and Gap-PCR.
a) MLPA peak patterns for the β-globin gene green probe set (β-globin^{XS} MLPA Kit; ServiceXS. This set comprises of 17 probes and two control probes (marked as C). The upper window shows a normal control and the lower window a patient heterozygous for the HPFH-2 deletion. The HPFH-2 deletion is shown by the reduced peak heights of probes 19-28 b) Gap-PCR for the HPFH-2 deletion mutation. M= 1kb marker. Lane 1= normal control, Lane 2= heterozygous control, Lane 5= patient with the HPFH-2 deletion by MLPA.

Table 3.1 Haematological indices and MLPA results of common β- globin cluster deletions confirmed by Gap-PCR.

Deleted Genes (MLPA)	Size kb	n	HbA_2 %	HbF %	RBC 10^{12}/l	Hb g/dl	MCV fl	MCH pg	Ethnicity	Confirmed Deletion
δβ	106	2	1.6-2	31-32.6	4.2-5.0	8.5-12.2	71-73.2	20.2-20.8	Black	HPFH-1 (Black) (Collins *et al.*, 1987)
δβ	105	25	2.2-3.1	18-36.1	4.03-5.8	9.2-37.2	73.1-91.4	22.3-31.2	Black/Mixed	HPFH-2 (Ghanaian) (Collins *et al.*, 1987)
δβ	48.5	1	2.1	28.6	5.10	13.6	81.7	26.7	Mediterranean	HPFH-3 (Indian) (Kutlar *et al.*, 1984)
δβ	20.6	1	2.5	8.3	5.77	12.0	70.0	20.9	Not given	Turkish $(\delta\beta)^0$ thal (Kulozik *et al.*, 1992)
δβ	13.4	11	2.4-3.6	4.2-14.1	4.79-5.8	9.4-13.3	62.9-74.2	19.6-24.8	Mixed/Caucasian	Sicilian $(\delta\beta)^0$ thal (Ottolenghi *et al.*, 1982)
$^{A}\gamma\delta\beta$	100	5	2.4-3.1	13.2-17.8	5.96-6.4	12-16.3	73.1-77.3	22.6-25.2	Asian/Chinese	Chinese $(\delta\beta)^0$ thal (Jones *et al.*, 1981b)
$^{A}\gamma\delta\beta$	23.4	7	2.1-3.2	14.1-17.8	5.1-6.5	10-14.5	66-70.2	20.8-22.9	Indian	Indian $^{G}\gamma(^{A}\gamma\delta\beta)^0$ thal (Jones *et al.*, 1981a)

Table 3.2 Haematological indices and MLPA results of rare/novel β– globin cluster deletions.

Deleted Genes (MLPA)	Size kb	n	HbA_2 %	HbF %	RBC 10^{12}/l	Hb g/dl	MCV fl	MCH pg	Ethnicity	Previously reported candidate deletion
$\varepsilon^G\gamma^A\gamma\delta\beta$	>200	1	3.7	0.9	6.7	13.1	63.0	19.0	White British	Scottish-Irish $(\varepsilon^G\gamma^A\gamma\delta\beta)^0$ thal (Trent *et al*., 1990)
$\varepsilon^G\gamma^A\gamma\delta\beta$	>160	4*	3.0-3.7	0.0-0.7	5.6-6.8	11-13.7	58-69	17.3-20.3	Asian Indian	None - Novel
δβ	25-45	3	3.1-3.3	12-25	4.9-5.7	11-13.7	67.4-71.3	22-24.1	Indian	Indian $^G\gamma^A\gamma(\delta\beta)^0$ thal (Mishima *et al*., 1989)
δβ	8-12	1	2.9	11.5	5.9	13.6	69.1	23	White British	East European $^G\gamma^A\gamma(\delta\beta)^0$ thal (Palena *et al*., 1994)
δβ	10-12	1	2.3	21.9	5.24	12.7	73.6	24.3	Thai	Vietnamese $^G\gamma^A\gamma(\delta\beta)^0$ thal (Craig *et al*., 1994)
δβ	30-40	1	2.8	16.4	6.1	16.3	77.7	25.2	African	None - Novel
β	>45	6	7.3-8.0	1.6-4.0	4.4-6.6	9.4-13.9	62-66.7	18.7-21.6	Filipino	Filipino β^0 thal (Waye *et al*., 1994)
β	10-12	1	8.3	3.1	4.6	9.7	61.9	21.3	Not given	Australian β^0 thal (Motum *et al*., 1992)
β	0.2-1	3*	8.6-9.0	1.7-3.9	5.6-6.2	9.7-10.5	61-61.9	18.6-21.3	Afghanistan	None - Novel
β	2-8	2*	7.3	8.4	4.54	11.2	72	24.7	Afro-Caribbean	None - Novel

* all members of the same family

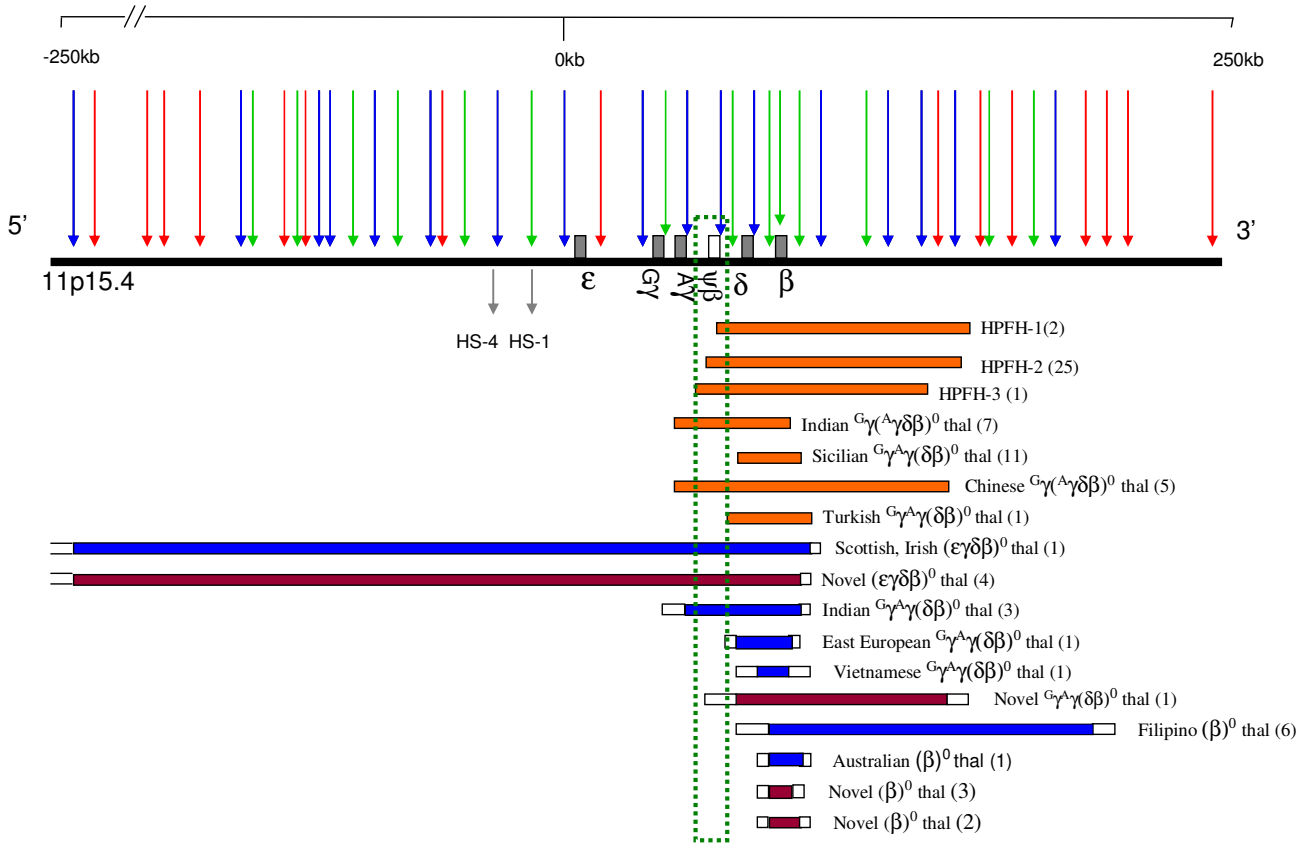


Figure 3.2 The human β-globin gene cluster on chromosome 11p.

The ε,^{G}γ, ^{A}γ, δ and β genes are indicated by grey boxes. The vertical arrows indicate the positions of the MLPA probe sets on the cluster and the colour of the arrow corresponds to the fluorescent dye label (green, blue or red) used in the MLPA reaction. The grey vertical arrows represent the hypersensitivity sites 1 and 4 for which two of the MLPA probes target. The positions of the 17 different deletions are shown below the β–globin cluster. The orange horizontal bars represent the common deletions which were confirmed by Gap-PCR. The blue horizontal bars show the rare deletions which have a possible candidate. The burgundy horizontal bars show the deletions which are possibly novel. The white boxes represent the positions between the probes with which further characterisation is needed in order to confirm the precise break points. The green hatched box represents the location of the 3.5kb ^{A}γ-δ globin intergenic region important in γ-globin gene silencing. Deletions removing this region had the highest levels of HbF.

3.6 Discussion

This study showed the efficacy of MLPA at identifying β-globin gene cluster deletions in a routine diagnostic laboratory setting. Fifty two of the deletion mutations detected were subsequently confirmed by Gap-PCR which validated the accuracy and sensitivity of the method. The technique only required the use of standard DNA laboratory equipment (thermal cyclers and capillary electrophoresis) and was found to be robust in routine use. It was also much quicker than conventional Southern blot analysis in that results could be obtained within two days. The main limitation of the method was that it could only detect deletions/duplications that alter overall copy number; other rearrangements such as simple inversions will be missed. Also, the MLPA technique on its own cannot define the exact breakpoints of a deletion; however this is not usually required in a diagnostic setting. If full characterisation of a particular deletion is required, a simple approach is to design sets of PCR primers which flank the deleted region indicated by the MLPA results. The primer pairs are then used to amplify across the deletion as Gap-PCR primers. The resulting PCR product can then be sequenced and the breakpoints determined by comparison with the normal β-globin gene cluster sequence in this region.

These results showed that β-globin gene cluster deletions were present at significant frequencies in the UK population. Over the 3 year period of the study the 75 patients identified with a deletion comprised 11% of the total number of patients/families referred for investigation of β-thalassaemia. All 7 of the common β-globin gene deletion/rearrangement mutations described by Craig *et*

al (1994) were observed and comprised 69% of the total mutations found. The most common deletions were HPFH-2, the Sicilian δβ-thalassaemia deletion, the Indian inversion δβ-thalassaemia mutation and the Chinese δβ-thalassaemia deletion (Table 3.I). Much rarer were the HPFH-1, HPFH-3 and the Turkish δβ-thalassaemia deletions which were each found only once or twice. The remaining 31% (23) of patients had one of 10 different rare or novel deletion mutations (Table 3.2). These would have been missed by a conventional Gap-PCR screen for the 7 common mutations. These results showed that rare/novel mutations comprise a significant proportion of the β-globin gene cluster deletion mutations in the UK population and indicate that the use of MLPA to detect them is warranted.

Two of the rare/novel mutations identified were particularly interesting as they had a deletion of the ε,γ,δ and β-globin genes, and thus the patients exhibited the rare phenotype of $(\varepsilon^{G}\gamma^{A}\gamma\delta\beta)^{0}$- thalassaemia trait. At the time of performing this analysis only 16 different $(\varepsilon^{G}\gamma^{A}\gamma\delta\beta)^{0}$- thalassaemia mutations had previously been reported in the literature (Rooks *et al*., 2005, Furuya *et al*., 2008). Of these 16 large deletions, 11 remove all or a greater part of the β-globin gene cluster but 5 leave the β-globin gene intact but inactivated. The two deletion mutations described here were of the category with the β-globin gene deleted. The first matched the previously described Scottish-Irish mutation. The second was found in an Asian Indian family and is most likely novel.

Two hundred and forty one of the 316 samples studied did not have a large deletion of the β-globin gene cluster. Beta-thalassaemia point mutations and

α-thalassaemia mutations (both deletion and non-deletion mutations) had previously been excluded in all samples. Therefore the reasons for the raised HbF and/or thalassaemic red cell indices in these 241 patients remained unclear. It was difficult to ascertain whether comprehensive iron studies had been carried out on all these patients and was therefore likely that at least some will simply have iron deficiency anaemia. It was also possible that some will have a β-globin gene cluster rearrangement that cannot be detected by MLPA; these include small deletions that lie between the MLPA probes and gene rearrangements such as a simple inversion which effectively silences the genes but does not alter gene copy number. Other patients may have thalassaemia or high HbF levels due to as yet unidentified genetic determinates.

A significant proportion of the samples in which a mutation was not identified had normal red cell indices and a modest increase in HbF consistent with non-deletion HPFH. Sequence analysis of the γ-globin genes in these patients would be necessary in order to eliminate any variants or point mutations such as the *Xmn*I-$^{G}\gamma$ mutation which has been shown to be a major locus responsible for increasing HbF (Gilman and Huisman, 1985). It has been suggested that >50% of the variance seen in HbF levels is unlinked to the β-globin gene locus and two other major loci have been identified and are unlinked to chromosome 11p15 (Garner *et al*., 2000). These are the *HBS1L-MYB* intergenic region on chromosome 6q23 and *BCL11A* on chromosome 2p15 (Thein *et al*., 2007, Menzel *et al*., 2007). The most significant of these is the *BCL11A* gene on chromosome 2 which was identified as a critical factor in the down regulation of HbF expression in adult erythrocytes (Sankaran *et al*., 2008). Therefore it was

also highly probable that some of the samples in this group may be linked to one of the above loci unlinked to the β-globin gene locus.

Although it is important to detect β-globin gene cluster deletion mutations for clinical reasons it is also useful for research purposes. By mapping novel mutations and studying the genotype-phenotype relationship of these deletions our understanding of the mechanisms involved in the control of the fetal and adult globin gene expression may be enhanced. This study alone has identified 4 possible novel deletions within the UK population, the simplicity of the MLPA technique makes it likely that many more will be detected.

In summary the results of this study have demonstrated the wide spectrum of β-globin gene cluster deletion mutations in the UK population, in which 13 known deletion mutations have been identified (22% of the published spectrum of 60 for all ethnic groups) plus four novel deletions. The MLPA method is simple and reliable and is recommended as the first choice method for screening for these clinically significant large deletion mutations.

CHAPTER FOUR: MAPPING OF PRECISE BREAKPOINTS IN NOVEL DELETIONS

4.1 Introduction

The results from the work described in chapter 3 demonstrated that the technique of MLPA was able to identify deletions in the β-globin gene clusters of patients referred to our laboratory for haemoglobinopathy investigations. As discussed this had importance for clinical reasons and was also useful for research purposes. The mapping of novel deletions may aid in the understanding of the genotype-phenotype relationships and also shed some light on the mechanisms involved in the fetal to adult haemoglobin switching process. The work described in chapter 3 led to the identification of four possible novel deletions in the β-globin gene clusters of patients from the UK population. However, the technique of MLPA was not able to define the exact breakpoints and therefore mapping using Gap-PCR methodology would need to be undertaken to characterise the deletions. The work described in this chapter aimed to fully characterise the breakpoints of the four novel deletions (deletion 1-4).

4.2 Deletion 1

This was the first reported case of a β-globin gene deletion mutation in a family of Afghan ancestry. The proband was a 10 year old Afghan female who had been referred for β-thalassaemia genotyping. She was transfusion dependent and had the phenotype of β-thalassaemia major with a complete absence of HbA (Table 4.1). Her parents are consanguineous (first cousins) and both had reduced red cell indices and elevated HbA_2 levels consistent with β-thalassaemia trait.

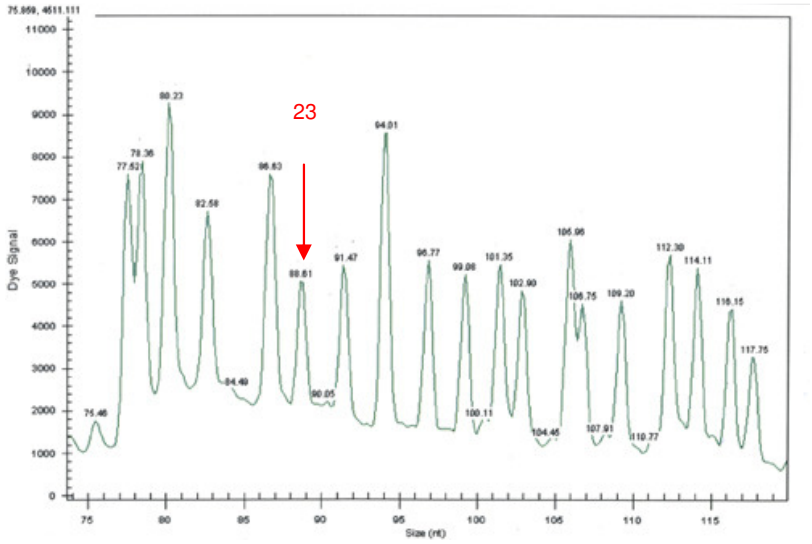
Table 4.1 Haematological parameters for deletion 1.

	Hb g/dl	RBC $x10^{12}$/l	MCV fl	MCH pg	HbA_2 %	HbF %
Mother	10.5	5.66	61.0	18.6	9.0	1.7
Father	13.2	6.01	65.0	17.9	8.6	3.9
Proband*	5.6	2.21	64.2	19.0	3.3	95.9

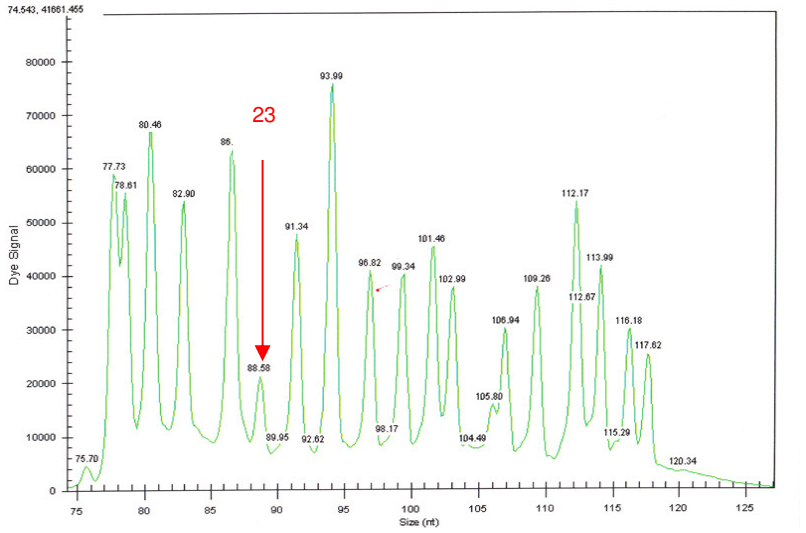
*Pre-transfusion sample

Initial molecular investigations by ARMS-PCR and β-globin gene sequence analysis (from the promoter to the polyadenylation signal region) failed to identify a β-thalassaemia point mutation. Further investigations using the technique of MLPA revealed a reduction in the height of a single probe for both parents and a complete absence of the same probe in the proband (Figure 4.1). The deleted probe was located in the β-globin gene promoter region. All other surrounding probes were normal. Cases of a reduction in the height of a single probe must be interpreted with caution as the reduced peak height could be due the presence of a rare sequence variant under the ligation site. However in this case the proband's β-globin gene promoter region had been sequenced previously and confirmed to be normal. Therefore the results indicated that the parents are carriers of a small deletion at the 5' end of the β-globin gene, and the proband was homozygous for this deletion mutation. The MLPA results were not consistent with any previously reported β-globin gene deletion.

a)



b)



c)

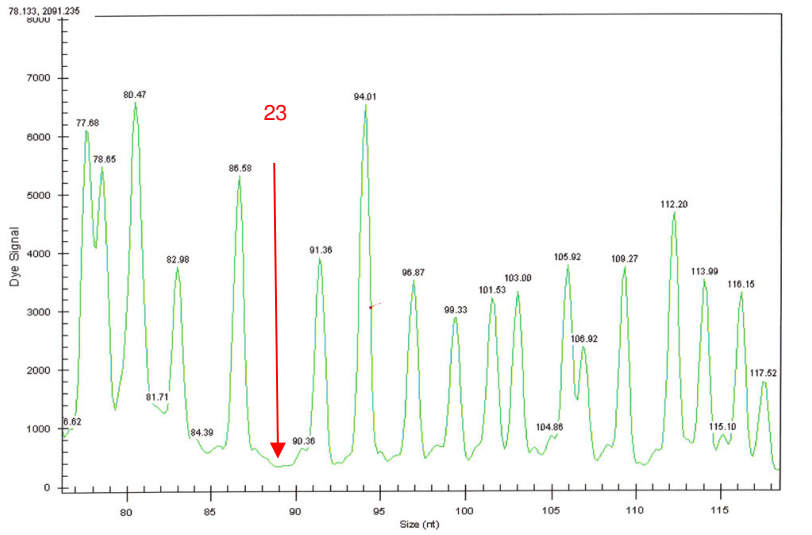


Figure 4.1 MLPA results for deletion 1 (β-globin^{xs} MLPA Kit; ServiceXS).

a) Normal control (arrow indicates normal probe height for probe 23)
b) Mother/father (arrow indicates reduced height for probe 23)
c) Proband (arrow indicates complete absence of probe 23)

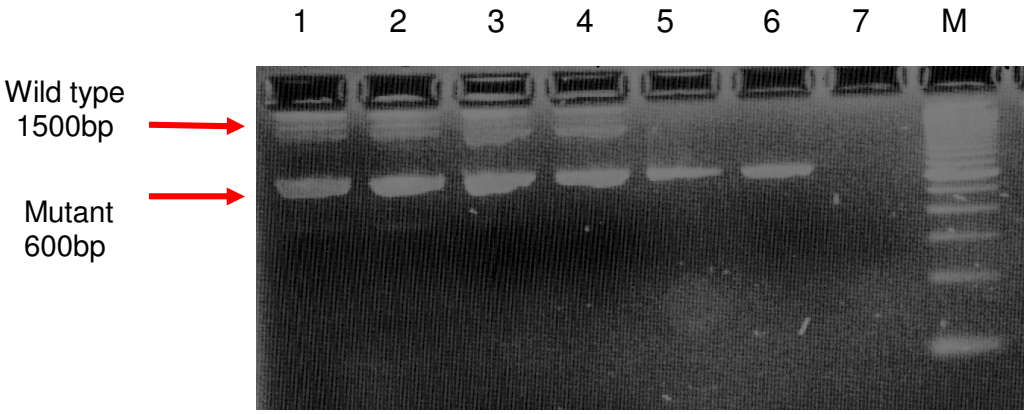
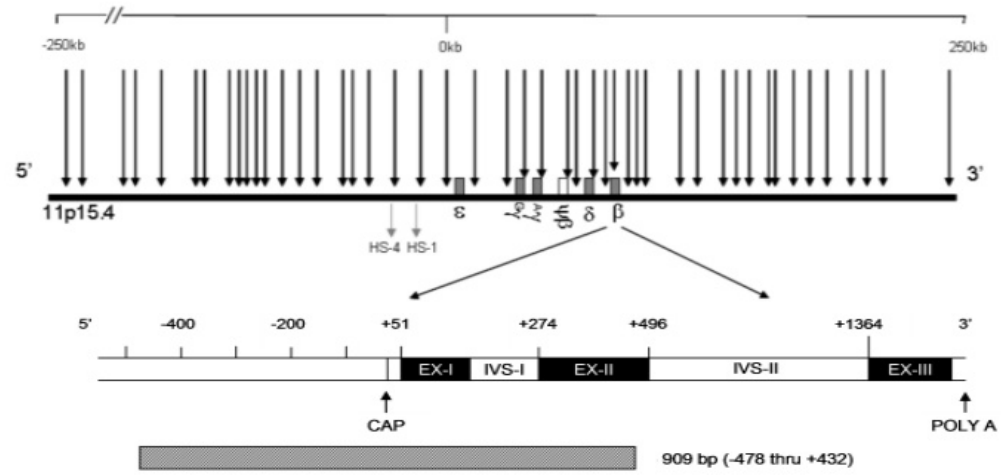


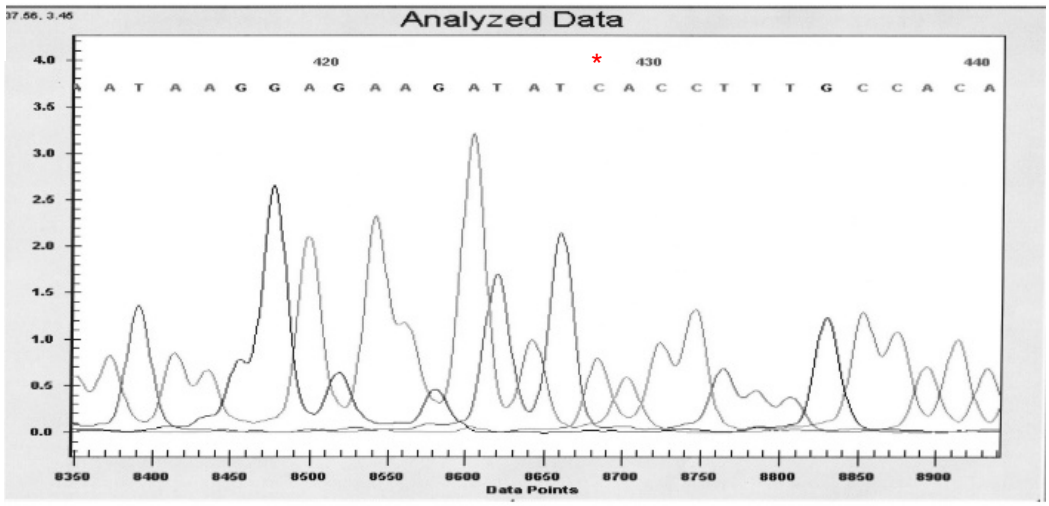
Figure 4.2 Gap-PCR amplification products in deletion 1.
Lane 1 & 2= mother, lane 3 & 4= father, lane 5 & 6= proband, lane 7= blank lane M= 100bp DNA ladder (New England BioLabs).

PCR primers were designed (forward TAATCTGAGCCAAGTAGAAG and reverse CACTGATGCAATCATTCGTC) to flank the deleted region indicated by the MLPA results. The primer pair was then used to amplify across the deletion as Gap-PCR primers (Figure 4.2). Amplifications were performed in 25µl tubes with the standard PCR reagents (see section 2.5). The PCR conditions involved an initial denaturation of 3 minutes at 95^0C, then 30 cycles of 95^0C for 1 minute, 56^0C for 2 minutes, 72^0C for 2 minutes and ending with a final extension of 72^0C for 10 minutes. Some of the resulting product (10µl) was electrophoresed through a 1.5% agarose, 1 x TBE gel at 100 Volts for one hour (Figure 4.2). The remaining product (15µl) was sequenced using a Beckman Coulter DNA sequencer and the breakpoints determined by comparison with the normal β-globin gene cluster sequence (Figure 4.3).

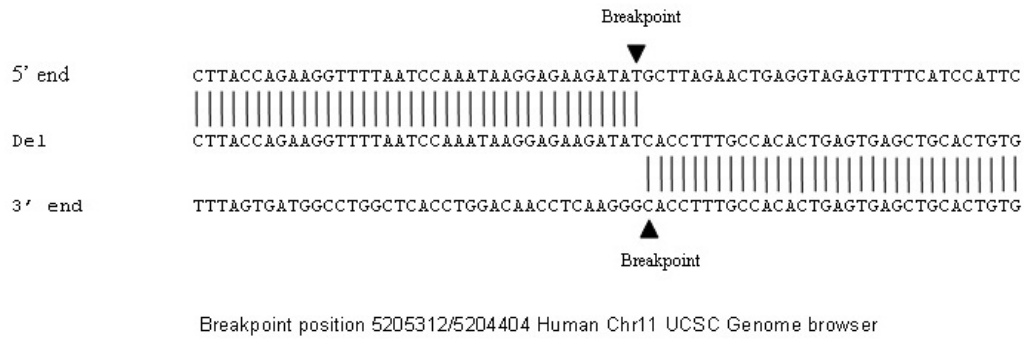
a)



b)



c)



Breakpoint position 5205312/5204404 Human Chr11 UCSC Genome browser

Figure 4.3 Sequence analysis results (β-globin gene) for deletion 1.

a) Diagram to illustrate the structure of the β-globin gene cluster and the positions of all the MLPA probes (black vertical arrows) used in the analysis. Below the cluster is a diagram showing the size and location of the deletion in relation to the β-globin gene.

b) Sequence chromatogram for the proband, where * denotes the deletion breakpoint.

c) The sequence of the junction fragment showing the 5' normal, mutant and 3' normal sequences. The vertical bars illustrate the positions at which the sequences are identical.

The proband was found to be homozygous for a novel 909 bp deletion (-478 through to + 432) removing the β-globin gene promoter, exon I and IVS-I entirely and most of exon II of the β-globin gene (Figure 4a). The 5' breakpoint was demonstrated to lie at position 5,205,311 and the 3' breakpoint at 5,204,404 according to the UCSC human chromosome 11 genome browser. Examination of the nucleotide sequences upstream and downstream of the junction breakpoints did not reveal any significant sequence homology suggesting that the deletion was caused by a non homologous recombination event (Figure 4.3b, 4.3c). However, there were AT-rich stretches of sequences around the 5' breakpoint which have been previously postulated as significant in the generation of β-globin gene deletions (Thein *et al*., 1989, Huang *et al*., 2008).

Very little is known regarding the frequency and range of thalassaemia mutations in Afghanistan. The frequency of β-thalassaemia carriers had been estimated at 3% but there was no published data on the percentage of affected individuals (Galanello *et al*., 2003). Several unusual or novel β^0-thalassaemia mutations had been previously reported in immigrant Afghan families (Giordano *et al*., 1999, Krugluger and Hopmeier, 2002, Kornblit *et al*., 2005) but there were no published studies of the spectrum of thalassaemia mutations in this population (Henderson *et al*., 2009). This novel mutation was the first reported case of a β-globin gene deletion mutation in the Afghan population and indicated that the Afghani population had a complex heterogeneous array of haemoglobin disorders due to invasions of different ethnic groups during various periods in history. The impact of racially heterogeneous immigrant

populations in European countries was responsible for a large increase in the number and variety of mutations encountered, creating challenging problems in carrier diagnosis, genetic counselling and prenatal diagnosis.

4.3 Deletion 2

An Afro-Caribbean antenatal lady was referred for prenatal diagnosis for a possible risk to her fetus of HbSickle/β-thalassaemia. Her partner was a sickle cell carrier and she had been identified on antenatal screening as having reduced red cell indices and an elevated HbA_2 level consistent with β-thalassaemia trait. A sample was referred from the mother and an amniotic fluid sample from the fetus but the father refused any further testing.

Table 4.2 Haematological parameters on the mother in deletion 2.

	Hb g/dl	RBC $x10^{12}$/l	MCV fl	MCH pg	HbA_2 %	HbF %
Mother	11.2	4.5	72.0	24.7	7.3	8.4

Initial molecular investigations by ARMS-PCR and β-globin gene sequence analysis failed to identify a β-thalassaemia point mutation in both the mother and the fetus. The fetus also tested negative for the sickle mutation by ARMS-PCR confirming that the fetus had not inherited the father's mutation. Further investigations using β-globin gene cluster MLPA probe sets (β-globin^{XS} MLPA Kit; ServiceXS and the P102-B2 HBB; MRC-HOLLAND) revealed a reduction in height of the probes shown in Figure 4.4 for both the mother and the fetus.

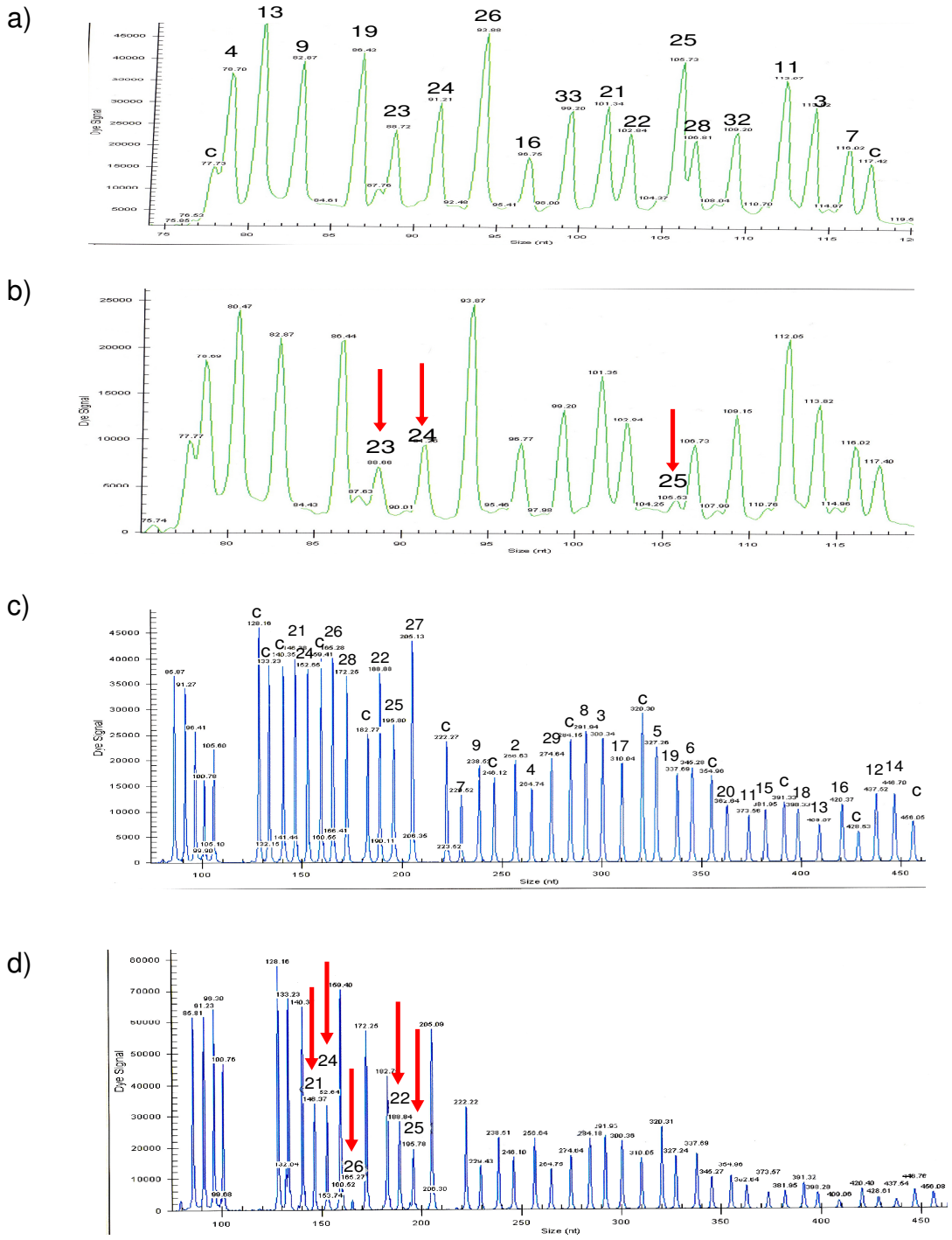


Figure 4.4 β-globin gene cluster MLPA results for deletion 2.
Numbered peaks represent individual probes along the β-globin gene cluster. Peaks denoted by the letter c= control peaks. a) Normal control patient sample analysed with the ServiceXS MLPA kit b) Mother's sample analysed with the ServiceXS MLPA kit c) Normal control patient sample analysed with the MRC-Holland MLPA kit d) Mother's sample analysed with the MRC-Holland MLPA kit. Red arrows indicate probes reduced in height when compared to the normal control and control peaks within each analysis.

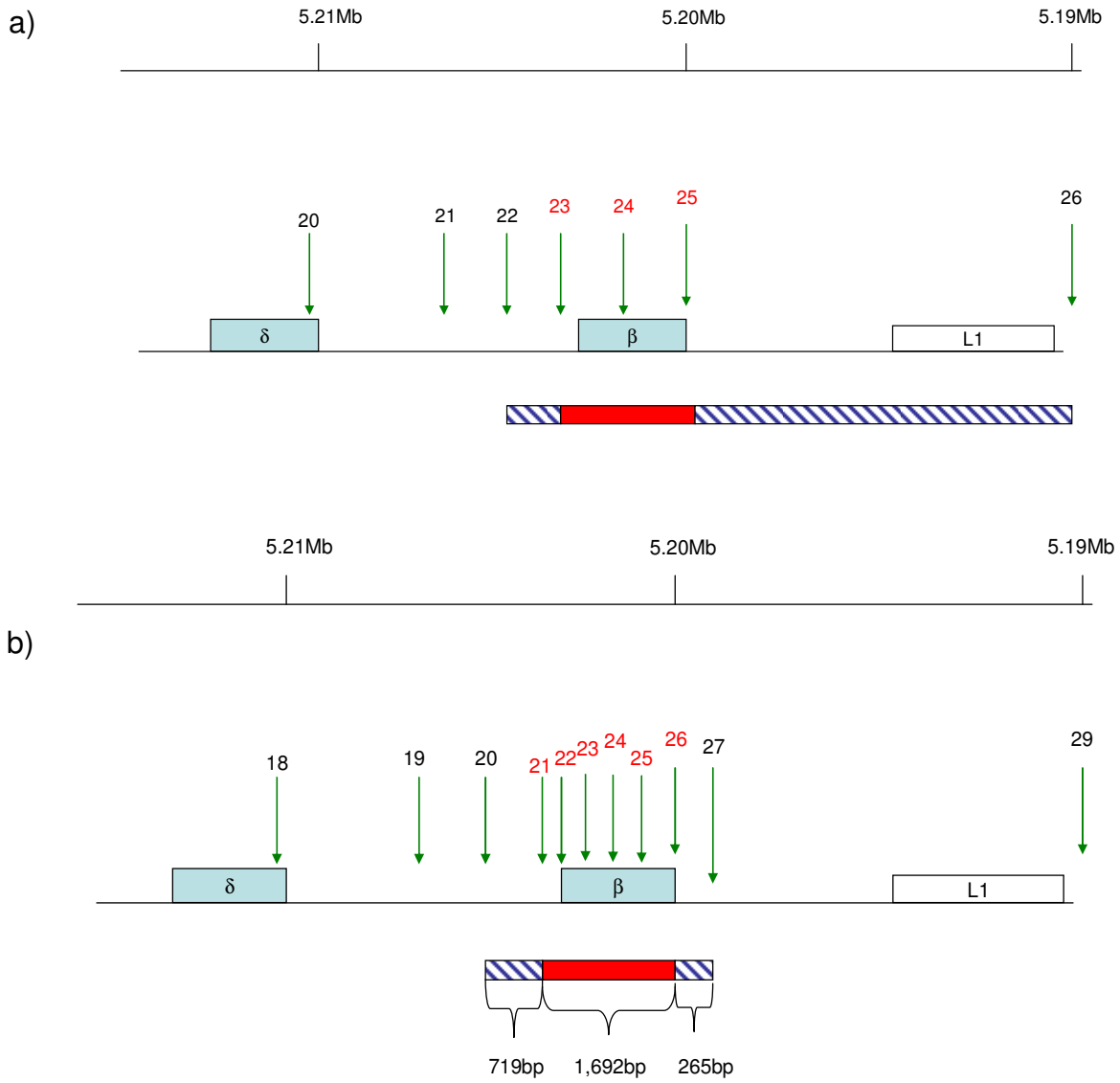


Figure 4.5 Schematic representations of the MLPA probes deleted in deletion 2. a) ServiceXS MLPA probe set and b) MRC-Holland P102-B2 MLPA probe set. The genes are shown as light blue boxes; arrows denote the probe position with the numbered red probes having reduced height. The deletion is illustrated below the genes as a red bar and the hatched blue areas show the possible breakpoint regions.

The deleted probes show that the whole of the β-globin gene is removed by the deletion. All other surrounding probes were normal. The MLPA results were not consistent with any previously reported β-globin gene deletion.

PCR primers were designed to flank the deleted region indicated by the MLPA results (forward TAATCTGAGCCAAGTAGAAG and reverse CCTCTACTTGAATCCTTTTC) with the maximum size of the deletion being 2,676bp and the minimum deletion size of 1,692bp (Figure 4.5). The primer pair was then amplified across the deletion as Gap-PCR primers using the QIAGEN Mastermix protocol (section 2.5) with an annealing temperature of 60^0C. The resulting fragment was sequenced using an ABI-PRISM 3100 automated DNA sequencer (Applied Biosystems) and the breakpoints were determined by comparison with the normal β-globin gene cluster sequence (Figure 4.6).

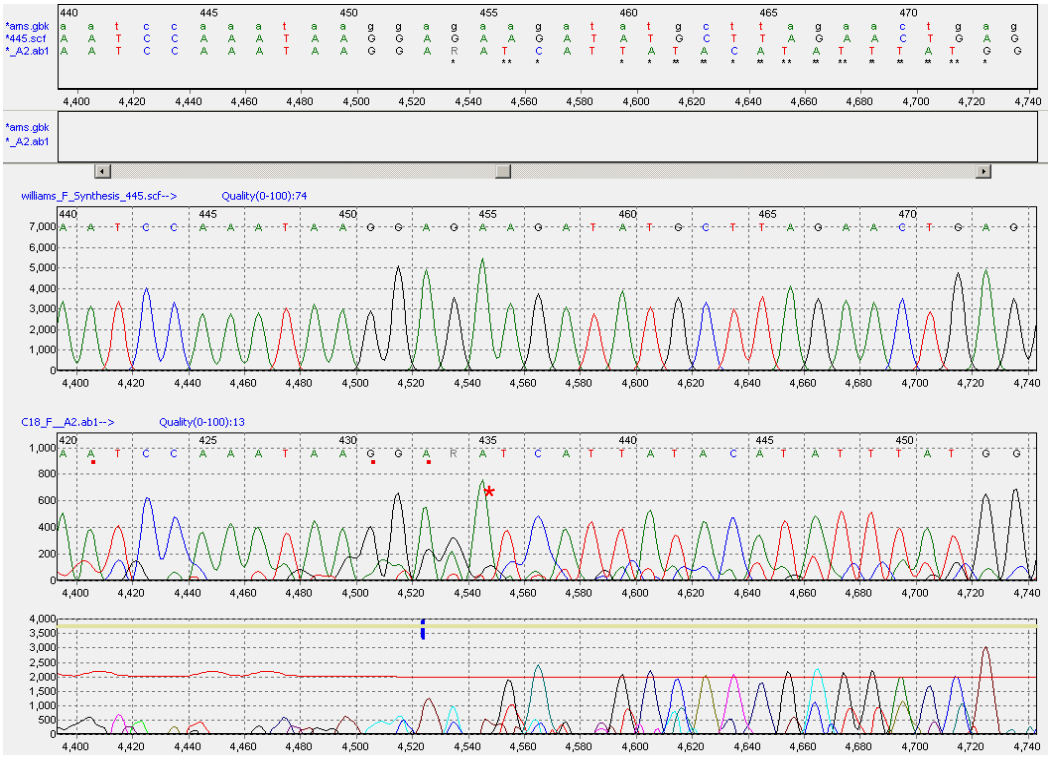


Figure 4.6 Analysed sequence data (Mutation Surveyor®) on the mother showing the deletion breakpoint. The top electropherogram is the reference trace, the middle the test sample trace and the bottom is the mutation electropherogram which highlights any differences between the reference and test sample traces as spikes. * denotes the possible deletion breakpoint.

The sequence analysis results showed a deletion of 1,393bp (HGVS: NG_000007.3:g.70060_71452del1393) (Figure 4.7) which fitted exactly with a previously reported deletion found in African-American and British patients (Anand *et al*., 1988, Thein *et al*., 1989). The deletion extends from 485bp 5' to the mRNA CAP site to the middle of the second intervening sequence.

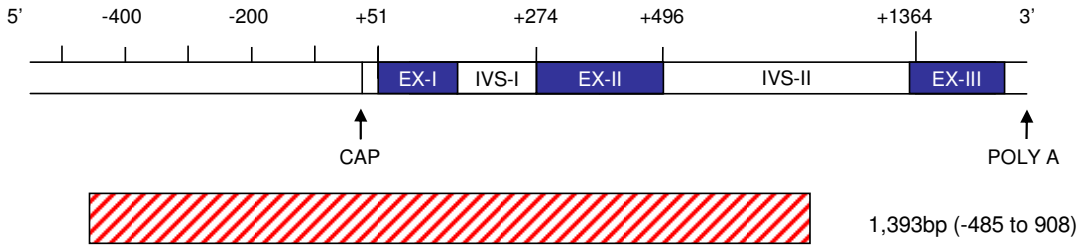


Figure 4.7 Diagram showing the size and location of the 1,393bp deletion in relation to the β-globin gene. The hatched red bar represents the size of the deletion.

Both the ServiceXS and MRC-Holland MLPA probe sets sized the deletion at a minimum size of 1,692bp, which was greater than its actual size of 1,393bp. This difference was most likely due to a common polymorphism in the target sequence which resides where probe 25 (ServiceXS) and 26 (MRC-Holland) anneal. It has been well documented than one of the limitations of the MLPA technique is that probes will be deleted in the presence of a polymorphism (Stuppia *et al*., 2012). This can lead to deletions appearing larger than they actually are and highlights why only having one probe reduced in height is not always a reliable indicator of a deletion.

4.4 Deletion 3

An African man was referred from his GP as he had an unexplained raised HbF level and a persistently low MCV (Table 4.3).

Table 4.3 Haematological parameters on deletion 3.

Hb g/dl	RBC $x10^{12}$/l	MCV fl	MCH pg	HbA_2 %	HbF %
16.3	6.1	77.7	25.2	2.8	16.4

His phenotype was consistent with a diagnosis of δβ−thalassaemia trait. MLPA results (Figure 4.8) revealed heterozygosity for a large deletion in the β-globin gene cluster which as predicted removed the δ and β-globin genes and possibly the $^{A}\gamma$-gene. Both the ServiceXS and MRC-Holland probe sets showed the same deleted region. All three ServiceXS probe sets (green, red and blue) were analysed with only the green and blue sets showing probes with reduced height. From the results the size of the deletion was calculated to be between 32,275bp and 44,303bp (Figure 4.9).

Thirty different primer pairs were designed spanning the breakpoint regions and long range PCR (section 2.5.1) was performed in an attempt to produce an amplification product for sequence analysis. None of the primer pairs amplified which might be explained by the deletion size being incorrect. If there was a polymorphism in the target sequence of one of the probes, like in deletion 2, the size may actually be smaller.

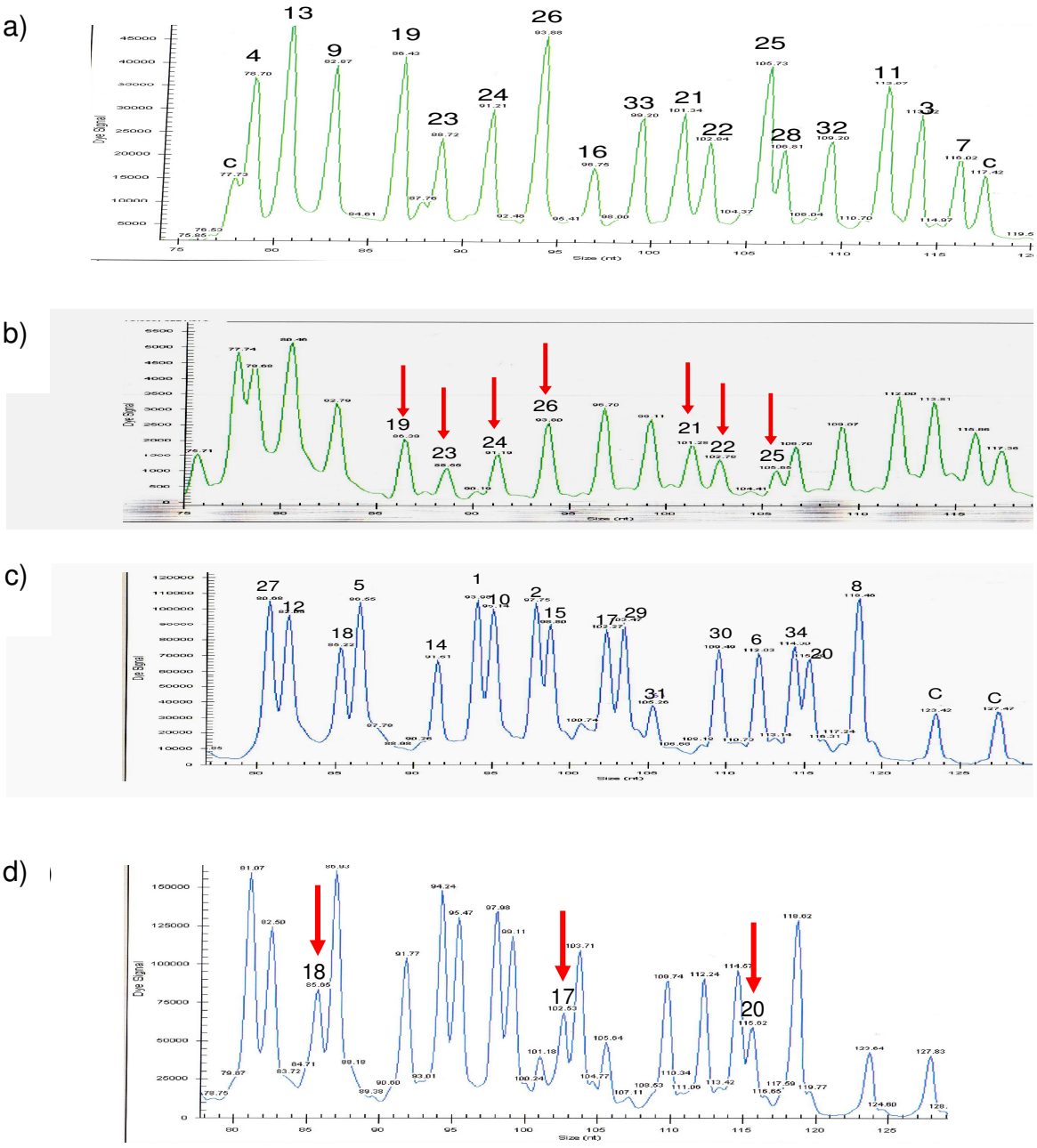


Figure 4.8 ServiceXS MLPA results for deletion 3. Numbered peaks represent individual probes along the β-globin gene cluster. Peaks denoted by the letter c= control peaks. a) Normal control patient sample analysed with the ServiceXS green probe set b) Patient sample analysed with the ServiceXS green probe set c) Normal control patient sample analysed with the ServiceXS blue probe set d) Patient sample analysed with the ServiceXS blue probe set. Red arrows indicate probes reduced in height compared to the normal control and control peaks within each analysis.

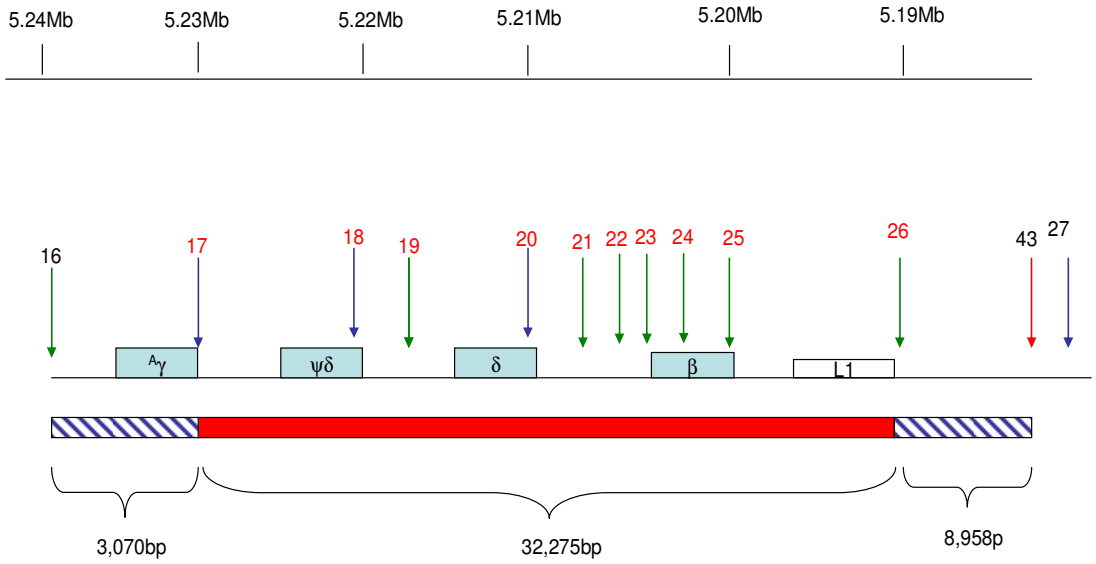


Figure 4.9 Schematic representation of the probes reduced in height in deletion 3 (ServiceXS MLPA set). The genes are shown as light blue boxes and the arrows denote the probe position with the probes numbered red having reduced height. The deletion is illustrated below the genes as a red bar and the hatched blue areas show the breakpoint regions.

As precise characterisation of the breakpoints in deletion 3 was proving difficult it was decided to try another approach. The development of microarray technology over the last few years has proven to be an essential tool in looking for copy number variation (Phylipsen *et al*., 2012). It was decided to analyse this sample on an iScan microarray (Illumina®) using the HumanOmni1-Quad chip with the analysis performed by Dr Ruth Clifford (http://www.illumina.com/products). This chip had over one million strategically selected markers therefore offering dense genome-wide coverage and also included data from the 1000 genomes project. The data was processed using GenomeStudioV2009.2 (Illumina, Inc., San Diego, California, USA) and then analysed using Nexus 6.1 Discovery Edition (BioDiscovery, Inc., El Segundo, California, USA).

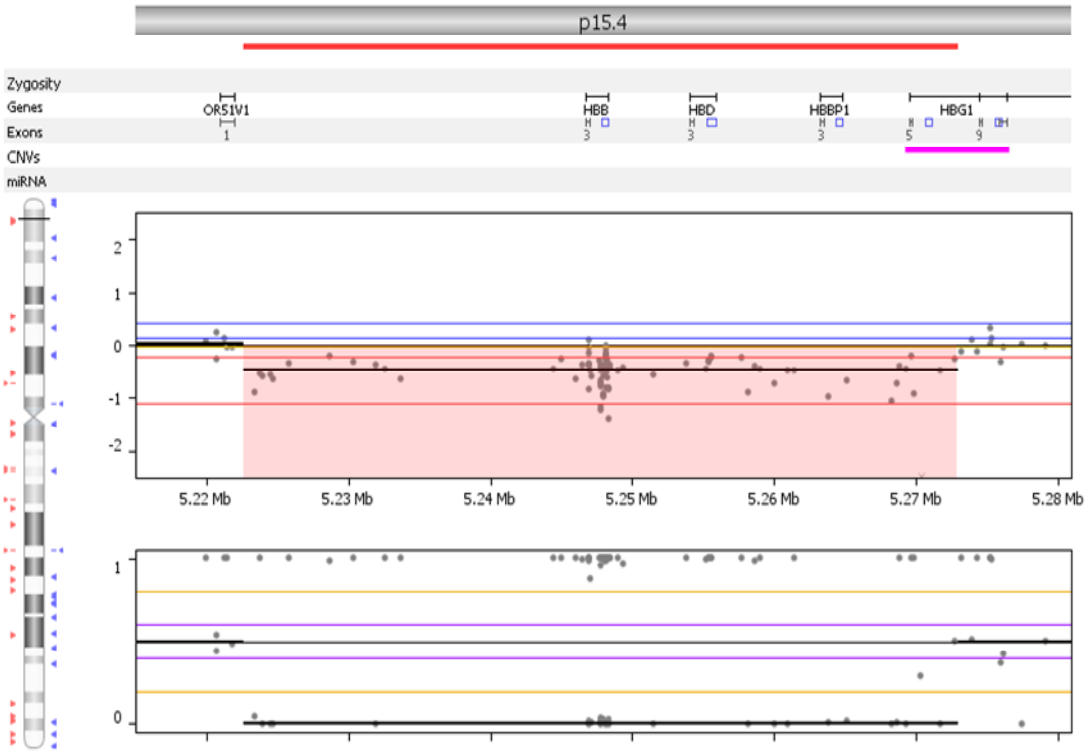


Figure 4.10 Microarray (iScan) results for deletion 3. The top panel shows the position of the deletion represented by a red bar with the genes affected below. The middle panel is the Log_2 Ratio track and the lower panel the B allele frequency plot. Both plots show heterozygosity for a large deletion. Single nucleotide polymorphism (SNP) probes are represented by grey dots.

The results of the microarray confirmed the presence of a heterozygous deletion removing part of the $^{A}\gamma$-gene and all of the δ- and β-globin genes (Figure 4.10). Forty three probes covered the deleted region and the size of the deletion ranged from 49,247bp to 51,322bp. The microarray results suggested that the MLPA results had underestimated the size of the deletion. The 5' end of the deletion seemed comparable with both techniques, removing part of the $^{A}\gamma$-gene however the 3' end appeared on the microarray to be larger than the MLPA results. Probe 43 with the ServiceXS MLPA probe set (Figure 4.9 and 4.11) was not reduced in height yet would appear to be deleted on the microarray.

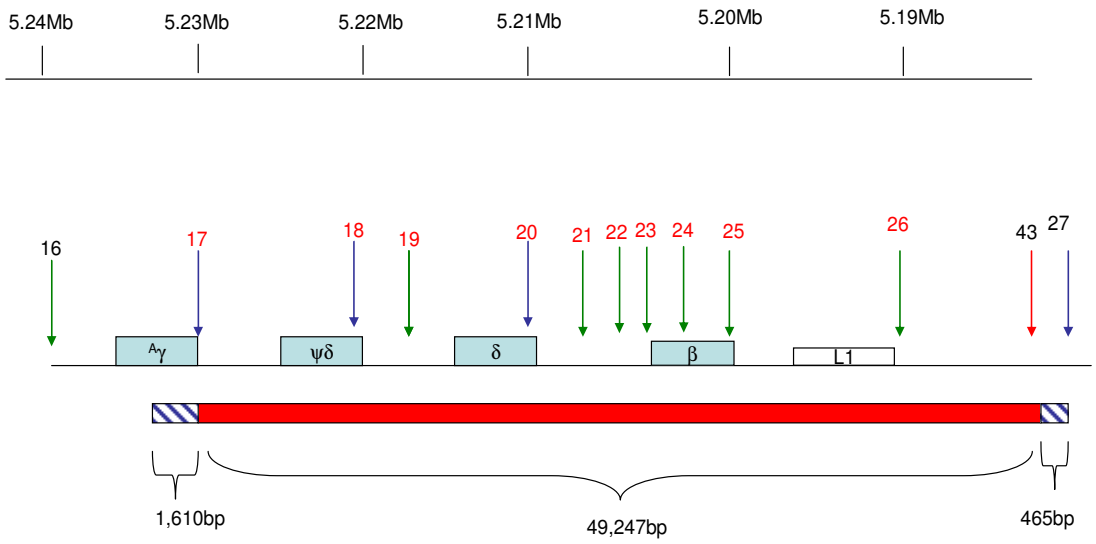


Figure 4.11 Schematic representation of the probes reduced in height in deletion 3 (ServiceXS) with the size of the deletion from the microarray. The genes are shown as light blue boxes and the arrows denote the probe positions with the probes numbered red having reduced height. The SNP array deletion size is illustrated below the genes as a red bar and the hatched blue areas show the breakpoint regions.

In order to explain why probe 43 failed to show any reduction in height, 5 known positive control samples with HPFH-2 (which would be expected to remove probe 43) were analysed using the same ServiceXS MLPA kit. The results showed that probe 43 failed to be reduced in height in all 5 control samples and therefore was unreliable. One possible explanation for this finding was that probe 43 was annealing to an alternative site in the genome.

In the literature 14 $^{A}\gamma\delta\beta$-thalassaemias had been reported and found in most ethnic groups (Table 1.2). Only one $^{A}\gamma\delta\beta$-thalassaemia had been reported in Africans with a deletion size of 35.81kb (Henthorn *et al.*, 1985) which was smaller in size than deletion 3. However there was one $^{A}\gamma\delta\beta$-thalassaemia with a deletion size of 50kb which was similar to the size of deletion 3. This deletion

had been found previously in Belgians (Losekoot *et al*., 1991). Therefore deletion 3 could be a novel African $^{A}\gamma\delta\beta$-thalassaemia or the Belgian $^{A}\gamma\delta\beta$-thalassaemia found for the first time in an individual of African origin. The design of primers flanking the deletion breakpoints determined from the microarray data should enable confirmation of its precise size.

4.5 Deletion 4

Four Asian Indian family members were referred for investigation as their 2 month old son (case 19) had unexplained persistent anaemia in the newborn period. The child's grandmother (case 2) was under the care of the same clinician and had been prescribed oral iron for 10 years without a response in her red cell indices. The β-globin gene cluster MLPA results revealed heterozygosity for a large deletion consistent with the rare genotype of $(\varepsilon\gamma\delta\beta)^{0}$-thalassaemia (Figure 4.13). The deletion removed most of the β-globin gene cluster including the β-LCR, ε, γ, δ, and β-globin genes.

Adult heterozygotes for the rare condition of εγδβ-thalassaemia have a typical β-thalassaemia trait blood picture but with normal HbA_2 and HbF levels. However in the neonatal period heterozygotes for this disorder are anaemic and can be so severely affected that they require blood transfusions (Rooks *et al*., 2005, Driscoll *et al*., 1989). The condition improves as they get older (3-6 months after birth) and the anaemia becomes less severe. The homozygous state for this disorder has not been encountered, presumably because it is not compatible with fetal survival. It is difficult to understand why this condition is more severe in fetal life, particularly as infants have two out of the four γ-globin

genes intact. The exact mechanisms involved are still not clear but the condition has clinical variability, even within the same family (Trent *et al*., 1990).

In view of the clinical significance of this deletion the referring clinician then referred a further 20 family members. MLPA results revealed that a total of 12 family members had inherited the same $(\varepsilon\gamma\delta\beta)^0$-thalassaemia deletion (Figure 4.12). The haematological parameters for all 12 patients which proved to be heterozygote for the deletion are shown in Table 4.4.

Table 4.4 Haematological parameters for the pedigree with deletion 4.

Case	Age (years)	Hb g/dl	RBC $x10^{12}$/l	MCV fl	MCH pg	HbA_2 %	HbF %
2	55	9.4	4.87	59.0	19.3	3.0	0.4
4	34	13.7	6.75	64.0	20.3	3.0	<1.0
6	32	12.9	6.54	60.0	19.7	3.2	0.7
9	24	11.6	6.06	59.0	19.1	2.8	0.2
12	24	10.3	5.16	62.0	20.2	3.3	0.3
13	21	10.2	5.16	61.0	19.8	3.4	1.0
14	19	11.6	5.75	61.0	20.1	3.1	1.2
15	16	12.5	6.10	62.0	20.5	3.2	0.5
17	10	10.2	5.81	58.0	17.6	3.7	0.5
19	2months	5.5	2.77	71.0	20.0	0	55.2
20	8	11.0	6.00	55.0	18.3	2.9	0.3
21	5	9.7	5.62	68.7	17.3	3.2	0.7

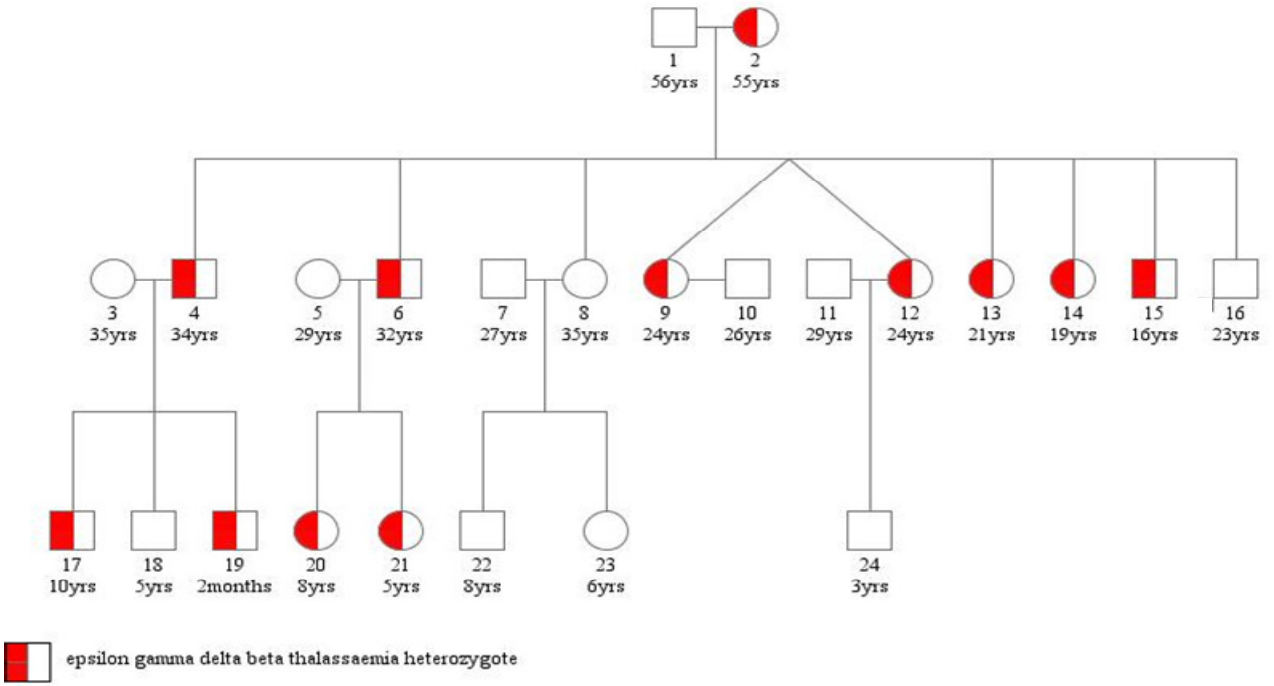


Figure 4.12 Diagram showing the pedigree of the Asian Indian family with $(\varepsilon\gamma\delta\beta)^0$-thalassaemia (deletion).

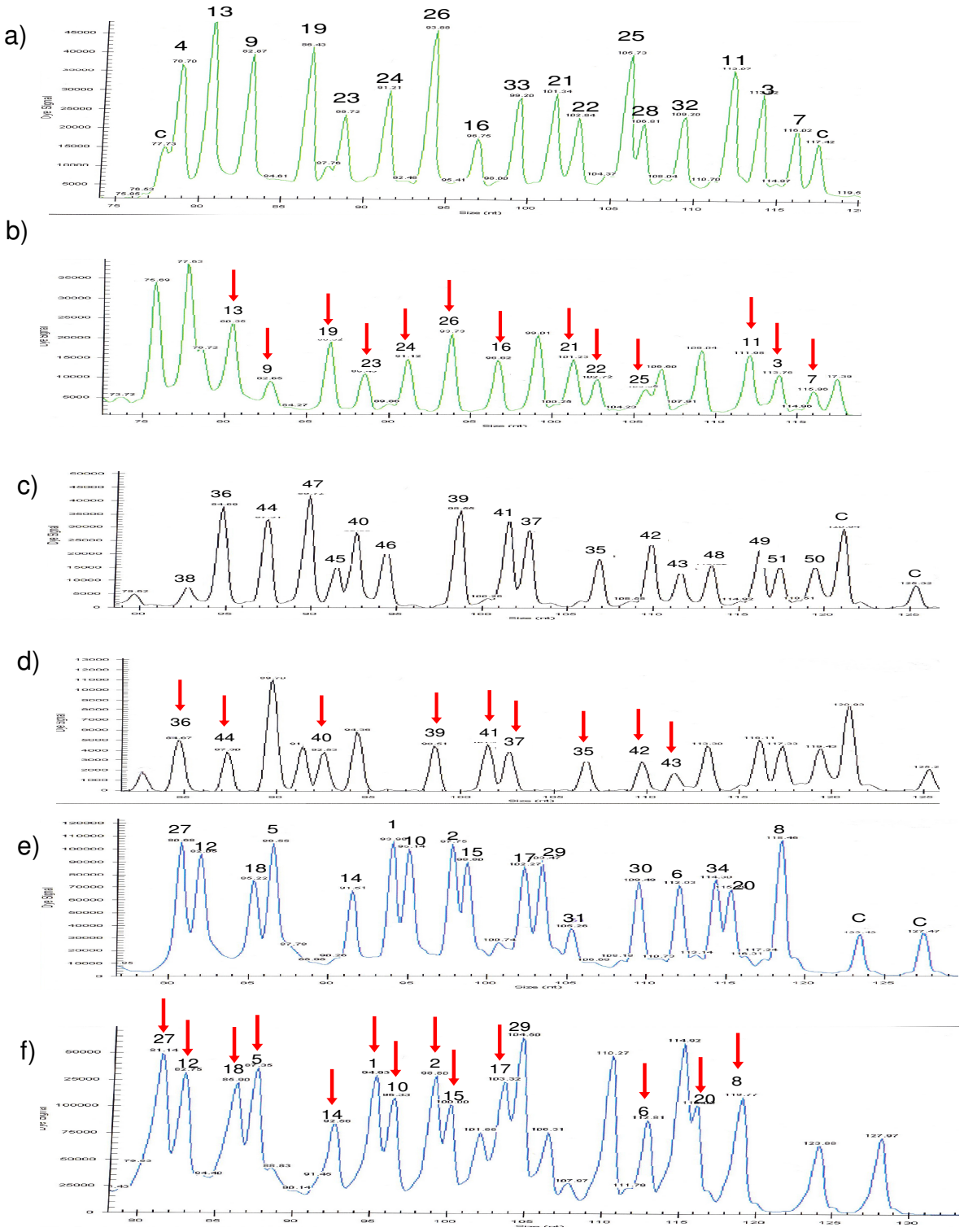


Figure 4.13 ServiceXS MLPA results for deletion 4. Numbered peaks represent probes along the β-globin gene cluster. C= control peaks. a) Normal control analysed with the green probe set b) Patient sample analysed with the green probe set c) Normal control analysed with the red probe set d) Patient sample analysed with the red probe set e) Normal control analysed with the blue probe set f) Patient sample analysed with the blue probe set. Red arrows indicate probes reduced in height compared to the normal control and control peaks within each analysis.

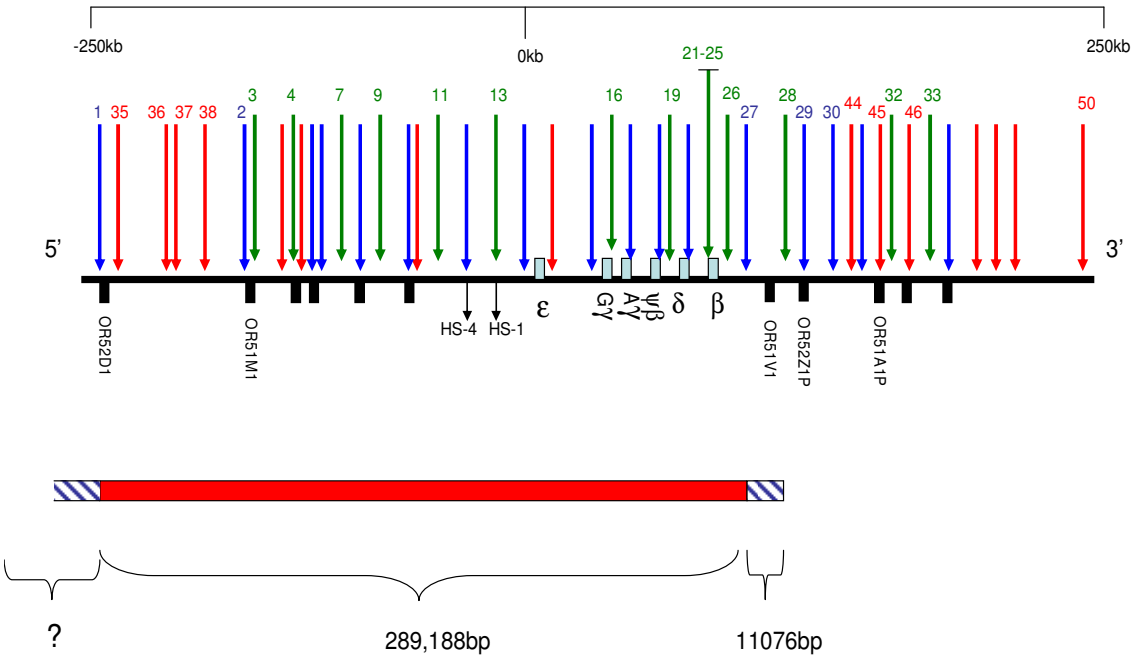


Figure 4.14 Schematic representation of the probes reduced in height in deletion 4 (ServiceXS). The genes are shown as light blue boxes and the arrows denote the probe position with the numbered red probes having reduced height. The SNP array deletion size is illustrated below the genes as a red bar and the hatched blue areas show the breakpoint regions.

Both the ServiceXS and MRC-Holland probe sets showed the same deleted region. All three ServiceXS probe sets (green, red and blue) were analysed with probes 1 to 44 showing reduced height and the entire set of probes in the MRC-Holland kit being reduced (1-29). Probes 28, 29 and 30 which were located 5' to probe 44 were not reduced in height. It was possible that probe 44 was showing reduced height due to a polymorphism in the target sequence and therefore probe 27 was taken as the last probe. From the MLPA results the smallest deletion size was calculated as 289,188bp with the largest size unknown as the last 5' probe was deleted (Figure 4.14). Without the presence of an intact probe at the 5' end of the β-globin gene cluster it was not possible to characterise the 5' breakpoint. Therefore, further characterisation of this

deletion was attempted using microarray technology. DNA from case 2 was analysed using the iScan (Illumina®) microarray with the HumanOmni1-Quad chip. The results of the microarray confirmed the presence of a heterozygous deletion which removed all the globin genes on the cluster (Figure 4.15). The deleted region was covered by 604 probes and the size of the deletion from the results was calculated to be between 502,051bp and 507,634bp.

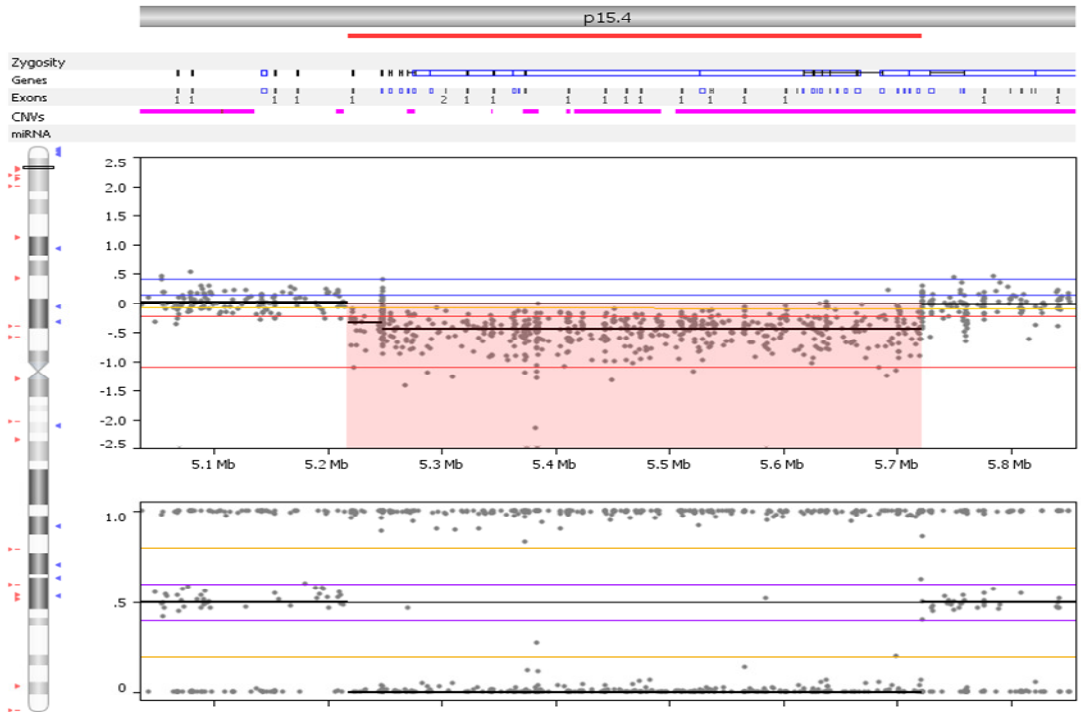


Figure 4.15 SNP microarray (iScan) results for deletion 4. The top panel shows the position of the deletion represented by a red bar with the genes affected below. The middle panel is the Log_2 Ratio track and the lower panel the B allele frequency plot. Both plots show heterozygosity for a large deletion. SNP probes are represented by grey dots.

The microarray results suggested that the actual size of the deletion was twice as large as expected from the MLPA results. The 3' end does remove the olfactory gene *OR51V1* as expected and probe 44 should not have been reduced on the MLPA results. The 5' end removes two more olfactory genes

(*OR52H1* and *OR52B6*) and three TRIM genes (*TRIM6*, *TRIM5* and *TRIM22*) which code for a family of proteins whose function is to mediate interferon.

As stated in chapter 3, at the time of performing this analysis (2008) only 16 different $(\varepsilon^{G}\gamma^{A}\gamma\delta\beta)^{0}$-thalassaemia mutations had been reported in the literature. This deletion was the largest $(\varepsilon\gamma\delta\beta)^{0}$-thalassaemia mutation reported at that time and the first found in the Asian Indian population. Since then however 30 different $(\varepsilon^{G}\gamma^{A}\gamma\delta\beta)^{0}$-thalassaemia deletion mutations have been reported with one deleting a large region of 1.7Mb (Shalev *et al*., 2013). Yet only 13 of the reported 29 different $(\varepsilon^{G}\gamma^{A}\gamma\delta\beta)^{0}$-thalassaemia deletion mutations have been fully characterised due to the difficulties in mapping breakpoints in large deletions. Recently a 506kb $(\varepsilon^{G}\gamma^{A}\gamma\delta\beta)^{0}$-thalassaemia deletion mutation (termed Pakistani I) found in a Pakistani family was fully characterised having taken the researchers over 20 years to define the exact breakpoints (Rooks *et al*., 2012). Eventually the breakpoints were defined using array comparative genome hybridization (aCGH) experiments, a similar approach as was carried out with deletion 4. In the Pakistani I deletion there was no sequence homology between the 5' and 3' breakpoint region. This type of non-homologous recombination has been observed in other deletions of the β-globin gene cluster. However the break point region was found in a palindrome region, a finding which has been associated with other deletion breakpoints and demonstrates that palindromes may play a large role in the formation of deletions. The exact breakpoints of the Pakistani I deletion match the deleted region from the microarray results in deletion 4. It is therefore possible that deletion 4 is in fact the Pakistani I deletion.

4.6 Discussion

Fully characterising deletions is important for comprehensive prenatal genetic diagnosis (as seen in deletion 2) and because it can lead to defining novel regulatory regions. By examining the breakpoints regions and comparing them between different deletions we improve our understanding of the biological processes that created them. The work described in this chapter illustrates the technical challenges in mapping large deletions. It demonstrates that there is a need for a single methodology similar to next generation sequencing techniques that can fully characterise deletions and rearrangements in a single analytical process. Since carrying out this work others have used fine-tiling aCGH technology to detect small and large rearrangements in both the α- and β-globin gene clusters with high resolution (Phylipsen *et al.*, 2012). The recent advance of microarray based technology brings hope that methods of detecting deletions and rearrangements of the β-globin gene cluster will be improved. Further design of primers flanking the microarray deletion breakpoints in deletion 3 and 4 should enable confirmation of their precise size.

CHAPTER FIVE: GAMMA GLOBIN GENE PROMOTER MUTATIONS ASSOCIATED WITH INCREASED HBF LEVELS IN ADULTS

5.1 Introduction

Among the factors known to influence increases in HbF production are a number of point mutations in the promoter region of the γ-globin genes. The work described in this chapter investigated whether γ-globin gene promoter mutations were a significant cause of the increased HbF levels observed in the patient samples referred to our laboratory for haemoglobinopathy investigations. A number of γ-globin gene promoter mutations had been discovered in individuals in other countries (Table 1.4) but this cause of HPFH had not been widely investigated in the UK. There was also little evidence to show whether these mutations were responsible for the high HbF levels observed in patients with a haemoglobinopathy as well as in patients with a non-deletion HPFH phenotype only. Therefore samples from patients with an unexplained raised HbF and a haemoglobinopathy were included in this study along with patients with a high HbF phenotype only (collected from the study in chapter 3). This work aimed to provide data regarding the spectrum of γ-globin gene promoter mutations in the UK and increase our knowledge of these genetic conditions.

5.2 Study subjects

A total of 131 patient samples were selected for this study. Forty one samples had increased HbF levels only (ranging from 1.5-25%) and came from the cohort of 316 patient samples analysed for a β-globin gene cluster deletion in chapter 3. The other 90 samples were referred for investigations into an unexplained elevated HbF level with a co-existing haemoglobinopathy. All samples had been analysed by β-globin gene cluster MLPA and there was no

evidence of a deletion mutation. Patient ages in this group ranged from aged 1-81 years and they were from all the major UK population ethnic groups with 36 being African, 36 white British, 28 Indian, 14 South-East Asian, 7 white European, 7 Mediterranean and 3 from the Middle East. Haemoglobinopathy phenotypes consisted of 55 samples with co-existing α-thalassaemia trait, 6 carriers for sickle cell trait, 28 β-thalassaemia carriers, 1 case with HbE disease and 41 samples with no evidence of any other haemoglobinopathy apart from the elevated HbF level.

5.3 Laboratory procedures

Haematological parameters were ascertained and DNA extraction performed as previously described (sections 2.2 and 2.3). DNA samples underwent PCR amplification and Sanger sequence analysis for both the $^{G}\gamma$- and $^{A}\gamma$-promoter regions using published primers and PCR conditions (Huang *et al.*, 2000).

5.4 Results

A total of 131 selected samples were analysed by Sanger sequencing for a non-deletion HPFH point mutation in the promoter of either of the γ-globin genes. Twenty eight (21%) were found to be heterozygous for a mutation (Table 5.1 and 5.2). Of these, 22 had one of the previously reported mutations with 6 different point mutations being identified, 2 were in the $^{G}\gamma$- and 4 in the $^{A}\gamma$-promoter. In the majority of cases the mutation was found in the same ethnic group and the HbF levels were similar to those which had been previously reported. The remaining 6 samples had unreported mutations. Three different novel mutations were identified; one in the $^{G}\gamma$- promoter and two in the

Table 5.1 Haematological parameters and genotypes in patients with mutations detected in the $^{G}\gamma$ globin gene promoter region. One of the mutations detected is previously unreported (marked with an *).

Case	Gene	Ethnic Origin	Hb (g/dl)	RBC (10^6/mm^3)	MCV (fl)	MCH (pg)	HbA2 (%)	HbF (%)	β-globin phenotype	α-globin genotype	$^{G}\gamma$ globin genotype
1	$^{G}\gamma$	African	16.2	6.52	74.5	24.2	2.4	14.1	HPFH	-α3.7/−α3.7	-202 C>G Heterozygous
2	$^{G}\gamma$	African	12.6	5.7	71.1	22.4	1.9	12.7	HPFH	-α3.7/−α3.7	-202 C>G Heterozygous
3	$^{G}\gamma$	African	13.4	4.7	88.0	28.8	1.8	17.9	HPFH	-α3.7/αα	-202 C>G Heterozygous
4	$^{G}\gamma$	African	13.5	6.2	69.9	22.0	1.6	8.8	HPFH	-α3.7/−α4.2	-200 +C Heterozygous
5	$^{G}\gamma$	African	11.3	5.1	73	22.1	2.3	6.0	HPFH	-α3.7/−α3.7	-194 C>T *Heterozygous
6	$^{G}\gamma$	African	11.8	4.1	94.2	28.6	3.5	6.3	HPFH	αα/αα	-194 C>T *Heterozygous

Table 5.2 Haematological parameters and genotypes in patients with mutations detected in the $^{A}\gamma$ globin gene promoter region. Two of the mutations detected are previously unreported (marked with an *).

Case	Gene	Ethnic Origin	Hb (g/dl)	RBC (10^6/mm3)	MCV (fl)	MCH (pg)	Hb A2 (%)	Hb F (%)	β-globin phenotype	α-globin genotype	$^{A}\gamma$ globin genotype
7	$^{A}\gamma$	White British	12.8	4.2	91.2	30.4	2.5	8.2	HPFH	αα/αα	-198 T>C
8	$^{A}\gamma$	White British	11.4	4.1	95.0	28.0	2.1	6.8	HPFH	αα/αα	-198 T>C
9	$^{A}\gamma$	White British	12.7	4.17	96.0	30.5	2.4	7.2	HPFH	αα/αα	-198 T>C
10	$^{A}\gamma$	White British	12.9	4.8	92.0	26.7	2.3	6.3	HPFH	αα/αα	-198 T>C
11	$^{A}\gamma$	White British	12.4	4.5	95.1	27.4	3.1	6.7	HPFH	αα/αα	-198 T>C
12	$^{A}\gamma$	White British	11.0	3.4	93.8	32.0	2.6	9.5	HPFH	αα/αα	-198 T>C
13	$^{A}\gamma$	White British	14.4	4.8	91.0	30.0	2.4	8.0	HPFH	αα/αα	-198 T>C
14	$^{A}\gamma$	White British	12.6	4.2	83.9	29.6	3.0	6.0	HPFH	αα/αα	-198 T>C
15	$^{A}\gamma$	White British	11.7	3.9	89.6	30.4	2.5	8.5	HPFH	αα/αα	-198 T>C
16	$^{A}\gamma$	White British	11.1	3.8	85.7	29.4	2.4	10.4	HPFH	αα/αα	-198 T>C
17	$^{A}\gamma$	White British	12.8	4.2	95.5	30.3	2.8	11.3	HPFH	αα/αα	-198 T>C
18	$^{A}\gamma$	White British	13.0	4.2	89.7	30.9	2.8	10.1	HPFH	αα/αα	-198 T>C
19	$^{A}\gamma$	Chinese	8.2	5.2	58.0	15.7	2.4	15.1	HPFH	αα/αα	-196 C>T
20	$^{A}\gamma$	Spanish	13.1	4.3	93.1	30.2	2.7	7.9	HPFH	αα/αα	-195 C>G
21	$^{A}\gamma$	White Italian	11.1	5.0	70.0	22.4	4	20.5	β-thalassaemia trait	αα/αα	-158 C>T
22	$^{A}\gamma$	White British	13.0	4.1	90.1	31.4	3	7.4	HPFH	αα/αα	-158 C>T
23	$^{A}\gamma$	White British	12.8	4.4	89.5	29.2	2.8	5.8	HPFH	αα/αα	-158 C>T
24	$^{A}\gamma$	African	9.1	4.7	64.7	19.4	1.8	10.0	HPFH	-α3.7/–α3.7	-158 C>T
25	$^{A}\gamma$	Asian-Indian	16.2	5.4	88.8	29.8	2.6	3.2	HPFH	αα/αα	-130 C>G *
26	$^{A}\gamma$	Asian-Indian	12.2	4.7	78.3	25.9	2.5	9.7	HPFH	αα/αα	-130 C>G *
27	$^{A}\gamma$	Asian-Indian	12.1	4.3	86.0	28.1	2.5	5.6	HPFH	αα/αα	-130 C>G *
28	$^{A}\gamma$	South-East Asian	12.8	4.0	93.7	32.2	2	9.8	HPFH	αα/αα	-117 G>C *

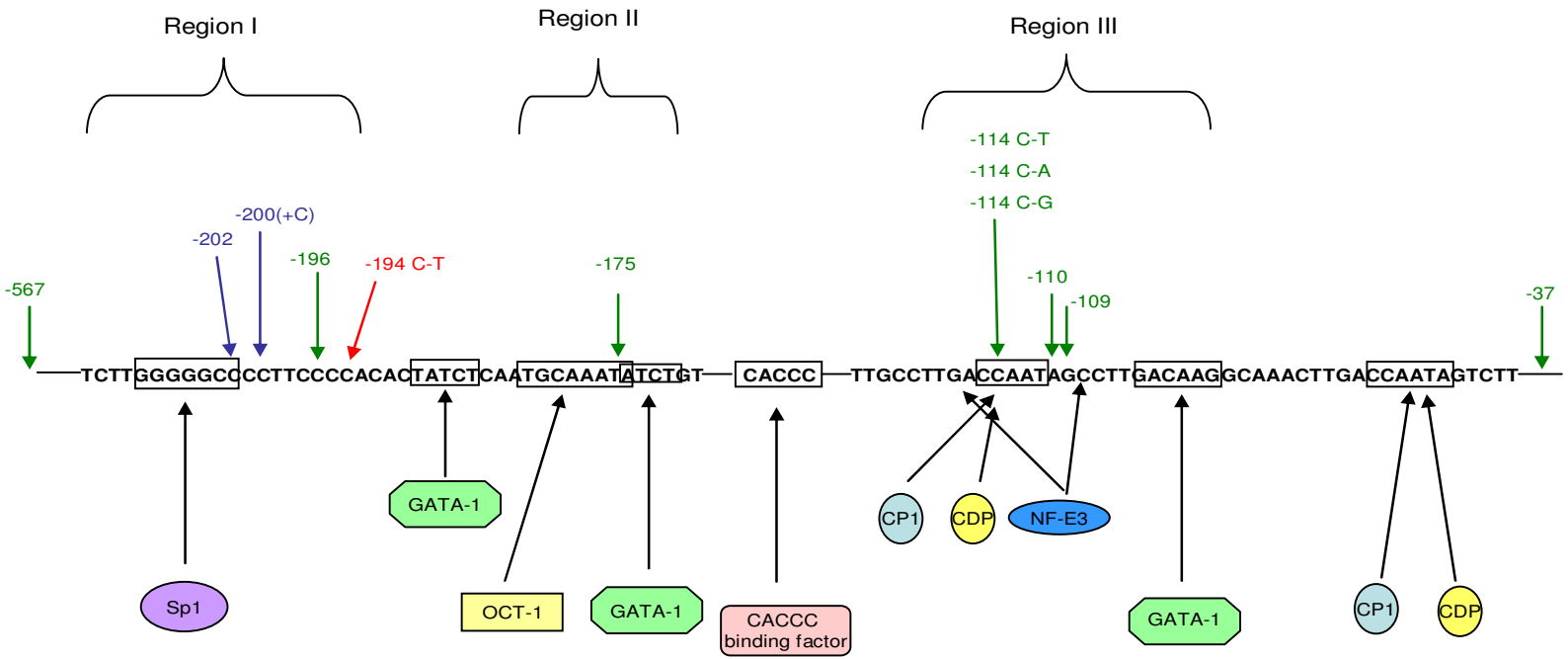


Figure 5.1 Positions of point mutations in the $^{G}\gamma$-globin gene promoters.
Boxed areas show conserved sequences which are binding sites for the transcription factors indicated below the sequence.

Key: ↓ - previously reported mutations not found in this study
↓ - previously reported mutations found in this study
↓ - mutations identified in this study which are novel

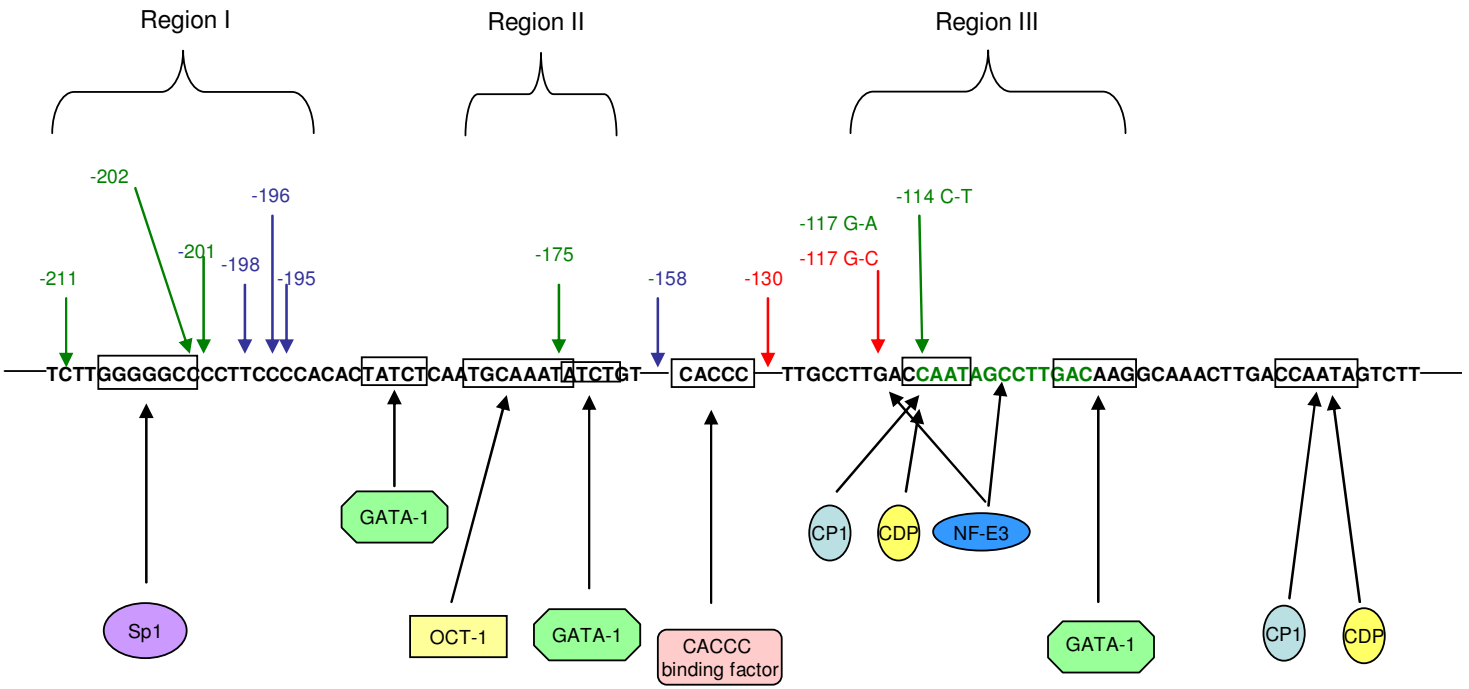


Figure 5.2 Positions of point mutations in the $^{A}\gamma$-globin gene promoters.
Boxed areas show conserved sequences which are binding sites for the transcription factors indicated below the sequence.

Key: ↓ - previously reported mutations not found in this study
↓ - previously reported mutations found in this study
↓ - mutations identified in this study which are novel

$^{A}\gamma$ (Figure 5.1 and 5.2). Twenty one of the 28 patients with a mutation were from the cohort of 41 samples with no evidence of any other haemoglobinopathy apart from an elevated HbF level. The remaining 7 patients with a mutation consisted of 1 patient with β-thalassaemia trait, 5 patients homozygous for α^{+}-thalassaemia and one patient heterozygous for α^{+}-thalassaemia. All six of the patients with a γ-globin gene promoter mutation and α-thalassaemia were of African ancestry. The ethnic group where the most mutations were identified was the cohort of 36 white British individuals. A total of 14 (39%) of these patients were identified as having a γ-promoter mutation.

5.5 Discussion

The range and frequency of non-deletion HPFH point mutations in the UK population was unknown. The results of this study show that mutations in the γ-globin gene promoters are likely to be present at significant frequencies in the UK population. Twenty eight patients were identified with a point mutation which comprised 21% of the total number of patients referred for investigation of a high HbF level in which a large deletion had been excluded. Six of the 25 previously reported mutations were identified. Two of the six mutations were in the $^{G}\gamma$-promoter and 4 in the $^{A}\gamma$-promoter.

The most common mutation identified was the -198 (T-C) $^{A}\gamma$-promoter mutation found in twelve unrelated patients of white British decent with HbF levels ranging from 6-11.3%. This is the only mutation reported in the literature to have been identified in patients of British origin and importantly this mutation has not been reported in any other ethnic groups. In 1986 Tate *et al* identified a

patient who was homozygous for the -198 (T-C) $^{A}\gamma$-promoter mutation and who had an elevated HbF level of 21%. However, since this time there has been no reported data on the frequency of this mutation in the British population. This study identified that 33% of our British cohort with elevated HbF levels have the -198 (T-C) $^{A}\gamma$-promoter mutation indicating that this mutation was likely to be a common cause of non-deletion HPFH in the British population. As this mutation had not been identified in other ethnic groups it was expected that it has arisen independently (therefore not through racial admixture) and was an example of the founder effect. Another interesting finding was that 6 of the 12 patients with this mutation came from the north of England with the other 6 being scattered across the midlands (Figure 5.3). In 1989 Higgs *et al* discovered that several individuals with the exclusively British determinant of α^{0}-thalassaemia (--BRIT) resided in the north west of England and that the mutation had become established through the UK through genetic drift (Higgs *et al*., 1985, Bhavnani *et al*., 1989).It is likely that the -198 (T-C) $^{A}\gamma$-promoter mutation has arisen in the same way.

Four patients were identified as having the previously described -158 (C-T) $^{A}\gamma$ mutation, one patient being white Italian, one African and two were of white British decent. This mutation had previously been reported to be found in Greeks at a range of 2.9-5.1%. Therefore this study identified a second γ-globin gene mutation that was also found in the British population. As this mutation had been identified in the Mediterranean it was possible that the African and two white British patients have some Mediterranean ancestry.

The other 2 previously reported $^{A}\gamma$-mutations (-196 C-T and -195 C-G) were identified in single patients with the HbF levels being in the same range as previously identified and in the same ethnic groups. The two $^{G}\gamma$-mutations identified (-202 C-G and -200+C) were both found in patients of African decent and with HbF levels matching with the reported phenotypes in the literature.

Figure 5.3 Map of the UK showing where the 12 patients with the British -198 $^{G}\gamma$ point mutation originate. Adapted from www.map-of-UK.com

Of the three novel mutations identified, one lies in the $^{G}\gamma$-promoter at position -194 C-T and was identified in two unrelated African patients with modest HbF levels ranging from 6-6.3%. This mutation resides in region 1 (Figure 5.1 and 5.2) of the $^{G}\gamma$-globin gene promoter and mutations close to this mutation such as the -196 (C-T) have been shown to affect the binding of the ubiquitous factor Sp1 (Gumucio *et al*., 1991). One theory on how this mutation increases HbF is that it creates a higher affinity or new binding sites for Sp1 and the stage selector protein (SSP). The higher affinity Sp1 binding site is associated with an increase in promoter activity and one explanation for the increase in HbF is that these new binding sites can overcome the suppression of the γ-globin gene promoter in the adult stage (Ronchi *et al*., 1989). A second theory is that mutations in region 1 alter the DNA conformation of a specific region of the promoter (-206 to -217). This location is thought to be a binding site for a repressor complex but this complex gets displaced by the transcription factors that bind to the new sequence created by the γ-globin gene promoter mutation (Ulrich *et al*., 1992). Therefore it is possible that the novel -194 C-T mutation has the same mechanism in increasing HbF levels as described above for the -196 C-T point mutation.

The second novel mutation was identified in the promoter of the $^{A}\gamma$-globin gene at position -130 (C-G). Three unrelated patients all of Asian Indian ancestry had this novel mutation with modest HbF levels ranging from 3.2-9.7% (Table 5.3). Between positions -160 and -130 lies the CACCC box motif which is highly conserved in all globin gene promoters. This element has been widely studied and shown to be functionally important as it stabilises the

interaction between the γ-globin gene promoter and the β-LCR (Li *et al*., 2001). Normally, in the adult, the CACCC box may become inactivated contributing to the down-regulation of γ-globin gene expression but mutations in this motif can produce an alternative CACCC box that allows the interaction of the γ-promoter with the β-LCR, resulting in the phenotype of HPFH (Li *et al*., 2001). Therefore it is possible that the -130 (C-G) mutation interferes with the CACCC box motif via a similar mechanism to cause a HPFH phenotype or as with other γ-globin gene promoter mutations it alters a binding site for a particular transcription factor.

The third novel mutation was also found in the promoter of the ^{A}γ-globin gene at position -117 (G-C) in a Southeast Asian patient with an HbF level of 9.8%. A γ-globin gene promoter point mutation has been reported extensively in the literature at this -117 position of the ^{A}γ-globin gene but the nucleotide substitution was G-A not G-C. The -117 (G-A) mutation has been reported in Greeks, Italians and Africans before but not in Southeast Asians. It is likely that this nucleotide substitution will have a similar effect as has been shown with the reported -117 (G-A) point mutation as the -114 ^{G}γ-globin gene also has been found to cause the HPFH phenotype with different nucleotide substitutions (Table 5.1). The -117 G-A mutation affects the distal CCAAT box of the ^{A}γ-globin gene and has been the subject of large numbers of studies (Li *et al*., 2001). The mechanism proposed to explain the HPFH phenotype is that the mutation increases transcription by increasing promoter strength. Several binding proteins (e.g. CPI and CDP) bind to the CCAAT box while sites for GATA-1 lie nearby and another erythroid specific protein NFE-3. The -117

mutation increases both CPI and CDP binding together with a decrease in GATA-1 and NFE-3 binding (Mantovani *et al*., 1988, Gumucio *et al*., 1988).

This study also showed that the γ-globin gene promoter mutations are inherited with both α- and β-thalassaemia. Of the 131 patients with high HbF levels, 55 had α-thalassaemia with the majority of these patients being of African ancestry. It was not surprising therefore that the six African patients with a γ-globin gene promoter mutation were also identified as having co-existing α-thalassaemia trait. α^+-thalassaemia trait is common in Africa with one third of the population being a carrier. One β-thalassaemia patient out of the 28 carriers in the cohort was also identified as having a γ-globin gene promoter mutation. Mutations were not seen in patients with sickle haemoglobin. However further studies would have to be carried out to determine the frequency in this group as only 6 sickle cell trait patients were included in this study.

In summary, the results show that mutations in the promoter regions of the γ-globin genes are a significant cause of HPFH in the UK population (21% of cases) where deletional HPFH has been excluded.

CHAPTER SIX: DNA POLYMORPHISMS ASSOCIATED WITH INCREASED HBF LEVELS IN ADULTS

6.1 Introduction

The work described in this chapter investigated the three major polymorphisms identified as being associated with the variation in HbF levels in adults. These are polymorphisms in *XmnI-HBG2*, *HBS1L-MYB* intergenic region on chromosome 6q23 and in the *BCL11A* gene on chromosome 2p16. These three loci have been shown previously to account for 20-50% of the variation in HbF levels in patients (Thein *et al*., 2009). Several SNPs in the *BCL11A* gene and *HBS1L-MYB* intergenic region have previously been identified as being linked to increases in HbF. Two have been shown to have a strong association in healthy Europeans, African sickle cell patients and β-thalassaemia patients; rs11886868 in exon 2 of *BCL11A* gene and rs9399137 in the *HBS1L-MYB* intergenic region (Thein *et al*., 2009).

6.2 Study subjects

One hundred and thirty one patient samples collected from the study described in chapter 5 had increases in HbF levels (ranging from 1.5-25%) but did not have a large deletion mutation. Patient ages in this group ranged from 1-81 years and they were from all the major UK population ethnic groups with 36 being African, 36 white British, 28 Indian, 14 South-East Asian, 7 white European, 7 Mediterranean and 3 from the Middle East. Haemoglobinopathy phenotypes consisted of 55 samples with α-thalassaemia carriers, 6 carriers for sickle cell trait, 28 β-thalassaemia carriers, 1 case with HbE disease and 41 samples with no evidence of any other haemoglobinopathy apart from the elevated HbF level.

6.3 Laboratory procedures

PCR amplification and Sanger sequence analysis of the $^{G}\gamma$-promoter region using published primers (Huang *et al*., 2000) was carried out to identify the SNP rs7482144 also known as the *XmnI-HBG2* polymorphism. PCR products were sequenced using the ABI-PRISM 3100 automated DNA sequencer (Applied Biosystems). Sequencing of this region also detected whether the rare polymorphism G→A at -161 of the $^{G}\gamma$-gene was present in any of the samples which is also known to increase HbF levels. Pyrosequencing assays were designed for both rs11886868 and rs9399137 SNPs and all 131 samples were analysed by this method (section 2.8). Statistical analysis was undertaken using Analyse-it® Microsoft Excel software version 2.20 (Analyse-it Software Ltd, Leeds, UK) and significant associations between the different polymorphisms and HbF was tested using a one-way ANOVA test. Groups showing an overall significance by the ANOVA test were then analysed using the Dunnett's test (Appendix 4).

6.4 Results

All 131 selected samples were analysed by Sanger sequencing and pyrosequencing for the three major SNPs (rs7482144 *XmnI-HBG2*, rs9399137 *HBS1L-MYB*, and rs11886868 *BCL11A*) associated with slight increases in HbF levels in adults. The samples were divided into the following genotype groups: β-thalassaemia trait (29 samples), α-thalassaemia trait (55 samples), sickle cell trait (6 samples) and no significant haemoglobinopathy but increased HbF (41 samples). Statistical analysis of HbF levels was carried out with the above groups for each SNP genotype (wild type, heterozygous and homozygous).

6.4.1 rs7482144 *XmnI-HBG2*

The whisker box plots in figure 6.1 show that the rs7482144 polymorphism is associated with increased HbF levels in both the β- and α-thalassaemia genotype groups with homozygotes having a higher HbF percentage. This polymorphism was not demonstrated in the sickle cell trait group so statistical analysis could not be performed. Unlike the thalassaemia genotype groups, the isolated high HbF group did not see an associated increased HbF with either heterozygosity or homozygosity for the rs7482144 polymorphism. Table 6.1 shows the mean value of HbF for each group with the different SNP genotypes. Statistical differences between the group means was given by the ANOVA *p* value. The β-thalassaemia group showed the most significant association with the rs7482144 polymorphism and HbF levels, with a *p* value of 0.0169. The Dunnett's test results showed borderline significance for both the heterozygous and homozygous groups and given the small sample numbers it is reasonable to expect that more data could yield more significant results. The other groups did not show a significant increase in HbF levels with this polymorphism and when the total HbF levels for each group were analysed for each SNP genotype the *p* value was 0.059.

6.4.2 rs9399137 *HBS1L-MYB*

The whisker box plots in figure 6.2 show that there is an association with higher HbF levels and the rs9399137 polymorphism in the β-thalassaemia group, with homozygotes having the greatest HbF level. Table 6.2 shows that the statistical significance for this difference in HbF levels was not shown with a *p* value of 0.3067. This was probably due to the small sample size in the homozygous group (n=4). The isolated high F group saw a significant difference in the

means (p = 0.0098) but with only heterozygotes having higher HbF levels. The homozygote group had a lower mean HbF level than the wild type group. The Dunnett's test results showed that there was a statistical significance in HbF levels between the heterozygous and wild type group (p = 0.0201). The α-thalassaemia and sickle cell trait groups did not show an association with higher HbF levels and the rs9399137 polymorphism.

6.4.3 rs11886868 *BCL11A*

The whisker box plots in figure 6.3 and the statistical analysis in table 6.3 show that the β-thalassaemia group is the only group to have a significant difference between the three SNP genotypes (p=0.0364), with heterozygotes having higher HbF levels but homozygotes have less HbF than the wild type. The Dunnett's test results show that the difference was between the heterozygous and wild type group, however the difference was not significant (p=0.111). Unlike the other two SNPs (rs7482144 and rs9399137) the C allele or minor allele for the rs11886868 polymorphism seems to be more prevalent than the T allele or major allele. Heterozygotes (TC) and homozygotes (CC) were more frequent through all genotype groups than the wild type allele (TT).

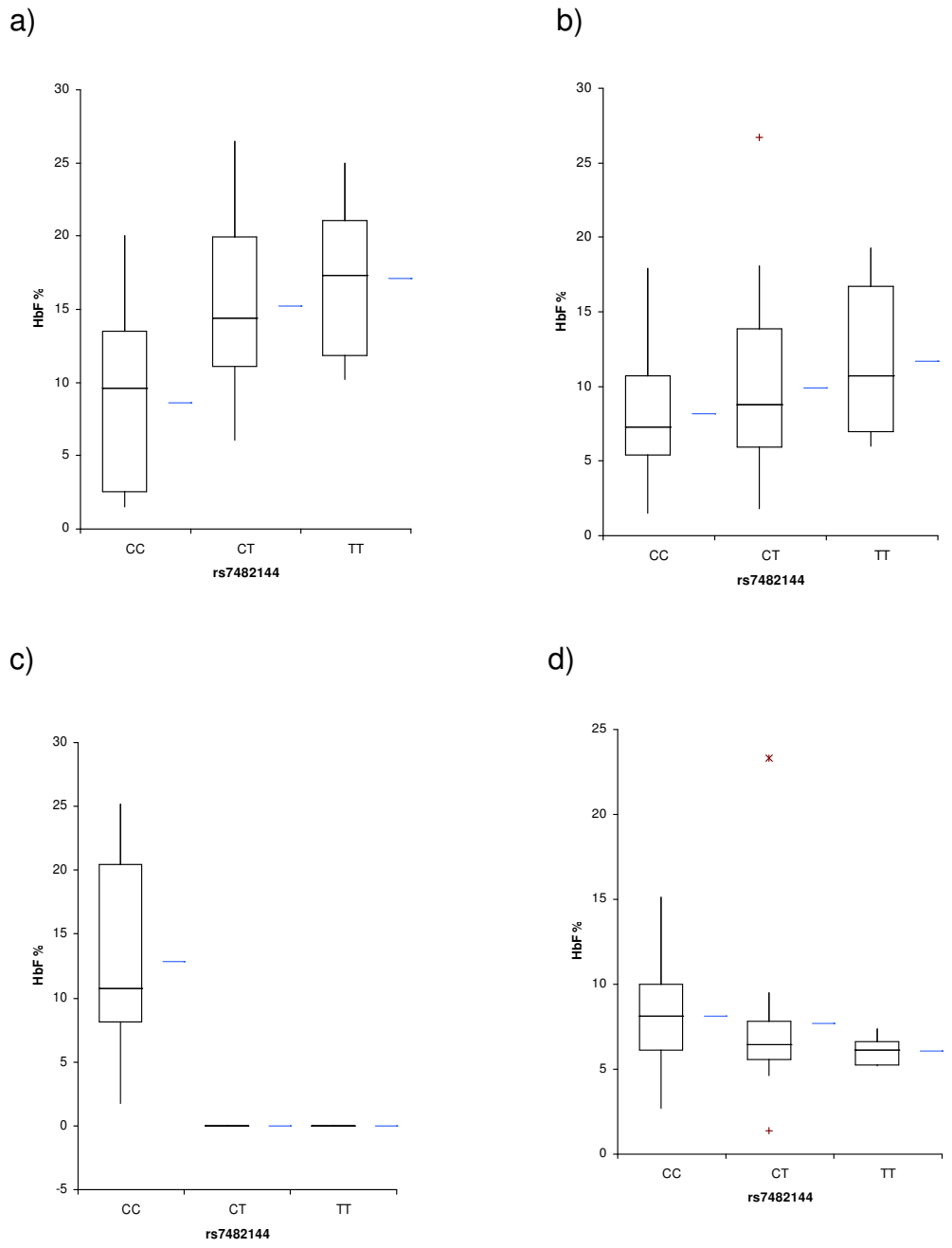


Figure 6.1 Association between rs7482144 genotypes and the proportion of HbF. a) β-thalassaemia trait b) α-thalassaemia trait c) sickle cell trait and d) isolated high HbF subjects. Boxes have lines at the lower quartile, median and upper quartile. The plot whiskers extend up and down from the median distance 1.5 times the interquartile range of the boxes.
Key: — mean, + outliers

Table 6.1 Mean value of HbF % associated with the rs7482144 (*XmnI*) polymorphism for each SNP genotype. Values in red indicate results showing a significant statistical difference.

rs7482144 (*XmnI* polymorphism)					
	Mean value of HbF % for genotype			ANOVA results	Dunnett's results
	CC (n) (wild type)	CT (n) (heterozygous)	TT (n) (homozygous)	*p* value	*p* adjusted (0.05 significance)
β-thalassaemia trait	8.6 (14)	15.3 (9)	17.1 (6)	0.0169	0.0501 – CC vs CT 0.0582 – CC vs TT
α-thalassaemia trait	8.2 (32)	9.9 (19)	11.7 (4)	0.2924	-
Sickle cell trait	12.9 (6)	-	-	-	-
High HbF	8.1 (24)	7.7 (12)	6.0 (5)	0.5346	-
Total	8.6 (76)	10.5 (40)	12.0 (15)	0.0591	-

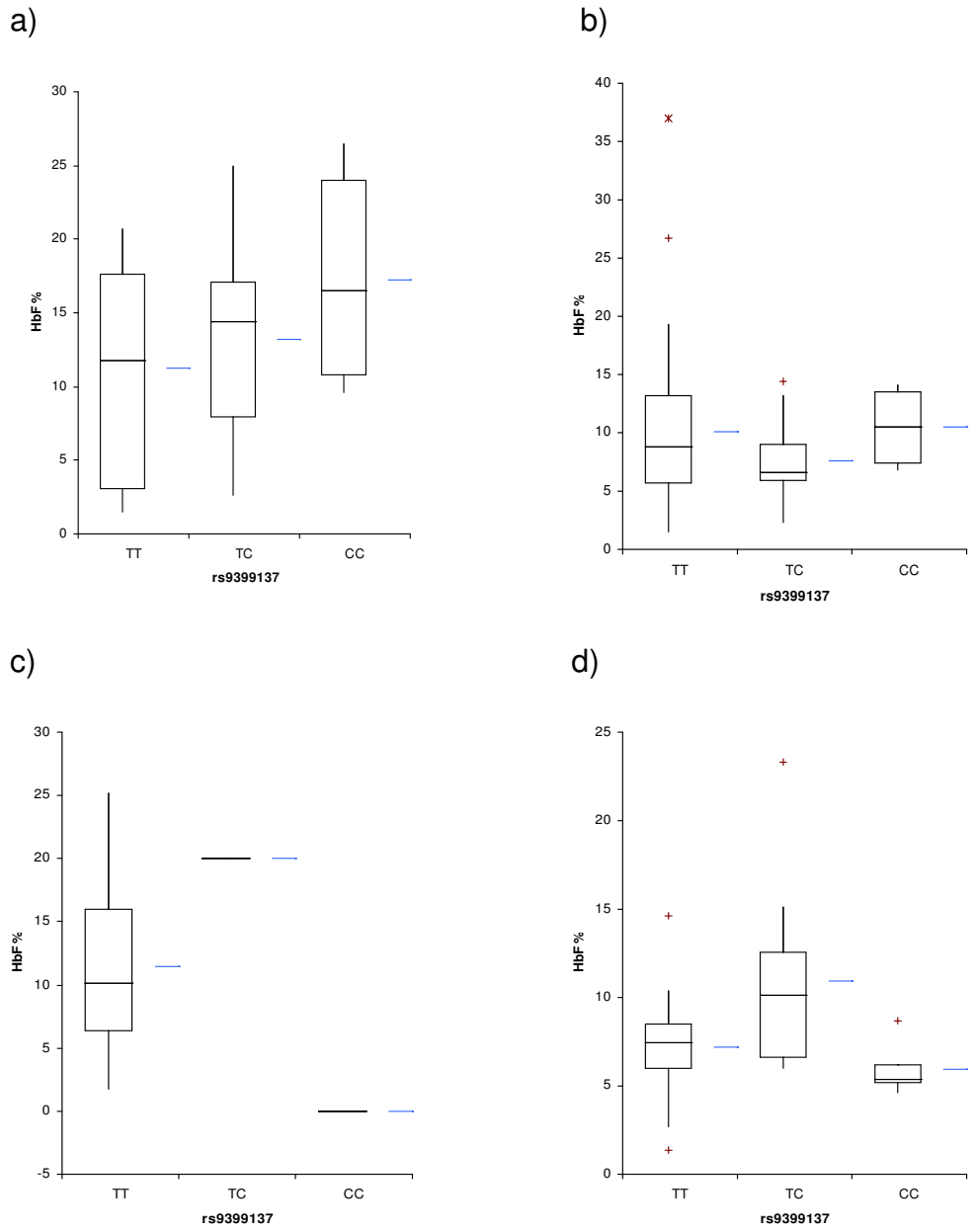


Figure 6.2 Association between rs9399137 genotypes and the proportion of HbF. a) β-thalassaemia trait b) α-thalassaemia trait c) sickle cell trait and d) isolated high HbF subjects. Boxes have lines at the lower quartile, median and upper quartile. The plot whiskers extend up and down from the median distance 1.5 times the interquartile range of the boxes.
Key: — mean, + outliers

Table 6.2 Mean value of HbF % associated with the rs9399137 (intergenic region of *HBS1L-MYB*) polymorphism with each SNP genotype. Values in red indicate results showing a significant statistical difference.

rs9399137 (intergenic region of *HBS1L-MYB*)					
	Mean value of HbF % for genotype			ANOVA	Dunnett's Test
	TT (n) (wild type)	TC (n) (heterozygous)	CC (n) (homozygous)	*p* value	*p* adjusted (0.05 significance)
β-thalassaemia trait	11.2 (16)	13.2 (9)	17.3 (4)	0.3067	-
α-thalassaemia trait	10.1 (41)	7.6 (12)	10.5 (2)	0.4772	-
Sickle cell trait	11.4 (5)	20 (1)	-	0.4120	-
High HbF	7.2 (25)	10.9 (9)	5.9 (7)	0.0098	0.0201 – TT vs TC
Total	9.6 (87)	10.6 (31)	10.1 (13)	0.7371	-

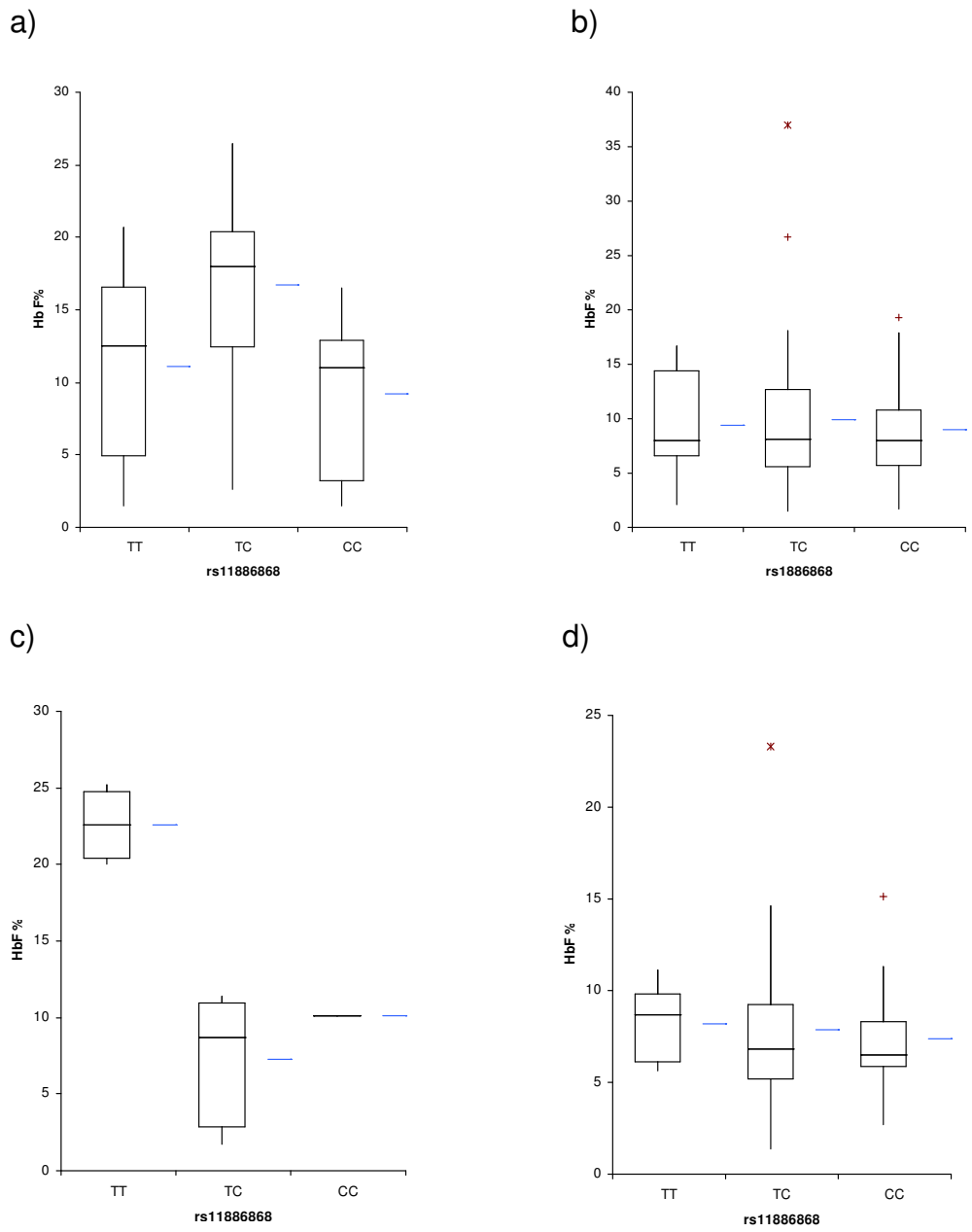


Figure 6.3 Association between rs11886868 genotypes and the proportion of HbF. a) β-thalassaemia trait b) α-thalassaemia trait c) sickle cell trait and d) isolated high HbF subjects. Boxes have lines at the lower quartile, median and upper quartile. The plot whiskers extend up and down from the median distance 1.5 times the interquartile range of the boxes.
Key: — mean, + outliers

Table 6.3 Mean value of HbF % associated with the rs11886868 (on the *BCL11A* gene) polymorphism with each SNP genotype. Values in red indicate results showing a significant statistical difference.

rs11886868 (on the *BCL11A* gene)					
	Mean value of HbF % for genotype			ANOVA	Dunnett's Test
	TT (n) (wild type)	TC (n) (heterozygous)	CC (n) (homozygous)	*p* value	*p* adjusted (0.05 significance)
β-thalassaemia trait	11.1 (9)	16.8 (11)	9.2 (9)	0.0364	0.111 – TT vs TC
α-thalassaemia trait	9.4 (10)	9.9 (26)	9 (19)	0.8974	-
Sickle cell trait	22.6 (2)	7.3 (3)	1 (1)	-	-
High HbF	8.2 (9)	7.9 (15)	7.4 (17)	0.8590	-
Total	10.4 (30)	10.6 (55)	8.5 (46)	0.1900	-

6.5 Discussion

Three major loci (*XmnI-HBG2, HBS1L-MYB* and *BCL11A*) have been identified and shown to effect HbF expression in adults. The results of this study were aimed at identifying whether the three polymorphisms associated with the above loci were found to have increased HbF levels in the 131 study samples.

Of the 131 samples, the rs7482144 (*XmnI-HBG2*) polymorphism was found in 42% with 40 patients being heterozygous and 15 patients homozygous. Homozygotes were seen in all groups apart from the subjects with sickle cell trait where the polymorphism was not present in any of the 6 samples. The percentage of heterozygotes and homozygotes across all genotype groups was comparable. In the β-thalassaemia trait cohort a significant statistical association (Table 6.1) was identified between the heterozygous and homozygous groups compared to the wild type. Previous studies have shown that the rs7482144 (*XmnI-HBG2*) polymorphism plays an important role in ameliorating the clinical severity in sickle cell anaemia and β-thalassaemia major patients by increasing HbF levels (Gilman and Huisman, 1985, Labie *et al*., 1985). Several studies have concluded that this polymorphism only exerts its effects of increasing HbF under conditions of erythropoietic stress and that no conclusive evidence was demonstrated for its influence on HbF levels in heterozygous β-thalassaemia and in normal individuals (Thein *et al*., 1987, Efremov *et al*., 1987). However this study has shown HbF levels to be higher in β-thalassaemia trait individuals with this polymorphism. The results also showed a modest effect in HbF levels in

α-thalassaemia cohort although no significant statistical association could be demonstrated. In order to fully elucidate this finding, larger sample groups would need to be collected and analysed. In the group with only a high HbF phenotype, associations between the rs7482144 polymorphism and increased HbF expression was not demonstrated and homozygotes had lower HbF levels than individuals without the polymorphism (-/-). Results did confirm that this is a common polymorphism found extensively in different population groups. However only modest increases in HbF were seen in the thalassaemia trait cohorts and this polymorphism was unable to explain the variance in HbF levels seen in the sickle cell trait and high HbF only groups. Only 6 samples had sickle cell trait and it might be that none of these patients had the Senegal or Arab-Indian sickle cell haplotype that are known to contain the *XmnI-HBG2* variant.

The rs9399137 polymorphism found in the intergenic region between *HBS1L* and *MYB* have been previously shown to be associated with HbF expression in Chinese, Europeans and Africans (Farrell *et al*., 2011, Menzel *et al*., 2007, Lettre *et al*., 2008, Creary *et al*., 2009). Several other SNPs in this region have also been associated with HbF expression, however the rs9399137 polymorphism was prevalent in a wider spread of populations and the 131 samples varied greatly in ethnicity so this polymorphism became the most likely candidate for this study. Results from figure 6.2 showed that homozygotes were seen in all groups apart from the subjects with sickle cell trait where the polymorphism was only found once in the heterozygous form. The percentage of heterozygotes and homozygotes across all genotype groups was

comparable. In the β-thalassaemia trait group the polymorphism was associated with increased HbF levels with homozygotes having the highest levels although no statistical significance could be shown. In the α-thalassaemia group heterozygotes had lower HbF levels than the wild type. In the high HbF group heterozygotes had higher HbF levels than both the homozygotes and the wild type group. Statistical association was shown between the heterozygous group and samples without the polymorphism but further larger studies with more homozygotes would need to be carried out to confirm whether this finding is significant.

The polymorphism rs11886868 in intron 2 of the *BCL11A* gene has been shown to be in association with HbF expression in Europeans, Africans and Sardinians (Uda *et al*., 2008). Results described in this chapter showed that homozygotes and heterozygotes were found in each genotype group however the frequencies were not as expected. Of the 131 samples analysed, 71% had the rs11886868 polymorphism with 35% being homozygous. Uda *et al* (2008) found the same finding when they directly genotyped 1,412 samples for this SNP. They found in the 134 HPFH samples genotyped, a 2 fold enrichment in the C allele in heterozygotes and a 5 fold enrichment in the C/C genotype when compared to individuals with normal HbF levels. No significant association was seen in any of the groups with this polymorphism and increases in HbF expression. In all groups homozygotes had lower HbF levels than individuals without the polymorphism.

Therefore the strongest association with the three polymorphisms and HbF expression was seen in β-thalassaemia trait subjects with the rs7482144 *XmnI-HBG2* polymorphism. Several studies involving different genotype groups, for example sickle cell anaemia and β-thalassaemia intermedia patients have been undertaken studying the polymorphisms in these three loci. One such study in Mediterraneans with β-thalassaemia intermedia also failed to show statistical correlations with HbF with SNPs in *BCL11A* and *HBS1L-MYB* (Nguyen *et al*., 2010). The results of their study suggested that *BCL11A* and *HBS1L-MYB* have a minor effect on HbF levels compared to the *XmnI-HBG2* polymorphism in β-thalassaemia intermedia patients.

Further studies would need to be carried out with larger sample sizes in order to accurately determine whether *BCL11A* and *HBS1L-MYB* are involved in increased HbF expression in our individuals with an HPFH phenotype. The interest in these loci has led researchers to primarily study disease phenotypes rather than milder conditions such as HPFH. Two large studies, one using meta-analysis data from 2,040 sickle cell anaemia patients and the other using two large sickle cell disease cohorts from America have both shown statistical association with HbF levels and the three major loci in African Americans with sickle cell disease (Lettre *et al*., 2008, Bae *et al*., 2012). It appears from this and other studies that associations between the loci and HbF can be revealed by large GWAS studies but often the exact polymorphisms involved will vary between cohorts. Recent studies now use these genetic modifiers to predict the type of thalassaemia (Galanello *et al*., 2009, Badens *et al*., 2011). Clinicians are now looking at scoring β-thalassaemia major and intermedia dependent on

whether ameliorating alleles such as the *XmnI-HBG2*, *BCL11A* and *HBS1L-MYB* polymorphisms are present in order to act as an early predictor of the type of thalassaemia the patient will develop. When further validated, this prediction tool of severity may have implications for genetic counselling and also therapeutic decisions such as stem cell transplantation (Badens *et al*., 2011).

CHAPTER SEVEN: *KLF1* GENE ASSOCIATED WITH INCREASED HBF LEVELS IN ADULTS

7.1 Introduction

The work described in chapter 6 introduced single nucleotide polymorphisms (SNPs) as being associated with increased HbF expression in adults. The three major loci (*XmnI-HBG2* on chromosome 11p15.5, *HBS1L-MYB* inter-genic region on chromosome 6q23 and *BCL11A* on chromosome 2p16.1) being reported in the literature as accounting for 50% of the variance seen in HbF levels thus indicating that additional loci must be involved (Thein *et al*., 2009).

Recent studies have shown that the erythroid transcription factor *KLF1* could play a critical role in regulating the switch between fetal and adult haemoglobin gene expression. In a Maltese study, 10 out of 27 family members exhibited HPFH due to a single point mutation in the *KLF1* gene. This mutation was found to remove the DNA binding domain which resulted in haploinsufficiency of the *KLF1* gene (Borg *et al*., 2010). Carriers of the mutation (p.K288X) had high HbF levels although variability was shown throughout the family with a range of 3.3%-19.5%. In a similar study from Sardinia, three heterozygous carriers of *KLF1* mutations (p.S270X or p.K332Q) had HbF levels ranging from 0.9%-1.2%. However two family members were compound heterozygous for both of the *KLF1* mutations and had very high HbF levels of 22.1% and 30.9% (Satta *et al*., 2011). The individuals with compound heterozygosity for the *KLF1* mutations also had very high levels of zinc protoporphyrin in their circulation which they suggest confirms the relevance of *KLF1* in the control of the erythropoietic pathway leading to haem biosynthesis (Satta *et al*., 2011). They conclude that their results do not confirm that the HPFH phenotype is caused by

mutations giving rise to *KLF1* haploinsufficiency. Results of only these two studies link the HPFH phenotype to mutations in the *KLF1* gene and therefore further studies need to be carried out with a large cohort of patients with the HPFH phenotype in order to confirm these findings.

The aim of the work described in this chapter is to investigate whether *KLF1* mutations are associated with the elevated HbF levels observed in the 131 samples studied in the work in chapters 5 and 6.

7.2 Study subjects

7.2.1 Elevated HbF level and matched control group

131 blood samples analysed in the work described in chapter 5 and six (subject age 1-81 years) with an elevated HbF level (range 1.5 – 25%) were tested for *KLF1* mutations and compared against a matched control group of 121 samples that had also been referred for haemoglobinopathy investigation but had normal HbF levels (<1%). Haemoglobinopathy phenotypes in the study group with an elevated HbF level consisted of 55 samples with co-existing α-thalassaemia trait, 6 carriers for sickle cell trait, 28 β-thalassaemia carriers, 1 case of HbE disease and 41 samples with no evidence of any other haemoglobinopathy apart from the elevated HbF level. Haemoglobinopathy phenotypes in the control group with normal HbF levels consisted of 41 α-thalassaemia carriers, 9 sickle cell carriers, 29 carriers for a haemoglobin variant, 20 β-thalassaemia carriers and 22 samples with no evidence of any haemoglobinopathy.

7.2.2 Sickle cell disease

A cohort of 55 patient samples with sickle cell disease (all homozygous for the sickle cell mutation) were studied to investigate whether *KLF1* mutations could be involved in HbF level variation in this group. Twenty patients had HbF levels below 10% and 35 had HbF levels >10%.

7.3 Laboratory procedures

Peripheral blood erythrocyte indices were determined using an automated cell counter (Sysmex XE 2100™). Haemoglobin identification and quantifications were carried out using a high performance liquid chromatography system (VARIANT™, Bio-Rad Laboratories, USA) and isoelectric focusing gel electrophoresis (RESOLVE®, PerkinElmer, USA). DNA was extracted from peripheral blood leukocytes by conventional phenol chloroform extraction or on an automated DNA extractor (Chemagen, Baesweiler, Germany). Genomic DNA samples underwent PCR to amplify the human *KLF1* gene (NT_086897.1: 4090501-4093981) using previously published primers (Singleton *et al*., 2008). PCR products were sequenced using the ABI-PRISM 3100 automated DNA sequencer (Applied Biosystems). Samples containing a mutation in the *KLF1* gene had the promoter regions of both the $^{A}\gamma$- and $^{G}\gamma$-genes amplified using previously published primers (Huang *et al*., 2000) and the PCR products sequenced as above. The common 3.7 and 4.2 Kb single α^{+} thalassaemia globin gene deletion mutations were identified by Gap-PCR (Liu *et al*., 2000) and β-globin gene cluster deletions excluded by MLPA in all samples. PolyPhen-2 and SIFT were used to predict the effects of any mutations on protein structure and function (Ramensky *et al*., 2002, Ng and Henikoff, 2003).

7.4 Results

7.4.1 Elevated HbF subjects

KLF1 mutations which are predicted to effect gene function (PolyPhen-2 and SIFT) were identified in 11 out of 131 (8.4%) subjects with increased HbF levels (Table 7.1). Ten had a single heterozygous mutation and one individual was compound heterozygous for two mutations. In total, eleven different *KLF1* mutations were identified (Table 7.1). Deletion mutation in the β-globin gene cluster and mutations in the γ-globin gene promoter sequences were excluded as a cause of the increased HbF level in all 11 subjects. Functionally effective *KLF1* mutations were not identified in the matched cohort of 121 samples with normal HbF levels. Nine of the 11 mutations identified were previously unreported (Figure 7.1).

Eight were mis-sense mutations, one in exon 1 (L51R) and seven in the zinc finger domains (R301C, R301H, W313C, R328H, R328L, T334K and T334R) which would be expected to disrupt DNA binding (Figure 7.1). Two were frame shift mutations in exon 2, an 11bp deletion (K54PfsX9) producing a new stop codon 8 nucleotides downstream and a 7bp insertion (G176AfsX179) producing a stop codon 178 nucleotides downstream. The latter mutation was identified in two patients, singly and compound heterozygous with the L51R mis-sense mutation. The final mutation identified in this group was a 1bp nucleotide substitution (c.913+1G>A) at the 3' end of exon 2 which would be predicted to disrupt splicing.

7.4.2 Sickle cell anaemia subjects

Out of the 55 sickle cell disease patients studied one further unreported functionally effective *KLF1* mutation was identified (c.914-4_914-1 del CTAG) in a sickle cell disease patient with an elevated HbF level of 20.3%. Mutations in the γ-globin gene promoter sequences and the *XmnI-HBG2* polymorphism were excluded as the cause of the increased HbF in this patient. This *KLF1* mutation is a 4bp deletion and is located close to the start of the second zinc finger domain in exon 3 and is likely to result in aberrant splicing.

7.4.3 Tolerated SNPs

Two patients in the HPFH cohort and one patient in the sickle cell disease cohort were found to have SNPs in the coding regions of the *KLF1* gene (G5K, G160K, and G250A). PolyPhen-2 and SIFT analyses suggest that they are neutral substitutions which are tolerated and therefore not pathogenic. In support of this G5K and G160K were also found in 2 samples in the normal HbF level control group.

Table 7.1 Haematological parameters and genotypes in subjects with mutations detected in the *KLF1* gene.
Cases 1-11 are subjects from the elevated HbF cohort and case 12 was from the sickle cell anaemia cohort. Only 2 of the 13 mutations detected had been reported previously (marked with an *). The presence of a wild type allele in the *KLF1* genotype is indicated by [=].

Case	Ethnic group	Hb (g/dl)	RBC (10^6/mm^3)	MCV (fl)	MCH (pg)	HbA2 (%)	HbF (%)	β-globin phenotype/ genotype	α globin genotype	*KLF1* genotype	*KLF1* protein change
1	White British	14.3	6.18	69.4	23.1	3.3	3.0	HPFH	aa/aa	c.[159_169 del GAAGTCTGAGG]+ [=]	K54PfsX9
2	Pakistani	12.7	4.72	79.4	26.9	2.4	14.6	HPFH	aa/aa	c.[901C>T]+ [=]	R301C
3	Black African	10.6	5.43	61	19	2.4	6.6	HPFH	-a3.7/aa	c.[902G>A]+ [=]	R301H
4	Black African	12.1	4.74	81.0	25.5	2.8	7.3	HPFH	-a3.7/aa	c. [939G>T]+ [=]	W313C
5	Thai	12.6	5.51	75.2	22.8	2.4	5.9	HPFH	-a3.7/aa	c.[983G>A]+[=]	R328H*
6	Slovakian	13.0	4.29	91.7	30.3	3.2	8.7	HPFH	aa/aa	c.[983G>T]+[=]	R328L*
7	Black African	10.7	4.24	80.4	25.2	3.3	6.8	HPFH	-a3.7/aa	c. [1001C>A]+ [=]	T334K
8	Black African	8.9	-	73.0	23.0	3.8	6.7	HPFH	-a3.7/aa	c.[913+1G>A]+[=]	-
9	Vietnamese	13.7	5.38	77.6	25.4	2.8	9.5	HPFH	-a3.7/aa	c. [526_527Ins CGGCGCC]+ [152T>G]	G176AfsX179 L51R
10	Korean	13.8	5.34	78	24.5	4.0	1.7	HPFH	-a3.7/aa	c.[526_527Ins CGGCGCC]+[=]	G176AfsX179
11	Thai	11.8	5.80	62.1	20.4	-	11.0	EE	aa/aa	c.[1001C>G]+[=]	T334R
12	Nigerian	12.7	6.0	67.0	21.0	-	20.3	SS	-a3.7/ -a3.7	c.[914-4_914-1 del CTAG]+ [=]	-

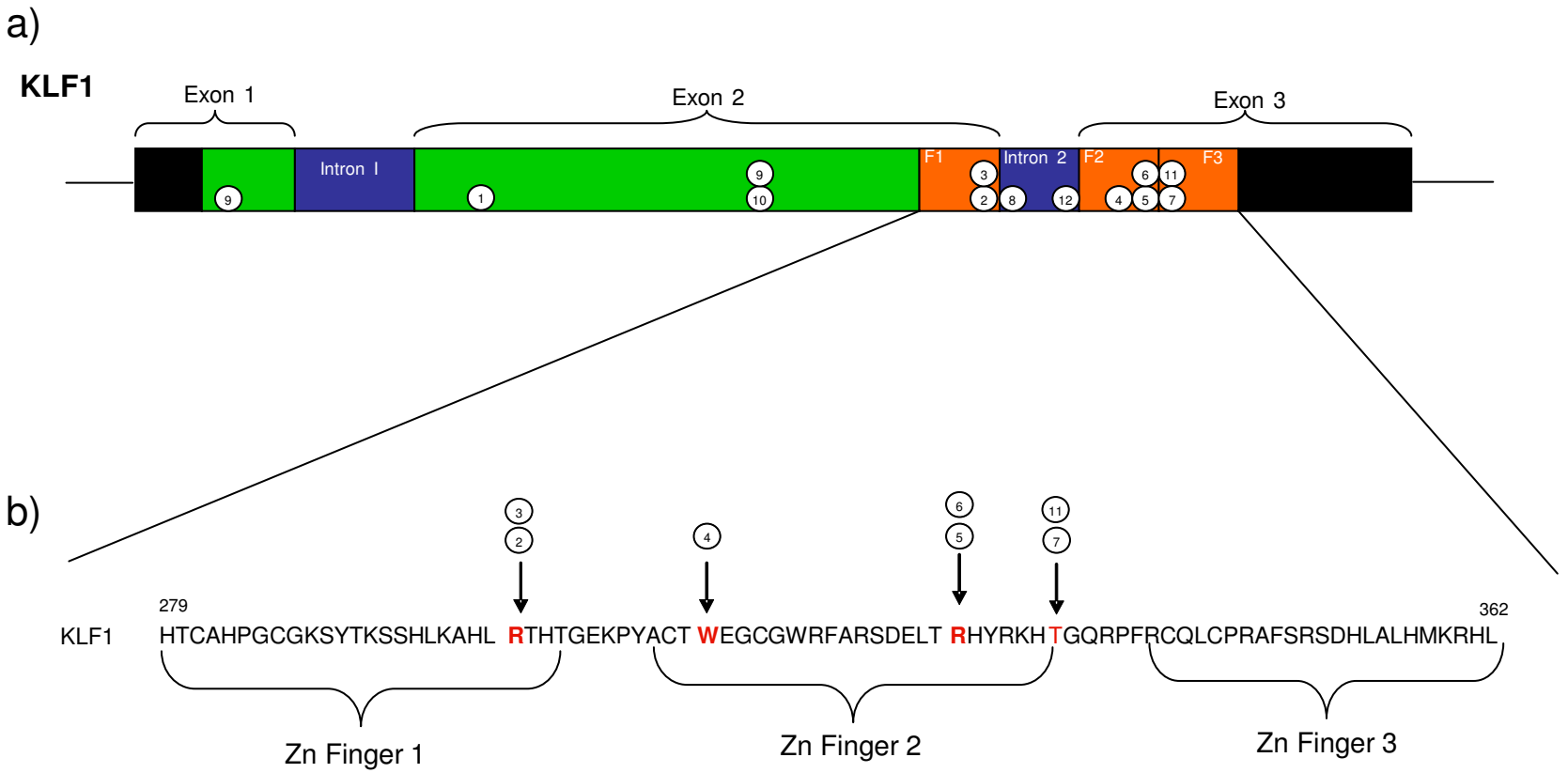


Figure 7.1 Diagram showing the position of the mutations identified in the *KLF1* gene.
a) Schematic representation of the *KLF1* gene with numbered circles corresponding to the case number in table 7.1 (untranslated regions= black, coding regions= green, zinc finger domains= orange and introns= blue). b) Amino acid sequence of the three zinc fingers (Zn) in *KLF1* (NCBI's Homologene) in Homosapiens (Sayers *et al.*, 2011). Arrows show the positions of mis-sense mutations within the zinc finger domains.

7.5 Discussion

Recent reports have identified mutations in the *KLF1* gene which are associated with a variety of phenotypes in humans. These include the Lutheran blood group (Singleton *et al*., 2008), congenital dyserythropoietic anaemia (Arnaud *et al*., 2010), hereditary spherocytosis (Heruth *et al*., 2010), high levels of zinc protoporphyrin (Satta *et al*., 2011), HPFH in two families (Borg *et al*., 2010, Satta *et al*., 2011) and more recently borderline increases in HbA_2 levels (Perseu *et al*., 2011). The UK population is ethnically diverse and our laboratory receives requests for haemoglobinopathy investigations for individuals who originate from all the malarial regions of the world. This study identified *KLF1* mutations in a significant proportion of these referrals with increased HbF levels. *KLF1* mutations predicted to have an effect were found in 11 out of 131 referrals with increased HbF levels but *KLF1* mutations were not identified in 121 haemoglobinopathy referrals with normal HbF levels. This strongly suggests that the *KLF1* mutations are associated with the observed increased HbF levels in these patients. All the mutations identified were heterozygous indicating that a single altered *KLF1* allele can elevate HbF.

The increased HbF levels observed in our *KLF1* mutation positive subjects ranged from 1.7% to 14.4%. Well established factors known to cause HPFH (deletions in the β-globin cluster or mutations in the γ-globin gene promoters) were excluded in all cases. An interesting finding was that 4 out of our 11 cases also had borderline increased HbA_2 levels (3.3 – 4.0%). All of these samples had normal β-globin gene sequences, most likely excluding β-thalassaemia as the cause of the elevated HbA_2 level. This finding concurs with a recent report

that shows that mutations in the *KLF1* gene are associated with elevated δ-globin gene expression which gives rise to borderline increased HbA_2 levels (Perseu *et al*., 2011). Significantly, a proportion of the cohort in that study had increased HbF levels as well as increased HbA_2 levels. The proposed mechanism for *KLF1* mutations increasing HbF levels is reduced activation of *BCL11A* by *KLF1* which in turn results in inefficient repression of γ-globin synthesis. The δ-globin gene has no *KLF1* binding sites therefore the increase in δ-globin gene expression is most likely to be due to indirect effects. Probably impaired looping of the β-LCR with the β-globin gene that results in increased expression of the competing δ-globin gene (Perseu *et al*., 2011). Whether a *KLF1* mutation produces a HPFH phenotype or an increased HbA_2 level (or a combination of both phenotypes) will possibly depend on the balance between these two effects, which in turn will most likely depend on factors specific to a particular *KLF1* mutation and other interacting factors.

The majority of our subjects with a *KLF1* mutation had hypochromic red cells (MCH <27pg), however this could mostly be explained by the co-existing presence of the extremely common 3.7kb single α-globin gene deletion (Table 7.1). Exceptions to this were cases 1, 3 and 5. Cases 3 and 5 had an MCH lower than normally observed with a single α-globin gene deletion. Case 1 was of white British descent and had markedly thalassaemic indices but had tested negative for all types of α- and β-thalassaemia mutations. The *KLF1* mutation (K54PfsX9) identified in this individual would be predicted to be more severe than the other mutations in that it results in the loss of all three zinc finger domains and most of exon 2. It is possible that this could result in severe

impairment of the β-globin gene's association with the β-LCR resulting in a marked reduction in β-globin expression producing a β-thalassaemia type phenotype. Only one mutation was found more than once (G176AfsX179), in case 10 it was associated with a HbF level of 1.7% whilst in case 9 where it was found in combination with the L51R mis-sense mutation the HbF level was 9.5% suggesting that the effects of these mutations may be additive.

A previously unreported *KLF1* mutation (c.914-4_914-1 del CTAG) was identified in one of the 55 patients investigated who were homozygous for the sickle cell mutation, which suggests *KLF1* mutations are not particularly common in sickle disease. However interestingly, the Nigerian sickle cell disease patient identified with this mutation was completely asymptomatic and maintained a haemoglobin level of 12.7g/dl with a HbF level of 20.3%. It is possible the *KLF1* mutation is ameliorating the phenotype by increasing the HbF level via reduced γ-globin gene suppression. However the patient also has homozygous α^+-thalassaemia, α-thalassaemia is known to have a complex interaction with sickle cell disease but does increase the overall haemoglobin level slightly (Ballas, 2001). It is therefore likely that complex mechanisms including multiple gene interactions are involved in the maintenance of this patient's robust haemoglobin level and asymptomatic phenotype.

In summary *KLF1* mutations were found in 8.4% of our elevated HbF cohort, predominantly in individuals of African, Indian and Southeast Asian descent. This indicates *KLF1* mutations could be a widespread cause of HPFH in malarial regions where haemoglobinopathies are common, possibly making a

significant contribution to HbF variance in these populations. Also, the identification of *KLF1* mutations in individuals with a thalassaemia carrier phenotype and a particularly mild form of sickle cell disease indicates the effects of these mutations are likely to be heterogeneous and complex.

7.6 Case 12 – KLF1 mutation in a sickle cell disease patient

A particularly noteworthy case was the occurrence of a previously unreported *KLF1* mutation (c.914-4_914-1 del CTAG) in a 36 year old Nigerian male with a sickle cell disease genotype. Further investigations highlighted that this man had no family history of sickle cell disease and that his relatives were all clinically well and based in Nigeria. He had no clinical history of sickling and had never had a blood transfusion. Analysis of his HPLC and IEF results showed a complete absence of HbA (Figure 7.2). MLPA analysis and β-globin gene sequencing had confirmed the genotype of HbSS and Gap-PCR homozygosity for the 3.7 kb single α^{+}-globin gene deletion mutation. Co-existing α-thalassaemia in sickle cell disease has been shown to modify the phenotype (Steinberg and Sebastiani, 2012) and although this offers some explanation as to why his Hb level was normal it does not explain why he had an asymptomatic phenotype. If the *KLF1* mutation is responsible for the dramatic phenotype amelioration observed in this case, it would be highly significant as it would provide *in-vivo* evidence that controlled reduction of *KLF1* expression could be an effective treatment for sickle cell disease.

a)

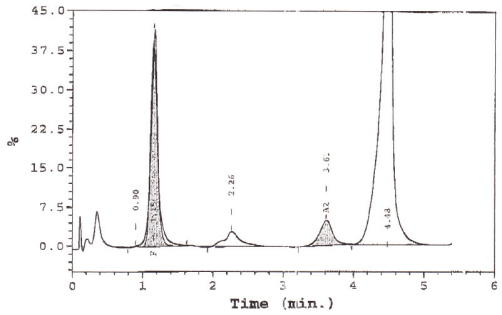
Peak Name	Calibrated Area %	Area %	Retention Time (min)	Peak Area
Unknown	---	0.1	0.90	1534
F	20.3*	---	1.15	291323
Ao	---	3.7	2.26	53027
A2	4.9*	---	3.61	69533
S-window	---	70.8	4.48	1005590

Total Area: 1421007

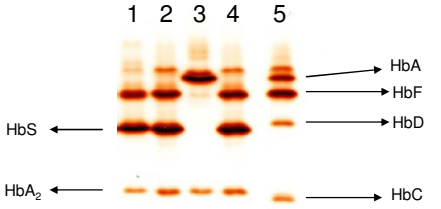
F Concentration = 20.3* %
A2 Concentration = 4.9* %

*Values outside of expected ranges

Analysis comments:



b)



Lane 1 – HbSS control
Lane 2 – case 12
Lane 3 – HbA control
Lane 4 – case 12
Lane 5 – HbAFDC control

Figure 7.2 HPLC and IEF results for case 12. a) HPLC chromatogram showing the major haemoglobin peaks S (70.8%) and F (20.3%). The small peak at the retention time consistent with HbA (3.7%) was a product of degradation and not HbA. This was confirmed by the IEF result below b) IEF gel showing clearly that case 12 (lanes 2 and 4) had no visible HbA and only major bands in the positions of HbS, HbF and HbA_2.

KLF1 is considered an important player in the regulation of haemoglobin switching but many other genes are now also thought to play a part (e.g. *SOX6*, *MYB*, *GATA-1*, *NuRD*, *BCL11A* and *HDAC1/2*) (Sankaran, 2011). Moreover, there are likely to be other as yet unidentified genes involved. Therefore in order to establish whether mutations in other genes are contributing to this patient's

mild phenotype, analysis of his DNA by next generation sequencing techniques that employ whole exome sequencing would be required

The referring clinician was contacted and the gentleman who resides in Aberdeen, Scotland consented to donating more blood and hair follicles. All the above analysis was repeated on DNA extracted from both samples (section 2.3.3) and confirmed the results of the original findings with the presence of the (c.914-4_914-1 del CTAG) *KLF1* mutation. DNA was also analysed on a whole genome SNP array (illumina iScan) by Dr Ruth Clifford. Analysis revealed that part of chromosome 11p showed an area with copy neutral loss of heterozygosity (cnLOH) or uniparental disomy (UPD) (Figure 7.3). UPD occurs when a person receives two copies of a chromosome, or part of a chromosome, from one parent and no copies from the other parent (an error in meiosis). It can be the result of 1) heterodisomy, in which a pair of non-identical chromosomes are inherited from one parent (meiosis I error) essentially benign or 2) isodisomy, in which a single chromosome from one parent is duplicated (meiosis II error). It can lead to the duplication of harmful recessive genes, therefore being potentially dangerous (Figure 7.4). Cases of UPD have been reported to be found in sickle cell disease and β-thalassaemia major (Swensen *et al*., 2010, Harteveld *et al*., 2013).

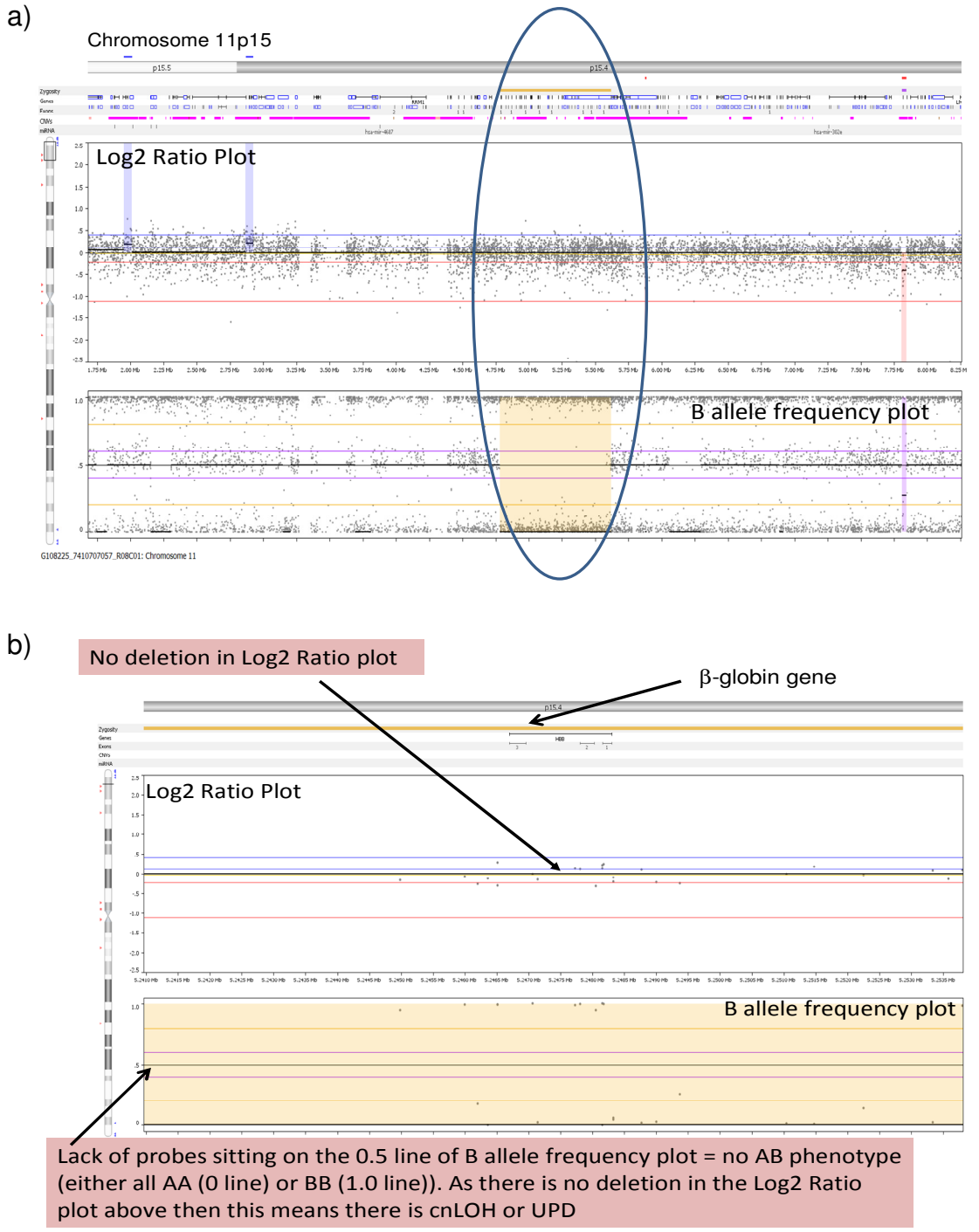


Figure 7.3 SNP array B allele frequency plots for case 12.
a) the blue oval highlights a large area on the plot showing a complete lack of probes sitting on the 0.5 B allele frequency plot b) Expanding the plot from a) above shows that there is no deletion and that the cnLOH region affects the whole of the β-globin gene cluster at a size of approximately 0.84Mb at location p15.4.

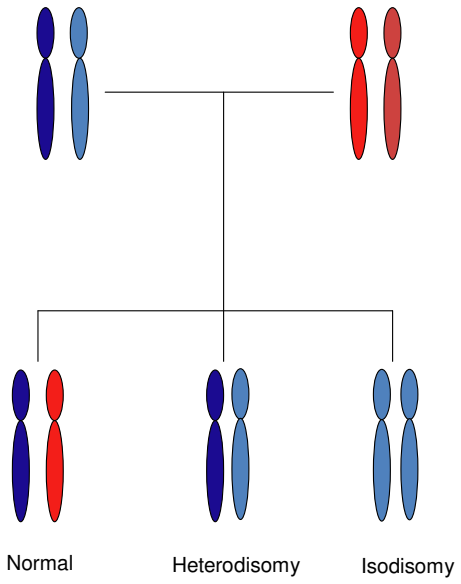


Figure 7.4 Diagram showing the inheritance of uniparental disomy.

These results explain why there was no family history of sickle cell disease as he had inherited two copies of the sickle gene from one parent (isodisomy). However this does not offer any explanation as to why he was asymptomatic.

In order to identify any other mutations in other genes associated with haemoglobin switching, DNA was analysed by Dr Lorna Gregory at The Wellcome Trust Centre for Human Genetics, Oxford by whole exome sequencing using a HiSeq® 2500 sequencing system (Illumina, Inc., San Diego, California, USA). The analysis showed a run yield of 37Gb of mapped data with 85.52% of reads being mapped to exonic regions and 98% coverage of the whole exome. After filtering and annotation using ANNOVAR, 348 predicted variants were left (specifically, insertions or deletions predicted to alter the reading frame, non-synonymous amino acid changes or loss or gain of a stop codon).

Table 7.2 HiSeq 2500 whole exome sequence analysis results for case 12.

No	Function	Gene	Amino Acid Change	dbSNP135	SIFT	SIFT Pred	Poly Phen2	Poly Phen2 Pred	Chr	Start	End	Reference	Observed
1	exonic	CCND3	NM_001136125:c.T559G:p.S187A	rs1051130	0.5	T	0	B	chr6	41903782	41903782	A	C
2	exonic	CDAN1	NM_138477:c.C2735T:p.A912V	rs143086237	0.61	T	0.267	P	chr15	43020919	43020919	G	A
3	exonic	CDAN1	NM_138477:c.C2671T:p.R891C	rs8023524	0.96	D	0.914	P	chr15	43020983	43020983	G	A
4	exonic	CDAN1	NM_138477:c.G1969A:p.G657S	rs61747153	0.25	T	0.003	B	chr15	43023161	43023161	C	T
5	exonic	CDAN1	NM_138477:c.A1787G:p.Q596R	rs12917189	0.55	T	0	B	chr15	43023482	43023482	T	C
6	exonic	HBB	NM_000518:c.A20T:p.E7V	rs77121243	0.97	D	0.001	P	chr11	5248232	5248232	T	A
7	exonic	FBXO7	NM_001033024:c.T35G:p.L12R	rs8137714	0	T	0	B	chr22	32871383	32871383	T	G
8	splicing	KLF1	-	-	-	-	-	-	chr19	12995875	12995878	CTAG	-
9	exonic	NT5C3B	NM_052935:c.C638G:p.S213C	rs1046404	0.98	D	0.725	P	chr17	39983808	39983808	G	C
10	exonic	NT5C3B	NM_052935:c.C626T:p.A209V	rs1046403	0	T	0	B	chr17	39983820	39983820	G	A
11	exonic	RPS24	NM_001142282:c.C391A:p.Q131K	rs7899453					chr10	79799959	79799959	C	A
12	exonic	TMPRSS6	NM_153609:c.T2207C:p.V736A	rs855791	0.97	D	0	B	chr22	37462936	37462936	A	G
13	exonic	TMPRSS6	NM_153609:c.A757G:p.K253E	rs2235324	0	T	0	B	chr22	37485724	37485724	T	C

Visual inspection of the 348 variant sequence calls revealed 13 to be valid, the others being invalid due to low coverage or resulting from misalignment. Table 7.2 shows the 13 possible sequence variants. Further filtering removed 11 sequence variants as these proved to all be tolerated benign SNPs, leaving 2 possible pathogenic variants. Of these 2 variants, 1 confirmed the homozygosity for the sickle mutation (number 6) and 1 heterozygosity for the *KLF1* splice site mutation (number 8).

In summary, as no further mutations have been found to explain this gentleman's asymptomatic phenotype, further family studies are now required to try and ascertain the phenotype of the *KLF1* mutation on its own and with the sickle cell gene. This patient has contacted his relations in Nigeria as he has 4 siblings and both of his parents are still alive and well. They have agreed to go to a local clinic in Nigeria and donate blood. As well as family studies, functional analysis and gene expression of *KLF1* with and without the mutation would confirm whether the *KLF1* splice site mutation in case 12 is responsible for the increased HbF level seen.

CHAPTER EIGHT: HAEMOGLOBIN SWITCHING OCCURS EARLIER IN ASIAN-INDIAN NEWBORNS THAN NEWBORNS FROM OTHER ETHNIC GROUPS

8.1 Introduction

Newborns in the UK are screened for sickle cell disease and β-thalassaemia major at 5-8 days of age as part of the national dried blood spot screening programme (Streetly *et al*., 2009). In the Thames Valley region of the UK we currently screen approximately 30,000 newborn blood spots a year. Screening is carried out using a combination of high-performance liquid chromatography (HPLC) and isoelectric focusing (IEF). The predominant haemoglobin at birth is HbF with HbA levels in normal term infants being in the range of 5-40%. Whilst performing newborn screening for sickle cell and thalassaemia in the Thames Valley region I observed that the ratio of HbA to HbF appeared to show some variation with ethnic origin. Asian Indian (Indian, Pakistani and Bangladeshi) newborns had higher HbA levels than newborns from other ethnic groups. β- and α-thalassaemia are prevalent within this ethnic group, however, neither explains this finding. The reason for the increased HbA level observed in Asian Indian newborns was unclear.

8.2 Differences in haemoglobin switching in newborns

HbA levels lower than normal (slowing of the switching process) have been shown in many pathological conditions. These include premature newborns, maternal anoxia, placental insufficiency, intrauterine growth retardation, some chromosomal abnormalities, hyperinsulinaemia secondary to maternal diabetes mellitus and sudden infant death syndrome (Galacteros *et al*., 1991). Heterozygosity for β-thalassaemia also decreases HbA concentrations at birth

(Galanello *et al*., 1981). A previous small study in 1987 had shown that newborns from the Far East (Japan and China) had slightly lower HbA levels at birth than infants from other ethnic groups (Kutlar *et al*., 1987). However this study did not include Asian Indian newborns.

Higher than normal HbA levels (accelerated Hb switching) can be an indication of a possible red blood cell transfusion, gestations >40 weeks or with larger than average birth weights (Galacteros *et al*., 1991, Smith and Cauchi, 1982). It has also been reported to occur in some chromosomal abnormalities, γ-thalassaemia traits and in the $^{A}\gamma$-$^{A}\gamma$-globin gene rearrangement (Wilson *et al*., 1968, Weller *et al*., 1966, Huisman *et al*., 1983). In the $^{A}\gamma$-$^{A}\gamma$-globin gene rearrangement instead of having $^{G}\gamma$ -$^{A}\gamma$-globin genes there is a nucleotide substitution at codon 136 and both $^{A}\gamma$-globin genes are inherited.

8.3 Accurate HPLC quantitation of haemoglobins in newborns

Newborn screening HPLC analysers for haemoglobinopathy identification are able to accurately identify the presence of clinically significant haemoglobin variants important in newborn screening such as sickle haemoglobin. The major objective of all newborn haemoglobinopathy screening programmes is Hb identification rather than accurate quantitation. The HPLC used in this study was a Bio-Rad Vnbs which adopts a set of integration rules to quantitate the percentage of haemoglobin present in a sample. These integration rules were derived from those generated by the Genetic Disease Laboratory, California, USA after they analysed 2.5 million newborns by HPLC and are known to give slight variations in the quantitation of major haemoglobins within a given sample

run on separate occasions (Eastman *et al*., 1996). Therefore in order to determine whether Asian Indians have increased HbA% when compared with other ethnic groups, accurate quantitation of HbA would be required. Discussions with the HPLC manufacturer, Bio-Rad about this problem led to the trial of some new improved integration software in development for the Vnbs analyser, termed Valley-to-Valley integration. This method provided consistent peak integration for each analysis. The new Valley-to-Valley software was run in parallel with the old integration method for all newborn blood spots analysed during the work described in this chapter.

8.4 Study subjects and laboratory procedures

The ratio of HbA to HbF was studied in 2,286 term newborns (38-40 weeks gestation) from a variety of ethnic groups; all had a normal haemoglobin pattern (i.e. absence of any variant haemoglobin) and none of the newborns had received a blood transfusion. The blood was collected from the heel of the baby between 5-8 days after birth and placed on a dried blood spot screening card by the midwife. Of the 2,286 samples, 1686 were of Caucasian origin, 496 were Asian Indian and 104 were of black African origin. Along with the ethnic groups the birth weight, gestation and HPLC results were collected on all samples. The blood spots for each newborn were run twice and in parallel with both software packages in order to ascertain which integration method gave the most reliable result. Results were statistically analysed by means of an independent *t*-test using the software package Analyse -It®.

8.5 HbA levels in Caucasian, Asian Indian and black African newborns born at 38-40 weeks gestation

The results showed that the Valley-to-Valley software integration gave HbA% approximately 0.2% higher for all samples than the existing software. The results were reproducible and so the Valley-to-Valley software results were used to test the hypothesis. Newborns born at gestation between 38-40 weeks were selected for analysis so as to eliminate premature newborns in the Caucasian or black African groups being responsible for the differences in HbA levels observed. The results revealed an average level of HbA in Asian Indian newborns of 20.0%, which was significantly higher (p=<0.001) than the average level of 17.4% observed in Caucasian babies, and 18.6% observed in babies of African origin (Table 8.1). The standard deviation of the HbA levels in Asian Indian newborns (SD 7.2) was significantly greater than in Caucasian newborns (SD 6.0). This finding indicated that the increased average HbA level was due to a sub-group of infants with higher HbA levels (Fig 8.1).

Table 8.1 HbA % in newborns from Caucasian, Asian Indian and black African origins

Ethnicity	n	Average HbA%*	Significance	*p* value
Caucasian	1686	17.4 ± 6.01 (4.2 - 42.5)	-	-
Asian Indian	496	20. 0 ± 7.16 (5.5 - 43.3)	Indian/Caucasian	<0.001
Black African	104	18.6 ± 7.05 (5.9 - 37.9)	Black/Caucasian	<0.05

*Average Hb A % ± SD and range between parentheses

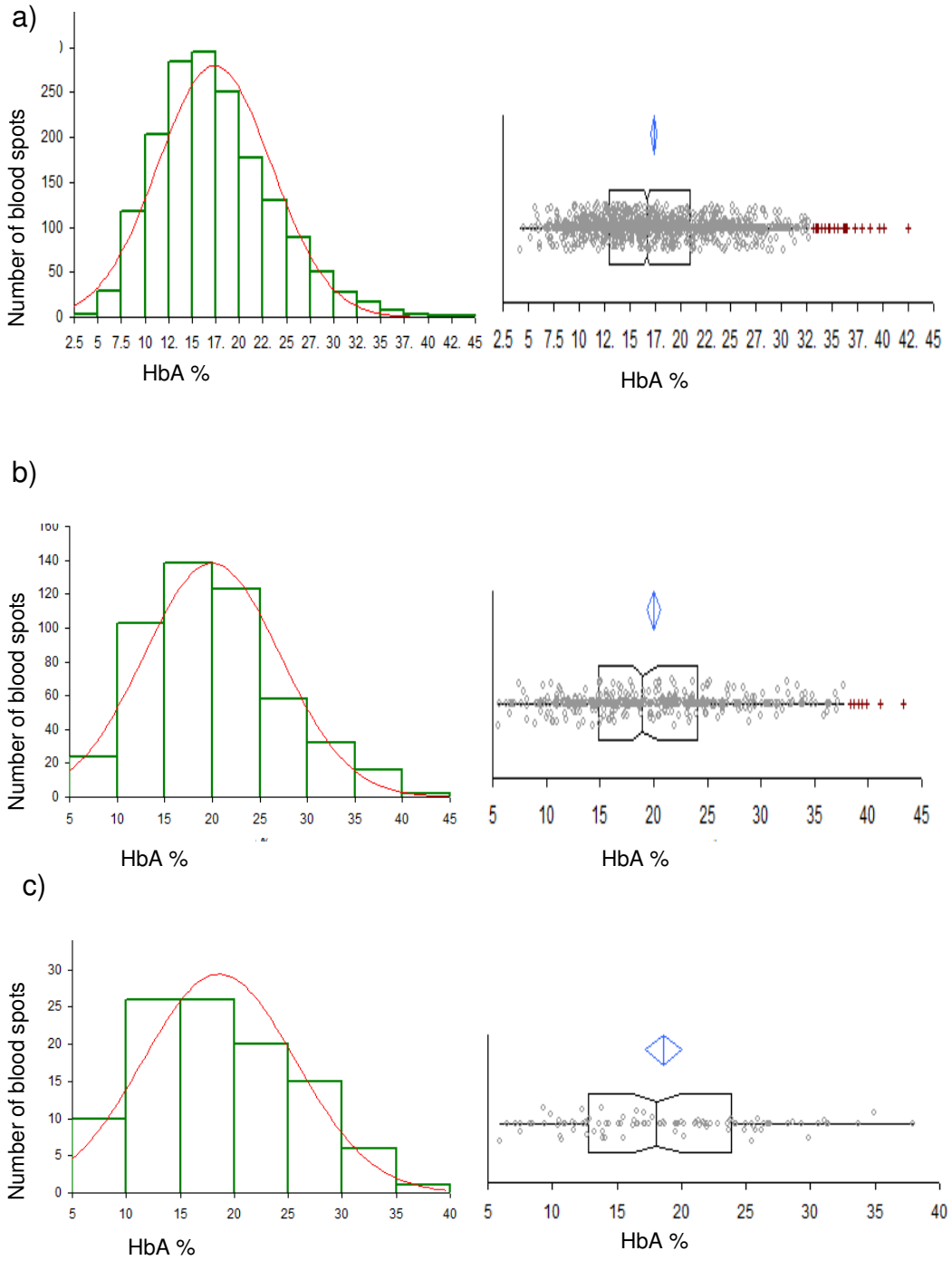


Figure 8.1 Histogram and boxplot showing HbA levels in each ethnic group.
a) Caucasian b) Asian Indian c) black African

Key: Normal fit distribution + Outliers

95% CI Outlier boxplot 95% CI mean

8.6 Birth weight

Differences in birth weight could be responsible for the increased HbA% seen in the Asian Indian newborns. The birth weights of newborns born at gestation 38-40 weeks were analysed. Results showed a statistical significance (*p*=<0.001) between the birth weights in the Asian Indian group compared to the Caucasian group (Table 8.2). However the difference shown was that the Asian Indian newborns had the lowest birth weights (mean 3.1kg) with the Caucasian newborns being the heaviest (3.4 kg) (Table 8.2 and Figure 8.2). A recent large study in the UK confirmed the same findings and found that newborns born from Asian Indian parents had considerably smaller birth weight babies than those born from European parents (Wells *et al*., 2013).

Table 8.2 Average birth weight at term (38-40 wks gestation) from Caucasian, Asian Indian and black African newborns

Ethnicity	n	Average Birth weight (kg)*	Significance	*p* value
Caucasian	1686	3.4 ± 0.45 (1.8 – 5.0)	-	-
Asian Indian	496	3.1 ± 0.47 (1.4 – 5.0)	Indian/Caucasian	<0.001
Black African	104	3.3 ± 0.45 (2.5-4.2)	Black/Caucasian	0.309 (not sig)

* Average birth weight (kg) ± SD and range between parentheses

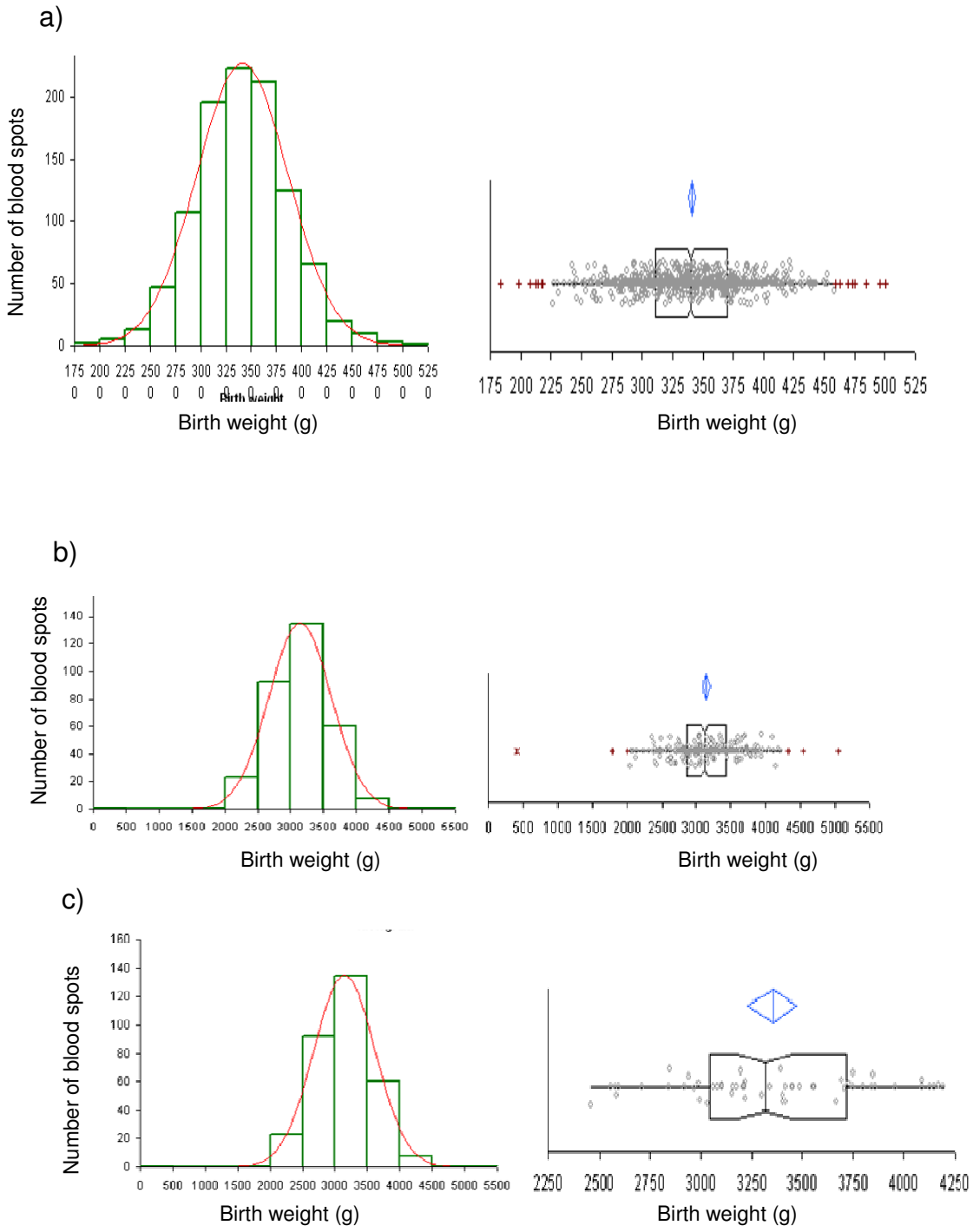


Figure 8.2 Histogram and boxplot showing the birth weight (g) at 38-40 weeks gestation in each ethnic group.
a) Caucasian b) Asian Indian c) black African

Key: Normal fit distribution + Outliers

95% CI Outlier boxplot 95% CI mean

8.7 γ-thalassaemia

Increased birth weight and gestation have been ruled out as the reason for the increased HbA levels observed in Asian Indian newborns but could this finding be explained by a cohort of Indian infants who have increased HbA levels at birth because they are carriers for γ-thalassaemia? Cases of γ-thalassaemia have been described in most populations (Chinese, Indian, European, Japanese and African). Huisman *et al* (1983) showed it occurred at very low frequencies in India; however they only looked at 208 newborns from Bombay. A further small study looked at 200 Indians from the Gond tribe of central India and failed to observe γ-thalassaemia among the tribe although other haemoglobinopathies such as sickle haemoglobin and α-thalassaemia were evident (Gupta *et al*., 1991).

Only one type of γ-thalassaemia has been described (Figure 8.3) and it arises from a deletion of approximately 5kb which is the result of an unequal crossover between the $^{G}\gamma$- and $^{A}\gamma$-globin genes forming a $^{GA}\gamma$-hybrid gene (Sukumaran *et al*., 1983). Deletion of a γ-gene removes competition for the upstream β-LCR leading to increased interaction of the β-LCR with the β-gene in *cis*, thus enhancing its expression. This hybrid produces a protein product identical to the $^{A}\gamma$-gene but at a level normally seen for the $^{G}\gamma$-gene which explains why this deletion is not symptomatic even in the homozygous condition. Cases of γ-thalassaemia are also rarely detected because they would only be observed for a short period of time after birth when the unusually high HbA% is evident.

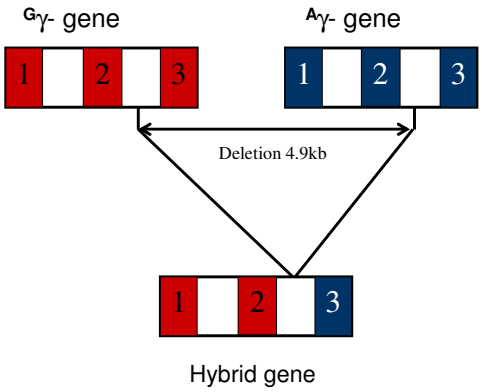


Figure 8.3 Schematic representation of deletional γ-thalassaemia.
Sequencing the junction fragment reveals the presence of a hybrid gene identical to $^{G}\gamma$-gene before the polymorphic TG repeat in IVS2 and identical to the $^{A}\gamma$-gene after this repeat. The size of the deletion corresponds exactly to the size of DNA which separates the γ-genes.

8.7.1 Identification of γ-thalassaemia by MLPA

As already discussed in chapter 3, the MLPA technique detects large deletions and rearrangements in the β-globin gene cluster. There are three probes spanning the γ-globin genes in the ServiceXS β-globin gene cluster probe set, with one probe (probe 16) targeting the intergenic region between the $^{G}\gamma$- and $^{A}\gamma$-gene. Therefore it was possible to detect heterozygosity and homozygosity for γ-thalassaemia using the ServiceXS β-globin gene cluster MLPA probe set. MLPA is best suited to good quality DNA but only small amounts of dried blood spot samples were left with the majority being over a month old. DNA was extracted from 10 newborn dried blood spots analysed from the 496 Asian Indian cohort who had HbA% ranging from 35-45%. Results for the MLPA β-globin gene cluster probe set (β-globin^{XS} MLPA Kit; ServiceXS) showed that 6 out of the 10 samples had normal peak heights for all probes and therefore no evidence of the deletional γ-thalassaemia already described. The remaining 4

samples failed to achieve good quality MLPA results and had low signal strength. This was most likely due to the age of the dried blood spot and the small sample size.

8.7.2 Case of γ-thalassaemia in an Asian Indian family

Whilst undertaking this study a family were referred to the laboratory from a hospital in Leeds, UK for investigations into a newborn Pakistani presenting at birth (40 weeks gestation) with a HbA level of 70% (Table 8.3). No other abnormalities were evident and the proband was clinically well. The pregnancy and birth were uncomplicated and the proband had two normal siblings. The proband had not received a red blood cell transfusion and had a healthy birth weight (3.21kg). The mother and father had no evidence of a haemoglobinopathy on antenatal screening. EDTA whole blood samples were received from the mother, father and proband. Haematological parameters were ascertained and DNA extraction performed as previously described (section 2.2 and 2.3).

Table 8.3 Haematological data in family with γ-thalassaemia.

	Hb g/dl	RBC 10^6/mm^3	MCV fl	MCH pg	HbA %	HbF %	HbA_2 %
Mother	12	4.5	93	27	87	0.2	2.8
Father	15	5.4	93	28	86	0.2	2.6
Proband	10	3.7	92	27	71.2	15	1.2

Sequence analysis of the α- and β-globin genes was performed on all three samples and revealed no mutations. The very high HbA% in the proband with no apparent haemoglobinopathy in the parents could be explained by the inheritance of γ-thalassaemia. This is because γ-thalassaemia would only be

detected in adults by DNA analysis as the haematological parameters would be normal. MLPA analysis using the ServiceXS β-globin gene cluster probe set was carried out on the family and the results showed a heterozygous deletion in probe 16 in both the proband and the father (Figure 8.4). The mother showed no evidence of any deletions or rearrangements in the β-globin gene cluster. Probe 16 sits between the $^{G}\gamma$- and $^{A}\gamma$-gene and could indicate inheritance of the only deletional γ-thalassaemia described to date.

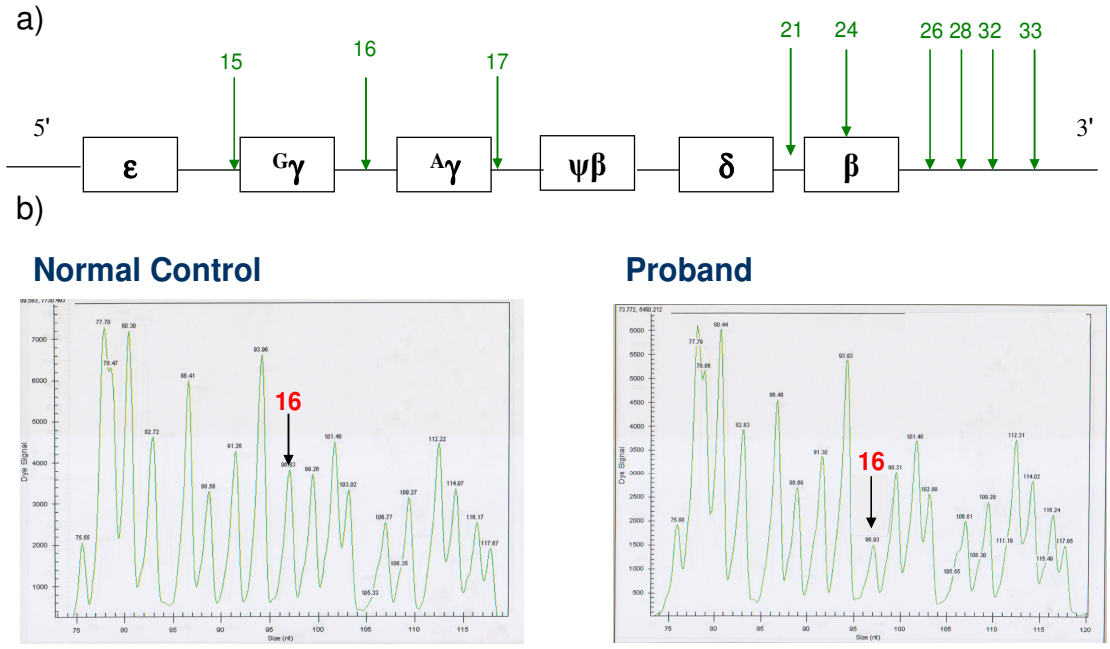


Figure 8.4 β-globin gene cluster MLPA results (ServiceXS) on the Asian Indian family. a) Schematic diagram of the position of the sequence where probe 16 anneals on the β-globin gene cluster (green arrows represent the positions of some of the probes) b) MLPA results of a normal control and the proband. Probe 16 is reduced by half height in the proband when compared to the control indicating the presence of a heterozygous deletion.

The MLPA results showed the presence of a heterozygous deletion however the proband had what would appear to be almost homozygous levels of HbA (70%). Reports of patients with heterozygous γ-thalassaemia have shown HbA levels of 45% so further studies on this family need to be carried out to see whether they also have a γ-globin gene rearrangement or γ-globin variant as well as γ-thalassaemia.

8.7.3 Identification of γ-thalassaemia by Gap-PCR

No further DNA was available from the Asian Indian newborn dried blood spots as all the blood spots had been utilised. Therefore in order to determine if the cohort of Indian infants who have increased HbA levels at birth are carriers for γ-thalassaemia, alternative samples were required to determine the frequency of γ-thalassaemia in the Asian Indian population. Whole blood samples from 300 adult Asian Indian (Pakistani, Bangladeshi and Indian) patients referred to the laboratory for either routine antenatal screening or partner testing were collected and the DNA extracted. A full blood count and haemoglobinopathy screen by HPLC had been carried out. Two hundred and forty three samples showed no evidence of a haemoglobinopathy, 48 had possible α^+-thalassaemia, 7 β-thalassaemia trait and 2 HbE trait.

A Gap-PCR was designed to flank the deleted region of the γ-thalassaemia deletion mutation already described. Primers were designed to flank the $^{G}\gamma$- and $^{A}\gamma$-gene: forward 5' TGTTGCTTTATAGGATTTTTCACT 3' and reverse 5' TGATGAGCAAATAAAAGCAGTAAA 3', giving a mutant band of 2,526bp. A set of control primers were also added which targeted a sequence in the β-

globin gene and therefore acted as non γ-gene specific primers: forward 5' TAATCTGAGCCAAGTAGAAG 3' and reverse 5' CACTGATGCAATCATTCGTC 3' and gave a control band size of 1,500bp. The father and proband of the Asian Indian family with the deletional γ-thalassaemia were used as a positive control (Figure 8.5). The PCR reaction was performed in 25μl reaction volume containing 100-250ng of genomic DNA, 2 units of *Taq* polymerase, 10pmol of the γ-gene primer, 0.5pmol of the control primers, 0.2mM dNTPs, 2.0mM $MgCl_2$,1 X Buffer IV (Abgene). The cycling conditions were 95^0C for 5 min, 30 cycles at 95^0C for 1 min, 62^0C for 2 min, 72^0C for 2 min and a final extension 72^0C for 10 mins.

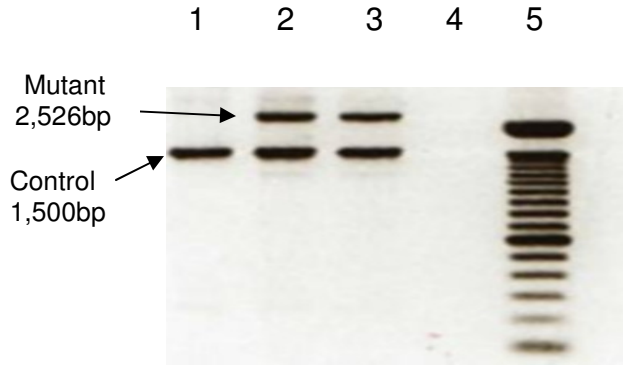


Figure 8.5 Gap-PCR showing the 4.9kb γ-thalassaemia deletion mutation.
Lane 1: Negative control
Lane 2: Proband
Lane 3: Father
Lane 4: Blank
Lane 5: 100bp marker

Of the 300 adult Asian Indians analysed by Gap-PCR all were negative for the 4.9kb γ-thalassaemia deletion mutation. This confirmed the low frequency of the 4.9kb γ-thalassaemia deletion mutation in the Asian Indian population.

8.7.4 Chromosomal abnormalities

Chromosomal abnormalities have been associated with the differences in the timing of haemoglobin switching in newborns. In 1966 there was a report on a child with a severe chromosomal abnormality (C/D translocation) who had started to switch from fetal to adult haemoglobin long before birth (Weller *et al*., 1966). His HbF% at 24 days was 9%. Two years later in 1968 the haemoglobin levels of 12 infants with Down's syndrome were compared against 48 normal term newborns (Wilson *et al*., 1968). The infants with Down's syndrome were found to have low HbF levels of 60%. Discussions with the newborn screening midwive in Oxford confirmed that chromosomal abnormalities were not the reason for the increased HbA levels seen in the Asian Indian newborns in this study. There was no evidence that any of the Asian Indian babies with elevated HbA levels had been referred to a paediatric clinic or had further tests for a possible chromosomal abnormality.

8.8 Discussion

Whilst performing newborn screening in Oxford I observed that Asian Indian newborns had higher HbA% when analysed by HPLC than other ethnic groups. Several studies in the 1980s showed that there were differences in the timing of haemoglobin switching amongst newborns. Many of these studies focused on the ratios of $^{G}\gamma$- and $^{A}\gamma$-chains at birth rather than HbF and HbA. They confirmed that some newborns had higher or lower $^{G}\gamma$- and $^{A}\gamma$-chain ratios than others. Only one report looked at the differences in the mean HbA values at birth in newborns from different countries but did not include Asian Indians (Kutlar *et al.*, 1987). They did however show that Far East Asian newborns (Chinese and Japanese) had lower HbA% when compared to Mediterranean and black African newborns. Other studies went on to report that γ-thalassaemia was present in these ethnic groups which would be expected to increase HbA levels. Therefore it was unclear why the 496 Asian Indian newborns analysed had had higher mean HbA levels (20.0%) at birth compared with the Caucasian (17.4%) and Black (18.6%) groups. Statistical analysis using the independent *t*-test showed that the standard deviation in the Asian Indian newborns was high (SD 7.16) and from the results it could be seen that there was a sub-group of Asian newborns with high HbA levels at birth.

Birth weight and gestational age are known factors that can increase HbA levels in newborns. Gestational differences were excluded by selecting newborns born at term (38 to 40 weeks). The birth weights of the 2,286 newborns were collated and statistical analysis of the mean birth weights within each ethnic group showed that the Asian Indians newborns were lighter at birth. A previous

large study of 6,123 newborns in France showed that babies with an increased birth weight have a higher HbA level at birth (Galacteros *et al*., 1991). Therefore it would not be expected that smaller birth weight newborns would have higher HbA% as evident in the Asian Indian group.

Increased birth weight, gestational age and chromosomal abnormalities were not able to explain the finding of higher HbA levels in Asian Indian newborns but maybe it could be explained by a cohort of Asian Indian newborns with γ-thalassaemia. Although γ-thalassaemia had been reported to be rare in most populations, accurate frequency data within many populations did not exist and only small sample numbers from Asian Indians from one particular region or tribe had been studied (Huisman *et al*., 1983, Huisman *et al*., 1991, Huisman *et al*., 1985b). Only 6 of the 10 samples chosen with high HbA levels in the Asian Indian cohort had sufficient amounts of good quality DNA for MLPA analysis of the β-globin gene cluster. Results showed no evidence of deletional γ-thalassaemia or a γ-globin gene rearrangement. Further studies on the differences in newborn haemoglobin levels between the ethnic groups would ideally require consent at birth for another dried blood spot or a cord blood sample to be taken from the newborn. DNA analysis of the γ-globin genes by MLPA would then be possible and the frequency of γ-thalassaemia and γ-globin gene rearrangements in each ethnic group determined. Obtaining further blood samples and undertaking family studies on specific cases with high HbA levels such as in the Asian Indian family referred from Leeds with γ-thalassaemia would be valuable in trying to determine the cause of the differences seen in HbA levels in newborns.

γ-thalassaemia is analogous to the 3.7kb α^+-globin gene deletion mutation and like the α-globin genes the γ-globin genes have highly homologous flanking sequences which represent a potential target area for genetic rearrangements. The 3.7kb α^+-thalassaemia deletion is very frequent in Asian Indians and therefore it would be thought that γ-thalassaemia is too. DNA was extracted from 300 adult Asian Indians and a Gap-PCR designed to detect the 4.9kb deletional γ-thalassaemia in order determine what the frequency of this mutation was in the Asian Indian population. The referral of the Asian Indian family with the 4.9kb γ-thalassaemia deletion mutation also indicated that these mutations exist within this population. All 300 adult samples gave negative results by Gap-PCR for the deletional 4.9kb γ-thalassaemia mutation. Results showed that the frequency of this mutation is low within Asian Indians but it might be that other novel types of γ-thalassaemia are prevalent. It is unlikely that there is only one type of γ-thalassaemia mutation and with the emergence of DNA techniques like MLPA, which is in routine use in most laboratories now, gene deletions and rearrangements can be more easily detected allowing for the detection of novel mutations. Therefore the hypothesis that γ-thalassaemia might be responsible for the increased HbA levels seen in several Asian Indian newborns can only be answered by further large scale studies.

Further DNA analysis by γ-globin gene sequencing would also be to identify the presence of the $^{A}\gamma$ -$^{A}\gamma$-globin gene rearrangement. Although reported to be relatively rare it has been found in black African, Chinese, Japanese and Mediterranean newborns (Huisman *et al*., 1983). This rearrangement is also

responsible for lowering HbF levels and therefore increasing HbA levels in newborns.

In summary, this is the first study to identify a cohort of Asian Indian newborns that undergo haemoglobin switching earlier than other ethnic groups. It was not possible to identify the precise mechanisms involved and large scale DNA studies would need to be carried out to determine whether γ-globin gene deletions or rearrangements are responsible for this finding. As well as looking at the γ-globin genes it would be interesting to study the other known factors involved in haemoglobin switching such as the transcription factors *KLF1* and *BCL11A*.

CHAPTER NINE: DISCUSSION

9.1 Summary of findings

The work described in this thesis aimed to increase our knowledge and understanding of what the genetic reasons are for the variations in HbF levels seen in the UK population with a view to further understanding some of the molecular mechanisms involved.

One well documented cause of HPFH is the inheritance of deletions in the β-globin gene cluster. Yet the frequency and range of deletional β-globin gene cluster mutations in the UK population was unknown. The work described in chapter 3 used samples taken from a large diverse ethnic population with an HPFH phenotype to determine that β-globin gene cluster deletions are present at significant frequencies in the UK population. 24% of the selected samples were found to have a deletion. A wide spectrum of mutations were identified and found in all major UK population ethnic groups. This study also evaluated the efficacy of the recently available technique of MLPA and concluded that the technique should be the recommended as the first choice method for screening for these clinically significant large deletion mutations in a routine diagnostic laboratory setting.

The work described in chapter 3 identified 4 possible novel deletions in the β-globin gene clusters of patients from the UK population. The work described in chapter 4 used several DNA techniques to try and determine the exact breakpoints in each deletion in order to gain an understanding of the molecular mechanisms involved. Two of the 4 deletions were characterised fully.

Deletion 1, a novel 909 bp β-globin gene mutation was the first reported case of a β-globin gene deletion in the Afghan population. The deletion was caused by a non homologous recombination event; however there were AT-rich stretches of sequences around the 5' breakpoint thought to be significant in the generation of β-globin gene deletions. Deletion 2 was the previously reported 1,393bp β-globin gene deletion, highlighting the fact that the MLPA technique cannot size deletions accurately when polymorphisms in the target sequence are present. A similar problem arose with deletion 3 as repeated efforts to map the breakpoints failed possibly due to incorrect sizing of the deletion by the MLPA probes. Microarray analysis showed the presence of a heterozygous $^{A}\gamma\delta\beta$-thalassaemia deletion. Further primer walking with the new array deletion breakpoints should enable confirmation of its precise size. Deletion 4 was a rare $(\varepsilon\gamma\delta\beta)^{0}$-thalassaemia deletion found in 12 family members of a large Asian Indian pedigree. All MLPA probe sets were reduced making identifying the exact breakpoints difficult. Microarray analysis confirmed a large $(\varepsilon\gamma\delta\beta)^{0}$-thalassaemia mutation which could be the Pakistani I deletion mutation.

Among the factors known to influence increases in HbF production are a number of point mutations in the promoter region of the γ-globin genes. The range and frequency of non-deletion HPFH γ-globin gene point mutations in the UK population was unknown. The work described in chapter 5 demonstrated that these mutations are a frequent cause of non-deletion HPFH in the UK and were found in 28 (21%) patients. Six of the mutations identified were previously reported point mutations and three were novel mutations. Interestingly most mutations were found in white British individuals and is a likely cause of

non-deletion HPFH in the British population. It is expected that this mutation has arisen independently and has been established through genetic drift. The frequency of γ-globin gene promoter mutations in patients with an unexplained increased HbF and a haemoglobinopathy were low. The three novel mutations were found in important regions for transcription factor binding. This suggested that the molecular mechanism responsible for the upregulation of HbF involved altering the binding site for a particular transcription factor.

The work described in chapter 6 looked at three polymorphisms in the major loci (*XmnI-HBG2, HBS1L-MYB* and *BCL11A*) that have been identified and shown previously to account for 20-50% of the variation in HbF levels in patients. The strongest statistical association with the three polymorphisms and HbF expression was seen in β-thalassaemia trait subjects with the *XmnI-HBG2* polymorphism. The SNPs in *BCL11A* and *HBS1L-MYB* failed to show statistical correlations with HbF. These results confirm the findings of previous studies that have suggested that *BCL11A* and *HBS1L-MYB* have a minor effect on HbF levels compared to the *XmnI* polymorphism in β-thalassaemia patients. Further large scale studies would need to be performed in order to accurately determine whether *BCL11A* and *HBS1L-MYB* are involved in increased HbF expression in our individuals with an HPFH phenotype.

Recent reports have identified mutations in the *KLF1* gene which are associated with a variety of phenotypes in humans. The work described in chapter 7 identified *KLF1* mutations in patients with increased HbF levels. Mutations predicted to have an effect were found in 11 (8.4%) out of 131 referrals with increased HbF levels and one patient with HbSS. All the mutations identified

were heterozygous indicating that a single altered *KLF1* allele can elevate HbF. Mutations were identified in individuals predominantly of African, Indian and Southeast Asian descent demonstrating that *KLF1* mutations could be a widespread cause of HPFH in malarial regions where haemoglobinopathies are common. The identification of a *KLF1* mutation in an individual with a particularly mild form of sickle cell disease required further investigation. Apart from the inheritance of co-existing α-thalassaemia there was no explanation as to why his Hb level was normal and he had an asymptomatic phenotype. If the *KLF1* mutation is responsible for his dramatic phenotype amelioration then it would be highly significant as it would provide *in-vivo* evidence that controlled reduction of *KLF1* expression could be an effective treatment for sickle cell disease. Whole genome SNP array analysis revealed that part of chromosome 11p showed an area with uniparental disomy (UPD). These results explained why there was no family history of sickle cell disease but did not offer any explanation as to why he was asymptomatic. Whole exome sequence analysis confirmed the two variants already identified but failed to identify any further mutations in other genes which could explain this unusual phenotype. Further family studies are required to ascertain the phenotype of the *KLF1* mutation on its own and with the sickle cell gene.

Finally, the finding that Asian Indian newborns appeared to undergo haemoglobin switching earlier than other ethnic groups was investigated in chapter 8. The reason for the increased HbA level observed in Asian Indian newborns was unclear. The ratio of HbA to HbF was studied in 2,286 term newborns from a variety of ethnic groups and Asian Indian newborns analysed

had higher mean HbA levels at birth compared with the Caucasian and black groups. Statistical analysis showed that there was a sub-group of Asian Indian newborns with high HbA levels at birth. Increased birth weight, gestation and chromosomal abnormalities were shown not to be the cause of the high HbA levels. The frequency of γ-thalassaemia which could be responsible for the earlier switching in the in the Asian Indian newborns was unknown. Six Asian Indian newborns with high HbA levels were analysed by MLPA for γ-thalassaemia and the results failed to show any evidence of deletional γ-thalassaemia or a γ-globin gene rearrangement. The recent referral of an Asian Indian family with the 4.9kb γ-thalassaemia deletion mutation indicated that these mutations exist within this population. Therefore 300 adult Asian Indian samples were analysed for the deletional 4.9kb γ-thalassaemia mutation and all gave a negative result for this deletion. The frequency of this mutation in Asian Indian is therefore low but what cannot be excluded is that other novel types of γ-thalassaemia might be prevalent. This is the first study to identify a cohort of Asian Indian newborns undergoing haemoglobin switching earlier than other ethnic groups. It was not possible to identify the precise mechanisms involved and large scale DNA studies would need to be carried out to try and determine whether γ-globin gene deletions or rearrangements are responsible for this finding.

9.2 Conclusions

Figure 9.1 shows a summary of the results (chapters 3-7) from the cohort of patients with increased HbF levels and patients with a haemoglobinopathy and unexplained elevated HbF. The results determine the frequency of deletional and non-deletion HPFH in the UK population which was otherwise unknown. As our laboratory is the UK National Haemoglobinopathy Reference Centre (NHRL) this study benefitted from being able to research patient samples from a wide variety of ethnic groups. This enabled the identification of the range of mutations present in all ethnic groups residing in the UK with the discovery of 16 novel mutations. Since the introduction of MLPA for detecting deletions in the globin gene clusters many new deletion mutations are being identified showing that deletions in these clusters occur more frequently than was first thought.

Figure 9.2 shows a summary of results of the frequency of non-deletion HPFH (chapters 5-7) in both patients with high HbF only and patients with a haemoglobinopathy and an unexplained raised HbF level. Being able to study these two cohorts of patients enabled the determination of whether some mutations are associated with disease traits. One interesting finding was that γ-promoter mutations appear to occur more frequently in populations without a significant haemoglobinopathy and that it is a likely cause of HPFH in white British individuals with only two mutations having been found in this ethnic group. There is a selective advantage for having increased HbF levels as HbF provides protection from malaria by the retardation of parasite growth. Therefore it is likely that the British γ-promoter mutations have arisen from

genetic drift and that similar to British α^0-thalassaemia these mutations originally came from malarial regions. Figure 9.2 also shows that *KLF1* mutations are responsible for increased HbF levels in non-deletion HPFH in both patients with and without a haemoglobinopathy. As discussed, mutations in the *KLF1* gene have been associated with several different phenotypes but this study also identified *KLF1* mutations with disease traits (HbEE and HbSS). Another interesting finding was that mutations in the *KLF1* gene were not linked to γ-promoter mutations.

This thesis has revealed the genetic factors responsible for an HPFH phenotype in 40% of the patient samples studied. Although the frequency of mutations was higher than originally suspected there are still a large proportion of patient samples (60%) with an unexplained raised HbF for which no genetic determinant could be found.

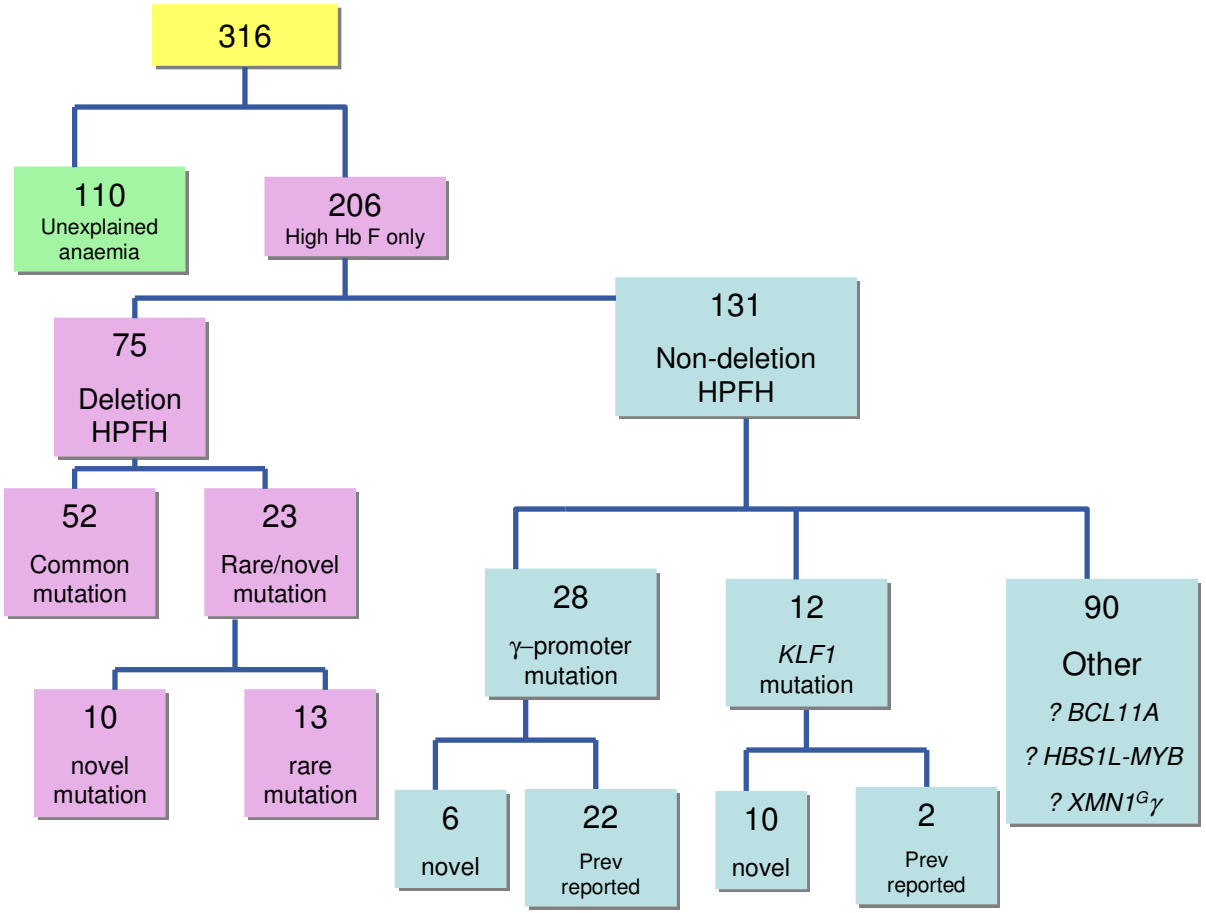


Figure 9.1 Summary of results showing the frequency of deletional and non deletion HPFH in the UK population.
The numbers in each box represent the number of patients with the each type of HPFH.

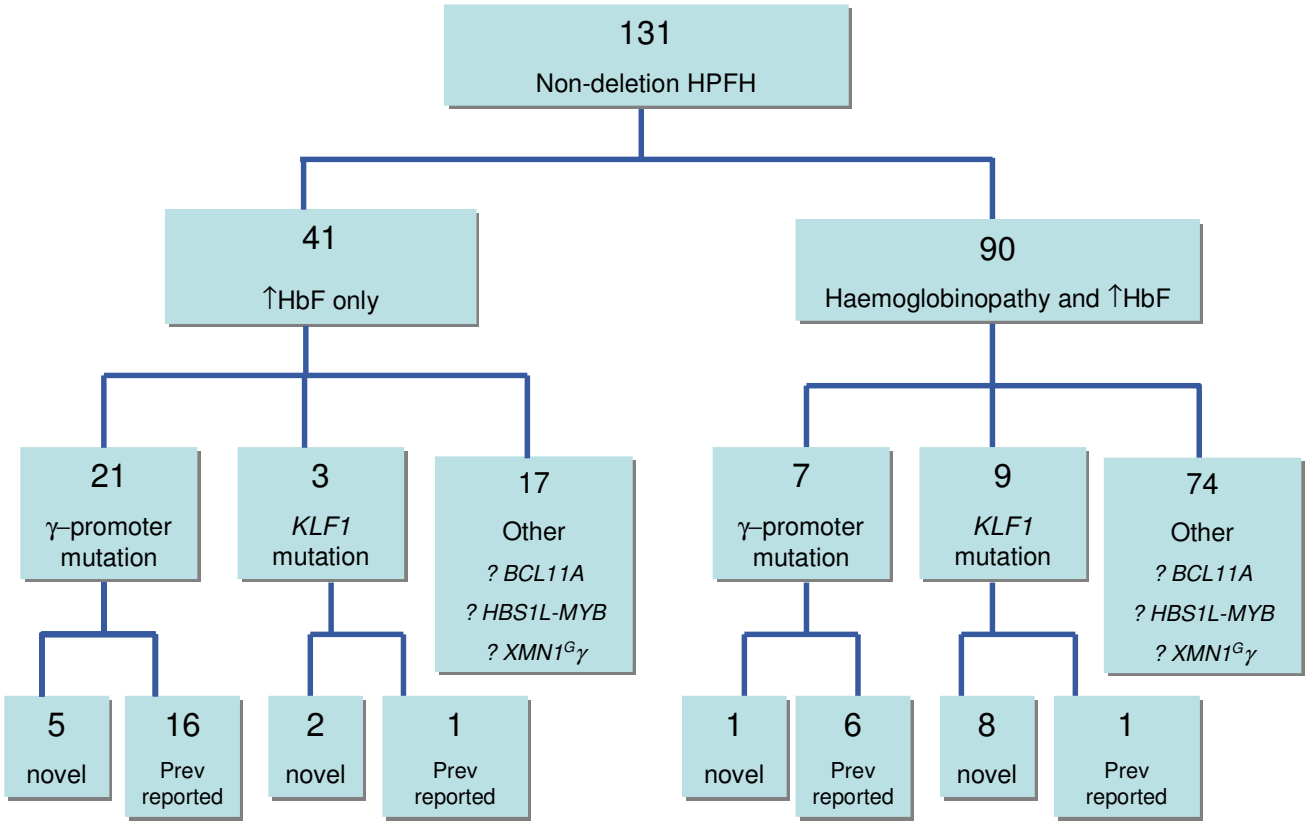


Figure 9.2 Summary of results showing the frequency of non deletion HPFH in both patients with high HbF only and patients with a haemoglobinopathy and an unexplained raised HbF level. The numbers in each box represent the number of patients with the each type of HPFH.

9.3 Future work

There are still many unanswered questions which require more extensive investigations. It was not possible with the small cohorts of samples in this study to fully evaluate the three loci (*XmnI-HBG2*, *HBS1L-MYB*, and *BCL11A*) shown in previous large GWAS studies to be responsible for variation in HbF levels. New research has also suggested that further genetic factors might play a role in the regulation of HbF (HDAC1/2, miRNA-15A/16-1, PRMT1, TR2, TR4 and COUP-TFII) (Sankaran, 2011). The fact that there are so many genetic factors recently identified and thought to be involved in the regulation of HbF lends promise to this field. In order to evaluate some of these potential factors the use of targeted next-generation sequencing (NGS) could be employed. Specific panels designed to screen the regions of interest for rare and common mutations causing red cell disorders (rare anaemias, red cell enzymopathies and the haemoglobinopathies) could be implemented in specialist centres. This would offer clinicians a one-stop effective test for screening patients and not only will discover known mutations but will allow scope for the discovery of new mutations and perhaps new genes. This new era of NGS technology holds a number of potential advantages over traditional methods, including the ability to fully sequence large numbers of genes (hundreds to thousands) in a single test. NGS can detect deletions, insertions, copy number alterations, translocations, and exome-wide base substitutions in all the known red cell related genes. Continuing advances in NGS technology should provide us with more insight into the molecular mechanisms involved with the potential of revealing new targets for therapeutic intervention.

REFERENCES

ALBITAR, M., CARE, A., PESCHLE, C. & LIEBHABER, S. A. (1992) Developmental switching of messenger RNA expression from the human alpha-globin cluster: fetal/adult pattern of theta-globin gene expression. *Blood,* 80**,** 1586-91.

ANAGNOU, N. P., PAPAYANNOPOULOU, T., NIENHUIS, A. W. & STAMATOYANNOPOULOS, G. (1988) Molecular characterization of a novel form of (A gamma delta beta)zero-thalassaemia deletion with a 3' breakpoint close to those of HPFH-3 and HPFH-4: insights for a common regulatory mechanism. *Nucleic Acids Res,* 16**,** 6057-66.

ANAGNOU, N. P., PAPAYANNOPOULOU, T., STAMATOYANNOPOULOS, G. & NIENHUIS, A. W. (1985) Structurally diverse molecular deletions in the beta-globin gene cluster exhibit an identical phenotype on interaction with the beta S-gene. *Blood,* 65**,** 1245-51.

ANAND, R., BOEHM, C. D., KAZAZIAN, H. H., JR. & VANIN, E. F. (1988) Molecular characterization of a beta zero-thalassemia resulting from a 1.4 kilobase deletion. *Blood,* 72**,** 636-41.

ARENDS, A. (2004) The Venezuelan type of non deletion HPFH is due to a C>T mutation at nt-211 of the Agamma globin gene. American Society of Hematology. Abstract 3728, *Blood* 104(11) Suppl 1.

ARNAUD, L., SAISON, C., HELIAS, V., LUCIEN, N., STESCHENKO, D., GIARRATANA, M. C., PREHU, C., FOLIGUET, B., MONTOUT, L., DE BREVERN, A. G., FRANCINA, A., RIPOCHE, P., FENNETEAU, O., DA COSTA, L., PEYRARD, T., COGHLAN, G., ILLUM, N., BIRGENS, H., TAMARY, H., IOLASCON, A., DELAUNAY, J., TCHERNIA, G. & CARTRON, J. P. (2010) A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. *Am J Hum Genet,* 87**,** 721-7.

BADENS, C., JOLY, P., AGOUTI, I., THURET, I., GONNET, K., FATTOUM, S., FRANCINA, A., SIMEONI, M. C., LOUNDOU, A. & PISSARD, S. (2011) Variants in genetic modifiers of beta-thalassemia can help to predict the major or intermedia type of the disease. *Haematologica,* 96**,** 1712-4.

BAE, H. T., BALDWIN, C. T., SEBASTIANI, P., TELEN, M. J., ASHLEY-KOCH, A., GARRETT, M., HOOPER, W. C., BEAN, C. J., DEBAUN, M. R., ARKING, D. E., BHATNAGAR, P., CASELLA, J. F., KEEFER, J. R., BARRON-CASELLA, E., GORDEUK, V., KATO, G. J., MINNITI, C., TAYLOR, J., CAMPBELL, A., LUCHTMAN-JONES, L., HOPPE, C., GLADWIN, M. T., ZHANG, Y. & STEINBERG, M. H. (2012) Meta-analysis of 2040 sickle cell anemia patients: BCL11A and HBS1L-MYB are the major modifiers of HbF in African Americans. *Blood,* 120**,** 1961-2.

BALLAS, S. K. (2001) Effect of alpha-globin genotype on the pathophysiology of sickle cell disease. *Pediatr Pathol Mol Med,* 20**,** 107-21.

BAUER, D. E. & ORKIN, S. H. (2011) Update on fetal hemoglobin gene regulation in hemoglobinopathies. *Curr Opin Pediatr,* 23**,** 1-8.

BHAVNANI, M., WICKHAM, M., AYYUB, H. & HIGGS, D. R. (1989) Alpha-thalassaemia in the north west of England. *Clin Lab Haematol,* 11**,** 293-7.

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

BOLLEKENS, J. A. & FORGET, B. G. (1991) Delta beta thalassemia and hereditary persistence of fetal hemoglobin. *Hematol Oncol Clin North Am,* 5**,** 399-422.

BORG, J., PAPADOPOULOS, P., GEORGITSI, M., GUTIERREZ, L., GRECH, G., FANIS, P., PHYLACTIDES, M., VERKERK, A. J., VAN DER SPEK, P. J., SCERRI, C. A., CASSAR, W., GALDIES, R., VAN IJCKEN, W., OZGUR, Z., GILLEMANS, N., HOU, J., BUGEJA, M., GROSVELD, F. G., VON LINDERN, M., FELICE, A. E., PATRINOS, G. P. & PHILIPSEN, S. (2010) Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet,* 42**,** 801-5.

BOUVA, M. J., HARTEVELD, C. L., BAKKER-VERWEIJ, G., VAN DELFT, P. & GIORDANO, P. C. (2006) Ggamma -37 (A-->T): a new nondeletional hereditary persistence of fetal hemoglobin determinant associated with the rare codon 91 (+T) delta0-thalassemia. *Hemoglobin,* 30**,** 371-7.

BOYER, S. H., BELDING, T. K., MARGOLTE, L., NOYES, A. N., BURKE, P. J. & BELL, W. R. (1975) Variations in the frequency of fetal hemoglobin-bearing erythrocytes (F-cells) in well adults, pregnant women, and adult leukemics. *Johns Hopkins Med J,* 137**,** 105-15.

CAMASCHELLA, C., OGGIANO, L., SAMPIETRO, M., GOTTARDI, E., ALFARANO, A., PISTIDDA, P., DORE, F., TARAMELLI, R., OTTOLENGHI, S. & LONGINOTTI, M. (1989) The homozygous state of G to A--117A gamma hereditary persistence of fetal hemoglobin. *Blood,* 73**,** 1999-2002.

CAMASCHELLA, C., SERRA, A., GOTTARDI, E., ALFARANO, A., REVELLO, D., MAZZA, U. & SAGLIO, G. (1990) A new hereditary persistence of fetal hemoglobin deletion has the breakpoint within the 3' beta-globin gene enhancer. *Blood,* 75**,** 1000-5.

CAO, A. & MOI, P. (2002) Regulation of the globin genes. *Pediatr Res,* 51**,** 415-21.

CAPPELLINI, M. D., FIORELLI, G. & BERNINI, L. F. (1981) Interaction between homozygous beta (0) thalassaemia and the Swiss type of hereditary persistence of fetal haemoglobin. *Br J Haematol,* 48**,** 561-72.

CAVAZZANA-CALVO, M., PAYEN, E., NEGRE, O., WANG, G., HEHIR, K., FUSIL, F., DOWN, J., DENARO, M., BRADY, T., WESTERMAN, K., CAVALLESCO, R., GILLET-LEGRAND, B., CACCAVELLI, L., SGARRA, R., MAOUCHE-CHRETIEN, L., BERNAUDIN, F., GIROT, R., DORAZIO, R., MULDER, G. J., POLACK, A., BANK, A., SOULIER, J., LARGHERO, J., KABBARA, N., DALLE, B., GOURMEL, B., SOCIE, G., CHRETIEN, S., CARTIER, N., AUBOURG, P., FISCHER, A., CORNETTA, K., GALACTEROS, F., BEUZARD, Y., GLUCKMAN, E., BUSHMAN, F., HACEIN-BEY-ABINA, S. & LEBOULCH, P. (2010) Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature,* 467**,** 318-22.

CHANG, Y. C., SMITH, K. D., MOORE, R. D., SERJEANT, G. R. & DOVER, G. J. (1995) An analysis of fetal hemoglobin variation in sickle cell disease: the relative contributions of the X-linked factor, beta-globin haplotypes, alpha-globin gene number, gender, and age. *Blood,* 85**,** 1111-7.

CHASSANIDIS, C., KALAMARAS, A., PHYLACTIDES, M., POURFARZAD, F., LIKOUSI, S., MAROULIS, V., PAPADAKIS, M. N., VAMVAKOPOULOS, N. K., ALEPOROU-MARINOU, V., PATRINOS, G. P. & KOLLIA, P. (2009) The Hellenic type of nondeletional hereditary persistence of fetal hemoglobin results from a novel mutation (g.-109G>T) in the HBG2 gene promoter. *Ann Hematol,* 88**,** 549-55.

CHEN, Z., LUO, H. Y., BASRAN, R. K., HSU, T. H., MANG, D. W., NUNTAKARN, L., ROSENFIELD, C. G., PATRINOS, G. P., HARDISON, R. C., STEINBERG, M. H. & CHUI, D. H. (2008) A T-to-G transversion at nucleotide -567 upstream of HBG2 in a GATA-1 binding motif is associated with elevated hemoglobin F. *Mol Cell Biol,* 28**,** 4386-93.

CLOSE, J., GAME, L., CLARK, B., BERGOUNIOUX, J., GEROVASSILI, A. & THEIN, S. L. (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**,** 33.

COLEMAN, M. B., ADAMS, J. G., 3RD, STEINBERG, M. H., PLONCZYNSKI, M. W., HARRELL, A. H., CASTRO, O. & WINTER, W. P. (1993) G gamma A gamma (beta+) hereditary persistence of fetal hemoglobin: the G gamma -158 C-->T mutation in cis to the -175 T-->C mutation of the A gamma-globin gene results in increased G gamma-globin synthesis. *Am J Hematol,* 42**,** 186-90.

COLLINS, F. S., COLE, J. L., LOCKWOOD, W. K. & IANNUZZI, M. C. (1987) The deletion in both common types of hereditary persistence of fetal hemoglobin is approximately 105 kilobases. *Blood,* 70**,** 1797-803.

COLLINS, F. S., METHERALL, J. E., YAMAKAWA, M., PAN, J., WEISSMAN, S. M. & FORGET, B. G. (1985) A point mutation in the A gamma-globin gene promoter in Greek hereditary persistence of fetal haemoglobin. *Nature,* 313**,** 325-6.

COLLINS, F. S., STOECKERT, C. J., JR., SERJEANT, G. R., FORGET, B. G. & WEISSMAN, S. M. (1984) G gamma beta+ hereditary persistence of fetal hemoglobin: cosmid cloning and identification of a specific mutation 5' to the G gamma gene. *Proc Natl Acad Sci U S A,* 81**,** 4894-8.

COSTA, F. F., ZAGO, M. A., CHENG, G., NECHTMAN, J. F., STOMING, T. A. & HUISMAN, T. H. (1990) The Brazilian type of nondeletional A gamma-fetal hemoglobin has a C----G substitution at nucleotide -195 of the A gamma-globin gene. *Blood,* 76**,** 1896-7.

CRAIG, J. E., BARNETSON, R. A., PRIOR, J., RAVEN, J. L. & THEIN, S. L. (1994) Rapid detection of deletions causing delta beta thalassemia and hereditary persistence of fetal hemoglobin by enzymatic amplification. *Blood,* 83**,** 1673-82.

CRAIG, J. E., ROCHETTE, J., FISHER, C. A., WEATHERALL, D. J., MARC, S., LATHROP, G. M., DEMENAIS, F. & THEIN, S. (1996) Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nat Genet,* 12**,** 58-64.

CREARY, L. E., ULUG, P., MENZEL, S., MCKENZIE, C. A., HANCHARD, N. A., TAYLOR, V., FARRALL, M., FORRESTER, T. E. & THEIN, S. L. (2009) Genetic variation on chromosome 6 influences F cell levels in healthy individuals of African descent and HbF levels in sickle cell patients. *PLoS One,* 4**,** e4218.

DE ANGIOLETTI, M., LACERRA. G., CARESTIA. C. (1997) *The 6th International Conference on Thalassaemia and the Haemoglobinopathies, Malta*; abstract 148

DONALD, J. A., LAMMI, A. & TRENT, R. J. (1988) Hemoglobin F production in heterocellular hereditary persistence of fetal hemoglobin and its linkage to the beta globin gene complex. *Hum Genet,* 80**,** 69-74.

DOVER, G. J., SMITH, K. D., CHANG, Y. C., PURVIS, S., MAYS, A., MEYERS, D. A., SHEILS, C. & SERJEANT, G. (1992) Fetal hemoglobin levels in sickle cell disease and normal individuals are partially controlled by an X-linked gene located at Xp22.2. *Blood,* 80**,** 816-24.

DRISCOLL, M. C., DOBKIN, C. S. & ALTER, B. P. (1989) Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proc Natl Acad Sci U S A,* 86**,** 7470-4.

EASTMAN, J. W., WONG, R., LIAO, C. L. & MORALES, D. R. (1996) Automated HPLC screening of newborns for sickle cell anemia and other hemoglobinopathies. *Clin Chem,* 42**,** 704-10.

EFREMOV, D. G., DIMOVSKI, A. J. & HUISMAN, T. H. (1994) The -158 (C-->T) promoter mutation is responsible for the increased transcription of the 3' gamma gene in the Atlanta type of hereditary persistence of fetal hemoglobin. *Blood,* 83**,** 3350-5.

EFREMOV, G. D., FILIPCE, V., GJORGOVSKI, I., JURICIC, D., STOJANOVSKI, N., HARANO, T., NAKATSUJI, T., KUTLAR, A., KUTLAR, F., BAKIOGLU, I. (1986a) G gamma A gamma(delta beta)zero-thalassaemia and a new form of gamma globin gene triplication identified in the Yugoslavian population. *Br J Haematol,* 63**,** 17-28.

EFREMOV, G. D., GJORGOVSKI, I., STOJANOVSKI, N., DIAZ-CHICO, J. C., HARANO, T., KUTLAR, F. & HUISMAN, T. H. (1987) One haplotype is associated with the Swiss type of hereditary persistence of fetal hemoglobin in the Yugoslavian population. *Hum Genet,* 77**,** 132-6.

EFREMOV, G. D., NIKOLOV, N., HATTORI, Y., BAKIOGLU, I. & HUISMAN, T. H. (1986b) The 18- to 23-kb deletion of the Macedonian delta beta-thalassemia includes the entire delta and beta globin genes. *Blood,* 68**,** 971-4.

FARRELL, J. J., SHERVA, R. M., CHEN, Z. Y., LUO, H. Y., CHU, B. F., HA, S. Y., LI, C. K., LEE, A. C., LI, R. C., YUEN, H. L., SO, J. C., MA, E. S., CHAN, L. C., CHAN, V., SEBASTIANI, P., FARRER, L. A., BALDWIN, C. T., STEINBERG, M. H. & CHUI, D. H. (2011) A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression. *Blood,* 117**,** 4935-45.

FEINGOLD, E. A. & 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**,** 2178-86.

FENG, W. C., SOUTHWOOD, C. M. & BIEKER, J. J. (1994) Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem,* 269**,** 1493-500.

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

FRASER, P., GRIBNAU, J. & TRIMBORN, T. (1998) Mechanisms of developmental regulation in globin loci. *Curr Opin Hematol,* 5**,** 139-44.

FRIEDMAN, S. & SCHWARTZ, E. (1976) Hereditary persistence of foetal haemoglobin with beta-chain synthesis in cis position (Ggamma-beta+-HPFH) in a negro family. *Nature,* 259**,** 138-40.

FUCHAROEN, S., SHIMIZU, K. & FUKUMAKI, Y. (1990) A novel C-T transition within the distal CCAAT motif of the G gamma-globin gene in the Japanese HPFH: implication of factor binding in elevated fetal globin expression. *Nucleic Acids Res,* 18**,** 5245-53.

FURUYA, C., YAMASHIRO, Y., HATTORI, Y., HINO, M., NISHIOKA, H., SHIMIZU, Y., OKANO, K. & HORIBE, K. (2008) A novel epsilon gamma delta beta thalassemia of 1.4 Mb deletion found in a Japanese patient. *Am J Hematol,* 83**,** 84-6.

GALACTEROS, F., GUILLOUD-BATAILLE, M. & FEINGOLD, J. (1991) Sex, gestational age, and weight dependancy of adult hemoglobin concentration in normal newborns. *Blood,* 78**,** 1121-4.

GALANELLO, R., BARELLA, S., SATTA, S., MACCIONI, L., PINTOR, C. & CAO, A. (2002) Homozygosity for nondeletion delta-beta(0) thalassemia resulting in a silent clinical phenotype. *Blood,* 100**,** 1913-4.

GALANELLO, R., ELEFTHERIOU, A., TRAEGER-SYNODINOS, J., OLD, J., PETROU, M., ANGASTINIOTIS, M. (2003) Prevention of Thalassaemias and other Haemoglobin Disorders. *Thalassaemia International Federation Publications*; 1, 150.

GALANELLO, R., MELIS, M. A., RUGGERI, R. & CAO, A. (1981) Prospective study of red blood cell indices, hemoglobin A2, and hemoglobin F in infants heterozygous for Beta-thalassemia. *J Pediatr,* 99**,** 105-8.

GALANELLO, R., SANNA, S., PERSEU, L., SOLLAINO, M. C., SATTA, S., LAI, M. E., BARELLA, S., UDA, M., USALA, G., ABECASIS, G. R. & CAO, A. (2009) Amelioration of Sardinian beta0 thalassemia by genetic modifiers. *Blood,* 114**,** 3935-7.

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 Genet,* 62**,** 1468-74.

GARNER, C., TATU, T., REITTIE, J. E., LITTLEWOOD, T., DARLEY, J., CERVINO, S., FARRALL, M., KELLY, P., SPECTOR, T. D. & THEIN, S. L. (2000) Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood,* 95**,** 342-6.

GELINAS, R., BENDER, M., LOTSHAW, C., WABER, P., KAZAZIAN, H., JR. & STAMATOYANNOPOULOS, G. (1986) Chinese A gamma fetal hemoglobin: C to T substitution at position-196 of the A gamma gene promoter. *Blood,* 67**,** 1777-9.

GELINAS, R., ENDLICH, B., PFEIFFER, C., YAGI, M. & STAMATOYANNOPOULOS, G. (1985) G to A substitution in the distal CCAAT box of the A gamma-globin gene in Greek hereditary persistence of fetal haemoglobin. *Nature,* 313**,** 323-5.

GEORGE, E., FARIDAH, K., TRENT, R. J., PADANILAM, B. J., HUANG, H. J. & HUISMAN, T. H. (1986) Homozygosity for a new type of G gamma (A gamma delta beta)zero-thalassemia in a Malaysian male. *Hemoglobin,* 10**,** 353-63.

GIARDINE, B., VAN BAAL, S., KAIMAKIS, P., RIEMER, C., MILLER, W., SAMARA, M., KOLLIA, P., ANAGNOU, N. P., CHUI, D. H., WAJCMAN, H., HARDISON, R. C. & PATRINOS, G. P. (2007) HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. *Hum Mutat,* 28**,** 206.

GIGLIONI, B., CASINI, C., MANTOVANI, R., MERLI, S., COMI, P., OTTOLENGHI, S., SAGLIO, G., CAMASCHELLA, C. & MAZZA, U. (1984) A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and beta-thalassemia. *EMBO J,* 3**,** 2641-5.

GILMAN, J. G. & HUISMAN, T. H. (1985) DNA sequence variation associated with elevated fetal G gamma globin production. *Blood,* 66**,** 783-7.

GILMAN, J. G., JOHNSON, M. E. & MISHIMA, N. (1988) Four base-pair DNA deletion in human A gamma globin-gene promoter associated with low A gamma expression in adults. *Br J Haematol,* 68**,** 455-8.

GILMAN, J. G., MISHIMA, N., WEN, X. J., KUTLAR, F. & HUISMAN, T. H. (1988a) Upstream promoter mutation associated with a modest elevation of fetal hemoglobin expression in human adults. *Blood,* 72**,** 78-81.

GILMAN, J. G., MISHIMA, N., WEN, X. J., STOMING, T. A., LOBEL, J. & HUISMAN, T. H. (1988b) Distal CCAAT box deletion in the A gamma globin gene of two black adolescents with elevated fetal A gamma globin. *Nucleic Acids Res,* 16**,** 10635-42.

GIORDANO, P. C., HARTEVELD, C. L., BOK, L. A., VAN DELFT, P., BATELAAN, D., BEEMER, F. A. & BERNINI, L. F. (1999) A complex haemoglobinopathy diagnosis in a family with both beta zero- and alpha (zero/+)-thalassaemia homozygosity. *Eur J Hum Genet,* 7**,** 163-8.

GUMUCIO, D. L., ROOD, K. L., BLANCHARD-MCQUATE, K. L., GRAY, T. A., SAULINO, A. & COLLINS, F. S. (1991) Interaction of Sp1 with the human gamma globin promoter: binding and transactivation of normal and mutant promoters. *Blood,* 78**,** 1853-63.

GUMUCIO, D. L., ROOD, K. L., GRAY, T. A., RIORDAN, M. F., SARTOR, C. I. & COLLINS, F. S. (1988) Nuclear proteins that bind the human gamma-globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. *Mol Cell Biol,* 8**,** 5310-22.

GUPTA, R. B., TIWARY, R. S., PANDE, P. L., KUTLAR, F., ONER, C., ONER, R. & HUISMAN, T. H. (1991) Hemoglobinopathies among the Gond tribal groups of central India; interaction of alpha- and beta-thalassemia with beta chain variants. *Hemoglobin,* 15**,** 441-58.

HARTEVELD, C. L., REFALDI, C., GIAMBONA, A., RUIVENKAMP, C. A., HOFFER, M. J., PIJPE, J., DE KNIJFF, P., BORGNA-PIGNATTI, C., MAGGIO, A., CAPPELLINI, M. D. & GIORDANO, P. C. (2013) Mosaic segmental uniparental isodisomy and progressive clonal selection: a common mechanism of late onset beta-thalassemia major. *Haematologica,* 98**,** 691-5.

HARTEVELD, C. L., VOSKAMP, A., PHYLIPSEN, M., AKKERMANS, N., DEN DUNNEN, J. T., WHITE, S. J. & GIORDANO, P. C. (2005) Nine unknown rearrangements in 16p13.3 and 11p15.4 causing alpha- and beta-thalassaemia characterised by high resolution multiplex ligation-dependent probe amplification. *J Med Genet,* 42**,** 922-31.

HENDERSON, S., TIMBS, A., MCCARTHY, J., GALLIENNE, A., VAN MOURIK, M., MASTERS, G., MAY, A., KHALIL, M. S., SCHUH, A. & OLD, J. (2009) Incidence of haemoglobinopathies in various populations - the impact of immigration. *Clin Biochem,* 42**,** 1745-56.

HENTHORN, P. S., SMITHIES, O. & MAGER, D. L. (1990) Molecular analysis of deletions in the human beta-globin gene cluster: deletion junctions and locations of breakpoints. *Genomics,* 6**,** 226-37.

HENTHORN, P. S., SMITHIES, O., NAKATSUJI, T., FELICE, A. E., GARDINER, M. B., REESE, A. L. & HUISMAN, T. H. (1985) (A gamma delta beta)^{0}-Thalassaemia in Blacks is due to a deletion of 34 kbp of DNA. *Br J Haematol,* 59**,** 343-56.

HERUTH, D. P., HAWKINS, T., LOGSDON, D. P., GIBSON, M. I., SOKOLOVSKY, I. V., NSUMU, N. N., MAJOR, S. L., FEGLEY, B., WOODS, G. M., LEWING, K. B., NEVILLE, K. A., CORNETTA, K., PETERSON, K. R. & WHITE, R. A. (2010) Mutation in erythroid specific transcription factor KLF1 causes Hereditary Spherocytosis in the Nan hemolytic anemia mouse model. *Genomics,* 96**,** 303-7.

HIGGS, D. R., AYYUB, H., CLEGG, J. B., HILL, A. V., NICHOLLS, R. D., TEAL, H., WAINSCOAT, J. S. & WEATHERALL, D. J. (1985) Alpha thalassaemia in British people. *Br Med J (Clin Res Ed),* 290**,** 1303-6.

HIGGS, D. R., ENGEL, J. D. & STAMATOYANNOPOULOS, G. (2012) Thalassaemia. *Lancet,* 379**,** 373-83.

HOFFBRAND, A. V. & PETTIT, J. E. (1993) Essential Haematology. *Blackwell Scientific Publications (3rd Ed)*, pp95.

HUANG, C. H., CHANG, Y. Y., CHEN, C. H. & KO, T. M. (2008) Molecular characterization of a beta-globin gene deletion of 1357 bp in a Taiwanese beta-thalassemia carrier. *Hemoglobin,* 32**,** 498-504.

HUANG, H. J., STOMING, T. A., HARRIS, H. F., KUTLAR, F. & HUISMAN, T. H. (1987) The Greek A gamma beta+-HPFH observed in a large black family. *Am J Hematol,* 25**,** 401-8.

HUANG, X. D., YANG, X. O., HUANG, R. B., ZHANG, H. Y., ZHAO, H. L., ZHAO, Y. J., HUANG, C. L., HOU, C. J. & ZHANG, J. W. (2000) A novel four base-pair deletion within the Agamma-GLOBin gene promoter associated with slight increase of Agamma expression in adult. *Am J Hematol,* 63**,** 16-9.

HUISMAN, T. H., CHEN, S. S., NAKATSUJI, T. & KUTLAR, F. (1985a) A second family with the Atlanta type of HPFH. *Hemoglobin,* 9**,** 393-8.

HUISMAN, T. H., KUTLAR, F. & GU, L. H. (1991) Gamma chain abnormalities and gamma-globin gene rearrangements in newborn babies of various populations. *Hemoglobin,* 15**,** 349-79.

HUISMAN, T. H., KUTLAR, F., NAKATSUJI, T., BRUCE-TAGOE, A., KILINC, Y., CAUCHI, M. N. & ROMERO GARCIA, C. (1985b) The frequency of the gamma chain variant A gamma T in different populations, and its use in evaluating gamma gene expression in association with thalassemia. *Hum Genet,* 71**,** 127-33.

HUISMAN, T. H., REESE, A. L., GARDINER, M. B., WILSON, J. B., LAM, H., REYNOLDS, A., NAGLE, S., TROWELL, P., ZENG, Y. T., HUANG, S. Z., SUKUMARAN, P. K., MIWA, S., EFREMOV, G. D., PETKOV, G., SCIARRATTA, G. V. & SANSONE, G. (1983) The occurrence of different levels of G gamma chain and of the A gamma T variant of fetal hemoglobin in newborn babies from several countries. *Am J Hematol,* 14**,** 133-48.

INDRAK, K., INDRAKOVA, J., KUTLAR, F., POSPISILOVA, D., SULOVSKA, I., BAYSAL, E. & HUISMAN, T. H. (1991) Compound heterozygosity for a beta zero-thalassemia (frameshift codons 38/39; -C) and a nondeletional Swiss type of HPFH (A----C at NT -110, G gamma) in a Czechoslovakian family. *Ann Hematol,* 63**,** 111-5.

JAWAID, K., WAHLBERG, K., THEIN, S. L. & BEST, S. (2010) Binding patterns of BCL11A in the globin and GATA1 loci and characterization of the BCL11A fetal hemoglobin locus. *Blood Cells Mol Dis,* 45**,** 140-6.

JOLY, P., LACAN, P., GARCIA, C., COUPRIE, N. & FRANCINA, A. (2009) Identification and molecular characterization of four new large deletions in the beta-globin gene cluster. *Blood Cells Mol Dis,* 43**,** 53-7.

JONES, R. W., OLD, J. M., TRENT, R. J., CLEGG, J. B. & WEATHERALL, D. J. (1981a) Major rearrangement in the human beta-globin gene cluster. *Nature,* 291**,** 39-44.

JONES, R. W., OLD, J. M., TRENT, R. J., CLEGG, J. B. & WEATHERALL, D. J. (1981b) Restriction mapping of a new deletion responsible for G gamma (delta beta)o thalassemia. *Nucleic Acids Res,* 9**,** 6813-25.

JOSHI, D. D., NICKERSON, H. J. & MCMANUS, M. J. (2004) Hydrops fetalis caused by homozygous alpha-thalassemia and Rh antigen alloimmunization: report of a survivor and literature review. *Clin Med Res,* 2**,** 228-32.

KIEFER, C. M., HOU, C., LITTLE, J. A. & DEAN, A. (2008) Epigenetics of beta-globin gene regulation. *Mutat Res,* 647**,** 68-76

KORNBLIT, B., TAANING, P. & BIRGENS, H. (2005) Beta-thalassemia due to a novel nonsense mutation at codon 37 (TGG-->TAG) found in an Afghanistani family. *Hemoglobin,* 29**,** 209-13.

KOSTEAS, T., PALENA, A. & ANAGNOU, N. P. (1997) Molecular cloning of the breakpoints of the hereditary persistence of fetal hemoglobin type-6 (HPFH-6) deletion and sequence analysis of the novel juxtaposed region from the 3' end of the beta-globin gene cluster. *Hum Genet,* 100**,** 441-5.

KRUGLUGER, W. & HOPMEIER, P. (2002) Identification of a compound beta-thalassemia homozygosity [codon 10 (GCC-->GCA) and codon 16 (-C)] in an Afghan family. *Hemoglobin,* 26**,** 317-20.

KULOZIK, A. E., BELLAN-KOCH, A., KOHNE, E. & KLEIHAUER, E. (1992) A deletion/inversion rearrangement of the beta-globin gene cluster in a Turkish family with delta beta zero-thalassemia intermedia. *Blood,* 79**,** 2455-9.

KUTLAR, A., GARDINER, M. B., HEADLEE, M. G., REESE, A. L., CLEEK, M. P., NAGLE, S., SUKUMARAN, P. K. & HUISMAN, T. H. (1984) Heterogeneity in the molecular basis of three types of hereditary persistence of fetal hemoglobin and the relative synthesis of the G gamma and A gamma types of gamma chain. *Biochem Genet,* 22**,** 21-35.

KUTLAR, F., KUTLAR, A., GU, Y. C. & HUISMAN, T. H. (1987) Adult hemoglobin levels in newborn babies from different countries and in babies with some significant hemoglobinopathies. *Acta Haematol,* 78**,** 28-32.

LABIE, D., DUNDA-BELKHODJA, O., ROUABHI, F., PAGNIER, J., RAGUSA, A. & NAGEL, R. L. (1985) The -158 site 5' to the G gamma gene and G gamma expression. *Blood,* 66**,** 1463-5.

LABIE, D. & ELION, J. (1996) Sequence polymorphisms of potential functional relevance in the beta-globin gene locus. *Hemoglobin,* 20**,** 85-101.

LEDERER, C. W., BASAK, A. N., AYDINOK, Y., CHRISTOU, S., EL-BESHLAWY, A., ELEFTHERIOU, A., FATTOUM, S., FELICE, A. E., FIBACH, E., GALANELLO, R., GAMBARI, R., GAVRILA, L., GIORDANO, P. C., GROSVELD, F., HASSAPOPOULOU, H., HLADKA, E., KANAVAKIS, E., LOCATELLI, F., OLD, J., PATRINOS, G. P., ROMEO, G., TAHER, A., TRAEGER-SYNODINOS, J., VASSILIOU, P., VILLEGAS, A., VOSKARIDOU, E., WAJCMAN, H., ZAFEIROPOULOS, A. & KLEANTHOUS, M. (2009) An electronic infrastructure for research and treatment of the thalassemias and other hemoglobinopathies: the Euro-mediterranean ITHANET project. *Hemoglobin,* 33**,** 163-76.

LEONOVA, J., KAZANETZ, E. G., SMETANINA, N. S., ADEKILE, A. D., EFREMOV, G. D. & HUISMAN, T. H. (1996) Variability in the fetal hemoglobin level of the normal adult. *Am J Hematol,* 53**,** 59-65.

LETTRE, G., SANKARAN, V. G., BEZERRA, M. A., ARAUJO, A. S., UDA, M., SANNA, S., CAO, A., SCHLESSINGER, D., COSTA, F. F., HIRSCHHORN, J. N. & ORKIN, S. H. (2008) DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc Natl Acad Sci U S A,* 105**,** 11869-74.

LI, Q., DUAN, Z. J. & STAMATOYANNOPOULOS, G. (2001) Analysis of the mechanism of action of non-deletion hereditary persistence of fetal hemoglobin mutants in transgenic mice. *EMBO J,* 20**,** 157-64.

LIU, Y. T., OLD, J. M., MILES, K., FISHER, C. A., WEATHERALL, D. J. & CLEGG, J. B. (2000) Rapid detection of alpha-thalassaemia deletions and alpha-globin gene triplication by multiplex polymerase chain reactions. *Br J Haematol,* 108**,** 295-9.

LOSEKOOT, M., FODDE, R., GERRITSEN, E. J., VAN DE KUIT, I., SCHREUDER, A., GIORDANO, P. C., VOSSEN, J. M. & BERNINI, L. F. (1991) Interaction of two different disorders in the beta-globin gene cluster associated with an increased hemoglobin F production: a novel deletion type of (G) gamma + ((A) gamma delta beta)(0)-thalassemia and a delta(0)-hereditary persistence of fetal hemoglobin determinant. *Blood,* 77**,** 861-7.

LUNTER, G. & GOODSON, M. (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res,* 21**,** 936-9.

MANTOVANI, R., MALGARETTI, N., NICOLIS, S., RONCHI, A., GIGLIONI, B. & OTTOLENGHI, S. (1988) The effects of HPFH mutations in the human gamma-globin promoter on binding of ubiquitous and erythroid specific nuclear factors. *Nucleic Acids Res,* 16**,** 7783-97.

MARTI, H. R. & BUETLER, R. (1961) [Hemoglobin F and hemoglobin A2 increase in the Swiss population]. *Acta Haematol,* 26**,** 65-74.

MASON, K. P., GRANDISON, Y., HAYES, R. J., SERJEANT, B. E., SERJEANT, G. R., VAIDYA, S. & WOOD, W. G. (1982) Post-natal decline of fetal haemoglobin in homozygous sickle cell disease: relationship to parenteral Hb F levels. *Br J Haematol,* 52**,** 455-63.

MATSUNAGA, E., KIMURA, A., YAMADA, H., FUKUMAKI, Y. & TAKAGI, Y. (1985) A novel deletion in delta beta-thalassemia found in Japan. *Biochem Biophys Res Commun,* 126**,** 185-91.

MENZEL, S., GARNER, C., GUT, I., MATSUDA, F., YAMAGUCHI, M., HEATH, S., FOGLIO, M., ZELENIKA, D., BOLAND, A., ROOKS, H., BEST, S., SPECTOR, T. D., FARRALL, M., LATHROP, M. & THEIN, S. L. (2007) A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet,* 39**,** 1197-9.

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**,** 2776-86.

MISHIMA, N., LANDMAN, H., HUISMAN, T. H. & GILMAN, J. G. (1989) The DNA deletion in an Indian delta beta-thalassaemia begins one kilobase from the A gamma globin gene and ends in an L1 repetitive sequence. *Br J Haematol,* 73**,** 375-9.

MIYOSHI, K., KANETO, Y., KAWAI, H., OHCHI, H., NIKI, S., HASEGAWA, K., SHIRAKAMI, A. & YAMANO, T. (1988) X-linked dominant control of F-cells in normal adult life: characterization of the Swiss type as hereditary persistence of fetal hemoglobin regulated dominantly by gene(s) on X chromosome. *Blood,* 72**,** 1854-60.

MOTUM, P. I., DENG, Z. M., HUONG, L. & TRENT, R. J. (1994) The Australian type of nondeletional G gamma-HPFH has a C-->G substitution at nucleotide -114 of the G gamma gene. *Br J Haematol,* 86**,** 219-21.

MOTUM, P. I., HAMILTON, T. J., LINDEMAN, R., LE, H. & TRENT, R. J. (1993) Molecular characterisation of Vietnamese HPFH. *Hum Mutat,* 2**,** 179-84.

MOTUM, P. I., LINDEMAN, R., HAMILTON, T. J. & TRENT, R. J. (1992) Australian beta zero-thalassaemia: a high haemoglobin A2 beta zero-thalassaemia due to a 12 kb deletion commencing 5' to the beta-globin gene. *Br J Haematol,* 82**,** 107-13.

NG, P. C. & HENIKOFF, S. (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res,* 31**,** 3812-4.

NGUYEN, T. K., JOLY, P., BARDEL, C., MOULSMA, M., BONELLO-PALOT, N. & FRANCINA, A. (2010) The XmnI (G)gamma polymorphism influences hemoglobin F synthesis contrary to BCL11A and HBS1L-MYB SNPs in a cohort of 57 beta-thalassemia intermedia patients. *Blood Cells Mol Dis,* 45**,** 124-7.

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.

NUINOON, M., MAKARASARA, W., MUSHIRODA, T., SETIANINGSIH, I., WAHIDIYAT, P. A., SRIPICHAI, O., KUMASAKA, N., TAKAHASHI, A., SVASTI, S.,MUNKONGDEE,T., MAHASIRIMONGKOL, S., PEERAPITTAYAMONGKOL, C., VIPRAKASIT, V., KAMATANI, N., WINICHAGOON, P., KUBO, M., NAKAMURA, Y. & FUCHAROEN, S. (2010) A genome-wide association identified the common genetic variants influence disease severity in beta0-thalassemia/hemoglobin E. *Hum Genet,* 127**,** 303-14.

OLD, J. M., AYYUB, H., WOOD, W. G., CLEGG, J. B. & WEATHERALL, D. J. (1982) Linkage analysis of nondeletion hereditary persistence of fetal hemoglobin. *Science,* 215**,** 981-2.

ONER, R., KUTLAR, F., GU, L. H. & HUISMAN, T. H. (1991) The Georgia type of nondeletional hereditary persistence of fetal hemoglobin has a C---T mutation at nucleotide-114 of the A gamma-globin gene. *Blood,* 77**,** 1124-5.

ONER, R., ONER, C., ERDEM, G., BALKAN, H., OZDAG, H., ERKAN, M., GUMRUK, F., GURGEY, A. & ALTAY, C. (1996) A novel (delta beta)(0)-thalassemia due to a approximately 30-kb deletion observed in a Turkish family. *Acta Haematol,* 96**,** 232-6.

OTTOLENGHI, S., CAMASCHELLA, C., COMI, P., GIGLIONI, B., LONGINOTTI, M., OGGIANO, L., DORE, F., SCIARRATTA, G., IVALDI, G., SAGLIO, G. (1988) A frequent A gamma-hereditary persistence of fetal hemoglobin in northern Sardinia: its molecular basis and hematologic phenotype in heterozygotes and compound heterozygotes with beta-thalassemia. *Hum Genet,* 79**,** 13-7.

OTTOLENGHI, S., GIGLIONI, B., TARAMELLI, R., COMI, P., MAZZA, U., SAGLIO, G., CAMASCHELLA, C., IZZO, P., CAO, A., GALANELLO, R., GIMFERRER, E., BAIGET, M. & GIANNI, A. M. (1982) Molecular comparison of delta beta-thalassemia and hereditary persistence of fetal hemoglobin DNAs: evidence of a regulatory area? *Proc Natl Acad Sci U S A,* 79**,** 2347-51.

PALENA, A., BLAU, A., STAMATOYANNOPOULOS, G. & ANAGNOU, N. P. (1994) Eastern European (delta beta) zero-thalassemia: molecular characterization of a novel 9.1-kb deletion resulting in high levels of fetal hemoglobin in the adult. *Blood,* 83**,** 3738-45.

PATRINOS, G. P., GIARDINE, B., RIEMER, C., MILLER, W., CHUI, D. H., ANAGNOU, N. P., WAJCMAN, H. & HARDISON, R. C. (2004) Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. *Nucleic Acids Res,* 32**,** D537-41.

PATRINOS, G. P., KOLLIA, P., LOUTRADI-ANAGNOSTOU, A., LOUKOPOULOS, D. & PAPADAKIS, M. N. (1998) The Cretan type of non-deletional hereditary persistence of fetal hemoglobin [A gamma-158C-->T] results from two independent gene conversion events. *Hum Genet,* 102**,** 629-34.

PERRINE, S. P. (2005) Fetal globin induction--can it cure beta thalassemia? *Hematology Am Soc Hematol Educ Program***,** 38-44.

PERSEU, L., SATTA, S., MOI, P., DEMARTIS, F. R., MANUNZA, L., SOLLAINO, M. C., BARELLA, S., CAO, A. & GALANELLO, R. (2011) KLF1 gene mutations cause borderline HbA2. *Blood*.

PETERSON, K. R., LI, Q. L., CLEGG, C. H., FURUKAWA, T., NAVAS, P. A., NORTON, E. J., KIMBROUGH, T. G. & STAMATOYANNOPOULOS, G. (1995) Use of yeast artificial chromosomes (YACs) in studies of mammalian development: production of beta-globin locus YAC mice carrying human globin developmental mutants. *Proc Natl Acad Sci U S A,* 92**,** 5655-9.

PHYLIPSEN, M., AMATO, A., CAPPABIANCA, M. P., TRAEGER-SYNODINOS, J., KANAVAKIS, E., BASAK, N., GALANELLO, R., TUVERI, T., IVALDI, G., HARTEVELD, C. L. & GIORDANO, P. C. (2009) Two new beta-thalassemia deletions compromising prenatal diagnosis in an Italian and a Turkish couple seeking prevention. *Haematologica,* 94**,** 1289-92.

PHYLIPSEN, M., CHAIBUNRUANG, A., VOGELAAR, I. P., BALAK, J. R., SCHAAP, R. A., ARIYUREK, Y., FUCHAROEN, S., DEN DUNNEN, J. T., GIORDANO, P. C., BAKKER, E. & HARTEVELD, C. L. (2012) Fine-tiling array CGH to improve diagnostics for alpha- and beta-thalassemia rearrangements. *Hum Mutat,* 33**,** 272-80.

PISSARD, S., M'RAD, A., BEUZARD, Y. & ROMEO, P. H. (1996) A new type of hereditary persistence of fetal haemoglobin (HPFH): HPFH Tunisia beta + (+C-200)G gamma. *Br J Haematol,* 95**,** 67-72.

RAMENSKY, V., BORK, P. & SUNYAEV, S. (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res,* 30**,** 3894-900.

RIMMER, A., MATHIESON, I., LUNTER, G. & MCVEAN, G. *(2012)* Platypus: An Integrated Variant Caller *(www.well.ox.ac.uk/platypus).*

ROCHETTE, J., CRAIG, J. E. & THEIN, S. L. (1994) Fetal hemoglobin levels in adults. *Blood Rev,* 8**,** 213-24.

RONCHI, A., NICOLIS, S., SANTORO, C. & OTTOLENGHI, S. (1989) Increased Sp1 binding mediates erythroid-specific overexpression of a mutated (HPFH) gamma-globulin promoter. *Nucleic Acids Res,* 17**,** 10231-41.

ROOKS, H., BERGOUNIOUX, J., GAME, L., CLOSE, J. P., OSBORNE, C., BEST, S., SENIOR, T., HEIGHT, S., THOMPSON, R., HADZIC, N., FRASER, P., BOLTON-MAGGS, P. & THEIN, S. L. (2005) Heterogeneity of the epsilon gamma delta beta-thalassaemias: characterization of three novel English deletions. *Br J Haematol,* 128**,** 722-9.

ROOKS, H., CLARK, B., BEST, S., RUSHTON, P., OAKLEY, M., THEIN, O. S., CUTHBERT, A. C., BRITLAND, A., RUF, A. & THEIN, S. L. (2012) A novel 506kb deletion causing epsilongammadeltabeta thalassemia. *Blood Cells Mol Dis,* 49**,** 121-7.

RUTLAND, P. C., PEMBREY, M. E. & DAVIES, T. (1983) The estimation of fetal haemoglobin in healthy adults by radioimmunoassay. *Br J Haematol,* 53**,** 673-82.

SAGLIO, G., CAMASCHELLA, C., SERRA, A., BERTERO, T., REGE CAMBRIN, G., GUERRASIO, A., MAZZA, U., IZZO, P., TERRAGNI, F., GIGLIONI, B. & ET AL. (1986) Italian type of deletional hereditary persistence of fetal hemoglobin. *Blood,* 68**,** 646-51.

SALLER, E., MORADKHANI, K., DUTLY, F., VINATIER, I., PREHU, C., FRISCHKNECHT, H. & GOOSSENS, M. (2012) Comparison of two known chromosomal rearrangements in the deltabeta-globin complex with identical DNA breakpoints but causing different Hb A(2) levels. *Hemoglobin,* 36**,** 177-82.

SAMPIETRO, M., THEIN, S. L., CONTRERAS, M. & PAZMANY, L. (1992) Variation of HbF and F-cell number with the G-gamma Xmn I (C-T) polymorphism in normal individuals. *Blood,* 79**,** 832-3.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SANKARAN, V. G. (2011) Targeted therapeutic strategies for fetal hemoglobin induction. *Hematology Am Soc Hematol Educ Program,* 459-65.

SANKARAN, V. G., MENNE, T. F., XU, J., AKIE, T. E., LETTRE, G., VAN HANDEL, B., MIKKOLA, H. K., HIRSCHHORN, J. N., CANTOR, A. B. & ORKIN, S. H. (2008) Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science,* 322**,** 1839-42.

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

SANKARAN, V. G., XU, J., RAGOCZY, T., IPPOLITO, G. C., WALKLEY, C. R., MAIKA, S. D., FUJIWARA, Y., ITO, M., GROUDINE, M., BENDER, M. A., TUCKER, P. W. & ORKIN, S. H. (2009) Developmental and species-divergent globin switching are driven by BCL11A. *Nature,* 460**,** 1093-7.

SANKARAN, V. G., XU, J., BYRON, R., GREISMAN, H. A., FISHER, C., WEATHERALL, D. J., SABATH, D. E., GROUDINE, M., ORKIN, S. H., PREMAWARDHENA, A. & BENDER, M. A. (2011) A functional element necessary for fetal hemoglobin silencing. *N Engl J Med,* 365**,** 807-14.

SATTA, S., PERSEU, L., MOI, P., ASUNIS, I., CABRIOLU, A., MACCIONI, L., DEMARTIS, F. R., MANUNZA, L., CAO, A. & GALANELLO, R. (2011) Compound heterozygosity for KLF1 mutations associated with remarkable increase of fetal hemoglobin and red cell protoporphyrin. *Haematologica,* 96**,** 767-70.

SAYERS, E. W., BARRETT, T., BENSON, D. A., BOLTON, E., BRYANT, S. H., CANESE, K., CHETVERNIN, V., CHURCH, D. M., DICUCCIO, M., FEDERHEN, S., FEOLO, M., FINGERMAN, I. M., GEER, L. Y., HELMBERG, W., KAPUSTIN, Y., LANDSMAN, D., LIPMAN, D. J., LU, Z., MADDEN, T. L., MADEJ, T., MAGLOTT, D. R., MARCHLER-BAUER, A., MILLER, V., MIZRACHI, I., OSTELL, J., PANCHENKO, A., PHAN, L., PRUITT, K. D., SCHULER, G. D., SEQUEIRA, E., SHERRY, S. T., SHUMWAY, M., SIROTKIN, K., SLOTTA, D., SOUVOROV, A., STARCHENKO, G., TATUSOVA, T. A., WAGNER, L., WANG, Y., WILBUR, W. J., YASCHENKO, E. & YE, J. (2011) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res,* 39**,** D38-51.

SCHOUTEN, J. P., MCELGUNN, C. J., WAAIJER, R., ZWIJNENBURG, D., DIEPVENS, F. & PALS, G. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res,* 30**,** e57.

SCHROEDER, W. A., HUISMAN, T. H. & SUKUMARAN, P. K. (1973) A second type of hereditary persistence of foetal haemoglobin in India. *Br J Haematol,* 25**,** 131-5.

SEDGEWICK, A. E., TIMOFEEV, N., SEBASTIANI, P., SO, J. C., MA, E. S., CHAN, L. C., FUCHAROEN, G., FUCHAROEN, S., BARBOSA, C. G., VARDARAJAN, B. N., FARRER, L. A., BALDWIN, C. T., STEINBERG, M. H. & CHUI, D. H. (2008) BCL11A is a major HbF quantitative trait locus in three different populations with beta-hemoglobinopathies. *Blood Cells Mol Dis,* 41**,** 255-8.

SHALEV, H., LANDAU, D., PISSARD, S., KRASNOV, T., KAPELUSHNIK, J., GILAD, O., BROIDES, A., DGANY, O. & TAMARY, H. (2013) A novel epsilon gamma delta beta thalassemia presenting with pregnancy complications and severe neonatal anemia. *Eur J Haematol,* 90**,** 127-33.

SHIOKAWA, S., YAMADA, H., TAKIHARA, Y., MATSUNAGA, E., OHBA, Y., YAMAMOTO, K. & FUKUMAKI, Y. (1988) Molecular analysis of Japanese delta beta-thalassemia. *Blood,* 72**,** 1771-6.

SIATECKA, M. & BIEKER, J. J. (2011) The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood*.

SINGER, S. T., STYLES, L., BOJANOWSKI, J., QUIROLO, K., FOOTE, D. & VICHINSKY, E. P. (2000) Changing outcome of homozygous alpha-thalassemia: cautious optimism. *J Pediatr Hematol Oncol,* 22**,** 539-42.

SINGLETON, B. K., BURTON, N. M., GREEN, C., BRADY, R. L. & ANSTEE, D. J. (2008) Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. *Blood,* 112**,** 2081-8.

SMITH, M. B. & CAUCHI, M. N. (1982) Analysis of rates of globin synthesis in normal human fetuses and neonates and in twins. *Am J Hematol,* 12**,** 121-30.

STAMATOYANNOPOULOS, G., VEITH, R., AL-KHATTI, A., FRITSCH, E. F., GOLDWASSER, E. & PAPAYANNOPOULOU, T. (1987) On the induction of fetal hemoglobin in the adult; stress erythropoiesis, cell cycle-specific drugs, and recombinant erythropoietin. *Prog Clin Biol Res,* 251**,** 443-53.

STAMATOYANNOPOULOS, G., GROSVELD, F. (2001) Hemoglobin switching. In: Stamatoyannopoulos G, Majerus PW, Perlmutter RM, Varmus H, eds. *Molecular Basis of Blood Diseases*. 3rd ed. Philadelphia, PA:WB Saunders;:pp135-182

STAMATOYANNOPOULOS, G. (2005) Control of globin gene expression during development and erythroid differentiation. *Exp Hematol,* 33**,** 259-71.

STEINBERG, M. H. & SEBASTIANI, P. (2012) Genetic modifiers of sickle cell disease. *Am J Hematol,* 87**,** 795-803.

STEINBERG, M. H. & NAGEL, R. L. (2009) Genetic modulation of sickle cell disease and thalassaemia. In Disorders of Hemoglobin,, STEINBERG, M. H., FORGET, B. G., HIGGS, D. R. & WEATHERALL, D J. *Cambridge University Press*, 2nd Ed, pp638-657

STOMING, T. A., STOMING, G. S., LANCLOS, K. D., FEI, Y. J., ALTAY, C., KUTLAR, F. & HUISMAN, T. H. (1989) An A gamma type of nondeletional hereditary persistence of fetal hemoglobin with a T----C mutation at position -175 to the cap site of the A gamma globin gene. *Blood,* 73**,** 329-33.

STREETLY, A., LATINOVIC, R., HALL, K. & HENTHORN, J. (2009) Implementation of universal newborn bloodspot screening for sickle cell disease and other clinically significant haemoglobinopathies in England: screening results for 2005-7. *J Clin Pathol,* 62**,** 26-30.

STUPPIA, L., ANTONUCCI, I., PALKA, G. & GATTA, V. (2012) Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases. *Int J Mol Sci,* 13**,** 3245-76.

SUKUMARAN, P. K., NAKATSUJI, T., GARDINER, M. B., REESE, A. L., GILMAN, J. G. & HUISMAN, T. H. (1983) Gamma thalassemia resulting from the deletion of a gamma-globin gene. *Nucleic Acids Res,* 11**,** 4635-43.

SVASTI, S., PAKSUA, S., NUCHPRAYOON, I., WINICHAGOON, P. & FUCHAROEN, S. (2007) Characterization of a novel deletion causing (deltabeta)0-thalassemia in a Thai family. *Am J Hematol,* 82**,** 155-61.

SWENSEN, J. J., AGARWAL, A. M., ESQUILIN, J. M., SWIERCZEK, S., PERUMBETI, A., HUSSEY, D., LEE, M., JOINER, C. H., PONT-KINGDON, G., LYON, E. & PRCHAL, J. T. (2010) Sickle cell disease resulting from uniparental disomy in a child who inherited sickle cell trait. *Blood,* 116**,** 2822-5.

TASIOPOULOU, M., BOUSSIOU, M., SINOPOULOU, K., MORAITIS, G., LOUTRADI-ANAGNOSTOU, A. & KARABABA, P. (2008) G gamma-196 C-->T, A gamma-201 C-->T: two novel mutations in the promoter region of the gamma-globin genes associated with nondeletional hereditary persistence of fetal hemoglobin in Greece. *Blood Cells Mol Dis,* 40**,** 320-2.

TATE, V. E., WOOD, W. G. & WEATHERALL, D. J. (1986) The British form of hereditary persistence of fetal hemoglobin results from a single base mutation adjacent to an S1 hypersensitive site 5' to the A gamma globin gene. *Blood,* 68**,** 1389-93.

THEIN, S. L. (2008) Genetic modifiers of the beta-haemoglobinopathies. *Br J Haematol,* 141**,** 357-66.

THEIN, S. L. & CRAIG, J. E. (1998) Genetics of Hb F/F cell variance in adults and heterocellular hereditary persistence of fetal hemoglobin. *Hemoglobin,* 22**,** 401-14.

THEIN, S. L., HESKETH, C., BROWN, J. M., ANSTEY, A. V. & WEATHERALL, D. J. (1989) Molecular characterization of a high A2 beta thalassemia by direct sequencing of single strand enriched amplified genomic DNA. *Blood,* 73**,** 924-30.

THEIN, S. L. & MENZEL, S. (2009) Discovering the genetics underlying foetal haemoglobin production in adults. *Br J Haematol,* 145**,** 455-67.

THEIN, S. L., MENZEL, S., LATHROP, M. & GARNER, C. (2009) Control of fetal hemoglobin: new insights emerging from genomics and clinical implications. *Hum Mol Genet,* 18**,** R216-23.

THEIN, S. L., MENZEL, S., PENG, X., BEST, S., JIANG, J., CLOSE, J., SILVER, N., GEROVASILLI, A., PING, C., YAMAGUCHI, M., WAHLBERG, K., ULUG, P., SPECTOR, T. D., GARNER, C., MATSUDA, F., FARRALL, M. & LATHROP, M. (2007) Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. *Proc Natl Acad Sci U S A,* 104**,** 11346-51.

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**,** 214-28.

THEIN, S. L., WAINSCOAT, J. S., SAMPIETRO, M., OLD, J. M., CAPPELLINI, D., FIORELLI, G., MODELL, B. & WEATHERALL, D. J. (1987) Association of thalassaemia intermedia with a beta-globin gene haplotype. *Br J Haematol,* 65**,** 367-73.

TIMBS, A. T., RUGLESS, M. J., GALLIENNE, A. E., HAYWOOD, A. M., HENDERSON, S. J. & OLD, J. M. (2012) Prenatal diagnosis of hemoglobinopathies by pyrosequencing: a more sensitive and rapid approach to fetal genotyping. *Hemoglobin,* 36**,** 144-50.

TRENT, R. J., JONES, R. W., CLEGG, J. B., WEATHERALL, D. J., DAVIDSON, R. & WOOD, W. G. (1984) (A gamma delta beta) thalassaemia: similarity of phenotype in four different molecular defects, including one newly described. *Br J Haematol,* 57**,** 279-89.

TRENT, R. J., SVIRKLYS, L. & JONES, P. (1988) Thai (delta beta)0-thalassemia and its interaction with gamma-thalassemia. *Hemoglobin,* 12**,** 101-14.

TRENT, R. J., WILLIAMS, B. G., KEARNEY, A., WILKINSON, T. & HARRIS, P. C. (1990) Molecular and hematologic characterization of Scottish-Irish type (epsilon gamma delta beta)zero thalassemia. *Blood,* 76**,** 2132-8.

TUAN, D., FEINGOLD, E., NEWMAN, M., WEISSMAN, S. M. & FORGET, B. G. (1983) Different 3' end points of deletions causing delta beta-thalassemia and hereditary persistence of fetal hemoglobin: implications for the control of gamma-globin gene expression in man. *Proc Natl Acad Sci U S A,* 80**,** 6937-41.

TUAN, D., SOLOMON, W., LI, Q. & LONDON, I. M. (1985) The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A,* 82**,** 6384-8.

UDA, M., GALANELLO, R., SANNA, S., LETTRE, G., SANKARAN, V. G., CHEN, W., USALA, G., BUSONERO, F., MASCHIO, A., ALBAI, G., PIRAS, M. G., SESTU, N., LAI, S., DEI, M., MULAS, A., CRISPONI, L., NAITZA, S., ASUNIS, I., DEIANA, M., NAGARAJA, R., PERSEU, L., SATTA, S., CIPOLLINA, M. D., SOLLAINO, C., MOI, P., HIRSCHHORN, J. N., ORKIN, S. H., ABECASIS, G. R., SCHLESSINGER, D. & CAO, A. (2008) Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A,* 105**,** 1620-5.

ULRICH, M. J., GRAY, W. J. & LEY, T. J. (1992) An intramolecular DNA triplex is disrupted by point mutations associated with hereditary persistence of fetal hemoglobin. *J Biol Chem,* 267**,** 18649-58.

VORUGANTI, I., ENG, B. & WAYE, J. S. (2009) Molecular characterization of a novel 55.1 kb (G)gamma((A)gammadeltabeta)(0)-thalassemia deletion in two Canadian families. *Hemoglobin,* 33**,** 422-7.

WAHLBERG, K., JIANG, J., ROOKS, H., JAWAID, K., MATSUDA, F., YAMAGUCHI, M., LATHROP, M., THEIN, S. L. & BEST, S. (2009) The HBS1L-MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells. *Blood,* 114**,** 1254-62.

WANG, K., LI, M. & HAKONARSON, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res,* 38**,** e164.

WAYE, J. S., ENG, B., HUNT, J. A. & CHUI, D. H. (1994) Filipino beta-thalassemia due to a large deletion: identification of the deletion endpoints and polymerase chain reaction (PCR)-based diagnosis. *Hum Genet,* 94**,** 530-2.

WEATHERALL, D. J. (2000) Single gene disorders or complex traits: lessons from the thalassaemias and other monogenic diseases. *BMJ,* 321**,** 1117-20.

WEATHERALL, D. J., CARTNER, R., CLEGG, J. B., WOOD, W. G., MACRAE, I. A. & MACKENZIE, A. (1975) A form of hereditary persistence of fetal haemoglobin characterized by uneven cellular distribution of haemoglobin F and the production of haemoglobins A and A2 in homozygotes. *Br J Haematol,* 29**,** 205-20.

WEATHERALL, D. J. & CLEGG, J. B. (2001) Inherited haemoglobin disorders: an increasing global health problem. *Bull World Health Organ,* 79**,** 704-12.

WEATHERALL, D. J., WOOD, W. G., JONES, R. W. & CLEGG, J. B. (1985) The developmental genetics of human hemoglobin. *Prog Clin Biol Res,* 191**,** 3-25.

WEATHERALL, D. J., HIGGS, D. R., BUNCH, C., OLD, J. M., HUNT, D. M., PRESSLEY, L., CLEGG, J. B., BETHLENFALVAY, N. C., SJOLIN, S., KOLER, R. D., MAGENIS, E., FRANCIS, J. L. & BEBBINGTON, D. (1981) Hemoglobin H disease and mental retardation: a new syndrome or a remarkable coincidence? *N Engl J Med,* 305**,** 607-12.

WELLER, S. D., APLEY, J. & RAPER, A. B. (1966) Malformations associated with precocious synthesis of adult haemoglobin. A new chromosomal anomaly syndrome. *Lancet,* 1**,** 777-9.

WELLS, J. C., SHARP, G., STEER, P. J. & LEON, D. A. (2013) Paternal and maternal influences on differences in birth weight between Europeans and Indians born in the UK. *PLoS One,* 8**,** e61116.

WHO (1994) Guidelines for the control of haemoglobin disorders. In: *Report of the VIth Annual Meeting of the WHO Working Group on Haemoglobinopathies, Cagliari, Sardinia, 8-9 April, 1989.* World Health Organisation, Geneva.

WILSON, M. G., SCHROEDER, W. A. & GRAVES, D. A. (1968) Postnatal change of hemoglobins F and A2 in infants with Down's syndrome (G trisomy). *Pediatrics,* 42**,** 349-53.

WOLK, M., MARTIN, J. E. & REINUS, C. (2006) Development of fetal haemoglobin-blood cells (F cells) within colorectal tumour tissues. *J Clin Pathol,* 59**,** 598-602.

WOOD, W. G. (1989) HbF production in adult life. *Prog Clin Biol Res,* 316B**,** 251-67.

WOOD, W. G. (1993) Increased HbF in adult life. *Baillieres Clin Haematol,* 6**,** 177-213.

WOOD, W. G., WEATHERALL, D. J., CLEGG, J. B., HAMBLIN, T. J., EDWARDS, J. H. & BARLOW, A. M. (1977) Heterocellular hereditary persistence of fetal haemoglobin (heterocellular HPFH) and its interaction with beta thalassaemia. *Br J Haematol,* 36**,** 461-73.

XU, J., SANKARAN, V. G., NI, M., MENNE, T. F., PURAM, R. V., KIM, W. & ORKIN, S. H. (2010) Transcriptional silencing of γ-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev,* 24**,** 783-98.

YAMASHIRO, Y., HATTORI, Y., OKAYAMA, N., SHINODA, E., SUYAMA, N., TANAKA, T. & OHI, S. (2005) A novel (g)gamma(a)gamma(deltabeta)zero-thalassemia with a 27 kb deletion. *Hemoglobin,* 29**,** 197-208.

YANG, K. G., LIU, J. Z., KUTLAR, F., KUTLAR, A., ALTAY, C., GURGEY, A. & HUISMAN, T. H. (1986) Beta zero-thalassemia in association with a gamma-globin gene quadruplication. *Blood,* 68**,** 1394-7.

ZENG, Y. T., HUANG, S. Z., CHEN, B., LIANG, Y. C., CHANG, Z. M., HARANO, T. & HUISMAN, T. H. (1985) Hereditary persistence of fetal hemoglobin or (delta beta) zero-thalassemia: three types observed in South-Chinese families. *Blood,* 66**,** 1430-5.

ZERTAL-ZIDANI, S., DUCROCQ, R., WEIL-OLIVIER, C., ELION, J. & KRISHNAMOORTHY, R. (2001) A novel delta beta fusion gene expresses hemoglobin A (HbA) not Hb Lepore: Senegalese delta(0)beta(+) thalassemia. *Blood,* 98**,** 1261-3.

ZERTAL-ZIDANI, S., MERGHOUB, T., DUCROCQ, R., GERARD, N., SATTA, D. & KRISHNAMOORTHY, R. (1999) A novel C-->A transversion within the distal CCAAT motif of the G gamma-globin gene in the Algerian G gammabeta+-hereditary persistence of fetal hemoglobin. *Hemoglobin,* 23**,** 159-69.

ZHANG, J. W., SONG, W. F., ZHAO, Y. J., WU, G. Y., QIU, Z. M., WANG, F. N., CHEN, S. S. & STAMATOYANNOPOULOS, G. (1993) Molecular characterization of a novel form of (A gamma delta beta)zero thalassemia deletion in a Chinese family. *Blood,* 81**,** 1624-9.

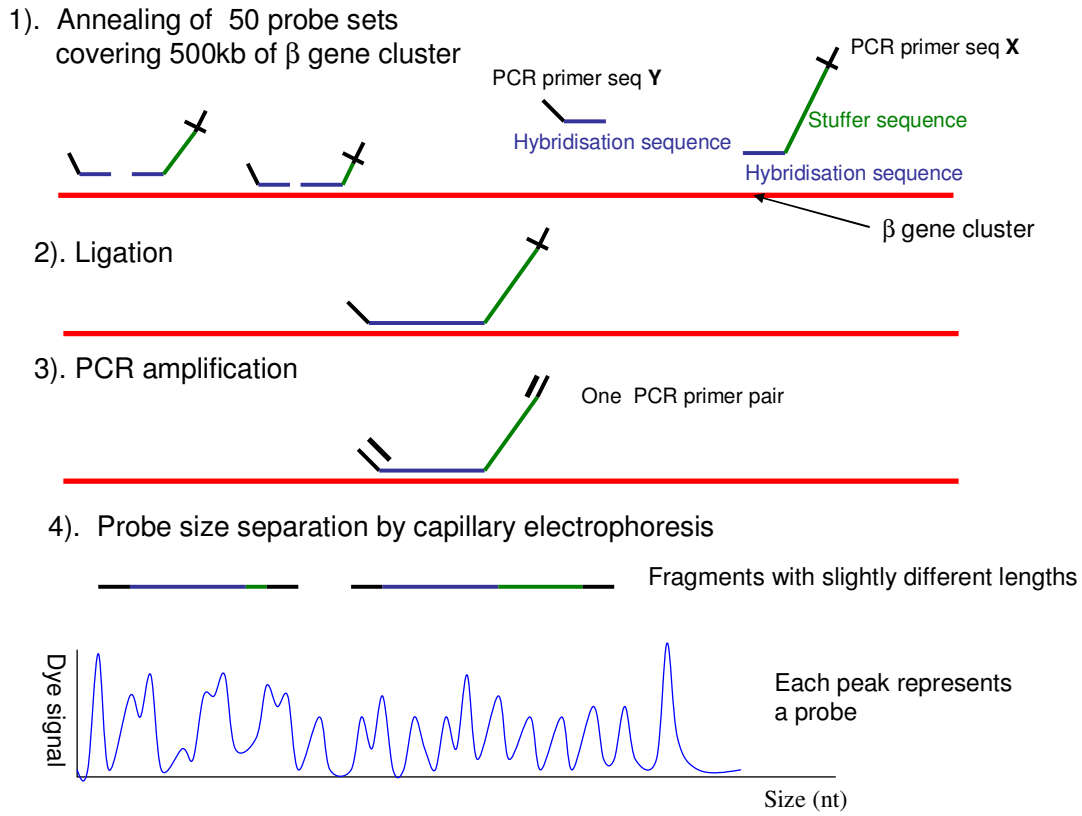
ZHANG, J. W., STAMATOYANNOPOULOS, G. & ANAGNOU, N. P. (1988) Laotian (delta beta) (0)-thalassemia: molecular characterization of a novel deletion associated with increased production of fetal hemoglobin. *Blood,* 72**,** 983-8.

ZHOU, D., LIU, K., SUN, C. W., PAWLIK, K. M. & TOWNES, T. M. (2010) KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. *Nat Genet,* 42**,** 742-4.

APPENDICES

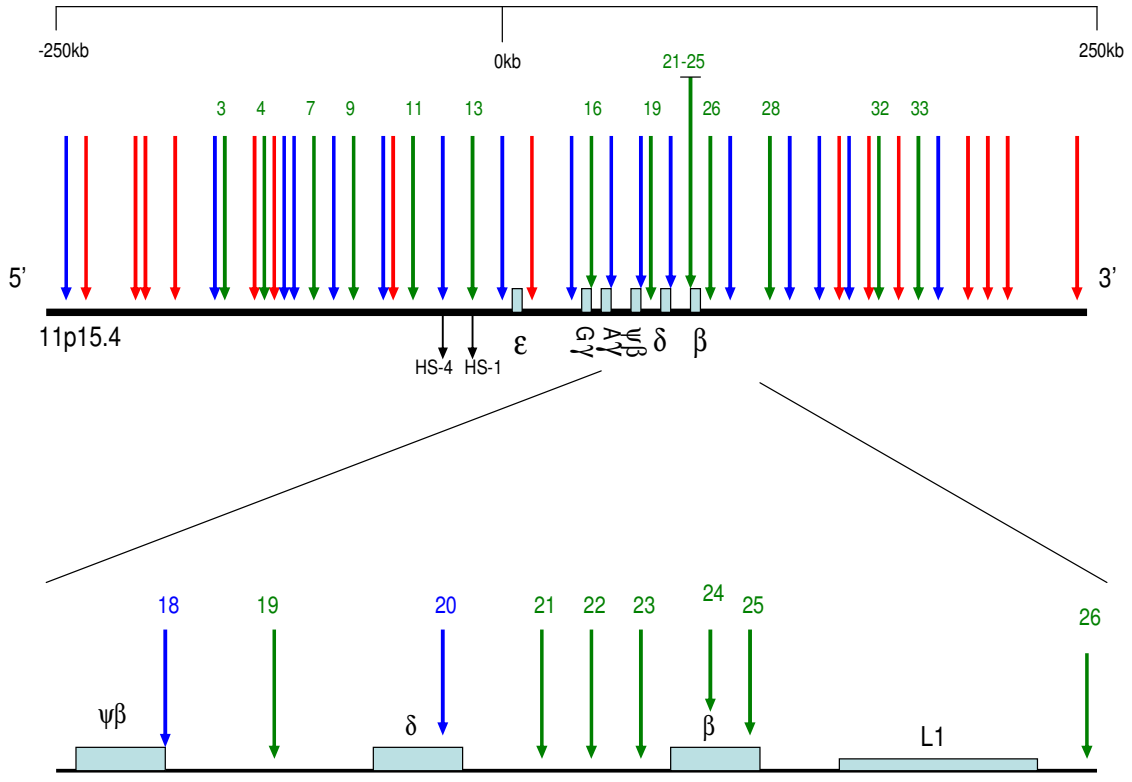
APPENDIX 1: Diagram showing the principle of the MLPA technique

Outline of the MLPA reaction (adapted and modified from Schouten *et al.*, 2002 and www.mrc-holland.com)



APPENDIX 2: Schematic representation of β-globin gene cluster MLPA probe set designed by Harteveld et al (2005).

The genes throughout the β-globin gene cluster are represented by light blue boxes. The three (green, blue and red) probe set positions in the cluster are indicated by the coloured arrows. The bottom diagram shows a closer representation of the probes whose target sequence is in the region where the δ- and β-globin genes are found. (adapted and modified from Harteveld *et al*., 2005)



APPENDIX 3: Pyrosequencing assay design rs 11886868

a) Pyromark Assay Design Software 2.0 showing the primer set properties for SNP rs 11886868

chr2:60720083+60720375

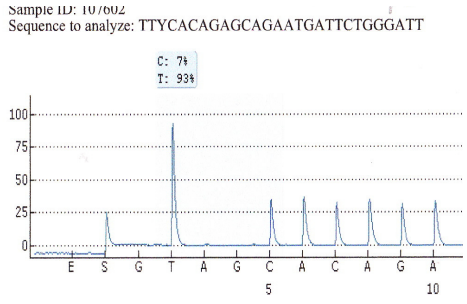
Primer Set 1	Score: 100
General Warnings	

	F1	TGCCCCTTTGCTGTCAAT
	R1	ACCATGGATGAATCCCAGAAT
	S1	CGTCTTTTGTGTTTAATTTC

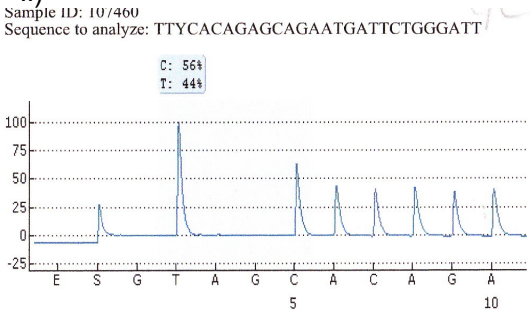
	PCR Product	Forward PCR Primer, F1	Reverse PCR Primer, R1	Sequencing Primer, S1
Length, bp	183	18	21	20
Position, 5'- 3'		17 - 34	199 - 179	142 - 161
Warnings				
Tm, ºC		69.2	69.3	52.0
%GC	44.3	50.0	42.9	30.0
Sequence to Analyze	TTNCACAGAG CAGAATGATT CTGGGATT			

b) Results of pyrosequencing rs 11886868: i) wild type (TT) ii) heterozygote (TC) and iii) homozygote (CC) samples.

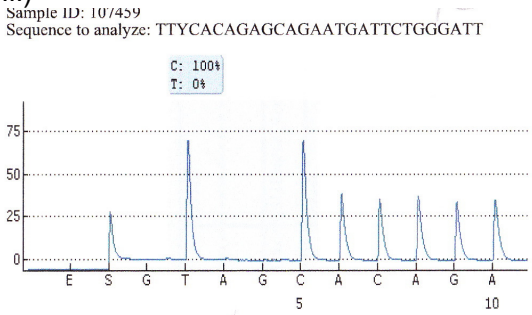
i)



ii)



iii)



APPENDIX 4: Pyrosequencing assay design rs 9399137

a) Pyromark Assay Design Software 2.0 showing the primer set properties for SNP rs 9399137

chr6:135418885+135419101

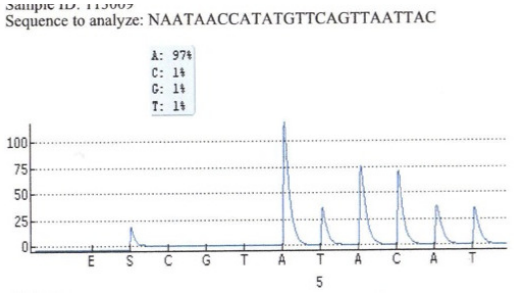
Primer Set 1	Score: 100
General Warnings	

→	F1	CAACATCACCTTAAAAGGCGGTAT
←	R1	GCAGGGTTGCTTGTGAAAAAA
←	S1	TGCTTGTGAAAAAACTGT

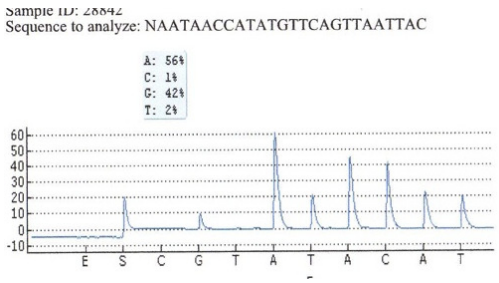
	PCR Product	Forward PCR Primer, F1	Reverse PCR Primer, R1	Sequencing Primer, S1
Length, bp	94	24	21	18
Position, 5'- 3'		66 - 89	159 - 139	152 - 135
Warnings				
Tm, ºC		71.2	71.2	51.7
%GC	33.0	41.7	42.9	33.3
Sequence to Analyze	NAATAACCAT ATGTTCAGTT AATTAC			

b) Results of pyrosequencing rs 11886868: i) wild type (TT) ii) heterozygote (TC) and iii) homozygote (CC) samples.

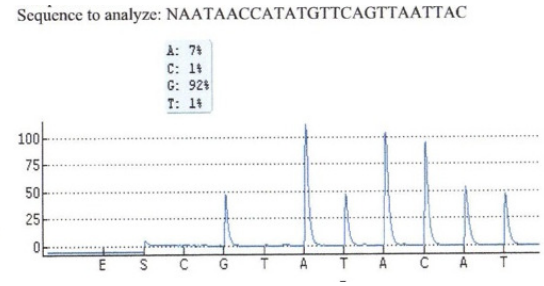
i)



ii)



iii)



APPENDIX 5: Statistical analysis

Chapter 6 required the use of two statistical tests, one-way ANOVA and Dunnett's test. Chapter 8 analysed data using an independent *t*-test. All statistics were performed using the software package Analyse-it®.

ANOVA

Analysis of variance (ANOVA) test was used to analyse the differences between the group means and their associated "variation" in HbF among the groups. ANOVA provides a statistical test of whether or not the means of several groups are equal, and therefore generalizes the *t*-test to more than two groups. ANOVAs are useful in comparing three or more means (groups or variables) for statistical significance.

Dunnett's test

ANOVA is able to say whether there was an overall difference between the groups analysed but cannot tell which specific groups differed. Therefore a further test is required. The Dunnett's test was chosen as it is the most powerful method at performing pair-wise comparisons. This test is able to make comparisons between the groups and also compares groups against a control (in this study the wild type allele).

Independent *t*-test

The independent *t*-test compares the means between two unrelated groups on the same continuous, dependent variable.

APPENDIX 6: Publications arising from this work

Publications

1. **GALLIENNE, A. E.**, DREAU, H. M., MCCARTHY, J., TIMBS, A. T., HAMPSON, J. M., SCHUH, A., OLD, J. M. & HENDERSON, S. J. (2009) Multiplex ligation-dependent probe amplification identification of 17 different beta-globin gene deletions (including four novel mutations) in the UK population. *Hemoglobin,* 33**,** 406-16.

2. **GALLIENNE, A. E.**, IBERSON, N. M., DREAU, H. M., JACKSON, H., BIGNELL, P. A., OLD, J. M., SCHUH, A. & HENDERSON, S. J. (2010) Characterisation of a novel deletion causing beta-thalassemia major in an Afghan family. *Hemoglobin,* 34**,** 110-4.

3. **GALLIENNE, A. E.**, DREAU, H. M., SCHUH, A., OLD, J. M. & HENDERSON, S. (2011) Ten novel mutations in the erythroid transcription factor *KLF1* gene associated with increased fetal hemoglobin levels in adults. *Haematologica,* 97**,** 340-3.

Abstracts

1. **GALLIENNE, A. E.**, HAMPSON, J., BIGNELL, P., OLD, J. M., SCHUH, A., & HENDERSON, S. J. (2009) Application of MLPA and gap-PCR mapping to characterise a novel beta globin gene cluster deletion mutation in a family of Afghan ancestry. *British Journal of Haematology*, 145, Abstract #62

Prize Winner

2. **GALLIENNE, A. E.**, EGLINTON-HARRIS, J., WELLS, R., OLD, J. M., SCHUH, A., & HENDERSON, S. J (2010) Indian newborns have higher adult haemoglobin levels than newborns from other ethnic groups. *The 17th Hemoglobin Switching Meeting, Oxford,Abstract* #28

3. **GALLIENNE, A. E.**, DREAU, H. M., SCHUH, A., HATTON, C., OLD, J. M. & HENDERSON, S. (2011) Nine Novel Mutations in the Erythroid Transcription Factor KLF1 Gene Associated with Increased Fetal Hemoglobin Levels in Adults. *Blood* (ASH Annual Meeting Abstracts), Nov 2011; 118:3191

Oral Presentations

1. **GALLIENNE, A. E.**, HENDERSON, S. J., MCCARTHY, J., MOLYNEUX, A. T., HAYWOOD, A., & OLD, J. M. (2008) Multiple Ligase-dependent Probe Amplification (MLPA) for the detection and pre-natal diagnosis of β globin gene cluster deletions in the UK population. *British Society for Haematology,* 48th Annual Scientific Meeting, April 7-9; 141: 183(Suppl1)

2. **GALLIENNE, A. E.**, EGLINTON-HARRIS, J., WELLS, R., FARRAR, L., SCHUH, A., OLD, J. M., & HENDERSON, S. J. (2011) Haemoglobin switching occurs earlier in Asian-Indian newborns than newborns from other ethnic groups. *12th International Conference on Thalassaemia and The Haemoglobinopathies;* Antaly, Turkey: May 11-14; OP-071