Genetic heterogeneity of *KLF1* deficiency and the pleiotropy of haemoglobin phenotypes

Laura Grech

September 2018

Submitted to the Faculty of Medicine and Surgery, University of Malta in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

Supervisor: Prof Alex E. Felice M.D. PhD

To my daughter Julia

"*Life is not easy for any of us. But what of that? We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained*" Marie Curie

Acknowledgements

My heartfelt thanks go out to all those who have supported me throughout these past 4 years. I would like to thank my supervisor Prof. Alex E. Felice MD PhD for trusting me this project and believing in my competence throughout the practical and theoretical period and for giving me the opportunity to work at the Laboratory of Molecular Genetics at the Department of Biomedical Sciences, University of Malta. My sincere gratitude goes to Prof Joseph Borg PhD, who has provided me with constant help and support throughout my PhD studies. I would like to thank Prof Sjaak Philpsen at the Department of Cell Biology, Erasmus MC, Rotterdam, Prof Marieke von Lindern, Dr. Emile van den Akker and Steven Heshusius PhD student at the Department of Haematopoiesis, Sanquin Research, Amsterdam. Erasmus MC and Sanquin helped with all external resources to conduct RNA-sequencing and ATAC sequencing.

Special mention goes to Prof Anthony Fenech PhD and Dr. Robert Formosa PhD for their help with cell culture and promoter assays. I wish to express my sincere gratitude to Ms. Ruth Galdies B.Pharm., M.Sc, Ms. Wilma Cassar B.Pharm and Ms. Jeanesse Scerri M.Sc, Ms. Jessica Debattista M.SC and Mr. Jeremy Cutajar for their constant help in the laboratory practice and technical advice in the laboratory. I would also like to thank Dr. Chris Saliba PhD, Dr. Melissa Formosa PhD, Dr. Nikoli Pace MD, PhD and Dr. Jean Paul Ebejer PhD for their help in laboratory practice. I am also very grateful to Ms. Mary Rose Caruana and Ms. Patrica Sammut for her help in patient management. I also direct my most heartfelt thanks towards all the families that participated in this research.

My sincere thanks go to my husband Mark for his patience and support. I am forever indebted to my parents for their encouragement and motivation.

This work was partially supported by the Dean Funds, Faculty of Medicine and Surgery.

Abstract

Kruppel like factor 1 (KLF1) also known as the 'Master regulator of erythropoiesis' is located on the short arm of chromosome 19. *KLF1* activates a diverse set of genes that have an important role in the regulation of key pathways such as erythropoiesis, cell membrane and cytoskeleton, apoptosis, heme synthesis & transport, cell cycling, iron procurement and globin chain production. To date over 65 molecular variants have been recorded. Their haematological phenotypes range from the clinically benign In(Lu) type of Lu(a-b-) blood group, variability in the HbA_2 levels, congenital dyserythropoietic anaemia (CDA) and in most extreme cases hydrops foetalis secondary to profound anaemia. In 2010 sequencing revealed a nonsense mutation in *KLF1* in a large Maltese family with hereditary persistence of foetal haemoglobin (HPFH). The p.Lys288Ter mutation was found to ablate the DNA binding domain of the key erythroid transcription factor.

Through the Thalassemia and Molecular Genetics clinic, we identified 5 other families with the same *KLF1* p.Lys288Ter truncation mutation but with normal or slightly increased HbF levels and borderline HbA_2. The phenotypic variability of HbF was not readily explained by the co-inheritance of know HbF- modulating variants in the *HBB*, *HBS1L-MYB* and *BCL11A* loci. To investigate the phenotypic variability of *KLF1* deficiency, we took advantage of primary erythroid cultures of individuals carrying the *KLF1* p.Lys288Ter mutation with variable HbF and unrelated Maltese individuals wildtype for this mutation. A combined analysis of whole genome sequencing (WGS), gene expression (RNA-seq), chromatin accessibility assays (ATAC-seq) and promoter activity tests was performed to explore the molecular basis of the heterogeneity in HbF levels displayed by *KLF1* p.Lys288Ter individuals.

WGS identified two *KLF1* promoter mutations, one in *cis* (-1133C>A) and one in *trans* (-251C>G) to the truncation mutation. A significant decrease in the luciferase expression was noted in constructs with the -1133C>A and -251C>G mutation when compared to construct with the wildtype promoter. After carrying out EMSA, for the -1133C>A mutation a super-shift in the wildtype type oligo was noticed. This super shift was absent in the mutant oligo. Using regulomeDB it was shown that YY1 transcription factor binds to the wildtype sequence, but it does not bind to the mutant sequence. After combining, the results from the promoter assays, the RNA-sequencing and chromatin accessibility studies, the best explanation from HbF variability in *KLF1* deficient subjects is the allelic variation of the wildtype *KLF1*.

A retrospective search in the Malta biobank for undiagnosed subjects with borderline HbA_2 and β-thalassaemia heterozygotes was carried out. Four-hundred and twenty-six subjects were collected together with 200 controls. One hundred and thirty-two subjects were pseudo-thalassemia while 294 were β-thalassemia heterozygotes. More than 55% of pseudo-thalassemia, had a mutation in the KLF1 gene while KLF1 mutations were present in 40% of the β-thalassemia mutation. Most of the *KLF1* variants were closer to the 5' end of the gene spanning from the promoter region to approximately 60% of exon 2. Three SNPs were found at the 3' UTR. The four promoter mutations showed a statistically significant difference in the luciferase expression of constructs containing the mutant KLF1 promoter when compared to the wildtype.

Furthermore, from a clinical point of view, the data presented, highlighted the importance of *KLF1* sequencing in patients with haematological features resembling β-thalassaemia for differential diagnosis of microcytosis and for genetic counselling.

Abbreviations

A	ACH	Active Chromatin Hub
	ATAC	Chromatin accessibility assay
	AD5	Adenovirus type 5
B	BCL11A	B-cell lymphoma/leukaemia 11A protein
	Bp	Base pair
	BSA	Bovine serum albumin
C	CtBP	Carboxy-terminal binding protein
	ChIP-seq	Chromatin ImmunoPrecipitation sequencing
	CDA	Congenital Dyserythropietic Anaemia
	CD	Cluster of Differentiation
	CNSHA	Chronic non-spherocytic haemolytic anaemias
	CO_2	Carbon dioxide
	CO	Carbon oxide
	CBC	Complete Blood Count
	Ct	Cycles to Threshold
	CaCl2	Calcium Chloride
	CML	Chronic myeloid leukaemia
D	dsDNA	Double stranded DNA
	DNA	Deoxyribonucleic acid
	Dex	Dexamethasone
	DMEM	Dulbeccos's Modified Eagle's Medium
	DPP	Digital Pulse Processing
	DLR	Dual luciferase reporter
E	EDTA	Ethylenediaminetetraacetic
	Epo	Erythropoietin
	EV	Empty Vector
	EMSA	Electrophoretic mobility shift assay
F	FBS	Foetal bovine serum
G	gDNA	Genomic DNA
H	HPFH	Hereditary persistence of foetal haemoglobin
	HbF	Foetal haemoglobin
	HbE	Haemoglobin E
	Hb	Haemoglobin
	HbA	Adult haemoglobin
	HS	Hypersensitivity site
	HbC	Haemoglobin C
	HMG	High mobility group
	HPLC	High Performance liquid chromatography
	Hb	Haemoglobin
	HRM	High Resolution Melting
	HGMD	Human Gene Mutation Database
	HEPs	Human Erythroid Progenitor cells
	HEK293T	Human embryonic Kidney cells
	HUDEP-2	Human Umbilical-cord blood derived progenitor-2

I	IVS	Intervening sequence
	(In)Lu	Inhibitor of Lutheran antigen
J	JMML	Juvenile myelomonocytic leukaemia
K	KLF	Kruppel-like factors
	kb	Kilo bases
L	LCR	Locus Control Region
	LB	Luria Broth
M	MEL	Murine erytroleukemia
	MCH	Mean Cell Haemoglobin
	MCV	Mean Cell Volume
	MYB	v-myb myeloblastosis viral oncogene
	MCS	Multiple cloning region
	MgCl2	Magnesium Chloride
N	NSHA	Non-spherocytic haemolytic anaemia
	NO	Nitric Oxide
	NaCl	Sodium Chloride
	NTC	Non-template Control
	NaOH	Sodium Hydroxide
	NGS	Next Generation Sequencing
O	O_2	Oxygen
	OMIM	Online Mendelian Inheritance in Man
P	PBS	Phosphate buffer saline
	PCR	Polymerase chain reaction
	PCMV	Promoter of cytomegalovirus
	P/S	Penicillin/streptomycin
	PLB	Passive lysis buffer
R	RNAi	RNA interference
	RT-qPCR	Real-time quantitative technique
	Rluc	Renilla luciferase
	RFLP	Restriction Fragment Length Polymorphism
	RPMI	Roswell Park Memorial Institute medium
S	SCD	Sickle cell disease
	SDS	Sodium Dodecyl Sulfate
	ssDNA	Single stranded DNA
	SCF	Stem cell factor
T	TAD	Transactivation domain
	Tfrc	Transferrin receptor
	TE	Tris-Ethylenediminetetraacetic
U	UTR	Untranslated Region
	UIBC	Unsaturated iron binding capacity
	UTR	Untranslated Region
V	V	Volts
W	WGS	Whole Genome Sequencing
Z	ZnPP	Zinc protoporphyrin
	4-OHT	4-hydroxytamoxifen
	δ	Delta
	β	Beta

	α	Alpha
	ε	Epsilon
	γ	Gamma

Table of Contents

List of Figures

List of Tables

1.1 Kruppel Like factors (KLFs)

Kruppel-like factors (KLFs) are zinc finger-containing transcription factors whose role is to control a diversity of biological processes including proliferation, differentiation, growth, development, survival, and responses to external stress. The KLF family is made up of 17 members homologous to the Drosophila melanogaster protein (Schuh et al., 1986) and they also share homology with the Sp1 transcription factor (Kadonaga et al., 1987). The KLF proteins can be divided into three distinct groups on the basis of their functional characteristics. KLF3, KLF8 and KLF12 act as transcription repressors through their interaction with the carboxy-terminal binding protein (CtBP). KLF1, KLF2, KLF4, KLF5, KLF6 and KLF7 are transcription activators while KLF9, KLF10, KLF13, KLF14 and KLF16, through their interaction with the transcriptional corepressor Sin3A, have repressor activity (McConnell & Yang, 2010). As shown in figure 1.1 numerous human diseases, including metabolic disorders, cardiovascular diseases, haematological diseases and cancer have been associated with mutations in the KLFs.

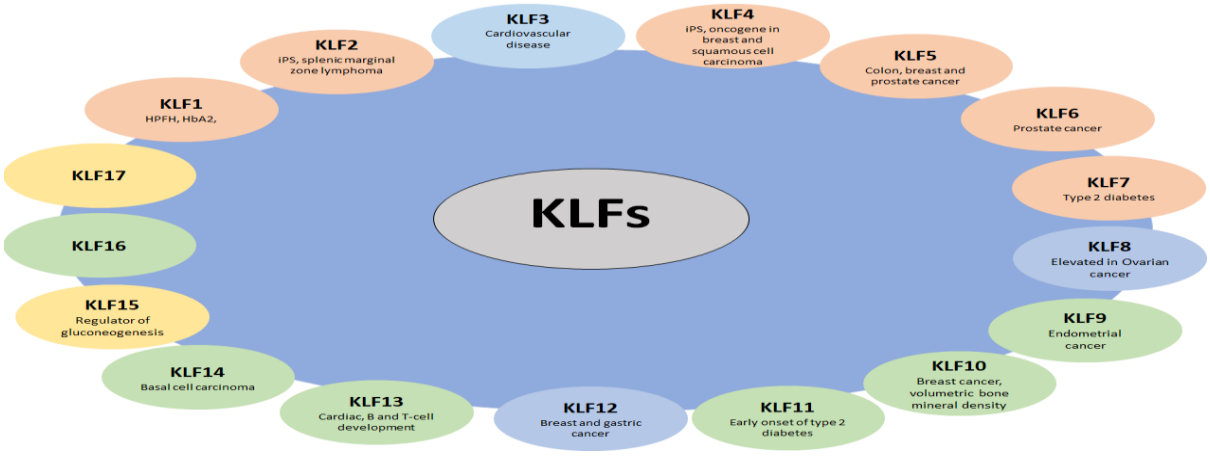

Figure 1.1: The pathobiology of KLF1 family proteins. The KLF1 proteins are grouped according to common structural and functional domains. KLFs1, 2, 4, 5, 6 and 7 are able to bind acetylases, KLFs 3, 8 and 12 have a CtBP-binding site, KLFs 9, 10, 11, 13, 14 and 16 have a Sin3A binding site. KLFs 15 and 17 are more distantly related based on phylogenetic analysis.

Our main focus is on Kruppel like factor 1 (KLF1) also known as the 'Master Regulator of erythropoiesis' (Borg et al., 2010). The *KLF1* gene spans approximately 3 kilo bases (kb) of genomic DNA and is located on the short arm of chromosome 19. The KLF1 gene contains 3 exons encoding a 363 amino-acid protein (Bieker, 1996). At the carboxyl terminus of the protein, one finds three highly conserved classical Cys2/His2 zinc fingers. The 1st and 2nd zinc finger contains 23 residues while the 3rd finger contains 21 residues. These zinc fingers are connected together by the characteristic Kruppel-link which consists of seven amino acid spacer TGEKP(Y/F)X. The role of the link is to enable KLF1 to bind to GC and CACCC DNA boxes. As shown in figure 1.2 each zinc finger chelates a single zinc ion-co-ordinated by two cysteine and two histidine residues (Dang et al., 2000).

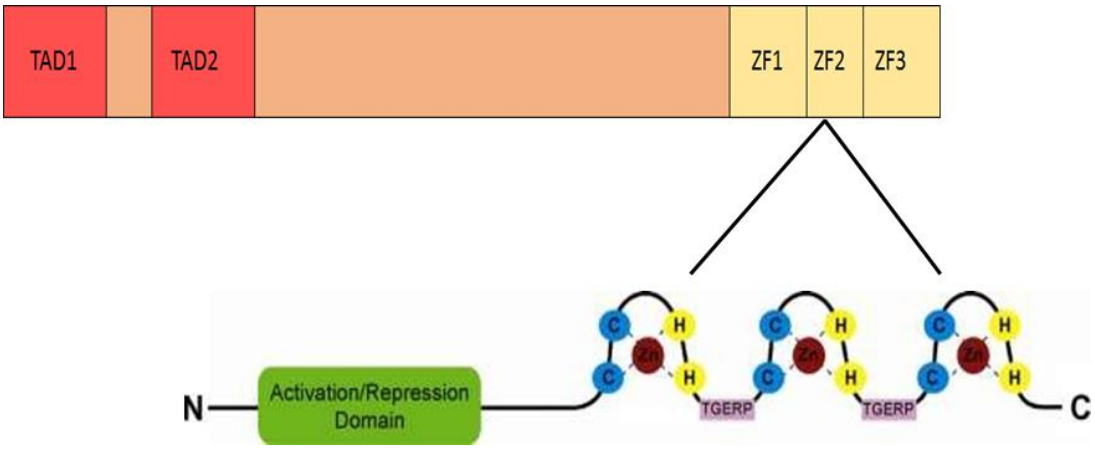

Figure 1.2: Model of the KLF1 gene. The KLF1 protein contains two N-terminal transactivation domains (TAD1 and TAD2) which are required for KLF1 to act as a transcriptional activator. The 3 zinc fingers (ZF1, ZF2, and ZF3) are linked together by a"TGERP"-like motif. Adapted from Pearson et al., (2008).

At the N terminal, one finds the transactivation domain (TAD). It was shown by Chen & Bieker (1996) that the first 40 residues of the TAD recruit cofactors required for β-globin activation while residues 50 to 90 are essential for β-globin gene activation. Based on sequence analysis, Mas et al., (2011) divided the KLF TAD into two functional subdomains; mainly TAD1 (1 to 40 residues) and TAD2 (51 to 90 residues). They also demonstrated that TAD2 binds to the amino-terminal PH domain of the Tfb1/p62 subunit of the TFIIH which is a rate-limiting factor during the elongation phase of β-globin gene expression (Bird et al., 2004). The TAD2 also

binds to four domains (TAZ1/CH1, KIX, the TAZ2/CH3, and IBID) of CBP/p300 which are essential for p53-dependent activation (Mas et al., 2011).

1.2 Biological Roles of KLF1

KLF1 acts as a global regulator of erythroid production and integrity. As shown in figure 1.3 KLF1 activates a diverse set of genes that have an important role in the regulation of key pathways such as erythropoiesis, cell membrane and cytoskeleton, apoptosis, heme synthesis & transport, cell cycling, iron procurement and globin chain production (Tallak & Perkins, 2010).

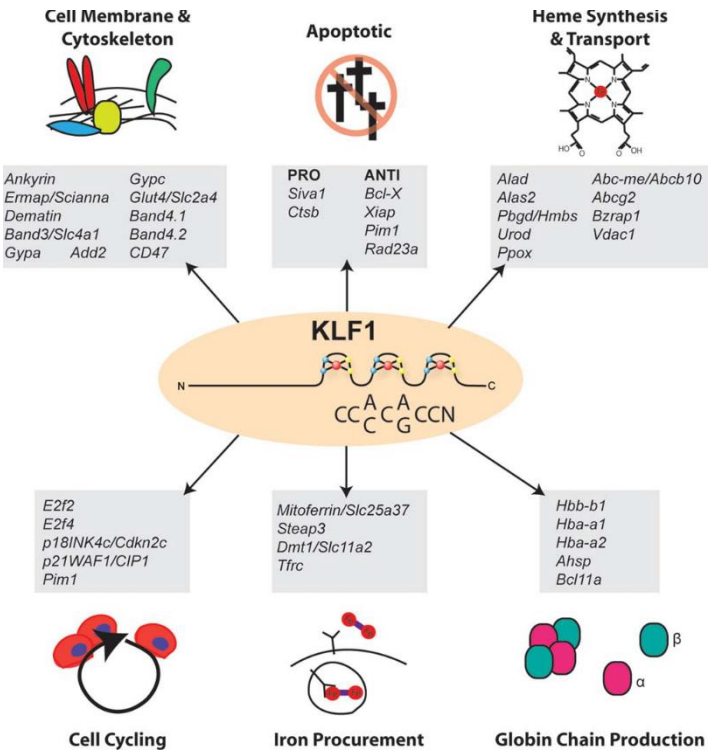

Figure 1.3: Biological Roles of KLF1. KLF1 binds to a subset of CACC-box motifs in *vivo* as refined by ChiP-seq analysis. As shown when KLF1 binds to the CACC-box motifs it activates a diverse set of genes involved in key pathways. Adapted from Tallak & Perkins (2010).

1.2.1 Role of KLF1 in Erythropoiesis

Erythropoiesis is a dynamic multistep process involving the differentiation of early erythroid progenitors to enucleated red blood cells. A large number of transcription factors such as GATA-1, GATA-2, Lmo2, FOG, c-myb, TAL-1, BCL11A, RUNX-1, PU.1 together with KLF1 are involved in the formation of hematopoietic cell lineages. KLF1 plays an important role both during early as well as during terminal differentiation (Gnanapragasam & Bieker, 2017). The role of KLF1 in erythropoiesis has been studied extensively in mice. It was shown that mice heterozygous for KLF1 appear completely healthy but in absence of KLF1 the foetuses develop a fatal anaemia and die at around E14. In absence of KLF1 the embryonic globin gene expression was not influenced, but when embryos switched to definitive erythropoiesis in the foetal liver, $KLF^{-/-}$ mice rapidly developed anaemia because of deficit in β-globin expression (Nuez et al., 1995; Perkins et al., 1995). Analysis of the foetal liver of $KLF^{-/-}$ mice showed that erythroid cells had an abnormal morphology and most of them retained the nucleus. The $KLF^{-/-}$ could not be rescued by the expression of exogenous human γ-globin (Perkins et al., 2000). This indicated that β-globin genes are not the only genes regulated by KLF1, in fact genome wide expression analysis of $KLF^{-/-}$ erythroid cells, showed that KLF1 also controls genes associated with cell cycle and erythroid cell membrane (Tallak et al., 2007).

1.2.2 Role of KLF in regulation of the Erythroid membrane and cytoskeleton

Circulating primitive erythrocytes from $KLF1^{-/-}$ embryos showed a ruffled and wrinkled phenotype at E11.5 and E12.5, due to a loss of Dematin/Epb4.9 expression (Drissen et al., 2005; Hodge et al., 2006). Dematin is an actin binding protein that helps in the erythroid

membrane structural stability, in-fact full length disruption of dematin led to severe anaemia and altered erythroid membrane stability and morphology in mice (Lu et al., 2016). ChIP-seq in primary erythroid cells confirmed that apart from dematin, Ankyrin, Ermap, glycophorin A, glycophorin C and ankyrin 1 are also KLF1 target genes. These play an important role in the maintenance and integrity of the erythroid cell membrane and cytoskeleton (figure 1.3) (Nilson et al., 2006; Ye et al., 2000; Tallak et al., 2010).

1.2.3 Role of KLF1 in Cell Cycle

The role of KLF1 in the regulation of the cell cycle was first suggested by Coghill et al., (2001). Erythroid progenitors from KLF1$^{-/-}$ embryos were immortalized and retrovirally transduced with a KLF1 expression construct to produce cells containing a synthetic KLF1-ERTM fusion protein that was able to be activated by the addition of 4-hydroxytamoxifen (4-OHT). These cells are known as K1-ER cells. Reactivation of these K1-ER cells led to a cessation in proliferation and an engagement in differentiation. This indicated in order for KLF1 to promote differentiation it acted as a negative regulator in the cell. Using murine erythroleukemia (MEL) cells it was shown that during the differentiation process, KLF1 regulates the changes in the cell cycle underlying erythroid maturation by inducing the expression of p21WAF1/CIP1 independent of p53. KLF1 activates the p21 by binding directly to the GC rich region in the p21 proximal promoter and also by occupying a phylogenetically conserved region containing CACCC core motifs located downstream of the p21 TATA box (Siatecka et al., 2010). It was also demonstrated that KLF1 activates the expression of E2f2 and E2f4 genes, which act to drive proliferation (Pilon et al., 2008; Tallak et al., 2009). By using imaging flow cytometry Gnanapragasam et al., (2016), suggested that during erythropoiesis in KLF1$^{-/-}$ foetal livers, there is a 2-stage block in the cell cycle. The first block occurs at the progenitor stage, resulting

in a premature cell cycle exit that leads to impaired entry into terminal differentiation. The second occurs at the orthochromatic erythroblast stage, where cells showed decreased cell cycle exit, with significant number of cells still in the S phase. This results in late-stage erythroblasts that are not able to enucleate and can even become binucleated.

1.2.4 Role of KLF1 in heme synthesis and iron procurement

KLF1 is involved in the co-ordination of heme synthesis and iron procurement. It regulates the expression of a large number of the enzymes that are required for the conversion of succinyl-CoA and glycine into heme. It also controls the expression of the transferrin receptor (Tfrc) and other proteins that are required for iron uptake and processing by erythroid cells (Drissen et al., 2005; Hodge et al., 2006; Tallack et al., 2010). The expression of two important mitochondrial membrane transporters mainly Abcb10 and Abcg2 are also regulated by KLF1. Klf1 activates the expression of these two genes resulting in the export of iron bound-heme from the mitochondria to the cytoplasm. Once in the cytoplasm of the erythroid cell, heme is complexed with the globin protein subunits to produce haemoglobin (Tallack & Perkins, 2010).

1.2.5 Role of KLF1 in β-globin gene expression

The role of KLF1 in β-globin gene expression was first shown by Nuez et al., (1995) & Perkins et al., (1995). They showed that disruption of *KLF1* gene in mice led to significant loss of adult-type β-globin expression giving rise to anaemia and resulted in embryonic lethality of homozygous null animals at E14-E16. These mice exhibited molecular and morphological features typical of severe β-thalassaemia with significantly elevated α/β ratio altough the

embryonic β-like globin genes βH1 and Epsilon were not affected by the loss of KLF1. This showed that KLF1 is required for high-level expression of the adult β-globin. KLF1 interacts with the mammalian SWI/SNF chromatin remodelling complex also known as E-RC1. The interaction of KLF1 to the E-RC1 generates a DNase I hypersensitive, transcriptionally active, chromatinized β-globin promoter template in vitro (Armstrong et al., 1998). It was proposed by Drissen et al., (2004) that KLF1 facilitates the formation of the Active Chromatin Hub (ACH), bringing the β-globin gene in contact with the Locus Control Region (LCR) (figure 1.4).

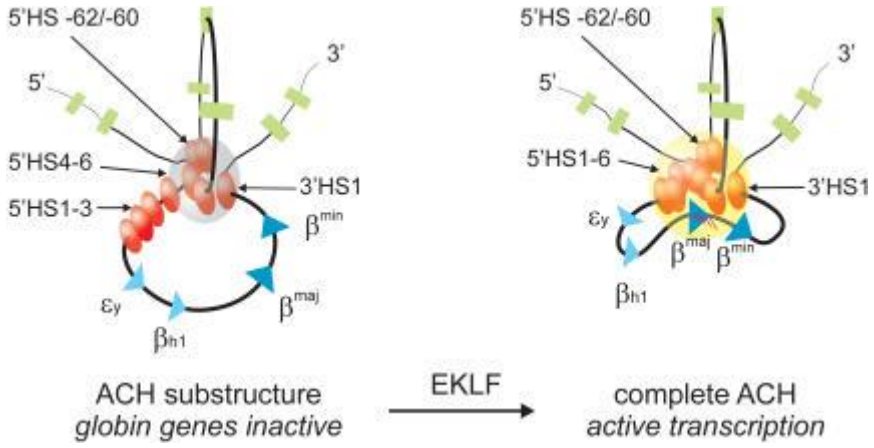

Figure 1.4: Formation of active ACH. This figure shows a two-dimensional representation of the ACH (grey sphere) which is a nuclear compartment dedicated to RNA polymerase III transcription. It is formed by cis-regulatory elements of the β-globin locus. In erythroid cells, a substructure of the ACH (yellow sphere) consisting of the 5'HS-62/60, 3'HS1 and HS at the 5' of the LCR is formed independently of KLF1. Progression of the substructure into a fully function ACH which includes the HS at the 3' side of the LCR and the active β-globin gene, only occurs in the presence of KLF1. Adapted Drissen et al., (2004).

1.3 ROLE of KLF1 in diseases

As already illustrated above, *KLF1* plays an important role in many physiological processes and therefore the downstream effects of *KLF1* mutations can give rise to a wide range of haematological conditions. After the finding by Borg et al., (2010) that KLF1 mutations can give rise to hereditary persistence of foetal haemoglobin (HPFH), over 65 molecular variants in KLF1 have been reported. According to Perkins et al., (2016) these variants can be grouped

into four functional classes. The first class consists of variants with no or minor functional consequences. These are usually missense mutations located outside the DNA binding domain. The 2nd class consists of reduced function hypomorphic variants which are usually missense variants or small in-frame deletions interfering with the normal function of KLF1. The 3rd class contains truncating loss-of-function variants which are usually stop codons or frame-shift variants resulting in truncation of the KLF1 protein. These variants affect one allele and cause haploinsufficiency for KLF1. Class 4 are dominant variants and only one case was reported so far. Various haematological conditions are associated with these mutations as shown below.

1.3.1 Rare IN(LU) blood group

The first reported mutation in the *KLF1* gene was associated with the In(Lu) phenotype. The Lutheral blood antigens are red cell membrane glycoproteins encoded by the *BCAM* gene. In rare occasions, the Lutheran antigens can be either weakly expressed or absent on the red cells giving rise to Lu(a-b-) phenotype. The Lu(a-b-) phenotype in most cases is due to recessive BCAM mutations, X-linked mutations or In(Lu) mutations. Singleton et al., (2008) showed that erythroblasts from persons with the Lu(a-b-) had reduced transcription levels in many erythroid specific genes when compared to normal erythroblasts. This suggested that this phenotype can be due to an abnormality in a transcription factor. Twenty one out of 24 subjects with In(Lu) phenotype had a mutation in the KLF1 gene. In these subjects one KLF1 allele was functional while the other was defective. These subjects had no reported pathology, indicating that one functional KLF1 allele was adequate for erythropoiesis.

Helias et al., (2013) found another seven different mutations in KLF1 gene which were responsible for the In(Lu) phenotype. Most of the mutations were loss of function mutations but three were missense mutations in the DNA binding domain resulting in loss of

transactivation capacity of KLF1 in *vitro*. In contrast to subjects from Singelton et al., (2008) study, these subjects with In(Lu) phenotype had increased foetal haemoglobin (HbF) and HbA_2 levels. In Japan, 120 out of 481,322 blood donors had the In(Lu) phenotype. Out of these 120 donors, 100 donors also had a mutation in the KLF1 gene. Thirty-four mutations were identified, out of which 21 were novel (Kawai et al., 2017). It was shown that other blood group antigens are also direct targets of KLF1, in-fact in KLF1 null subjects, there was a reduced expression of Kell, Duffy, Kidd, RhD, RhAG, Scianna and LW blood group antigens (Hodge et al., 2006; Tallack et al., 2012; Magor et al., 2015).

1.3.2 Hereditary Persistence of Foetal Haemoglobin

HPFH is characterized by persistence of HbF production in adulthood. In a large Maltese family, ten out of twenty-seven family members exhibited HPFH with HbF levels ranging from 3.3% to 19.5%. The HPFH individuals had a truncated mutation (p.K288X) in the KLF1 gene. Gene expression profiles conducted on cultured primary erythroid cells from subjects with HPFH showed that these had decreased expression of the foetal haemoglobin repressor *BCL11A* and upregulation of the gamma globin gene. Knockdown of *KLF1* in cultured erythroid progenitor cells gave similar gene expression results. On restoration of KLF1 activity in these cells, the RNA and protein *BCL11A* levels were upregulated. Thus, this shows that KLF1 has a dual role (figure 1.5) in the regulation of foetal-to-adult globin gene switching. First, it acts directly on the *HBB* locus as a preferential activator of the *HBB* gene and secondly it acts indirectly by activating the expression of BCL11A which, in turns, represses the gamma globin genes (Wijgerde, 1996; Borg 2010; Zhou 2010).

Satta et al., (2011) reported two Sardinian brothers with normochromic, normocytic anaemia, elevated HbF levels (one 30.9% and the other 22.1%) and elevated red blood cell zinc

protoporphyrin. Both parents and the third brother had a normal haematological phenotype, normal HbF levels and normal red blood cell zinc protoporphyrin. The brothers with high HbF levels had a nonsense mutation (p.S270X) in the KLF1 at exon 2 and another missense mutation (p.K332Q) at exon 3. They inherited the missense mutation from their mother and the nonsense mutation from their father. It was shown that the nonsense mutation completely ablates the zinc finger domain and reduces the ability of the KLF1 to interact with DNA. The other missense mutation further reduces the KLF1 function.

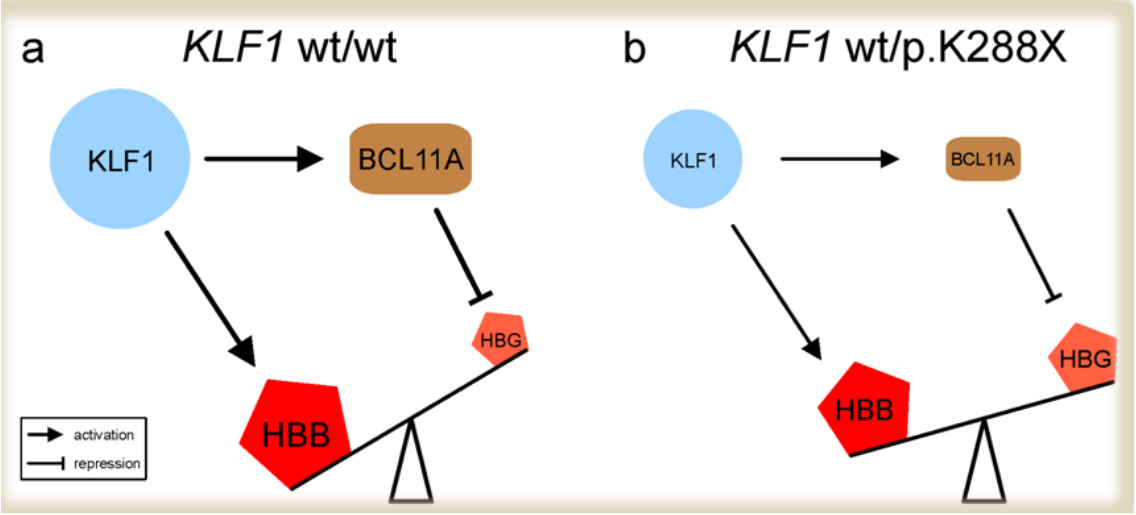

Figure 1.5: Model for the regulation of β-like globin expression by KLF1 in adults. Figure a show that KLF1 preferentially activates the HBB gene and the BCL11A gene, while the BCL11A protein silences the HBG1/HBG2 (HBG) genes in normal adults. Figure b shows that in patients with KLF1 p.K288X mutation KLF1 activity is reduced. This decreases expression of BCL11A and the diminished amount of BCL11A protein alleviates repression of the HBG1/HBG2 genes. Adapted from Borg et al., (2010).

A boy of Greek descent presented with pallor and anaemia, with a haemoglobin of 8.0g/dL, low mean cell volume (MCV) and mean corpuscular haemoglobin (MCH), an HbF value of 83.4% and HbA_2 of 4.3%. The patient was splenectomised at the age of 4 and one-year post-splenectomy his HbF went up to 97.5% and his HbA_2 went down to 2.5%. His parents were found to be both β-thalassaemia heterozygotes, the father had an HbF of 4.3% and HbA_2 of 6.7% while the mother had an HbF of 7.1% and HbA_2 of 4.3%. Genetic studies on the β-globin gene showed that the proband was compound heterozygote for codon 39 and also had a nonsense mutation in the proximal CACC box. These two mutations on their own could not

explain the nontransfusion dependent thalassaemia phenotype with high foetal haemoglobin. Sequencing of the KLF1 gene revealed two mutations. One a non-synonymous variant in exon 2 and the other a missense mutation in exon 3 (Ghallagher et al., 2016).

Hariharan et al., (2017) reported a case of a six-year-old un-transfused Indian girl who presented with persistent anaemia, recurrent jaundice and hepatoslenomegaly, HbA_2 of 1.6% and HbF of 97.9%. Both parents were heterozygote for the β-thalassaemia mutation codon 15 (G→A) mutation and the daughter was also found to be homozygous for this β-mutation. Since the mother and daughter had high HbF (mother HbF 8.6%) genetic testing on Xmn1, KLF1 and BCL11A were performed. A novel mutation in the *KLF1* coding region was discovered. This mutation in codon 211 A→G (c.632 A>G) changes the amino acid from glutamine to arginine and by using bioinformatic software this mutation was predicted to have a damaging effect on the protein structure. Weatherall et al., (1975) described a British family with HPFH and the -198T>C mutation in the promoter of the γ globin gene was responsible for this type of HPFH also known as the British-type HPFH. Through a combination of *in vitro* and *in vivo* studies, it was shown that this mutation creates *de novo* binding site for KLF1 resulting in significant increase in HbF levels (Wienert et al., 2017).

1.3.3 Borderline HbA_2

In normal adults, the level of minor Hb (HbA_2; $\alpha_2\delta_2$) compromises about 2.5% of the total haemoglobin. Increased HbA_2 is one of the parameters used for the identification of beta thalassaemia heterozygotes subjects. Although the exact mechanisms of the elevation of HbA_2 in beta-thalassaemia heterozygotes is not yet known, evidence shows that several factors are involved. In beta-thalassaemia heterozygotes there was an increased availability of α-globin chains due to β-globin deficit. This results in a post-transcriptional mechanism that increases HbA_2 (Steinberg et al., 2015). If there is inefficient binding of the transcription factors to the

altered or deleted β-promoters, more factors would be free to trigger the δ-promoter both on the cis and trans-locus. This increases the output of both the δ-globin genes (Menzel et al., 2013).

Persue et al., (2011) carried a study on a cohort of 145 Italian patients with borderline HbA_2 between 3.3% and 4.1% who had no mutations in the beta globin and the alpha globin. These subjects were tested for *KLF1* mutations. Thirty-six percent (36%) of these subjects had a mutation in the *KLF1* gene. Most of these subjects had the non-sense pSer270X mutation in the N-terminal domain. The other mutations were deletions, frameshift mutations, and missense mutations. In these subjects the HbF level was quite variable, ranging from 0.2% and 5.8% while MCV and MCH were normal in all subjects. In the remaining 93 subjects' with borderline HbA_2, complete sequencing of the *KLF1* gene did not reveal any mutation. Persue et al., (2011) showed that *KLF1* variants are not only linked to extremely rare cases of red cell disorders but can also give rise to milder phenotypes. Other studies showed that the co-inheritance of haemoglobinopathies and KLF1 variants is common in Southern China (Liu et al., 2015).

In 2014, Liu et al., studied the incidence of KLF1 mutations in 2 different Chinese populations; 3839 subjects from a thalassaemic endemic region in south China and 1190 subjects from a non-thalassaemic endemic region in north China. A cohort of 79 subjects with borderline HbA_2 (3.3% and 4.1%) and/or elevated HbF (>1.5%) were also recruited from a large random Guangxi population. In Southern China, 48 out of the 3839 subjects had a mutation in the *KLF1* gene giving a heterozygote frequency of 1.25% in this population. In contrast, in Northern China only 1 subject was heterozygote for a *KLF1* mutation giving a 0.08% frequency. From the 79 subjects with borderline HbA_2 and/or elevated HbF, 16 subjects had a mutation in the *KLF1* gene. Out of these 64 subjects, 41 were *KLF1* heterozygotes, 11 were *KLF1*

heterozygotes co-inherited with β-thalassaemia heterozygotes, while 12 were *KLF1* heterozygotes co-inherited with β-thalassaemia homozygote or compound heterozygote. Ten of the mutations were missense mutations in exon 1, 2 and 3; 1 was a frameshift mutation in exon 2, one a splicing mutation in intron 2, 5 were synonymous mutations in exon 2 and 3 and five were conserved mutations in the 3' Untranslated region (UTR).

Lou et al., (2014) also recruited 165 adults from China with borderline HbA_2 levels (3.3% and 4.0%). Fifteen of these subjects had decreased red cell indices while the other 150 subjects had normal red cell indices. Out of these 165 subjects, 20 were heterozygote for a *KLF1* mutation. Eight types of mutations were identified, two were shift mutations in exon 2 and exon 3 and six were missense mutations also in exon 2 and exon 3. No KLF1 mutations were found in the other 80 subjects with normal HbA_2. In 2015, a 30-year old Japanese male presented with non-iron deficiency anaemia, mild microcytosis and mild erythrocytosis. The serum iron, unsaturated iron binding capacity (UIBC) and ferritin were in normal range. The HbA_2 was 4.1% and HbF level was 1.3%. No mutations in the β-globin gene were observed but the patient was found to be heterozygote for the p.C316Y KLF1 mutation in the 2^{nd} zinc finger (Nitta et al., 2015)

Paglietti et al., (2016) carried out a study on a cohort of subjects with borderline HbA_2 (3.3% to 4.1%) in Cagliari. Two hundred and twenty-three subjects (223) were collected over a period of 3 years and six months. These were separated into 3 groups (i) Group A subjects showing borderline HbA_2 levels associated with reduced MCH and/or MCH (ii) Group B consisted of subjects showing borderline HbA_2 levels associated with normal MCV and MCH while (iii) Group C consisted of subjects with borderline HbA_2 levels and normal red cell indices without a β-thalassaemia heterozygous partner. Molecular analysis for β, δ, α and KLF1 genes was performed on these subjects.

In Group A, a β-globin gene mutation was present in 6 out of 11 subjects with significant microcytosis. Two of these subjects with HbA_2 <4.0% were also carriers for a δ mutation. Three out of the 5 subjects who were negative for a pathological β-globin gene mutation were found to be carriers for the non-sense pSer270X KLF1 mutation. Two of these subjects also had α-thalassaemia deletions and higher HbA_2 level than typically found in α-thalassaemia carriers. The higher HbA_2 levels in these subjects was justified by the presence of the KLF1 mutation. In Group B, out of 26 subjects only 1 was a carrier for a mutation in β-globin gene and out of the remaining 25 subjects only 3 were positive for the p.Ser270X mutation. In group C, among the 28 subjects with borderline HbA_2 and normal red cell indices, 13 had a mutation in the KLF1 gene. Eleven were heterozygote for the p.Ser270X mutation, 1 was heterozygote for the p.R319Efs*34 frameshift mutation and 1 heterozygote for the T280-H283 deletion.

As seen in figure 1.6, KLF1 variants do not only affect the HbA_2, HbF and MCV levels in normal individuals and subjects with β-thalassaemia traits but it also has an effect on subjects with α-thalassaemia trait. In a study carried out by Yu et al., (2015), 1468 Chinese subjects with α-thalassaemia were recruited and molecular analysis of the KLF1 gene was carried out. Seventy-two (72) subjects had a mutation in the *KLF1* gene. It was noted that the subjects with KLF1 mutations had lower MCH and MCV levels and higher HbA_2 than the other α-thalassaemia subjects with no KLF1 mutations. In Sardinia, 26 subjects out of 195 α-thalassaemia carriers also had a mutation in the KLF1 gene. Most of these subjects with KLF1 mutations had higher HbA_2 level and lower MCV and MCH than the other α-thalassaemia carriers with no KLF1 mutation (Satta et al., 2017). As seen in figure 1.6 (b) the globin chain balance is only affected in subjects with KLF1 variants co-inherited with β-thalassaemia and not in subjects with KLF1 variants or subjects with KLF1 variants co-inherited with α-thalassaemia (Perkins et al., 2016).

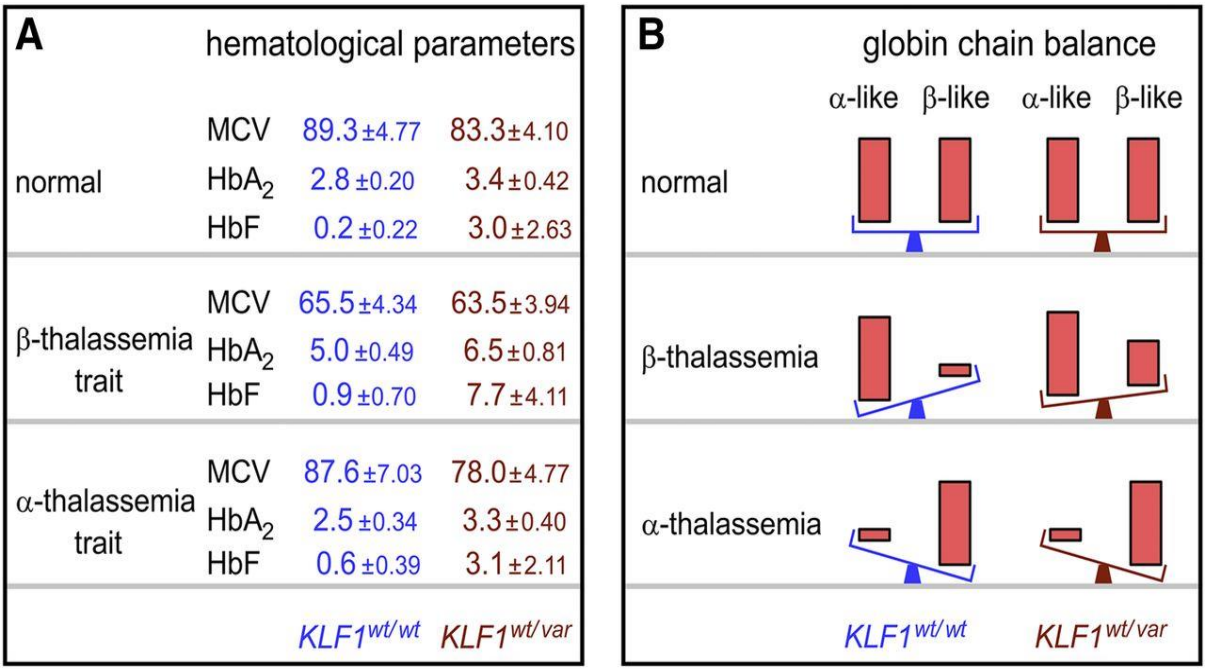

Figure 1.6: Impact of KLF1 variants on clinical severity of hemoglobinopathies. Figure A shows the haematological indices – HbA_2, HbF and MCV in controls, β- and α-thalassaemia trait individuals with or without KLF1 mutations. Figure B shows the effect of KLF1 variants on globin chain balance in healthy controls, β or α-thalassaemia subjects. Adapted from Perkins et al., (2016).

The role of KLF1 variants in subjects with haemoglobin E (Hb E) disorder was also studied by Tepakhan et al., (2016) in South-east Asia. Hb E is caused by a single nucleotide substitution at codon 26 of the β-globin gene. This mutation gives rise to a β^+ thalassaemia phenotype by creating an abnormal splicing within the exon 1 of β-globin gene (Rees et al., 1998). The study was carried out on 953 subjects, 575 homozygotes for Hb E, 278 heterozygotes for Hb E and 100 normal controls. Out of the 575 homozygous Hb E, 48 had a mutation in the *KLF1* gene while out of the 278 subjects' heterozygotes for Hb E, 49 had a mutation in the *KLF1* gene. No KLF1 mutation was detected in the normal controls. It was noted that subjects with *KLF1* mutation had higher HbF levels than other Hb E homozygotes or heterozygotes subjects without a KLF1 mutation.

1.3.4 The role of KLF1 in anaemias

Anaemia is a major public health problem and the major causes include malnutrition particularly due to iron and folate deficiency. However, a considerable percentage of patients with anaemia have inherited red cell defects due to mutations in genes encoding for red cell enzymes, membrane proteins, heme and globins. It was shown by Arnuald et al., (2010) and Viprakasit et al., (2013) that mutations in the KLF1 gene can also give rise to anaemia such as congenital dyserythropoietic anaemia (CDA) and haemolytic anaemia.

1.3.4.1 Congenital dyserythropoietic anaemia

CDA are a group of rare hereditary disorders characterized by ineffective erythropoiesis, late erythroblasts and the development of secondary hemochromatosis. Patients usually present with anaemia, jaundice and splenomegaly. The CDAs are classified into three major types (I to III) and several minor subgroups (IV – VII) according to the morphological abnormalities of erythroblast nuclei observed in the bone marrow (Heimpel et al., 1968; Iolascon et al., 2013). Arnuald et al., (2010) described a missense mutation (E325K) in the KLF1 gene which was responsible for CDA in 5 patients. Some of these patients, together with the red blood cell abnormalities, also had persistent expression of embryonic ζ and ε globin chains, high foetal haemoglobin, intra erythroblastic and intra erythrocytic inclusions and deficiency of erythroid CD44 and aquaporin I. It was shown that this mutation has a dominant-negative effect on the KLF1 transcriptional activity and also abolishes the expression of the adhesion molecule CD44 and the water channel AQP1. This type of CDA was classified as CDA IV.

1.3.4.2 Non-spherocytic haemolytic anaemia (NSHA)

NSHA is a term used to describe inherited anaemias characterized by shortened red blood cell survival, abnormal morphology, erythroid hyperplasia in the marrow and haemolysis. Viprakasit et al., (2014) reported 8 unrelated patients with severe transfusion-dependent neonatal anaemia with red cell abnormalities. These abnormalities ranged from thalassaemia like morphology with hypochromic and microcytic red cells to abnormalities usually associated with chronic non-spherocytic haemolytic anaemias (CNSHA) with anisopoikilocytosis, fragmented cells and high reticulocyte count. All the eight probands had remarkable persistence of embryonic and foetal globin synthesis. All eight subjects were found to be compound heterozygotes for two mutations in the KLF1 gene. One was a p.R301H substitution in exon 2 and the other was a 7bp insertion resulting in a frameshift and it disrupted the protein. In one proband the level of *BCL11A* mRNA was measured and it was shown the *BCL11A* expression was reduced when both alleles of *KLF1* gene were mutated.

1.3.4.3 Neonatal anaemia

Magor et al., (2015) described a case of a boy with severe neonatal anaemia born at 38 weeks of gestation and with mild hydrops fetalis. The boy also showed marked hepatomegaly, splenomegaly, and jaundice. Despite frequent exchange transfusions and phototherapy, the bilirubin level continued to increase. A bone marrow biopsy showed marked erythroid hyperplasia with dyserythropoiesis, normal megakaryocytes and normal myeloid cell differentiation. The HbF level remained >70% beyond 1 year of age and the HbA_2 was below normal level. Both the mother and the father had mildly elevated HbF levels while the mother also had elevated HbA_2 level. Both parents had reduced expression of CD44 and normal expression of CD235 (GPA) and they also had an In(Lu) serologic phenotype. Sequencing of the KLF1 gene detected a frameshift mutation in exon 3 in the mother. This mutation resulted

in a truncated version of *KLF1* that cannot bind DNA. The father also had a novel mutation at the beginning of exon 2 of the *KLF1* gene. The severe neonatal anaemia with kernicterus in the proband was caused by compound heterozygosity for the null mutations in the *KLF1*. This was the first described case of *KLF1*-null in humans.

1.3.5 KLF1 and Juvenile Myelomonocytic leukaemia

The role of KLF1 methylation in juvenile myelomonocytic leukaemia (JMML) was first reported by Fluhr et al., (2017). JMML is an aggressive malignant hematopoietic disorder which usually occurs in infants and young children. In JMML there is the differentiation of the leukemic stem/progenitor cells clone not only toward granulocytes and monocytes but also to the red cell lineage. One of the main hallmarks of the disease is increase in HbF. Increase in HbF at diagnosis is a bad prognostic factor and it is linked with higher risk of treatment failure and shorter survival (Busque et al., 1995; Niemeyer et al., 1997). Spleen cells from 14 patients diagnosed with JMML were collected while peripheral blood from healthy volunteers were also collected. Genomic DNA was isolated and bisuflite - converted and analysed for HBG1/HBG2, BCLL1A and KLF1 methylation. It was found that an upstream *KLF1* enhancer sequence was highly sensitive to DNA methylation and in JMML patients is was hypermethylated and not methylated in controls (Fluhr et al., 2017).

1.4 Haemoglobin

Haemoglobin (Hb) is the oxygen-carrying molecule in the body. It is chemically best considered as duplex of globin heterodimers: the main adult haemoglobin (HbA) is composed of two alpha (α) and two beta (β) chains assembled in two αβ dimers ($\alpha_1\beta_1$ and $\alpha_2\beta_2$), while HbF is composed of two alpha-gamma (αγ) dimers. Their main function is to transport oxygen

(O2) from the lungs to tissues, but it also specifically interacts with three other gases, carbon dioxide (CO2), carbon monoxide (CO) and nitric oxide (NO).

1.4.1 Globin genetics

The alpha (α) globin cluster lies closely to the telomere of the short arm of chromosome 16 in a GC-rich and Alu-dense gene segment. It contains one copy of the zeta (ζ) globin gene and two functional copies of the α-globin gene (HBA1 and HBA2). The ζ globin gene is expressed only during early embryogenesis, while α1 and α2 are expressed simultaneously in both foetal and adult stages. In addition, the α globin gene cluster also contains two pseudo-genes known are ψζ1 and ψα1 (Deisseroth et al., 1977). The remote regulatory region, known as hypersensitivity site 40 (HS-40) controls the expression of ζ, α1 and α2. The HS-40 is located approximately 40 kilobases (kb) upstream of the 5' end of the ζ-globin gene (figure 1.7) (Higgs et al., 1990).

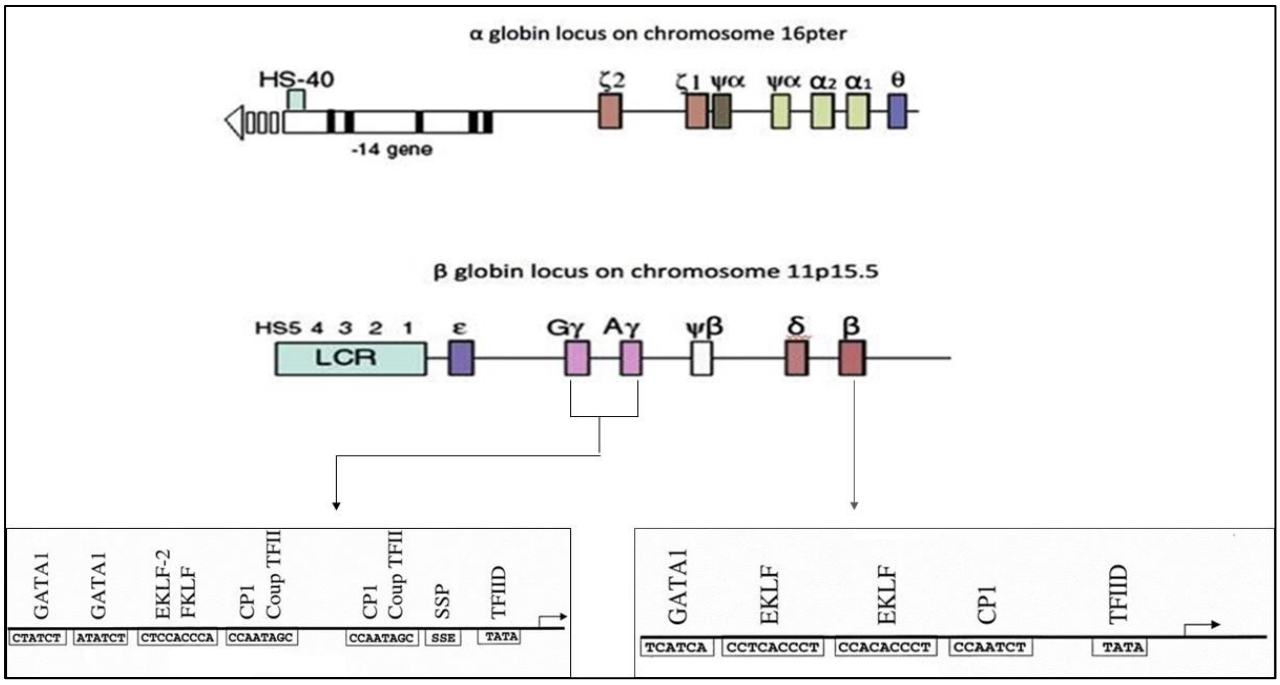

Figure 1.7: Schematic representation of the globin gene cluster. The α-like globin cluster (top) is located on the short arm of chromosome 16 and includes the ζ, α2, and α1 globin genes. These are under the control of an upstream remote regulatory region called HS-40. The β-like globin cluster (bottom) is located on the short arm of chromosome 11. The expression of the ε, γ, δ and β are under the control of the locus control region. The bottom part shows the organization of the γ and β promoters.

The β globin locus is located on chromosome 11 and it contains five genes sequentially arranged from 5' to 3' as they are expressed during the development of the human being. The epsilon (ε) is the first gene expressed during embryogenesis, two foetal genes, $^{G}\gamma$ and $^{A}\gamma$ (these two foetal genes code for the same protein except at amino acid position 136, Gy carries glycine and Ay carries alanine), and the adult delta (δ) and β genes (figure1.7) (Efstratidis et al., 1980). The promoters of the β-globin gene share remarkable homology but unique sequences in each gene are responsible for the developmental stage specificity of each promoter. A prominent difference is seen between the γ promoter and the β-promoter. As shown in figure 1.7 the γ-promoter has a duplication of the CAAT box and a single CACCC box while on the other hand the β-promoter has duplicated CACCC box and a single CAAT box. As illustrated in figure 1.7, several transcriptional factors such as KLF1, GATA1 CP1 bind to these sequences (Jane et al., 1992). Coordination expression of these five genes is dependent on the critical regulatory element located upstream of the genes knows as the LCR.

The LCR is made up of 5 DNase hypersensitivity sites (HS) denoted by HS1, HS2, HS3, HS4 and HS5 with the HS5 being the most 5' (Forrester et al. 1986). HS1 to HS4 are erythroid lineage specific, while HS5 functions as a developmental stage specific border (Farrell, West & Felsenfed, 2002; Wai *et al.*, 2003). Each HS has a core sequence which is approximately 250 nucleotides long (Stamatoyannopouls, 2005) and its role is to bind to different transcription factors and chromatin remodelling proteins such as GATA-1, KLF1, NF-E2, Tal1, Sp1 and USF (Liang et al. 2008).

1.5 Haemoglobinopathies

Mutations in the HBB locus can give rise to haemoglobinopathies. Haemoglobinopathies are the commonest monogenic disorders, affecting about 7% of the world's populations (Manca & Masala, 2008). The haemoglobinopathies can be classified as, (i) the structural variants such as Sickle cell disease (HbS), haemoglobin C (HbC) and Haemoglobin E (HbE), (ii) the thalassaemia which can be further sub-classified into α, β, δβ and εγδβ thalassaemia depending on the particular globin chain or chains that is/are ineffective synthesized and (iii) HPFH in which there is a defect in the normal switch from foetal to adult haemoglobin (Weatherall, 2001). Table 1.1 shows the number of haemoglobin variants known to date.

Query	Number of reported mutations
Total entries in database	1744
Total haemoglobin variant entries	1317
Total thalassaemia entries	494
Total entries in both variant and thalassaemia	49
Entries involving the alpha 1 gene	358
Entries involving the alpha 2 gene	442
Entries involving the beta gene	915
Entries involving the delta gene	127
Entries involving the Agamma gene	61
Entries involving the Ggamma gene	77
Entries with an insertion mutation	84
Entries with a fusion gene mutation	11
Entries with a deletion mutation	228
Entries with a substitution mutation	1435
Entries with a high oxygen affinity	100
Entries with a low oxygen affinity	48
Unstable haemoglobins	151
Methemoglobins	13

Table 3.1: Summary of number of variants in the database of human haemoglobin and thalassaemia. Retrieved on 21st May 2018 from http://globin.bx.psu.edu/cgi-bin/hbvar/counter.

1.5.1 β-thalassaemia

The β-thalassaemia's are caused by a spectrum of mutations resulting in a quantitative reduction of structurally normal β globin chains. When no β globin is produced, β thalassaemia alleles are considered as β^0/β^0 and these subjects are referred to as β thalassaemia homozygote. When some β globin is produced at a reduced level, the β alleles are considered either as β^0/β^+ or β^+/β^+ also known as β thalassaemia intermedia or β/β^+ or β/β^0 known as β thalassaemia trait. The diagnostic feature of β thalassaemia is the hypochromic microcytic red cells accompanied by an elevated level of HbA_2 with a normal or increased of HbF.

More than 390 mutations have now been associated with β thalassaemia (http://www.ithanet.eu/db/ithagenes?action=list&hem=2) but out of these only 40 account for 90% or more of the β thalassaemia worldwide (Kountouris et al., 2014). The most common mutations can be divided into non-deletions and deletions. The non-deletional mutations are usually single bases substitutions, small insertions, or deletions of one to a few bases located within the gene and its immediate flanking sequences. The non-deletion forms of β thalassaemia are further subdivided into transcriptional mutations, mutations affecting RNA processing and translational mutations (Huisman et al., 1997). Transcriptional mutants usually involve the conserved DNA sequences that form the β globin promoter or the stretch of 50 nucleotides in the 5'UTR. These transcriptional mutations usually include the functionally important CACCC, CCAAT and ATAA boxes and causes a mild to minimal reduction of β globin output (Chen et al., 2007).

Mutations affecting the RNA processing can be further subdivided into (i) mutations of the splice site junction – these mutations usually occur in sequences critical in the splicing process such as the invariant dinucleotides GT at the 5' and AG and the 3' splice junctions in the introns (Atweh et al., 1985). These mutations usually completely abolish the normal splicing and give

rise to β^0 phenotype; (ii) mutations of splice site consensus sequence – these are usually mutations that include the last three nucleotides of the exon and the first six nucleotides of the intron for the 5' donor site while for the 3' acceptor site they include the last 10 nucleotides of the intro and the first nucleotide of the exon. These mutations reduce the efficiency of normal splicing to a varying degree. The β-thalassaemia phenotype associated with this type of mutation depends on the position of the variant. For example, a mutation at the intervening sequence (IVS)1 position 5 give a severe β-thalassaemia phenotype while a substitution in the adjacent nucleotide; IVS1 position 6 give a milder β-thalassaemia phenotype (Treisman et al., 1983); (iii) mutations that create new alternative splice sites in introns – five such mutations have been identified in the β-globin gene two in IVS1 and three in IVS2. The phenotype associated with these mutations may be either β^+ or β^0 depending on the site and nature of the mutation (Metherall et al., 1986) and (iv) mutations that create alternative splice site in exons – four mutations in exon 1 have been identified that give rise to the activation of a cryptic or alternative splice site. One of the mutations is the GAG → AAG in codon 26 in exon 1 which give rise to HbE (Orkin et al., 1982).

Approximately half of the β thalassaemia alleles are the result of the introduction of a premature termination codon either due to direct mutations that create a stop codon or a frameshift mutation that cause a premature termination. One of the first nonsense mutation that was extensively studied was the mutation at codon 39 (CAG → TAG). This is the second most common mutation in the Mediterranean population (Maquat, 2004). As already stated above, β thalassaemia is rarely caused by deletions. In-fact only eighteen deletions restricted to the HBB gene have been described. These deletions range from 25bp to 6kb (Thein, 2017).

1.5.2 Silent β thalassaemia

Subjects heterozygotes for 'silent' β thalassaemia do not show any evident haematological phenotype (normal CBC and normal HbA_2). The only abnormality is the mild imbalance of globin chain synthesis. Silent β thalassaemia alleles are not common except for the C → T mutation at position -101 of the β globin gene which account for most of the milder forms of β-thalassaemia in the Mediterranean (Gonzalez-Redondo et al., 1989). Several other mutations in the 5' and 3' UTR in the β globin gene have also been associated with 'silent' β-thalassaemia. Berg et al., (1991) reported that the sequence variation [TA]x[T]y at position -530 of the β-globin gene may also be responsible for 'silent' β-thalassaemia and that the reduced β-globin expression may be related to increased binding of the BP1 repressor protein.

1.5.3 δβ Thalassaemia

δβ Thalassaemia is a term used to describe a disorder characterized by decreased or absent β-globin production with a variable compensatory increase in γ-chain synthesis. The first type of δβ was described by Gerald & Diamond (1958) and it was known as Hb Lepore. As shown in figure 1.8 Hb Lepore contains a fusion δβ-globin chain produced by a misaligned crossover between the δ and β genes (Baglioni, 1962). Different cross over points give rise to different Hb Lepore such as Hb Lepore-Boston and Hb Lepore-Hollandia. This recombination event also give rise to an anti-Lepore chromosome, which contains a normal δ and β loci and a fused (βδ) gene (figure 1.8). A variety of anti-Lepore-like haemoglobins have been discovered including Hb Miwada, Hb P-Congo, Hb Lincol Park and Hb P-Nilotic (Felice et al.1982). The δβ thalassaemias can be subdivided into those that produce both $^{G}\gamma$ and $^{A}\gamma$ chains known as $^{G}\gamma^{A}\gamma(\delta\beta)^{0}$ thalassaemia and those that only contain Gγ chains known as $^{G}\gamma(^{A}\gamma\ \delta\beta)^{0}$.

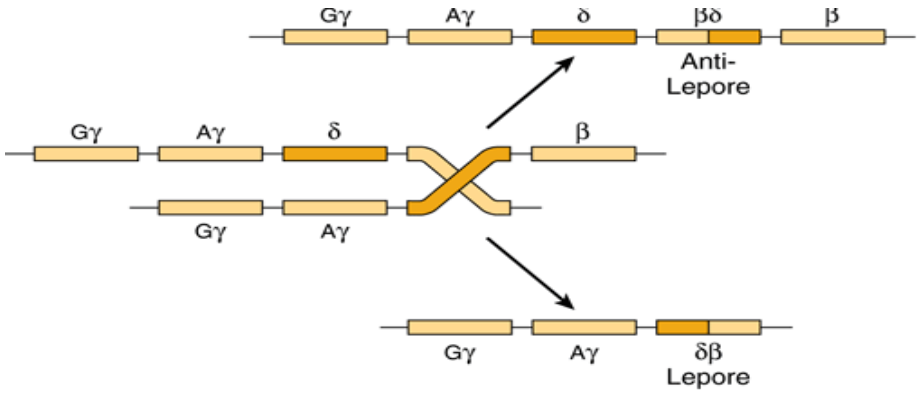

Figure 1.8: The abnormal crossing over mechanism involved in the generation of (a) the Lepore and anti-Lepore haemoglobins.
Adapted from: Recent Developments in Molecular Genetics of Human Haemoglobin by Weatherall & Clegg (1979).

1.5.4 Hereditary Persistence of Foetal Haemoglobin

Hereditary Persistence of Foetal Haemoglobin describes a range of conditions with a persistence of HbF production in adult life in the absence of any haematological disorder. Classification of HPFH is based on the type of cellular distribution of HbF (pancellular or heterocelluar), the type of globin chain produced (Gγ, Aγ or both) and the type of molecular defect present (deletion or non-deletion) (Bollekens & Forget, 1991).

Huisman et al. (1972) reported an abnormal haemoglobin associated with HPFH designated as Haemoglobin Kenya, resulted from an unequal crossing over between the Aγ gene and the β gene. The deletion of the Aγ to β gamma region gives rise to the production of a new fusion globin containing amino acids 1 to 80 of the Aγ chain and amino acids 81 to 146 of the β chain. Heterozygotes for this abnormal Hb, produce 7% to 25% Hb Kenya and 5% to 15% HbF. The HbF only contains α and Gγ chains (Kendall et al., 1973).

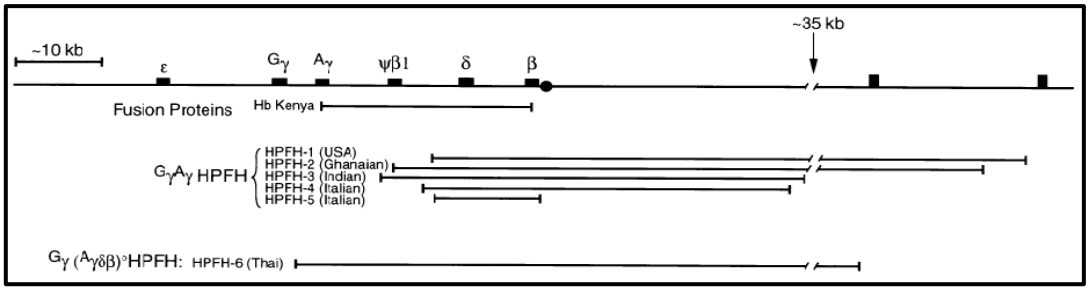

Figure 1.9: Deletions of the β-globin gene cluster associated with fusion proteins and HPFH. Adapted from Molecular Basis of Hereditary Persistence of Fetal Haemoglobin by Forget, B. (1998).

Six types of deletional HPFH (HPFH1-HPFH6) (figure 1.9) produced as a result to partial deletions of the δ and β globin genes have been identified. Deletional HPFH are characterized by pancellular homogenous distribution of HbF among all red cells and in heterozygotes and an increase in HbF levels between 20% to 30% is noted (Bollekens & Forget, 1991). Non-deletion HPFH are caused by point mutations affecting the promoter of the Gγ or Aγ genes (Stamatoyannopoulos & Grosveld, 2001). In heterozygotes, the HbF levels vary from 1% to 35% with prevalence of Gγ or Aγ chains according to the gene promoter containing the mutation (Thein, 2008). These point mutations have been grouped in three distinct regions of the 5'flanking DNA of the affected γ-globin genes. The first region is located approximately 200 bp from the site of transcription initiation of the γ genes. The second region is in the γ promoter while the third region is in a well-known regulatory element of globin and other genes: the CCAAT box sequence (Forget, 1998).

1.6 Haemoglobin switching

One of the most studied events that occurs in the β-globin cluster is haemoglobin switching. It involves the suppression of the γ-globin gene accompanied by the increase in the expression of the β-globin gene. Understanding this process has major therapeutically implications for patients with haemoglobinopathies such as β-thalassaemia and sickle cell disease (SCD) as increase in γ globin can functionally substitute the defective β-globin.

In humans two globin switches occur (figure 1.10); the embryonic to the foetal globin switch and the foetal to the adult switch. The embryonic haemoglobin is composed of Hb Gower 1 ($\alpha_2\varepsilon_2$), Hb Gower 2 ($\zeta_2\varepsilon_2$) and Hb Portland ($\zeta_2\gamma_2$). As the expression of ζ- and ε- globin begins to cease after the first two months of gestation, the first switch occurs giving rise to the synthesis of Hb F ($\alpha_2{}^G\gamma_2$) and ($\alpha_2{}^A\gamma_2$). The site of erythropoiesis also changes from the yolk sac and para-aortic region to the foetal liver (Dover & Boyer, 1980).

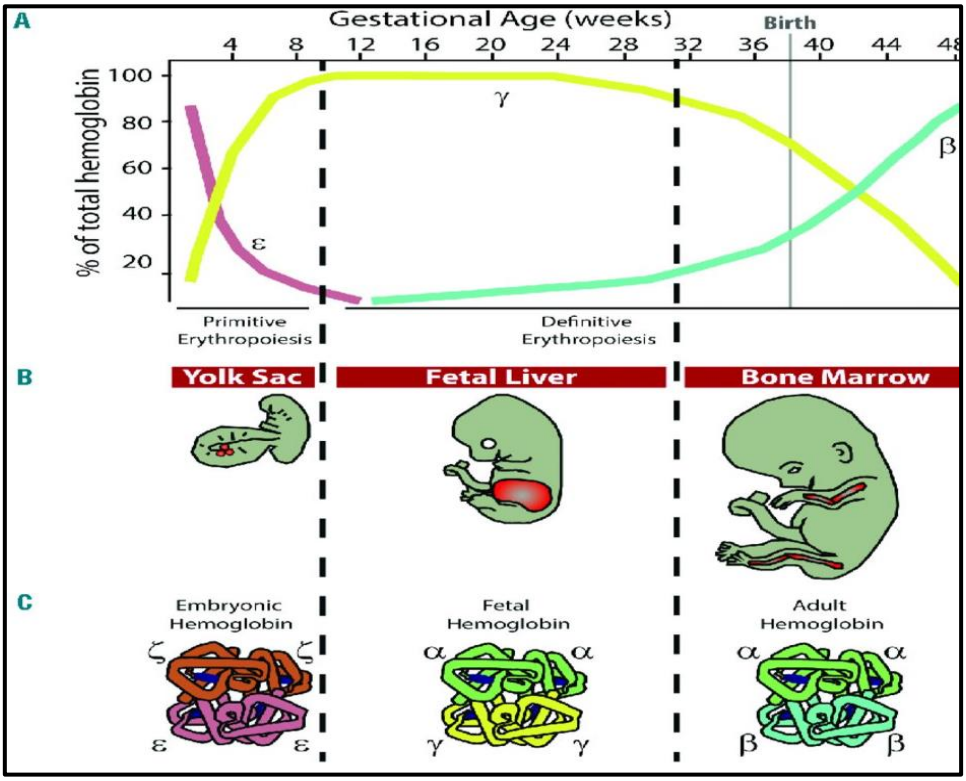

Figure 1.10: Haemoglobin switching– Figure A shows the two haemoglobin switches. At around 6th week of gestation, embryonic globin is silenced, and foetal globin starts to be expressed. At around the time of birth, foetal to adult switch occurs. Figure B shows the major anatomical sites of hematopoiesis during development. Erythropoiesis occurs in the blood islands of the yolk sac in the first 8 weeks of gestation. Between the 8th and 32nd week it continues in the foetal liver and after the 32nd week in the bone marrow. Figure C shows the main human haemoglobins expressed during development. Adapted from Flicking the switch: Adult hemoglobin expression in erythroid cells derived from cord blood and human induced pluripotent stem cells by Cantu & Philipsen (2014).

The second globin switch which occurs at the time of birth, involves the decline of Hb F synthesis coupled with increased synthesis of adult haemoglobin composed of Hb A ($\alpha_2\beta_2$) with a minor Hb A_2 ($\alpha_2\delta_2$) (Brinkman and Jonxis, 1935; Weinberg et al., 1983). This switch is

also accompanied by a change in the site of erythropoiesis from the foetal liver and spleen to the bone marrow. Residual amounts of Hb F continue to be synthesized throughout adult life and expressed by F-erythrocytes (Kleihauer, Braun & Betke, 1957; Hosoi, 1965).

1.7 Role of Cis And trans acting factors in globin gene switching

Globin gene switching is controlled by a complex interaction between the cis and trans regulators as shown in figure 1.11. The cis regulators consist of the LCR together with the downstream globin sequences while the trans-acting regulators consist primarily of transcription factors such as KLF1, BCL11A, HBS1L-MYB and GATA-1.

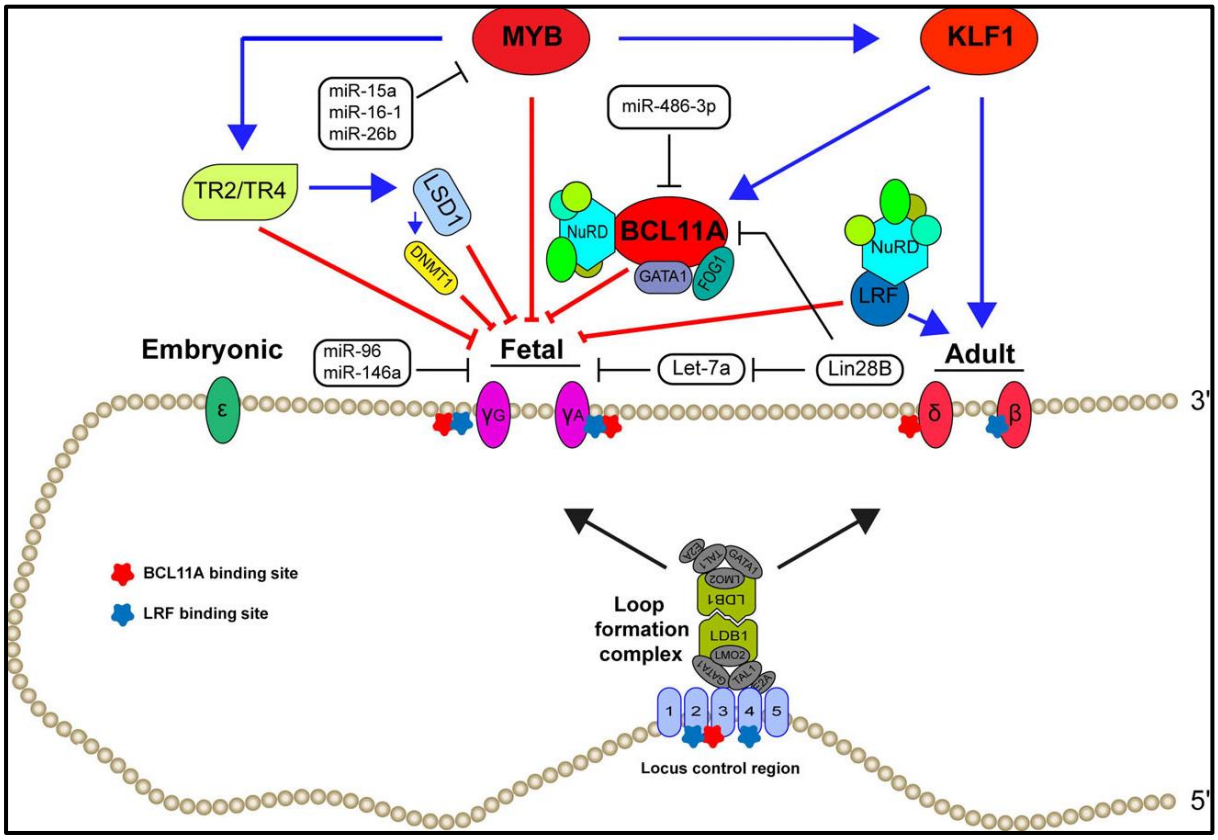

Figure 1.11: Regulation of foetal haemoglobin production. Globin switching cis and trans regulators and chromatin loop formation complex Adapted from Fetal haemoglobin induction in sickle cell disease by Paikari & Sheehan (2017).

1.7.1 Switching mechanisms

The discovery of the LCR played an important role in the investigation of switching mechanisms. Studies in transgenic mice revealed that two main mechanisms (i) gene silencing and (ii) gene competition control haemoglobin switching.

1.7.1.1 Gene silencing

The concept of gene silencing came from studying transgenic mice carrying the embryonic human ε gene. These experiments showed that these mice carrying the ε gene fail to express that gene at any stage of development but when the ε gene is linked to the LCR there is abundant ε expression. This showed that the expression of this gene is dependent on the presence of LCR. The expression of the ε gene is totally restricted in the embryonic yolk sac cells and all the sequences required for silencing this gene are contained in sequences flanking the gene. This shows that the control of ε gene is autonomous (Raich et al., 1990; Shih et al., 1990).

1.7.1.2 Gene competition

Behringer et al., (1996) and Enver et al., (1990) carried out experiments in transgenic mice and it was shown that there is a competitive mechanism of γ to β globin gene switching. When the LCR was linked to either the γ or the β gene alone, developmental control was lost, but when the LCR was linked to both of the genes, developmental control was restored. Following these experiments, it was suggested that in the embryonic stage there is the interaction of the LCR with the ε-globin gene and by gene competition the downstream genes are turned off. In the foetus, the ε-gene is silenced and the LCR interacts with the Gγ and Aγ genes, while in an adult the γ gens are silenced and the LCR interacts with the β-globin gene.

The interaction of the gene with the LCR depends on two conditions which are (i) the predominant transcriptional environment and (ii) the distance from the LCR (Hanscombe et

al., 1991; Peterson & Stamatoyannopouls, 1993). Visualization of the interaction between the LCR and globin genes was made possible by in situ hybridization methods. It showed that the LCR interacts with only one promoter of the locus at a given time and during switching there is a change in frequency of interaction of the LCR with a given gene (Wijgerde et al., 1995; Fraser & Grosveld, 1998). Technologies such as 3C and RNA trap assays showed that a chromatin loop is formed when the globin gene is enhanced by the LCR. The formation of this loop between the LCR and the β-globin gene requires erythroid specific trans-acting factors such as KLF1, GATA-1 and FOG-1 (Carter et al., 2002; Palstra et al., 2003; Drissen et al., 2004; Vakoc et al., 2005). Breda et al., (2016) showed that a complex combination of factors including GATA1, TAL1, E2A, LMO2 and LBD1 mediates the formation of the loop between the LCR and globin promoters.

1.7.1.3 Other cis-acting factors

Sequences *in cis* to the β globin locus include the X*mnI* and the $(AT)_XT_Y$. The XmnI (C→T) polymorphism at position -158 of the HbG2 was shown to promote the expression of foetal haemoglobin (Gilman & Huisman, 1985). Different studies have demonstrated that subjects with the T allele have increased HbF and individuals with SCD and β thalassaemia had milder symptoms (Labie et al.1985; Thein et al.1987). In adults under erythropoietic stress, the -158 Gγ C→T activates the X*mn*I site contributing to an overall higher HbF expression levels. Since the -158 substitution is near a DNase I hypersensitive site located 50 to 150bp 5' of the γ gene CAP sites, its mechanism may be structural. *In vitro* it was evident that the chromatin of this region is accessible to DNase I, since it will have an 'open' structure. In *vivo* the -158 substitution increases the probability that the chromatin of this region will be more accessible to components of the transcription apparatus of the adult erythroid cell (Gilman & Huisman, 1985).

The (AT)x(T)y motif, located 0.50 kb 5' to the β globin gene, is involved in the regulation of the β gene. This region exhibits a high content of adenine (A) and thymidine (T) residues (Berg et al.1989). Typically, this sequence consists of 7 consecutive pairs of (AT) dinucleotides followed by 7 (T) nucleotides (Chebloune et al.1988). The most common sequence is the $(AT)_7T_7$ and is considered as the reference sequence. Other configurations have also been described and these include $(AT)_9T_5$, $(AT)_8T_4$, $(AT)_8T_5$,$(AT)_8T_6$, $(AT)_6T_9$, $(AT)_{10}T_3$, and $(AT)_{11}T_3$. The $(AT)_xT_y$ likely acts as a DNA binding site for the BP1 protein which represses the adult delta and beta globin genes (Chase et al, .2002; Fu et al., 2001). Variable severities of sickle cell disease have been linked to polymorphic motifs of the $(AT)_xT_y$ (Elion et al., 1992; Lemsaddek et al., 2004). Shorter $(AT)_xT_y$ motifs, such as the $(AT)_6T_9$, produce more Hb S because they bind less BP1 protein. On the other hand, cases with longer $(AT)_xT_y$ sequences, such as $(AT)_9T_5$ and $(AT)_{11}T_3$, express significantly less Hb S vs Hb F (Elion et al., 1992).

1.7.2 Trans-acting factors in globin gene switching

1.7.2.1 HBS1-like protein – v-myb myeloblatosis viral oncogen (HBS1L-MYB)

Linkage analysis in a large Indian family with HPFH showed that the genetic determinant for high HbF segregated independently from the β-globin cluster (Thein & Weatherall, 1989). Via a painstaking mapping exercise, the genomic location for high HbF in this family was mapped to chromosome 6q23 (Garner et al.1998). Thein (2007) found that it was due to variants in the region between the HBS1L and MYB genes. The *HBS1L* gene is expressed in haematopoietic cells and is involved in regulation of several cellular processes, while the *MYB* gene is involved in oncogenesis and plays an important role in erythropoiesis (Thein et al., 2007).

Jiang (2006) found that high HbF phenotype was associated with lower cMYB levels in erythroid cultures. Foetal livers null for cMYB gene had only megakaryocytes and macrophages present and it was noted that reduced cMYB expression increases γ-globin

expression via a process of accelerated erythroid maturation. Homozygotes mutations in the HBS1L-MYB gene are linked with higher HbF levels in healthy subjects or subjects' heterozygotes for β thalassaemia (Craig et al., 1996). In healthy Northern European subjects SNVs in HBS1L-MYB contributes to 19% of the variations in the HbF levels (Thein et al., 2009). The exact role of MYB in globin switching is still not clear. However, it was established that disruption of MYB in mice give rise to an increase in ε and γ globin expression. This showed that MYB plays a role in gamma globin silencing (Suzuki et al., 2013).

1.7.2.2 B-Cell Lymphoma/leukaemia (BCL11A)

Genome-wide association studies (GWAS) led to the identification of *BCL11A* on chromosome 2p15 as a potential modifier of HbF levels (Menzel et al., 2007 & Uda et al., 2008). *BCL11A* was known to be a zinc finger transcriptional repressor protein, an oncogene in B-cell malignancies (Satterwhite et al., 2001) and to play an important role in gene expression during B-lymphopoiesis and neurogenesis (Avram et al., 2000). Menzel (2007) showed that BCL11A is expressed in erythroid precursors while knockdown of BCL11A by RNA interference (RNAi) in Human erythroid progenitor cells (HEPS) showed an increase in HbF expression without any unfavourable effect on erythropoiesis (Sankaran et al., 2008). Knockdown of BCL11A in genetically engineered SCD mice reversed the phenotype of these mice, showing the therapeutic potential of targeting BCL11A (Xu et al.2011 & Esteghamat et al., 2013).

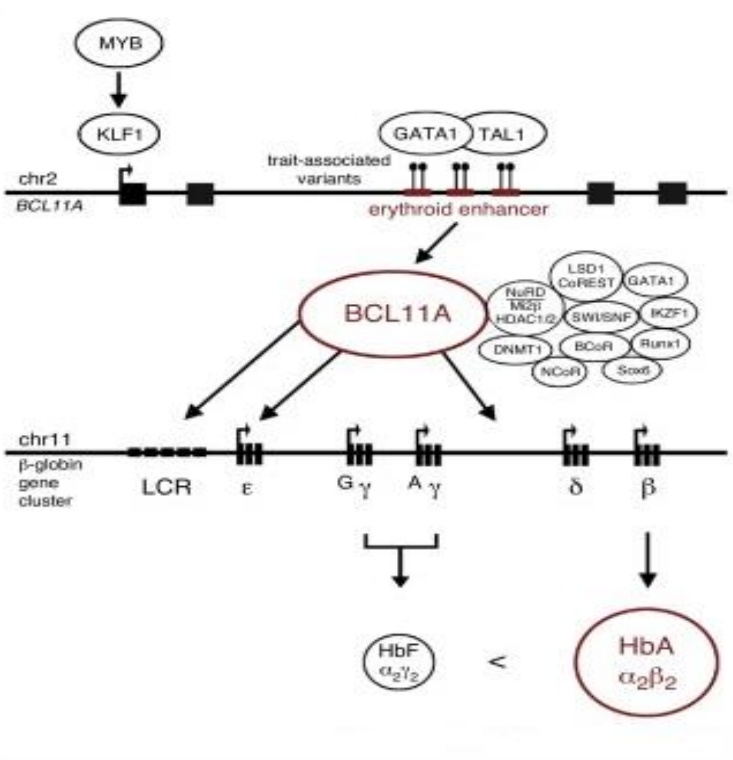

Figure 1.12: The BCL11A-HbF repression axis. Adapted from Haemoglobin switching's surprise: the versatile transcription factor BCL11A is a master repressor of foetal haemoglobin by Bauer & Orkin (2015).

BCL11A participates in multiprotein transcriptional complexes in erythroid cells (Figure 1.12). Its protein partners include DNA-binding erythroid TFs such as FOG1, GATA-1, RUNX1, IKZF1 and SOX6. It also binds to numerous transcriptional corepressors and chromatin regulators such as NuRD, SIN3 and SWI/SNF complexes (Sankaran et al., 2008 & Xu 2013). At the β-globin gene cluster, the *BCL11A* occupies the erythroid chromatin and it binds at numerous sites within the LCR distal enhancer elements and the ε-globin gene (Sankaran et al., 2008). BCL11A in erythroid cells is required for configuring the β globin locus by promoting long range interactions between the LCR and the β globin gene at the expense of the interaction between the LCR and the γ globin (Xu et al., 2010).

Like many SNPs highlighted by GWAS, the common genetic variants in *BCL11A* associated with HbF level fall in a non-coding region, mainly in the second intron. The variants which are co-inherited as haplotypes, cluster at a region marked by an erythroid enhancer chromatin

signature. The BCL11A variant most highly associated with HbF is the rs1427407. This variant disrupts a composite half E-box/GATA motif, a TF recognition site highly bound by GATA-1/TAL1 complexes in erythroid cells. In primary human erythroid precursors, this variant was associated with modest reduction in TF binding and down regulation of *BCL11A* (Bauer et al., 2013).

1.7.2.3 SOX-6 & Gata-1

Sox6 belongs to the family of Sry-related high mobility group (HMG) box transcription factors and is expressed in several tissues including, cartilage, testis, neuronal and erythropoietic tissues (Conner, et al., 1995). Inactivation of *Sox6* gene in mice results in neonatal death secondary to cardiac or skeletal myopathy (Smits et al., 2001). At the foetal liver stage, *Sox6* deficient mice showed elevation in embryonic globins while the Hbb-bh1 expression was downregulated in late foetal levels (Yi et al., 2006). Transplantation of foetal liver cells from *Sox6* deficient mice into wild-type adult mice resulted in repression of the εγ-globin (Cohen-Barak, 2007) by binding directly to the εγ-promoter (Yi, 2006). In mice *Sox-6* also enhance definitive erythropoiesis by stimulating erythroid cell survival, proliferation, and terminal maturation (Dumitriu et al., 2006). *Sox6* is co-expressed with *BCL11A* during erythroid development. It interacts physically with *BCL11A* and Gata-1 to occupy the β-globin cluster. The physical interaction between Sox-6 and *BCL11A* may help recruit *BCL11A* and NuRD repressor complexes to the proximity of the γ-globin genets. *Sox6* may act as a cofactor of *BCL11A* in globin gene switching and silencing of γ-globin gene (Xu, 2010).

Gata-1 forms part of the GATA family of proteins made up of zinc-finger TFs (Orkin, 1992). Gata-1, which was discovered as a β-globin locus-binding protein (Evans & Felsenfeld, 1989), is essential for erythroid maturation in vivo in fact, Gata-1-null cells fail to mature beyond the proeythroblast stage (Penvy et al., 1995). Gata-1 has a role in activating expression of the adult

mouse *Hbb-b1* by recruiting RNA polymerase II to the *Hbb-b1* promoter (Johnson, 2002). It was suggested that Gata-1 may carry out haemoglobin silencing by inducing chromatin looping since Gata-1 was shown to bind to a region upstream of the *HBG-1* and *HBG-2* promoter in a FOG 1 depended manner. In doing so, it led to the recruitment of the NuRD repressive complex.

1.7.2.4 Leukaemia/lymphoma-related factor (LRF)

The leukaemia/lymphoma related factor (LRF) recently has been identified as a new transcription factor that represses γ-globin expression (Masuda et al., 2016). The LRF is encoded by the ZBTB7A gene and it is a ZBTB transcription factor that binds DNA through C-terminal C_2H_2-type zinc fingers. Through its N-terminal BTB domain it recruits a transcriptional repressor complex (Lee & Maeda, 2012). To assess the role of the LRF on the erythroid transcriptome, Masuda et al. (2016) inactivated the Zbtb7a gene in erythroid cells of adult mice and then RNA sequencing was performed using splenic erythroblasts from control and LRF conditional knockout mice. They found that the LRF transcription factor interacts with the γ-globin genes and maintains the nucleosome density necessary for γ-globin gene silencing in adults. It carries out its repressive activity through a NuRD repressor complex independent of *BCL11A*.

1.8 Specific Objectives

The main objectives of this thesis were:

To investigate the phenotypic variability of HbF in subjects with the same *KLF1* truncation mutation, the p.Lys288Ter.

To explore the occurrence of additional *KLF1* mutations in undiagnosed subjects with borderline HbA_2 and β-thalassaemia heterozygotes.

2.1 Sample collection

2.1.1 Families Recruitment

Approval for conducting research involving specimen and data collection and preservation was sought and obtained from the University of Malta Research Ethics Committee (UREC) (Appendix A)The Maltese family with HPFH reported by Borg et al., (2010) was extended and through the Thalassaemia & Genetics Clinic at Mater Dei Hospital, 5 other families with the same KLF1 mutation; the p.Lys288Ter (rs267607202) but with different level of foetal haemoglobin were identified. Throughout the thesis these families are referred to as Fam F1 to Fam F6 respectively.

Fam F1 was extended with baby IV-9 born at 40 weeks of gestation together with the testing of his siblings IV-7 and IV-8. In Fam F2 the proband (II-2) presented with normal haemoglobin, mean cell volume (MCV), low mean cell haemoglobin (MCH) and mean cell haemoglobin concentration (MCHC). In Fam F3 the proband (I-1) presented with low haemoglobin, low MCV, low MCH and low MCHC. In Fam 4 the proband (II-3), a 23-year old pregnant woman presented with normal haemoglobin, low MCV, MCH and MCHC. In Fam F5 and Fam F6 the probands were both antenatal cases; in Fam F5 the proband presented with low haemoglobin, low MCV, MCH and MCHC while in Fam F6 the proband presented with normal haemoglobin and MCV and low MCH and MCHC. The probands were asked to encourage further participation from other relatives.

Blood samples were collected in two purple vacutainers containing K2-EDTA as an anticoagulant. One sample was used to measure the haematological parameters together with HbA_2 and HbF while the other sample was used for DNA extraction. In case of children, for DNA extraction cheek cells were obtained, by rinsing their mouth with 25mL of mouth wash

solution for about 30 seconds. The individuals were advised not to brush their teeth before collecting the specimen and to do it first thing in the morning. The mouthwash was then transferred into a sterile conical tube and stored at room temperature prior to DNA extraction.

2.1.2 Collection of samples with HbA2 between 3.1% and 6.7%.

A retrospective search in the Malta Biobank was carried out for samples with HbA_2 between 3.1% and 6.7%. All pregnant woman and children under 2 years of age were excluded from the study. Four-hundred and twenty-six (426) samples were collected. Together with these 426 samples, 200 controls with normal complete blood count (CBC), HbA_2 less or equal to 3.0% and normal HbF levels were included in this study

Haematological parameters were measured using an automated haematology analyser (XT-20001, Sysmex, Canada) at the haematology laboratory at Mater Dei Hospital, while HbF and HbA_2 were determined by High Performance liquid chromatography (HPLC) (Variant IITM System, Bio-Rad, USA) at the Genetics Laboratory, University of Malta. The absolute HbF level in milligrams per decilitre (mg/dL) was calculated using the following formula:

$$\text{HbF (mg/dL)} = \frac{\text{HbF (\%) x Hb(g/dL)}}{100} \text{x1000}$$

Where the HbF (%) was determined by HPLC and the Hb (g/dL) was obtained from the blood count.

2.2 DNA extraction

Genomic DNA was extracted and purified from peripheral blood leucocytes by salting out technique (Miller et al., 1988). Genomic DNA from cheek cells was also extracted by salting out technique.

2.2.1 DNA extraction from whole blood

Reagents were prepared (Appendix B) and whole blood collected in EDTA tubes and stored at -20°C was allowed to thaw at room temperature for about one hour on a rotator mixer. DNA extraction tubes were labelled and 3mL of blood were transferred in each respective tube followed by the addition of 9mL of 1 x erythrocyte lysing buffer (0.155M NH_4Cl, 10mM $KHCO_3$, 0.1nM Na_2EDTA; pH 7.4). The mixture was kept cool for 15 minutes with intermittent mixing followed by centrifugation at 2500 x g for 10 minutes at 5°C. After centrifugation the supernatant was discarded, and the pellet of white blood cells was washed with 3mL 1 x erythrocyte lysing buffer, vortexed briefly and centrifuged as before. These steps were repeated until a clean white pellet was obtained. The pellet of white blood cells was resuspended in 1.5mL of 1 x SE buffer (75mM NaCl, 25mM Na_2EDTA; pH 8.0) containing 1% Sodium Dodecyl Sulfate (SDS w/v) and 1.5µL proteinase K (1µL proteinase K per 1mL SE/SDS). The tubes were incubated at 37°C overnight in a water bath. During this step the white blood cells were lysed by the SDS while the proteins were partially digested by proteinase K.

After the overnight incubation, 1.5mL of SE buffer together with 750µL of 6M Sodium Chloride (NaCl) were added to each tube, followed by the addition of 3.75mL chloroform. The tubes were vortexed vigorously for 20 seconds and then transferred on a rotator for 1 hour.

Every 15 minutes each tube was vortexed for 20 seconds. The emulsion was then centrifuged for 10 minutes at 2000 x g with the slowest breaking force. The upper aqueous phase (containing DNA) was aspirated without disturbing the inter-phase. This was carried out using a sterile transfer pipette. The upper aqueous phase was transferred to a sterile well labelled 15mL DNA extraction tube. DNA was precipitated by adding double volume (of the upper phase) of cold 100% ethanol to the aqueous phase followed by gentle swirling of the tube. Using a sterile disposable Pasteur pipette, the DNA was transferred to a new microfuge tube containing 1mL of cold 70% ethanol. The tube was centrifuged at 11,000 x g for 4 minutes and the resulting supernatant was decanted off. One millilitre of 70% ethanol was added to each tube followed by centrifugation as described above. This step was repeated another two times to purify the DNA from any salts. The DNA pellet was dried and was re-dissolved in 100µL of Tris- Ethylenediaminetetraacetic acid (TE) buffer (1M Tris-Cl, 0.5M EDTA; pH 8.0) and left overnight on a rotator. The DNA was stored at 4°C until analysed.

2.2.2 DNA extraction from cheek cells

The tubes containing the mouthwash were centrifuged at 2400 x g for 15 minutes and without disturbing the pellet, the supernatant was removed by means of a sterile pipette. The pellet was washed with 25mL of Phosphate buffer saline (PBS) and centrifuged at 2400 x g for 10 minutes. This washing step was repeated another two times to ensure complete removal of the mouthwash. The pellet was then resuspended in 1mL SE buffer containing 1µl proteinase K and 1% SDS and incubated at 37°C overnight in a water bath. The following day, 1 mL of SE buffer together with 500µL 6M NaCl and 2.5mL chloroform were added. The tubes were left to mix for 1 hr and vortexed for 20 seconds every 15 minutes. Centrifugation was then performed at 600 x g for 10 minutes with slowest breaking force being set and the supernatant

was then transferred to a clean tube and equal volume of ice-cold 100% ethanol were added to the supernatant to precipitate the DNA.

The DNA was spooled out and placed in a microfuge tube filled with 1mL of 75% ethanol and centrifuged for 4 minutes at 11,000 x g for 4 minutes. While taking extreme precaution not to lose the DNA, the ethanol was then decanted, and the microfuge tubes were left to dry. To the DNA pellet, 100µL of TE buffer were added and the DNA was stored at 4°C until analysed.

2.2.3 Determination of DNA concentration

DNA concentration and purity were determined using NanoDrop™2000/2000c(Thermo Scientific, US). The NanoDrop technology is an innovative sample retention system that uses the surface tension to hold and measure microvolume samples. The samples are measured between two optical pedestals without the use of capillaries or cuvettes. The microsample is placed directly on top of the detection surface and between the ends of the optical fibres. Due to the surface tension, a liquid column is created, forming a vertical optical path. The light source is provided by a xenon flash lamp and the light that passes through the sample is analysed by a spectrometer that utilizes an array (Desjardins, Hansen & Allen, 2009).

Using the appropriate buffer, 1µl of blank solution was pipetted onto the bottom pedestal, the arm was lowered, and blank was chosen. The blank was wiped off from the measurement pedestals using a dry, lint free laboratory wipe. The sample numbers were entered in the appropriate field and 1µl of sample was pipetted and measured.

For nucleic acid quantification the NanoDrop uses a modified Beer-Lambert equation which is: $c = (A * \varepsilon)/b$

where:

c – is the nucleic acid concentration in ng/µl

A – is the absorbance in Au

ε – is the wavelength-dependent extinction coefficient in ng-cm/µl

b – the path length in cm

The generally accepted extinction coefficients for double-stranded DNA are 50 ng-cm/µL while for RNA are 40 ng-cm/µL. The ratio of absorbance at 260 and 280 nm is used to assess the purity of DNA and RNA. For DNA a ratio of ~ 1.8 is generally accepted as pure for DNA while for RNA a ratio of ~ 2.0 is generally accepted as pure for RNA.

2.3 Polymerase chain reaction

Polymerase chain reaction (PCR) is a simple enzymatic assay allowing the amplification of a specific DNA fragment from a complex pool of DNA. Each PCR assay requires the presence of a template DNA, primers, nucleotides and DNA polymerase which is the key enzyme linking the individual nucleotides together forming the PCR product. It involves three stages. The first stage is the denaturation stage which takes place at 94°C. During this step the hydrogen bonds within the DNA molecule are disrupted giving rise to two single strands of DNA (ss-DNA). The second stage is the annealing stage where oligonucleotide primers anneal to the complementary sequences on the ssDNA. They bind specifically to sequences flanking the segment to be amplified. The annealing temperature depends on the melting temperature (T_m) of the primers. The T_m is calculated using the following equation:

$$T_m = 2(A+T) + 4(C+G)$$

The third stage is the extension stage where DNA polymerase using an ssDNA as template will extend the primers in the 5' to 3' direction. The time needed for the reaction to occur is determined by the length of sequence to be amplified. Usually 1000bp are amplified every 1 minute. The cycle is repeated between 25 to 35 times and the reaction comes to an end when all free nucleotides are used up and DNA polymerase is degraded (Mullis, 1980).

2.3.1 Primer design

For primer design a web application known as Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) (Untergasser et al., 2012) was used. Before designing of the primers, several factors were taken into consideration, which included:

(i) primer length – primers ideally should be between 18-22 base pair (bp).

(ii) primer melting temperature – primers must have a melting temperature (Tm) between 60 - 64°C with an ideal temperature of 62°C. Increasing the Tm extensively can give rise to secondary structure. Ideally the Tm of the 2 primers should not differ by more than 2°C for both primers to bind simultaneously and efficiently amplify the product.

(iii) GC content – the GC content must be between 35%-65% with an ideally content of 50%. This allows complexity but still maintains a unique sequence.

(iv) Amplicon length – the distance between the primers was chosen to be between 200bp and 900bp.

The primer sequences obtained from Primer 3 were analysed through Primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Ye et al., 2012) to check primer specificity. The primers were then checked using Oligo Analyzer. This software checks if the

sequences form any considerable secondary structures during the PCR. Primers were ordered from IDT, Belgium.

2.3.2 PCR of the β Globin Gene and the KLF1 gene

Amplification of the β-globin Locus Control Region (LCR) was carried out to check for the $(AT)_xT_y$ motif, amplification of the β-globin gene was carried out to check for thalassaemia mutations, while amplification of the HbG2 promoter was performed to check for the XMN1 polymorphism (figure 2.1a) together with the amplification of the *KLF1* promoter and exons (figure 2.1b). Table 2.1 shows the primers used, the length of the PCR products and the method of post-PCR analysis in which either restriction enzyme digest or DNA sequencing were used.

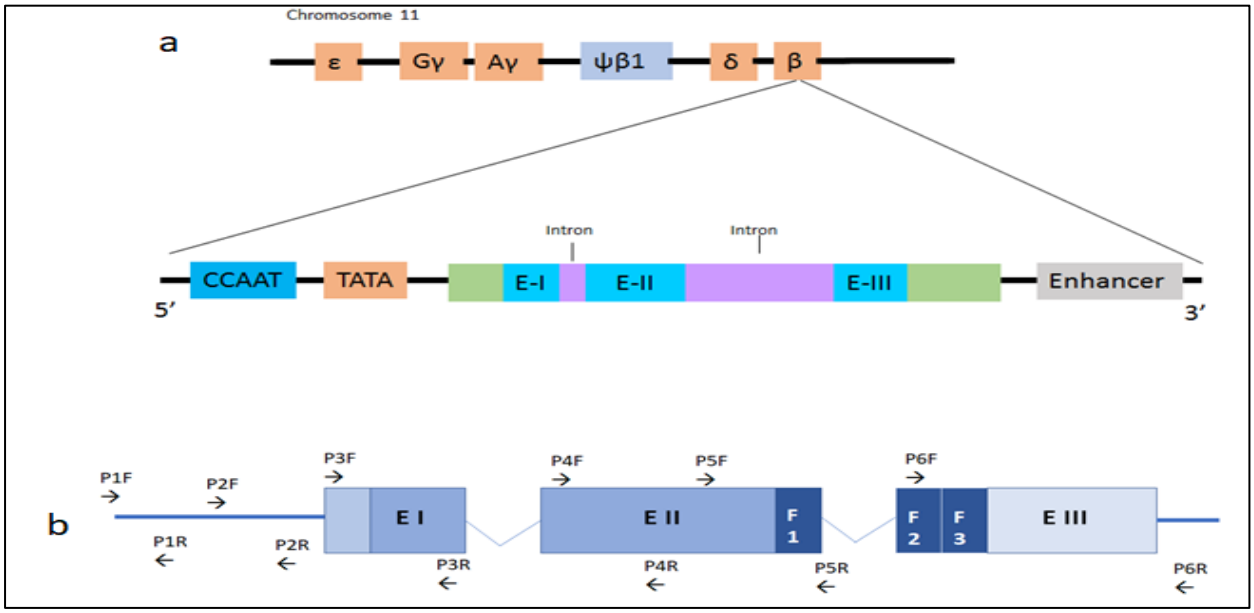

Figure 2.1: The Beta and KLF1 gene. Figure 2a shows the structure of the β globin locus and the β globin gene. Figure 2b shows a map of the KLF1 gene. EI, EII and EIII refer to exon 1, exon 2 and exon 3. The dark blue boxes (F1, F2 and F3) represent the DNA binding domains. P1 to P6 are the primers used to sequence all the KLF1 promoter and exons.

Primer Name	Sequence 5' → 3'	Fragment Size (bp)	Post-PCR analysis	Region
5' Gγ FWD 5' Gγ REV	ATC GGG TGC CTA CAT ACA TAC C GCT TGT GAT AGT AGC CTT GTC C	866	*XMNI* restriction enzyme digest	HbG2
5' β FWD 5' β REV	AAG TAA CTA ATG CAC AGA GC GCA CTG GCT TAG GAG TTG G	655	DNA sequencing	$(AT)_xT_y$
β FWD β REV	AAG GCT CAT GGC AAG AAA GTG C CTA AAA CGA TCC TGA GAC TTC C	281	DNA sequencing	β globin
5' *P1* FWD 5' *P1* REV	GCA GAG AGA GGG CTT CAT CC AAG TGA CCC TCT CTC CTC CT	653	DNA sequencing	KLF1 Promoter
5' *P2* FWD 5' *P2* REV	ATC CAA CGG TCC TAT CCC AC CTC GTG AAC TCT GAG GCT GT	768	DNA sequencing	KLF1 Promoter
P3 FWD 5' P3 REV	GCT TTG GAC ACA GGG TTA GT TCA GGT CAA GAT GCA GGT CT	400	DNA sequencing	KLF1 Exon1
P4 REV	TTC CAA AGC CTC TGC GTC AG ACG CCG CAG GCA CTG AAA G	570	DNA sequencing	KLF1 Exon 2
5' P5 FWD REV	GGG AGG AAG AGG ACG ATG A GGA CAA GGA AGC CAT AAG C	985	DNA sequencing	KLF1 Exon 2
FWD	AGG CTG AGT AAA GGG GTG TG ACC TTC AGG AGC CGC TTT CT	650	DNA sequencing	KLF1 Exon 3

Table 2.1: Primers used for the amplification of the β-LCR, β-globin locus, the HbG2 and KLF1 promoter and exons.

2.3.3 PCR of the β-globin gene, $(AT)_xT_y$ region and the gamma globin promoter

The PCR reaction mixture consisted of an appropriate buffer having the right concentration of $MgCl_2$, the four nucleotides, forward and reverse primers and genomic DNA. For the β-LCR, β-globin gene and HbG2 promoter the Thermo PCR Master Mix (2x) (Taq DNA polymerase 0.05U/µl, reaction buffer, 4nM $MgCl_2$ and 0.4nM of each dNTP) was used. The final volume of the reaction mixture used was 25µL. To each single PCR tube, 12µL of 2x Reddy mix was added together with 11µl sterile distilled water, 0.2µl of each forward and reverse primer (50pM) and 2µl of (100ng) genomic DNA.

With each batch of samples, a negative control was added. The plates were then placed into a thermal cycler. After carrying out gradient PCR to find the optimal annealing temperature for each set of primer, the following thermal profiles were used. The PCR for the β-globin LCR was carried out for 30 cycles with each cycle being optimized as follows; denaturation at 95°C

for 50 seconds, annealing at 58°C for 1 minute and elongation at 72°C for 1 minute. The PCR for the β-globin gene was carried out for 30 cycles with each cycle being optimized as follows, denaturation at 95°C for 50 seconds, annealing at 53°C for 1 minute, elongation at 72°C for 1 minute. The PCR for the HBG2 gene was carried out for 30 cycles with each cycle optimized as follows; denaturation at 95°C for 50 seconds, annealing at 61°C for 1 minute and elongation at 72°C for 1 minute. All PCR products were checked on a gel. The products of the β-globin LCR and β-globin gene were sent for sequencing as explained in section 2.5.2 while a restriction enzyme digest was performed on the products of the HBG2 PCR as explained in section 2.4.

2.3.4 PCR of the KLF1 gene

To sequence the KLF1 gene, 6 PCRs were carried out as illustrated in figure 2.1b. The Solis BioDyne 5x FirePol® Master Mix 5x Reaction Buffer B (0.4M Tris-HCl, 0.1M $(NH_4)_2SO_4$, 0.1% w/v Tween-20), 12.5nM $MgCl_2$ and 2nM of each dNTPs) was used for all the PCRs. The final volume of each reaction mixture was 30µl. To each single PCR tube, 6µl of 5x FirePol® Master mix was added together with 21µl of sterile distilled water, 0.5µl of each forward and reverse primer (10pM) and 2µl of (100ng) genomic DNA.

With each batch of samples, a negative control was added. The plates were then placed into a thermal cycler. Primer set 1 and Primer set 2 were used to amplify the promoter of the KLF1. For primer set 1 the PCR was carried out for 28 cycles with each cycle optimized as follows; denaturation at 95°C for 1 minute, annealing at 59°C for 1 minute and denaturation at 72°C for 40 seconds. For primer set 2 the PCR was carried out for 33 cycles with each cycle optimized as follows; denaturation at 95°C for 1 minute, annealing at 58.2°C for 1 minute and extension at 72°C for 1 minute 30 seconds. Primer set 3 was used to amplify exon 1, while primer sets 4 and 5 amplified exon 2 and primer set 6 amplified exon 3. For primer sets 3, 4, 5 and 6 the PCR

was carried out for 30 cycles with denaturation at 94°C for 1 minute and elongation at 72°C for 1 minute. The annealing temperature for primer set 3 was 56°C for 1 minute, that for primer set 4 was 60°C for 1 minute while that for primer set 5 and 6 was 58°C for 1 minute. All pcr products were checked on gel and sent for sequencing.

2.4 Restriction Enzyme Digest

Restriction Enzyme Digest was carried out on the HBG2 PCR product to check for the rs7482144 (XMN1) polymorphism. The XmnI Restriction Fragment Length Polymorphism (RFLP) creates a restriction enzyme recognition site which can be easily observed on agarose gel electrophoresis. The enzyme was bought from New England BioLabs Inc, United Kingdom and it recognizes the following site:

5'...GAANNNNTTC...3'

3'...CTTNNNNAAG...5'

XmnI recognition site

The reaction mixture for XmnI digest was prepared by pipetting 0.05µl of XmnI enzyme (20,000units/mL), 1µl of 1x cut smart Buffer (50nM Potassium Acetate, 20nM Tris-acetate, 10nM Magnesium acetate, 100µg/ml BSA), 6.95µl of water and 2µl of PCR product. The reaction was incubated at 37°C for 1 hour and then the enzyme was inactivated at 65°C for 20 minutes. A positive and negative control were added with every digest. The products were then seen on a 2.5% agarose gel.

2.5 PCR product Purification and DNA sequencing

2.5.1 PCR product purification

Prior to DNA sequencing, the PCR products were purified using microcon centrifugal filter devices. The aim of these devices is to concentrate and desalt DNA samples for downstream processes such as sequencing. A microcon sample reservoir was inserted into a 1.5mL vial and without touching the membrane, 20µL of PCR product were pipetted into the sample reservoir. The reservoir was sealed by attaching cap in place and the whole assembly was placed in a microcentrifuge and counterbalanced with a similar device. Centrifugation at 500 x g for 12 minutes was carried out. The sample reservoir was retained, and the filtrate was discarded. 20µl of sterile distilled water was pipetted into the sample reservoir and incubated for 5 minutes at room temperature. The sample reservoir was then inverted 180° to face down into a new 1.5mL microcentrifuge tube. The assembly was then centrifuged for 1500 x g for 5 minutes at which point the PCR product was collected at the bottom, desalted and purified.

2.5.2 DNA sequencing

The purified PCR products were sent to GATC for sequencing. Once DNA sequencing was complete, the data files were downloaded in *.ab1 format and opened using CodonCode Aligner software v5.0.1 (http://www.codoncode.com/aligner/). This software enabled easy to use interface and able to align multiple sequences at once saving time for additional testing.

2.6 High Resolution Melting Assay

High Resolution Melting (HRM) analysis is an innovative technique based on analysis of melting DNA. HRM characterizes DNA samples according to their dissociation behaviour as

they transit from double stranded (dsDNA) to single stranded DNA (ssDNA) with increasing temperature.

HRM assays were used to genotype the globin gene trans-regulators SNVs rs4671393 (Thein et al., 2008) in *BCL11A*, rs9399137, rs9494142, rs9376090 and rs9389269 in *HBS1L-MYB* (Thein et al., 2007). HRM genotyping was performed on the Rotor-Gene 6000 instrument (Qiagen, Hilden, Germany) using the EvaGreen® HRM mix from Solis Biodyne (Estonia) containing HOT FIREpol® DNA polymerase, EVvaGreen® dye, 12.5nM MgCl2, dNTPs, a buffer, BSA and a ROX control which is used to normalise well-to-well variations that might occur during quantitative real-time PCR. A 20 L reaction volume was prepared per sample and manually pipetted in 0.1mL strip tubes (Qiagen, Germany) which were loaded in a locking 72-well rotor. The following components were added in each reaction mix:

	Amount added per reaction
5X HOT FIREpol EvaGreen HRM mix (with ROX)	4 L
Forward primer (10 M):	0.5 L
Reverse primer (10 M):	0.5 L
Genomic DNA (50ng):	2.0 L
Sterile distilled water:	13 L
Final volume:	**20 L**

The list of primers used is shown is table 2.2. Three positive controls including a wild type, heterozygote and homozygous mutant sample were included in each run. A 'no template control' (NTC) was used as a negative control to exclude the presence of contamination. The thermal profile for all five genotyped SNPs was as follows:

[redacted]

95 C for 10 minutes (HOT FIREPol polymerase activation)

95 C for 10 seconds (denaturation)

60 C for 30 seconds (annealing) } 45 cycles

72 C for 10 seconds (extension)

72 C for 10 minutes (final extension)

[redacted]

Ramp from 55 C to 95 C, rising by 0.1 C with a hold of 2 seconds at each step

After completion of the run, the qPCRs were analysed to confirm that all samples were amplified, and no outliers were present (figure 2.2.1) since outliers affect the HRM results. To determine the T_m of the product the melt curve was then analysed (figure 2.2.2). The HRM data was analysed by first adjusting the regions of normalisation exactly at the start and end of the melt plot (figure 2.2.3). The generated melting curves of all samples were compared to the start and end fluorescent signal levels. This helped in the interpretation and analysis of data. The 'normalised' graph was then displayed and by looking for a shift in the melt curve (figure 2.2.4) the genotypes were identified. The known samples (controls) were inserted in the 'HRM genotypes' list and unknown samples were compared to these known samples and results were displayed in the 'difference graph' (figure 2.2.5). Genotypes of unknown samples were automatically called and displayed in a separate table accompanied with a confidence value. Samples with a confidence value below the threshold were sent for Sanger sequencing.

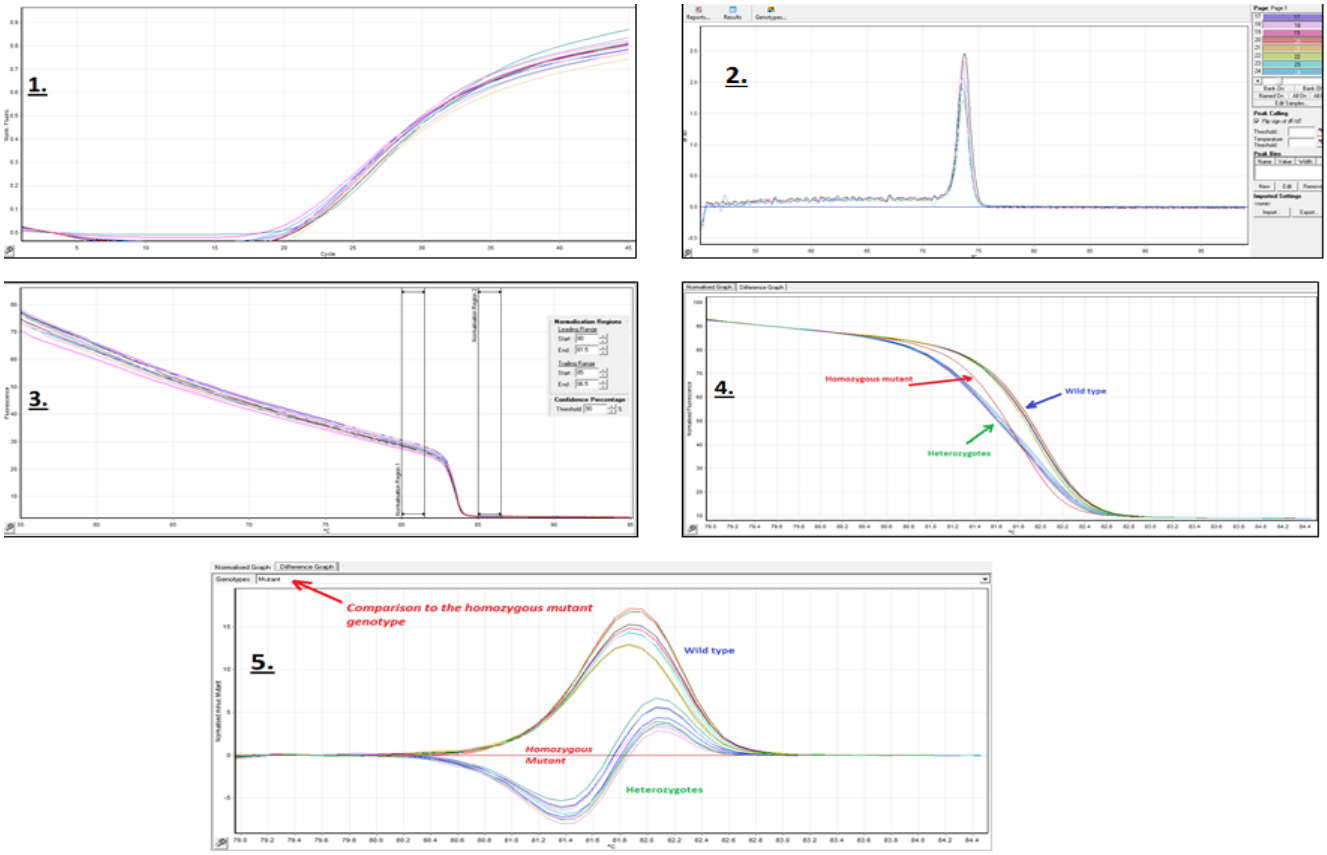

Figure 2.2: Different graphs generated and analysed during HRM analysis. Figure 2.2.1 shows quantitation analysis to determine amplification efficiency. Figure 2.2.2 shows Melt curve analysis to determine the melting temperature of the product. Figure 2.2.3: Selection of normalisation region. Figure 2.2.4 shows normalisation graph while figure 2.2.5 shows difference graph whereby unknown variants are compared to known genotypes.

Gene	Single nucleotide variation (SNV)	Primer sequence 5→3	PCR product size
MYB	rs9399137	FWD: CATCACCTTAAAGGCGGTATTGTATG REV: GATTCCACTTTCAGAACTTATCCCAAGA	148 bp
MYB	rs9494142	FWD: GTCTTTGGTCTCTCAGTCAATTCGA REV: CAACAAAAGAAAATGCAGTTATATAG	107 bp
MYB	rs9376090	FWD: GGCCAATACCATATGAAGCTAAGTCT REV: GCCTCATTTGTTCCTGGCAGTAT	70 bp
MYB	rs9389269	FWD: GCTCATATGAAGAAGACAGGCAGAA REV: AATGAAGAGTGGTGATCATTGGGT	78 bp
BCL11A	rs4671393	FWD: GCCCAGTGCTGTGGACA REV: CATGGCCAAGCTGATGAGGAT	88 bp

Table 2.2: Primers used for HRM assays

2.7 Whole genome sequencing

Whole genome sequencing (WGS) is becoming one of the most widely used application in next generation sequencing (NGS). Through WGS one can obtain the most comprehensive view of genomic information together with its associated biological implications (Cirulli & Goldstein, 2010). Whole genome sequencing was carried out on 12 Fam FI family members and the four family members of Fam F2. Genomic DNA was extracted from peripheral blood by a modified salting out technique and the concentration and purity were determined using the Thermo Scientific NanoDrop^{TM} while DNA integrity was checked by gel electrophoresis.

Whole genome sequencing was carried out at Complete Genomics, California USA using DNA Nano array protocol (figure 2.3). The first step involved the fragmentation of the sample DNA followed by ligation to a common adaptor set for clonal amplification and sequencing. The resulting circular fragments were then copied and compacted into palindrome-promoted coils of single-stranded DNA known as DNA nanoballs (DNBs). The DNBs were then absorbed onto a photolitohographically etched, surface modified silicone substrates with grid patterned arrays for DNA binding. High accuracy combinatorial probe anchor ligation (cPAL) sequencing chemistry is then used for DNA sequencing. (Drmanac et al., 2010).

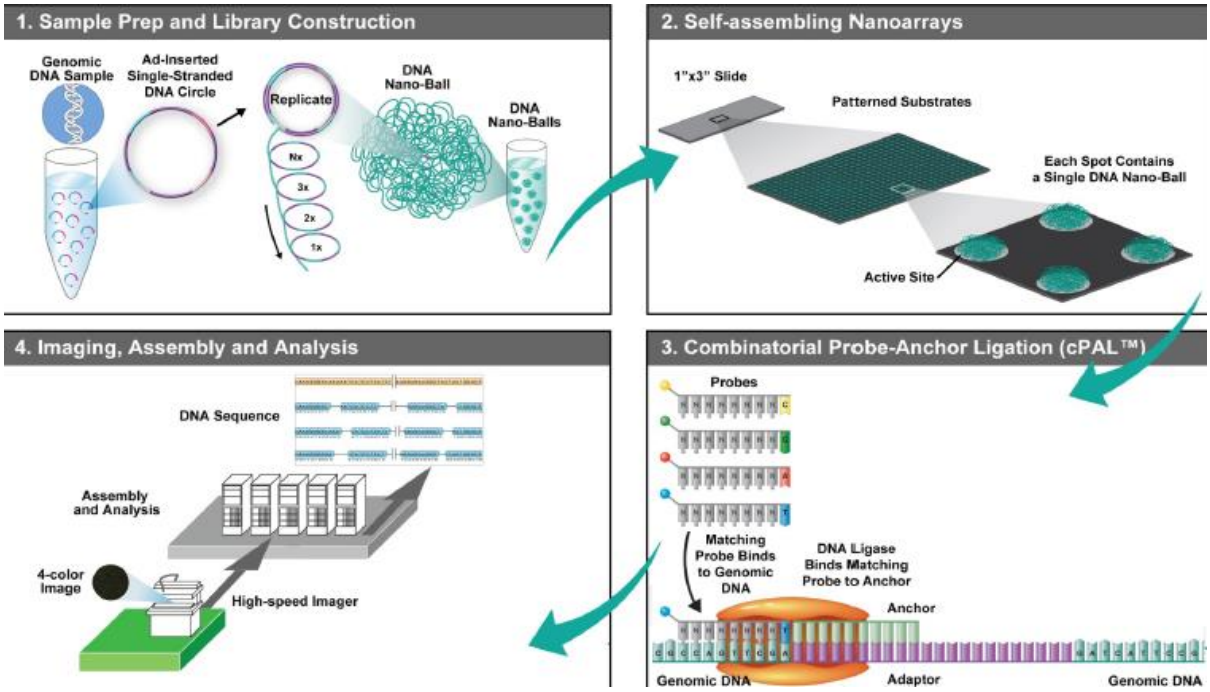

Figure 2.3: Diagram showing the process involved in Whole Genome sequencing by Complete Genomics

The WGS data was then aligned to the hg19 reference genome by Complete Genomics (CG) using their proprietary software. Preliminary Bioinformatics analysis on each genome was done by CG. A list of variants with information about their genomic position and occurrence in each genome was provided for each individual. Variants were represented as 11 for homozygous, 01 as a heterozygous and when a variant was not found it was marked as 00. Nucleotides which were not called by the sequencers or filtered out by quality checking were defined as No Calls (N).

Filtering techniques were used to analyse the 10 sequenced genomes from Fam F1 and the 4 sequenced genomes from Fam F2. In Fam F1, the 10 family members were divided into four categories according to their HbF levels. Subjects II-5 and II-1 who had an HbF of 2480 and 2465mg/dL were labelled as high HbF, Subjects III-4, III-6, III-15 and III-16 with an HbF between 1135 and 411 mg/dL were labelled as medium HbF, subject III-1 with an HbF of 281mg/dL was labelled as intermediate HbF while Samples III-17, II-6 and II-2 which an HbF

between 70 and 38 mg/dL were the controls. In this scenario filtering checked for heterozygote and homozygote mutations inherited with the p.K288X mutation.

In Fam F2, the HbA_2 was used as a biomarker for filtering. Subjects I-1 and II-2 with an HbA_2 of 3.4 and 3.3% of were labelled as cases while subjects I-2 and II-1 with an HbA_2 of 2.7 and 3.0% were labelled as controls. In this scenario filtering checked for heterozygote and homozygote mutations inherited with the p.K288X mutation

2.8 Culturing of Human Erythroid Progenitor cells (HEPs)

To investigate the phenotypic variability of KLF1 haploinsufficiency in Fam F1 and Fam F2 we took advantage of primary erythroid cultures of individuals carrying the p.Lys288Ter mutation and of other family members wildtype for this mutation. Figure 2.4a and figure 2.4b shows the family members selected for cell culturing. We performed a combined analysis of gene expression (RNA-sequencing), chromatin accessibility assays (ATAC-sequencing) and promoter activity tests to explore the molecular basis of the heterogeneity in HbF levels displayed by the p.Lys288Ter heterozygote individuals. HEPs were cultured from family member II-5, II-6, III-15, III-16 and III-17 from Fam F1 and from subjects I-1, I-2 and II-3 from Fam F2 together with another two independent controls who were wildtype for the p.Lys288Ter mutation.

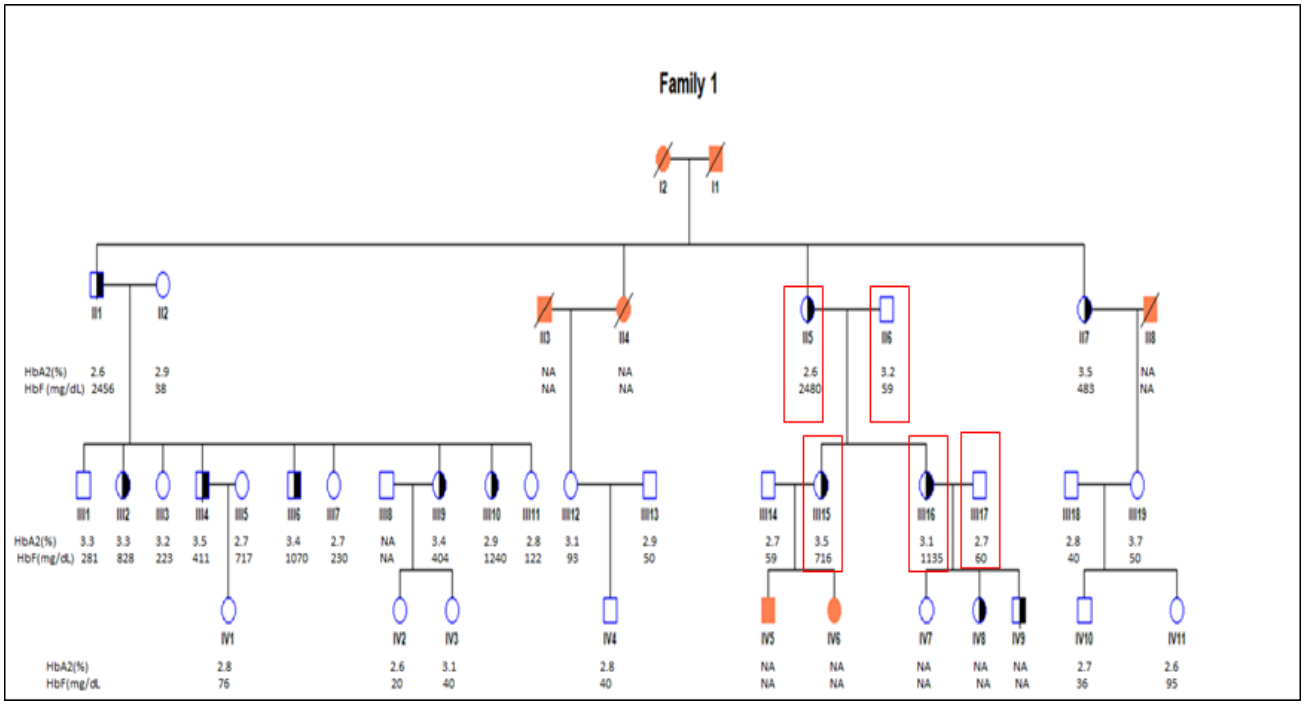

Figure 2.4a: Family tree of Fam F1. Cell culturing, RNA sequencing and ATAC-sequencing was carried out on family members with a red box.

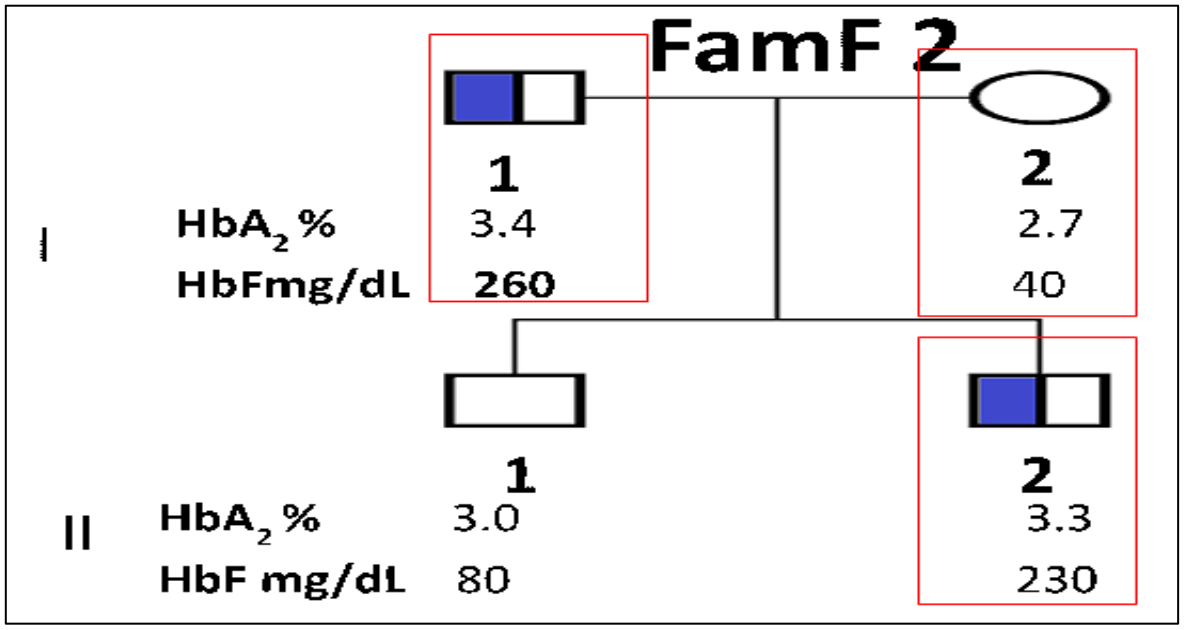

Figure 2.4b: Family tree of Fam F2. Cell culturing, RNA sequencing and ATAC-sequencing was carried out on family members with a red box.

Human erythroid progenitor cells (HEPs) were cultured as described in Leberbauer et al., (2005) & Van der Akker et al., (2010) in the presence of recombinant human erythropoietin

(Epo) (1unit/mL), recombinant human Stem Cell Factor (SCF) (50 ng/mL) and dexamethasone (5 x 10-7 M; Sigma-Aldrich, USA), The erythroblast were counted with an electronic cell counter (CASY-1, Scharfe System, Reutlingen, Germany) as explained in section 2.15.2.2.

2.8.1 Preparation of Culture Medium

HEPs were cultured in Cell-Quin Medium (home-made media gently donated by Emile van den Akker, Department of Hematopoiesis, Sanquin Research, Amsterdam) supplemented with 1% L-glutamine (Sigma-Aldrich, USA), 2 U/ml Erythropoietin (EPO), 50ng/ml Stem Cell Factor (SCF), 1 x 10^3 Dexamethasone (Sigma-Aldrich, USA) and 1% P/S. The media was kept refrigerated at 4°C until further use.

2.8.2 Isolation of White Blood cells from Buffy Coat

Erythroid cultures were performed from freshly isolated peripheral blood monocular cells in a two-step culture system. Using sterile conditions, 10 EDTA tubes were centrifuged at 1200 x g for 25 minutes at room temperature without brake. The resultant supernatant which consisted of platelet-rich plasma was discarded while the buffy coat was aspirated and transferred to a 25mL microcentrifuge tube. To the isolated buffy coat, 9ml of PBS were added followed by the gentle addition of 2.5ml of Histopaque (density: 1.077g/mL at 25°C) (Sigma-Aldrich, USA) below the cells. Centrifugation at 1200 x g for 25 minutes at room temperature without brake was carried out. This allowed the separation of the erythrocytes and the mononuclear cell fraction. The mononuclear cell fraction was collected, transferred to a labelled 25mL microcentrifuge tube and washed with PBS and centrifuged at 450 x g for 7 minutes. This step was repeated 3 times.

After the third time the cell pellet was suspended in complete Cellquin medium supplemented with 1% L-glutamine (Sigma-Aldrich, USA), 2 U/ml Erythropoietin (EPO), 50ng/ml Stem Cell Factor (SCF), 1 x 10^3 dexamethasone (Sigma-Aldrich, USA), 1% P/S. The T75 flasks were incubated at 37°C and 5% CO_2 and 98% humidity. The erythroid cells were kept in this medium for 10 days and the medium was refreshed every other day. Cells were maintained at $0.7x10^6$ cell/ml (Casy® Model TCC, Scharfe System GmbH; Reutlingen, Germany). In the second step erythroid differentiation was induced by eliminating SCF and Dexamethasone from the culture and the cells were kept in medium with high EPO (10 U/ml) and 5% human plasma supplemented by Blood Bank, Mater Dei Hospital. At the start of differentiation, cells were put at $2x10^6$ cells/ml and the medium was replenished every other day.

2.8.3 RNA extraction

Samples were collected from proliferating (T0) and differentiating (T48) erythroid cultures. RNA was isolated using TRIzol® RNA isolation reagent (ThermoFisher Scientific, United States). Cells were centrifuged at 650 x g for 8 minutes; the supernatant was aspirated, and the cells were washed with PBS. This step was repeated another 2 times. Tubes were vortexed and 100µL of chloroform was added, vortexed for 15 seconds and left at room temperature for 3 minutes. The tubes were centrifuged for 15 minutes at 15,000 x g at 4°C and after centrifugation the upper phase was transferred to a new labelled tube. This was followed by the addition of 250µL isopropanol and the tubes were mixed well and left at room temperature for 10 minutes. The tubes were centrifuged for 15 minutes at 15,000 x g at 4 °C, the supernatant was discarded, and the RNA was washed twice with 75% ethanol. The tubes were left to dry from the ethanol and the RNA was suspended in 20µL of RNAse free water. The quality and quantity of the RNA was checked using the Nanodrop as described in section 2.2.3 and stored at -80°C.

2.8.4 cDNA synthesis by Reverse Transcriptase

Reverse transcription involves the synthesis of a complementary strand of DNA produced from RNA and by using random primers that specifically hybridise to the poly-A-tail of mRNAs it is transcribed into singe stranded cDNA.

For reverse transcription the SuperScript™ II Reverse Transcriptase (Invitrogen, Life Technologies, USA) was used. The RNA samples were taken out for the -80°C freezer and put on ice. In a nuclease-free microcentrifuge tube 1µL of Oligo(dT)$_{12-18}$ (500µg/mL), 1ng of total RNA and 1µL dNTP Mix (10 nM each) were added. The mix was topped up to 12µL with sterile distilled water and heated to 65°C for 5 minutes using the Applied Biosystems® GeneAmp® PCR System 9700 Thermal Cycler followed by a quick chill on ice. The contents of the tube were collected by brief centrifugation and 4µL of 5X First-Strand Buffer (250nM Tris-HCL, pH 8.3 at room temperature, 375nM KCl; 15nM $MgCl_2$), 2µL of 0.1 M DTT and 1µL of RNaseOUT™ (40 units/µL) were added to the mixture. The contents of the tube were mixed gently and incubated at 42°C for 2 minutes and 1µL (200 units) of SuperScript™ RT were added. The mixture was mixed gently by pipetting up and down and incubated at 42°C for 50 minutes followed by heating at 70°C for 15 minutes to inactivate the reaction. The cDNA was stored at -20°C.

2.9 Real-Time quantitative PCR

The Real-time quantitative PCR (RT-qPCR) technique was used to amplify and detect the expression of the gamma globins, Beta, KLF1 and BCL11A in HEPs from proliferating (T0) and differentiating (T48). Real-time qPCR was performed on thermocycler (ABI 7300, Applied Biosystems, USA) using the 5x HOT FIREPol® EvaGreen® qPCR Mix Plus (ROX) from Solis Biodyne (Estonia) containing HOT FIREPol® DNA Polymerse, EvaGreen®

qPCR buffer, 12.5nM MgCl2, dNTPs, EvaGREEN® dye and ROX dye. A 20µl reaction volume was prepared per sample and manually pipetted in a 96-well plate (table 2.3).

Reagent	Volume (µL)
5X HOT FIREPol® EvaGreen® qPCR Mix	5
Oligonucleotide forward primer (10pmol/µL)	0.5
Oligonucleotide reverse primer (10pmol/µL)	0.5
Nuclease free water	13
cDNA	1
Total volume	**20**

Table 2.3: Reaction mixture for RT-qPCR

Name	Sequence	Purpose
GAPDH-FWD	GTGAAGGTCGGAGTCAACG	q-PCR
GAPDH-REV	TGAGGTCAATGAAGGGGTC	q-PCR
GAMMA-FWD	GATGCCATAAAGCAGCTGGATG	q-PCR
GAMMA-REW	TTGCAGAATAAAGCCTATCCTTGA	q-PCR
BETA-FWD	AACTGTGTTCACTAGCAACCTCAA	q-PCR
BETA-REV	GAGTGGACAGATCCCCAAAGGA	q-PCR
HBG1/2-FWD	AGGTGCTGACTTCCTTGGG	q-PCR
HBGI/2-REV	GGGTGAATTCTTTGCCGAA	q-PCR
KLF1-FWD	CCACAGCCGAGACCGCCTTGACC	q-PCR
KLF1-REV	CTCTCATCGTCCTCTTCCTCCC	q-PCR
BCL11A-FWD	CGAGCACAAACGGAAACAATG	q-PCR
BCL11A-FWD	GATTAGAGCTCCATGTGCACG	q-PCR

Figure 2.4: Primers used for RT-qPCR

The list of primers used is shown in table 2.4. cDNA previously obtained from reverse transcription was diluted with 80µL nuclease-free water and 1µL was pipetted in the respective wells. Each sample reaction was repeated three times. The 96-well optical plate

was sealed with an adhesive optical cover and care was taken to avoid contact with the optical surface. The plate was inserted in the RT-PCR thermocycler (ABI 7300, Applied Biosystems, USA) and the program with the appropriate PCR conditions (Table 2.5) was initiated.

Step	Temperature (°C)	Time (seconds)	Number of cycles
Hot start	95	600	1
Denaturation	95	30	40
Annealing	60	30	
Dissociation	95	15	1
Analysis	60	30	

Table 2.5: Thermocycler Temperature Profile for RT-PCR. Denaturation, annealing, extension and fluorescent detection steps were repeated for a total of 40 cycles after which they were followed by dissociation analysis.

The data obtained was analysed using the SDS software v1.3.1 (Applied Biosystems, USA). Ct-values were normalized against housekeeping genes 18S and GAPDH. The final value obtained ($\Delta\Delta C_T$) was used to compare the mRNA expression between p.Lys288Ter heterozygotes and wildtype at T0 and T48.

2.10 RNA sequencing

RNA-sequencing combines the aspects of transcript identification and the quantification of gene expression in a single high-throughput sequencing assay (Conesa et al. 2016). A typical RNA sequencing experiment (figure 2.5) involves the conversion of RNA (total or fractionated, such as poly(A)+) to a library of cDNA fragments with adaptors attached to one or both ends. Each amplified molecule is then sequenced in a high-throughput manner so that short sequences are obtained. The short sequences can be either from one end (single-end sequencing) or both ends (pair-end sequencing). The generated reads are typically between 30 – 400bp depending on the DNA sequencing technology used. Following sequencing, the

resulting reads are either aligned to a reference genome, reference transcripts, or else they can be assembled de novo without the genomic sequence. These produce a genome-scale transcription map consisting of both the transcriptional structure and level of expression of each gene (Wang et al., 2009).

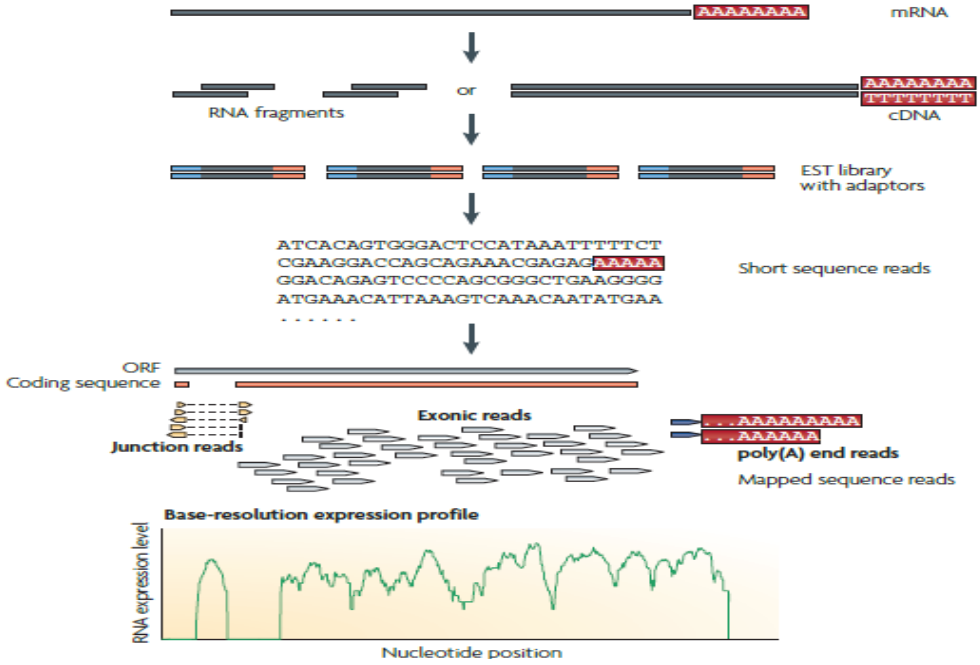

Figure 2.5: RNA-sequencing experiment. Long RNAs are first converted into a library of cDNA fragments. Sequencing adaptors shown in blue are subsequently added to each cDNA fragment and by using high-throughput sequencing technology a short sequence from each cDNA is obtained. The resulting sequence reads are then aligned with the reference genome and classified in three types (i) exonic reads, (ii) junction reads and (iii) poly(A) end reads. These three types of reads are then used to generate a base-resolution expression profile for each gene. Adapted from RNA-Seq: a revolutionary tool for transcriptomics by Wang, Gerstein & Snyder (2009), *Nature Review Genetics* (10), 59.

RNA sequencing was performed on Subjects II-5, II-6, III-15, III-16 and III-17 from Fam F1, and subjects I-2 and II-2 from Fam F2 together with unrelated controls. Erythroid cells of subject I-1 from Fam F2 died in culture, so it was eliminated. Samples from p.Lys288Ter carriers and controls were collected at low hemoglobinzed pro-erythroblast stage (T0) and after 48 hours of terminal differentiation (T48). RNA sequencing libraries were prepared using TruSeq Stranded mRNA library kit (Illumina Cambridge Ltd., UK) according to manufacturer's instructions at Erasmus Medical Centre (Rotterdam, Netherlands). Sequencing was performed on the Illumina HiSeq2500 sequencer. Samples were sequenced to a depth of

15x10^6 x 50bp single end reads. RNA sequencing reads were mapped to GRCh38 using STAR. Low expressed mRNA (<3 counts per million mapped reads in less than 4 samples) were filtered prior to differential expression analysis using EdgeR package. Analysis was performed at Erasmus MC.

2.11 Assay for Transposase Accessible Chromatin with high- throughput sequencing

Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-sequencing) combines next-generation sequencing with new biochemical techniques to enable genome-wide investigations of a broad range of chromatin phenomena. ATAC-seq was described by Buenrostro et al., (2013) where it was used to identify regions of open chromatin, identify nucleosome-bound and nucleosome free positions in regulatory regions and to infer the position of DNA-binding proteins using foot printing. ATAC-seq (figure 2.6) uses hyperactive Tn5 transposases which integrates its adaptor payload into regions of accessible chromatin, while steric hindrance in less accessible chromatin makes transposition of Tn5 less possible. Therefore, amplifiable DNA fragments suitable for high-throughput sequencing are preferentially generated at locations of open chromatin

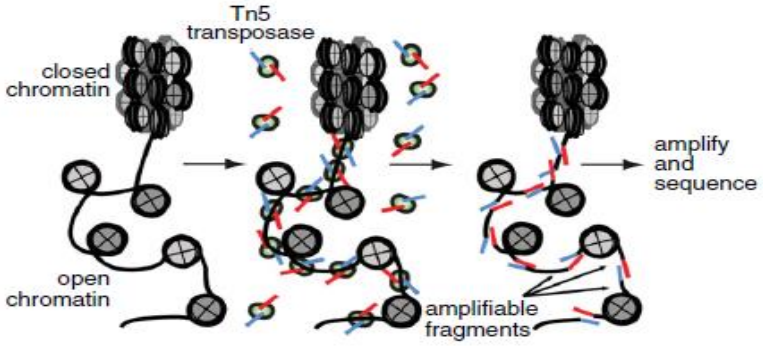

Figure 2.6: ATAC-sequencing reaction schematic. Transposases (green) are loaded with sequencing adaptors (red and blue), and these inserts only in regions of open chromatin (between nucleosomes in grey) and generates sequin-library fragments that are PCR-amplified. Adapted from Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position by Buenrostro (2013), Nature Methods, 10(12), 1214.

ATAC-seq was carried out at Erasmus MC. The method was optimized and carried out on cultured HEPs of subjects II-5, II-6, III-15, III-16 and III-17 from Fam F1, and subjects I-2 from Fam F2 together with unrelated controls. The procedure consisted of three major steps:

(i) Preparation of nuclei – around 50,000 cells were centrifuged at 500 x g for 5 minutes and washed using 50µl of cold 1x PBS followed by another centrifugation at 500 x g for 5 minutes. Cells were then lysed using cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2 and 0.1% IGEPAL CA-630). The nuclei were then centrifuged at 500 x g for 10 minutes using a refrigerated centrifuge.

(ii) Transpose and purify – The pellet was then re-suspended in the transposase reaction mix (25 µL 2× TD buffer, 2.5 µL transposase (Illumina) and 22.5 µL nuclease-free water) and incubated at 37°C for 30 minutes. The samples were then purified using a Minelute Purification kit (Qiagen).

(iii) PCR – The library fragments were then amplified using 1× NEBnext PCR master mix and 1.25 µM of custom Nextera PCR primers and conditions described by Buenrostre (2013). The libraries were then purified using a PCR clean-up kit (Qiagen) and quantified. The libraries had a concentration of approximately 30nM in 20µL. Libraries were then amplified for a total of 12 cycles and sequenced on the HiSeq 2500.

Reads were mapped using Bowtie2 (Langmead et al., 2012). Low abundant regions (<1 count per million mapped reads in less than 3 samples) were filtered prior to differential analysis in EdgeR. Analysis was performed at EMBL.

2.12 Expression vectors

For promoter functional studies several expression vectors were used.

2.12.1 pGL4.10 [luc2]Vector

The pGL4.10 *[luc2]* (Promega, USA) encodes for the luciferase reporter gene *luc2* (Photinus pyralis) and is specially designed for high expression and reduced anomalous transcription. As seen in figure 2.7, it is a basic vector with no promoter, however it contains a multiple cloning region (MCS) to allow the cloning of the promoter of choice. The expression of luciferase by the *luc+* gene is therefore under the transcriptional control of the DNA fragment which is cloned into the MCS, since this acts as a promoter to the *luc+*. The vector also has an ampicillin resistance (-lactamase) gene, which is useful for positive selection of transformed *E.Coli* during the incubation stages. In this study, this vector was used as the experimental reporter vector, where the multiple cloning region was used to clone the KLF1 promoter, both with the wildtype sequence and with the mutations of interest.

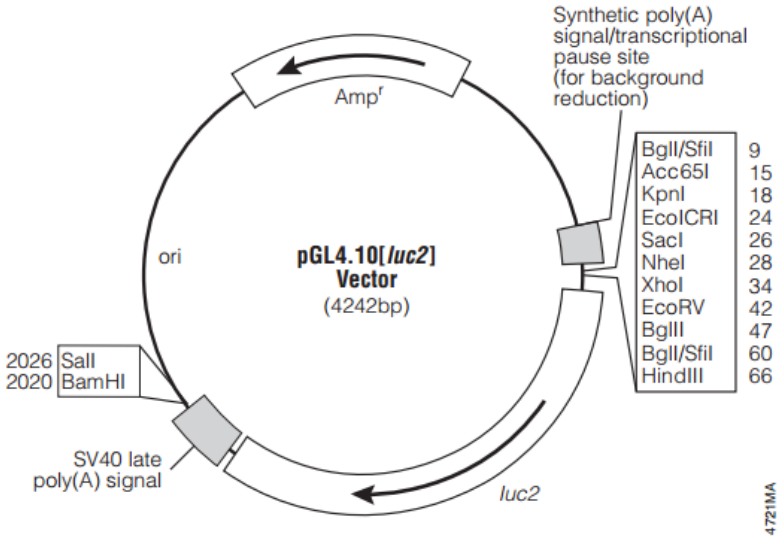

Figure 2.7: Schematic structure of the pGL4.10[luc2] vector

2.12.2 pRL-TK Vector and pRL SV40 Vector

The pRL-TK Vector (Promega, USA) and the pRL SV40 Vector (Promega, USA) contains a cDNA (Rluc) encoding *Renilla* luciferase which was originally cloned from the marine organism *Renilla reniformis* (Lorenz et al., 1991). The *Renilla* luciferase is a 36kDA monomeric protein that does not require any modification for activity, therefore immediately following translation, the enzyme may function as a genetic reporter gene (Matthews et al., 1977).

The pRL-TK vector was used as an internal control reporter vector. This vector was co-transfected with the firefly luciferase to provide a baseline luciferase reading against which the firefly luciferase activity can be normalized. The pRL SV40 vector (figure 2.8) was used as a positive control. It contains the SV40 enhancer together with early promoter elements to provide high-level expression of Renilla luciferase when co-transfected in mammalian cells. Both vectors also have an ampicillin resistance (-lactamase) gene. The pRL-TK vector and pRL SV40 vector were generously provided by Prof Anthony Fenech, from the Department of Clinical Pharmacology and Therapeutics, University of Malta.

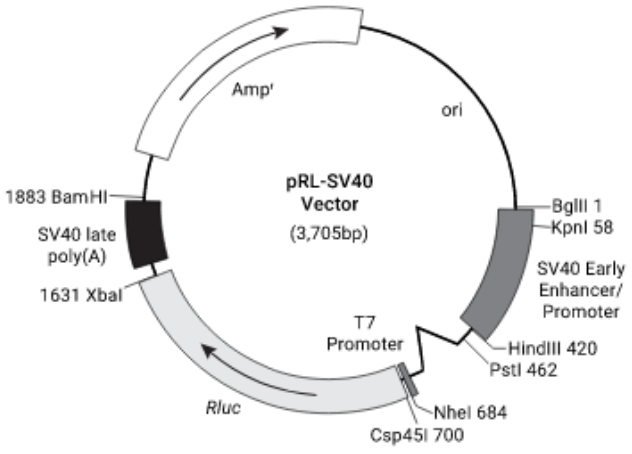

Figure 2.8: Schematic structure of pLR SV40 plasmid

2.12.3 pMAX GFP vector

pMAX GPF vector is a green fluorescent protein derived from copepod *Pontellina* sp. The vector backbone contains an early promoter of cytomegalovirus (PCMV), a chimeric intron for enhanced gene expression, the pUC origin of replication for propagation in *E.coli* and the kanamycin resistant gene. The GFP is located between the CMV promoter and the SV40 polyadenylation signal (SV40 poly A). This vector was used as a positive control to optimize and monitor transfection efficiency in our cells of interest.

2.13 Sub-cloning of synthetic DNA in vector

Attempts were made several times to Sub cloning the desired wildtype *KLF1* promoter sequence in the pGL4.10[*luc2*] vector, but optimization was proving to be difficult and lengthy, and in order not to compromise the project, Mutagenex Inc (USA) was contacted for the sub-cloning of synthetic DNA in the pGL4.10[l*uc2*] Vector (Promega). The sequences of the wildtype *KLF1* promoter together with the sequence of the mutation of interest namely -1133C>A (rs112943513), -251C>G (rs3817621), -1133C>A & -251C>G (rs112943513 & rs3817621), -145G>A (rs79334031), -154T>C (rs372651309), -43G>A (rs372359976) and the -102T>G (rs548543206) were sent to Mutagenex. Mutagenex carried out the synthesis of the DNA sequences and then they were sub cloned into the pGL4.10[*luc2*] vector between *Nhel* and *EcoRV*. The vectors containing the wildtype sequence and the mutant sequence were shipped to the University of Malta on agar plates (Appendix C).

Primer	Sequence
Vector Forward	5’ - TGAATCGATAGTACTAACATACGC
Internal Forward	5’ – CCAGAGCCAGGGCACTGGTC
Vector Reverse	5’ – GTCCCGTCTTCGAGTGGGTAG

Table 2.6 - Sequences of primers used for basic sequencing

Once we received them, sequencing of the vectors was carried out to confirm the wildtype and mutant sequences using 3 primers, as shown in table 2.6.

2.14 Bacterial Transformation

Standard bacterial cloning techniques were used to carry out amplification of all expression vectors, namely pRL-TK Vector and pRL SV40 Vector, the pGL4.10 [*luc2]* containing the wildtype sequence and the mutant sequence and the pMAX GFP vector.

2.14.1 Preparation of competent DH5α E.coli cells

The DH5α *E.coli* bacterial cells (Invitrogen, UK) were used as competent cells for the introduction of the circular plasmids. All solutions were prepared (Appendix D). Using aseptic techniques, a 5µl wire loop, DH5α cells were streaked on a Luria Broth (LB) Agar plate with no antibiotics. The plate was incubated at 37°C overnight. The following day, a single colony was picked and transferred to a 10ml Lb broth in a sterile universal container. The Lb broth was incubated at 37°C and 250rpm overnight. Following incubation, 1L conical flask containing 100ml Lb broth was inoculated with 2ml of the overnight 10ml broth and incubated at 37°C and 250rpm. Every hour, during incubation 1ml of this broth was aseptically removed and the turbidity was measured at 600nm. The Lb broth was used as blank. The procedure was repeated until the absorbance reached 0.6.

When the absorbance reached 0.6, the flask was removed from the incubator and the culture was transferred to two 50ml sterile centrifuged tubes and left for 10 minutes on ice. The tubes were then centrifuged at 3000 x g for 20 minutes at 4°C. The supernatant was decanted, and the bacterial pellets were drained as much as possible by inverting and tapping repeatedly on an absorbent tissue. Each pellet was suspended in 10ml of ice-cold 100mM $MgCl_2$ and placed on ice for 10 minutes. The suspension was centrifuged again at 3000 x g for 10 minutes, the

pellets were drained as before and were then resuspended in 10ml ice-cold 100mM $CaCl_2$ and left for 10 minutes on ice. The tubes were centrifuged again as before, and the cell pellet was then resuspended in 1ml of ice-cold $CaCl_2$ containing 15% glycerol. 200µl of the cell pellet together with $CaCl_2$ containing 15% glycerol were aliquoted into prechilled microcentrifuge tubes and the tubes were frozen immediately at -80°C.

2.14.2 Bacterial transformation of vectors

Amplification and cloning of all vectors were carried out using standard bacterial cloning techniques. Circular plasmids were introduced into competent DH5α using the heat shock method and then allowed to grow on an agar medium. After, DNA extraction was carried out to extract and purify plasmid DNA. The purified plasmid DNA was transfected into mammalian cells.

Using 50µl of DH5α per transformation, the required volume of competent bacterial and vectors were allowed to thaw on ice for 15 minutes. In a 1.5ml microcentrifuge tube, 1µl of plasmid DNA was added to each aliquot of DH5α cells. The mixture was gently swirled and kept on ice for 30 minutes. Heat shock at 42°C for 1 minute was carried out in a heat block (QDB2, Grant, UK). The microcentrifuge tubes were then placed again on ice for two minutes after which 1ml of LB broth (Sigma-Aldrich, USA) at 37°C was added to each tube. The tubes were placed in a shaking incubator (Innova® 42, New Brunswick Scientfiic) at 37°C for 1-hour swirling at 250rpm. After 1 hour, the mixture was added to 6ml LB broth together with 6µl of Kanamycin in the case of GFP while for the other vectors 6µl of ampicillin were added. The 50ml tubes were placed again in a shaking incubator at 37°C for 24 hours swirling at 250rpm. After 24 hours, 100µl of the mixture was added to 50ml LB broth together with 50µl kanamycin or ampicillin in a conical flask and left overnight in the shaking incubator at 37°C.

The next day the mixture was poured in 50ml centrifuge tubes and centrifuged at 2000 x g for 45 minutes at 4°C. The supernatant was decanted, and the cell pellet was kept at -20°C until plasmid extraction was carried out.

2.14.3 Plasmid extraction

For plasmid extraction the AccuPrep® Plasmid Mini Extraction Kit (Bionner, South Korea) was used. Each cell pellet was resuspended in 2500µl Buffer 1 (Resuspension Buffer). To completely resuspend the cells, vortexing was carried out. The mixture containing the resuspended cells was divided into 10 labelled microcentrifuge tubes, and each tube contained 250µl of resuspended cells. 250µl of Buffer 2 (Lysis) was added to each tube and the tubes were mixed by inverting gently 3-4 times. At this stage it is important that one does not vortex the tubes, because vortexing can result in shearing of genomic DNA. Buffer 3 (Neutralization buffer) was added and the tubes were mixed gently 3-4 times. Following the addition of the lysis buffer and neutralization buffer to the lysate, the chromosomal DNA and cell debris formed an insoluble aggregate.

Centrifugation at 18,000 x g at 4°C for 10 minutes was carried out in a micro-centrifuge. After centrifugation, the white protein aggregate appeared at the bottom of the tube and the cleared lysate was transferred to the DNA binding column tube. This clear lysate contains a chaotropic salt, originating from neutralization buffer, which helps the binding of the plasmid DNA on the membrane surface. This was followed by another centrifugation at 18,000 x g for 1 minute. The flow-through was poured off and the DNA binding filter column was re-assembled with the 2.0ml collection tube. 700µl of Buffer 4 (Washing buffer) was added to the DNA binding column tube and centrifugation at 18,000 x g for 1 minute was carried out. This eliminated any salt and precipitates. The flow-through was removed and the DNA-binding filter column was

re-assembled with the 2.0ml collection tube. To remove any residual ethanol, the DNA-binding filter column was dried by additional centrifugation at 18,000 x g for 1 minute. The DNA binding filter column was transferred to a new 1.5ml microcentrifuge tube and 70µl of pre-warmed (about 60°C) elution buffer was added, and the plasmid DNA was eluted by centrifugation at 18,000 x g for 1 minute.

2.15 Cell Culturing

For this thesis 3 cell lines; the Human Embryonic Kidney cells (HEK 293T), K562 cells and HUDEP-2 cells were used together with Human Erythroid Progenitor (HEP) cells.

2.15.1 HEK293T Cells

The HEK293 cell line was derived in 1973 by exposing the human primary embryonic kidney cell culture of an aborted embryo to the mechanically sheared DNA of adenovirus type 5 (AD5) (Graham et al., 1977). A 4-kbp adenoviral genome fragment which encodes for the E1A/E1B proteins was integrated in chromosome 19. These proteins interfere with the cell cycle control pathways and counteract apoptosis (Louis et al., 1997). Mainly the two derived cell lines are (i) 293T cells which expresses a temperature-sensitive allele of the SV40 T antigen, (ii) 293S cell line which is adapted to grow in suspension. For this study we cultured the 293T cells which were maintained in Dulbecco's Modified Eagle's Medium (DMEM) (Sigma Aldrich, USA, D5648), supplemented with 10% foetal bovine serum (FBS) (Sigma Aldrich, USA) and 1% P/S (Sigma Aldrich, USA). HEK293T cells were cultured in T-75 flasks at 37°C with 5% CO_2 and 96% humidity.

2.15.1.1 Preparation of DMEM

The DMEM was available in powder form and 13.4g were dissolved in 1 litre of distilled water. 3.7g of sodium hydrogen carbonate were also added. The mixture was stirred until all components were completely dissolved. The medium was sterilized immediately by filtering through a 0.22µM sterile filter (Millipore, USA) which was maintained under negative pressure. The media was stored at 4°C for further use.

2.15.1.2 Passaging of HEK293T cells

The cells were checked under the light microscope to determine an approximate degree of confluency and morphology. The condition of the adherent cells was continuously monitored to ensure that there was enough space and medium to support further growth. When the cell confluency exceeded 80%, the HEK293T cells were split and supplied with fresh media for further proliferation.

The spent culture medium was removed from the T-75 flask and placed in a 15mL falcon tube. Then, 4mL of 1X trypsin- Ethylenediaminetetraacetic (EDTA) solution (Sigma-Aldrich, USA; 5mL 10X trypsin/EDTA with 45mL 1X PBS) were added to the side of the T-75 flask. The flask was rocked gently to get complete coverage of the monolayer, followed by a short period of incubation at 37°C. The cells were observed under the microscope to see that they were completely detached and 5mL of the spent medium was added to the flask for the deactivation of trypsin. The cell suspension was transferred in a 15mL tube and centrifuged at 450 x g for 5 minutes. The supernatant was discarded and 12mL of supplemented DMEM was used to re-suspend the cell pellet. The cell culture flask was placed in the incubator at 37°C with 5% CO_2 and 96% humidity.

2.15.1.3 Counting of HEK293T cells

For HEK293T cells, cell counting was carried out using a haemocytometer. Following trypsinization and centrifugation, 100µL of the re-suspended cells were placed in a micro-centrifuge tube together with 100µL of 0.4% trypan blue solution (Sigma Aldrich, USA). A clean haemocytometer (Hawksley, United Kingdom) was covered with a clean cover slip and a small drop of the stained, diluted cell suspension was applied to the surface of the haemocytometer at the edge of the coverslip. Each chamber was allowed to fill by capillary action. The haemocytometer was placed on the stage of the light microscope and the cells present in the four outside squares of the haemocytometer (coloured in red) (figure 2.9) were counted.

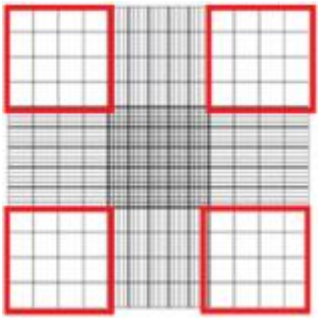

Figure 2.9: Haemocytometer grid. The cells present in the red boxes were counted.

The number of cells counted in the four boxes were added and the value obtained was then divided by 2 (dilution factor of trypan blue) (x 10^4cells/mL). The value obtained was then used to measure the number of cells required per well for every experiment.

2.15.2 K562 Cells

K562 cell line established by Lozzio & Lozzio (1975) are immortalized erythroleukemia cells derived from the pleural aspirate of a patient with chronic myeloid leukaemia (CML). This cell line is characterized by aneuploidy and dysregulated constitutive expression of the BCR-ABL

due to the presence of Philadelphia chromosome (t9:22). The cells are bipotential, meaning they can be induced to express either erythroid or megakaryocytic properties, depending on the inducer.

K562 cells were maintained in Roswell Park Memorial Institute medium (RPMI) 1640 Medium (ThermoFisher Scientific, USA), supplemented with 10% FBS (Sigma Aldrich, USA) and 1% P/S (Sigma Aldrich, USA). K562 cells were cultured in T-75 flasks at 37°C with 5% CO_2 and 96% humidity.

2.15.2.1 Passaging of K562 cells

The cells were checked under the light microscope. When the cell culture reached approximately 0.7 - 0.8 x 10^6 cells/ml, the cells were transferred to a 15ml tube and centrifuged at 290 x g for 5 minutes. The supernatant was decanted, and the cell pellet was resuspended in complete RPMI 1640 and split into two T-75 flasks. The flasks were then incubated at 37°C with 5% CO_2 and 96% humidity.

2.15.2.2 Counting of K562

K562 cells were counted using an electronic cell counter CASY-1 (Scharfe System, Reutlingen, Germany). Casy® - Technology combines the Resistance Measurement Principle with Digital Pulse Processing (DPP) to assess the cell concentration, viability and diameter. Measurement was performed by suspending 100µL of K562 cells in 10mL CASY® ton which is an electrolyte. The cells are then aspirated through a measuring pore at a constant speed. During the measurement process, a pulsed low voltage field with 1 MHz is applied to the measuring pore via two platinum electrodes. While the cells are passing through the measuring pore, they displace a quantity of electrolyte corresponding to their volume. Living intact cells are usually considered isolators and therefore, during their passage through the measuring pore,

an increase level of resistance is achieved. On the other hand, the cell membrane of dead cells no longer acts as an electrical barrier, therefore they are recorded by the size of their nucleus. The resulting electrical signals generated by the cell as it passes through the measuring pore, are analysed by amplitude, pulse width, course of time and resulting pulse area. The analysed pulse areas of cell signals are cumulated and assigned in a calibrated multi-channel analyser that has 512,000 differentiated size channels.

2.15.2.3 Addition of Hemin to K562 cells

Hemin was added to the K562cells to induce differentiation and increase the expression of *KLF1* (Addya et al., 2004). Hemin (Sigma-Aldrich, USA) was prepared by dissolving 25mg in 1ml of 1.4M Sodium Hydroxide (NaOH) and stored at -20°C. A working solution of 10nM was prepared and the required amount was pipetted in the corresponding well as shown in table 2.6.

Hemin Concentration	96 well plate	24 well plate	6 well plate
10µM	0.1µL	0.5µL	1.0µL
20µM	0.2µL	1.0µL	2.0µL
30µM	0.3µL	1.5µL	3.0µL
40µM	0.4µL	2.0µL	4.0µL
50µM	0.5µL	2.5µL	5.0µL

Table 2.7: Addition of hemin to K562 cells

2.15.3 Human Umbilical-cord blood Derived Erythroid Progenitor-2 cells

The Human Umbilical-cord blood derived erythroid Progentoir-2 (HUDEP-2) cell line is an immortalized cell line created by the insertion of a tetracycline inducible HPV16-E6/E7 expression system via lentiviral transduction into human umbilical cord blood cells. These cells normally express adult haemoglobin ($\alpha_2\beta_2$) (Kurita et al., 2013). HUDEP-2 cell line was purchased from Yukio Nakamura, RIKEN Tsukuba Branch, Japan.

HUDEP-2 cell line were cultured in Cell-Quin Medium (home-made media gently donated by Emile van den Akker, Department of Haematopoiesis, Sanquin Research, Amsterdam), supplemented with SCF (50ng/ml), EPO (3IU/ml), dexamethasone (1×10^3) (Sigma-Aldrich, USA) doxycycline (1µg/mL) (Sigma-Aldrich, USA), 1% L-Glutamine (Sigma-Aldrich, USA) and 1% P/S (Sigma-Aldrich, USA). HUDEP-2 cells were cultured in T-25 flasks at 37°C with 5% CO_2 and 96% humidity.

2.15.3.1 Passaging of HUDEP-2 cells

The cells were checked under the light microscope and counted, using an electronic cell counter CASY-1 (Scharfe System, Reutlingen, Germany). The cells were maintained at a concentration of 1x106 cells/ml and every second day, the doxycline was replenished.

2.15.4 Proliferation assay

Proliferation assays are often used to determine if a compound has any effect on cell proliferation or to see if that compound has a direct cytotoxic effect on the cells. An MTT tetrazolium reduction assay (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide; thiazolyl blue) was used to determine the effect of hemin on cell proliferation. The MTT substrate, which is prepared in a physiologically balanced solution, is added to

the cells in culture and incubated for up to 4 hours. When added to the cells, the active mitochondrial dehydrogenases convert the dissolved MTT to an insoluble purple formazan precipitate, with an absorbance maximum near 570nm. The signal generated is directly proportional to the number of viable cells. When the cells die, they are not able to convert MTT into formazan, therefore the colour formation serves as a useful convenient marker for viable cells (Mosmann, 1983).

2.15.4.1 MTT assay following hemin addition to K562 cells

The MTT solution (Sigma-Aldrich, USA) was prepared by dissolving 5mg in 1mL 1 x PBS, covered with foil and stored in the dark at 4°C. On day 1, cell counting was carried out, K562 cells were seeded in two 96-well plates at a concentration of 0.2 x 10^5 cells/well and incubated at 37°C with 5% CO_2 and 96% humidity. One plate for each time interval; 24 and 48 hours. Twenty-four hours after seeding, hemin was added to the wells, as shown in figure 2.10, at a concentration of 10μM, 20μM, 30μM, 40μM and 50μM. Following the addition of hemin, the plates were swirled for a few minutes and incubated at 37°C with 5% CO_2 and 96% humidity. After 24 hours one 96-well plate was transferred to the sterile laminar flow hood and 10μL of 5mg/ml MTT solution was added to each well. The plate was placed back in the incubator for 4 hours. After 4 hours, the plates were centrifuged at 14,000 x g for 10 minutes. The supernatant was decanted and 100μL of DMSO was added to each plate. The aim of the DMSO is to dissolve the formazan produced inside the viable cells. The plate was then placed on a Thermomixer (Eppendorf, UK) for 2 minutes at 300rpm, to ensure that all crystals are dissolved and then transferred to a spectrophotometer (Mithras Multimode Microplate Reader LB 940, Berthold Technologies) and the absorbance at 570nm was measured. The same procedure was repeated the next day for the 48 hours' time interval. This procedure was repeated twice.

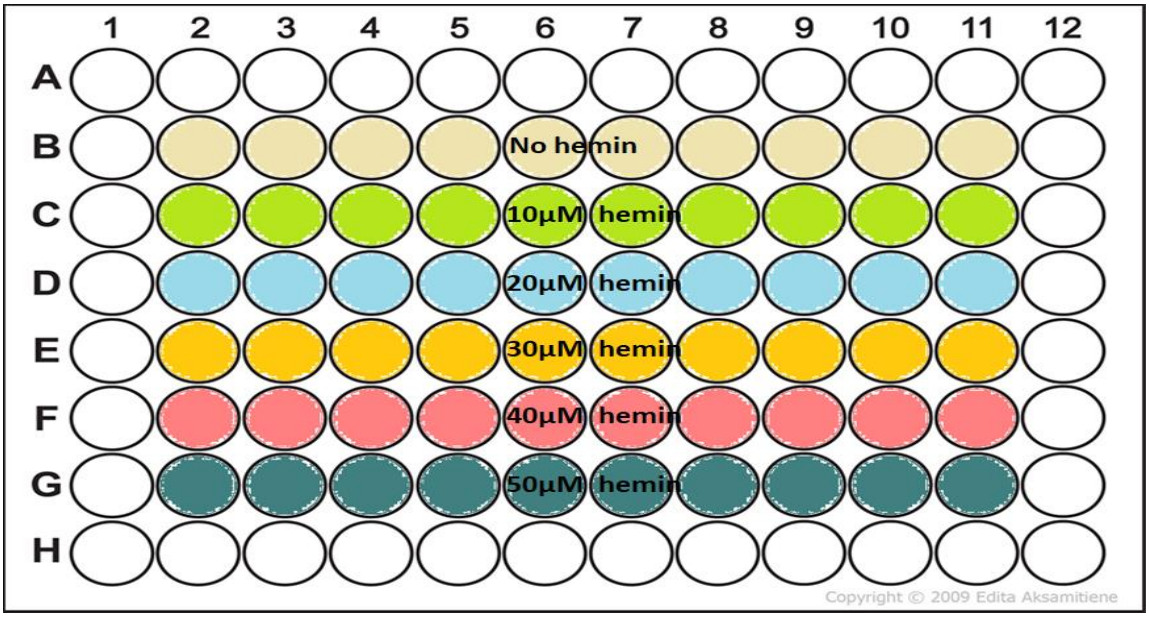

Figure 2.10: MTT assays of K562 cells treated with different hemin concentrations. The cream wells show K562 with no addition of hemin. The other wells show different concentration of hemin ranging from 10µM to 50µM The peripheral wells were not used and were therefore filled with PBS.

2.16 Transient Transfections of cell lines

Transfection introduces foreign nucleic acids into cells, producing a genetically modified cell. The introduction of the genetic material can either be done by transient transfection or by stable transfection. In case of a stable transfection, the introduced genetic material will be integrated into the host genome, even after the host cells replicate. In the case of transient transfection, the transfected genes are only expressed for a limited period of time and are not integrated into another genome (Recillas-Targa, 2006). Many transfection methods have been developed and each method uses a different approach that must be considered depending on cell type and purpose. For our cell lines, transient transfection was carried out.

Prior to transfection, the whole process was optimized using different methods such as FuGENE® HD (Promega, USA), Lipofectamine™ (Invitrogen, USA) and Magnetofaction.

Transfected cells were visualized using the pMAX-GFP expression plasmid which gave off green luminescence under the microscope. The best method that gave satisfactory transfection results was by using Magentofection™. In the case of HEK293, best transfection rate was achieved by Magnetofection™, using the NeuroMag reagent, while in case of K562, it was achieved by Magnetofection™, using the PolyMag reagent.

The principle of Magnetofection is the association of nucleic acids, transfection reagents or virus with specific cationic magnetic nanoparticles. Using the appropriate magnetic field, the resulting molecular complexes are then concentrated and transported into cells (figure 2.11). The magnetic nanoparticles are made up of fully biodegradable iron oxide, coated with specific cationic proprietary molecules. Their association with the gene vector is achieved by salt-induced colloidal aggregation and electrostatic interaction. By the influence of an external magnetic field, generated by the magnetic plate, the magnetic particles are then concentrated on cells. The cellular uptake of genetic material is achieved by endocytosis and pinocytosis. In contrast with other physical transfection methods that create hole or electroshock to the cell membrane, with Magnetofection the membrane architecture and structure of the cell remain intact. The nucleic acids are then released into the cytoplasm by different mechanisms, depending on the formulation used. In our case the formulation used was that of the proton sponge effect, where the cationic polymers coated on the nanoparticles promotes endosome osmotic swelling. This results in the disruption of the endosomal membrane and the intracellular release of DNA.

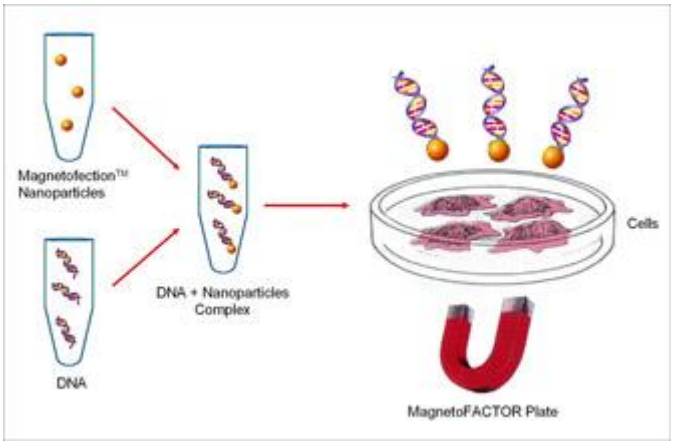

Figure 2.11: Magnetofection principle

2.16.1 Transfection of K562 cells

One day prior to transfection, K562 were seeded in a 24 well plate at a density of $0.75x10^5$ cells per well. Cells in 75mL culture flasks were centrifuged, the supernatant decanted and the cell pellet was resuspended in 10mL of fresh RPMI, supplemented with 10% FBS and 1% PS. Cell counting was carried out, using an automated CASY system and depending on the cell count obtained, the cell suspension was either diluted or concentrated, to achieve the correct cell density for seeding in 500µl medium. In the case of the 24-well plate, the peripheral wells were not used for transfection purposes and therefore they were filled with PBS. The plates were placed overnight in the incubator set at 37°C with 5% CO_2 and 96% humidity.

The next day, cells were checked under the light microscope to ensure that they had reached a confluence of about 80%. The DNA/PolyMag complexes were prepared in RPMI with no other supplements added. After optimization, it was found that the ideal amount of DNA to PolyMag reagent ratio was that of 1:1 (0.5µg of DNA and 0.5µl of PolyMag reagent). The mixture was mixed gently by pipetting up and down and was incubated at room temperature for 20 minutes. A total volume of 25µl of the complex mixture was added to the K562 cells. The 24 well plate was swirled gently, to ensure a uniform distribution of the mixture and was placed on a

magnetic device (Magnetofect nano II™, nanoTherics), set at an oscillating frequency of 2Hz, and amplitude of 0.2mm and incubated at 37°C for 20 minutes. After 20 minutes the plate was removed from the magnetic plate and was incubated at 37°C with 5% CO_2 and 96% humidity for further studies.

2.16.2 Transfection of HEK293T cells

One day prior to transfection, HEK293T cells were seeded in a 24 well plate at a concentration of $0.5x10^5$ cells per well. The HEK293 cells in a 75mL culture flask were trypsinized, spun and resuspended in 10mL fresh DMEM with 10% FBS and 1% PS. Cell count was carried out using a haemocytometer and depending on the cell count obtained, the cell suspension was either diluted or concentrated to achieve the correct cell density for seeding in 500µl medium. In the case of the 24-well plate, the peripheral wells were not used for transfection purposes and therefore they were filled with PBS. The plates were placed overnight in the incubator set at 37°C with 5% CO_2 and 96% humidity.

The next day, cells were checked under the light microscope to ensure that cells had reached a confluence of about 80%. The DNA/NeuroMag complexes were prepared in DMEM, with no other supplements added. After optimization, it was found that the ideal amount of DNA to NeuroMag reagent ratio was that of 2:1 (0.5µg of DNA and 0.25µl of NeuroMag reagent). The mixture was mixed gently by pipetting up and down and was incubated at room temperature for 20 minutes. A total volume of 25µl of the complex mixture was added to the HEK293 cells. The 24 well plate was swirled gently to ensure a uniform distribution of the mixture and was placed on a magnetic device (Magnetofect nano II™, nanoTherics), set at an oscillating frequency of 2Hz, and amplitude of 0.2mm and incubated at 37°C for 30 minutes. After 30 minutes the plate was removed from the magnetic plate and incubated at 37°C with 5% CO_2 and 96% humidity for further studies.

2.16.3 Transfection of HUDEP-2

Several transient transfection methods were tried on HUDEP-2 cells without achieving a satisfactory transfection result. Therefore HUDEP-2 cells were not used for promoter assays.

2.17 Dual Luciferase Reporter (DLR) Assay

Luciferase Reporter Assays are used to study gene expression at the transcriptional level. The transcriptional activity of a transcription factor is quantified through the targeted expression of the firefly luciferase gene. The firefly luciferase gene is characterized by its ability to generate a stabilized luminescent signal, which can be detected, quantified and interpreted. The value obtained is directly proportional to the expression of the studied gene. The Dual-Luciferase® Reporter (DLR™) assay system (Promega, USA) uses the quantification of two distinct luciferase enzymes that are co-expressed within the same cells, but, due to their different origin, they can be measured separately. The primarily luciferase enzyme is the firefly (*Photinus pyralis*), while the second luciferase enzyme is the Renilla (*Renilla reniformis,* also known as the sea pansy) luciferase. The DLR assay allows firstly the quantification of the firefly luciferase luminescence, which is then quenched, to allow the measurement of the Renilla luciferase luminesce.

The aim of this experiment was to study the effect of the KLF1 promoter mutations on the luciferase expression when compared to the wildtype promoter. The HEK293T cells and the K562 cells were transfected with the plasmid containing the EV, wildtype and mutants KLF1 promoter, the reporter plasmid pRL-SV40 and the pRL-TK vector. The baseline response of the pRL-SV40 plasmid containing the Renilla firefly was used as an internal control to

normalize the data. The pRL-TK vector was used as a control to test the efficiency of transfection.

2.17.1 Transfections of expression vectors in HEK293T cells and K562 cells

One day prior to transfections, the HEK293T cells and the K562 cells were seeded in 24-well plates, as already described in section 2.16.1 and 2.16.2, for each respective plasmid. After 24 hours the cells were checked under the microscope to verify that the required confluence was reached.

Expression Plasmid	Amount of Plasmid (µg plasmid well)	pRL-SV40 (µg plasmid per well)	NeuroMag (µl/well)	Total volume in DMEM (no FBS or penicillin) (µL/well)
pRL-TK	0.5	0.0125	0.25	25
pGL4.10[luc 2] -EV	0.5	0.0125	0.25	25
pGL4.10[*luc2*] – WT KLF1 promoter	0.5	0.0125	0.25	25
pGL4.10[*luc2*] -1133C>A substitution	0.5	0.0125	0.25	25
pGL4.10[*luc2*] –251 C>G substitution	0.5	0.0125	0.25	25
pGL4.10[*luc2*] -1133C>A & –251 C>G substitutions	0.5	0.0125	0.25	25
pGL4.10[*luc2*] –148 G>A substitution	0.5	0.0125	0.25	25
pGL4.10[*luc2*] –154 C>T substitution	0.5	0.0125	0.25	25
pGL4.10[*luc2*] –43 G>A substitution	0.5	0.0125	0.25	25
pGL4.10[*luc2*] –102 T>G substitution	0.5	0.0125	0.25	25

Table 2.8: The ten master mixes prepared containing different expression plasmids transfected in HEK293T cell

Expression Plasmid	Amount of Plasmid (µg plasmid well)	pRL-SV40 (µg plasmid per well)	PolyMag (µl/well)	Total volume in RPMI (no FBS or penicillin) (µL/well)
pRL-TK	0.5	0.0125	0.5	25
pGL4.10[*luc2*] -EV	0.5	0.0125	0.5	25
pGL4.10[luc2] – WT KLF1 promoter	0.5	0.0125	0.5	25
pGL4.10[l*uc2*] -1133C>A substitution	0.5	0.0125	0.5	25
pGL4.10[*luc2*] –251 C>G substitution	0.5	0.0125	0.5	25
pGL4.10[*luc2*] -1133C>A & –251 C>G substitutions	0.5	0.0125	0.5	25
pGL4.10[*luc2*] –148 G>A substitution	0.5	0.0125	0.5	25
pGL4.10[*luc2*] –154 C>T substitution	0.5	0.0125	0.5	25
pGL4.10[*luc2*] –43 G>A substitution	0.5	0.0125	0.5	25
pGL4.10[*luc2*] –102 T>G substitution	0.5	0.0125	0.5	25

Table 2.9: The ten master mixes prepared containing different expression plasmids transfected in K562

Transfection was carried out using either the pGL4.10 [*luc2*] with the wildtype KLF1 promoter or the mutated promoters, EV and the pRL-TK vector in combination with the pRL SV40 Renilla Luciferase plasmid. For HEK293T cells and K562 cells, in total, 10 master mixes were prepared, as shown in table 2.8 and 2.9. The necessary dilutions of each plasmid were carried out to achieve the desired working concentrations, followed by the addition of the NueroMag in case of HEK293T cells and PolyMag in the case of K562 cells. The reaction mixture was incubated for 30 minutes in the case of the NeuroMag and 20 minutes in the case of PolyMag. A total of 25µl of the reaction mixture was pipetted in the respective well and the plates were put on a magnetic device sequentially and incubated at 37°C. In the case of HEK293T cells,

the plate was incubated for 30 minutes, while in the case of K562, the plate was incubated for 20 minutes. After the incubation, the cells were removed from the magnetic plate and incubated at 37°C with 5% CO_2 and 96% humidity for 48 hours. For each expression plasmid, 3 wells were prepared, and the experiment was repeated 4 times.

2.17.2 Luciferase reporter assay for dose response curve of hemin

K562 cells were transfected with either the pGL4.10 [l*uc2*] empty vector, the wildtype KLF1 promoter together with the pRL SV40 Renilla Luciferase. Following transfection, the cells were treated with different concentrations of hemin to determine the optimal concentration to use for further studies. Each concentration of hemin was repeated 4 times, for both the empty vector and the vector containing the wildtype KLF1 promoter.

K562 were seeded in a 24-well plate and after 24 hours the cells were checked under the light microscope and transfection was carried out, as described in section 2.16.2. After transfection the cells were removed from the magnetic plate and incubated at 37°C with 5% CO_2 and 96% humidity. Twenty-four hours after transfection, different concentration of hemin, including 10μM, 20μM, 30μM, 40μM and 50μM, were pipetted into each respective well, as shown in figure 2.12. Twenty-four hours after addition of hemin, cell lysis was performed, as explained in section 2.14.4.2. It was found that the 50μM of hemin results in the largest luciferase activity.

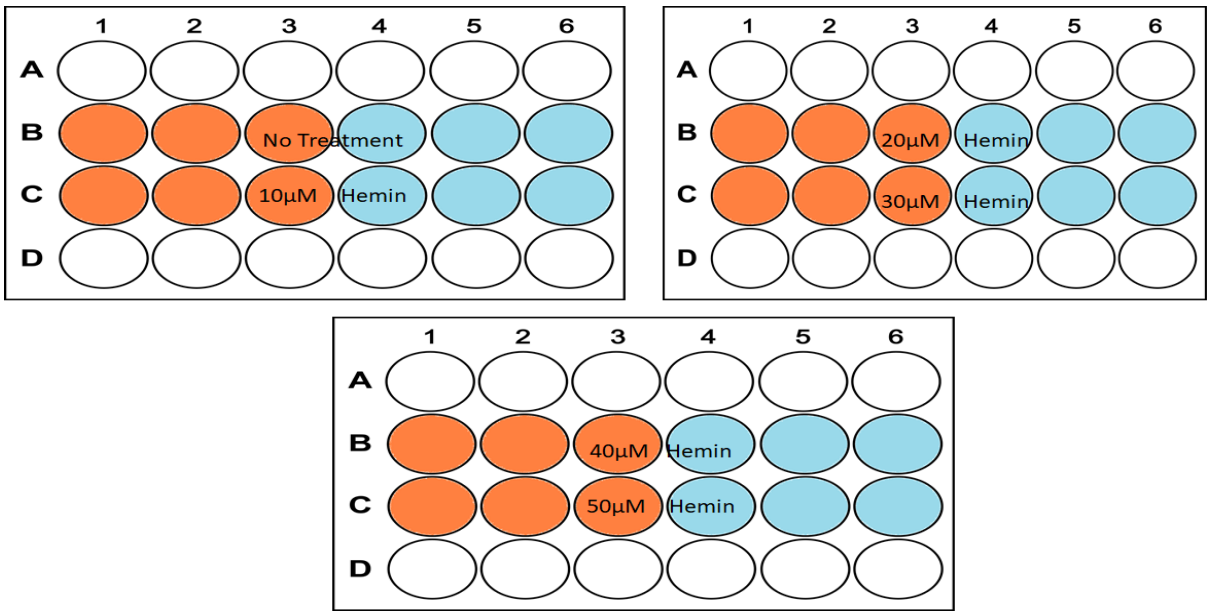

Figure 2.12: Orientation of 24-well plates used for DLR assay after transfection with the EV and wildtype KLF1 promoter, and treatment with different hemin concentrations. Orange-coloured wells indicate K562 cells transfected with the pGL4.10 [luc2] empty vector and blue-coloured wells indicate K562 cells transfected with the wildtype KLF1 promoter-pGL4.10 [luc2] vector. The cells were subjected to different hemin concentrations (10µM - 50µM).

2.17.3 Transfection of Expression Vectors in K562 Cells followed by addition on 50µM hemin

K562 cells were seeded in a 24 well plate and transfected with each respective expression vector, as explained in section 2. Twenty-four hours after transfection, 50µM of hemin was pipetted into each well, except in the pRL-TK wells. The plates were incubated at 37°C with 5% CO_2 and 96% humidity. Twenty-four hours after addition of hemin, cell lysis was performed, as explained in section 2. For each expression plasmid, 3 wells were prepared, and the experiment was repeated 4 times.

2.17.4 Passive Lysis of cells

2.17.4.1 Passive Lysis of HEK293T Cells

For passive lysis of cells, the passive lysis buffer (PLB) supplied with the Dual Luciferase Reporter Assay kit (Promega, UK) was used to enhance lysis of HEK293T cells and K562 cells, while maintaining the optimum performance and stability of the firefly and *Renilla* luciferase reporter enzymes. In the case of HEK293T cells, 48 hours after transfection, the plates were removed from the incubator and placed in a Class II laminar flow. The medium was gently aspirated, and the wells were washed with 300µl PBS. The PBS was then aspirated and PLB was added to the cells. The 1x PLB concentrate was prepared by 1mL of the 5x PLB with 4mL of sterile water and 100µl of the 1x PLB was pipetted in each well. The plates were placed on an orbital shaker for 10 minutes at 10 x g and frozen at -20°C before initiating the luciferase assay.

2.17.4.2 Passive Lysis of K562 cells

In the case of K562, 48 hours after transfection, the plates were removed from the incubator and placed in a Class II laminar flow. The cells were transferred to a 1.5mL microcentrifuge tube and centrifuged at 450 x g for 5 minutes at 4°C. The supernatant was decanted, and the cells pellet was washed with 150µl PBS. Centrifugation at 450 x g for 5 minutes at 4°C was performed, the supernatant was decanted and 100ul of 1x PLB was added to the cells. The cells were placed on an orbital shaker for 10 minutes at 10 x g and frozen at -20°C before initiating the luciferase assay.

2.17.5 Luciferase Enzyme Activity Measurement

Luciferase reporter activity was measured using both a manual luminometer and an automated luminometer. The dual-Luciferase Kit reagents were removed from the freezer and allowed to thaw at room temperature in a water-bath. In the meantime, the cell lysates were removed from the -20°C and placed on the orbital shaker to thaw. The Reagents were prepared according to the manufactures' instructions. The Luciferase Assay Reagent II (LAR II) was prepared by re-suspending the lyophilized Luciferase assay substrate in 10mL of the supplied Luciferase Assay Buffer II. The Stop & Glo® Substrate was prepared by adding 200µl of the 50x Stop and Glo® substrate to 10mL Stop & Glo® Buffer.

2.17.5.1 Luciferase Enzyme Activity Measurement using the manual luminometer

A sterile polypropylene tube was prepared for each cell lysate. 100µl of the LAR II was pipetted in each tube, followed by the addition of 20ul of the cell lysate and placed into a manual luminometer (TD-20/20, Tuner Design), to measure first the luciferase (firefly reading). The polypropylene tube was removed from the luminometer and 100µL of the 1x Stop & Glo® reagent was added to the tube. The polypropylene tube was placed again in the luminometer and Renilla luciferase activity reading was measured. The light emitted was directly related to the concentration of each luciferase enzyme. The tube was discarded, and this process was repeated for all cell lysates. The readings were computed and analysed.

All dual-luciferase assays were repeated using the LB 940 Mithras Multimode Microplate Reader (Berthold Technologies, USA).

All the luminescence data obtained from the dual Luciferase assays was analysed as follow.
The mean firefly luminescence and the mean Renilla luminescence of the non-transfected controls (NTC) were calculated. Each mean NTC luminescence value was subtracted from the

respective luminescence value of each lysate. This generated a set of background-corrected luminescence values. Using the background-corrected values, the ratio of firefly to renilla luminescence was calculated for each lysate in the plate. The mean of these ratios was then calculated, generating a mean value of firefly:renilla luciferase activity for each lysate. Each mean firefly:luciferase ratio was then divided by the ratio pertaining to the cell lysates, transfected with empty pGL4.10 [luc2] luciferase plasmid, thus expressing the ratio of the activities as a fold over luciferase expression in cells transfected with promoter-less pGL4.10 [luc2] luciferase plasmid.

Data was not normally distributed and therefore the Kruskall-Wallis used to test whether there is a significant difference in the median of the different independent groups. When a statistically significant Kurskal-Wallis value was obtained, the Post-Hoc Analysis was performed, to identify where the difference truly lies.

2.18 Electrophoretic mobility Shift assay

In addition to functional promoter assays, we wanted to see if the interaction of a putative transcription factor with the DNA sequence of interest is affected by the promoter mutations. Electrophoretic mobility shift assay (EMSA) was carried out. EMSA is based on the principle that a protein-DNA complex migrates more slowly through a native gel than the corresponding free DNA.

2.18.1 Nuclear protein extraction

The first step involved nuclear protein extraction from K562 cells using the Abcam nuclear extraction kit (Cambridge, UK). K562 were grown to 2 x 10^6/ml and collected into a 15ml conical tube. The cells were centrifuged at 172 x g for 5 minutes and the supernatant was

discarded. The cells were washed with PBS and centrifuged again at 200 x g for 5 minutes and the supernatant was discarded, and the cell pellet was resuspended in 200µl 1x Pre-extraction buffer and the resuspended cell pellet was transferred to a microcentrifuge vial. The microcentrifuge vial was incubated on ice for 10 minutes. After incubation the vial was vortexed vigorously for 10 seconds and centrifuged at 15,000 x g for 1 minute. The cytoplasmic extract was carefully removed from the nuclear pellet. DTT solution and PIC were added to the extraction buffer at a ratio of 1:1000. 20µl of extraction buffer containing DTT and PIC were added to the nuclear pellet and the extract was incubated on ice for 15 minutes. Every 3 minutes the extract was vortexed for 5 seconds. The suspension was centrifuged for 10 minutes at 20,000 x g at 4°C and the supernatant was transferred to a new microcentrifuge tube. The protein concentration of the nuclear extract was measured, as explained in section 2.

2.18.2 Nuclear extract protein quantification

The protein concentration of the nuclear extracts was measured using Bradford assay. Six bovine serum albumin (BSA) (Sigma-Aldrich, United Kingdom) standards, at a concentration of 0.1, 0.3, 0.5, 0.7, 0.9 and 1mg/ml, were prepared by diluting the BSA in sterile distilled water. 10µl of each standard was dissolved in 1ml of diluted Bradford reagent (Sigma-Aldrich, United Kingdom) in 96 well plate and mixed by pipetting. 10µl of each nuclear extract were added to the diluted Bradford reagent and left for 10 minutes at room temperature. Each standard and sample were then measured in duplicate on the NanoDrop and the absorbance values of the standard were used to plot a calibration curve from which the protein concentration could be extrapolated.

2.18.3 Preparation of non-denaturing polyacrylamide gel

After preparation of the solutions (Appendix E) the sandwich glass panes were first washed with deionized water and then with 70% ethanol. The elastic sealer was then placed along the runners, with the groove facing upwards. The sandwich panes were then assembled and secured with clips at both ends. We checked for leaks by adding some water. A comb was inserted in the sandwich and marked approximately 5mm below, so that the gel would eventually be filled up roughly to this mark. Two gels, an 8% running gel and a 6% stacking gel were prepared as shown in table 2.10.

	8% Running Gel	6% Stacking gel
Denoised water	3.1ml	3.5ml
40% Acrylamide	1.6ml	750µl
Tris pH 8.8/6.8	3ml	630µl
10% APS	80µl	50µl
TEMED	16µl	10µl

Table 2.10: Preparation of non-denaturation polyacrylamide gel

The running gel was poured between the glass panes up to the mark. A layer of water-saturated butanol (to avoid the introduction of air bubbles) on top of the running gel was added and left to set for around 30 minutes. Once the running gel solidified, the butanol layer was poured off from the running gel and rinsed well with deionised water. The stacking was prepared and poured onto the running gel and the comb was placed and left to solidify.

2.18.4 Preparation of samples and controls

For EMSA, double stranded oligonucleotides of approximately 25bp-30bp containing the wildtype sequence and mutant sequence (table 2.11) were ordered from IDT, Belgium. For each promoter mutation, a reaction containing wildtype oligonucleotides, a reaction containing the mutant oligonucleotides and a control reaction was prepared.

Oligo	Sequence
rs112943513 Wildtype Forward	TTTGTTACCTCAGCCTCCTGAGTAGCTGGGA
rs112943513 Wildtype Reverse	TCCCAGCTACTCAGGAGGCTGAGGTAACAAA
rs112943513 Mutant Forward	TTTGTTACCTCAGCATCCTGAGTAGCTGGGA
rs112943513 Mutant Reverse	TCCCAGCTACTCAGGATGCTGAGGTAACAAA
rs381762 Wildtype Forward	TCAAACCCTGAACCCCCCAACCCTTGATATT
rs381762 Wildtype Reverse	AATATCAAGGGTTGGGGGGTTCAGGGTTTGA
rs381762 Mutant Forward	TCAAACCCTGAACCCCGCAACCCTTGATATT
rs381762 Mutant Reverse	AATATCAAGGGTTGCGGGGTTCAGGGTTTGA
rs79334031 Wildtype Forward	GAAACAGTGCCCCCCCGCCGCCTTGCCTTGC
rs79334031 Wildtype Reverse	GCAAGGCAAGGCGGCGGGGGGGCACTGTTTC
rs79334031 Mutant Forward	GAAACAGTGCCCCCCCACCGCCTTGCCTTGC
rs79334031 Mutant Reverse	GCAAGGCAAGGCGGTGGGGGGGCACTGTTTC
rs372651309 Wildtype Forward	CCCAGAAACAGTGCCCCCCCGCCGCCTTGC
rs372651309 Wildtype Reverse	GCAAGGCGGCGGGGGGGCACTGTTTCTGGG
rs372651309 Mutant Forward	CCCAGAAACAGTGCTCCCCCGCCGCCTTGC
rs372651309 Mutant Reverse	GCAAGGCGGCGGGGGAGCACTGTTTCTGGG
rs372359976 Wildtype Forward	GTTCACGAGGCAGCCGAGGAAGAGGAGGCT
rs372359976 Wildtype Reverse	AGCCTCCTCTTCCTCGGCTGCCTCGTGAAC
rs372359976 Mutant Forward	GTTCACGAGGCAGCCAAGGAAGAGGAGGCT
rs372359976 Mutant Reverse	AGCCTCCTCTTCCTTGGCTGCCTCGTGAAC
rs548543206 Wildtype Forward	TGCAGCCAATCAGCTAAGGACAGAGAGGAG
rs548543206 Wildtype Reverse	CTCCTCTCTGTCCTTAGCTGATTGGCTGCA
rs548543206 Mutant Forward	TGCAGCCAATCAGCGAAGGACAGAGAGGAG
rs548543206 Mutant Reverse	CTCCTCTCTGTCCTTCGCTGATTGGCTGCA

Table 4.11: List of Oligos used for EMSA

In wildtype and mutant reaction 1040ng of nuclear extract was added together with 2µl binding buffer (750nM KCl, 0.5nM dithiothreitol, 0.5nM EDTA, 50nM Tris, pH7.4) purchased from

Invitrogen, United States and 5µl of sterile distilled water and 1µl of (50pM) forward and 1µl of (50pM) reverse oligonucleotide strand. The control reaction was prepared by adding 1µl of (50pM) forward and 1µl of (50pM) reverse wildtype oligonucleotide strand, 2µl binding buffer and 6µl sterile distilled water. The reaction mixture was mixed gently and incubated for 20 minutes at room temperature. After 20 minutes, 2µl of 6x EMSA gel-loading solution (Invitrogen, United States) was added to each mixture and the samples and controls were loaded on the gel. The electrophoresis was run at 100V for approximately 45 minutes. After the gel was stained in TAE buffer containing 5µl of ethidium bromide for 30 minutes. After 30 minutes the gel was washed 3 times with distilled water and visualized under UV transilluminator.

2.19 Using RegulomeDB to assess the function of the promoter mutations

All promoter mutations were analyzed for potential regulatory functions using RegulomeDB database. RegulomeDB is a database that annotates SNPs with known and predicted regulatory elements such as promoter regions, binding sites of transcription factors in the intergenic regions of H.Sapiens genome. Source of these data include public datasets from GEO, the ENCODE project and published literature.

3.1 Families with p.Lys288Ter KLF1 variation

The Maltese family with HPFH reported by Borg et al., (2010) was further extended. Through the thalassaemia clinic at Mater Dei hospital 5 other families with the same *KLF1* variation but with different level of foetal haemoglobin were identified. Throughout the thesis, these families are referred to as Fam F1 to Fam F6 respectively (Figures 3.1 & 3.3 and table 3.1).

3.1.1 Fam F1

Fam F1 was extended with baby IV-9 (figure 3.1) born at 40 weeks of gestation with severe jaundice, high bilirubin (221.8µmol/l - range 0 - 137µmol/l), high gamma-glutamyl transferase (γGT) (181 U/l - range 8 – 61 U/l), normal levels of alkaline phosphatase (ALP) and alanine aminotransferase (ALT). The CBC showed a high neutrophil count (16.28 $x10^9$/L - range 2.7 – 14.4 $x10^9$/L), low MCV (97.4fL- range 100 – 125 fL), high MCH (35.2pg - range 27 – 34 pg), high MCHC (36.2g/dL – range 30 -36 g/dL), normal haemoglobin and normal platelet count. The baby was placed in paediatric intensive care on phototherapy. Bilirubin levels were reported daily (figure 3.2a) until the 26th day when the bilirubin level decreased to 31.4 umol/l. The CBC was repeated every two days (figures 3.2b-e) and after the 160th day the baby showed a typical KLF1 deficiency picture, that of normal haemoglobin, low MCV and MCH and high reticulocytes levels. After 1 year of age, the HbA_2 level was of 1.8%, while the HbF level was 2964mg/dl. Sequencing of the KLF1 gene confirmed the p.Lys288Ter pathogenic mutation , together with the p.Met39Leu, and he was also homozygote for the p.Ser102Pro. Haemoglobin electrophoresis showed the presence of the alpha St. Lukes.

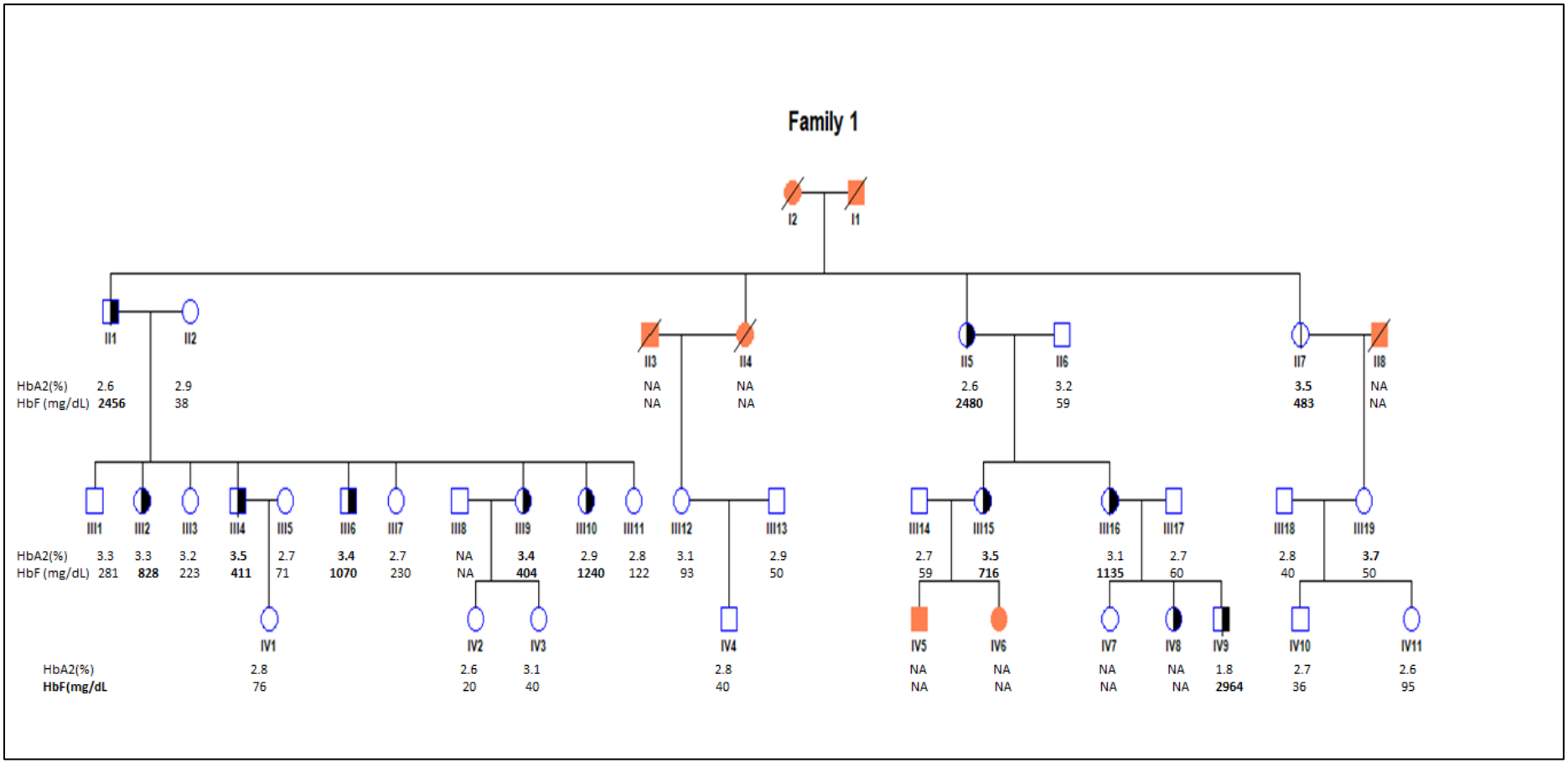

Figure 3.1: Extended Family tree of Fam F1. Heterozygotes for the p.Lys288Ter pathogenic mutation are represented by half-filled symbols. Open symbols show unaffected family members while orange symbols, purple symbols indicate Hb St. Luke's while grey symbols indicated unknown genotype. The HbA_2(%) and HbF(mg/dL) are beneath each family member. High HbA_2 (>3.3%) and HbF are bold.

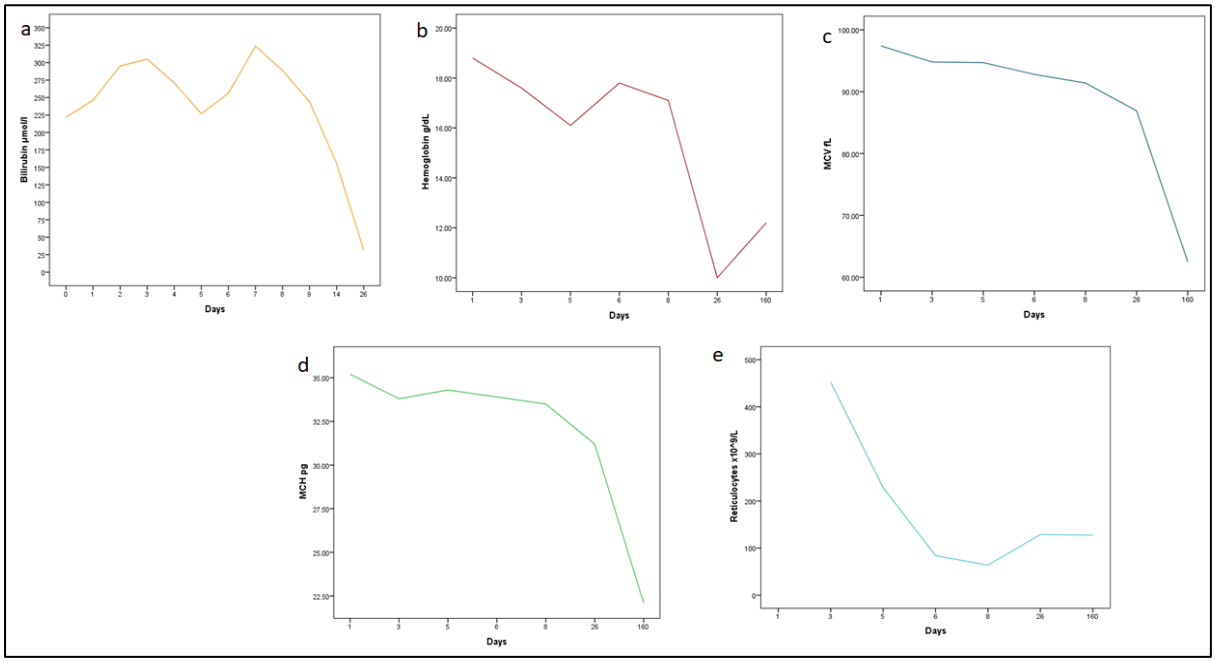

Figure 3.2: Haematological and biochemical values of IV-9. *The bilirubin, the haemoglobin, the MCV, the MCH and the reticulocyte levels of Baby IV-9. Figure A shows the bilirubin level in µmol/l during the first 26 days, figure B shows the haemoglobin level in g/dL during the first 160 days, figure C shows the MCV in fL during the first 160 days, figure D shows the MCH in pg during the first 160 days while figure E shows the Reticulocyte count in x109/L during the first 160 days*.

3.1.2 Fam F2

The proband (II-2) of Fam F2 had low MCH, HbF of 230mg/dL and HbA_2 of 3.3% (figure 3.3a). The brother (II-1) had normal haematological parameters, HbF of 80mg/dL and HbA_2 of 3.0%. The father (I-1) also had HbF of 260mg/dL and HbA_2 of 3.4% while the mother had MCH of 26.6pg, HbF of 40mg/dL and HbA_2 of 2.7% (table 3.1). *KLF1* gene sequencing of the proband and the father revealed three variations; the p.Met39Leu (rs112361212) and the p.Lys288Ter (rs267607202). The father was heterozygote for the p.Ser102Pro (rs2072597) and the proband was homozygote for this variation . The mother (I-2) and II-1 were heterozygote for the p.Ser102Pro. No α or β variations were detected.

3.1.3 Fam F3

The proband (II-1) of Fam F3, a one-year old girl, presented with microcytic anaemia (haemoglobin 9.6g/dl, MCV 62.0fL and MCH 20.0pg), HbF of 1210mg/dL and HbA_2 of 4.1%. The father had an MCV of 66.8fL and MCH of 20.9pg, HbF of 290mg/dL and HbA_2 of 4.9% while the mother had MCH of 26.3pg, HbF of 40mg/dL and HbA_2 2.8% (Table 3.1). Only the proband was available for sequencing. HBB sequencing revealed the IVS1,6C variation, while *KLF1* sequencing revealed the p.Lys288Ter pathogenic mutation . It was assumed that the proband inherited both the IVS1,6C and the p.Lys288Ter from her father, since the father had high HbA_2 and HbF. No α-variations were detected in the proband.

3.1.4 Fam F4

In Fam F4 (figure 3.3c), a 23-year-old pregnant woman (II-3) presented with MCV of 77.4fL, MCH of 26.4pg, HbF of 360mg/dL and HbA_2 of 3.6%. One of her children (III-4) died in utero at 36^{th} week of gestation. Her sister (II-2) also had an MCV of 79.6fL, MCH of 26.7pg, HbF of 90mg/dL and HbA_2 of 3.5%. *KLF1* sequencing revealed two variations; the p.Met39Leu and the p.Lys288Ter in both sisters. All children were wildtype for the p.Met39Leu and the p.Lys288Ter pathogenic mutation . No α or β variations were detected.

3.1.5 Fam F5

In Fam F5 (figure 3.3d), a 36 year old pregnant woman presented with HbA_2 of 3.4% and HbF of 168mg/dL, MVC of 77.2fL and MCH of 24.8pg. No variations were detected in the α and β-globin gene, but sequencing of the KLF1 gene revealed the p.Met39Leu, the p.Ser102Pro varations and p.Lys288Ter pathogenic mutations. The baby and the father had no variations in the KLF1 gene.

3.1.6 Fam F6

In Fam F6 (figure 3.3e), a 30 year old pregnant woman presented with HbA_2 of 3.4% and HbF of 283mg/dL. The MCV and MCH were low. No variation in the α and β-globin genes were detected. Sequencing of the KLF1 gene of proband and the baby revealved that they were heterozgyote for the p.Met39Leu, p.Ser102Pro variations and p.Lys288Ter pathogenic mutation. The father had no variations in the KLF1 gene. No α or β variations were detected.

Family	Member	HbA_2 (%)	HbF (mg/dL)	Hb (g/dL)	MCV (fL)	MCH (pg)	MCHC (g/dL)	p.Lys288 Ter (rs267607202)
Fam F2	I-1	3.4	260	15.1	84.5	27.8	32.9	+/-
	I-2	2.7	40	11.4	80.1	26.6	33.2	-/-
	II-1	3.0	80	14.5	87.6	28.5	32.6	-/-
	II-2	3.3	230	14.6	83.1	26.6	32.0	+/-
Fam F3	I-1	4.9	290	13.2	66.8	20.9	31.3	NA
	I-2	2.8	40	10.5	84.6	26.3	33.3	NA
	II-1	4.1	1210	9.6	62.0	20.0	32.3	+/-
Fam F4	II-1	2.7	40	14.8	85.9	34.0	33.0	-/-
	II-2	3.5	90	13.2	79.6	26.7	33.5	+/-
	II-3	3.6	360	13.5	77.4	26.4	34.5	+/-
	II-4	NA	NA	NA	NA	NA	NA	-/-
	III-1	NA	NA	NA	NA	NA	NA	-/-
	III-2	NA	NA	NA	NA	NA	NA	-/-
	III-3	NA	NA	NA	NA	NA	NA	-/-
Fam F5	I-1	2.6	60	14.8	85.7	31.0	34.0	-/-
	I-2	3.4	168	11.2	77.2	24.8	32.2	+/-
	II-1	NA	NA	NA	NA	NA	NA	-/-
Fam F6	I-2	3.0	150	12.5	86	29	34	NA
	II-1	2.9	384	12.0	77	26	33	+/-
	II-2	3.0	45	15	88	30	34	-/-
	III-1	2.7	48	12.1	86	29	34	-/-
	III-3	3.4	283	13.4	78	27	34	+/-
	III-4	2.8	37	14.5	87	30	34	-/-
	III-5	3.5	149	12.4	88	30	34	NA
	IV-1	2.8	47	11.8	81	27	33	-/-

Table 3.1: Table showing the CBC, the HbA_2(%) and HbF (mg/dL) together with the KLF1 of Fam F2 to Fam F6. NA means the data was not available.

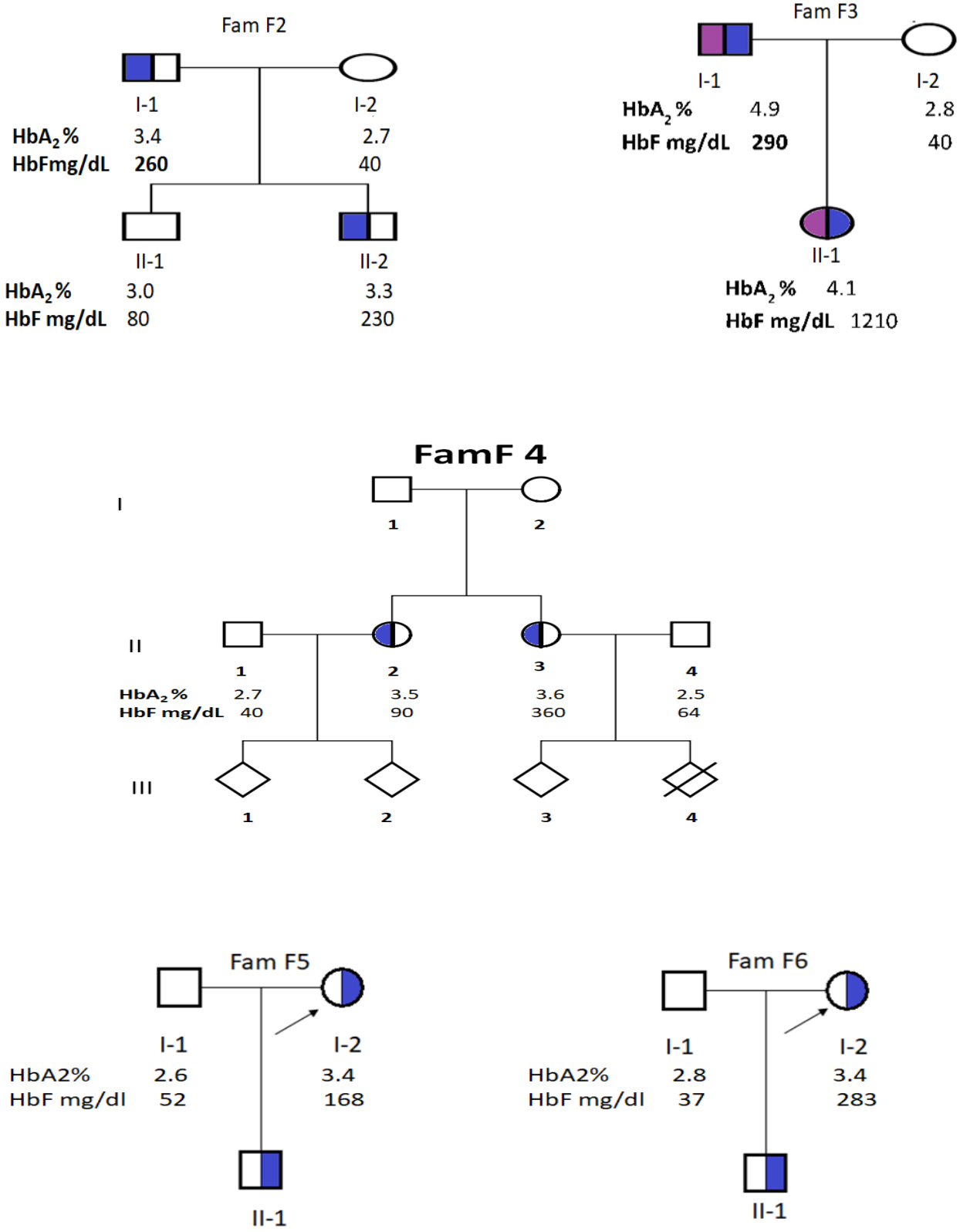

Figure 3.3: Family Trees of Fam F2, Fam F3, Fam F4, Fam F5 and Fam F6. Members heterozygote for the p.Lys288Ter pathogenic mutation are showed by half-filled symbols (blue). Members heterozygote for the IVS1,6C variation are shown by half-filled purple symbols. Open figures show unaffected family members.

3.2 Whole genome sequencing

The p.Lys288Ter KLF1 pathogenic mutation on its own does not fully explain the HbF and HbA_2 variability between same family members (Fam F1) and the other families (Fam F2 – Fam F6). Therefore a combined analysis of whole genome sequencing (WGS), gene expression (RNA-seq) and chromatin accessibility assays (ATAC-seq) were conducted to explore the molecular basis of the heterogeneity in HbF levels displayed by KLF1 p.Lys288Ter individuals on selected Fam F1 and Fam F2 family members

3.2.1 WGS analysis on Fam F1

3.2.1.1 WGS analysis on Fam F1 family members using HbF as a biomarker

The 10 Fam F1 family members were divided into four categories according to their HbF level, as shown in table 3.2.

FamF 1	Hbf mg/dL	HbA_2 (%)	-1133 C>A (rs112943513)	-251 C>G (rs3817621)	p.Met39Leu (rs112361212)	p.Ser102Pro (rs2072597)	p.Lys288Ter (rs267607202)
II-5	2480	2.6	+/-	+/-	+/-	+/+	+/-
II-1	2465	2.6	+/-	+/-	+/-	+/+	+/-
III-16	1135	3.1	+/-	-/-	+/-	+/-	+/-
III-6	1070	3.4	+/-	-/-	+/-	+/-	+/-
III-15	719	3.5	+/-	-/-	+/-	+/-	+/-
III-4	411	3.5	+/-	-/-	+/-	+/-	+/-
III-1	281	3.3	-/-	+/-	-/-	+/-	-/-
III-17	70	0.5	-/-	-/-	-/-	+/-	-/-
II-6	59	3.2	-/-	-/-	-/-	-/-	-/-
II-2	38	2.9	-/-	-/-	-/-	-/-	-/-

Table 3.2: Table showing the HbF in mg/dL, the HbA_2 in % and the klf1 variations in selected Fam F1 members. Subjects II-1 and II-5 marked in red have high HbF, subjects III-4, III-6, III-15 and III-16 marked in yellow have medium HbF levels, subject III-1 marked in blue have intermediate HbF level (between medium and low) while subjects II-2, II-6 and III-17 marked in green have low HbF level.

Looking closely in the KLF1 promoter, two variations were found. These were the -1133C>A (rs112943513 and the -251C>G (rs3817621). These variations were confirmed by Sanger Sequencing (figure 3.4). The -1133C>A mutation has a minor allele frequency (MAF) of 0.01 while the -251C>G has a MAF of 0.33.

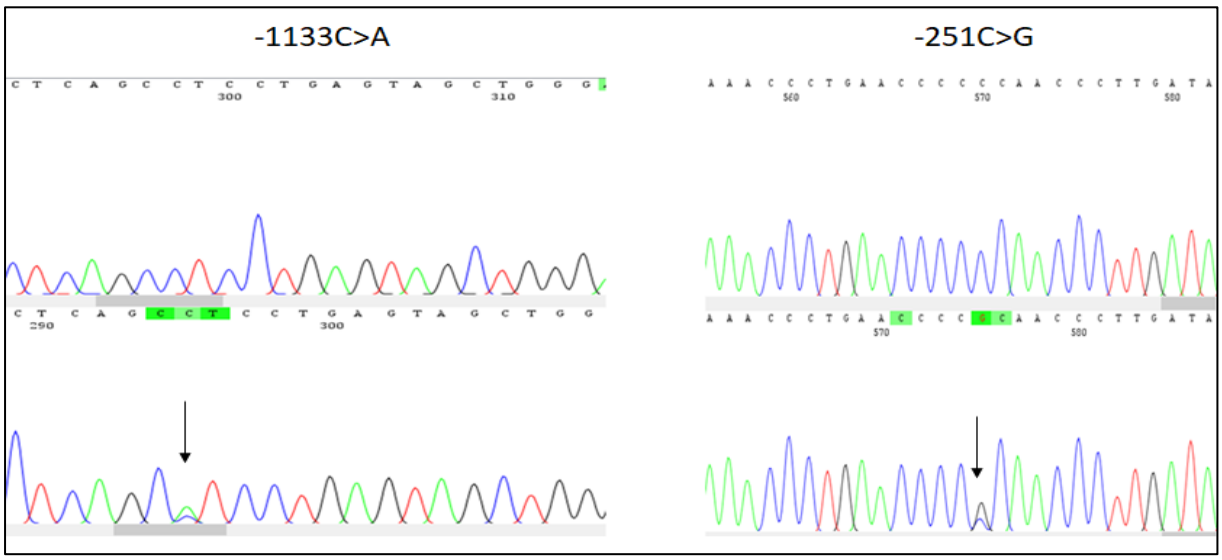

Figure 3.4: KLF1 promoter variations. The top sequence shows the wildtype sequence while the bottom sequence shows the SNVs marked with a black arrow. The sequences on the left shows the wildtype and variant -1133C>A sequence while the sequences on the right shows the wildtype and variant -251C sequence.

As shown in table 3.2, II-1 and II-5 were heterozygotes for both promoter variations, heterozygotes for the p.Met39Leu and the p.Lys288Ter and homozygote for the p.Ser102Pro had high foetal haemoglobin (2465 and 2480mg/dl). III-4, III-6, III-15 and III-16 were heterozygote for the -1133C>A, the p.Met39Leu, the p.Ser102Pro and the p.Lys288Ter and had medium level of HbF (411, 1070, 719 and 1135mg/dl). III-1, was heterozygote for the -251G>C variation and the p.Ser102Pro had intermediate HbF level (281mg/dl) while II-2, II-6 had no variation in the KLF1 gene and had low HbF level (38 and 59mg/dl). III-17 was heterozygote for the p.Ser102Pro and also had low HbF level (70mg/dl). Figure 3.5 shows a map of the KLF1 variations and figure 3.6 shows the KLF1 framework in Fam F1.

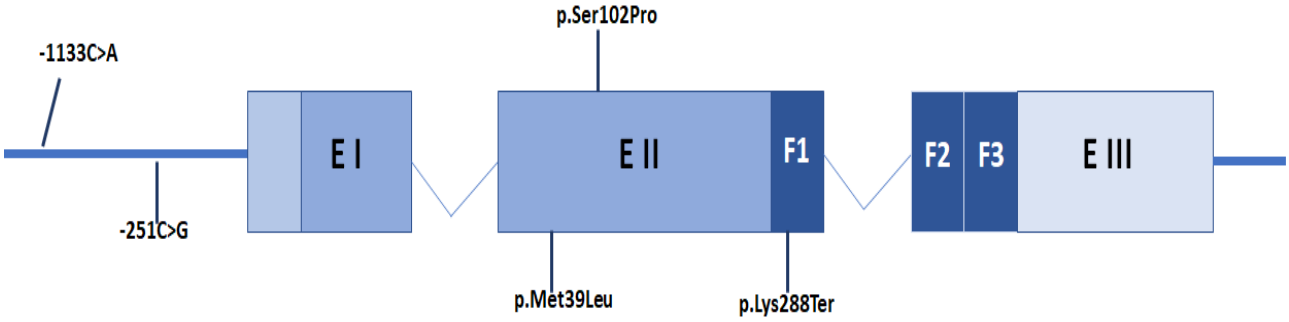

Figure 3.5: An extended map of the KLF1 gene showing all the SNV found in Fam F1. EI, EII and EIII refer to exon 1, exon 2 and exon 3. The dark blue boxes (F1, F2 and F3) represent the DNA binding domains

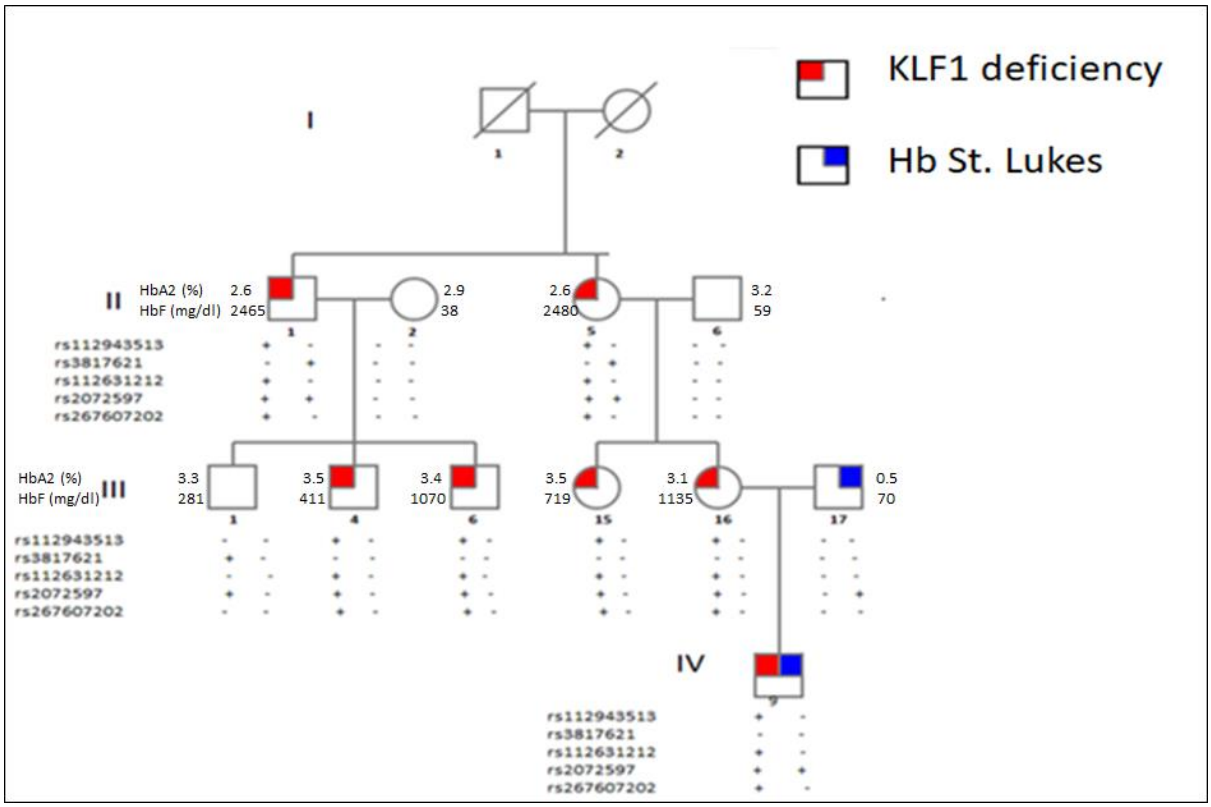

Figure 3.6: KLF1 framework of selected Fam F1 family members

3.2.2 WGS on Fam F2

Whole genome sequencing was performed on all Fam F2 members. When looking closely at the KLF1 gene, the father (I-1) was found to be heterozygote for the -1133 C>A KLF1 promoter variation, the mother (I-2) was found to be heterozygote for the -251C>G promoter variation . II-2 had no KLF1 promoter variations, while the proband (II-2) was heterozygote for both KLF1 promoter variations. The KLF1 framework are shown in figure 3.7.

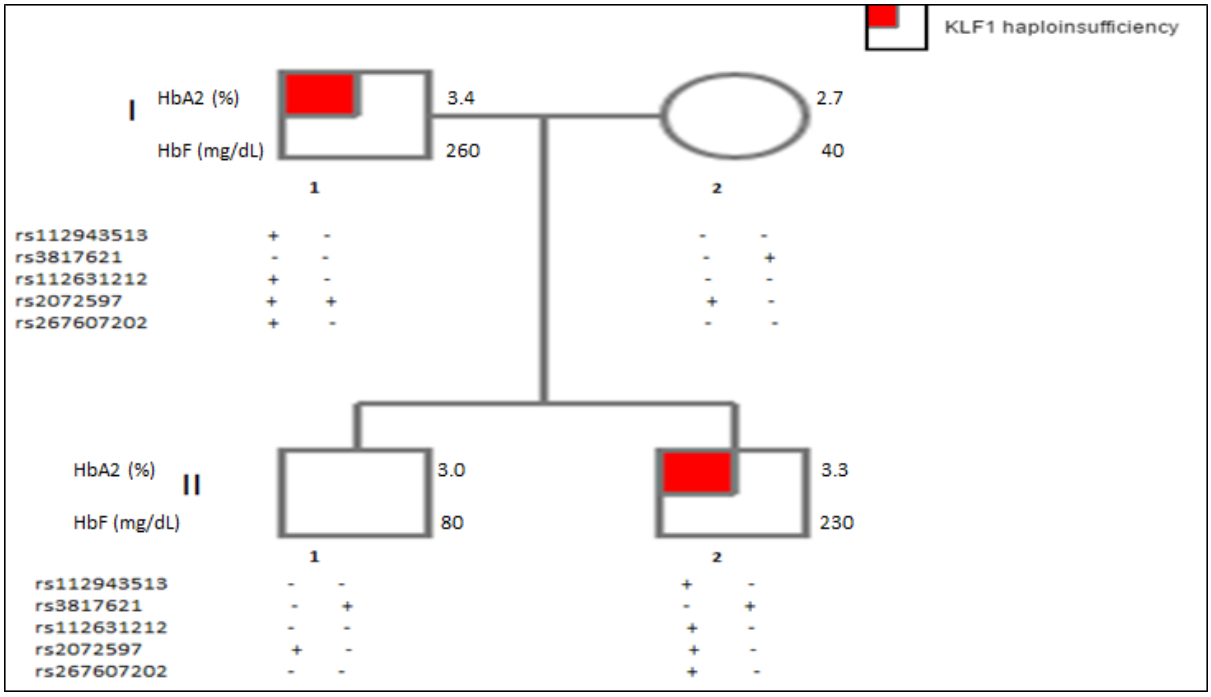

Figure 3.7: KLF1 haplotypes of FamF2 family members.

3.2.2.1 WGS analysis on Fam F2 using HbA_2 as biomarker

In Fam F2, I-1 and II-2 were heterozygote for the p.Lys288Ter pathogenic mutation but had intermediate level of HbF (260 and 230mg/dl) and borderline HbA_2 (3.4 and 3.3%). Using HbA_2 as a biomarker, when looking at genes related to foetal haemoglobin, I-1 and II-2 were homozygote or heterozygote for 12 variations that were absent in I-2 and II-1 (table 3.3). Two of these variations included the p.Lys288Ter pathogenic mutation and the p.Met39Leu

variation in the KLF1 gene. Other variations included, an SNP in the HBG1, 2 SNPs and insertion in the HBA1, 3 SNPs in SOX6, 2 SNPs in CREB1 and 1 SNP in DNASE2.

Chromosome	varType	xRef	MAF	I-1 3.4%	I-2 2.7%	II-1 3.0%	II-2 3.3%	symbol	component
chr11	snp	rs141669845	0.49	homo	wt	wt	het	HBG1	TSS-UPSTREAM
chr16	ins	rs71391119	NA	het	wt	wt	het	HBA1	TSS-UPSTREAM
chr16	snp	rs72763684	0.04	het	wt	wt	het	HBA1	TSS-UPSTREAM
chr16	snp	rs72763685	0.04	het	wt	wt	het	HBA1	TSS-UPSTREAM
chr11	snp	rs11023934	0.26	het	wt	wt	het	SOX6	TSS-UPSTREAM
chr11	snp	rs12221729	0.23	het	wt	wt	het	SOX6	TSS-UPSTREAM
chr11	snp			het	wt	wt	het	SOX6	TSS-UPSTREAM
chr2	snp	rs62189350	0.03	het	wt	wt	het	CREB1	TSS-UPSTREAM
chr2	snp	rs2253206	0.48	het	wt	wt	het	CREB1	TSS-UPSTREAM
chr19	snp	rs267607202		het	wt	wt	het	KLF1	CDS
chr19	snp	rs112631212	0.01	het	wt	wt	het	KLF1	CDS
chr19	snp	rs112348773	<0.01	het	wt	wt	het	DNASE2	CDS

Table 3.3: GWS analysis using HbA_2 as biomarker. Variations that are present in the heterozygote or homozygote state in family members I-1 and II-2 and absent in I-2 and II-1.

In Fam F1, the *DNASE2* variation (rs112348773) was also present in subjects with the p.Lys288Ter pathogenic mutation and the p.Met39Leu. The promoter variations were not found in Fam F3, Fam F4, Fam F5 and Fam F6.

3.3 Transcriptome Analysis

For RNA data analysis, samples were divided into 3 groups, according to their KLF1 variations and HbF in vitro, as showed in table 3.3. Subjects II-5 from Fam F1 and II-2 from Fam F2 were put in the High HbF group, III-15 and III-16 from Fam F1 were put in moderate HbF group, while II-6, II-7 from Fam F1 and I-2 from Fam F2 together with another independent control were put in the low HbF group (table 3.4).

Sample similarity was assessed by calculating inter sample Euclidean distance and performing multidimensional scaling. As shown in figure 3.8, this showed separate clusters of KLF1 p.Lys288Ter carriers and control samples within the timepoints T0 and T48.

At T48 the separation between the carriers and controls was more pronounced. At T0, two control samples grouped along with the carriers. This suggested that at T48 samples within the experimental group were more similar in gene expression. Therefore, for differential expression analysis we focused on T48 samples.

Sample	Family Member	HbF in vitro	-1133 C>A (rs112943513)	-251C>G (rs3817621)	p.Met39Leu (rs112361212)	p.Ser102Pro (rs2072597)	p.Lys288Ter (rs267607202)
1	II-5 Fam F1	High	+/-	+/-	+/-	+/-	+/-
2	II-2 Fam F2	High	+/-	+/-	+/-	+/-	+/-
3	III-15 Fam F1	Moderate	+/-	-/-	+/-	+/-	+/-
4	III-16 Fam F1	Moderate	+/-	-/-	+/-	+/-	+/-
5	II-6 Fam F1	Low	-/-	-/-	-/-	-/-	-/-
6	III-17 Fam F1	Low	-/-	-/-	-/-	+/-	-/-
7	I-2Fam F2	Low	-/-	+/-	-/-	+/-	-/-
8	Independent control	Low	-/-	-/-	-/-	-/-	-/-
9	Independent control	Low	-/-	-/-	-/-	-/-	-/-

Table 3.4: Grouping of subjects for RNA-sequencing data analysis. The table shows the HbF in vitro together with the KLF1 variants in selected Fam F1 and Fam F2 members. Subjects with high HbF are shown in pink, those with moderate HbF are shown in yellow while those with low HbF are shown in green.

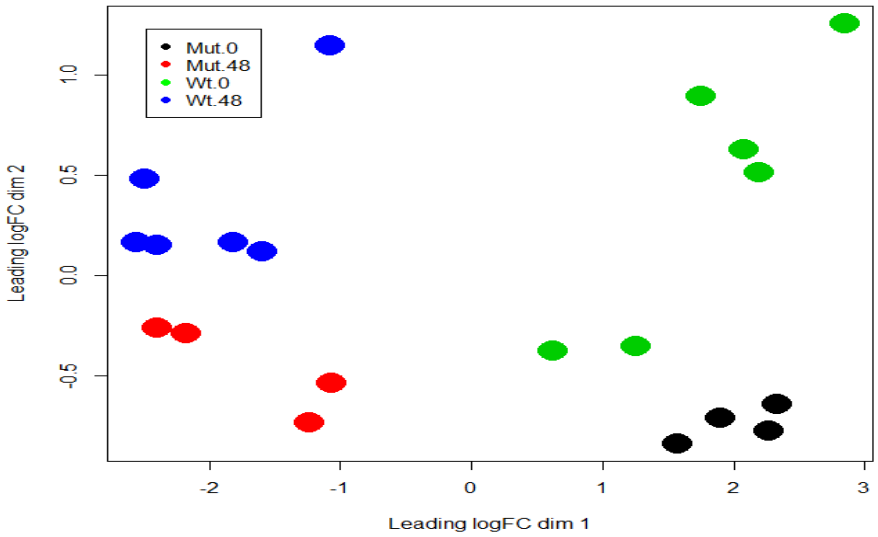

Figure 3.8: MDS-plots showing sample similarity. The black dots represent p.Lys288Ter heterozygotes at T0 while the green dots represent controls at T0. The blue dots show controls at T48 while the red dots show the p.Lys288Ter heterozygotes at T48.

3.3.1 Differential expression analysis between p.Lys288Ter carriers and controls

A total of 344 differential expressed genes were identified among the p.Lys288Ter carriers when compared to control samples at T48. The p.Lys288Ter heterozygotes had higher expression of gamma globin, embryonic zeta and epsilon globin, while they had lower expression of *BCAM*, *CD44*, *BCL11A* and beta globins (figure 3.9a). Other differentially expressed genes between p.Lys288Ter carriers and controls at T48 included *AHSP*, *KIT*, *GATA2*, *MYC*, *VEGFA* genes (figure 3.9b and Appendix F).

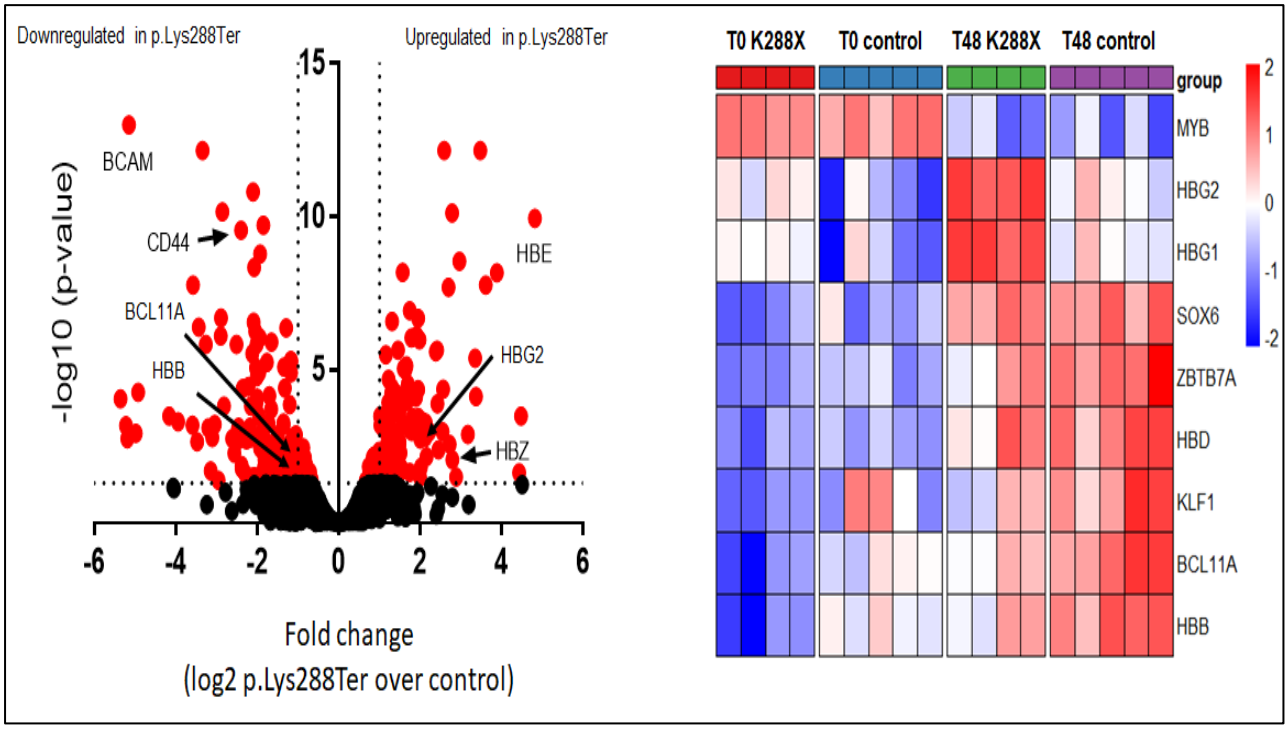

Figure 3.9: Volcano plot and heat map for differentially expressed genes. The figure on the left is Volcano plot with the 344 differential expressed genes identified in the p.Lys288Ter heterozygotes when compared to controls after 48 hours of differentiation. The figure on the right is a heatmap showing the expression of genes related to foetal haemoglobin switching in p.Lys288Ter heterozygotes and controls at T0 and T48.

3.3.3 Fam F1 & Fam F2 expression analysis between subjects with High, Medium and Low HbF in vitro

After 48 hours from differentiation, a difference in the expression of the globins, mainly a difference in HbA, HBB, HBG, HBZ and HBE, was noted between subjects with high HbF, medium HbF and low HbF. Gamma Globin, the embryonic zeta and epsilon were upregulated in high HbF and medium HbF subjects, while downregulated in subjects with low HbF (figure 3.10). The gamma globins were more or less similarly expressed in subjects with high HbF and medium HbF. In all controls, gamma globin was downregulated with the exception of control 8, where it was slightly upregulated. A difference of 4 log2-fold was seen in the upregulation of epsilon, between the high HbF and medium HbF, while in all controls it was downregulated.

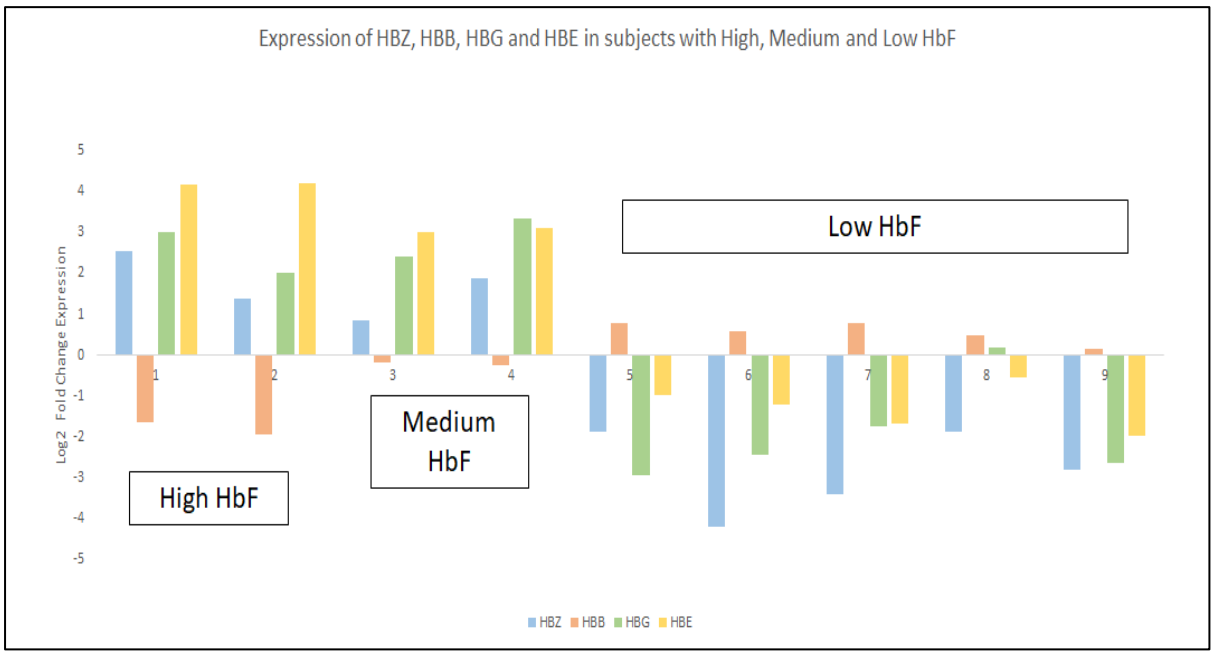

Figure 3.10: Log2 Fold change expression of HBZ, HBB, HBG and HBE in subjects with High HbF (1 &2), Medium HbF (4 & 5) and Low HbF (5, 6, 7, 8 & 9) in vitro after 48hours from differentiation (RNA sequencing data). Blue bars represent HBZ, pink bars represent HBB, green bars represent HBG while yellow bars represent HBE.

A difference in beta globin expression was noted between those with high HBF and those with Medium HbF. In HbF subjects, the beta globin was downregulated by a log fold change of 2, while in those with medium HbF, it was downregulated by a log fold change of 0.2. In all controls, HBB is upregulated. A significant decrease in KLF1 expression was observed in those with high HbF, while KLF1 is slightly downregulated in those with medium HbF (figure 3.11). In controls KLF1 is upregulated. The decrease in KLF1 was accompanied by a decrease in BCL11A. High HbF subjects showed the most significant downregulation of BCL11A with a log fold change of -1, while a log fold change of -0.3 was observed in medium HbF.

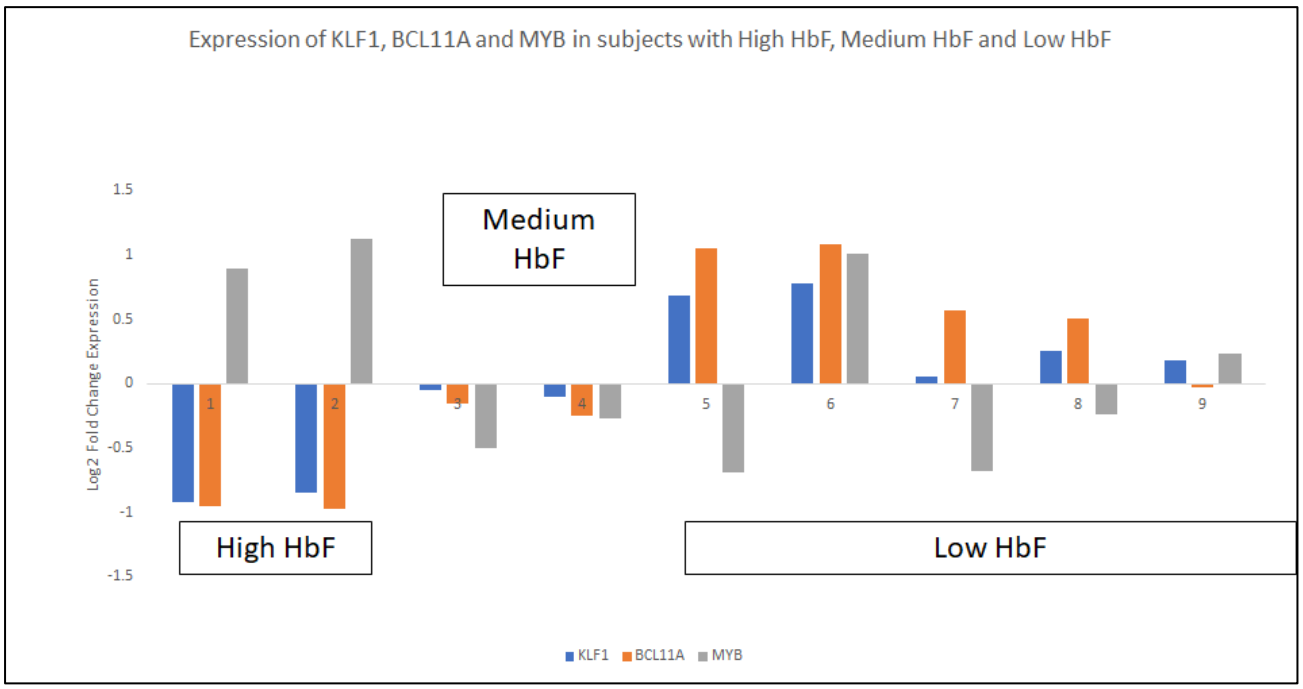

Figure 3.11: Fold change expression of KLF1, BCL11A and MYB in subjects with High HbF (1 & 2), Medium HbF (3 & 4) and Low HbF (5, 6, 7, 8 and 9) in vitro after 48hours from differentiation (RNA sequencing data). Blue bars represent KLF1, orange bars represent BCL11A while grey bars represent MYB.

In controls, except for control 9, *BCL11A* was upregulated. *MYB* was upregulated in high HbF subjects, while in subjects with medium HbF, *MYB* was slightly downregulated. In three controls, *MYB* was downregulated, while in two controls it was upregulated.

3.3.4 Validation of RNA results by qualitative PCR

RT-qPCR on cultured erythroid cells before differentiation (T0) and 48 hours after differentiation (T48) confirmed RNA-sequencing results (figure 3.12).

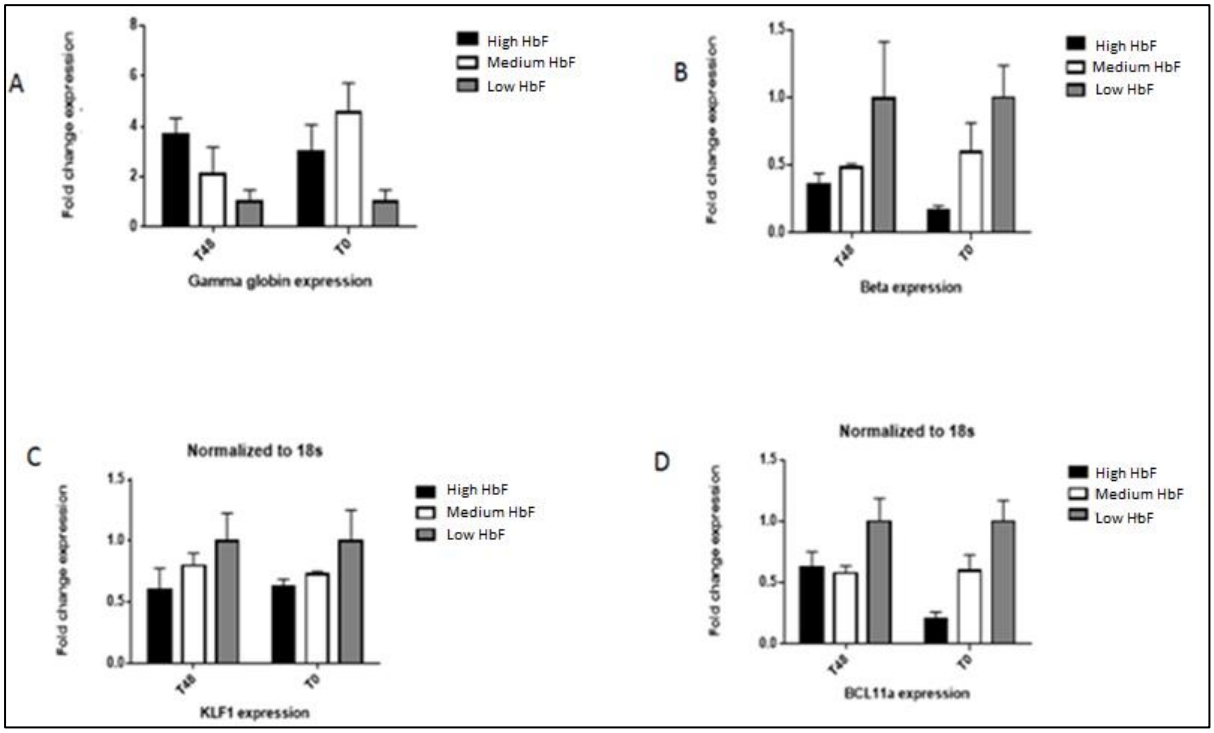

Figure 3.12: qPCR of Gamma, Beta, KLF1 and BCL11A to validate RNA sequencing results. The fold change expression of each gene is showed into three groups High HbF (black bars), medium HbF (white bars) and low HbF (grey bars) before differentiation (T0) and after differentiation (T48).

Gamma globin expression was increased in both Medium and High HbF subjects, when compared to control cells at both T0 and T48. The gamma globin expression increased more than 2 and 3-fold change compared to control cells (figure 3.12a). The increase in Gamma globin was accompanied by decrease in Beta globin (figure 3.12b) in both medium and high HbF, when compared to control cells in both T0 and T48. *KLF1* expression (figure 3.12c) was decreased in both medium and high HbF subjects when compared to controls at both T0 and T48. The decrease in *KLF1* was also accompanied by decreased *BCL11A* expression (figure 3.12d), in medium and high HbF subjects, at both T0 and T48.

3.3.5 Chromatin accessibility in KLF1 deficiency

ATAC sequencing was carried out on cultured cells to investigate the effect of chromatin accessibility on gene expression in KLF1 deficient subjects and controls. Data analysis for

ATAC sequencing was carried out at Sanquin, Amsterdam. To see if there were differences in specific regulatory elements between p.Lys288Ter heterozygotes and controls, peaks were called using the MACS2 algorithm (Zhang et al., 2008) and categorized based on the distance to the nearest transcription start site using the ALTRE R-package (Baskin et al., 2017). Fourteen thousand and six hundred and fifty-four (14,654) regulatory elements were found. Out of the 14,645 regulatory elements, 76 were found in both p.Lys288Ter heterozygotes and control samples, 4% were found only in the controls samples and 21% were found only in the p.Lys288Ter heterozygotes.

EdgeR was used for differential expression analysis on KLF1 p.Lys288Ter heterozygotes compared to control group. Five hundred and fifty-nine (559) different peaks with altered accessibility (FDR <0.05) were found. The p.Lys288Ter heterozygotes had a more open chromatin. In fact, 339 peaks were more accessible, while 220 peaks were less accessible, when compared to controls. The most pronounced peak was found in the HBB locus, where p.Lys288Ter heterozygotes showed increased accessibility of the HbG1 and HBG2 promoter (figure 3.13). No difference was noted in the other genes, nor the distal enhancers of the locus control region (LCR). Samples were then categorized according to their HbF level and a visible difference is seen in HBG1 and HBG2 between the high, medium and low HbF group. High HbF subjects showed the most increased accessibility for HBG1 and HBG2, followed by the medium HbF subjects. The difference in HbG2 was more pronounced (figure 3.14).

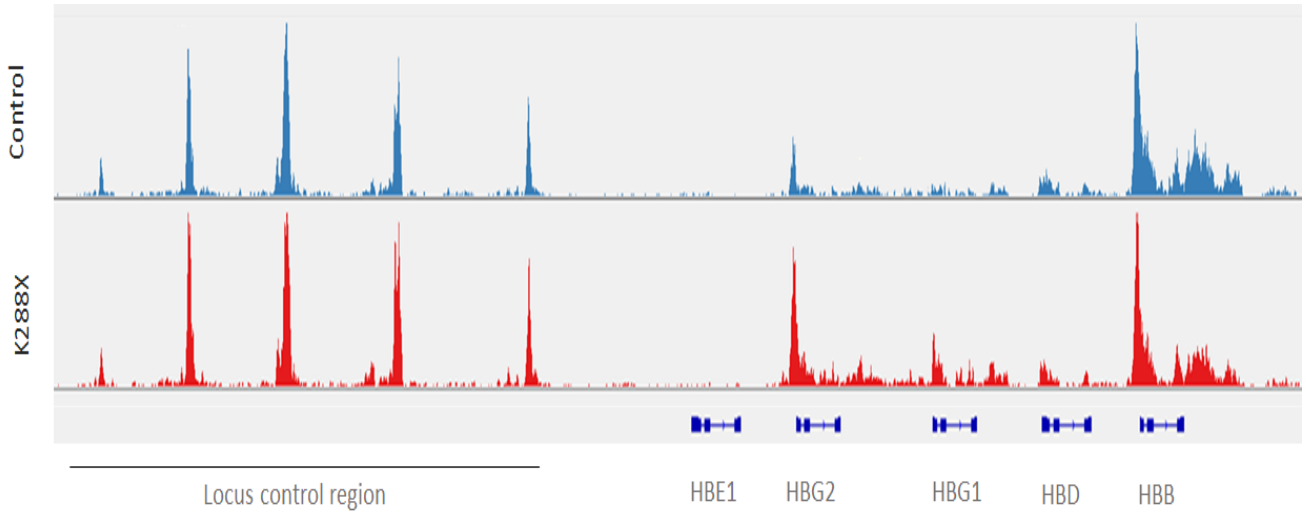

Figure 3.13: Sequence tracks from ATAC sequence. This figure shows the sequence tracks of the LCR, HBE1, HBE2, HBG1, HBD and HBB from ATAC sequence from controls (peaks in blue) and p.Lys288Ter heterozygotes (peaks in red). ATAC sequencing was carried out on erythroid cells after 48 hours from differentiation

.

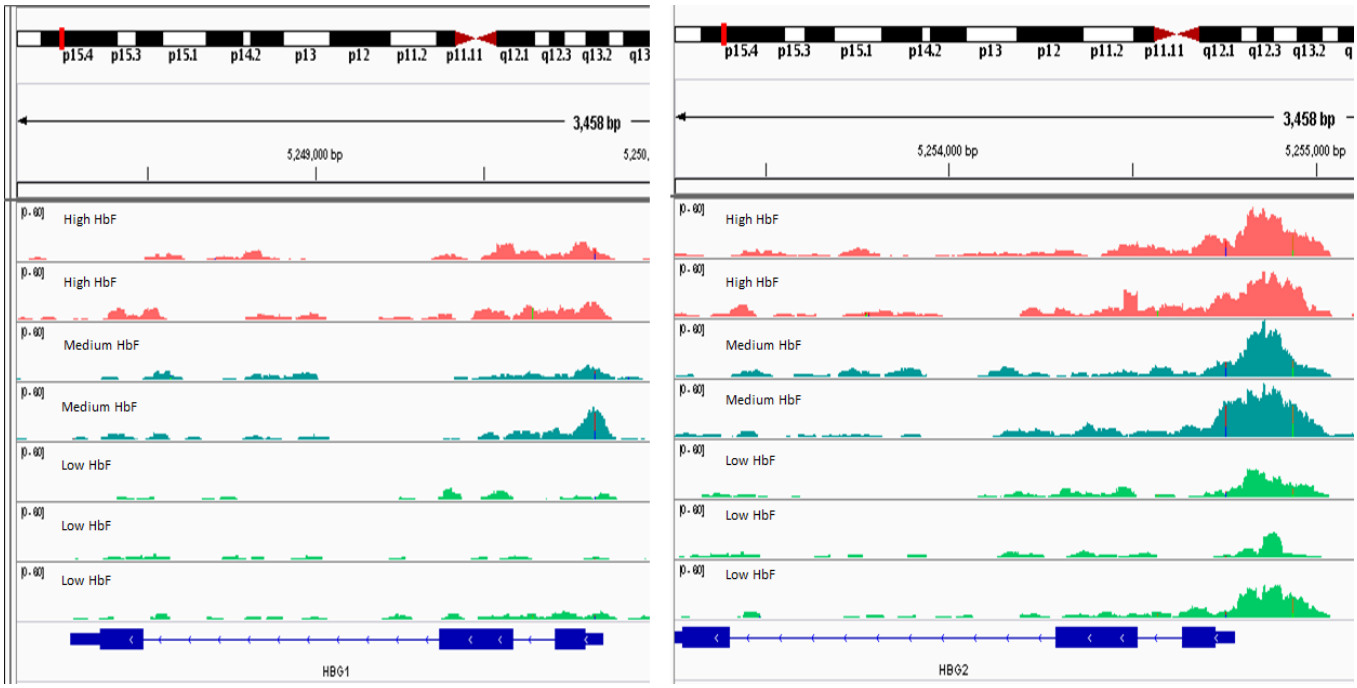

Figure 3.14: Sequence tracks from ATAC sequencing using IGV. The figure on the left shows HBG1 and the figure on the right shows HBG2. Peaks of subjects with High HbF are shown in red, those with medium HbF are shown in blue while those with low HbF are shown in green.

3.3.6 Integration of RNA Sequencing with ATAC sequencing

To address the effect of KLF1 haploinsufficiency on chromatin accessibility in relation to gene expression findings from RNA and ATAC sequencing were compared. The majority of the differentially expressed genes showed increased chromatin accessibility upon differentiation (T48) when compared to T0. The difference in chromatin accessibility between T0 and T48 was irrespective of sample genotype (figure 3.15)

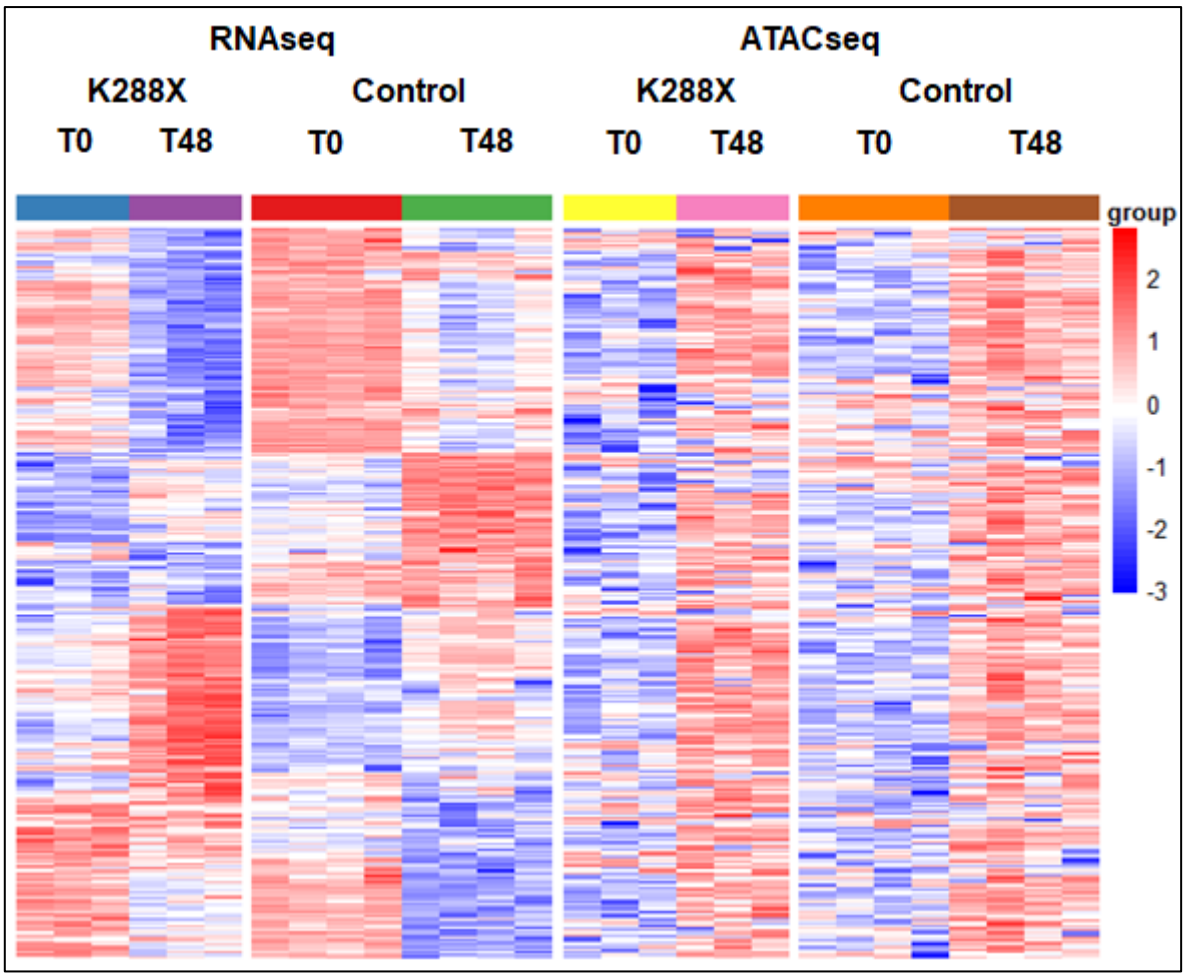

Figure 3.15: Integration of RNA sequencing and ATAC sequencing. Heatmap of RNA sequening and ATAC sequencing peak data. The heatmap on the left shows' RNA-seq data for K288X subjects and controls at T0 and T48. The heatmap on the right shows ATAC-seq data for K288X subjects and controls at T0 and T48.

When comparing Controls to p.K288X subjects the HbG1/2 promoters showed differential accessibility which was paired with differential expression. In the case of HBB and BCL11A

promoter the differential expression was not paired with difference in chromatin accessibility (figure 3.16).

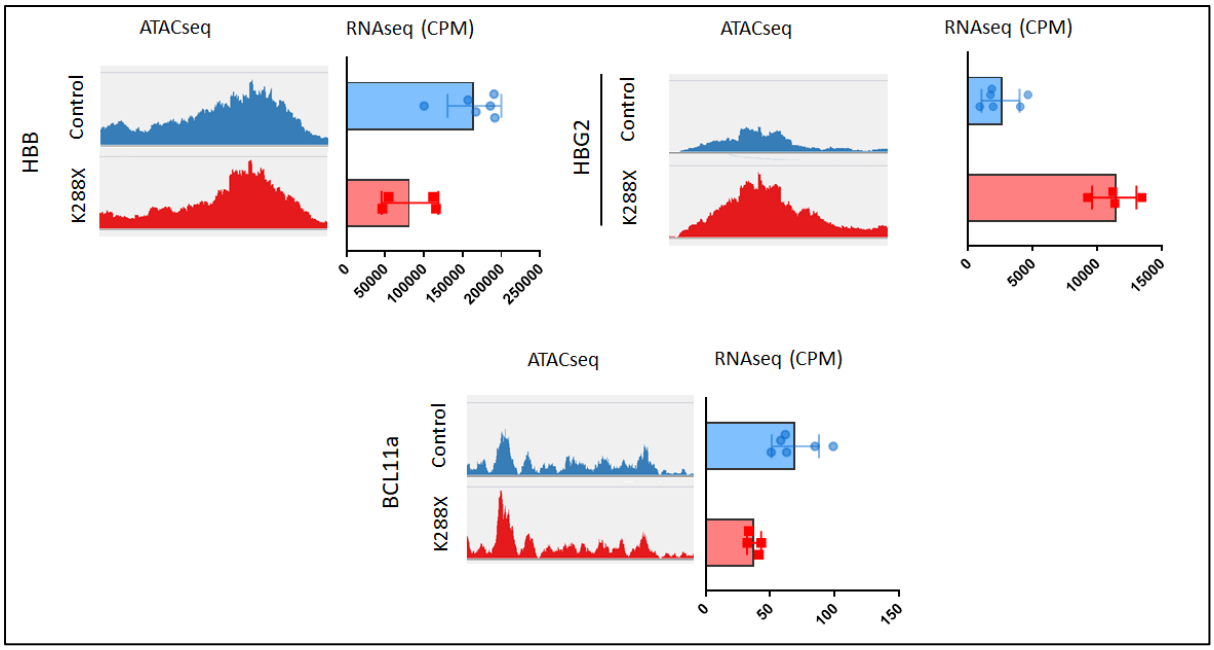

Figure 3.16: Integration of ATACseq and RNAseq. Sequencing tracks from ATACseq with corresponding CPM values from RNAseq in control and KLF1 p.K288X patient cells

3.4 Subjects with HbA_2 between 3.1% and 6.7%

We wanted to investigate the effect of *KLF1* variations together with other cis and trans regulators on subjects with HbA_2 between 3.1% and 6.7%.

3.4.1 Demographic and haematological Data of controls and subjects with HbA_2 between 3.1% and 6.7%

As explained in section 2.2, after carrying out a retrospective search in the Malta BioBank, four hundred and twenty-six (426) subjects with HbA_2 between 3.1% and 6.7%, together with 200 subjects with HbA_2 between 2.1% and 3.0% and normal CBC and normal ferritin levels were collected. β-globin gene sequencing was performed, and the 426 subjects were divided into

two categories according to their β genotyping as shown in figure 3. 17. Those subjects that had a phenotype resembling β-thalassaemia, but had no variation in the beta globin gene, are referred to as Pseudo-thalassaemia, while those that were heterozygote for a beta gene variation are referred to as β-thalassaemia heterozygotes.

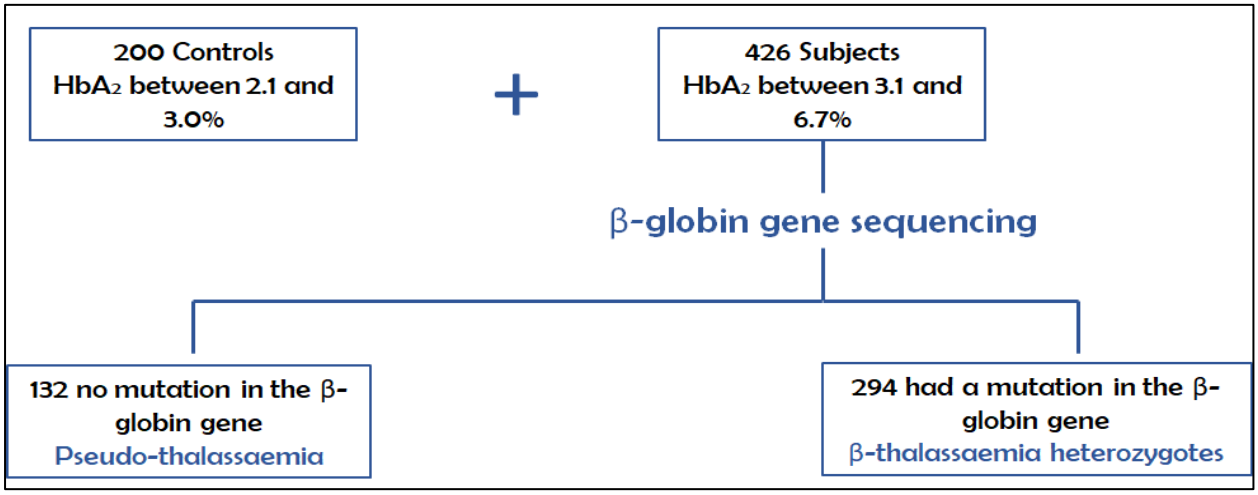

Figure 3.17: Classification of 426 subjects according to their -gene sequencing

3.4.1.1 Controls

Two-hundred (200) subjects with an HbA_2 between 2.1% and 3.0% (2.7% ±0.01) and HbF between 0.1 and 19.6mg/dL (4.94mg/dL ±0.21) were collected from the Malta BioBank as control group. One hundred and twenty-six (126) were males, while 74 were female (not pregnant), between 21 and 69 years. All the CBC parameters were normal, as shown in table 3.4; the average of Hb was 14.8 g/dL (±0.09), while the average mean of the MCV, MCH and MCHC were 85.1fL (±0.24), 29.3pg (±0.09) and 34.5g/dL (±0.06) respectively (table 3.5).

	Controls	Pseudo-Thalassaemia	B-thalassaemia heterozygotes
N	200	132	294
Males (n)	126	62	129
Females (n)	74	70	165
HbA_2 (%)	2.7 (±0.01)	3.21 (±0.01)	4.35 (±0.04)
HbF (mg/dL)	4.94 (±0.21)	83.8 (±8.14)	109.4 (±11.97)
Hb (g/dL)	14.8 (±0.09)	13.1 (±0.26)	12.2 (±0.76)
MCV (fL)	85.0 (±0.24)	80.2 (±1.42)	67.1 (±0.32)
MCH (pg)	29.3 (±0.09)	27.5 (±0.50)	21.9 (±0.12)
MCHC (g/dL)	34.5 (±0.06)	31.7 (±0.89)	29.7 (±0.49)

Table 3.5: Demographic and haematological data (mean and ±SE) in the control group, pseudo-thalassemia group and the β-thalassaemia heterozygote group. All controls and pseudo-thalassaemia subjects had normal ferritin levels and normal haemoglobin (>10g/dL). All pregnant females and children below 2 years of age were excluded from the study.

3.4.1.2 Pseudo-thalassaemia group

One hundred and thirty-two subjects (132) with HbA_2 between 3.1% and 3.9% and no variation in the β-globin gene. They were between 6 years and 90 years of age (mean 33 years). Out of the 134 subjects, 62 were males, while 70 were female. The HbF (mg/dL) ranged between 11.4mg/dL to 666.7 mg/dL (83.8 ±8.14), while the HbA_2 (%) was between 3.1 and 3.9 (3.21 ±0.01). The Hb (g/dL) was between 10.7 and 17.2 (13.1 ±0.26), while the MCV (fL) was between 61.2 and 101.2fL (80.2 ±1.42). The MCH (pg) was between 19.9 and 34.0 (27.5 ±0.50), while the MCHC (g/dL) was between 29.1 and 36.9 (31.7 ±0.89) (table 3.5).

3.4.1.3 β-thalassemia heterozygotes

Two hundred and ninety-four (294) beta thalassaemia heterozygotes with HbA_2 (%) between 3.1 and 6.7 (4.35 ±0.04) and HbF (mg/dL) between 10 and 2012.1 (109.4 ±11.97) were collected. One hundred and twenty-nine (129) were males, while 165 were females. They were between 3 and 86 years of age. The MCV (fL), MCH (pg) and MCHC (g/dL) were between

52.1 and 89.0 (fL) (67.1 ±0.32) 18.0 and 33.8 (pg) (21.9 ±0.12) and 23.8 and 36.1 (g/dL) (29.7± 0.49) respectively (table 3.5).

Two-hundred and twenty-seven (227) (77.2%) were IVS-I-6 (T>C), 39 (13.3%) were Codon 39 (C>T), 16 (5.4%) were IVS-I-110 (G>A), while 12 (4.1%) were IVS-II-1 (G>A). Table 3.6 shows the Mean and Standard Error of Mean (SEM) of HbA_2 (%), HbF(mg/dL), Hb (g/dL), MCV(fL), MCH(pg) and MCHC(g/dL) among the different β-thalassemia variations.

β-thalassaemia variation	HbA_2 (%)	HbF (mg/dL)	Hb (g/dL)	MCV (fL)	MCH (pg)	MCHC (g/dL)
IVS-I-6 (T>C) (n=227)	4.1 (±0.28)	103.4 (±14.82)	12.4 (±0.08)	68.4 (±0.33)	22.5 (±0.13)	30.0 (±0.55)
Codon 39 (C>T) (n=39)	5.3 (±0.10)	144.46 (±22.08)	11.4 (±0.17)	62.3 (±0.78)	19.9 (±0.34)	30.1 (±1.19)
IVS-I-110 (G>A) (n=16)	4.8 (±0.12)	75.5 (±20.82)	11.8 (±0.28)	64.2 (±0.86)	20.6 (±0.32)	30.1 (±1.69)
IVS-II-1 (G>A) (n=12)	5.6 (±0.22)	160.1 (±41.02)	11.3 (±0.26)	63.0 (±1.87)	20.4 (±0.71)	22.9 (±3.98)

Table 3.6: Demographic and haematological data (mean ± SEM) of HbA_2(%), HbF (mg/dL), Hb (g/dL), MCV (fL), MCH (pg) and MCHC (g/dL) in subjects with different Beta thalassaemia variations.

3.5 Spectrum of KLF1 variations in the Control Group, Pseudo-thalassaemia and β-thalassaemia heterozygotes

3.5.1 Spectrum of KLF1 variations in Control Group

Seventy-six (38%) had a single nuclear variation (SNP) in the *KLF1* gene. The most frequent variation was the pSer102Pro (rs2072597), present in 50 (25%), followed by the promoter

substitutions; -251C>G (rs3817621) present in 35 (17.5%) and the -148G>A present in 10 (5%). The pMet39Leu (rs112631212) was found in 5 (2.5%), the promoter variation 1133C>A (rs112943513) was in 2 (2.5%) while the variation pPhe182Leu (rs2072596) and the 3' UTR variants; 133G>A (rs922186305) and 277C>T (rs16978757) were present in 1, as shown in figure 3.18.

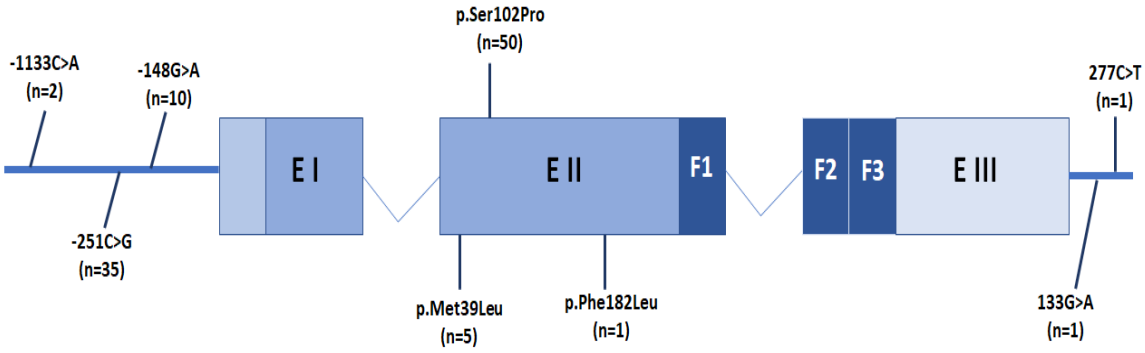

Figure 3.18 An extended map of the KLF1 gene showing all the SNV found in the control group. EI, EII and EIII refer to exon 1, exon 2 and exon 3. The dark blue boxes (F1, F2 and F3) represent the DNA binding domains

Fifty-one (51) (25.5%) were heterozygote for one variation, while 20 (10%) were compound heterozygote; 13 were compound heterozygote for the -251C>G and pSer102Pro, 4 were compound heterozygote for the pMet39Leu and pSer102Pro, 1 was compound heterozygote for the -148G>A and pSer102Pro, while another was compound heterozygote for the pSer102Pro and pPhe182Leu. One was compound heterozygote for the promoter substitution -251C>G and the 3'UTR 133G>A. Four had 3 variations in the *KLF1* gene, 2 had the -251C>G together with the -148G>A and the pSer102Pro, 1 had the -1133C>A substitution together with pMet39Leu and the pSer102Pro and 1 had the -251C>G substitution together with the pSer102Pro and 277C>T.

The Kurskal Wallis non-parametric showed no statistically significant difference in the HbA_2 and HbF levels between subjects with no KLF1 variations and subjects with a KLF1 variation.

3.5.1.1 Spectrum of KLF1 variations in Pseudo-thalassaemia group

Seventy-three (73 = 55.3%) had a variation in the *KLF1* gene. Most variations were present either in the promoter or in exon 2 (figure 3.19). No variations were found in the zinc fingers or in exon 3. The most frequent variation was the missense variant p.Ser102Pro (rs2072597), present in 65 (49.2%), followed by the promoter variant the -251C>G (rs3817621) found in 43 (32.5%) subjects.

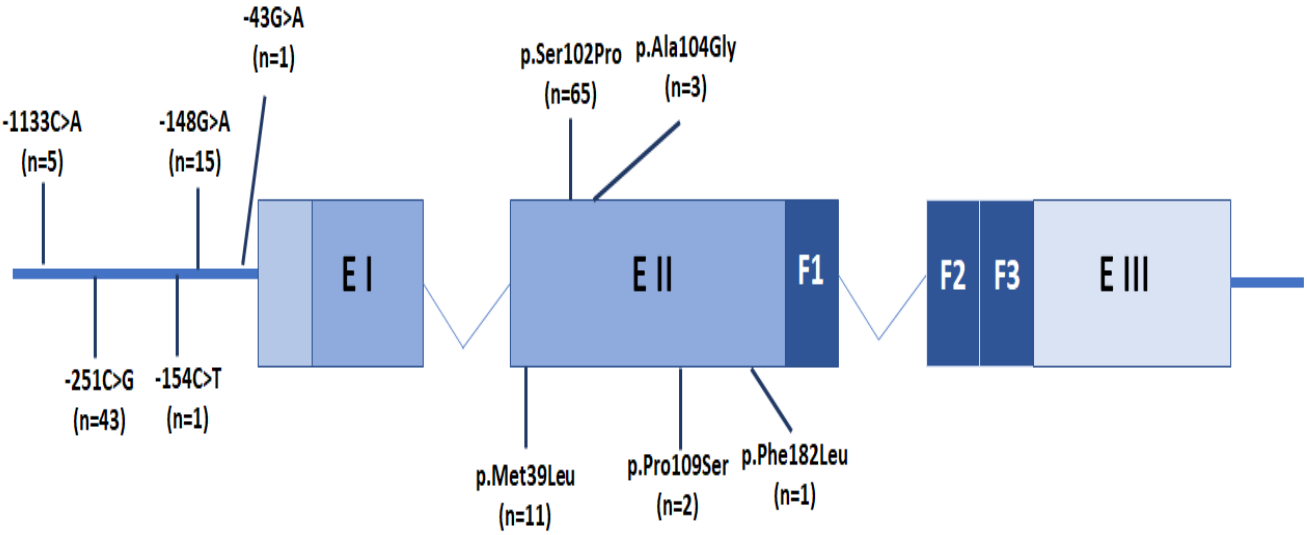

Figure 3.19: An extended map of the KLF1 gene showing all the SNV found in the Pseudo-thalassaemia group. EI, EII and EIII refer to exon 1, exon 2 and exon 3. The dark blue boxes (F1, F2 and F3) represent the DNA binding domains

Fifteen (11.4%) had the -148G>A (rs79334031) substitution, while 11 (8.3%) had the p.Met39Leu (rs112631212) substitution. Five subjects (3.8%) had the -1133C>A (rs112943513) substitution, while 3 subjects (2.3%) had the p.Ala104Gly (rs182276666) variation. The p.Pro109Ser (rs117351327) was found in 2 (1.5%) subjects, while the -154C>T (rs372651309), the -43G>A (rs372359976) and the p.Phe182Leu (rs2072596) was found in 1 subject.

Out of the 73 with *KLF1* variation, 38 (52.1%) were compound heterozygote, while 15 (20.5%) were triple heterozygote. Twenty-nine (29) were compound heterozygote for the -251C>G and p.Ser102Pro, 1 was compound heterozygote for the -1133C>A and p.Ser102Pro, 3 were compound heterozygote for the p.Met39Leu and p.Ser102Pro, 4 were compound heterozygote for the -148G>A and p.Ser102Pro, while 1 was compound heterozygote for the -251C>G and -148G>A. Out of the 15 triple heterozygotes, 8 had the -251C>G together with the -148G>A and pSer102Pro, 4 had the -1133C>A together with the p.Met39Leu and p.Ser102Pro, 1 had the -148G>A together with -43G>A and p.Ser102Pro, 1 had the -154C>T together with the pMet39Leu and p.Ser102Pro, while another was triple heterozygote for the p.Met39Leu together with p.Ser102Pro and p.Phe182Leu.

3.5.3 Spectrum of KLF1 variations in **Beta-thalassaemia heterozygous**

In the β-thalassaemia category, 119 (40.5%) had a variation in the *KLF1* gene. The most common variation was also the p.Ser102Pro (rs2072597), present in 107 (36.4%), followed by the promoter variation -251C>G (rs3817621), present in 59 (20.1%). Twenty-two (7.5%) had the p.Met36Leu (rs112631212), 17 (5.74%) had the -148G>A (rs79334031), while 13 (4.4%) had the -1133C>A (rs112943513). Nine (3.1%) had the p.Ala104Gly (rs182276666), while another 9 had the 277C>T (rs16978757) and the 296G>A (rs16978754) variations. One had the -102T>G (rs548543206) promoter variation, while another subject had the p.Pro109Ser (rs117351327) and another one had the p.Phe182Leu (rs2072596) as seen in figure 3.20.

In the β-heterozygote group, 58 were compound heterozygote for a KLF1 variation, 24 were triple heterozygote, while 4 subjects had four variations in the *KLF1* gene. Thirty-three were compound heterozygote for the -251C>G variation and the p.Ser102Pro, 7 were compound

heterozygote for the p.Met39Leu and p.Ser102Pro, 6 were compound heterozygote for the -148G>A and p.Ser102Pro, 6 were compound heterozygote for the 3'UTR variants 277C>T and 296G>A, 4 were compound heterozygote for the p.Ala104Gly and p.Ser102Pro, while one subject was compound heterozygote for the -251C>G and -148 G>A and another was compound heterozygote for the -102T>G promoter variation and p.Ser102Pro.

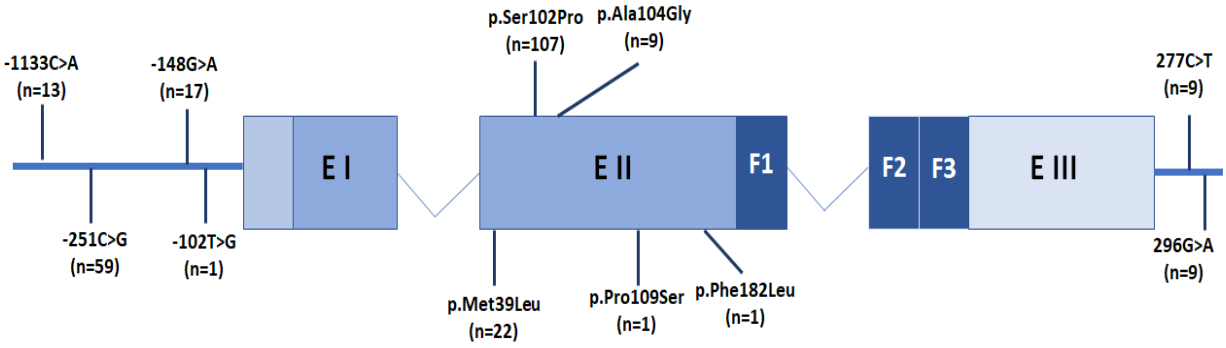

Figure 3.20: An extended map of the KLF1 gene showing all the SNV found in the β-thalassaemia heterozygotes. EI, EII and EIII refer to exon 1, exon 2 and exon 3. The dark blue boxes (F1, F2 and F3) represent the DNA binding domains

Eleven were triple heterozygote for the -1133C>A variation, together with the p.Met39Leu and the p.Ser102Pro, 9 were triple heterozygote for the -1133C>A, together with the -148G>A and the p.Ser102Pro, 1 had the p.Ser102Pro, together with the 277 C>T and 296G>A, another had the -251C>G, together with the p.Met39Leu and p.Ser102Pro, while another was triple heterozygote for the -251C>G, together with the p.Ser102Pro and the p.Phe182Leu. Two had the -251C>G and the -148G>A variations, together with the p.Met39Leu and the p.Ser102Pro, one subject had the -251C>G promoter variant, together with the p.Ser102Pro and the two 3'UTR variations; the 277 C>T and 296G>A, while one had the p.Met39Leu, together with the p.Ser102Pro and the two 3'UTR variations.

3.6 Effect of individual KLF1 SNVs on HbA_2 and HbF

We investigated the effect of the individual KLF1 single nucleotide variants on HbA_2 and HbF in the control, pseudo-thalassaemia and β-thalassaemia subjects.

3.6.1 Effect of the promoter variations -1133G>A and -251C>G on HbA_2 and HbF

We analysed the effect of the promoter variations; the -113G>A and -251C>G on HbA_2 and HbF, in the control group, the pseudo-thalassaemia group, and the β-thalassaemia group. Since the data was not normally distributed, the Kurskal-Wallis non-parametric test was used.

In the control group (table 3.7) 198 subjects were wildtype for the -1133C>A variation, while 2 were heterozygote. The HbA_2 level in the wildtype was that of 2.6%, while in the heterozygote it was 2.8%, while the HbF level in the wildtype was 4.93mg/dL while in the heterozygote it was 5.2mg/dL. No statistically significant difference was observed in the HbA_2 and HbF between the wildtype and heterozygote. In the Pseudo-thalassaemia group, 128 were wildtype for the -1133C>A variation, while 5 were heterozygote. A statistically significant difference in the HbA_2 ($p=0.04$) was observed between the two groups. The HbA_2 in the wildtype was 3.2%, while in the heterozygote it was 3.4%. The HbF in the wildtype was 84.45mg/dL, while in the heterozygotes it was 123.4mg/dL. In the β-thalassemia heterozygote group 281 were wildtype for the -1133C>A variation, while 13 were heterozygote. No difference was observed in the HbA_2 level between the 2 groups as both had an HbA_2 of 4.3%. The HbF level in the wildtype group was 102.3mg/dL, while the HbF level in the heterozygote group was 274mg/dL.

Promoter variation	Genotype	Control			Pseudo-thalassemia			β-thalassemia heterozygotes		
		N	**HbA_2 (%)**	**HbF (mg/dL)**	**N**	**HbA_2 (%)**	**HbF (mg/dL)**	**N**	**HbA_2 (%)**	**HbF (mg/dl)**
-1133C>A	CC	198	2.6 (±0.01)	4.93 (±0.21)	128	3.2 (±0.01)	84.45 (±8.57)	281	4.3 (±0.04)	102.3 (±10.48)
	CA	2	2.8 (±0.10)	5.2 (±0.38)	5	3.4 (±0.08)	123.4 (±71.80)	13	4.3 (±0.17)	274.2 (±151.7)
P value			0.4	0.46		**0.04**	0.93		0.46	0.10
-251C>G	CC	165	2.7 (±0.01)	4.81 (±0.23)	90	3.21 (±0.01)	83.34 (±12.05)	235	4.3 (±0.04)	115.7 (±14.6)
	CG	31	2.6 (±0.03)	4.99 (±0.43)	33	3.21 (±0.02)	72.79 (±8.18)	47	4.3 (±0.10)	93.3 (±21.3)
	GG	4	2.7 (±0.06)	10.05 (±2.64)	10	3.23 (±0.02)	127.95 (±35.64)	12	4.5 (±0.24)	69.1 (±23.6)
P value			0.52	**0.04**		0.43	0.06		0.82	0.40

Table 3.7: Effect of promoter variations -113C>A and -251C>C on HbA_2 (%) and HbF (mg/dL) in the control group, pseudo-thalassaemia group and β-thalassaemia heterozygote group. Statistically Significant p values are shown in bold.

In the control group, 165 subjects were wildtype for the -251C>G variation, 31 were the heterozygote and 4 were homozygote. The HbA_2 was 2.7% for both the wildtype and homozygote group and 2.6 for the heterozygote group. A statistically significant difference (p=0.04) was noted in the HbF. Subjects wildtype and heterozygote for the -251C>G variation had an HbF of 4.81mg/dL and 4.99mg/dL respectively, while those homozygotes had a higher HbF of 10.05mg/dL. In the β-thalassaemia heterozygote group, 235 were wildtype, 47 were heterozygote, while 12 were homozygote. The HbA_2 was 4.3% for both the wildtype and heterozygote, while the homozygotes had an HbA_2 of 4.5%. The HbF in the wildtype group was 115.7mg/dL, in the heterozygote group it was 93.3mg/dL while in the homozygote group it was 69.1mg/dL.

3.6.2 Effect of the other KLF1 SNVs on the HbA_2 and HbF in pseudo-thalassaemia and β-thalassaemia heterozygotes

We investigated the effect of the other KLF1 variations (excluding the -1133G>A variation and the -251 C>G variation) on the HbA_2 (%) and HbF (mg/dL) in the pseudo-thalassaemia and β-thalassaemia heterozygote group. Since Data is not normally distributed, the non-parametric Mann-Whitney U test was used to see if there is any statistically significant difference between the HbA_2 and HbF of the wildtype group and the other groups. No difference in the HbA_2 and HbF was noted in the control group between subjects with a KLF1 variation and subjects with no KLF1 variation.

3.6.2.1 Pseudo-thalassaemia group

The pseudo-thalassaemia group was divided into 12 categories according to the type of *KLF1* variation. As seen in table 3.8 and figure 3.21, the mean of the HbA_2 (%) for subjects with no KLF1 variation was 3.19 (±0.01), while that of the HbF (mg/dL) was 80.9 (±11.7). KLF1 heterozygotes had more or less the same HbA_2 level and HbF level, although the p.Pro109Ser heterozygotes had an HbA_2 of 3.4%. A statistically significant difference in HbA_2 level was noted in compound heterozygotes for the p.Met39Leu and the p.Ser102Pro (HbA_2 3.3%; p=0.013) and subject heterozygote for p.Met39Leu, p.Ser102Pro and p.Pro182Leu (HbA_2 3.9%; p=0.03).

KLF1 genotype	N	HbA2 (%)	P value	HbF (mg/dL)	P value
KLF1 wildtype	62	3.19 (±0.01)	-	80.9 (±11.7)	-
-148 G>A heterozygote	2	3.15 (±0.05)	0.671	41.1 (±2.4)	0.270
p.Ser102Pro heterozygote	35	3.19 (±0.01)	0.696	67.1 (±6.91)	0.707
p.Ala104Gly heterozygote	2	3.15 (±0.07)	0.671	62.7 (±10.3)	0.923
p.Pro109Ser heterozygote	1	3.4	0.127	66.4	0.984
-148 G>A & p.Ser102Pro	12	3.23 (±0.35)	0.175	81.3 (±18.67)	0.741
p.Met39Leu & p.Ser102Pro	9	3.32 (±0.05)	**0.013**	107.6 (±39.9)	0.717
p.Pro109Ser & p.Ser102Pro	1	3.20	0.825	55.2	0.762
p.Ser102Pro homozygote	5	3.24 (±0.04)	0.246	161.6 (±68.5)	0.075
-43G>A, -148G>A & p.Ser102Pro	1	3.10	0.444	604.8	**0.050**
-154C>T, p.Met39Leu & p.Ser102Pro	1	3.1	0.444	25.6	0.317
p.Met39Leu, p.Ser102Pro & p.Phe182Leu	1	3.9	**0.032**	14.7	0.254

Table 3.8: Association of KLF1 genotypes with HbA_2 and HbF in the pseudo-thalassaemia group. The table shows the different KLF1 genotypes, the mean and SEM of HbA_2 (%) and HbF (mg/dL). Statistically significant p values are marked in bold.

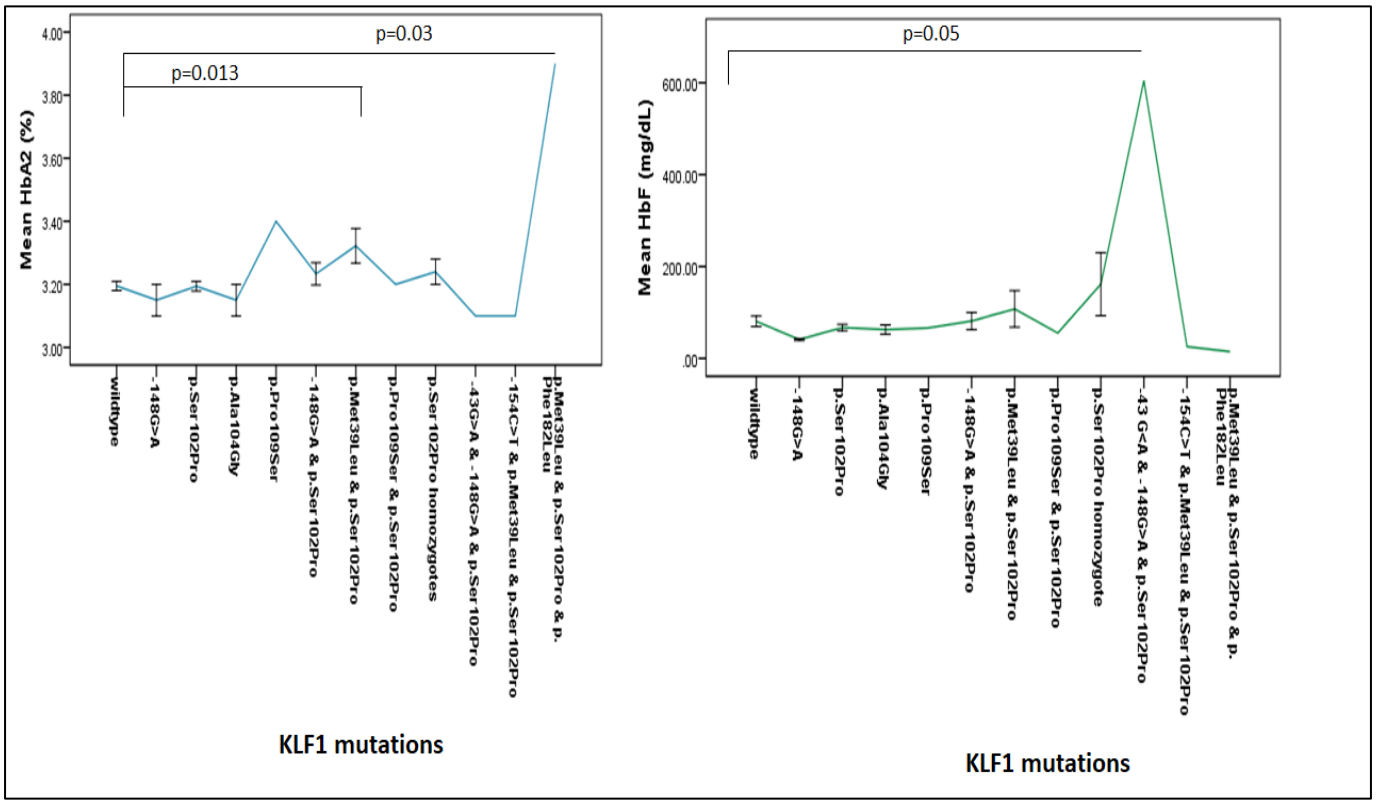

Figure 3.21: Line graph of the Mean HbA_2 (%) and HbF (mg/dL) according to the KLF1 variation in the pseudo thalassaemia group. The graph on the left represent the HbA_2 (%) while the graph on the right represent the HbF (mg/dL). The error bars show +/- 1 SE. Statistically significant difference between the wildtype group and the other groups is shown.

The HbF (mg/dL) was similar when comparing the different KLF1 genotypes to the wildtype, with the exception of subjects compound heterozygotes for the p.Met39Leu and p.Ser102Pro, where the HbF level was 107.6mg/dL (±39.9) while subjects homozygotes for the p.Ser102Pro had an HbF level of 161.6mg/dL (±68.5). The highest HbF that of 604.8mg/dL was seen in a triple heterozygote for the promoter variations -43G>A and -148G>A and the p.Ser102Pro.

3.6.2.2 β-thalassaemia heterozygotes

The β-thalassaemia heterozygotes were divided into 16 categories, according to their *KLF1* variation (table 3.9 and figure 3.22). One hundred and seventy-seven subjects had no variation in the *KLF1* gene and their HbA_2 was 4.35% (±0.05), while their HbF was 109.2mg/dL. Subjects heterozygotes for one KLF1 variation had similar HbA_2 and HbF levels to the wildtype group. An HbA_2 of 4.8% (±0.55) was seen in a triple heterozygote for the p.Ser102Pro, together with the two 3'UTR variations the 296G>A and the 277C>T, while an HbA_2 of 5.0% was observed in a subject who was IVS-I-6 heterozygote but was homozygote for both the 3' UTR variations. The highest HbA_2 level was noted in a subject heterozygote for the IVS-II-1 β-variation and compound heterozygote for the p.Ser102Pro and the p.Phe182Leu. This subject had an HbA_2 of 6.4% (p=0.045).

KLF1 genotype	N	HbA_2 (%)	P value	HbF (mg/dL)	P value
KLF1 wildtype	177	4.35 (±0.05)	-	109.2 (±13.2)	-
p.Ser102Pro heterozygote	56	4.34 (±0.09)	0.989	90.0 (±18.9)	0.271
p.Ala104Gly heterozygote	3	4.16 (±0.24)	0.867	52.3 (±20.4)	0.659
p.Pro109Ser heterozygote	1	3.9	0.562	64.5	0.978
-102T>G & p.Ser102Pro	1	4.2	0.966	20.1	0.393
-148 G>A & p.Ser102Pro	13	4.25 (±0.10)	0.869	55.6 (±9.88)	0.243
p.Met39Leu & p.Ser102Pro	19	4.35 (±0.14)	0.568	281.7 (±120.9)	0.143
p.Ala104Gly & p.Ser102Pro	6	4.51 (±0.35)	0.677	54.3 (±16.11)	0.05
p.Ser102Pro & p.Phe182Leu	1	6.4	**0.045**	203.4	0.213
296G>A & 277C>T heterozygote	5	4.26 (±0.39)	0.549	21.3 (±12.33)	0.012
p.Ser102Pro homozygote	6	4.1 (±0.17)	0.515	53.5 (±13.7)	0.466
296G>A & 277G>T homozygote	1	5.0	0.326	358.4	0.090
-148G>A, p.Met39Leu & p.Ser102Pro	2	4.00 (±0.10)	0.506	11.65 (±0.65)	0.015
p.Ser102Pro, 296G>A & 277G>T	2	4.8 (±0.55)	0.254	98.5 (±85.3)	0.989
p.Met39Leu, p.Ser102Pro, 296G>A & 277G>T	1	4.2	0.966	42.8	0.663

Table 3.9: Association of KLF1 genotypes with HbA_2 and HbF in the β-thalassaemia group. The table shows the different KLF1 genotypes, the mean and SEM of HbA_2 (%) and HbF (mg/dL). Statistically significant p values are marked in bold.

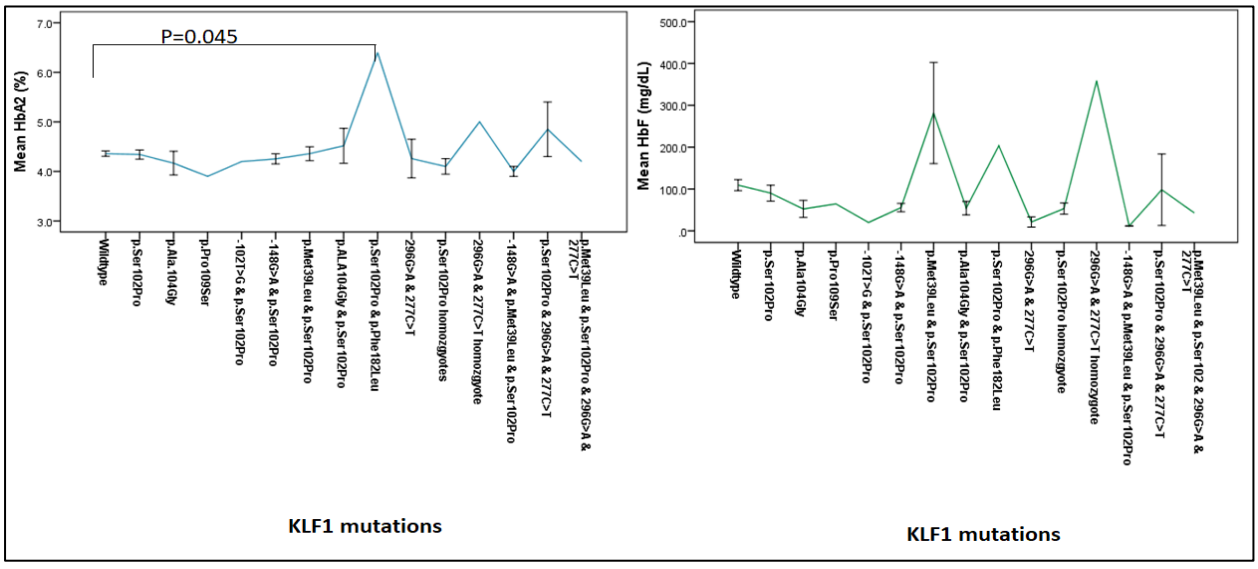

Figure 3.22: Line graph of the Mean HbA_2 (%) and HbF (mg/dL) according to the KLF1 variation in the β-thalassaemia heterozygote group. The graph on the left represent the HbA_2 (%) while the graph on the right represent the HbF (mg/dL). The error bars show +/- 1 SE. Statistically significant difference between the wildtype group and the other groups is shown.

The HbF in the wildtype group was similar to those heterozygotes for a KLF1 variation . An increase in HbF was seen in a compound heterozygote for the p.Ser102Pro and the p.Phe182Leu. This subject had an HbF of 203.4mg/dL. An HbF of 281.7mg/dL (±120.9) was seen in p.Ser102Pro homozygote variation, while the highest HbF observed was that of 358.4mg/dL in a homozygote for both the 3' UTR variations the 296G>A and the 277C>T.

3.7 Interplay of cis and trans-regulators on HbA_2 and HbF in pseudo-thalassaemia and β-thalassaemia heterozygote subjects

The effect of *cis*-regulators XMN1 (rs7482144) and $(AT)_xT_y$ motif, and the trans-regulators; BCL11A (rs4671393) polymorphisms, and the HBS1L-Myb haplotypes on HbA_2(%), HbF(mg/dL), in pseudo-thalassaemia and β-thalassaemia heterozygote was investigated. The effect of KLF1 together with the *cis* or *trans* regulator was also investigated. For statistical purposes we looked only at the largest groups. Since the data is not normally distributed, the Kruskal-Wallis non-parametric test was used.

3.7.1 Effect of cis-regulators on HbA_2 and HbF

3.7.1.1 Effect of XMN1 on HbA_2 and HbF in the pseudo-thalassaemia and β-thalassaemia group

In the pseudo-thalassaemia group, 98 subjects were wildtype for the XMN1 polymorphism, 32 were heterozygote, while 4 were homozygote for this polymorphism. No statistically significant difference was noted in the mean of HbA_2(%) between the wildtype, the heterozygote and the homozygote groups. A statistically significant difference (p=0.028) was seen in the HbF (mg/dL) between subjects' wildtype, heterozygote or homozygote for the XMN1 polymorphism (table 3.10). In the β-thalassemia heterozygote subjects, 230 were wildtype for the XMN1 polymorphism, while 60 were heterozygote and 4 were homozygote.

No statistically significant difference was observed in HbA_2 and HbF between the wildtype, heterozygote and homozygote.

Polymorphism	Genotype	Pseudo-thalassemia			β-thalassemia heterozygotes		
		N	HbA_2 (%)	HbF (mg/dL)	N	HbA_2 (%)	HbF (mg/dl)
XMN1 rs7482144	CC	98	3.2 (±0.01)	82.6 (±10.1)	230	4.37 (±0.04)	100.5 (±11.15)
	CT	32	3.18 (±0.01)	94.0 (±13.7)	60	4.25 (±0.07)	145.0 (±40.2)
	TT	4	3.2 (±0.06)	33.3 (±9.94)	4	4.8 (±0.35)	110.2 (±62.2)
P value			0.49	**0.028**		0.29	0.91

Table 3.10: Effect of XMN1 (rs7482144) on HbA_2 (%) and HbF (mg/dL) in the pseudo-thalassaemia group and β-thalassaemia heterozygote group. Statistically Significant p values are shown in bold

3.7.1.1.1 Effect of KLF1 together with XMN1 on HbA_2 and HbF

The effect of Xmn1 polymorphism combined with the effect of KLF1 variants on HbA_2 and HbF was studied.

KLF1	XMN1 Genotype	Pseudo-thalassemia				
		N	HbA_2 (%)	P Value	HbF (mg/dL)	P Value
KLF1 Wildtype	CC	42	3.2 (±0.01)	-	70.4 (±7.89)	-
	CT	16	3.23 (±0.07)	0.506	85.3 (±13.2)	0.227
	TT	2	3.15 (±0.07)	0.611	21.5 (±6.5)	0.076
p.Ser102Pro heterozygote	CC	20	3.19 (±0.09)	0.880	64.4 (±8.84)	0.531
	CT	12	3.16 (±0.04)	0.506	69.4 (±11.1)	0.677
	TT	2	3.30 (±0.10)	0.279	45.1 (±16.5)	0.357
-148G>A & p.Ser102Pro	CC	10	3.23 (±0.04)	0.499	61.8 (±17.1)	0.457
	CT	2	3.25 (±0.07)	0.444	146.4 (±55.2)	0.063
p.Met39Leu & p.Ser102Pro	CC	7	3.34 (±0.11)	0.188	123.0 (±50.8)	0.644

Table 3.11: Effect of XMN1 (rs7482144) together with KLF1 variants on HbA_2 (%) and HbF (mg/dL) in the pseudo-thalassaemia group.

In both the pseudo-thalassemia and β-thalassemia heterozygotes groups, most of the subjects were wildtype for the KLF1 and XMN1. In the pseudo-thalassaemia group, no difference in the HbA_2 and HbF was noted when compared to the wildtype group. While in the β-thalassaemia group a statistically significant difference was noted in subjects heterozygote for the KLF1 p.Ser102Pro variations and homozygote for the XMN1 polymorphism (tables 3.11 & 3.12).

KLF1	XMN1 Genotype	β-thalassaemia heterozygotes				
		N	HbA_2 (%)	P Value	HbF (mg/dL)	P Value
KLF1 Wildtype	CC	139	4.37 (±0.06)	-	103.8 (±13.9)	-
	CT	33	4.21 (±0.09)	0.647	124.5 (±39.5)	0.600
	TT	2	4.25 (±0.35)	0.979	155 (±128.0)	0.658
p.Ser102Pro heterozygote	CC	40	4.3 (±0.74)	0.981	91.5 (±24.3)	0.442
	CT	14	4.1 (±0.10)	0.326	102.4 (±30.9)	0.919
	TT	2	5.3 (±0.15)	**0.05**	65 (±53.0)	0.779
-148G>A & p.Ser102Pro	CC	14	4.22 (±0.24)	0.851	102.3 (±11.6)	0.993
	CT	7	4.12 (±0.28)	0.816	35.3 (±11.3)	0.111
p.Met39Leu & p.Ser102Pro	CC	14	4.35 (±0.15)	0.439	173.8 (±92.9)	0.239
	TT	3	4.7 (±0.55)	0.522	890.4 (±590.0)	0.328

Table 3.12: Effect of XMN1 (rs7482144) together with KLF1 on HbA_2 (%) and HbF (mg/dL) in the β thalassaemia group.

3.7.1.2 EFFECT OF THE $(AT)_xT_Y$ MOTIF ON HBA_2 AND HBF

In the pseudo-thalassaemia group and the β-thalassaemia group, the most common $(AT)_xT_y$ motif was the $(AT)_7T_7$ present in 84 of the pseudo-thalassaemia group and 226 of the β-

thalassemia heterozygote group. In the pseudo-thalassaemia group, the 2nd most common motif was the $(AT)_8T_5/(AT)_9T_5$ present in 20 subjects, while in the β-thalassaemia group it is the $(AT)_7T_7/(AT)_9T_5$ present in 40 subjects (table 3.13).

In the pseudo-thalassaemia group, a statistically significant difference (p=0.007) was observed for the HbA_2 between the different $(AT)_xT_y$ motifs. An HbA_2 of 3.1% was observed in the $(AT)_7T_7$, $(AT)_8T_5$ and $(AT)_7T_7/(AT)_9T_5$ motifs, while an HbA_2 of 3.2% was noted in the $(AT)_{10}T_3$, $(AT)_7T_7/(AT)_{10}T_4$and $(AT)_8T_5/(AT)_9T_5$ motifs. An HbA_2 of 3.3% was noted in the $(AT)_9T_5$ and $(AT)_7T_7/(AT)_8T_5$ motifs, while the $(AT)_9T_5/(AT)_{11}T_3$ motif had an HbA_2 of 3.4%. The highest HbA_2, that of 3.5%, was noted in the $(AT)_{11}T_3$ motif. In the β-thalassaemia group a statistically significant difference (p=0.019) was also observed for the HbA_2 between the different $(AT)_xT_y$ motifs. The lowest HbA_2, that of 3.7%, was noted in the $(AT)_7T_7/(AT)_{10}T_4$, followed by an HbA_2 of 3.8%, noted in the $(AT)_8T_5/(AT)_9T_5$ motif. Subjects in the $(AT)_{11}T_3$ motif had an HbA_2 of 4.0%, while the $(AT)_7T_7$, the $(AT)_7T_7/(AT)_8T_5$ and $(AT)_7T_7/(AT)_{10}T_4$ had an HbA_2 of 4.3, 4.4 and 4.5% respectively. The $(AT)_9T_5/(AT)_{11}T_3$ had an HbA_2 of 4.7%, while the $(AT)_7T_7/(AT)_{11}T_3$ had an HbA_2 of 4.8%. The highest HbA_2 that of 5.1% was noted in the $(AT)_9T_5$ motif.

$(AT)_xT_y$ motif	Pseudo-thalassaemia			β-thalassaemia heterozygote		
	N	HbA_2 (%)	HbF (mg/dL)	N	HbA_2 (%)	HbF (mg/dL)
$(AT)_7T_7$	84	3.19 (±0.01)	71.9 (±6.46)	226	4.3 (±0.04)	108.2 (±14.1)
$(AT)_8T_5$	5	3.18 (±0.02)	183.7 (±105.5)	0		
$(AT)_9T_5$	4	3.3 (±0.01)	65.9 (±16.2)	7	5.1 (±0.35)	277.1 (±184.7)
$(AT)_{10}T_3$	2	3.2 (±0.01)	104.9 (±16.0)	0		
$(AT)_{11}T_3$	2	3.5 (±0.10)	97.9 (±53.0)	1	4.0	26.0
$(AT)_7T_7/(AT)_8T_5$	3	3.36 (±0.03)	94.5 (±39.0)	2	4.4 (±0.01)	88.3 (±50.0)
$(AT)_7T_7/(AT)_9T_5$	6	3.15 (±0.01)	53.1 (±9.54)	40	4.5 (±0.13)	90.7 (16.4)
$(AT)_7T_7/(AT)_{10}T_4$	2	3.25 (±0.07)	108.2 (±38.2)	2	3.70 (±0.28)	34.5 (±4.50)
$(AT)_7T_7/(AT)_{11}T_3$	3	3.13 (±0.03)	75.9 (±36.0)	5	4.88 (±0.28)	129.7 (±62.2)
$(AT)_8T_5/(AT)_9T_5$	20	3.24 (±0.04)	72.0 (±15.5)	6	3.88 (±0.15)	112.7 (±69.3)
$(AT)_9T_5/(AT)_{11}T_3$	1	3.4	410.4	5	4.7 (±0.48)	127.6 (±28.6)
P value		**0.007**	0.337		**0.019**	0.478

Table 3.13: Effect of the (AT)xTy motifs on the HbA_2 and HbF in the pseudo-thalassaemia and the β-thalassaemia heterozygote group. The mean and ±SE of each motif is shown while statistically significant p values are shown in bold.

In the pseudo-thalassaemia group, the $(AT)_7T_7/(AT)_9T_5$, the $(AT)_9T_5$, the $(AT)_8T_5/AT_9T_5$, the $(AT)_7T_7/(AT)_{11}T_3$ and the $(AT)_7T_7$ motifs had the lowest HbF between 53.1 and 75.9mg/dL. The $(AT)_7T_7/(AT)_8T_5$, the $(AT)_{11}T_3$ the $(AT)_{10}T_3$ and the $(AT)_7T_7/(AT)_{10}T_4$ had an HbF of 94.5, 97.9mg/dL, 104.9 and 108.2mg/dL respectively. An HbF of 183.7mg/dL was noted in the $(AT)_8T_5$ motif, while the $(AT)_9T_5/(AT)_{11}T_3$ had the highest HbF of 410.4mg/dL. In the β-thalassaemia heterozygote group, the $(AT)_7T_7/(AT)_{10}T_4$ had the lowest HbF that of 34.5mg/dL, followed by the $(AT)_7T_7/(AT)_8T_5$ motif with an HbF of 88.3mg/dL and the $(AT)_7T_7/(AT)_9T_5$ with an HbF of 90.7mg/dL. The $(AT)_7T_7$ motif had an HbF of 108.2mg/dL, while the $(AT)_8T_5/(AT)_9T_5$, the $(AT)_9T_5/(AT)_{11}T_3$, and the $(AT)_7T_7/(AT)_{11}T_3$, had an HbF of 112.7, 127.6 and 129.7mg/dL. The highest HbF level was that of 277.1mg/dL in the $(AT)_9T_5$.

3.7.1.2.1 Effect of KLF1 together with (AT)xTy motif on HbA_2 and HbF

When investigating the effect of KLF1 together with the (AT)xTy motif no difference in HbA_2 and HbF was noted between the different groups in the pseudo thalassaemia group (table 3.) and β-thalassaemia heterozygote group (tables 3.14 & 3.15)

KLF1	(AT)xTy Motif	Pseudo-thalassemia				
		N	HbA_2 (%)	P Value	HbF (mg/dL)	P Value
KLF1 Wildtype	$(AT)_7T_7$	39	3.22 (±0.03)	-	39.4 (±52.5)	-
	$(AT)_{10}T_3$	2	3.25 (±0.05)	0.351	108.8 (±12.2)	0.073
	$(AT)_9T_5$	1	3.30	0.400	36.3	0.500
	$(AT)_7T_7/(AT)_9T_5$	5	3.14 (±0.05)	0.449	55.5 (±11.3)	0.449
	$(AT)_8T_5$	9	3.21 (±0.09)	0.435	84.7 (±30.1)	0.876
p.Ser102Pro heterozygote	$(AT)_7T_7$	4	3.15 (±0.05)	0.479	44.7 (±6.09)	0.181
	$(AT)_8T_5$	15	3.18 (±0.02)	0.885	67.3 (±10.05)	0.978
	$(AT)_8T_5/(AT)_9T_5$	3	3.18 (±0.05)	0.910	69.8 (±13.6)	0.743

Table 3.14: Effect of (AT)xTy motifs together with KLF1 variants on HbA_2 (%) and HbF (mg/dL) in pseudo-thalassaemia.

KLF1	$(AT)_xT_y$ Motif	β-thalassaemia heterozygotes				
		N	HbA_2 (%)	P Value	HbF (mg/dL)	P Value
KLF1 Wildtype	$(AT)_7T_7$	138	4.29 (±0.56)	-	107.9 (±16.42)	-
	$(AT)_9T_5$	3	5.8 (±0.45)	0.08	93.9 (±56.7)	0.966
	$(AT)_7T_7/(AT)_9T_5$	23	4.40 (±0.18)	0.539	118.0 (±21.3)	0.204
p.Ser102Pro heterozygote	$(AT)_7T_7$	47	4.19 (±0.07)	0.611	89.4 (±19.5)	0.705
	$(AT)_7T_7/(AT)_9T_5$	10	4.80 (±0.29)	0.074	68.4 (±39.6)	0.029

Table 3.15: Effect of (AT)xTy motifs together with KLF1 variants on HbA_2 (%) and HbF (mg/dL) in the β thalassaemia group.

3.7.2 Effect of trans-regulators on HbA_2 and HbF

3.7.1.1 Effect of BCL11A on HbA_2 and HbF in pseudo-thalassaemia and β-thalassaemia heterozygote group

Among the pseudo-thalassemia group, 90 subjects were wildtype for the *BCL11A* rs4671393 polymorphism, while 39 subjects were heterozygote and 5 subjects were homozygote. In the β-thalassaemia heterozygote group, 216 subjects were wildtype for the BCL11A polymorphism, 69 subjects were heterozygote, while 9 subjects were homozygote. No statistically significant difference was observed in the mean of HbA_2 and HbF between wildtype, heterozygote and homozygote subjects for the *BCL11A* polymorphism in the pseudo-thalassaemia and β-thalassaemia heterozygote group (table 3.16).

Polymorphism	Genotype	Pseudo-thalassemia			β-thalassemia heterozygotes		
		N	HbA_2 (%)	HbF (mg/dL)	N	HbA_2 (%)	HbF (mg/dl)
BCL11A (rs4671393)	GG	90	3.2 (±0.01)	87.1 (±11.32)	216	4.3 (±0.05)	108.8 (±16.23)
	GA	39	3.1 (±0.01)	76.4 (±10.1)	69	4.27 (±0.08)	101.7 (±16.75)
	AA	5	3.1 (±0.02)	84.5 (±24.9)	9	4.55 (±0.23)	88.8 (±22.4)
P value			0.142	0.536		0.241	0.543

Table 3:16: Effect of BCL11A (rs4671393) on HbA_2 (%) and HbF (mg/dL) in the control group, pseudo-thalassaemia group and β-thalassaemia heterozygote group.

3.7.1.2.2 Effect of KLF1 together with BCL11A polymorphism on HbA_2 and HbF

In the pseudo-thalassaemia group a statistically significant increase in HbF was noted in a subject who was compound heterozygote for the p.Met39Leu together with the p.Ser102Pro and homozygote for the rs4671393 variation . His HbF was that of 410.4 mg/dl. No difference in HbA_2 was noted (table 3.17)

KLF1	BCL11a Genotype	Pseudo-thalassemia				
		N	HbA_2 (%)	P Value	HbF (mg/dL)	P Value
KLF1 Wildtype	GG	39	3.19 (±0.01)	-	72.6 (±9.24)	-
	GA	18	3.26 (±0.06)	0.518	67.6 (±8.88)	0.625
	AA	4	3.12 (±0.02)	0.244	88.1 (±23.0)	0.383
p.Ser102Pro heterozygote	GG	18	3.23 (±0.09)	0.097	62.6 (±9.60)	0.596
	GA	14	3.15 (±0.05)	0.284	73.8 (±8.93)	0.323
	AA	1	3.1	0.450	14.3	0.250
-148G>A & p.Ser102Pro	GG	8	3.23 (±0.05)	0.380	77.3 (±27.7)	0.792
	GA	4	3.2 (±0.02)	0.319	73.1 (±9.78)	0.454
p.Met39Leu & p.Ser102Pro	GG	5	3.36 (±0.16)	0.293	65.7 (±21.7)	0.774
	GA	1	3.2	0.800	122.4	0.350
	AA	1	3.4	0.150	410.4	**0.05**

Table 3.17: Effect of BCL11A rs4671393 together with KLF1 variants on HbA_2 (%) and HbF (mg/dL) in pseudo-thalassaemia. Statistically Significant p values are shown in bold

In the β-thalassaemia group a statistically significant increase in HbA_2 was noted in subjects homozygote for the p.Ser102Pro and homozygote to rs4671393. These subjects had an HbA_2 of 5.35% when compared to subjects' wildtype for both KLF1 variations and *BCL11A* polymorphism (table 3.18). An increase in HbF was noted in a subject compound heterozygote for the p.Met39Leu and p.Ser102Pro and homozygote for the rs4671393. His HbF was 495.5mg/dl.

KLF1	BCL11a Genotype	β-thalassemia heterozygotes				
		N	HbA_2 (%)	P Value	HbF (mg/dL)	P Value
KLF1 Wildtype	GG	126	4.34 (±0.06)	-	99.0 (±14.84)	-
	GA	41	4.35 (±0.11)	0.750	135.5 (±34.02)	0.285
	AA	7	4.21 (±0.45)	0.805	116.4 (±31.3)	0.103
p.Ser102Pro heterozygote	GG	43	4.18 (±0.08)	0.311	85.7 (±22.4)	0.463
	GA	11	4.73 (±0.85)	0.101	123.4 (±41.0)	0.587
	AA	2	5.35 (±0.15)	**0.048**	89.0 (±29.0)	0.498
-148G>A & p.Ser102Pro	GG	7	4.17 (±0.11)	0.864	52.1 (±12.1)	0.378
	GA	4	4.22 (±0.04)	0.715	56.8 (±16.07)	0.787
p.Met39Leu & p.Ser102Pro	GG	12	4.35 (±0.17)	0.467	218.8 (±116.0)	0.408
	GA	5	4.56 (±0.31)	0.403	495.5 (±0.31)	**0.039**

Table 3.18: Effect of BCL11A rs4671393 together with KLF1 variants on HbA_2 (%) and HbF (mg/dL) in β-thalassaemia subjects. Statistically Significant p values are shown in bold

3.7.2 EFFECT OF HBS1L-MYB HAPLOTYPES ON HBA_2 AND HBF

The-HBS1L-MYB haplotype blocks were constructed using HaploView 4.2 program. As shown in figure 3.23, there is moderate linkage disequilibrium (LD) between the 4 SNP's rs9399137, rs9694142, rs9376090 and rs9389269.

To assess the effects of combinations of these SNPs on the HbA_2 and HbF level, we conducted a haplotype analysis. Eight distinct haplotypes were identified (tables 3.19 and 3.20), seven haplotypes; the TTTT, CCCC, TCCC, TCTT, CCTC, CTCC and TCCT were present, both in the pseudo-thalassaemia and the β-thalassaemia heterozygotes, while haplotype TTCT was present only in the pseudo-thalassaemia group. The most common haplotype in both groups

was the TTTT haplotype (haplotype 1), present in 39.5% of the pseudo-thalassaemia group and 56.8% of the β-thalassaemia group. In the pseudo-thalassaemia group, the 2nd most haplotype was the TCCC haplotype (haplotype 3), present in 14.2%, followed by the TCTT haplotype (haplotype 4), present in 13.4%. The CCCC haplotype (haplotype 2) was present in 11.2 %, while the CCTC haplotype (haplotype 5) and the CTCC haplotype (haplotype 6) was present in 6.8% of the subjects. The TCCT haplotype (haplotype 6) and the TTCT haplotype (haplotype 8) was present in 4.4% and 3.7% of the subjects respectively.

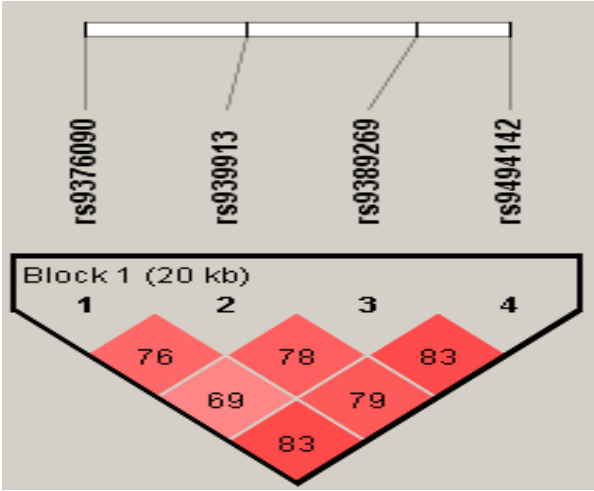

Figure 3.23: LD heatmap generated in HaploView to display the linkage disequilibrium (r2) between the 4 HBS1L-MYB SNPs.

In the β-thalassaemia heterozygote group, the 2nd most frequent haplotype was the CCCC haplotype, present in 28.5% of the subjects, followed by the TCCC haplotype, present in 5.9% of the subjects. The TCTT haplotype was present in 2.5% of the subjects, while the CCTC haplotype was present in 3.3% of the subjects. The least common haplotypes were the CTCC haplotype and the TCCT haplotype, present in 1.7% and 1.4% of the subjects.

Haplotype association analysis is shown in tables 3.4 and 3.5, while figure 3.8 shows the error graphs of the mean ± 1SE of the HbA_2 (%) and HbF (mg/dL) of the different haplotypes in the

pseudo-thalassaemia and β-thalassaemia heterozygote subjects. The non-parametric Mann-Whitney U test was performed to see if there is any statistical significance difference in the median of HbA_2 (%) and HbF (mg/dL) between the most common haplotype the TTTT haplotype (haplotype 1) and the other haplotypes in the pseudo-thalassaemia and β-thalassaemia haplotype.

	Pseudo-thalassaemia			β-thalassaemia heterozygote		
Haplotype	**Frequency**	**HbA_2 (%)**	**P value**	**Frequency**	**HbA_2 (%)**	**P value**
1.TTTT	39.5 %	3.21 (±0.01)	-	56.8%	4.29 (±0.05)	-
2.CCCC	11.2%	3.20 (±0.34)	0.681	28.5%	4.56 (±0.08)	**0.005**
3.TCCC	14.2%	3.21 (±0.34)	0.606	5.9%	4.21 (±0.06)	0.675
4.TCTT	13.4%	3.25 (±0.04)	0.600	2.5%	4.62 (±0.36)	0.295
5.CCTC	6.8%	3.24 (±0.50)	0.563	3.3%	4.13 (±0.07)	0.908
6.CTCC	6.8%	3.17 (±0.08)	0.451	1.7%	3.80 (±0.22)	0.118
7.TCCT	4.4%	3.20 (±0.03)	0.956	1.4%	3.77(±0.20)	**0.012**
8.TTCT	3.7%	3.14 (±0.02)	0.173	-	-	-

Table 3.19: Frequency of the 4 SNP haplotype and the mean ±1 SE of HbA_2 (%) in the pseudo thalassaemia group and the β-thalassaemia group. P value is the level of significance when comparing the mean of HbA_2 (%) between the wildtype haplotype (TTTT haplotype) and the other haplotypes.

	Pseudo-thalassaemia			β-thalassaemia heterozygote		
Haplotype	**Frequency**	**HbF (mg/dL)**	**P value**	**Frequency**	**HbF (mg/dL)**	**P value**
1.TTTT	39.5 %	95.5 (±15.4)	-	56.8%	81.3 (±10.96)	-
2.CCCC	11.2%	51.4 (±4.43)	0.08	28.5%	168.9 (±31.2)	**0.001**
3.TCCC	14.2%	75.3 (±9.33)	0.916	5.9%	151.9 (±77.5)	0.151
4.TCTT	13.4%	109.6 (±36.6)	0.571	2.5%	50.1 (±12.41)	0.60
5.CCTC	6.8%	61.6 (±16.5)	0.408	3.3%	93.1 (±12.94)	**0.031**
6.CTCC	6.8%	74.0 (±12.88)	0.942	1.7%	64.7 (±18.31)	0.670
7.TCCT	4.4%	85.6 (±19.1)	0.751	1.4%	59.7 (±18.4)	0.715
8.TTCT	3.7%	58.9 (±21.4)	0.416	-	-	-

Table 3.20: Frequency of the 4 SNP haplotype and the mean ±1 SE of HbF (mg/dL) in the pseudo thalassaemia group and the β-thalassaemia group. P value is the level of significance when comparing the mean of HbF (mg/dL) between the wildtype haplotype (TTTT haplotype) and the other haplotypes.

In the pseudo thalassaemia group, subjects with the TCTT haplotype (haplotype 4) had the highest HbA_2 and HbF (mg/dL) when compared to the other haplotypes. The other haplotypes had more or less the same HbA_2 level, with subjects in the TTCT haplotype (haplotype 8) having the least HbA_2(%). The level of HbF (mg/dL) varied between the different haplotypes, with the CCCC haplotype (haplotype 2) having the least HbF (mg/dL) (figure 3.24, tables 3.13

&3.14). No statistically significant difference was seen in the HbA_2 and HbF level between the TTTT haplotype and the other haplotypes in the pseudo-thalassaemia group.

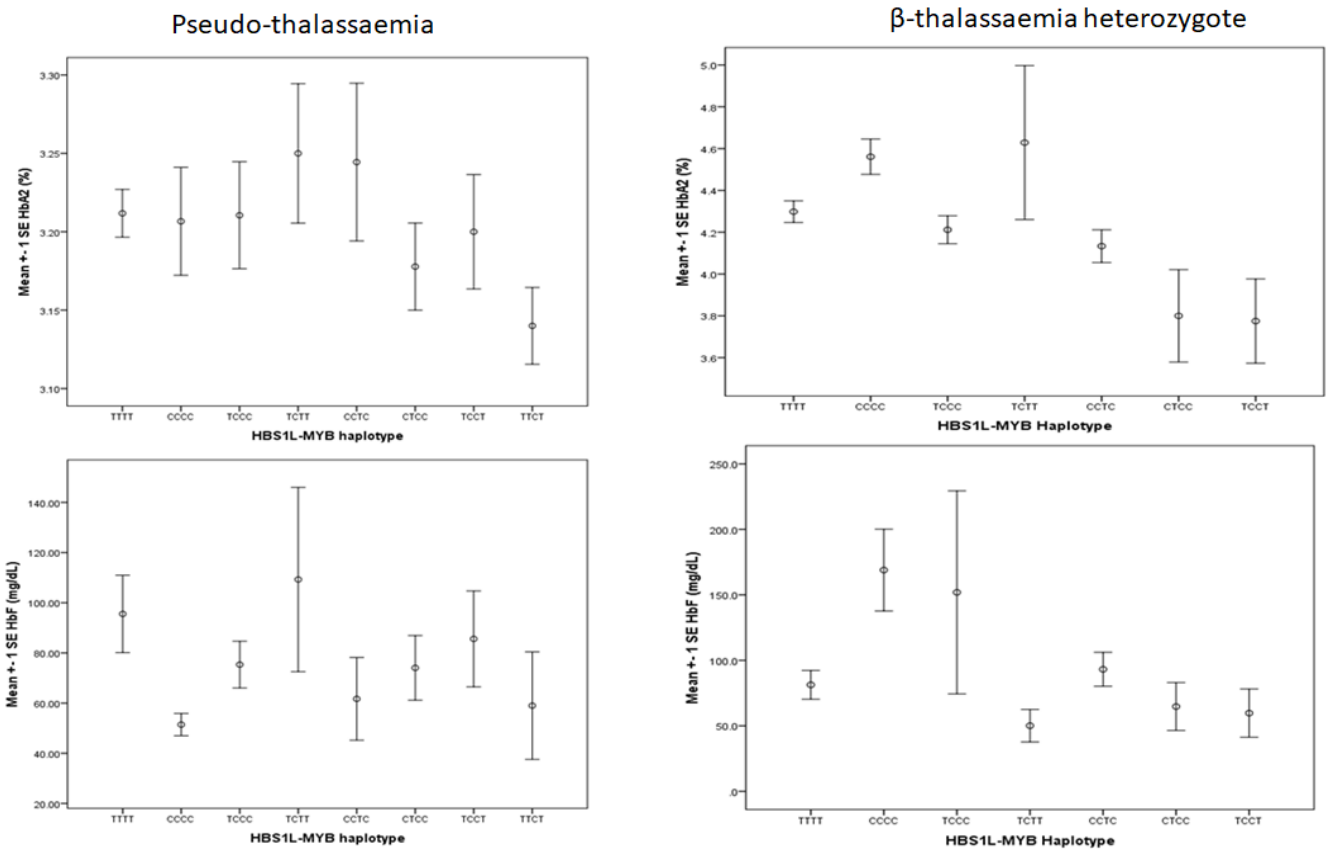

Figure 3.24: Error bar graphs showing the mean ±1 SE of HbA_2 (%) and HbF (mg/dL) of the different HBS1L-MYB haplotypes in the pseudo-thalassaemia group and the β-thalassaemia heterozygotes. The graph of the left-hand side shows the mean ±1SE of HbA_2(%) and the mean ±1SE of HbF (mg/dL) of the pseudo-thalassaemia group while the figure on the right shows the mean ±1SE of HbA_2(%) and the mean ±1SE of HbF (mg/dL) of the β- thalassaemia group.

In the β-thalassaemia heterozygote group, subjects with the TCCT haplotype (haplotype 4) had the highest HbA_2 level when compared to other haplotypes, followed by the CCCC haplotype (haplotype 2). Subjects with the TCCT haplotype (haplotype 7) had the least HbA_2 level. No statistically significant difference was noted in the HbA_2 level between the different haplotypes. Subjects with the CCCC haplotype (haplotype 2) had the highest HbF level, followed by the TCCC haplotype (haplotype 3). Subjects with the TCCT haplotype (haplotype 7) had the minimum HbF level (figure 3.22, tables 3.13 &3.14). A statistically significant

difference was seen in the HbF level, between the TTTT haplotype and the CCCC haplotype (p=0.001) and the TTTT haplotype and the CCTC haplotype (p=0.031).

3.7.1.2.1 Effect of *KLF1* together with *HBS1L-MYB* Haplotype on HbA_2 and HbF

When investigating the effect of KLF1 together with the TTTT and CCCC HbS1L-MYB haplotype no difference in HbA_2 was noted in the pseudo-thalassaemia group and β-thalassaemia heterozygote group (tables 3.21 & 3.22). An increase in HbF was noted in subjects compound heterozygote for the p.Met39Leu & p.Ser102Pro having the CCCC haplotype in both the pseudo-thalassaemia and β-thalassaemia group (tables 3.21 & 3.22).

KLF1	HBS1L-MYB Haplotype	Pseudo-thalassemia				
		N	**HbA_2 (%)**	**P Value**	**HbF (mg/dL)**	**P Value**
KLF1 Wildtype	TTTT	21	3.23 (±0.05)	-	84.8 (±14.6)	-
	CCCC	3	3.20 (±0.06)	0.680	67.6 (±8.88)	0.625
p.Ser102Pro heterozygote	TTTT	10	3.16 (±0.02)	0.635	68.3 (±9.41)	0.693
	CCCC	7	3.22 (±0.03)	0.348	72.4 (±19.4)	0.582
-148G>A & p.Ser102Pro	TTTT	9	3.23 (±0.04)	0.349	69.3 (±23.3)	0.792
p.Met39Leu & p.Ser102Pro	TTTT	3	3.46 (±0.26)	0.234	104.9 (±32.8)	0.452
	CCCC	2	3.4 (±0.05)	0.158	227.2 (±183.1)	0.640

Table 3.21: Effect of BCL11A rs4671393 together with KLF1 variants on HbA_2 (%) and HbF (mg/dL) in pseudo-thalassaemia

KLF1	HBS1L-MYB Haplotype	B-thalassaemia heterozygote				
		N	HbA_2 (%)	P Value	HbF (mg/dL)	P Value
KLF1 Wildtype	TTTT	105	4.28 (±0.69)	-	116.9 (±18.1)	-
	CCCC	63	4.41 (±0.16)	0.180	95.7 (±21.6)	0.346
p.Ser102Pro heterozygote	TTTT	25	4.48 (±0.15)	0.248	122.0 (±40.31)	0.312
	CCCC	13	4.31 (±0.14)	0.518	63.9 (±10.1)	0.787
-148G>A & p.Ser102Pro	TTTT	7	4.21 (±0.08)	0.664	55.4 (±14.3)	0.327
p.Met39Leu & p.Ser102Pro	TTTT	12	4.39 (±0.14)	0.236	284.4 (±164.5)	0.248
	CCCC	2	4.35 (±1.25)	0.899	112.8 (±58.3)	0.578

Table 3.22: Effect of BCL11A rs4671393 together with KLF1 variants on HbA_2 (%) and HbF (mg/dL) in β-thalassaemia subjects.

3.9 Proliferation analysis of K562 cells using different hemin concentrations

Absorbance readings of K562 cells treated with different hemin concentrations were obtained on two consecutive days, as explained in section 2.12. Data was normally distributed and therefore the Parametric Independent Sample t- test was used to compare means between untreated and different hemin concentrations.

3.9.1 Proliferative assay 24 hours after addition of hemin

The mean absorbance values for the untreated, 10µM, 20µM, 30µM, 40µM and 50µM hemin added to the K562 cells were reported as 0.52, 0.49, 0.48, 0.47. 0.46 and 0.48 respectively (figure 3.25). The mean absorbance values between K562 cells with no addition of hemin and those with different concentration of hemin were compared. One can notice a decrease in K562 cell proliferation on addition of different hemin concentration. The most notable difference in

mean absorbance was observed between the untreated K562 cells and K562 treated with 40µM hemin. However, it did not reach statistical significance (0.52 vs 0.46; p=0.06). One can also notice a slight increase in mean absorbance between the 40µM hemin and 50µM hemin (0.46 vs 0.48).

3.9.1 Proliferative assay 48 hours after addition of hemin

The mean absorbance values for the untreated, 10µM, 20µM, 30µM, 40µM and 50µM hemin added to K562 cells were reported as 0.86, 0.83, 0.81, 0.80, 0.79 and 0.82 respectively, as shown in figure 3.25. When comparing the mean absorbance of K562 with no hemin and those with different concentration of hemin, one can notice that there is a general slight increase in the mean absorbance value. The smallest mean absorbance value was that of K562 cells with 40µM hemin, as also observed after 24 hours. However, it was not statistically significant. A slight increase in absorbance value was also noted between K562 with 40µM hemin and those with 50µM hemin (0.79 vs 0.82).

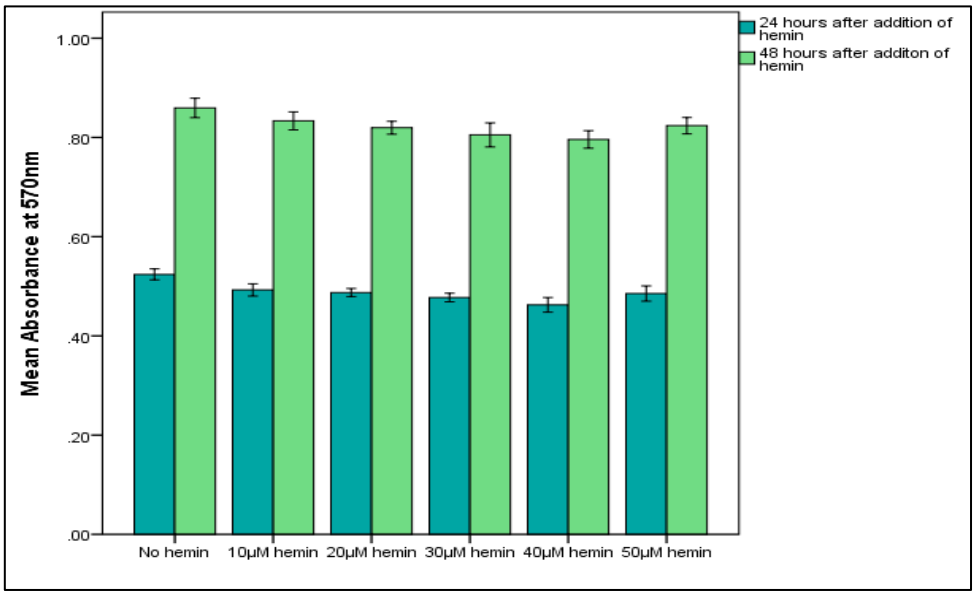

Figure 3.25: Proliferative ability of K562 24-hours (blue) and 48-hours (green) after addition of different concentrations of hemin. The bar chart represents the mean absorbance value of K562 cells either treated or untreated with hemin. Error bars represent the ± 1 SE.

3.10 Dual Luciferase Assay

3.10.1 Hemin response curve using luciferase assay

The luciferase reporter activity was measured in K562 cells transfected with the KLF1 wildtype promoter and EV in the presence of Renilla Luciferase, as explained in section 2.17.2. Transfected K562 cells were treated with 10μM, 20μM, 30μM, 40μM and 50μM hemin, or left untreated. Data was not normally distributed and therefore the non-parametric Mann-Whitney U-test and Kruskal Wallis Test were used to test whether there is a significant difference in the median of the different independent groups. When a statistically significant Kruskal-Wallis value was obtained, the Post-Hoc Analysis was performed, to identify where the difference truly lies.

3.10.1.1 K562 Cells transfected with empty vector:

As shown in figure 3.26 (red line) the mean fold change for K562 transfected with the empty vector was reported as 1.05, 0.78, 0.68, 0.91, 1.31 and 2.07 for K562 cells transfected with the empty vector and treated with no hemin, 10μM, 20μM, 30μM, 40μM and 50μM hemin. A statistically significant fold change was observed between untreated K562 and K562 treated with 10μM hemin (U= 3.00, p=0.048) and treated with 20μM hemin (U=4.00, p=0.01).

3.10.1.2 K562 Cells transfected with KLF1 wildtype promoter

As shown in figure 3.22 (black line), there was a slight increase in mean fold change between K562 cells transfected with the wildtype promoter with no addition of hemin and those treated with 10μM hemin, 20μM hemin and 30μM hemin. The mean fold change was reported as 31.2 for untreated K562, while 37.0, 47.2 and 47.3 for those treated with 10μM, 20μM and 30μM hemin respectively. A statistically significant increase in the fold change was noted between

untreated K562 and those treated with 40µM hemin and 50µM hemin. In fact, the mean fold change of K562 treated with 40µM hemin was 87.1 (U= 0.000, p=0.000), while that for K562 treated with 50µM hemin was 123.9 (U=0.000, p=0.000).

Upon combing cell proliferation assay results and hemin-dose response curve, it was concluded that the 50µM hemin concentration was activating the KLF1 without causing detrimental effects on cell proliferation. This concentration was further used for functional analysis.

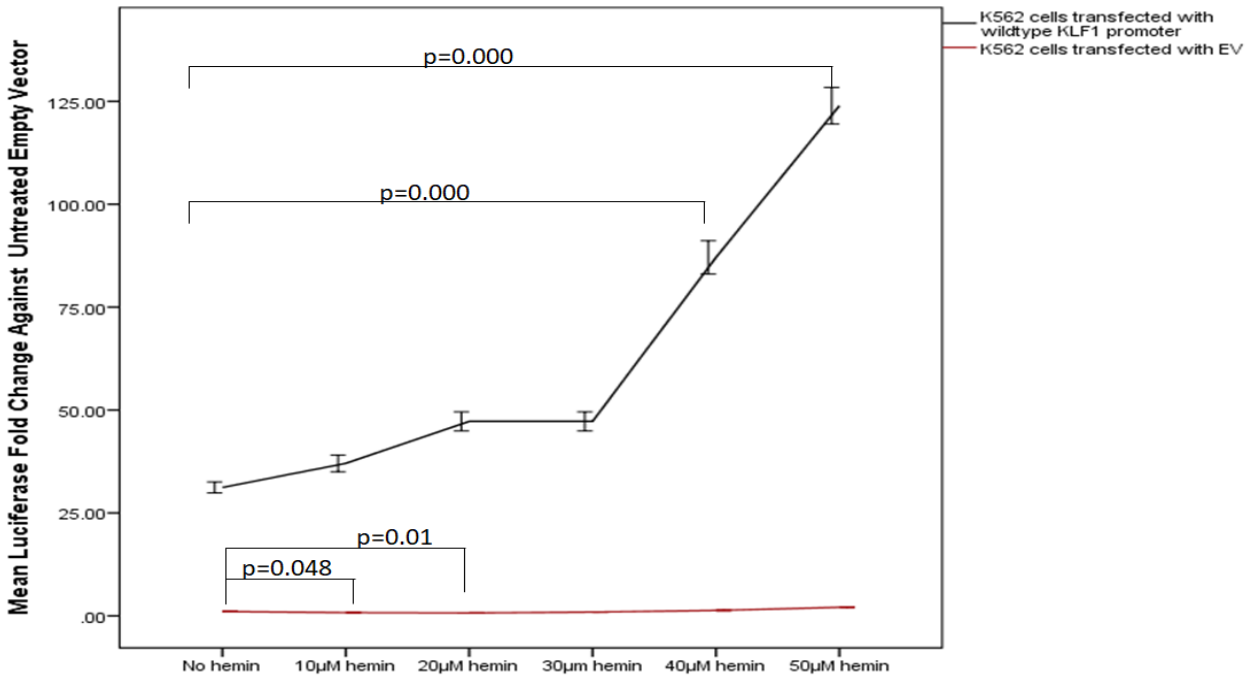

Figure 3.26: Hemin response curve using different hemin concentrations in K562 cells transfected with either the empty vector or the vector with the KLF1 wildtype promoter. Hemin concentration ranges from 0 to 50µM. The graph shows the mean fold changes of the EV and the KLF1 wildtype promoter- transfected cells exposed to different hemin concentrations. The error bars represent SE ±1 and the statistically significant difference (p<0.05) is also shown

3.10.2 Luciferase reporter activity assay in transfected HEK293T cells

There was a general decrease in the mean luciferase when comparing HEK293T cells transfected with the wild type KLF1 promoter and those transfected with the variant promoter (figure 3.27). The mean luciferase fold change of the wildtype KLF1 promoter was 109.1, that for promoter with the -1133C>A substitution was 74.1 while that for the promoter with the

-251 C>G substitution was 75.4. The mean fold change for the promoter containing both the -1133C>A substitution and -251 C>G substitution was 96.2. For the promoter containing the -148 G>A substitution was 84.1 while for the promoters containing the-154 G>A substitution, the -43 G>A substitution, and -102 T>G substitution the mean luciferase fold changes were 64.2, 96.6 and 87.3 respectively. Statistically significant difference was seen in the mean luciferase of the HEK293T transfected with the wildtype promoter and those transfected with the -1133 C>A substitution (p=0.01), -251 C>G substitution (p=0.03) and the -154 G>A substitution (p=0.00).

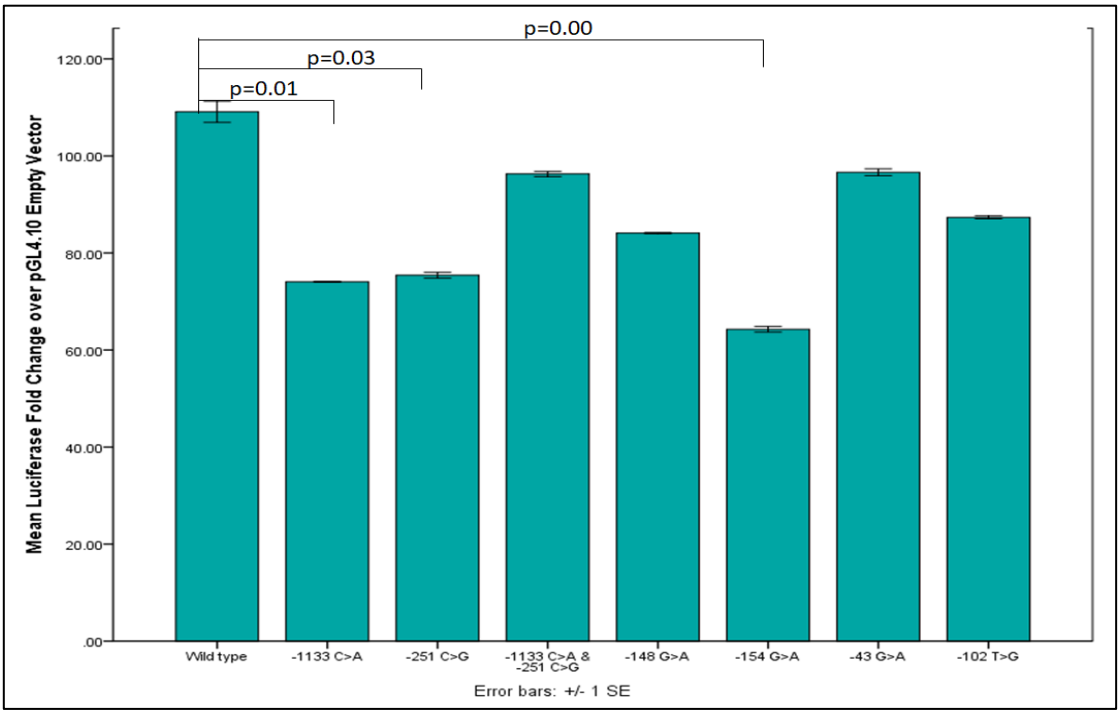

Figure 3.27: Dual-luciferase reporter assay of wildtype KLF1 promoter and variant constructs transfected in HEK293T cells. Firefly luciferase activity was normalized to the Renilla luciferase activity and the data is represented as the relative fold change increase compared to p.GL4.10 empty vector.

3.10.2.1 Luciferase reporter activity assay in transfected K562 cells

In K562 there was also a significant decrease in the mean luciferase, when comparing K562 cells transfected with the wildtype KLF1 promoter and those transfected with the variant promoter. The decrease in the mean fold change between the K562 cells transfected with the wildtype promoter and those with the variant promoters is more pronounced than that seen in HEK293T cells. The mean luciferase fold change of the wildtype KLF1 promoter was 242.2, while for promoter with the -1133C>A substitution was 125.9 and for the promoter with the -251 C>G substitution was 54.9. The mean luciferase fold change for the promoter containing both the -1133C>A substitution and the -251 C>G substitution was 102.9. The mean luciferase fold change for promoters containing the -148 G>A substitution, the -154 G>A substitution, the -43 G>A substitution and the -102 T>G substitution were 51.2, 96.7, 81.4 and 68.2 respectively. A statistically significant difference was also noted in K562 transfected with wildtype promoter, those transfected with the -251 C>G substitution, those transfected with the -148 G>A substitution (p=0.000) and those transfected with the -102 T>G substitution (p=0.006) (figure 3.28).

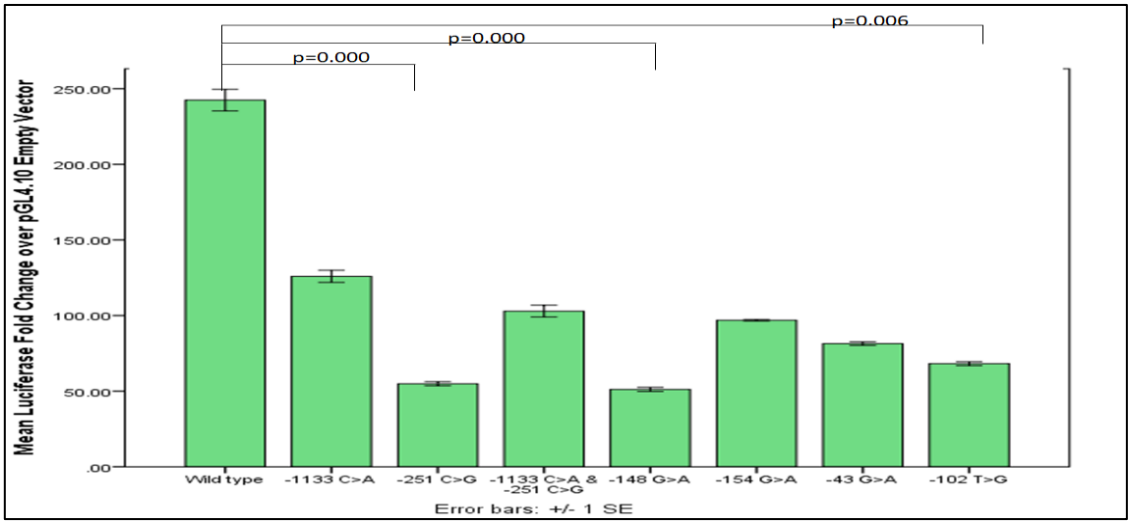

Figure 3.28: Dual-luciferase reporter assay of wildtype KLF1 promoter and variant constructs transfected in K562 cells. Firefly luciferase activity was normalized to the Renilla luciferase activity and the data is represented as the relative fold change increase compared to p.GL4.10 empty vector.

3.10.3 Luciferase reporter activity assays – K562 cells treated with 50µM hemin

Luciferase reporter activity assays were repeated in K562 cells transfected with either the EV, wildtype KLF1 promoter and KLF1 promoters containing one of the substitutions followed by the addition of 50µM hemin. An increase in the mean luciferase fold change was observed transfected K562 on addition of 50µM of hemin. The mean fold change of each treated and untreated expression vector was compared. A statistically significant difference was detected between the treated and untreated KLF1 wildtype promoter (p=0.004) and all the other treated and untreated KLF1 promoters containing one of the substitutions (p=0.004) (figure 3.29).

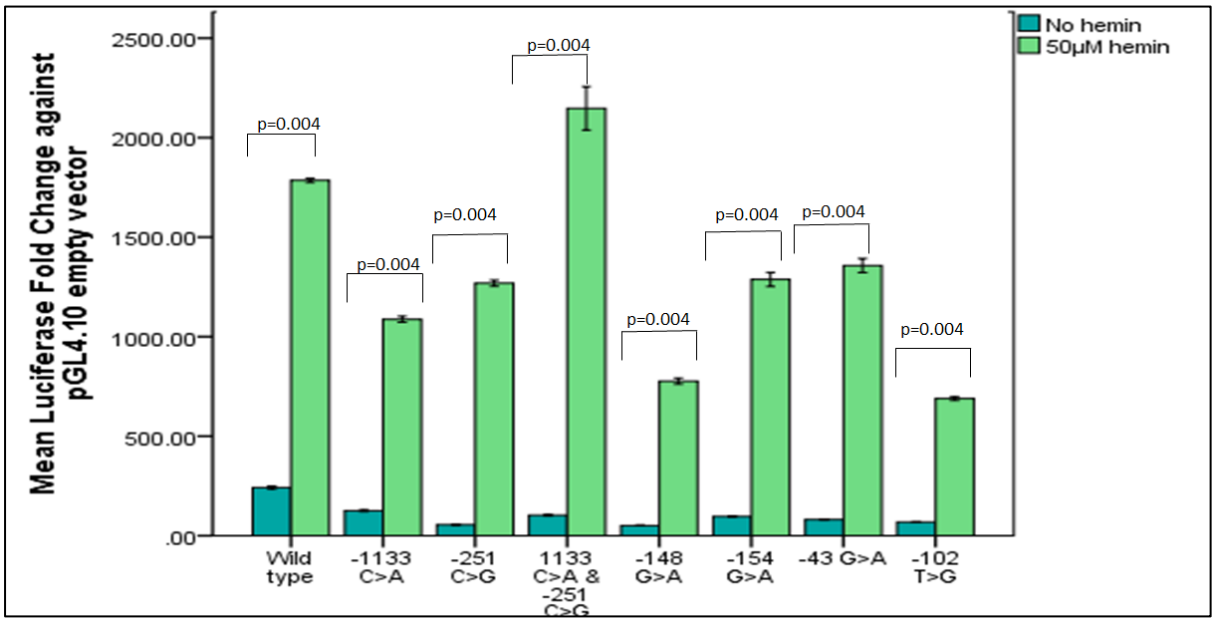

Figure 3.29: Dual luciferase reporter assays of K562 cells transfected with the wildtype promoter or variant constructs before addition of hemin and after addition of 50µM hemin. Firefly luciferase activity was normalized to the Renilla luciferase activity and the data is represented as the relative fold change increase compared to p.GL4.10 empty vector.

After addition of 50µM hemin, except for K562 transfected with the -1133C>A and -251C>C, all cells transfected with a variant promoter had a decrease in the mean luciferase fold change when compared to the wildtype promoter. The mean fold change of the wildtype promoter was 1785, while that for the -1133C>A substitution was 1087, for the promoter containing -251 C>G substitution it was 1269, for the promoter containing both the -1133C>A substitution and the -251C>G substitution it was 1269, while for promoters containing the -148G>A

substitution, the -154G>A substitution, the -43G>A substitution and the -102 T>G substitution they were 737, 1287, 1357 and 689 respectively (figure 3.30). A statistically significant difference was observed between the wildtype promoter and the promoter containing the -148G>A substitution (p=0.006), and between the wildtype promoter and the promoter containing the -102T>G substitution (p=0.000).

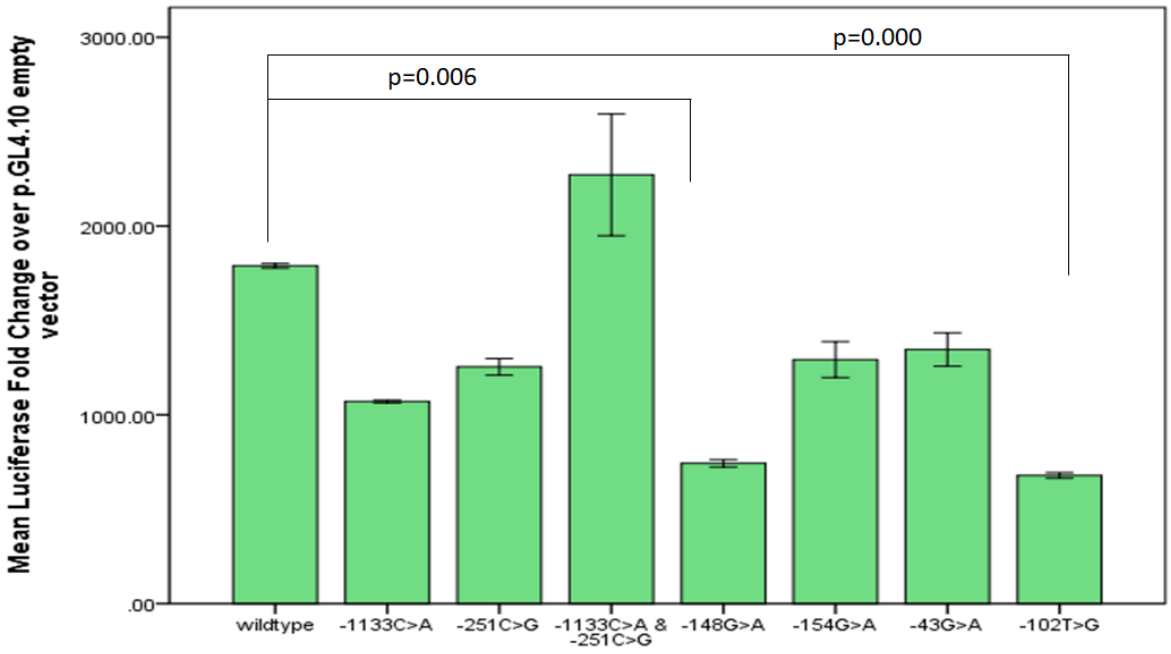

Figure 3.30: Dual luciferase reporter assays of K562 cells transfected with the wildtype promoter or variant constructs after addition of 50μM hemin. Firefly luciferase activity was normalized to the Renilla luciferase activity and the data is represented as the relative fold change increase compared to p.GL4.10 empty vector

3.11 Electrophoretic mobility shift assays

Electrophoretic mobility shift assay (EMSA) were performed using total nuclear protein extracts from human K562 and oligonucleotides bearing the wildtype and the variant KLF1 promoter sequence. The most evident difference was noted between the oligonucleotides wildtype for the KLF1 promoter and oligonucleotide bearing the -1133C>A (figure 3.31). It seems that something is binding to the wildtype sequence, but it is absent in the variant

sequence. A difference in the intensity of the bands in the wildtype oligonucleotides and oligonucleotides variant for the -154G>A was also noted (figure 3.32).

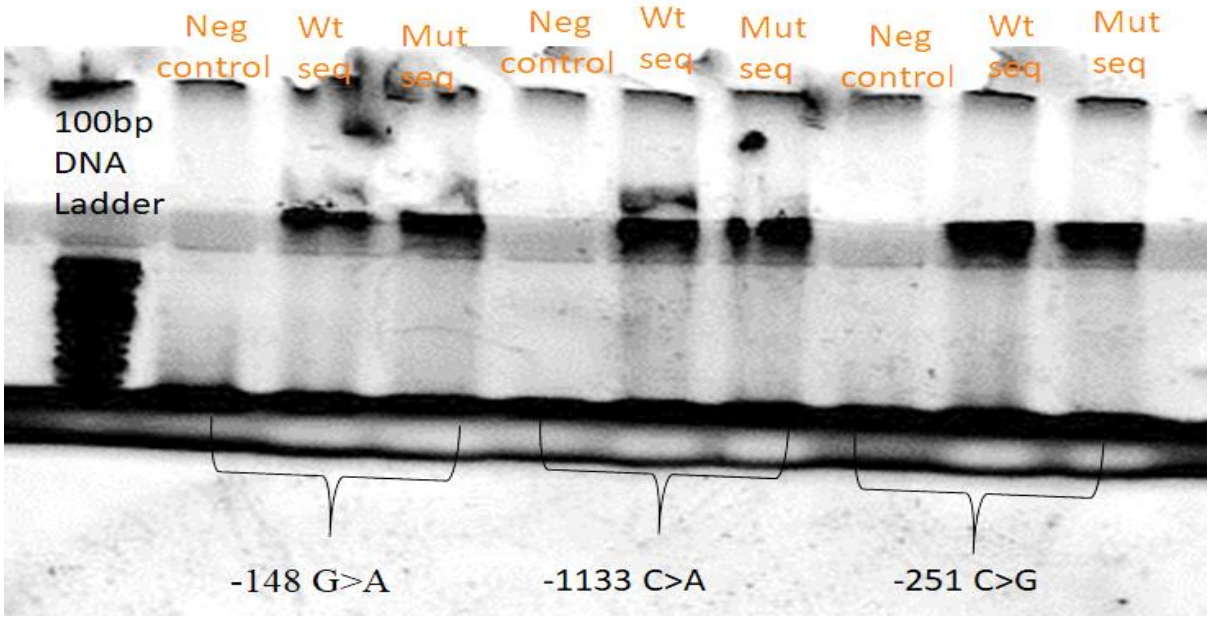

Figure 3.31: EMSA analysis using the wildtype oligonucleotides and variant probes for rs799334031, rs112943513 and rs3817621 and nuclear extracts isolated from K562 cells. A negative control without nuclear extracts was included with every variation .

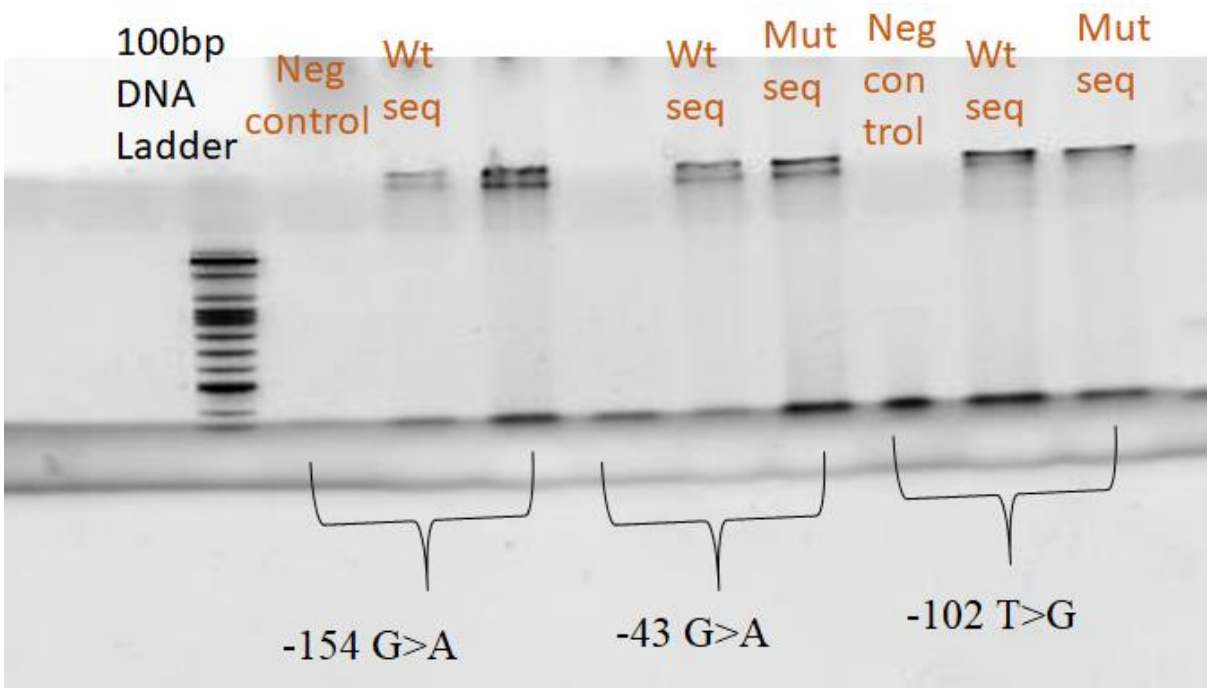

Figure 3.32: EMSA analysis using the wildtype oligonucleotides and variant probes for rs372651309, rs372359976 and rs548543206 and nuclear extracts isolated from K562 cells. A negative control without nuclear extracts was included with every variation.

3.11.1 Using RegulomeDB to assess the function of the promoter mutations

All promoter mutations were analyzed for potential regulatory functions using RegulomeDB database. RegulomeDB showed that five out of the six promoter mutations were likely to affect the binding of a number of transcription factors in K562 cell line. The -148G>A mutation was found to affect the binding of a number of transcription factors such as MYC, GATA2 and ZBTB7A (LRF) while the -1133C>A mutation was found to affect the binding of the YY1 transcription factor (figure 3.33).

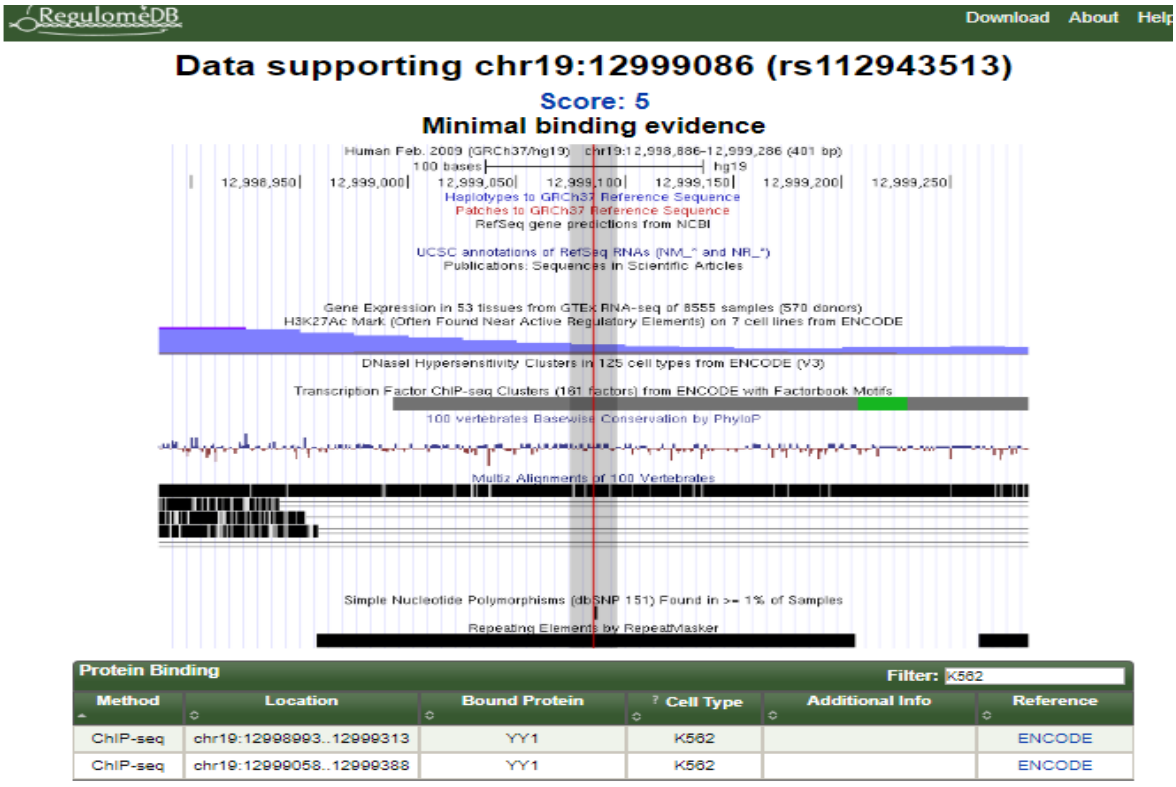

Figure 3.33: RegulomeDB output for -1133C>A mutation

The -251C>G mutation was found to affect the binding of POLR2, NFYB and MYC while the mutations -154G>A and -43G>A were found to affect the binding of LRF, POLR2, TAL1 and GATA2.

The National haemoglobin testing coupled with patient referral and follow-up of previous research conducted in the last 5 years has uncovered yet additional families with the KLF1 p.Lys288Ter mutation albeit with variable levels of HbF and at times even normal values. The data, as observed previously, suggested a dominant allele with variable penetrance due to *KLF1* deficiency.

Families such as the ones being described in this project, although rare, have been tremendously informative with respect to globin gene switching, as was already observed in the past (Huisman et al. 1972; Borg et al., 2010). Sequences *in cis* such as the XmnI (Weatherall et al., 1975) and $(AT)_XT_Y$ (Berg et al., 1989) and *in trans* such as *BCL11A*, *c-Myb* and *SOX6, LRF* have been discovered (Lettre et al., 2008, Yi et al., 2006, Masuda et al., 2016). Dual testing by iso-electric focusing and reverse phase liquid chromatography in population based haemoglobin testing has proven a powerful tool for gene discovery in particular when combined with extensive family studies and advanced genomics coupled with transcriptomics and functional assays as in these cases.

A candidate region on chromosome 19 was identified in Fam F1 members with high HbF by using genome wide-linkage SNP scan, followed by linkage analysis. Sequencing reveled a non-sense mutation in the *KLF1* gene that ablated the DNA binding domain. Functional assays suggested that *KLF1* in addition to its established role in erythropoiesis (Nuez et al., 1995; Perkins et al., 1995) and regulation of adult globin (Donze et al., 1995, Wijgerde et al., 1995), activated the expression of *BCL11A,* that in turn repressed the gamma globin genes (Borg et al., 2010). Through the Thalassemia and Molecular Genetics clinic, 5 other families with the same KLF1 truncation mutation, but with normal or slightly increased HbF levels, have now been identified (figure 3.3). The phenotypic variability was not readily explained by the co-

inheritance of known HbF- modulating variants in the *cis* and *trans* regulators (*HBB*, *HBS1L-MYB* and *BCL11A* loci).

WGS identified two new *KLF1* promoter mutations, one in *cis* (-1133C>A) and one in *trans* (-251C>G) to the truncation mutation. Fam F1 family members (II-2, II-5) with high HbF level were compound heterozygote for the promoter mutations, while subjects with medium HbF (II-15, II-16) had only the one *in cis*. Although usually we concentrate more on mutations present in the coding sequence, it is a well-known fact that between 1% to 2% of the disease-causing mutations are point mutations located in the promoter. They can disrupt the normal process of gene activation and transcriptional initiation resulting in the decreased amount of mRNA and therefore protein (Cooper et al., 2002). Several promoter mutations have been classified as disease causing mutations, including mutations in the *HBB* promoter giving rise to β-thalassaemia, mutations in the promoter of *F9* gene giving rise to Haemophilia B, while promoter mutations in the *GP1BB* and *PKLR* give rise to Bernard-Soulier syndrome and pyruvate kinase deficiency (de Vooght et al., 2009).

Promoter mutations that can putatively give rise to a disease/phenotype must be characterized to evaluate the significance of the DNA sequence variation in relation to the disease/phenotype (de Vooght et al., 2009). A combination of functional promoter assays, including luciferase in constructs with the -1133C>A and -251C>G mutation when compared to the construct with the wildtype promoter. After EMSA, for the -1133C>A mutation a super-shift in the wildtype oligo was noticed. This super-shift was absent in the mutant oligo. Using regulomeDB (http://regulomedb.org/snp/chr19/12999086) it was shown that the YY1 transcription factor bounds to the wildtype sequence, but it did not bind to the mutant sequence.

The YY1 is a ubiquitously expressed transcription factor belonging to the human GLI-Kruppel family of nuclear proteins (Shi et al., 1991). YY1 plays a role in the regulation of cell proliferation and differentiation. Deficiency of YY1 in mice resulted in peri-implantation lethality during embryonic development (Zhang et al., 1994). It can either act as a repressor or activator of transcription depending upon the promoter context (Hahn, 1992). YY1 represses transcription of several cellular genes including α-actin (Lee at al., 1992), β-casein (Meier & Groner, 1994), c-Fos (Gualberto, 1992) and ε-globin (Raich, 1995) while it activates c-Myc (Riggs et al., 1993), ribosomal protein (Hariharan et al., 1991) and γ-globin genes (Roversei et al., 2010). A decreased interaction between the YY1 and the HbG1 gene promoter region was noted in a Brazilian nondeletion HPFH (-195 C→G) mutation (Roversei et al., 2010).

The subjects with high HbF had a more open gamma globin chromatin than those with lower HbF levels (figure 3.10). Collectively the variability in HbF could probably be counted by allelic variation of the wildtype *KLF1* (Figure 4.1). In many situations the wildtype allele compensates for a gene deficiency such as in δ-β^+ thalassaemia. It seems that the KLF1 allele in trans to the truncation mutation compensates unless it is impaired by a promoter mutation. The redundancy may serve to protect KLF1 levels in early stages of erythropoiesis.

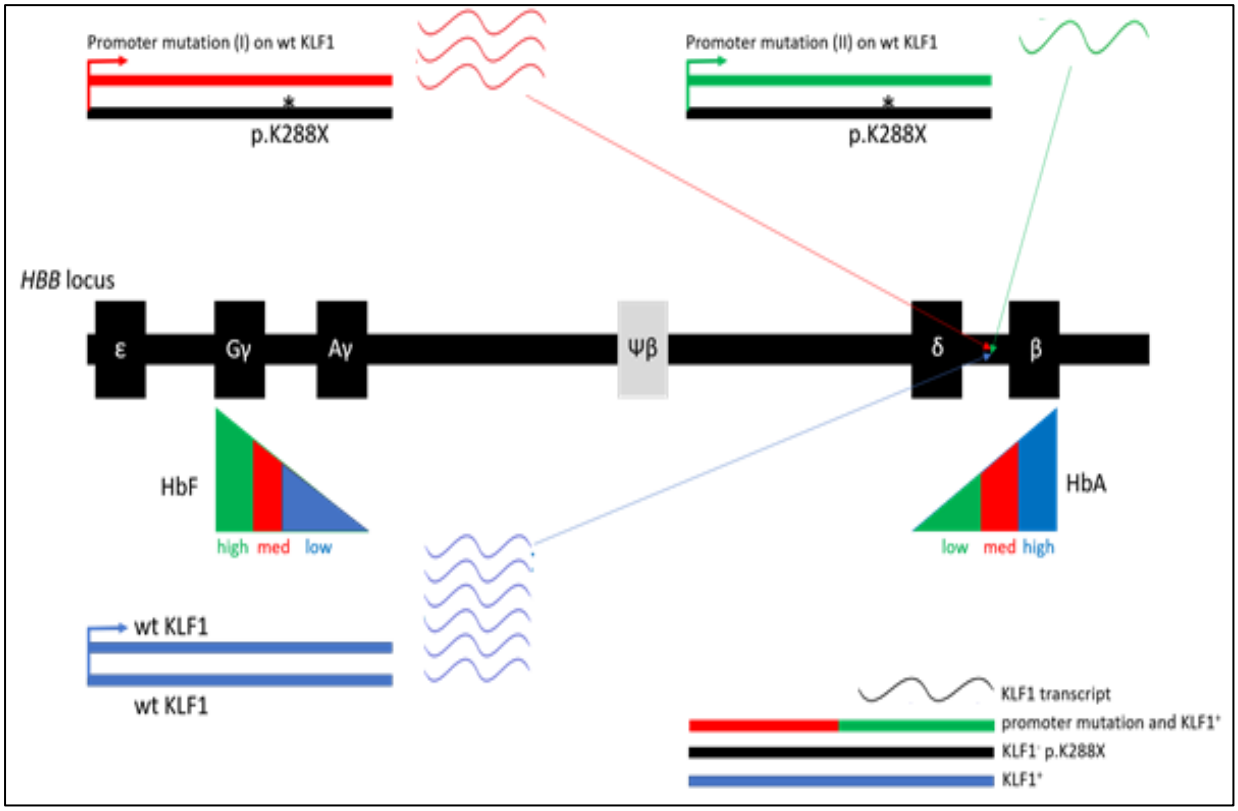

Figure 4.1: Hypothesized mechanism model of globin gene switching by differential KLF1 experession. Normal beta globin gene expression with an expected repression of gamma globin expression is encountered when both KLF1 alleles are wildtype and free of any functional promoter mutations. In the presence of one defective KLF1 allele (p.K288X) the remaining wildtype KLF1 will carry out its function and this will give rise to mild phentoype such as increase in HbF levels. When the presence of one defective KLF1 allele and a promoter mutation of the other KLF1 allele, the KLF1 expression will be much lower, giving rise to higher incease in HbG1/2 expression

Fam F2, Fam F4, Fam F5 and Fam F6 presented with a beta thalassemia picture having low MCV, a borderline HbA_2 and normal or elevated HbF, but without β-globin gene mutation. *KLF1* sequencing revealed the truncation mutation together with another 2 missense variants; the p.Met39Leu and the p.Ser102Pro. It was considered useful to define these subjects as having pseudo-thalassemia. Here, one needs to distinguish between silent-thalassemia and pseudo-thalassemia. In silent thalassemia, they do not show any hematological phenotype (normal CBC, normal HbA_2 and normal HbF), although they still have a mutation in the β-globin gene usually the C→T mutation at position -101 (Thein et al., 2013) and IVS 1,6C heterozygotes (Scerri, 1998). In pseudo-thalassemia, cases present with a picture of beta thalassemia picture (low MCV, low MCH, increase in HbA_2, normal or elevated HbF) without

mutations in the beta genes. Fam F3 was unique, in the sense that it was the only family where the proband and the father were IVS1,6C heterozygotes together with the *KLF1* truncation mutation. The results from these families showed the pleiotropic effects of *KLF1* mutations.

The retrospective search in the Malta biobank was useful to uncover mutations in the *KLF1* in undiagnosed subjects with borderline HbA_2 and β-thalassaemia heterozygotes. More than 55% of pseudo-thalassemia had a mutation in the *KLF1* gene, while KLF1 mutations were present in 40% of the β-thalassemia heterozygotes. Most of the *KLF1* variants were closer to the 5' end of the gene (figure 4.2) spanning from the promoter region to approximately 60% of exon 2. Three SNPs were found at the 3' UTR. The presence of mutations closer to the 5' end of the KLF1 gene were also noted by Persue et al., (2011) and Lou et al., (2014).

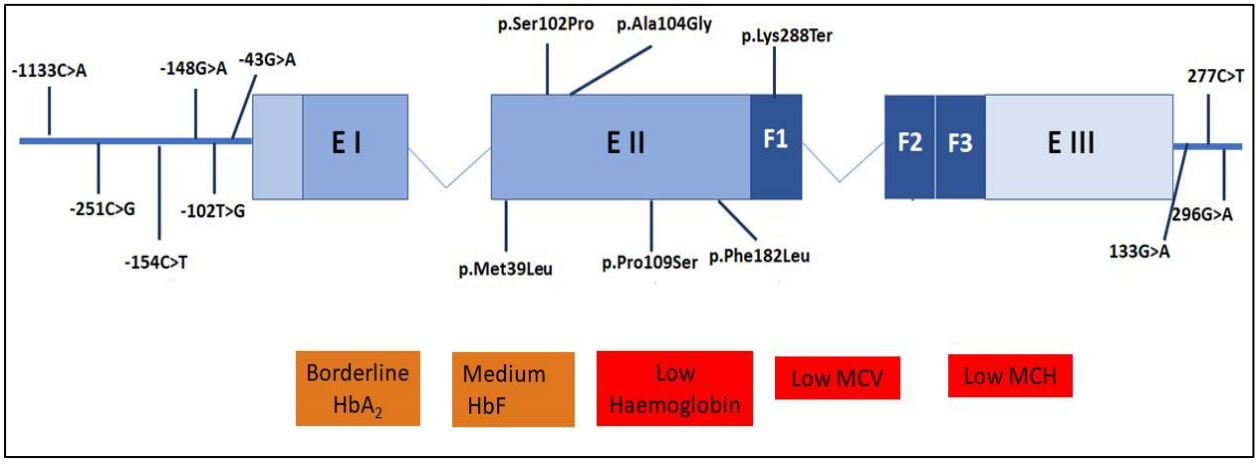

Figure 4.2: An extended map of the KLF1 gene showing all the SNV found in our study together with the phenotypes associated with these mutations. EI, EII and EIII refer to exon 1, exon 2 and exon 3. The dark blue boxes (F1, F2 and F3) represent the DNA binding domains.

Contrary to earlier views, KLF1 mutation could be classified as autosomal recessive. Heterozygotes mutations had no effect on the phenotype, but compound heterozygotes or homozygotes gave rise to more prominent phenotypes.

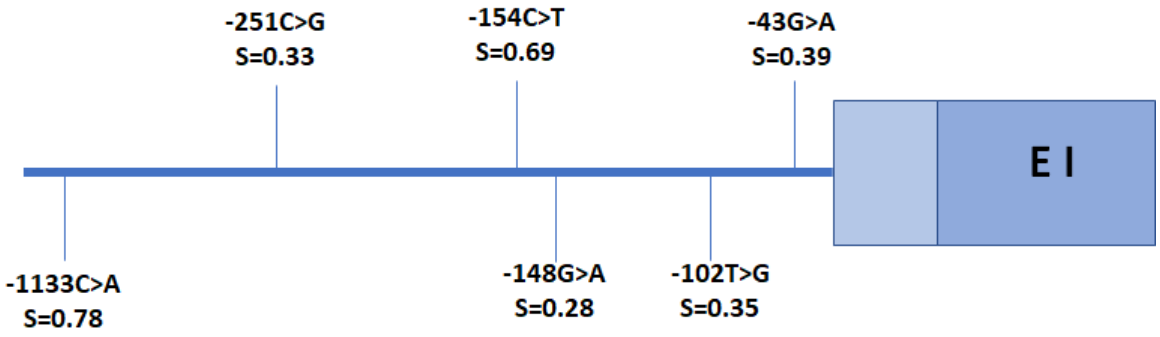

Figure 4.3: KLF1 promoter mutations. A predicative score (s) for luciferase expression is given for each mutation. The score was calculated by the luciferase expression of K562(mutant/wildtype) x HEK293 (wildtype/mutant). The smaller the score the more significant the difference between luciferase expression and to the wildtype promoter.

Six mutations were found in the promoter region (figure 4.2). The most frequent was the -251C>G. Heterozygotes or homozygotes had the same HbF as the wildtype. Since this was a retrospective study it was not possible to measure the expression of *KLF1* in subjects with this mutation. β°thalassaemia and HbE patients were found to have higher HbF levels when they were homozygote for the -251C>G mutation (Khamphikham et al., 2017). It was shown that this promoter mutation was associated with differences in reticulocyte count (Astle et al., 2016) and it was also present in subjects with myeloproliferative neoplasm (MPN), myelodysplastic syndrome (MDS) and Acute myeloid leukaemia (AML) (Gnanapragasam et al., 2018).

The second most common promoter mutation was the -148G>A. This was the first *KLF1* promoter mutation to be reported in a Serbian adult female who had normal haematological parameters and elevated HbF levels (11%). Together with the -148G>A mutation, she was found to be heterozygote for the p.Ser102Pro. Reporter gene experiments showed a drastically decreased expression in the -148G>A construct when compared to the construct with the

wildtype sequence (Radmilovic et al., 2013). In this study decreased luciferase expression (predictive score 0.28) was also noted in the cells transfected with -148G>A construct relative to the wildtype promoter.

This mutation was absent in the Serbia general population and was also completely absent in normal subjects of Greek origin (n=100). Radmilovic et al., (2013) pointed out that it could be responsible for higher HbF and the subject with -148G>A mutation had decreased KLF1. In the Maltese, the -143G>A mutation was also present in the control group without showing any effect on the HbF level. An increase in HbF was only noted in triple heterozygote for this promoter mutation together with the p.Ser102Pro and the -43G>A mutation.

After performing EMSA from nuclear protein extracts from K562 with 32P labelled oligonucleotides bearing wildtype and the -148C>A promoter Radmilovic et al., (2013), showed that the -148G>A mutation resided in a Sp1 binding site and the mutation led to the ablation of Sp1 binding. This was not reflected in our study, but these experiments need to be further optimized. Using Transfac (http://genexplain.com/transfac/) it was shown that two other transcription factors, Churchill TF also known as TRIM28 and RNF96, did not bind in the presence of the mutation.

The rare -154 C>T mutation (predictive score 0.69) in the promoter region was first reported in a patient with severe, transfusion-dependent neonatal anaemia with red cell abnormalities. The subjects who had microcytic anaemia had the -154C>T mutation, together with the p.Ala298Pro in the 1st zinc finger. The HbF and HbA_2 of the patient were not available. His mother, who also had HbE, was found to be heterozygote for the -154C>T mutation and had an HbF of 4.3% (Viprakasit et al., 2014). In this study, the -154 C>T was found in a 6-year-old Maltese boy with HbA_2 (3.3%), HbF (140mg/dl), and Hb (12.2g/dL), but low MCV

(71.2fL) and MCH (25.7pg). This compares well with the haematological indices of the subject described by Viprakasit et al., (2014), except for the very high HbF and critically low Hb levels. Due to its low frequency, data about this mutation could not be supported statistically, but only interpreted with respect to the data available on the individual involved. Dual luciferase assays showed a decrease in the luciferase expression in both HEK293T cells and K562cells transfected with the -154C>T consturct when compared to those transfected with the widltype promter. Using *in-silco* programmes, this muation was predicted to modify the binding of tumour suppressor gene p53, paired box 5, as well as other transcription factors thought to occupy that promoter location.

The -43G>A rare promoter muation was found in a 36 year old male with HbA_2 of 3.1% and a high HbF of 604.8mg/dl. The subject was not anaemic and was also heterozgyote for the -148G>A and the p.Ser102Pro. This mutation is 2bp upstream of the transcription initiation site and is therefore likely to affect *KLF1* expression and increase in HbF. A large decrease in luciferase expression in K562 (predicitve score 0.39) was noted in constructs containing the -43G>A mutation, when compared to wildtype promoter.

The -102T>G promoter mutation was found in a 93-year-old β-thalassaemia heterozygote female with an HbA_2 of 4.2%, HbF of 20mg/dl and microcytic anaemia (Hb 7.2g/dL, MCV 67.6fL; MCH 21pg). It is interesting to point out that this mutation is just 5bp downstream from the CCAAT motif, which could potentially effect transcription of the gene. Functional studies carried out have attributed this mutation to a lower *KLF1* promoter activity (predictive score 0.35 figure 4.3).

The missense variant p.Ser102Pro found in exon 2, was the most common SNV present both in the heterozygotes and homozygotes. This mutation, which appeared to be non-pathogenic on its own, was found in a patient with CDA, but was also present in his healthy mother

(Arnaud et al., 2010). Satta et al., (2011) reported the presence of this mutation in subjects with high zinc protoporphyrin levels and borderline HbF levels. The p.Ser102Pro was found to be the most common *KLF1* mutation in the Sardinian population (Persue et al, 2011; Paglietti et al., 2016), while from this study it can be concluded that it is also the most common *KLF1* mutation in the Maltese population. The p.Met39Leu was also frequently present in all three groups. Subjects carrying this mutation were always compound heterozygote for the p.Ser102Pro. In the 6 Fam Fs the p.Met39Leu was also always in LD with the p.Ser102Pro. This mutation is probably a neutral substitution as mouse *KLF1* contains a leucine at this position (Miller et al., 1993).

Other mutations in exon 2 included the p.Ala104Gly, p.Pro109Ser and the p.Phe182Leu. The p.Ala104Gly and the p.Pro109Ser were first associated with CDA and In(Lu) phenotype (Arnuad et al., 2010), but later studies showed that the wildtype allele was compensating for these missense mutations, because they were also found in healthy subjects (Helias et al., 2013; Khamphikham et al., 2017; Gnanapragasam et al., 2018;). This was also confirmed in our study. The p.Phe182Leu, which is the closest mutation to the zinc finger, was found only in three subjects. One from the control group, who presented with normal haematological indices and two subjects, one in the pseudo-thalassaemia and one in β-thalassaemia heterozygotes group, who presented with increased in HbA_2. Phenotypically two of these subjects with the p.Phe182Leu were similar to the subjects with the same mutations reported by Singleton et al., (2008) and Liu et al., (2014).

The 3' UTR plays an important role in the stabilisation, localisation and translation of messengers RNA (mRNA). Mutations in the 3'UTR can result in the disruption of the functional elements of the 3'UTR, such as polyadenylation signal, micribonucleic acid (miRNA) target sites and AU-rich elements can give rise to non-functional proteins or reduced

amount of functional proteins. Mutations in 3'UTR were associated with diseases such as Haemoglobin H disease, Parkinson disease, breast cancer, atopic dermatitis, papillary thyroid carcinoma, spinocerebellar ataxia and much more (Chatterjee & Pal, 2009).

Three mutations in the 3'UTR were identified in this study (figure 4.2). When heterozygous, subjects showed no phenotype related to the mutation, but when the subject was compound heterozygous for 277C>T and 296G>A an increase in HbF was noted. The 277C>T and the 296G>A mutations were extremely rare and in literature were detected only in the African, Caribbean and African American population, while they were absent in the European population (Gnanapragasam et al., 2018). The presence of these two mutations in our population indicates some degree of gene migration, most probably from the Middle East via South Italy. When studying the contribution of *KLF1* mutations in haematological malignancies, the 277C>T and 296G>A mutations were present in one subject with MPN and one subject with MDS, but no association was found (Gnanapragasam et al., 2018).

The implication of *KLF1* mutations on HbA_2 was first reported in a cohort of 145 Italian subjects with HbA_2 between 3.3 and 4.1 %. Fifty-two subjects had a mutation in the *KLF1* gene (Perseu et al., 2011). In China, 165 subjects with borderline HbA_2 were recruited, 15 had decreased red cell indices, while the other 150 subjects had normal red cell indices. Out of the 165 subjects, 20 had a mutation in the *KLF*1 gene, while *KLF1* mutations were absent from 80 subjects with normal HbA_2 (Lou et al., 2014).

HbA_2 is a tetramer of α and δ globin chains, forming about 2-3% of the total haemoglobin. It has no known physiological role, but it helps in providing diagnostic clues for the β-thalassaemia heterozygotes. Although there is a strong sequence homology in the coding regions of the HBB and HBD, the expression levels of these genes are remarkably different

(Rieder et al., 1965). In-fact HBD transcription is half as efficient as that of HBB (Humphries et al., 1982). The difference can be attributed to the differences between their promoters. The canonical CCAAT motif of the HBD is replaced by CCAAC motif, which binds CP1 less avidly and the CACC sequence in the HBB, that is a binding site for KLF1 is missing in HBD (Adams et al., 1982).

In our study, more than 55% of previously undiagnosed pseudo-thalassaemia had a mutation in the *KLF1* gene. A statistically significant difference in the HbA_2 was noted in those who were either compound heterozygote or homozygote for a KLF1 mutation. These results show the importance of KLF1 sequencing in subjects who have the picture of a beta-thalassaemia in the absence of beta globin mutations. Families can now have a diagnosis that they could not have before.

Until recently, *KLF1* variants were considered only to be extremely rare, but high throughput DNA sequencing prompted the identification of numerous *KLF1* mutations. The presence of *KLF1* variants were found to be remarkably high in a thalassaemia endemic region in Southern China, whereas in Northern China, a non-endemic region was found to be low (Liu et al., 2014). These findings suggested that *KLF1* mutations occurred selectively in the presence of β-thalassemia to increase the production of HbF, which in turn ameliorated the severity of the disease.

Moreover, in our study 40% of the β-thalassaemia heterozygotes had a mutation in the KLF1 gene. The most common mutation was the p.Ser102Pro. Liu et al, (2014) found that co-existence of β-thalassaemia with *KLF1* mutations resulted in higher levels of HbA_2. Persue et al, (2011) reported that the interaction of *KLF1* mutations with β-thalassaemia only resulted in very high HbA_2 levels, without any other clinical or haematological effect. In our study a

statistically significant increase in HbA_2 was noted in a β-thalassaemia heterozygote subjects who were also compound heterozygotes or homozygotes for KLF1 mutations. It was shown that expression of HbF remains very high in β-thalassemia homozgyotes patients who are also compound heterozygotes for *KLF1* mutations that affected a highly conserved codon in the zinc finger DNA-binding domain (Satta et al., 2011; Huang et al., 2015). A statistically significant difference in HbF levels was not noted between the β-thalassaemia heterozygotes for *KLF1* mutations when compared to those with no *KLF1* mutations. This study also showed that KLF1 interaction with β-thalassaemia only resulted in more elevated levels of HbA_2 without any other clinical or haematological effect. This information is important for genetic counselling in couples carrying HBB mutations together with KLF1 mutations.

Genetic regulation of foetal haemoglobin was first associated with mutations in the *XMN1* (Gilman & Huisman, 1985), while genome wide association studies (GWAS) have identified SNPs in trans-acting elements such as in *BCL11A* and *HbS1L-MYB* that account for more than 20% of the variation in HbF levels (Lettre et al., 2008; figure 4.4). The effect of other *cis*-and *trans* regulators together with the *KLF1* variants on HbA_2 and HbF was investigated. The *cis*-regulators were the XMN1 and the (AT)xTy while the *trans* regulators were the *BCL11A* and *HbS1L-MYB.*

The XmnI (C→T) polymorphism at position -158 of the HBG2 was found to promote the expression of foetal haemoglobin under stress erythropoiesis (Gilman & Huisman, 1985). The XMN1 polymorphism had little effect on the γ -globin gene expression among neonate Hb F-Malta-I heterozygotes (Felice et al., 2007), however, it correlated with γ globin levels in anaemic adult β thalassemia homozygotes (Felice et al., 2007). Pereira et al., (2015) showed that β-thalassaemia heterozygotes from Portugal, who were also heterozygote or homozygote for the XMN1 polymorphism, had an increase in foetal haemoglobin. In our study, no

statistically significant difference in HbF was noted in subjects' heterozygotes or homozygotes for the XMN1 polymorphism in both the pseudo-thalassaemia and β-thalassaemia group. Our findings were consistent with other studies that showed that there was no correlation in the HbF levels between subjects' wildtype and heterozygote or homozygote for the XMN1 polymorphism (Persue et al., 2011; Cyrus et al., (2017). No difference in HbF levels was noted when studying the effect of both XMN1 polymorphism combined with KLF1 variants.

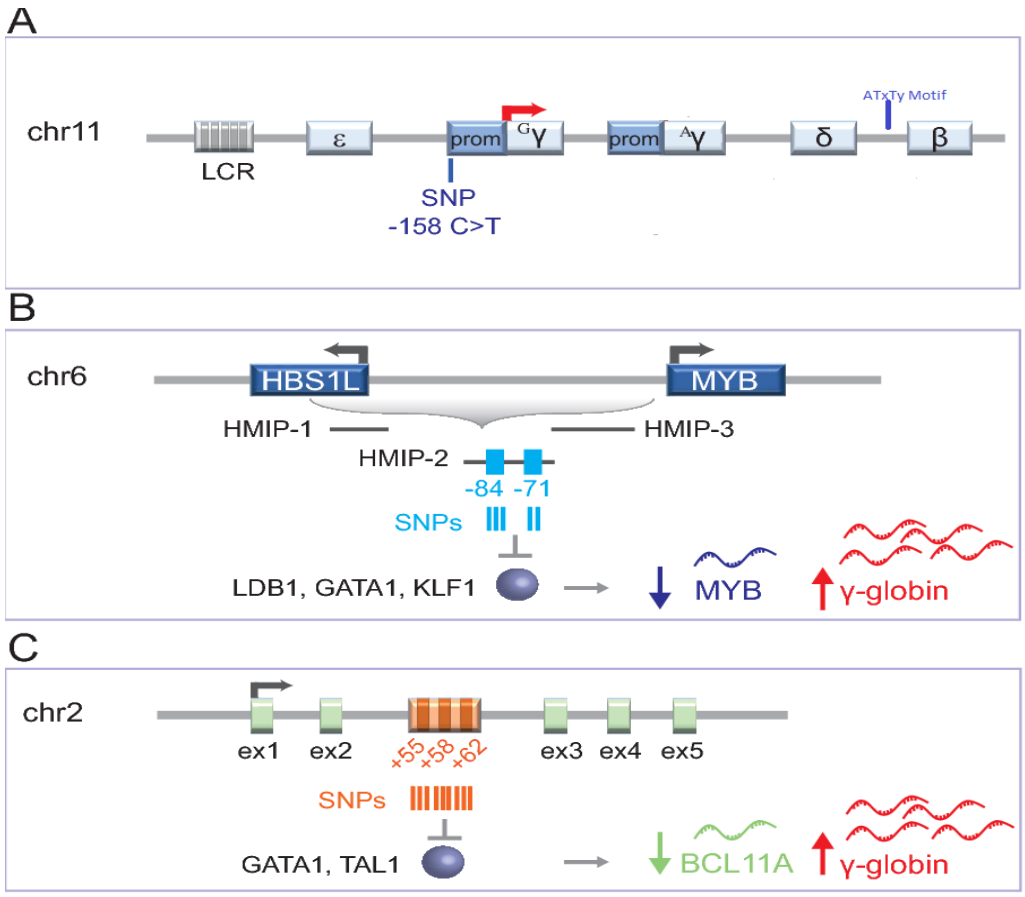

Figure 4.4: Genetic variants in the b-globin locus, HBS1L-MYB, and BCL11A loci influence foetal haemoglobin (HbF) levels. (A): Schematic representation of the b-globin locus on chromosome 11. The XMN1 polymorphism position -158 of the HBG2 and the (AT)xTy motif upstream of the β-globin gene were associated with moderately increase in foetal haemoglobin

(B): The region between HBS1L and MYB genes on chromosome 6 contains three HMIP blocks 1, 2, and 3. SNPs in these regions were associated with high HbF. These variants reduce LDB1, GATA1, and KLF1 occupancy, thus decreasing MYB expression and, therefore, increasing HbF expression. (C): Representation of the BCL11A gene on chromosome2. Several SNPs, associated with high HbF levels map to three erythroid-specific intronic enhancers located 55, 58, and 62 kb downstream of the BCL11A TSS. A SNP within the 162 kb enhancer impairs GATA1 and TAL1 binding, thus leading to a reduction of BCL11A expression and increase of HbF levels. Targeted disruption of a GATA1 binding site in the 158 kb enhancer is also associated with decreased BCL11A levels and high HbF expression.

The (AT)xTy motif, situated -540bp upstream of the β-globin gene, acts as a DNA binding site for the BP1 protein and repressed the δ and β-globin genes (Berg et al., 1989; Fu et al., 2001; Chase et al., 2002). Typically, this sequence consisted of 7 consecutive pairs of (AT) dinucleotides followed by 7 (T) nucleotides (Chebloune et al.1988). The most common sequence is the $(AT)_7T_7$ (n=21) and is considered as the reference sequence. In vitro studies on cell lines showed longer (AT)xTy motifs such as $(AT)_9T_5$ and $(AT)_{11}T_3$ bind strongly to the negative regulator BP1 protein, while shorter motifs such as $(AT)_6T_9$ bind less BP1. HbS homozgyotes who have the $(AT)_9T_9$ motif produce more HbS, whereas those with the $(AT)_9T_5$ and the $(AT)_{11}T_3$ motif express significantly less HbS versus HbF and therefore they have a better clinical picture (Elion et al., 1992). The most common motif in both the pseudo-thalassemia and β-thalassaemia heterozygotes was the $(AT)_7T_7$. Other common motifs included the $(AT)_8T_5$/$(AT)_9T_5$ in the pseudo-thalassaemia group and the $(AT)_7T_7$/$(AT)_9T_5$ in the β-thalassaemia group. No effect on HbF was found when combing the (AT)xTy motif together with the KLF1 mutations.

The effect of the rs4671393 residing in the intron 2 of the *BCL11A* gene was shown to have the strongest effect on HbF levels among subjects with sickle cell disease (Lettre et al., 2008). It was also reported in Fam F1 by Borg et al., (2010). Chaouch et al., (2016) found that subjects with SCA heterozygote or homozygote for the rs4671393 mutation presented with a HbF of >15%. Here no correlation between rs4671393 and HbF level was found. Our findings were in accordance with those of Nguyen et al., (2010) and Eljali, et al, (2013). Nguyen et al., (2010) showed that there was no statistically significant correlation between rs4671393 and HbF level in 57 β-thalassaemia intermedia patients. When studying Maltese β-thalassaemia heterozygotes Eljali et al., (2013), showed that there was no significant association between the *BCL11A*

mutation and HbF in both β-thalassaemia subjects independent of Hb levels. Recent studies carried out by Cyrus et al., (2017) and Lai et al, (2017) also showed no relation between the rs4671393 and HbF levels in β-thalassaemia intermedia and transfusion dependent β-thalassaemia.

When investigating the effect of BCL11A rs4671393 together with *KLF1* variants, an increase in HbF in subjects compound heterozygotes for the p.Met39Leu and the p.Ser102Pro and homozygote for the rs4671393 was noted in both the pseudo-thalassaemia and β-thalassaemia heterozygotes. The KLF1 variants together with the rs4671393 showed a cumulative effect on HbF. Lai et al., (2017) showed that on its own the rs4671393 has no effect on HbF, but when tested together with two XMN1 and two HbS1L-MYB mutations they showed a cumulative effect on raising HbF.

Linkage analysis in a large family with HPFH of Gujarati/North Indian descent had shown that the genetic determinant for high HbF segregated independently from the β-globin cluster (Thein & Weatherall, 1989). Via a painstaking mapping exercise, the genomic location for high HbF in this family was mapped to chromosome 6q23 (Garner et al.1998). Thein (2007) found that it was due to variants in the region between the HBS1L and MYB genes. The *HBS1L* gene is expressed in haematopoietic cells and is involved in regulation of several cellular processes, while the *MYB* gene is involved in oncogenesis and plays an important role in erythropoiesis (Thein et al., 2007). Twelve variants that were originally reported to be strongly associated with HbF persistence by Thein et al., (2007) were also found to modulate HbF levels in healthy subjects of African and East Asian descent in SCA and β-thalassaemia heterozygotes and homozygotes (Gibney et al., 2008; Makani et al., 2010; Solovieff et al., 2010; Farrell et al., 2011; Bae et al., 2012). Menzel et al., (2014) showed that there was a close linkage disequilibrium between these 12 SNPs and they gave rise to two major haplotypes (figure 4.5)

which were present in 92% of their European cohort. Seventy percent (70%) had the haplotype associated with low HbF, while 22% had the haplotype associated with high HbF.

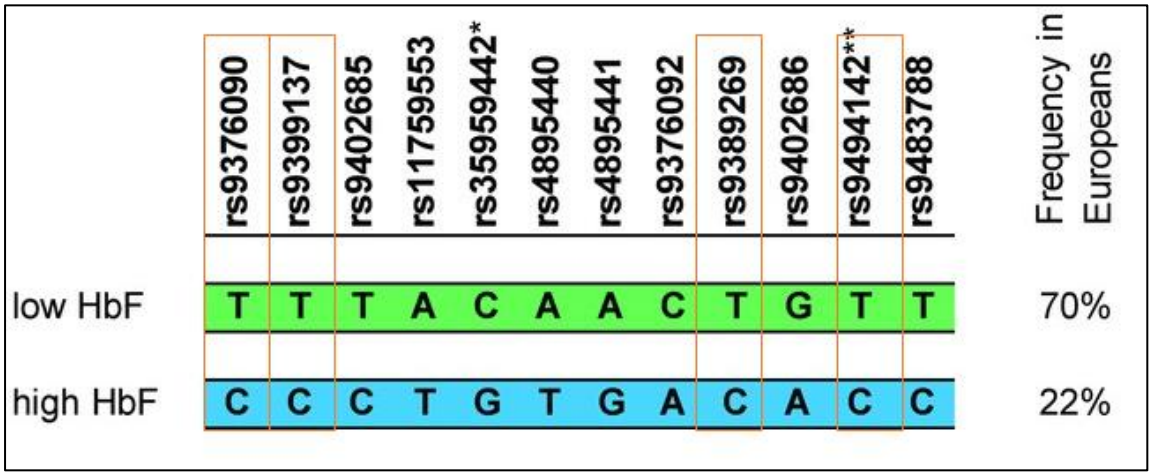

Figure 4.5: Composition of the two main haplotypes involving HbF- associated variants at HBS1L-MYB in healthy Europeans. When investigating the 12 SNPs originally reported to be strongly associated with HbF persistence, Menzel et al., found a very close LD between them and this resulted in two major haplotypes – one associated with low HbF (green) and one associated with high HbF (blue). The 4 SNPs included in our study are marked by an orange box.

In our study we investigated the effect of 4 of these SNPs (marked in orange in figure 4.5) on HbF. In the pseudo-thalassaemia group, the most frequent haplotypes were the TTTT and the CCCC haplotype. No difference in the HbF level of these haplotypes was noted. In the β-thalassaemia heterozygote, the most frequent haplotypes were also the TTTT and the CCCC haplotype. In this group a statistically significant difference in the HbF level was noted between subjects with TTTT haplotype and those with CCCC haplotype. Those with CCCC haplotype had higher HbF levels.

Conclusion

In this study, phenotypically the KLF*1* autosomal recessive mutations were not associated with any phenotypes. Compound heterozygotes or homozygotes mutation resulted in borderline HbA_2, low MCV and low MCH.

The pleotropic effects of KLF1 mutations can be attributed in part to the quantitative and qualitative effects on KLF structure and function (Gallagher, 2011). Most mutations in *KLF1* are missense variants as also shown in this study, that alter the protein coding sequences. The effect of the missense variant on the phenotype to some extent depends on its location and the change it imposes on the KLF1 protein structure and function. The data on Fam F1 and Fam F2 clearly demonstrated the huge redundancy in the regulation of *KLF1*. The outcomes of possible deficiencies were well compensated by the wildtype allele. The promoter mutations aborted the compensatory mechanisms. This observation reflected the critical importance of *KLF1* in erythropoiesis and globin gene switching.

Furthermore, from a clinical point of view, the data presented, highlighted the importance of *KLF1* sequencing in patients with haematological features resembling β-thalassaemia for differential diagnosis of microcytosis and for genetic counselling.

Adams, J. G.,3rd, Morrison, W. T., & Steinberg, M. H. (1982). Hemoglobin parchman: Double crossover within a single human gene. *Science (New York, N.Y.), 218*(4569), 291-293.

Addya, S., Keller, M. A., Delgrosso, K., Ponte, C. M., Vadigepalli, R., Gonye, G. E., & Surrey, S. (2004). Erythroid-induced commitment of K562 cells results in clusters of differentially expressed genes enriched for specific transcription regulatory elements. *Physiological Genomics, 19*(1), 117-130. doi:10.1152/physiolgenomics.00028.2004 [doi]

Armstrong, J. A., Bieker, J. J., & Emerson, B. M. (1998). A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell, 95*(1), 93-104. doi:S0092-8674(00)81785-7 [pii]

Arnaud, L., Saison, C., Helias, V., Lucien, N., Steschenko, D., Giarratana, M. C., . . . Cartron, J. P. (2010). A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. *American Journal of Human Genetics, 87*(5), 721-727. doi:10.1016/j.ajhg.2010.10.010 [doi]

Astle, W. J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A. L., . . . Soranzo, N. (2016). The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell, 167*(5), 1415-1429.e19. doi:S0092-8674(16)31463-5 [pii]

Atweh, G. F., Anagnou, N. P., Shearin, J., Forget, B. G., & Kaufman, R. E. (1985). Betathalassemia resulting from a single nucleotide substitution in an acceptor splice site. *Nucleic Acids Research, 13*(3), 777-790.

Bae, H. T., Baldwin, C. T., Sebastiani, P., Telen, M. J., Ashley-Koch, A., Garrett, M., . . . 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*(9), 19611962. doi:10.1182/blood-2012-06-432849 [doi]

Baglioni, C. (1962). The fusion of two peptide chains in hemoglobin lepore and its interpretation as a genetic deletion. *Proceedings of the National Academy of Sciences of the United States of America, 48,* 1880-1886.

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., & Shendure, J. (2011). Exome sequencing as a tool for mendelian disease gene discovery. *Nature Reviews.Genetics, 12*(11), 745-755. doi:10.1038/nrg3031 [doi]

Behringer, R. R., Ryan, T. M., Palmiter, R. D., Brinster, R. L., & Townes, T. M. (1990). Human gamma- to beta-globin gene switching in transgenic mice. *Genes & Development, 4*(3), 380389.

Berg, P. E., Mittelman, M., Elion, J., Labie, D., & Schechter, A. N. (1991). Increased protein binding to a -530 mutation of the human beta-globin gene associated with decreased betaglobin synthesis. *American Journal of Hematology, 36*(1), 42-47.

Berg, P. E., Williams, D. M., Qian, R. L., Cohen, R. B., Cao, S. X., Mittelman, M., & Schechter, A. N. (1989). A common protein binds to two silencers 5' to the human beta-globin gene. *Nucleic Acids Research, 17*(21), 8833-8852.

Bieker, J. J. (1996). Isolation, genomic structure, and expression of human erythroid kruppel-like factor (EKLF). *DNA and Cell Biology, 15*(5), 347-352. doi:10.1089/dna.1996.15.347 [doi]

Bird, G., Zorio, D. A., & Bentley, D. L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Molecular and Cellular Biology, 24*(20), 8963-8969. doi:10.1128/MCB.24.20.89638969.2004 [doi]

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods, 10*(12), 12131218. doi:10.1038/nmeth.2688 [doi]

Busque, L., Gilliland, D. G., Prchal, J. T., Sieff, C. A., Weinstein, H. J., Sokol, J. M., . . . Sokol, L. (1995). Clonality in juvenile chronic myelogenous leukemia. *Blood, 85*(1), 21-30.

Cantu, I., & Philipsen, S. (2014). Flicking the switch: Adult hemoglobin expression in erythroid cells derived from cord blood and human induced pluripotent stem cells. *Haematologica, 99*(11), 1647-1649. doi:10.3324/haematol.2014.116483 [doi]

Chaouch, L., Moumni, I., Ouragini, H., Darragi, I., Kalai, M., Chaouachi, D., . . . Abbes, S. (2016). rs11886868 and rs4671393 of BCL11A associated with HbF level variation and modulate clinical events among sickle cell anemia patients. *Hematology (Amsterdam, Netherlands), 21*(7), 425-429. doi:10.1080/10245332.2015.1107275 [doi]

Chase, M. B., Fu, S., Haga, S. B., Davenport, G., Stevenson, H., Do, K., . . . Berg, P. E. (2002). BP1, a homeodomain-containing isoform of DLX4, represses the beta-globin gene. *Molecular and Cellular Biology, 22*(8), 2505-2514.

Chebloune, Y., Pagnier, J., Trabuchet, G., Faure, C., Verdier, G., Labie, D., & Nigon, V. (1988). Structural analysis of the 5' flanking region of the beta-globin gene in african sickle cell anemia patients: Further evidence for three origins of the sickle cell mutation in africa. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(12), 4431-4435.

Chen, X., & Bieker, J. J. (1996). Erythroid kruppel-like factor (EKLF) contains a multifunctional transcriptional activation domain important for inter- and intramolecular interactions. *The EMBO Journal*, *15*(21), 5888-5896.

Chen,X,W, Mo, Q. H., Zeng, R., & Xu, X. M. (2007). A novel mutation of -73(A-->T) in the CCAAT box of the beta-globin gene identified in a patient with the mild beta-thalassemia intermedia. *Annals of Hematology, 86*(9), 653-657.

Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews.Genetics, 11*(6), 415-425. doi:10.1038/nrg2779 [doi]

Coghill, E., Eccleston, S., Fox, V., Cerruti, L., Brown, C., Cunningham, J., . . . Perkins, A. (2001). Erythroid kruppel-like factor (EKLF) coordinates erythroid cell proliferation and hemoglobinization in cell lines derived from EKLF null mice. *Blood, 97*(6), 1861-1868.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., . . . Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology, 17,* 13-016-0881-8. doi:10.1186/s13059-016-0881-8 [doi]

Cooper, D. N. (2002). Human gene mutation in pathology and evolution. *Journal of Inherited Metabolic Disease, 25*(3), 157-182.

Cyrus, C., Vatte, C., Borgio, J. F., Al-Rubaish, A., Chathoth, S., Nasserullah, Z. A., . . . Ali, A. K. (2017). Existence of HbF enhancer haplotypes at HBS1L-MYB intergenic region in transfusion-dependent saudi beta-thalassemia patients. *BioMed Research International, 2017,* 1972429. doi:10.1155/2017/1972429 [doi]

Dang, D. T., Pevsner, J., & Yang, V. W. (2000). The biology of the mammalian kruppel-like family of transcription factors. *The International Journal of Biochemistry & Cell Biology, 32*(11-12), 1103-1121. doi:S1357-2725(00)00059-5 [pii]

de Vooght, K. M., van Wijk, R., & van Solinge, W. W. (2009). Management of gene promoter mutations in molecular diagnostics. *Clinical Chemistry, 55*(4), 698-708. doi:10.1373/clinchem.2008.120931 [doi]

Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W. F., Ruddle, F., . . . Kucherlapati, R. (1977). Localization of the human alpha-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. *Cell, 12*(1), 205218. doi:0092-8674(77)90198-2 [pii]

Desjardins, P., Hansen, J. B., & Allen, M. (2009). Microvolume protein concentration determination using the NanoDrop 2000c spectrophotometer. *Journal of Visualized Experiments : JoVE, (33). pii: 1610. doi*(33), 10.3791/1610. doi:10.3791/1610 [doi]

Donohoe, M. E., Zhang, X., McGinnis, L., Biggers, J., Li, E., & Shi, Y. (1999). Targeted disruption of mouse yin yang 1 transcription factor results in peri-implantation lethality. *Molecular and Cellular Biology, 19*(10), 7237-7244.

Donze, D., Townes, T. M., & Bieker, J. J. (1995). Role of erythroid kruppel-like factor in human gamma- to beta-globin gene switching. *The Journal of Biological Chemistry, 270*(4), 19551959.

Dover, G. J., & Boyer, S. H. (1980). Quantitation of hemoglobins within individual red cells: Asynchronous biosynthesis of fetal and adult hemoglobin during erythroid maturation in normal subjects. *Blood, 56*(6), 1082-1091.

Drissen, R., Palstra, R. J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., & de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes & Development, 18*(20), 2485-2490. doi:18/20/2485 [pii]

Drissen, R., von Lindern, M., Kolbus, A., Driegen, S., Steinlein, P., Beug, H., . . . Philipsen, S. (2005). The erythroid phenotype of EKLF-null mice: Defects in hemoglobin metabolism and membrane stability. *Molecular and Cellular Biology, 25*(12), 5205-5214. doi:25/12/5205 [pii]

Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., . . . Reid, C. A. (2010). Human genome sequencing using unchained base reads on selfassembling DNA nanoarrays. *Science (New York, N.Y.), 327*(5961), 78-81. doi:10.1126/science.1181498 [doi]

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., . . . Proudfoot, N. J. (1980). The structure and evolution of the human beta-globin gene family. *Cell, 21*(3), 653-668. doi:0092-8674(80)90429-8 [pii]

Elijali, S. (2013). *Does summation of HMIP SNPs differentiate between the hb F of the βoCodon39 and the β+IVS-I- 6C thalassaemia heterozygotes* (MSc).

Elion, J., Berg, P. E., Lapoumeroulie, C., Trabuchet, G., Mittelman, M., Krishnamoorthy, R., . . . Labie, D. (1992). DNA sequence variation in a negative control region 5' to the beta-globin gene correlates with the phenotypic expression of the beta s mutation. *Blood, 79*(3), 787792.

Enver, T., Raich, N., Ebens, A. J., Papayannopoulou, T., Costantini, F., & Stamatoyannopoulos, G. (1990). Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature, 344*(6264), 309-313. doi:10.1038/344309a0 [doi]

Farrell, C. M., West, A. G., & Felsenfeld, G. (2002). Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Molecular and Cellular Biology, 22*(11), 3820-3831.

Farrell, J. J., Sherva, R. M., Chen, Z. Y., Luo, H. Y., Chu, B. F., Ha, S. Y., . . . Chui, D. H. (2011). A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression. *Blood, 117*(18), 4935-4945. doi:10.1182/blood-2010-11317081 [doi]

Felice, A. E., Borg, J., Pizzuto, M., Cassar, W., Galdies, R., Wettinger, S. B., . . . Scerri, C. A. (2017). A review of cis-trans interplay between DNA sequences 5′ to the G γ- and β-globin genes among hb F-malta-I Heterozygotes/Homozygotes and β-thalassemia Homozygotes/Compound heterozygotes, and the effects of hydroxyurea on the hb F/Ferythrocyte; the need for large multicenter trials. *Haemoglobin, 31*(2), 279-288.

Fluhr, S., Krombholz, C. F., Meier, A., Epting, T., Mucke, O., Plass, C., . . . Flotho, C. (2017). Epigenetic dysregulation of the erythropoietic transcription factor KLF1 and the beta-like globin locus in juvenile myelomonocytic leukemia. *Epigenetics, 12*(8), 715-723. doi:10.1080/15592294.2017.1356959 [doi]

Forrester, W. C., Thompson, C., Elder, J. T., & Groudine, M. (1986). A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proceedings of the National Academy of Sciences of the United States of America, 83*(5), 1359-1363.

Fraser, P., & Grosveld, F. (1998). Locus control regions, chromatin activation and transcription. *Current Opinion in Cell Biology, 10*(3), 361-365. doi:S0955-0674(98)80012-4 [pii]

Fu, S., Stevenson, H., Strovel, J. W., Haga, S. B., Stamberg, J., Do, K., & Berg, P. E. (2001). Distinct functions of two isoforms of a homeobox gene, BP1 and DLX7, in the regulation of the beta-globin gene. *Gene, 278*(1-2), 131-139. doi:S0378111901007168 [pii]

Gallagher, P. G. (2011). HbA_2: At the borderline of the KLF. *Blood, 118*, 4301-4302.

Gallagher,P.G, Maksimova,Y, Schulz,V.P, Forget,B.G. (2016). Mutation in a highly conserved COOH-terminal residue of kruppel-like factor 1 associated with elevated HbF in a compound heterozygous β-thalassemia patient with a nontransfusion-dependent thalassemia phenotype. *Hemoglobin, 40*(5), 361-364.

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. *American Journal of Human Genetics, 6*(1468), 1474.

Gerald, P. S., & Diamond, L. K. (1958). The diagnosis of thalassemia trait by starch block electrophoresis of the hemoglobin. *Blood, 13*, 61-69.

Gibney, G. T., Panhuysen, C. I., So, J. C., Ma, E. S., Ha, S. Y., Li, C. K., . . . Chui, D. H. (2008). Variation and heritability of hb F and F-cells among beta-thalassemia heterozygotes in hong kong. *American Journal of Hematology, 83*(6), 458-464. doi:10.1002/ajh.21150 [doi]

Gilman, J. G., & Huisman, T. H. (1985). DNA sequence variation associated with elevated fetal G gamma globin production. *Blood, 66*(4), 783-787.

Gnanapragasam, M. N., & Bieker, J. J. (2017). Orchestration of late events in erythropoiesis by KLF1/EKLF. *Current Opinion in Hematology, 24*(3), 183-190. doi:10.1097/MOH.0000000000000327 [doi]

Gnanapragasam, M. N., Crispino, J. D., Ali, A. M., Weinberg, R., Hoffman, R., Raza, A., & Bieker, J. J. (2018). Survey and evaluation of mutations in the human KLF1 transcription unit. *Scientific Reports, 8*(1), 6587-018-24962-3. doi:10.1038/s41598-018-24962-3 [doi]

Gonzalez-Redondo, J. M., Stoming, T. A., Kutlar, A., Kutlar, F., Lanclos, K. D., Howard, E. F., . . . Gurgey, A. (1989). A C----T substitution at nt--101 in a conserved DNA sequence of the promotor region of the beta-globin gene is associated with "silent" beta-thalassemia. *Blood, 73*(6), 1705-1711.

Graham, F. L., Smiley, J., Russell, W. C., & Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *The Journal of General Virology, 36*(1), 59-74. doi:10.1099/0022-1317-36-1-59 [doi]

Gualberto, A., LePage, D., Pons, G., Mader, S. L., Park, K., Atchison, M. L., & Walsh, K. (1992). Functional antagonism between YY1 and the serum response factor. *Molecular and Cellular Biology, 12*(9), 4209-4214.

Hahn, S. (1992). The yin and the yang of mammalian transcription. *Current Biology : CB, 2*(3), 152-154. doi:0960-9822(92)90268-F [pii]

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., & Grosveld, F. (1991). Importance of globin gene order for correct developmental expression. *Genes & Development, 5*(8), 1387-1394.

Hariharan, N., Kelley, D. E., & Perry, R. P. (1991). Delta, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. *Proceedings of the National Academy of Sciences of the United States of America, 88*(21), 9799-9803.

Hariharan, P., Gorivale, M., Colah, R., Ghosh, K., & Nadkarni, A. (2017). Does the novel KLF1 gene mutation lead to a delay in fetal hemoglobin switch? *Annals of Human Genetics, 81*(3), 125-128. doi:10.1111/ahg.12191 [doi]

Heimpel, H., & Wendt, F. (1968). Congenital dyserythropoietic anemia with karyorrhexis and multinuclearity of erythroblasts. *Helvetica Medica Acta, 34*(2), 103-115.

Helias, V., Saison, C., Peyrard, T., Vera, E., Prehu, C., Cartron, J. P., & Arnaud, L. (2013). Molecular analysis of the rare in(lu) blood type: Toward decoding the phenotypic outcome of haploinsufficiency for the transcription factor KLF1. *Human Mutation, 34*(1), 221-228. doi:10.1002/humu.22218 [doi]

Higgs, D. R., Wood, W. G., Jarman, A. P., Sharpe, J., Lida, J., Pretorius, I. M., & Ayyub, H. (1990). A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes & Development, 4*(9), 1588-1601.

Hodge, D., Coghill, E., Keys, J., Maguire, T., Hartmann, B., McDowall, A., . . . Perkins, A. (2006). A global role for EKLF in definitive and primitive erythropoiesis. *Blood, 107*(8), 3359-3370. doi:2005-07-2888 [pii]

Huang, J., Zhang, X., Liu, D., Wei, X., Shang, X., Xiong, F., . . . Xu, X. (2015). Compound heterozygosity for KLF1 mutations is associated with microcytic hypochromic anemia and increased fetal hemoglobin. *European Journal of Human Genetics : EJHG, 23*(10), 13411348. doi:10.1038/ejhg.2014.291 [doi]

Huisman, T. H., Wrightstone, R. N., Wilson, J. B., Schroeder, W. A., & Kendall, A. G. (1972). Hemoglobin kenya, the product of fusion of amd polypeptide chains. *Archives of Biochemistry and Biophysics, 153*(2), 850-853. doi:0003-9861(72)90408-0 [pii]

Humphries, R. K., Ley, T., Turner, P., Moulton, A. D., & Nienhuis, A. W. (1982). Differences in human alpha-, beta- and delta-globin gene expression in monkey kidney cells. *Cell, 30*(1), 173-183. doi:0092-8674(82)90023-X [pii]

Iolascon, A., Heimpel, H., Wahlin, A., & Tamary, H. (2013). Congenital dyserythropoietic anemias: Molecular insights and diagnostic approach. *Blood, 122*(13), 2162-2166. doi:10.1182/blood-2013-05-468223 [doi]

Jane, S. M., Ney, P. A., Vanin, E. F., Gumucio, D. L., & Nienhuis, A. W. (1992). Identification of a stage selector element in the human gamma-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the betapromoter. *The EMBO Journal, 11*(8), 2961-2969.

Kadonaga, J. T., Carner, K. R., Masiarz, F. R., & Tjian, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell, 51*(6), 1079-1090. doi:0092-8674(87)90594-0 [pii]

Kendall, A. G., Ojwang, P. J., Schroeder, W. A., & Huisman, T. H. (1973). Hemoglobin kenya, the product of a gamma-beta fusion gene: Studies of the family. *American Journal of Human Genetics, 25*(5), 548-563.

Khamphikham, P., Sripichai, O., Munkongdee, T., Fucharoen, S., Tongsima, S., & Smith, D. R. (2018). Genetic variation of kruppel-like factor 1 (KLF1) and fetal hemoglobin (HbF) levels in beta(0)-thalassemia/HbE disease. *International Journal of Hematology, 107*(3), 297-310. doi:10.1007/s12185-017-2357-3 [doi]

Kountouris, P., Lederer, C. W., Fanis, P., Feleki, X., Old, J., & Kleanthous, M. (2014). IthaGenes: An interactive database for haemoglobin variations and epidemiology. *PloS One, 9*(7), e103020. doi:10.1371/journal.pone.0103020 [doi]

Kurita, R., Suda, N., Sudo, K., Miharada, K., Hiroyama, T., Miyoshi, H., . . . Nakamura, Y. (2013). Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PloS One, 8*(3), e59890. doi:10.1371/journal.pone.0059890 [doi]

Lai, Y., Zhou, L., Yi, S., Chen, Y., Tang, Y., Yi, S., . . . He, S. (2017). The association between four SNPs (rs7482144, rs4671393, rs28384513 and rs4895441) and fetal hemoglobin levels in Chinese Zhuang beta-thalassemia intermedia patients. *Blood Cells, Molecules & Diseases, 63*, 52-57. doi:S1079-9796(16)30226-1 [pii]

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923 [doi]

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923 [doi]

Leberbauer, C., Boulme, F., Unfried, G., Huber, J., Beug, H., & Mullner, E. W. (2005). Different steroids co-regulate long-term expansion versus terminal differentiation in primary human erythroid progenitors. *Blood, 105*(1), 85-94. doi:10.1182/blood-2004-03-1002 [doi]

Lee, S. U., & Maeda, T. (2012). POK/ZBTB proteins: An emerging family of proteins that regulate lymphoid development and function. *Immunological Reviews, 247*(1), 107-119. doi:10.1111/j.1600-065X.2012.01116.x [doi]

Lee, T. C., Shi, Y., & Schwartz, R. J. (1992). Displacement of BrdUrd-induced YY1 by serum response factor activates skeletal alpha-actin transcription in embryonic myoblasts. *Proceedings of the National Academy of Sciences of the United States of America, 89*(20), 9814-9818.

Lettre, G., Sankaran, V. G., Bezerra, M. A., Araujo, A. S., Uda, M., Sanna, S., . . . 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. *Proceedings of the National Academy of Sciences of the United States of America, 105*(33), 11869-11874. doi:10.1073/pnas.0804799105 [doi]

Liang, S., Moghimi, B., Yang, T. P., Strouboulis, J., & Bungert, J. (2008). Locus control region mediated regulation of adult beta-globin gene expression. *Journal of Cellular Biochemistry, 105*(1), 9-16. doi:10.1002/jcb.21820 [doi]

Liu, D., Zhang, X., Yu, L., Cai, R., Ma, X., Zheng, C., . . . Xu, X. (2014). KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of beta-thalassemia. *Blood, 124*(5), 803-811. doi:10.1182/blood-2014-03-561779 [doi]

Lorenz, W. W., McCann, R. O., Longiaru, M., & Cormier, M. J. (1991). Isolation and expression of a cDNA encoding renilla reniformis luciferase. *Proceedings of the National Academy of Sciences of the United States of America, 88*(10), 4438-4442.

Lou, J. W., Li, D. Z., Zhang, Y., He, Y., Sun, M. N., Ye, W. L., & Liu, Y. H. (2014). Delineation of the molecular basis of borderline hemoglobin A2 in chinese individuals. *Blood Cells, Molecules & Diseases, 53*(4), 261-264. doi:10.1016/j.bcmd.2014.04.005 [doi]

Louis, N., Evelegh, C., & Graham, F. L. (1997). Cloning and sequencing of the cellular-viral junctions from the human adenovirus type 5 transformed 293 cell line. *Virology, 233*(2), 423-429. doi:S0042-6822(97)98597-4 [pii]

Lozzio, C. B., & Lozzio, B. B. (1975). Human chronic myelogenous leukemia cell-line with positive philadelphia chromosome. *Blood, 45*(3), 321-334.

Lu, Y., Hanada, T., Fujiwara, Y., Nwankwo, J. O., Wieschhaus, A. J., Hartwig, J., . . . Chishti, A. H. (2016). Gene disruption of dematin causes precipitous loss of erythrocyte membrane stability and severe hemolytic anemia. *Blood, 128*(1), 93-103. doi:10.1182/blood-2016-01692251 [doi]

Magor, G. W., Tallack, M. R., Gillinder, K. R., Bell, C. C., McCallum, N., Williams, B., & Perkins, A. C. (2015). KLF1-null neonates display hydrops fetalis and a deranged erythroid transcriptome. *Blood, 125*(15), 2405-2417. doi:10.1182/blood-2014-08-590968 [doi]

Makani, J., Komba, A. N., Cox, S. E., Oruo, J., Mwamtemi, K., Kitundu, J., . . . Williams, T. N. (2010). Malaria in patients with sickle cell anemia: Burden, risk factors, and outcome at the outpatient clinic and during hospitalization. *Blood, 115*(2), 215-220. doi:10.1182/blood2009-07-233528 [doi]

Manca, L., & Masala, B. (2008). Disorders of the synthesis of human fetal hemoglobin. *IUBMB Life, 60*(2), 94-111. doi:10.1002/iub.4 [doi]

Maquat, L. E. (2004). Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics. *Nature Reviews.Molecular Cell Biology, 5*(2), 89-99. doi:10.1038/nrm1310 [doi]

Mas, C., Lussier-Price, M., Soni, S., Morse, T., Arseneault, G., Di Lello, P., . . . Omichinski, J. G. (2011). Structural and functional characterization of an atypical activation domain in erythroid kruppel-like factor (EKLF). *Proceedings of the National Academy of Sciences of the United States of America, 108*(26), 10484-10489. doi:10.1073/pnas.1017029108 [doi]

Masuda, T., Wang, X., Maeda, M., Canver, M. C., Sher, F., Funnell, A. P., . . . Maeda, T. (2016). Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science (New York, N.Y.), 351*(6270), 285-289. doi:10.1126/science.aad3312 [doi]

Matthews, J. C., Hori, K., & Cormier, M. J. (1977). Purification and properties of renilla reniformis luciferase. *Biochemistry, 16*(1), 85-91.

McConnell, B. B., & Yang, V. W. (2010). Mammalian kruppel-like factors in health and diseases. *Physiological Reviews, 90*(4), 1337-1381. doi:10.1152/physrev.00058.2009 [doi]

Meier, V. S., & Groner, B. (1994). The nuclear factor YY1 participates in repression of the betacasein gene promoter in mammary epithelial cells and is counteracted by mammary gland factor during lactogenic hormone induction. *Molecular and Cellular Biology, 14*(1), 128137.

Menzel, S., Garner, C., Rooks, H., Spector, T. D., & Thein, S. L. (2013). HbA2 levels in normal adults are influenced by two distinct genetic mechanisms. *British Journal of Haematology, 160*(1), 101-105. doi:10.1111/bjh.12084 [doi]

Menzel, S., Rooks, H., Zelenika, D., Mtatiro, S. N., Gnanakulasekaran, A., Drasar, E., . . . Thein, S. L. (2014). Global genetic architecture of an erythroid quantitative trait locus, HMIP-2. *Annals of Human Genetics, 78*(6), 434-451. doi:10.1111/ahg.12077 [doi]

Metherall, J. E., Collins, F. S., Pan, J., Weissman, S. M., & Forget, B. G. (1986). Beta zero thalassemia caused by a base substitution that creates an alternative splice acceptor site in an intron. *The EMBO Journal, 5*(10), 2551-2557.

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. *Molecular and Cellular Biology, 13*(5), 2776-2786.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research, 16*(3), 1215.

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *Journal of Immunological Methods, 65*(1-2), 55-63. doi:0022-1759(83)90303-4 [pii]

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American, 262*(4), 56-61, 64-5.

Ngueyen, T., Joly, P., Moulsma, M., Bonello-Palot, N., & Francina, A. (2010). The XMNI Gγ polymorphism influences hemoglobin F synthesis contrary to BCL11A and HBS1L-MYB SNPs in a cohort of 57 β-thalassemia intermedia patients. *Blood Cells, Molecules & Diseases, 52*(2), 124-127.

Niemeyer, C. M., Arico, M., Basso, G., Biondi, A., Cantu Rajnoldi, A., Creutzig, U., . . . Zimmermann, M. (1997). Chronic myelomonocytic leukemia in childhood: A retrospective analysis of 110 cases. European working group on myelodysplastic syndromes in childhood (EWOG-MDS). *Blood, 89*(10), 3534-3543.

Nilson, D. G., Sabatino, D. E., Bodine, D. M., & Gallagher, P. G. (2006). Major erythrocyte membrane protein genes in EKLF-deficient mice. *Experimental Hematology, 34*(6), 705712. doi:S0301-472X(06)00126-3 [pii]

Nitta, T., Kawano, F., Yamashiro, Y., Takagi, F., Murata, T., Tanaka, T., . . . Hattori, Y. (2015). A new kruppel-like factor 1 mutation (c.947G > A or p.C316Y) in humans causes betathalassemia minor. *Hemoglobin, 39*(2), 121-126. doi:10.3109/03630269.2015.1008702 [doi] 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*(6529), 316-318. doi:10.1038/375316a0 [doi]

Orkin, S. H., Kazazian, H. H.,Jr, Antonarakis, S. E., Ostrer, H., Goff, S. C., & Sexton, J. P. (1982). Abnormal RNA processing due to the exon mutation of beta E-globin gene. *Nature, 300*(5894), 768-769.

Paglietti, M. E., Satta, S., Sollaino, M. C., Barella, S., Ventrella, A., Desogus, M. F., . . . Origa, R. (2016). The problem of borderline hemoglobin A2 levels in the screening for betathalassemia carriers in sardinia. *Acta Haematologica, 135*(4), 193-199. doi:10.1159/000442194 [doi]

Pereira, C., Relvas, L., Bento, C., Abade, A., Ribeiro, M. L., & Manco, L. (2015). Polymorphic variations influencing fetal hemoglobin levels: Association study in beta-thalassemia carriers and in normal individuals of portuguese origin. *Blood Cells, Molecules & Diseases, 54*(4), 315-320. doi:10.1016/j.bcmd.2015.02.001 [doi]

Perkins, A. C., Peterson, K. R., Stamatoyannopoulos, G., Witkowska, H. E., & Orkin, S. H. (2000). Fetal expression of a human agamma globin transgene rescues globin chain imbalance but not hemolysis in EKLF null mouse embryos. *Blood, 95*(5), 1827-1833.

Perkins, A. C., Sharpe, A. H., & Orkin, S. H. (1995). Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature, 375*(6529), 318-322. doi:10.1038/375318a0 [doi]

Perseu, L., Satta, S., Moi, P., Demartis, F. R., Manunza, L., Sollaino, M. C., . . . Galanello, R. (2011). KLF1 gene mutations cause borderline HbA(2). *Blood, 118*(16), 4454-4458. doi:10.1182/blood-2011-04-345736 [doi]

Peterson, K. R., & Stamatoyannopoulos, G. (1993). Role of gene order in developmental control of human gamma- and beta-globin gene expression. *Molecular and Cellular Biology, 13*(8), 4836-4843.

Pilon, A. M., Arcasoy, M. O., Dressman, H. K., Vayda, S. E., Maksimova, Y. D., Sangerman, J. I., . . . Bodine, D. M. (2008). Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2. *Molecular and Cellular Biology, 28*(24), 7394-7401. doi:10.1128/MCB.01087-08 [doi]

Radmilovic, M., Zukic, B., Petrovic, M. S., Bartsakoulia, M., Stankovic, B., Kotur, N., . . . Pavlovic, S. (2013). Functional analysis of a novel KLF1 gene promoter variation associated with hereditary persistence of fetal hemoglobin. *Annals of Hematology, 92*(1), 53-58. doi:10.1007/s00277-012-1625-9 [doi]

Raich, N., Enver, T., Nakamoto, B., Josephson, B., Papayannopoulou, T., & Stamatoyannopoulos, G. (1990). Autonomous developmental control of human embryonic globin gene switching in transgenic mice. *Science (New York, N.Y.), 250*(4984), 1147-1149.

Recillas-Targa, F. (2006). Multiple strategies for gene transfer, expression, knockdown, and chromatin influence in mammalian cell lines and transgenic animals. *Molecular Biotechnology, 34*(3), 337-354. doi:MB:34:3:337 [pii]

Rees, D. C., Styles, L., Vichinsky, E. P., Clegg, J. B., & Weatherall, D. J. (1998). The hemoglobin E syndromes. *Annals of the New York Academy of Sciences, 850*, 334-343.

Riggs, K. J., Saleque, S., Wong, K. K., Merrell, K. T., Lee, J. S., Shi, Y., & Calame, K. (1993). Yin-yang 1 activates the c-myc promoter. *Molecular and Cellular Biology, 13*(12), 74877495.

Riggs, K. J., Saleque, S., Wong, K. K., Merrell, K. T., Lee, J. S., Shi, Y., & Calame, K. (1993). Yin-yang 1 activates the c-myc promoter. *Molecular and Cellular Biology, 13*(12), 74877495.

Roversi, F., Ferreira Cunha, A., Lanaro, C., Brugnerotto, A., Favero, M., Albuquerque, A. and Ferreira Costa, F. (2010). The -195C→G substitution in Brazilian hereditary persistence of fetal hemoglobin decreases NF-E1/YY1 binding and increases PAX1 binding to the A gamma globin promoter region. *Blood*, *116*(2006)

Satta, S., Paglietti, M. E., Sollaino, M. C., Barella, S., Moi, P., Desogus, M. F., . . . Origa, R. (2017). Changes in HbA2 and HbF in alpha thalassemia carriers with KLF1 mutation. *Blood Cells, Molecules & Diseases, 64*, 30-32. doi:S1079-9796(16)30332-1 [pii]

Satta, S., Perseu, L., Moi, P., Asunis, I., Cabriolu, A., Maccioni, L., . . . Galanello, R. (2011). Compound heterozygosity for KLF1 mutations associated with remarkable increase of fetal hemoglobin and red cell protoporphyrin. *Haematologica, 96*(5), 767-770. doi:10.3324/haematol.2010.037333 [doi]

Scerri, C. A. (1998). *Clinical and molecular pathology of the beta IVS1-6C thalassaemia in Malta* (PhD).

Schuh, R., Aicher, W., Gaul, U., Cote, S., Preiss, A., Maier, D., . . . Kemler, R. (1986). A conserved family of nuclear proteins containing structural elements of the finger protein encoded by kruppel, a drosophila segmentation gene. *Cell, 47*(6), 1025-1032. doi:00928674(86)90817-2 [pii]

Shi, Y., Seto, E., Chang, L. S., & Shenk, T. (1991). Transcriptional repression by YY1, a human GLI-kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell, 67*(2),

Shih, D. M., Wall, R. J., & Shapiro, S. G. (1990). Developmentally regulated and erythroidspecific expression of the human embryonic beta-globin gene in transgenic mice. *Nucleic Acids Research, 18*(18), 5465-5472.

Siatecka, M., Lohmann, F., Bao, S., & Bieker, J. J. (2010). EKLF directly activates the p21WAF1/CIP1 gene by proximal promoter and novel intronic regulatory regions during

erythroid differentiation. *Molecular and Cellular Biology, 30*(11), 2811-2822. doi:10.1128/MCB.01016-09 [doi]

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*(5), 2081-2088. doi:10.1182/blood-2008-03-145672 [doi]

Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. *Experimental Hematology, 33*(3), 259-271. doi:S0301472X(04)00409-6 [pii]

Suzuki, M., Yamazaki, H., Mukai, H. Y., Motohashi, H., Shi, L., Tanabe, O., . . . Yamamoto, M. (2013). Disruption of the Hbs1l-myb locus causes hereditary persistence of fetal hemoglobin in a mouse model. *Molecular and Cellular Biology, 33*(8), 1687-1695. doi:10.1128/MCB.01617-12 [doi]

Tallack, M. R., & Perkins, A. C. (2010). KLF1 directly coordinates almost all aspects of terminal erythroid differentiation. *IUBMB Life, 62*(12), 886-890. doi:10.1002/iub.404 [doi]

Tallack, M. R., Keys, J. R., & Perkins, A. C. (2007). Erythroid kruppel-like factor regulates the G1 cyclin dependent kinase inhibitor p18INK4c. *Journal of Molecular Biology, 369*(2), 313-321. doi:S0022-2836(07)00342-7 [pii]

Tallack, M. R., Keys, J. R., Humbert, P. O., & Perkins, A. C. (2009). EKLF/KLF1 controls cell cycle entry via direct regulation of E2f2. *The Journal of Biological Chemistry, 284*(31), 20966-20974. doi:10.1074/jbc.M109.006346 [doi]

Tepakhan, W., Yamsri, S., Sanchaisuriya, K., Fucharoen, G., Xu, X., & Fucharoen, S. (2016). Nine known and five novel mutations in the erythroid transcription factor KLF1 gene and phenotypic expression of fetal hemoglobin in hemoglobin E disorder. *Blood Cells, Molecules & Diseases, 59*, 85-91. doi:10.1016/j.bcmd.2016.04.010 [doi]

Thein, S. L., & Weatherall, D. J. (1989). A non-deletion hereditary persistence of fetal hemoglobin (HPFH) determinant not linked to the beta-globin gene complex. *Progress in Clinical and Biological Research, 316B*, 97-111.

Thein, S. L., Menzel, S., Peng, X., Best, S., Jiang, J., Close, J., . . . 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. *Proceedings of the National Academy of Sciences of the United States of America, 104*(27), 11346-11351. doi:0611393104 [pii]

Treisman, R., Orkin, S. H., & Maniatis, T. (1983). Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature, 302*(5909), 591-596.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research, 40*(15), e115. doi:gks596 [pii]

van den Akker, E., Satchwell, T. J., Pellegrin, S., Daniels, G., & Toye, A. M. (2010). The majority of the in vitro erythroid expansion potential resides in CD34(-) cells, outweighing the contribution of CD34(+) cells and significantly increasing the erythroblast yield from peripheral blood samples. *Haematologica, 95*(9), 1594-1598. doi:10.3324/haematol.2009.019828 [doi]

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nature Reviews.Genetics, 10*(1), 57-63. doi:10.1038/nrg2484 [doi]

Weatherall, D. J. (2001). Phenotype-genotype relationships in monogenic disease: Lessons from the thalassaemias. *Nature Reviews.Genetics, 2*(4), 245-255. doi:10.1038/35066048 [doi]

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. *British Journal of Haematology, 29*(2), 205-220.

Wienert, B., Martyn, G. E., Kurita, R., Nakamura, Y., Quinlan, K. G. R., & Crossley, M. (2017). KLF1 drives the expression of fetal hemoglobin in british HPFH. *Blood, 130*(6), 803-807. doi:10.1182/blood-2017-02-767400 [doi]

Wijgerde, M., Grosveld, F., & Fraser, P. (1995). Transcription complex stability and chromatin dynamics in vivo. *Nature, 377*(6546), 209-213. doi:10.1038/377209a0 [doi]

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). PrimerBLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics, 13*, 134-2105-13-134. doi:10.1186/1471-2105-13-134 [doi]

Ye, T. Z., Gordon, C. T., Lai, Y. H., Fujiwara, Y., Peters, L. L., Perkins, A. C., & Chui, D. H. (2000). Ermap, a gene coding for a novel erythroid specific adhesion/receptor membrane protein. *Gene, 242*(1-2), 337-345. doi:S0378-1119(99)00516-8 [pii]

Yi, Z., Cohen-Barak, O., Hagiwara, N., Kingsley, P. D., Fuchs, D. A., Erickson, D. T., . . . Brilliant, M. H. (2006). Sox6 directly silences epsilon globin expression in definitive erythropoiesis. *PLoS Genetics, 2*(2), e14. doi:10.1371/journal.pgen.0020014 [doi]

Yu, L. H., Liu, D., Cai, R., Shang, X., Zhang, X. H., Ma, X. X., . . . Xu, X. M. (2015). Changes in hematological parameters in alpha-thalassemia individuals co-inherited with erythroid kruppel-like factor mutations. *Clinical Genetics, 88*(1), 56-61. doi:10.1111/cge.12443 [doi]

Zhang, L., & Nabel, G. J. (1994). Positive and negative regulation of IL-2 gene expression: Role of multiple regulatory sites. *Cytokine, 6*(3), 221-228. doi:1043-4666(94)90016-7 [pii]

Appendix A

Informed Consent Form – English

Informed Consent Form – Maltese

Appendix B – E

Supplementary information to the Methodology Section

Appendix F

Supplementary information to the Results Section

Formola ta' Kunsens Infurmat - Globin Bank

- Nixtieq nistidnek biex tipparteċipa f'dan l-istudju ta' riċerka fuq "**l-Epidemjoloġija Kilinka u tal-Popolazzjoni ta' Mard tal-Emoglobina u l-Ġenetika tal-Globin gene"**.

- Għandek tipparteċipa biss jekk tixtieq. Qabel ma tiddeċiedi jekk tixtieqx tieħu sehem, huwa importanti li tifhem għaliex ir-riċerka qeda issir

- Il-bank ta' l-emoglobina "Globin Bank" fih kollezzjoni ta' mard rari tad-demm (eż. Talessimija) li qed jiġi żviluppat għall riċerka speċifika u għodod dijanjostiċi, kif ukoll, metodi terapewtiċi ġodda għall-pazjenti li jsofru minn mard tal-emoglobina.

- Ir-riċerka li għandha titwettaq ser tkun taħt is-superviżjoni tal-Kumitat ta' l-Etika fl-Università ta' Malta (UREC). Id-drittijiet individwali tiegħek ser ikunu rispettati skont ir-regoli etiċi aċċettati internazzjonalment.

- Jekk inti taqbel li tipparteċipa f'dan il-proġett, madwar 3ml ta' demm għandu jkun meħud mid-driegħ tiegħek minn infermier kwalifikat u b'esperjenza jew phlebotomist. Jista' jkun hemm xi fsada, tbenġil, u / jew skumdita' fis-sit tal-injezzjoni.

- Il-kampjuni ser jinżammu inizjalment għal perjodu ta' għaxar (10) snin jew aktar taħt ir-responsabilta tal-investigatur prinċipali, il-Professur Alex Felice u/jew id-delegati tiegħu.

- DNA ser jiġi estratt mid-demm għall-analiżi ġenetika bħala parti mill-istudji fuq il-emoglobina, madankollu l-kampjuni jistgħu jintużaw fi proġetti ta' riċerka oħra li jirċievu l-approvazzjoni etika mill-UREC.

- Il-kampjuni bioloġiċi u d-data personali ser jiżammu separatament.

- Il-kampjuni kollha ser jkunu parzjalment anonimizzati b' kodiċi u jistgħu jiġu marbuta ma' informazzjoni personali tiegħek permezz ta' dan il-kodiċi biss mill-Kustodju tad-Data jew id-delegati tiegħu, sabiex jiżguraw il-privatezza tiegħek.

- Riċerkaturi, min iħaddem, organizzazzjonijiet mhux governattivi, kumpaniji ta' assigurazzjoni jew istituzzjonijiet edukattivi mhux se jkollhom aċċess għad-data personali.

- Ir-risultati miksuba se tkun ippubblikati biss b'mod anonimu u aggregata, bħal perċentwali jew data numerika mingħajr identifikazzjoni tal-parteċipant. Taħt l-ebda ċirkostanza d-data se tiġi pprovduta b'mod individwalizzat. Bħala riżultat, il-privatezza tiegħek se tkun żgurata u rispettata.

- Id-donazzjoni hija altruwistika u l-ebda benefiċċju ekonomiku għandhom jinkisbu jekk test dijanjostiku ġdid huwa żviluppat matul dan l-istudju.

- Il-Parteċipazzjoni tiegħek f'dan il-proġett hija volontarja u tista' tirtira l-kunsens tiegħek fi kwalunkwe ħin, mingħajr spjegazzjoni ulterjuri. Jekk tirtira l-parteċipazzjoni tiegħek, tista' tiddeċiedi jekk il-materjal diġà miġbur jibqax parti mill-istudju jew jinqered.

- Jekk inti teħtieġ aktar informazzjoni jew li tixtieq li tikkomunika kull bidla fl-indirizz, inti tista' tikkuntattja lil Ms Joanna Vella fil- Malta BioBank waqt il-ħinijiet tal-uffiċċju fuq 23402170/23402774 jew bl-email: joanna.vella@um.edu.mt. Għal aktar informazzjoni żur il-website: www.um.edu.mt/biobank.

_____

Investigatur Prinċipali

Formola ta' Kunsens Infurmat (li għandha tinżamm mill- Malta BioBank) **Data:** _____

Thalassemia Testing	☐
Other Haemoglobin Testing	☐

Billi taċċetta li tieħu sehem fil-proġett "**Epidemjoloġija Klinika u tal-Popolazzjoni ta' Mard tal-Emoglobina u il-Ġenetika tal-Globin gene**", inti tagħti il-kunsens tiegħek:

1. Lill-Investigatur Prinċipali u / jew id-delegati tiegħu biex jagħmlu l-osservazzjonijiet xierqa jew testijiet jew it-tnejn u li jieħdu l-kampjuni bioloġici meħtieġa.
2. Lill-Investigatur Prinċipali u / jew id-delegati tiegħu biex jiksbu aċċess għar-reġistri mediċi u oħrajn relatati mas-saħħa għall-ħażna fit-tul u l-użu ta' dan u informazzjoni oħra għal skopijiet ta' riċerka relatata mas-saħħa (anki wara l-inkapaċita' tiegħek jew mewt). Din l-informazzjoni ser tkun trattata bl-akbar kunfidenzjalita' u ser tinzamm f'post sigur.

		Iva	Le
1.	Jiena qrajt u fhmit l-istudju "**Epidemjoloġija Klinika u tal-Popolazzjoni ta' Mard tal-Emoglobina u il-Ġenetika tal-Globin gene**" u kelli l-opportunita' li nsaqsi mistoqsijiet.	☐	☐
2.	Jiena naqbel li nieħu sehem fl-istudju "**Epidemjoloġija Klinika u tal-Popolazzjoni ta' Mard tal-Emoglobina u il-Ġenetika tal-Globin gene**" - li tinkludi analiżi Ġenetika.	☐	☐
3.	Nagħti il-permess biex il-kampjuni tiegħi jiġu mħazzna fill-Malta BioBank	☐	☐
4.	**Użu tal-Kampjuni**		
a.	Jiena nipprovdi il-permess għall-użu tal-kampjuni tiegħi fil-proġetti ta' riċerka oħra li huma etikament approvati.	☐	☐
b.	Nixtieq li niġi ik-kuntattjat qabel il-kampjuni tiegħi jiġu użati fil-proġetti ta' riċerka oħra li huma etikament approvati.	☐	☐
5.	**Komunikazzjoni tar-riżultati:**		
	Nixtieq inkun infurmat bir-riżultati mil-Klinika Ġenetika a. Iva fil każijiet kollha jew b. Iva fil każ ta' trattament preventiv biss	☐ ☐	☐

Dettalji tad-donatur	**Firma**	**Dettalji tar-Rappreżentant/i Legali**	**Firma**
Attach Label with ID, Name, Surname, Address & Signature		Attach Label with ID, Name, Surname, Address & Signature	

Code tal-Kampjun				**Code tal-Kampjun/i**			
Sess	Maskili ☐ Femminili ☐	**Eta' tad-Donatur**		**Sess**	Maskili ☐ Femminili ☐	**Eta' tar-Raprezzentant/i Legali**	
Telefon / Mobile Nru.				**Telefon / Mobile Nru.**			
Email address				**Email address**			
Timbru tal-Konsulent tal-Ġenetika				**Firma tal-Konsulent tal-Ġenetika**			
Isem tax-Xhud (Ittri Kapitali)				**Firma tax-Xhud**			

Informed consent – Globin bank

- We would like to invite you to participate in the research about **the Clinical and Population Epidemiology of Haemoglobin Disorders and in-vivo Developmental Genetics of Globin Gene Control**.
- You should only participate if you want to. Before you decide whether you want to take part, it is important for you to understand why the research is being done.
- The Globin Bank collection of rare blood disorders (e.g thalassemia) is being developed for the development of specific research and diagnostic tools, as well as, new therapeutic methods for patients suffering from haemoglobin diseases.
- All the research will be carried out under the supervision of the University of Malta Research Ethics Committee (UREC). Your individual rights will be respected in accordance with the internationally accepted ethical rules.
- If you agree to participate in this project, a sample of blood (3ml) will be taken by venipuncture by a qualified nurse or phlebotomist. There may be some bleeding, bruising or discomfort at the injection site.
- The samples will be stored in the Malta Biobank for an initial period of ten years or longer under the responsibility of the principle investigator, Professor Alex Felice and/or his delegates.
- DNA will be extracted from the blood for genetic analysis as part of the globin studies, however the samples may be used in other research projects that receive ethics approval by the UREC.
- The biological samples and personal data will be stored separately.
- All the samples will be partially anonymised by coding and can only be linked to your personal information through this code only by the Data Custodian or his delegate, to ensure your privacy.
- Researchers, employers, governmental organisations, insurance companies or educational institutions will not have access to personal data.
- Data obtained will only be published in an anonymous and aggregated way, such as percentages or numerical data without identification of the participant.
- Under no circumstances will data be provided in an individualized way. As a result, your privacy will be ensured and respected.
- The donation is altruistic and no economic benefit shall be obtained if a new diagnostic test is developed during this study.
- Participation in this project is voluntary and you may withdraw your consent at any time, without further explanation. Should you withdraw your participation, you can decide whether the material already collected can remain part of the study or destroyed.
- Should you need more information or wish to communicate any change of address, you can contact Ms Joanna Vella at the Malta Biobank during office hours on 23402170/23402774 or by email: joanna.vella@um.edu.mt. For further information visit the website: www.um.edu.mt/biobank.

_____

Principal Investigator

Informed Consent (to be kept by the Malta BioBank) **Date:** _____

Thalassemia Testing	☐
Other Haemoglobin Testing	☐

By agreeing to take part in the study: "**Clinical and Population Epidemiology of Haemoglobin Disorders and in-vivo Developmental Genetics of Globin Gene Control**" you are giving your consent:

1. To the Principal Investigator and his delegates to make the appropriate observations and/or tests and to take the necessary biological samples.
2. To the Principal Investigator and his delegates to gain access to medical and other health-related records and for the long-term storage and use of this and other information for health-related research purposes (even after your incapacity or death). This information will be treated with the utmost confidentiality and will be stored in a secure location.

		Yes	No
1.	I have read and understood what the project titled "**Clinical and Population Epidemiology of Haemoglobin Disorders and in-vivo Developmental Genetics of Globin Gene Control**" is about and have had the opportunity to ask questions.	☐	☐
2.	I agree to take part in "**Clinical and Population Epidemiology of Haemoglobin Disorders and in-vivo Developmental Genetics of Globin Gene Control**" which includes genetic analysis.	☐	☐
3.	I give my permission for my samples to be stored in the Malta BioBank.	☐	☐
4.	**Use of Samples:**		
a.	I provide permission for the use of my samples in other ethically approved research projects.	☐	☐
b.	I would like to be re-contacted before my samples are used in other research projects.	☐	☐
5.	**Communication of results:**		
	I would like to be informed of the results through the Genetics Clinic: a. Yes in all cases or b. Yes in the case of preventative treatment only	☐ ☐	☐

Donor's Details	**Signature**	**Legal Guardian/s Details**	**Signature**
Attach Label with ID, Name, Surname, Address & Signature		Attach Label with ID, Name, Surname, Address & Signature	

Sample Code				**Sample Code/s**			
Gender	Female ☐ Male ☐	**Age of donor**		**Gender**	Female ☐ Male ☐	**Age/s of Legal Guardian/s**	
Telephone / Mobile No.				**Telephone / Mobile No.**			
Email address				**Email address**			
Stamp of Consultant Medical Geneticist				**Signature of Consultant Medical Geneticist**			
Witness Name (Block Capital)				**Signature of Witness**			

L-UNIVERSITÀ TA' MALTA

Msida – Malta
Skola Medika
Sptar Mater Dei

UNIVERSITY OF MALTA

Msida – Malta
Medical School
Mater Dei Hospital

Ref No: **45/2014**

Thursday, 18th December 2014

Prof Alex Felice
Laboratory of Molecular Genetics
University of Malta

Dear Prof Felice

Please refer to your application submitted to the Research Ethics Committee in connection with your research entitled:

Clinical and Population Epidemiology of Haemoglobin Disorders and in-vivo Developmental Genetics of Globin Gene Control

The University Research Ethics Committee granted ethical approval for the above mentioned protocol.

Yours sincerely,

Dr. Mario Vassallo
Chairman
Research Ethics Committee

Appendix B

Preparation of Reagents for Salting Out for DNA extraction

10 x SE buffer

NaCl	750nM	21.95 g/L
Na_2EDTA	250nM	46.55 g/L

Dissolve completely in deionised water and adjust to pH 8.0.

1 x SE buffer

10 x SE	100 mls
Deionised water	900 mls

10 x Erythrocyte lysis buffer

NH_4Cl	1.55M	42.25g/l
$KHCO_3$	100nM	5.0g/l
Na_2EDTA	1nM	0.186g/l

Dissolve completely in deionised water and adjust to pH 7.4.

1 x Erythrocyte lysis buffer

10 x Erythrocyte lysis buffer	100 mls
Deionised water	900 mls

10% SDS

SDS (L4390, Sigma Aldrich, USA)	100g
Deionised water	800 mls

Solution topped up to 1 L deionised water

1 x SE + 1% SDS

10 x SE buffer	25 mls
10% SDS	25 mls

Top up to 250 mls of deionised water

6M NaCl

NaCl	175.32g

Top up to 500mls with distilled water

Appendix C &D

Synthetic DNA Sequences &

The following are the sequences of synthetic DNA subcloned in the pGL4.10[luc2] Vector (Promega), between NheI and EcoRV. The wildtype sequence is followed by the sequences containing SNPs where;

- **GCTAGC** is the site where the restriction enzyme NheI was used to cleave DNA

Figure 1 - Restriction site of NheI

- **GATATC** is the site where the restriction enzyme EcoRV was used to cleave DNA

Figure 2 - Restriction site of EcoRV

- Substitution SNPs are marked in **red**
- **ATG**GAAGATG...**Luc** marks the initiation site

Wildtype sequence (1,282bp)

CACCGCACCTGGCCTTGTTTTTGTTTGTGTCACCCAGGTTGATCTGCCAGGAATAAAGATGG

CCACCCAACTTTTTTTTTTTTCCAGCGACAGAGTCTTGCTTTCTGTGCCCCAGGCTGGAC

TGCAGTGGTGTGATTATAGTTCAGTGCAGTCCCGACCTCCTGGGCTCAGGCAATCCTCCC

ACCTCAGCCTCCTGAGTAGCTGGGACTATGGGCATGCACCACCACTCACAGCTTTTTGTT

GGAGAGATAGGTTTTATGTTGCCCAGGCTGGTCTTGAAATCCTGGTGTCAAGTGATCCTC

CTGCCTCAGTCTTCCAATGAGTGACTATGAGTTATGTGTCTGTAATACGTATCCATGTCC

CCTTCCCAGGCTCCCAGGGCACCAGAGTGGAGGTTCCTGTTGTAGAAACTCAGATCCTCT

CCTCATGTTGGGCAGAATCAAGGAGCAGCCAGGCCCAGAGCCAGGGCACTGGTCTCCTGC

AGGTCAGGTACTGCCTGTGGCTTGATCCAACGGTCCTATCCCACCCAGGAGGAGAGAGGG

TCACTTTTCCCTTGGCTGCCCCCATCGCACTAAAGCAGCTGGCACTGAACCAGCCTCCAT

GCAGTCCCATGCAGTGCCACCCAAGGGTCCCCAGTAGACAATGGTGGGCCAGTTGTCAGG

GGCTTCCTCCTGCTGCAGGGCTGAGACCCTGGGAGGTCCCAACCCAGGCAAATTGAACGC

CAGGCTAATTTGAAGACCCAACTCCCAGCCCTCCCCTTCACCGGAGGACAGAGCTCTAGC

TGGCCTGGGCCCCCACCTGATAGCAGCCTCCAACGTCTGGGGTGTCTGATAATGCTTGGC

GGGGAGCTCGTGCCAAGTCCCGCCATCAGCACGGTTGTTGCTGTTTACTGGGGAGGGGGA

GGGCTGTGGAGCCTCAATCAGGGGGACAGGGGGTCCCACAGCTTCTTCCCAGAATACCCT

TTCTGCCTTTTCCAGGAAAGTTAACTGAGGGAAGACCCCCAAGTCTCTCCTTCTTTGGAG

ACCCAATGTCTGTTTTTACCCAGCACCTGGACCCTCAAACCCTGAACCCCCCAACCCTTG

ATATTTGACTTGGCTTTGGACACAGGGTTAGTCTTTAACCCCAGCCCCAGACAGGCCAAC

GTGAAGTTTGTGCCCCAGAAACAGTGCCCCCCCGCCGCCTTGCCTTGCTTTGCCTTATCA

GAGGCTGCAGCCAATCAGCTAAGGACAGAGAGGAGCCCTCGAAGGGGCTATCACAGCCTC

AGAGTTCACGAGGCAGCCGA**GATATC**AAGATCTGGCCTCGGCGGCCAAGCTTGGCAATC

CGTACTGTTGGTAAAGCCACC**ATG**GAAGAT…Luc

Substitution -1133C>A (rs112943513)

CACCGCACCTGGCCTTGTTTTTGTTTGTGTCACCCAGGTTGATCTGCCAGGAATAAAGATGG

CCACCCAACTTTTTTTTTTTTCCAGCGACAGAGTCTTGCTTTCTGTGCCCCAGGCTGGAC

TGCAGTGGTGTGATTATAGTTCAGTGCAGTCCCGACCTCCTGGGCTCAGGCAATCCTCCC

ACCTCAGC**A**TCCTGAGTAGCTGGGACTATGGGCATGCACCACCACTCACAGCTTTTTGTT

GGAGAGATAGGTTTTATGTTGCCCAGGCTGGTCTTGAAATCCTGGTGTCAAGTGATCCTC

CTGCCTCAGTCTTCCAATGAGTGACTATGAGTTATGTGTCTGTAATACGTATCCATGTCC

CCTTCCCAGGCTCCCAGGGCACCAGAGTGGAGGTTCCTGTTGTAGAAACTCAGATCCTCT

CCTCATGTTGGGCAGAATCAAGGAGCAGCCAGGCCCAGAGCCAGGGCACTGGTCTCCTGC

AGGTCAGGTACTGCCTGTGGCTTGATCCAACGGTCCTATCCCACCCAGGAGGAGAGAGGG

TCACTTTTCCCTTGGCTGCCCCCATCGCACTAAAGCAGCTGGCACTGAACCAGCCTCCAT

GCAGTCCCATGCAGTGCCACCCAAGGGTCCCCAGTAGACAATGGTGGGCCAGTTGTCAGG

GGCTTCCTCCTGCTGCAGGGCTGAGACCCTGGGAGGTCCCAACCCAGGCAAATTGAACGC

CAGGCTAATTTGAAGACCCAACTCCCAGCCCTCCCCTTCACCGGAGGACAGAGCTCTAGC

TGGCCTGGGCCCCCACCTGATAGCAGCCTCCAACGTCTGGGGTGTCTGATAATGCTTGGC

GGGGAGCTCGTGCCAAGTCCCGCCATCAGCACGGTTGTTGCTGTTTACTGGGGAGGGGGA

GGGCTGTGGAGCCTCAATCAGGGGGACAGGGGGTCCCACAGCTTCTTCCCAGAATACCCT

TTCTGCCTTTTCCAGGAAAGTTAACTGAGGGAAGACCCCCAAGTCTCTCCTTCTTTGGAG

ACCCAATGTCTGTTTTTACCCAGCACCTGGACCCTCAAACCCTGAACCCCCCAACCCTTG

ATATTTGACTTGGCTTTGGACACAGGGTTAGTCTTTAACCCCAGCCCCAGACAGGCCAAC

GTGAAGTTTGTGCCCCAGAAACAGTGCCCCCCCGCCGCCTTGCCTTGCTTTGCCTTATCA

GAGGCTGCAGCCAATCAGCTAAGGACAGAGAGGAGCCCTCGAAGGGGCTATCACAGCCTC

AGAGTTCACGAGGCAGCCGA **GATATC**AAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCG

TACTGTTGGTAAAGCCACC**ATG**GAAGAT…Luc

Substitution -251 C>G (rs3817621)

CACCGCACCTGGCCTTGTTTTTGTTTGTGTCACCCAGGTTGATCTGCCAGGAATAAAGATGG
CCACCCAACTTTTTTTTTTTTCCAGCGACAGAGTCTTGCTTTCTGTGCCCCAGGCTGGAC
TGCAGTGGTGTGATTATAGTTCAGTGCAGTCCCGACCTCCTGGGCTCAGGCAATCCTCCC
ACCTCAGCCTCCTGAGTAGCTGGGACTATGGGCATGCACCACCACTCACAGCTTTTTGTT
GGAGAGATAGGTTTTATGTTGCCCAGGCTGGTCTTGAAATCCTGGTGTCAAGTGATCCTC
CTGCCTCAGTCTTCCAATGAGTGACTATGAGTTATGTGTCTGTAATACGTATCCATGTCC
CCTTCCCAGGCTCCCAGGGCACCAGAGTGGAGGTTCCTGTTGTAGAAACTCAGATCCTCT
CCTCATGTTGGGCAGAATCAAGGAGCAGCCAGGCCCAGAGCCAGGGCACTGGTCTCCTGC
AGGTCAGGTACTGCCTGTGGCTTGATCCAACGGTCCTATCCCACCCAGGAGGAGAGAGGG
TCACTTTTCCCTTGGCTGCCCCCATCGCACTAAAGCAGCTGGCACTGAACCAGCCTCCAT
GCAGTCCCATGCAGTGCCACCCAAGGGTCCCCAGTAGACAATGGTGGGCCAGTTGTCAGG
GGCTTCCTCCTGCTGCAGGGCTGAGACCCTGGGAGGTCCCAACCCAGGCAAATTGAACGC
CAGGCTAATTTGAAGACCCAACTCCCAGCCCTCCCCTTCACCGGAGGACAGAGCTCTAGC
TGGCCTGGGCCCCCACCTGATAGCAGCCTCCAACGTCTGGGGTGTCTGATAATGCTTGGC
GGGGAGCTCGTGCCAAGTCCCGCCATCAGCACGGTTGTTGCTGTTTACTGGGGAGGGGGA
GGGCTGTGGAGCCTCAATCAGGGGGACAGGGGGTCCCACAGCTTCTTCCCAGAATACCCT
TTCTGCCTTTTCCAGGAAAGTTAACTGAGGGAAGACCCCCAAGTCTCTCCTTCTTTGGAG
ACCCAATGTCTGTTTTTACCCAGCACCTGGACCCTCAAACCCTGAACCCC**G**CAACCCTTG
ATATTTGACTTGGCTTTGGACACAGGGTTAGTCTTTAACCCCAGCCCCAGACAGGCCAAC
GTGAAGTTTGTGCCCCAGAAACAGTGCCCCCCCGCCGCCTTGCCTTGCTTTGCCTTATCA
GAGGCTGCAGCCAATCAGCTAAGGACAGAGAGGAGCCCTCGAAGGGGCTATCACAGCCTC
AGAGTTCACGAGGCAGCCGA**GATATC**AAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCG
TACTGTTGGTAAAGCCACC**ATG**GAAGAT…Luc

Substitutions -1133C>A & -251C>G (rs112943513 & rs3817621)

CACCGCACCTGGCCTTGTTTTTGTTTGTGTCACCCAGGTTGATCTGCCAGGAATAAAGATGG

CCACCCAACTTTTTTTTTTTTCCAGCGACAGAGTCTTGCTTTCTGTGCCCCAGGCTGGAC

TGCAGTGGTGTGATTATAGTTCAGTGCAGTCCCGACCTCCTGGGCTCAGGCAATCCTCCC

ACCTCAGC**A**TCCTGAGTAGCTGGGACTATGGGCATGCACCACCACTCACAGCTTTTTGTT

GGAGAGATAGGTTTTATGTTGCCCAGGCTGGTCTTGAAATCCTGGTGTCAAGTGATCCTC

CTGCCTCAGTCTTCCAATGAGTGACTATGAGTTATGTGTCTGTAATACGTATCCATGTCC

CCTTCCCAGGCTCCCAGGGCACCAGAGTGGAGGTTCCTGTTGTAGAAACTCAGATCCTCT

CCTCATGTTGGGCAGAATCAAGGAGCAGCCAGGCCCAGAGCCAGGGCACTGGTCTCCTGC

AGGTCAGGTACTGCCTGTGGCTTGATCCAACGGTCCTATCCCACCCAGGAGGAGAGAGGG

TCACTTTTCCCTTGGCTGCCCCCATCGCACTAAAGCAGCTGGCACTGAACCAGCCTCCAT

GCAGTCCCATGCAGTGCCACCCAAGGGTCCCCAGTAGACAATGGTGGGCCAGTTGTCAGG

GGCTTCCTCCTGCTGCAGGGCTGAGACCCTGGGAGGTCCCAACCCAGGCAAATTGAACGC

CAGGCTAATTTGAAGACCCAACTCCCAGCCCTCCCCTTCACCGGAGGACAGAGCTCTAGC

TGGCCTGGGCCCCCACCTGATAGCAGCCTCCAACGTCTGGGGTGTCTGATAATGCTTGGC

GGGGAGCTCGTGCCAAGTCCCGCCATCAGCACGGTTGTTGCTGTTTACTGGGGAGGGGGA

GGGCTGTGGAGCCTCAATCAGGGGGACAGGGGGTCCCACAGCTTCTTCCCAGAATACCCT

TTCTGCCTTTTCCAGGAAAGTTAACTGAGGGAAGACCCCCAAGTCTCTCCTTCTTTGGAG

ACCCAATGTCTGTTTTTACCCAGCACCTGGACCCTCAAACCCTGAACCCC**G**CAACCCTTG

ATATTTGACTTGGCTTTGGACACAGGGTTAGTCTTTAACCCCAGCCCCAGACAGGCCAAC

GTGAAGTTTGTGCCCCAGAAACAGTGCCCCCCCGCCGCCTTGCCTTGCTTTGCCTTATCA

GAGGCTGCAGCCAATCAGCTAAGGACAGAGAGGAGCCCTCGAAGGGGCTATCACAGCCTC

AGAGTTCACGAGGCAGCCGA**GATATC**AAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCG

TACTGTTGGTAAAGCCACC**ATG**GAAGAT…Luc

Substitution -148G>A (rs79334031)

CACCGCACCTGGCCTTGTTTTTGTTTGTGTCACCCAGGTTGATCTGCCAGGAATAAAGATGG

CCACCCAACTTTTTTTTTTTTCCAGCGACAGAGTCTTGCTTTCTGTGCCCCAGGCTGGAC

TGCAGTGGTGTGATTATAGTTCAGTGCAGTCCCGACCTCCTGGGCTCAGGCAATCCTCCC

ACCTCAGCCTCCTGAGTAGCTGGGACTATGGGCATGCACCACCACTCACAGCTTTTTGTT

GGAGAGATAGGTTTTATGTTGCCCAGGCTGGTCTTGAAATCCTGGTGTCAAGTGATCCTC

CTGCCTCAGTCTTCCAATGAGTGACTATGAGTTATGTGTCTGTAATACGTATCCATGTCC

CCTTCCCAGGCTCCCAGGGCACCAGAGTGGAGGTTCCTGTTGTAGAAACTCAGATCCTCT

CCTCATGTTGGGCAGAATCAAGGAGCAGCCAGGCCCAGAGCCAGGGCACTGGTCTCCTGC

AGGTCAGGTACTGCCTGTGGCTTGATCCAACGGTCCTATCCCACCCAGGAGGAGAGAGGG

TCACTTTTCCCTTGGCTGCCCCCATCGCACTAAAGCAGCTGGCACTGAACCAGCCTCCAT

GCAGTCCCATGCAGTGCCACCCAAGGGTCCCCAGTAGACAATGGTGGGCCAGTTGTCAGG

GGCTTCCTCCTGCTGCAGGGCTGAGACCCTGGGAGGTCCCAACCCAGGCAAATTGAACGC

CAGGCTAATTTGAAGACCCAACTCCCAGCCCTCCCCTTCACCGGAGGACAGAGCTCTAGC

TGGCCTGGGCCCCCACCTGATAGCAGCCTCCAACGTCTGGGGTGTCTGATAATGCTTGGC

GGGGAGCTCGTGCCAAGTCCCGCCATCAGCACGGTTGTTGCTGTTTACTGGGGAGGGGGA

GGGCTGTGGAGCCTCAATCAGGGGGACAGGGGGTCCCACAGCTTCTTCCCAGAATACCCT

TTCTGCCTTTTCCAGGAAAGTTAACTGAGGGAAGACCCCCAAGTCTCTCCTTCTTTGGAG

ACCCAATGTCTGTTTTTACCCAGCACCTGGACCCTCAAACCCTGAACCCCCCAACCCTTG

ATATTTGACTTGGCTTTGGACACAGGGTTAGTCTTTAACCCCAGCCCCAGACAGGCCAAC

GTGAAGTTTGTGCCCCAGAAACAGTGCCCCCCCC**A**CCGCCTTGCCTTGCTTTGCCTTATCA

GAGGCTGCAGCCAATCAGCTAAGGACAGAGAGGAGCCCTCGAAGGGGCTATCACAGCCTC

AGAGTTCACGAGGCAGCCGA**GATATC**AAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCG

TACTGTTGGTAAAGCCACC**ATG**GAAGAT…Luc

Substitution -154C>T (rs372651309)

CACCGCACCTGGCCTTGTTTTTGTTTGTGTCACCCAGGTTGATCTGCCAGGAATAAAGATGG

CCACCCAACTTTTTTTTTTTTCCAGCGACAGAGTCTTGCTTTCTGTGCCCCAGGCTGGAC

TGCAGTGGTGTGATTATAGTTCAGTGCAGTCCCGACCTCCTGGGCTCAGGCAATCCTCCC

ACCTCAGCCTCCTGAGTAGCTGGGACTATGGGCATGCACCACCACTCACAGCTTTTTGTT

GGAGAGATAGGTTTTATGTTGCCCAGGCTGGTCTTGAAATCCTGGTGTCAAGTGATCCTC

CTGCCTCAGTCTTCCAATGAGTGACTATGAGTTATGTGTCTGTAATACGTATCCATGTCC

CCTTCCCAGGCTCCCAGGGCACCAGAGTGGAGGTTCCTGTTGTAGAAACTCAGATCCTCT

CCTCATGTTGGGCAGAATCAAGGAGCAGCCAGGCCCAGAGCCAGGGCACTGGTCTCCTGC

AGGTCAGGTACTGCCTGTGGCTTGATCCAACGGTCCTATCCCACCCAGGAGGAGAGAGGG

TCACTTTTCCCTTGGCTGCCCCCATCGCACTAAAGCAGCTGGCACTGAACCAGCCTCCAT

GCAGTCCCATGCAGTGCCACCCAAGGGTCCCCAGTAGACAATGGTGGGCCAGTTGTCAGG

GGCTTCCTCCTGCTGCAGGGCTGAGACCCTGGGAGGTCCCAACCCAGGCAAATTGAACGC

CAGGCTAATTTGAAGACCCAACTCCCAGCCCTCCCCTTCACCGGAGGACAGAGCTCTAGC

TGGCCTGGGCCCCCACCTGATAGCAGCCTCCAACGTCTGGGGTGTCTGATAATGCTTGGC

GGGGAGCTCGTGCCAAGTCCCGCCATCAGCACGGTTGTTGCTGTTTACTGGGGAGGGGGA

GGGCTGTGGAGCCTCAATCAGGGGGACAGGGGGTCCCACAGCTTCTTCCCAGAATACCCT

TTCTGCCTTTTCCAGGAAAGTTAACTGAGGGAAGACCCCCAAGTCTCTCCTTCTTTGGAG

ACCCAATGTCTGTTTTTACCCAGCACCTGGACCCTCAAACCCTGAACCCCCCAACCCTTG

ATATTTGACTTGGCTTTGGACACAGGGTTAGTCTTTAACCCCAGCCCCAGACAGGCCAAC

GTGAAGTTTGTGCCCCAGAAACAGTGC**T**CCCCCGCCGCCTTGCCTTGCTTTGCCTTATCA

GAGGCTGCAGCCAATCAGCTAAGGACAGAGAGGAGCCCTCGAAGGGGCTATCACAGCCTC

AGAGTTCACGAGGCAGCCGA**GATATC**AAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCG

TACTGTTGGTAAAGCCACC**ATG**GAAGAT…Luc

Substitution -43G>A (rs372359976)

CACCGCACCTGGCCTTGTTTTTGTTTGTGTCACCCAGGTTGATCTGCCAGGAATAAAGATGG

CCACCCAACTTTTTTTTTTTTCCAGCGACAGAGTCTTGCTTTCTGTGCCCCAGGCTGGAC

TGCAGTGGTGTGATTATAGTTCAGTGCAGTCCCGACCTCCTGGGCTCAGGCAATCCTCCC

ACCTCAGCCTCCTGAGTAGCTGGGACTATGGGCATGCACCACCACTCACAGCTTTTTGTT

GGAGAGATAGGTTTTATGTTGCCCAGGCTGGTCTTGAAATCCTGGTGTCAAGTGATCCTC

CTGCCTCAGTCTTCCAATGAGTGACTATGAGTTATGTGTCTGTAATACGTATCCATGTCC

CCTTCCCAGGCTCCCAGGGCACCAGAGTGGAGGTTCCTGTTGTAGAAACTCAGATCCTCT

CCTCATGTTGGGCAGAATCAAGGAGCAGCCAGGCCCAGAGCCAGGGCACTGGTCTCCTGC

AGGTCAGGTACTGCCTGTGGCTTGATCCAACGGTCCTATCCCACCCAGGAGGAGAGAGGG

TCACTTTTCCCTTGGCTGCCCCCATCGCACTAAAGCAGCTGGCACTGAACCAGCCTCCAT

GCAGTCCCATGCAGTGCCACCCAAGGGTCCCCAGTAGACAATGGTGGGCCAGTTGTCAGG

GGCTTCCTCCTGCTGCAGGGCTGAGACCCTGGGAGGTCCCAACCCAGGCAAATTGAACGC

CAGGCTAATTTGAAGACCCAACTCCCAGCCCTCCCCTTCACCGGAGGACAGAGCTCTAGC

TGGCCTGGGCCCCCACCTGATAGCAGCCTCCAACGTCTGGGGTGTCTGATAATGCTTGGC

GGGGAGCTCGTGCCAAGTCCCGCCATCAGCACGGTTGTTGCTGTTTACTGGGGAGGGGGA

GGGCTGTGGAGCCTCAATCAGGGGGACAGGGGGTCCCACAGCTTCTTCCCAGAATACCCT

TTCTGCCTTTTCCAGGAAAGTTAACTGAGGGAAGACCCCCAAGTCTCTCCTTCTTTGGAG

ACCCAATGTCTGTTTTTACCCAGCACCTGGACCCTCAAACCCTGAACCCCCCAACCCTTG

ATATTTGACTTGGCTTTGGACACAGGGTTAGTCTTTAACCCCAGCCCCAGACAGGCCAAC

GTGAAGTTTGTGCCCCAGAAACAGTGCCCCCCCGCCGCCTTGCCTTGCTTTGCCTTATCA

GAGGCTGCAGCCAATCAGCTAAGGACAGAGAGGAGCCCTCGAAGGGGCTATCACAGCCTC

AGAGTTCACGAGGCAGCC**A**A**GATATC**AAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCG

TACTGTTGGTAAAGCCACC**ATG**GAAGAT…Luc

Substitution -102T>G (rs548543206)

CACCGCACCTGGCCTTGTTTTTGTTTGTGTCACCCAGGTTGATCTGCCAGGAATAAAGATGG
CCACCCAACTTTTTTTTTTTTCCAGCGACAGAGTCTTGCTTTCTGTGCCCCAGGCTGGAC
TGCAGTGGTGTGATTATAGTTCAGTGCAGTCCCGACCTCCTGGGCTCAGGCAATCCTCCC
ACCTCAGCCTCCTGAGTAGCTGGGACTATGGGCATGCACCACCACTCACAGCTTTTTGTT
GGAGAGATAGGTTTTATGTTGCCCAGGCTGGTCTTGAAATCCTGGTGTCAAGTGATCCTC
CTGCCTCAGTCTTCCAATGAGTGACTATGAGTTATGTGTCTGTAATACGTATCCATGTCC
CCTTCCCAGGCTCCCAGGGCACCAGAGTGGAGGTTCCTGTTGTAGAAACTCAGATCCTCT
CCTCATGTTGGGCAGAATCAAGGAGCAGCCAGGCCCAGAGCCAGGGCACTGGTCTCCTGC
AGGTCAGGTACTGCCTGTGGCTTGATCCAACGGTCCTATCCCACCCAGGAGGAGAGAGGG
TCACTTTTCCCTTGGCTGCCCCCATCGCACTAAAGCAGCTGGCACTGAACCAGCCTCCAT
GCAGTCCCATGCAGTGCCACCCAAGGGTCCCCAGTAGACAATGGTGGGCCAGTTGTCAGG
GGCTTCCTCCTGCTGCAGGGCTGAGACCCTGGGAGGTCCCAACCCAGGCAAATTGAACGC
CAGGCTAATTTGAAGACCCAACTCCCAGCCCTCCCCTTCACCGGAGGACAGAGCTCTAGC
TGGCCTGGGCCCCCACCTGATAGCAGCCTCCAACGTCTGGGGTGTCTGATAATGCTTGGC
GGGGAGCTCGTGCCAAGTCCCGCCATCAGCACGGTTGTTGCTGTTTACTGGGGAGGGGGA
GGGCTGTGGAGCCTCAATCAGGGGGACAGGGGGTCCCACAGCTTCTTCCCAGAATACCCT
TTCTGCCTTTTCCAGGAAAGTTAACTGAGGGAAGACCCCCAAGTCTCTCCTTCTTTGGAG
ACCCAATGTCTGTTTTTACCCAGCACCTGGACCCTCAAACCCTGAACCCCCCAACCCTTG
ATATTTGACTTGGCTTTGGACACAGGGTTAGTCTTTAACCCCAGCCCCAGACAGGCCAAC
GTGAAGTTTGTGCCCCAGAAACAGTGCCCCCCCGCCGCCTTGCCTTGCTTTGCCTTATCA
GAGGCTGCAGCCAATCAGC**G**AAGGACAGAGAGGAGCCCTCGAAGGGGCTATCACAGCCTC
AGAGTTCACGAGGCAGCCGA**GATATC**AAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCG
TACTGTTGGTAAAGCCACC**ATG**GAAGAT…Luc

Appendix D

Preparation of competent DH5α *E.coli* cells

The DH5α *E.coli* bacterial cells (Invitrogen, UK) were used as competent cells for the introduction of the circular plasmids.

Preparation of solutions

i. LB broth

Twenty grams (20g) of LB broth powder (Sigma-Aldrich, USA) was dissolved in 1000ml of deionized water. The solution was then transferred to 500ml autoclavable glass bottles autoclaved and the sterile broth was then stored at room temperature.

ii. LB agar

Thirthy-five grams (35g) of LB Agar poweder (Sigma-Aldrich, USA) was dissolved in 1000ml of deionized water. The solution was then transferred to 500ml autoclavable glass bottles autoclaved and the sterile broth was then stored at room temperature.

iii. LB agar plates

The sterile LB agar prepared in (ii) above was melted completely in a microwave oven at a low power and allowed to cool below 50°C. Using sterile conditions by working closely to a Bunsen burner the LB agar was poured into sterile 85mm petri dishes. The petri-dishes were then covered immediately at allowed to set and stored for up to a month at 4°C.

iv. 100nM Magnesium Chloride (MgCl2) solution

The role of MgCl2 is to induce the ability of the cells to take up DNA by altering the permeability of the membranes. The negatively charged incoming DNA is repelled by the negatively charged macromolecues present on the bacteria outer surface. The addition of MgCl2 will neutralize these unfavourable interactions.

0.952g of anhydrous MgCL2 was dissolved in 100mL deionized water and the solution was autoclaved.

v. 100M Calcium Chloride (CaCL2) solution

CaCl2 increases the bacterial cell's ability to incorporate plasmid DNA and therefore facilitating genetic transformation. The addition of CaCl2 to cell suspension will allow the binding of the plasmid DNA to lipopolysaccharide (LPS) allowing the negatively charged DNA backbone and the LPS to come together during heat shock transformation.

1.11g of anhydrous CaCl2 was dissolved in 100ml of deionized water and the solution was autoclaved

vi. 100nM Calcium chloride with 15% glycerol

The role of the glycerol is to help in shielding the negative charges present on the DNA and host cell membrane.

1.11g of anhydrous CaCl2 was dissolved in 80ml of deionized water. 15ml of glycerol was added and the solution was mixed vigorously. The solution was topped up to 100ml with deionized water, shacked again and autoclaved.

Appendix E

Solutions for EMSA

1 Molar TRIS buffer pH6.8

- 12.11g TRIS (Sigma-Aldrich, UK)
- Top up to 100mL with deionised water
- Adjust to pH6.8 with a calibrated pH meter

Label clearly with content and date of production and store in fridge

1Molar TRIS buffer pH8.8

- 12.11g TRIS (Sigma-Aldrich, UK)
- Top up to 100mL with deionised water
- Adjust to pH8.8 with a calibrated pH meter

Label clearly with content and date of production and store in fridge

10% SDS

- 10g of SDS (Sigma-Aldrich, UK)
- 100mL deionised water

Extreme care should be taken as SDS in powder is extremely toxic

10% APS

- 0.1g of APS (Sigma-Aldrich, UK) directly in a 1.5mL Eppendorf tube
- 1mL of deionised water, mixing lightly until the powder dissolves

10% APS should only be prepared few minutes prior to use, as its chemical composition breaks down easily

10X running buffer

- 30.3g of Tris (Sigma-Aldrich, UK)
- 144.2g of glycine (Sigma-Aldrich, UK)
- 10g SDS (Sigma-Aldrich, UK)
- Top up to 1000mL deionised water
- Adjust to pH8.8 using a calibrated pH meter

1X running buffer

- 100mL of 10X running buffer
- 900mL of deionised water

Appendix F

p.K288X heterozygote vs p.K288X wildtype T0	p.K288X heterozygote vs p.K288X wildtype T0	p.K288X heterozygotes Medium HbF vs High HbF T48	p.K288X heterozygotes Medium HbF vs High HbF T48
downregulated	**upregulated**	**downregulated**	**upregulated**
BCAM	HBE1	TRIB3	RPS4Y1
CTSE	G0S2	PHGDH	EIF1AY
C10orf10	HBZ	PPP1R3B	KDM5D
ARG1	CELA2B	HIST1H4C	DDX3Y
MIR144	SMPX	HMOX1	DUSP2
ABCG2	CA3	ARG2	TPSG1
HIST1H4C	ESPN	HPGDS	SNORD47
PKDCC	LDLRAD3	C10orf10	NCF4
LRP1	CSF2RB	PITX1	CST7
IFI27	SLC6A6	TMCC2	LPAR5
TMCC2	SLC6A19	IFIT1B	PDIA5
MTRNR2L6	PLXNC1	FAM129A	HHEX
ATP1B2	IGFBP2	CEBPB	FABP5
IFIT1B	SLC37A2	DDIT3	SLCO4A1
IGFBP4	ARL15	TCP11L2	SNORD80
PHGDH	ENPP3	CTH	PTGES
RUNDC3A	TUBB1	TSC22D3	PLXND1
TSPO2	CREB3L1	ULBP1	HPGD
GLIPR2	ACTN1	OR2W3	CD320
PLCL2	ID2	SESN3	C12orf45
SNX22	PSTPIP2	PNRC1	HPDL
HBA2	HBG2	SESN2	TPSD1
BSPRY	TNFRSF12A	CRYBA4	CORO1B
CRIP1	CD69	TMSB4X	FAM195A
SQRDL	TAC3	ATF5	NT5C3B
EIF1AY	TNS3	DDIT4	FAM213A

HBB	SAMSN1	VAMP5	ACTN1
TCP11L2	LHFPL2	GDF15	FAM129B
SPHK1	HBBP1	RUNDC3A	CSF2RB
OPTN	SLC12A6	GABARAPL1	KAT2A
C17orf99	CMIP	S100A10	TPSB2
AHSP	HBG1	PKD1	LY6E
RPS4Y1	AK4	SEL1L3	GCSAML
ACSL6	B3GNT5	ABTB1	RRP9
HBA1	ELMO1	FBXO30	EARS2
DARC	KIT	TRIM58	POP1
MT1E	TBC1D14	SLC6A19	TPSAB1
TMEM86B	IQGAP2	BCL6	PYCARD
HBM	PIBF1	NUDT4	C1orf186
ZDHHC14	TRIM29	MTRNR2L6	IGFBP2
C1QA	RYK	CREBRF	ETS2
GNG2	KCNK6	R3HDM4	GNA15
PIGQ	DHTKD1	UBE2H	MAGEF1
PRDX2	CA1	CEBPG	CALB2
GSTM1	TPSD1	TCEANC	SMYD2
SYNGR1	VWA5A	YPEL3	NOP16
ART4	ZNF175	ULK1	PUS7
MOSPD3	NCKAP1	SEPP1	LDLRAD3
CD70	MYC	PNPLA2	NME1
PITX1	NAB2	CLN8	MYBBP1A
IMPA2	RELN	ZNF746	TMEM173
A4GALT	CBLL1	SLC2A4	ANXA1
BACE2	ANXA4	CCDC71L	LOC100128881
SNCA	ADCY7	PLCL2	EPCAM
MYL4	SEMA7A	MBNL2	IFITM2
GABARAPL1	KIFC3	MAST3	GDF11
TGM2	TMEM164	YPEL2	WDR77

CDKN2D	SPRED2	C1QB	SNORD36A
MGST3	EGR1	UBAP1	GYG1
PHOSPHO1	LCP2	C17orf103	HMGN3
ALAD	GAS5	FOXO3	AKR1C1
RASSF7	BEX1	ARRDC2	SCAMP3
ID1	PROS1	RAP1GAP	ALDH1B1
GSTP1	TGFBRAP1	CHAC1	ERO1L
DHRS13	GALNT5	TUBB2A	ICAM2
HES6	C11orf21	APP	MARVELD1
E2F2	IFITM2	PIK3IP1	FAM136A
AIFM2	TGIF1	GPCPD1	ORAI1
KIAA0930	APOE	KLF9	CISD3
FOSB	VEGFA	FAM46C	SRM
C14orf80	ATP2A2	SLC16A1	SMYD5
GSTM3	GATA2	RB1CC1	NLRP2
FAM173A	PI4K2A	CNN3	NOP2
SNORD68	FADS3	CREB3L1	NMRAL1
OTUD1	PCNXL4	IRS2	MGST1
ITSN1	ESF1	BBC3	PDIA4
NDRG1	PFKP	ASNS	POGK
ISG20	MYADM	HECA	AKAP1
SCPEP1	ATP2A3	MAN1A1	LITAF
AK1	GTF2I	RIOK3	GLRX
YPEL3	KRCC1	ARG1	MRM1
SLC14A1	RPL7L1	PKD1P1	SLC10A4
TMEM160	LPIN1	ACP5	BID
BCL6	MGAT3	ID2	TMEM177
SLC25A37	HLTF	TNNI3	PEBP1
BCYRN1	ACTN4	ATF3	LYN
ACP5	C2orf74	IL6ST	EBPL
HDHD3	TRIM5	CDC42EP1	SLC25A23

ARRDC3	CYB561	VCPIP1	SEMA4C
SNORA57	DIP2B	TBCEL	PPM1F
S100A4	TJP2	MICAL2	METTL1
PRR24	RALB	PSAT1	PPP1R14B
FAM214B	IPO11	ARL4A	SLC19A1
SAC3D1	FAM60A	TESPA1	SNORD22
BPGM	ADAM10	MEGF9	KIT
BCL11A	PRMT3	ISG20	DDX10
BNIP3	PDZD8	BCL2L1	PDCL3
GCHFR	GPD1L	RABAC1	HEBP2
MIR22HG	PIP5K1A	MIR22HG	MLC1
ANXA5	ATN1	C3orf58	CCDC86

First hundred upregulated and downregulated genes between (i) subjects wildtype for the p.K288X mutation and subjects heterozygotes for the p.K288X mutation (ii) subjects heterozygotes for the p.K288X mutation and medium HbF levels and subjects heterozygotes for the p.K288x mutation and high HbF levels.