UNIVERSITÀ DEGLI STUDI DI SASSARI

PhD School in
Biomolecular and Biotechnological Sciences
Curriculum: Biochemistry and Molecular Biology
Director: Prof. Claudia Crosio
"XXVI Ciclo"

Analysis of genetic determinants associated with Persistent Synthesis of Fetal Hemoglobin

Supervisor: **Monica Pirastru, PhD**

Director: **Prof. Claudia Crosio**

PhD Student: **Sandro Trova**

uniss
UNIVERSITÀ DEGLI STUDI DI SASSARI

Fondazione
Banco di Sardegna

INDEX

ABSTRACT **3**

INTRODUCTION **4**

1. Hemoglobin **4**

1.1 Structure and function of Hemoglobin 4
1.2 Structure of globin genes and their cluster organization 5
1.3 Genomic context of the α– and β–globin gene clusters 9

2. Globin gene switching **12**

2.1 Regulatory regions and transcription factors of globin genes 13
2.2 The β–Globin Locus Control Region (β–LCR) role in globin expression 20
2.3 Chromatin role in β–like globin gene expression: the PYR role 25
2.4 Summary on the fetal to adult switch 27

3. Classification of the Disorders of Hemoglobin **29**

3.1 The Thalassemia syndromes 29
3.1.1 α–thalassemias 30
3.1.2 β–thalassemias 33
3.2 Pathophysiology and phenotypes of β–thalassemia 36

4. Persistence of Fetal Hemoglobin **39**

4.1 dHPFH and δβ–thalassemias (HBFQTL1) 42
4.2 Non–Deletional Hereditary Persistence of Fetal Hb (ndHPFH) (HBFQTL1) 46
4.2.1 $^{A}\gamma$ −117 G→A (rs35378915) 48
4.2.2 $^{A}\gamma$ −196 C→T (rs35983258) 49
4.2.3 $^{G}\gamma$ −175 T→C (rs63750654) 51
4.2.4 $^{G}\gamma$ −158 C→T (rs7482144) 51
4.3 Modern approaches for the identification of genetic modifiers of β–thalassemia phenotype 52
4.4 HBFQTL2: Chromosome 6–linked 53
4.5 HBFQTL3: Chromosome X–linked 54
4.6 HBFQTL4: Chromosome 8–linked 54
4.7 HBFQTL5: Chromosome 2–linked and Role of BCL11A 55
4.8 HBFQTL6: Chromosome 19–linked and Role of KLF1 56

AIM OF THIS STUDY **58**

MATERIALS AND METHODS 60

5. Choice of the Samples 60

6. Preparation of the samples 60

6.1 Preparation of Hemolysates 60
6.2 DNA extraction and quantification 61

7. Biochemical analysis of the samples 61

7.1 CE–HPLC 61
7.2 IEF 62
7.3 Preparative IEF: HbF enrichment 62
7.4 AUT–PAGE 63
7.5 RP–HPLC 64

8. Molecular analysis of the samples 64

8.1 PCR 64
8.2 MLPA 68
8.3 GAP–PCR 69
8.4 PCR fragment purification and quantification 69
8.5 Sequence Reaction 69
8.6 Sequence Analysis 70

9. Expression Studies on γ promoters 70

9.1 Promoter isolation from genomic DNA 71
9.2 Cloning of PCR fragments 72
9.3 Reporter assay 75

RESULTS 77

DISCUSSION 84

TABLES 95

REFERENCES 110

PROTOCOLS 117

LIST OF ABBREVIATIONS 122

ABSTRACT

One of the principal ameliorating factors in β–thalassemias is the innate ability to produce HbF. Although the majority of adults shows traces of HbF, in rare cases, named Hereditary Persistence of Fetal Hemoglobin (HPFH), its value can reach levels of 15%–30%.

cis–acting variants like deletions, point mutations and SNPs within the β–cluster can explain some of the variability; however 50% of the variance is unlinked to the β–cluster.

In this study, healthy HPFH individuals were analyzed: Hb profile was defined by CE–HPLC, RP–HPLC, IEF and AUT–PAGE. Structural analysis of the γ genes promoters revealed 4 ndHPFH mutations, whose strength was studied by luciferase assays.

A large deletion on the β–cluster (Sicilian δβ–thalassemia) was found in one sample by MLPA and GAP–PCR.

One $^{A}\gamma$ ndHPFH sample, also showing $^{G}\gamma$ persistence, presented 2 new mutations on the $^{G}\gamma$ promoter: luciferase assays demonstrated that the persistence of $^{G}\gamma$ is not attributable to these mutations *per se*.

Complete sequence of β genes showed $\beta^{0}(39)$ and $\beta^{+}(-87)$ β–thalassemia alleles; $\beta^{+}(-87)$ was found linked to a *cis*–acting polymorphic configuration associated with high HbF.

Orkin haplotype analysis showed evidence of a unique association between $^{A}\gamma$ –117 mutation and haplotype VII.

The data presented in this thesis confirm the complexity of the globin switch and the heterogeneity of the molecular mechanisms underlying the persistence of HbF, emphasizing the importance of the molecular context of a mutation.

INTRODUCTION

1. HEMOGLOBIN

1.1 Structure and function of Hemoglobin

Hemoglobin, also spelled haemoglobin (Hb), is a heterotetrameric protein synthetized by red blood cells precursors. It is made up of four polypeptide chains, assembled in a quaternary structure in which two so called α–like chains (141 amino acid residues) associate with two β–like chains (146 amino acid).

During development, eight human globin genes (Para. 1.2) express six different Hb types (Fig. 1): Hb Gower I ($\zeta_2\varepsilon_2$), Gower II ($\alpha_2\varepsilon_2$), and Portland ($\zeta_2\gamma_2$) are found in the embryo; fetal Hb (HbF; $\alpha_2\gamma_2$) is present mainly in the fetus, but also in the embryo and adult, whereas HbA ($\alpha_2\beta_2$) and HbA_2 ($\alpha_2\delta_2$) are seen in adults (Chap. 2).
Also, all Hbs undergo post–translational modification forming minor Hb such as HbF_1 (acetylated form of HbF) or HbA_{1C} (non–enzymatic glicated form of HbA) (Steinberg *et al.*, 2009).

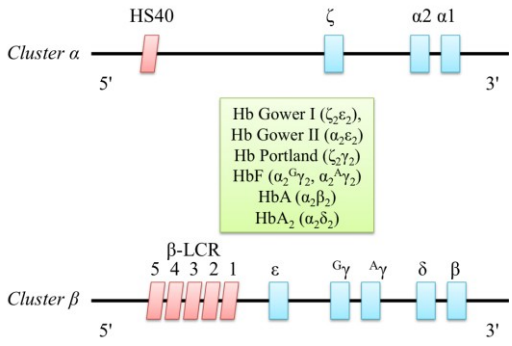


Figure 1
Basic organization of human globin gene clusters and composition of Hb produced.

Although the polypeptides have different sequences of amino acids, each protein chain arranges into a set of α–helix, folding up in a similar three–dimensional tertiary structure called "globin", comparable to the folding motif used in other proteins such as myoglobin. This particular folding creates a pocket for the prosthetic group heme (Fe–protoporphyrin IX); a polypeptide chain combined with a heme is called monomer or subunit of Hb. Heme allows the Hb in erythrocytes to bind O_2 reversibly and transport it from the lungs to respiring tissues (Steinberg *et al.*, 2009).

The folding is made up of seven or eight α–helix, named A to H starting from the amino end, while non–helical segments are named by the contiguous helix (e.g. AB, BC, …); residues within each segment are numbered from the amino end with an Arabic number (e.g. A1, AB1, and so on) (Watson and Kendrew, 1961).

Hb tetramer have two quaternary structures: the T or deoxy conformer that has low oxygen affinity and the R or oxy conformer that has high oxygen affinity. Further fine–tuning of Hb function comes from its allosteric behavior triggered by the binding of two small negative effector molecules: 2,3–BPG, and protons to specific sites on the T structure distant from the heme groups (Forget and Bunn, 2013).

The transport of O_2 differs in embryonic, fetal, and adult stages of development: early embryos obtain O_2 from the maternal interstitial fluid using embryonic Hbs; fetus obtains its O_2 via the placenta using fetal ones, while adult uses its adult Hbs.

O_2 affinity of HbF is higher than adult ones' due to its lower affinity for the allosteric effector 2,3–BPG; this allows HbF to take oxygen from maternal HbA in the placenta (Steinberg *et al.*, 2009).

1.2 Structure of globin genes and their cluster organization

The genes encoding the different globin chains of Hb are members of an ancient gene family and they share a common structural organization.

In humans, as in all vertebrate species, different α–like and β–like globin chains are synthetized at progressive stages of development: before precise knowledge of globin gene organization, it was clear that there must exist at least one gene for each of the different globin chain (ζ, α, ε, γ, β, δ). Further studies based on HbF heterogeneity and globin variants showed that α and γ genes are duplicated and structurally different: the two α genes encode the same protein, while γ genes are different only for one amino acid (136 Gly/Ala: $^{G}\gamma$ and $^{A}\gamma$, respectively).

The coding region of each globin gene is interrupted by two introns, also called intervening sequences (IVS1 and IVS2). In the β–like globin genes, the introns interrupt the coding sequence between codons 30 and 31 and between codons 104 and 105; in the α–globin gene family, IVS1 interrupt the coding sequence between codons 31 and 32 and IVS2 between codons 99 and 100 (Fig. 2 A).

The IVS1 is shorter than the IVS2 in both α– and β–globin genes, but IVS2 of the human β–globin gene is much longer than that of the α–globin gene.

Globin genes are transcribed into a pre–mRNA and IVSs are subsequently removed to produce the mature mRNA (**splicing**). This process requires different consensus sequences, which are almost universally found at the 5' (donor) and 3' (acceptor) splice sites of introns, and the consensus sequence surrounding the A (Adenosine) in the so called "branch of the introns". The dinucleotides GT (donor) and AG (acceptor) of the intron are essential and required for proper globin gene splicing (Steinberg *et al.*, 2009).

The mature mRNA is made up of a 5' untranslated region (UTR), which drives the translation, the coding region, a 3'UTR and the poly–A tail (Fig. 2 B).

.

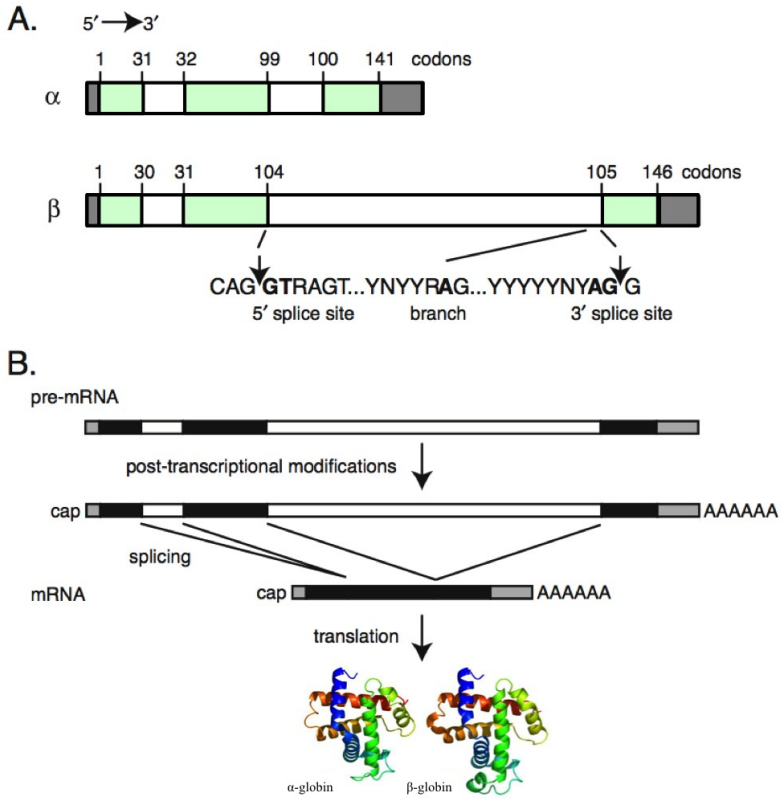


Figure 2
A) General structure of globin genes
B) The pathway for maturation and expression of mRNA of the globin genes
(Steinberg *et al.*, 2009)

α–like and β–like genes map on different chromosomes: the so called "β–globin cluster" maps on chromosome 11 (11p15.5), while "α–globin cluster" on chromosome 16 (16p13.3), 150 kb from the telomere of its short arm.

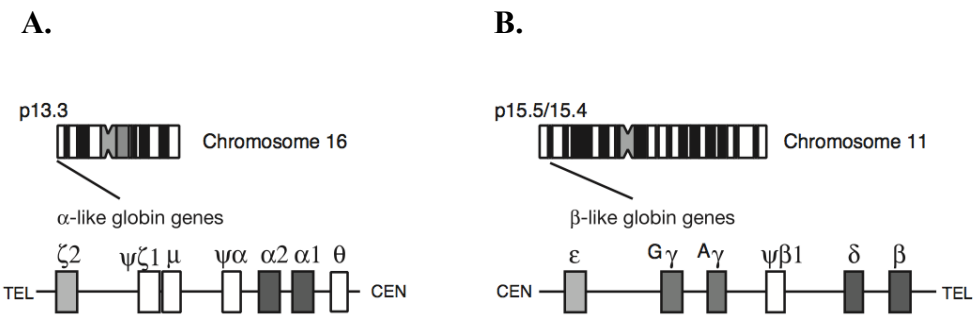


Figure 3
Organization of globin gene clusters. It is notable that the orientation with respect to the centromere (CEN) and telomere (TEL) are opposite.
A) α–globin cluster
B) β–globin cluster

(Steinberg *et al.*, 2009)

In each cluster the active genes are arranged in the same order as they are expressed during the development (expressed genes are indicated in **bold**):

- α–cluster (AC #: NG_000006). It spans about 30 kb and includes 4 functional genes and 3 pseudogenes. The order of genes is: 5'– **ζ2** (**HBZ**, embryonic) – ψζ1 (HBZP1 or HBZP) – μ or ψα2 (HBM or HBAP2) – ψα1 (HBAP1) – **α2** (HBA2, fetal and adult) – **α1** (HBA1, fetal and adult) – θ (HBQ1) – 3' (Fig. 3 A).

The α2 and α1 coding sequences are identical, however, the two genes differ slightly over the 5'–UTR and the introns, but significantly over the 3'–UTR.

Although the protein products of the α2– and α1–genes are identical, α2–mRNA in reticulocytes, both in in fetal and adult life, predominates over α1–mRNA by approximately 3:1, probably as a result of differences at the level of transcription of the two genes.

α2– and α1–mRNA have identical translational profiles and therefore should be produced with equal efficiencies, predicting a dominant role for the α2–globin locus. Two studies addressing this point come to different conclusions. In the first, it was found that α2 variants represented 24%–40% of total, whereas α1 variants

represented 11%–23%, suggesting a predominant role for the α2 gene at both mRNA and protein levels. The second study found that in heterozygotes the average proportion of stable variants resulting from α2 mutants (23.5%) was only slightly higher than from α1 mutations (19.7%), suggesting a less efficient translation of the α2–mRNA and a more equal contribution from the two genes at the protein level (Steinberg *et al.*, 2009).

The cluster shows three pseudo–globin genes: however in a recent work a transcript for a μ–globin was identified: it encodes for a 141 amino acid protein but generates non detectable μ–globin protein in human (Goh *et al.*, 2005). Its closest relatives are the α^{D}–globin genes actively expressed in red cells of reptiles and birds (Steinberg *et al.*, 2009). It is interesting that μ–globin gene is not the only one lacking a detectable product: the θ–globin gene is also transcribed, but not translated at least at detectable levels (Hsu *et al.*, 1988). Both genes are well conserved at the genomic level with appropriate splicing junctions and maintenance of open reading frames (ORFs). Like μ–globin, θ–globin has a highly regulated pattern of transcription in erythroid cells (Albitar *et al.*, 1992). In addition, both genes demonstrate only fractional levels of transcription compared with the dominant α genes, and their deletion in humans seems to not affect the clinical phenotype (Fei *et al.*, 1988).

- β–cluster: (AC #: NG_000007 and AC #: U01317). It spans about 45 kb and includes 5 functional genes and 1 pseudogene. The order of genes is: 5' – **ε** (**HBE**, embryonic) – ^{G}**γ** (**HBG2**, fetal) – ^{A}**γ** (**HBG1**, fetal) – ψβ1 (HBBP1) – **δ** (**HBD**, minor adult) – **β** (**HBB**, major adult) – 3' (Fig. 3 B).

HBG1 gene exist in two alleles coding for two globin that differ at codon 75: the mostly represented $^{A}\gamma^{I}$ (AC #: U01317, with Ile at codon 75) and the variant $^{A}\gamma^{T}$ (AC #: NG_000007, with Thr at cod. 75). The tetramer $\alpha_2{}^{A}\gamma^{T}{}_2$ is called HbF–Sardinia due to its high frequency in Sardinia (30% circa) (Manca and Masala, 2008).

The pseudogene ψβ1 is transcribed and spliced but the presence of mutations render it incapable of encoding a functional globin chain.

These β–like genes have the same overall exonic structure, leading to the conclusion that they are derived from one ancestral gene.

Note that the orientation with respect to the centromere (CEN) and telomere (TEL) are opposite: the α–like globin genes are transcribed toward CEN, whereas the β–like globin genes are transcribed toward TEL (Fig. 3 A and B) (Steinberg *et al.*, 2009).

1.3 Genomic context of the α– and β–globin gene clusters

The separation of α– and β–globin gene clusters to different chromosomes has allowed them to diverge into strikingly different genomic contexts, with consequential enigmatic understanding of their regulation. The β–globin gene cluster is A+T rich, with no CpG islands[1] (Collins and Weissman, 1984), while the α–like globin gene cluster is highly G+C rich, with multiple CpG islands (Fischel–Ghodsian *et al.*, 1987).

This brings to important differences in the structure and regulation of the two gene clusters. The β–globin gene cluster is subject to tissue–specific DNA methylation, while, due to the presence of CpG islands, the α–globin gene cluster is not methylated in any cell type.

β–like globin cluster have an increased accessibility of the chromatin only in expressing cells; in contrast, the α–like globin genes are in constitutively open chromatin (Steinberg *et al.*, 2009).

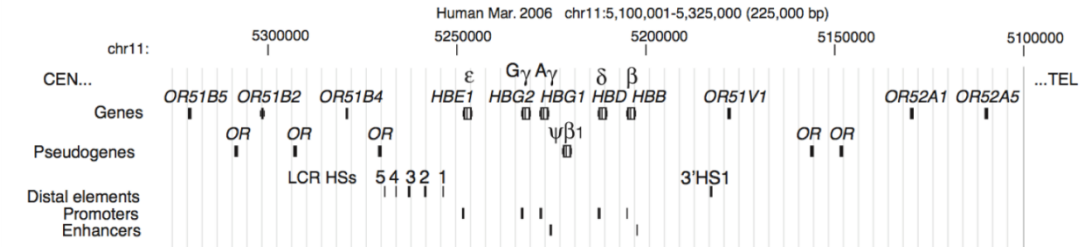
The types of genes that surround the α–like and β–like globin gene clusters are quite different (Fig. 4).

[1] The CpG islands are short regions (a few hundred base pairs) in which the CpG dinucleotides are not methylated; these have been associated with important functions such as promoters for transcription. Regions that are G+C rich tend to be enriched in genes expressed in an ample range of tissues; these genes also tend to have islands with an abundance of the dinucleotide CpG, in contrast to the rest of the genome, which has very few CpGs because these are the sites for DNA methylation when not organized in islands.

The **β–like** globin gene cluster is surrounded by olfactory receptor (OR) genes; OR gene cluster contain approximately 100 genes extending almost 1 million bp (Mb). Perhaps this has had an impact on natural selection for a powerful enhancer that could override the OR genes regulation in a chromosomal region devoted to olfactory–specific expression (Fig. 4 A).

The **α–like** globin genes are surrounded by widely expressed genes (Fig. 4 B), such as MPG (DNA repair enzyme methyl purine glycosylase) and POLR3K (subunit of RNA polymerase III).

A.



B.

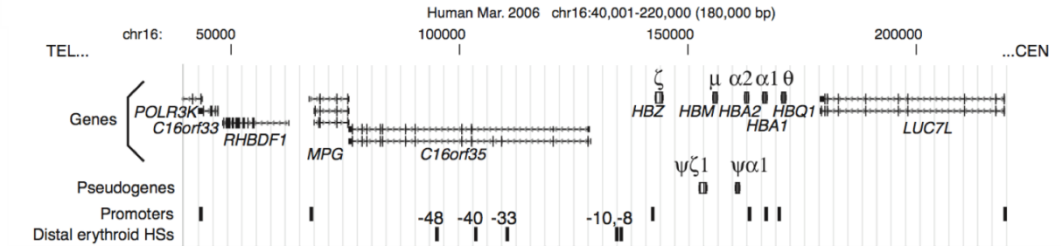


Figure 4
Genomic context of α– and β–cluster.

A) β–globin gene cluster
B) α–globin gene cluster

(Steinberg *et al.*, 2009)

Although the α–like globin gene cluster and surrounding DNA is in constitutively open chromatin, its histones are hyperacetylated (active chromatin) only in erythroid cells.

It is possible that this region encompasses all the sequences needed *in cis* for appropriate regulation of the α–like globin genes: in fact the regions surrounding the α–like globin gene cluster have undergone inter– and intrachromosomal rearrangements in various vertebrate lineages, while the genes from POLR3K through HBQ1 (θ gene) have remained together in all species.

Despite these many differences between the two clusters, the genes are expressed coordinately between the two loci, resulting in balanced production of α–like and β–like globins needed for the synthesis of Hbs, even though the mechanisms is still not clear (Steinberg *et al.*, 2009).

2. GLOBIN GENE SWITCHING

Hb switching is characteristic of all animal species that use Hb for O_2 transport. Most species have the embryonic to adult globin switch, while humans and a few other mammals have two globin gene switches, from embryonic to fetal globin and from fetal to adult globin formation, occurring around the perinatal period (Fig. 5; see Para. 1.1 and 1.2).

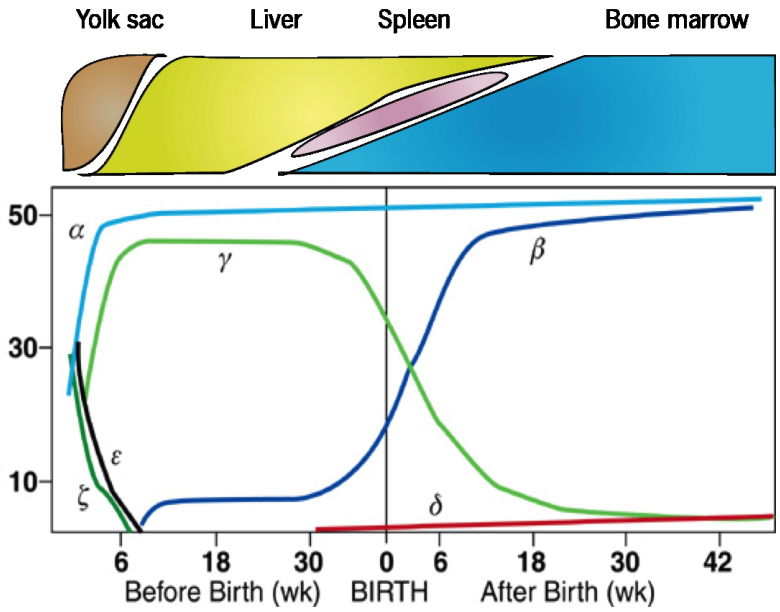


Figure 5
Switch of non–α globin genes and location of erythropoiesis.

The switch from ε– to γ–globin production begins very early in gestation (HbF can be detected in 5–week–old human embryos) and it is complete before the 10th week of gestation. β–globin expression starts early in human development, β–chain synthesis increases to approximately 10% of total Hb by 30–35 weeks of gestation (Steinberg *et al.*, 2009). At birth, HbF consist of 60%–80% of the total Hb, with a 70:30 $^{G}\gamma$:$^{A}\gamma$ ratio (Manca and Masala, 2008).

HbF takes approximately 2 years to reach the level of 0.5%–1% and is restricted to a few erythrocytes called "F cells": 3%–7% of erythrocytes are F cells and each contains approximately 4–8 pg of HbF, with $^{G}\gamma$:$^{A}\gamma$ ratio reverted to 40:60.

Understanding of the control of switching is expected to lead to the development of treatments of Hemoglobinopathies (Hbpathies) in which the

production of HbF can compensate for the loss of β–chain activity or can decrease imbalance between α:non–α globin ratio (Para. 3.2).

During ontogenetic development, there is a competition between regulatory sequences and active globin genes. These competitive interactions are influenced by variations in the relative concentrations and/or post–translational modifications of transcription factors that are expressed at all developmental stages (Steinberg *et al.*, 2009).

Nonetheless, since eukaryotic genes are packaged in the nucleosomes and that gene activation is always affected by many factors like the distal regulatory sequence, the proximal regulatory element, the chromatin structure and so on, globin gene regulation is destined to be very complex, involving intergenic transcription, nuclear localization and histone acetylation (Tang *et al.*, 2002).

2.1 Regulatory regions and transcription factors of globin genes

Promoter regions

For genes transcription one DNA segment have to interact with RNA polymerase II and its accessory factors (such as TFII–D and TFII–B): this DNA segment is called basal promoter.

Five motifs are present in human genes **basal promoters**: from 5'→3': **BRE** (–37 nucleotides 5' to CAP), **TATA box** (–31 nt), **Inr** (+1), **MTE** (+27) and **DPE** (+32); they are also required to drive the transcription of the globin genes (Fig. 6 A) (Maston *et al.*, 2006).

The upstream regulatory region in globin genes runs from approximately positions –40 to –250 (Fig. 6 B). Only one motif in this region is found in all the highly expressed globin genes: the **CCAAT box**; proteins such as **NF–Y** and **CP1** bind to this motif. It is missing from the δ–globin gene (HBD) promoter, but this gene is expressed at a low level (~1%–2% of HBB).

Two motifs are found in many but not all promoters. One is the **CACC box**, which is bound by transcription factors in the Krüppel–like zinc finger class (**KLF**).

Mutations in almost every position in the proximal CACC box of HBB promoter have been associated with β–thalassemia. Other KLFs may bind to the CACC boxes in other globin gene promoters, such as FKLF or KLF1366 for the HBG1 and HBG2 promoters.

The other motif occurring frequently in upstream regulatory regions is **WGATAR**, the binding site for **GATA1** and related proteins. GATA1 plays a critical role in both erythroid–specific *gene activation* and *repression*.

Furthermore, a different set of binding sites is distinctive to each type of gene. For instance, **βDRF60** and **BB1**–binding protein regulate the β–globin gene but not other globin genes (Fig. 6 B). Also, binding of **OCT1**, **SSP** and **γPE** has been shown for the upstream regions of γ–globin genes but not others.

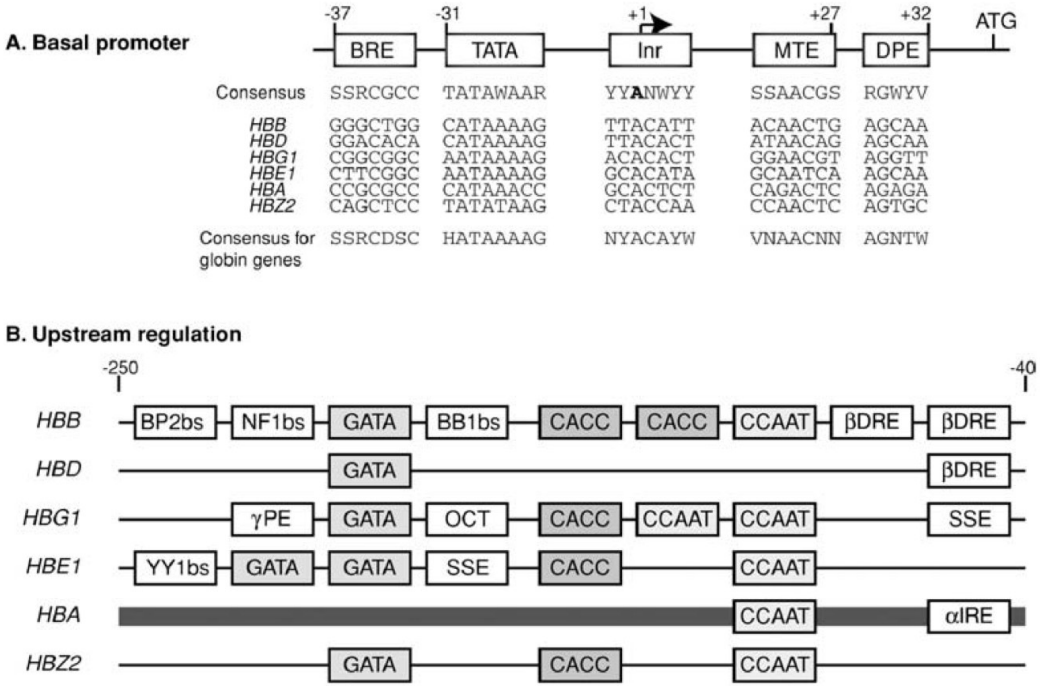


Figure 6

Motifs and binding sites in cis–regulatory modules of globin genes. The boxes were placed in the correct order but spacing is not indicated. The thick line for the HBA upstream regions (both HBA1 and HBA2) denotes that it is a CpG island.

A) Motifs in the basal promoter
B) Motifs in the regulatory regions immediately upstream of the basal promoters

(Steinberg *et al.*, 2009)

SSP is composed of the ubiquitously expressed factor CP2 and a recently cloned protein, NF–E4, which is erythroid specific and activates γ–gene expression in transfection experiments *in vitro*. These studies show that SSE has the ability to allow the γ–promoter to function in preference to a linked β–globin gene in plasmid constructs containing the HS2 portion of the β–LCR. SSP DNA binding activity appears to be relatively restricted to fetal erythroid cells. So, it is believed that γ–globin synthesis is stimulated in part by expression of SSP, which impart a competitive advantage for recruitment of the β–LCR to the γ–promoter. The SSE, however, is neither necessary nor sufficient for competitive inhibition of β–globin gene expression in immortalized erythroid cell lines (Steinberg *et al.*, 2009).

Another protein complex with potential roles in globin gene switching is direct repeat erythroid definitive (DRED). **DRED** was isolated through its affinity for direct repeat (**DR**) elements that cluster near the ε–globin and γ–globin promoters. DR sequences are of interest because mutations in this region are associated with several cases of Hereditary Persistence of Fetal Hb (HPFH, Chap. 4) (Steinberg *et al.*, 2009).

The most distinctive globin gene promoters are those of the α–globin genes (HBA1 and HBA2). These promoters are **CpG islands**, and among the Hb genes, only those encoding α–globin have this feature. It is still not clear what can prevent their expression in non–erythroid tissues and which sequences, in addition to the CpG island, lead to very high–level expression in erythroid cells.

No GATA1–binding site is found in the α–globin gene promoters, so sequence–directed binding of this protein to the proximal sequences is not the answer.

Several studies have shown that the CpG island is a key component of the *cis*–regulatory elements for the α–globin gene of humans and rabbits, possibly through its effects on chromatin structure (Steinberg *et al.*, 2009).

Enhancers

An enhancer is a DNA sequence able to increase the activity of a promoter; usually it is located on either side of a gene or internal to it (proximal enhancer), but it can also be located at significant distances (distal enhancer).

Two **proximal enhancers** have been found close to genes in the β–globin gene cluster, one that is 3' to HBB (3'HS1) and one that is 3' to HBG1. In both cases the enhancers are less than 1 kb downstream of the poly(A) additional signal for the respective genes.

Both enhancers contain binding sites for GATA1 and the HBG1 enhancer also binds to the γPE protein (Fig. 7).

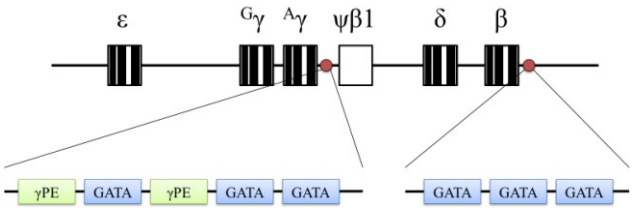


Figure 7
Motifs in the proximal enhancers at 3' $^{A}\gamma$ and β (3'HS1).

The **3'HS1** enhancer is located downstream from the poly(A) site of the β–globin gene. A deletion of sequences between 605 and 895 bp 3' to the poly(A) site, results in a 10 fold reduction in the level of expression of the β–globin gene in transgenic mouse. Also, a DNA fragment including β–globin 3'–flanking sequences (425–1480 bp from the poly(A) site), in either orientation, activates transcription from the otherwise silent $^{G}\gamma$–globin promoter in the mouse fetal liver. This enhancer, in the region approximately 600–900 bp 3' to the β–globin poly(A) site, contributes to the differential stage–specific expression of the β–globin and $^{G}\gamma$–globin genes (Trudel and Costantini, 1987).

The enhancer located downstream from the $^{A}\gamma$ gene was identified on the basis of transient transfection experiments. This element contains binding sites for various transcription factors, but it appears to have no effect on γ–globin gene expression *in vivo*. In transgenic mice, presence of this 3' element protects the γ gene from

position effects, suggesting that its likely role is stabilization of the interaction between the γ–globin gene and the β–LCR (Steinberg *et al.*, 2009).

Another enhancer has been recently found in intron 2 of the β gene. For β–LCR activity, the 5'HS3 element requires the interaction with the AT–rich region (ATR) in β–globin IVS2. The ATR consists in an inverted double GATA1 site composed of 2 non–canonical sequences (GATT and GATG) and an OCT1 consensus site (Bharadwaj *et al.*, 2003).

In addition to the TATA and initiator elements, conserved Enhancer box (E–box) motifs are located in the β–globin downstream promoter. Previous studies showed that E–box–binding proteins can participate in the formation of transcription complexes on both TATA–containing and TATA–less genes.

The downstream promoter region of the human β–globin gene contains three **E–box motifs** (consensus sequence CANNTG). Two of these E–boxes, the one overlapping the initiator (+1) and the distal E–box (+60), are conserved in human, mouse and rabbit, whereas the E–box located at +20 is only present in the human and rabbit genes.

One of the E–box motifs overlaps the initiator and this composite element interacts with USF1 and TFII–I in vitro. It is proposed that in cells not expressing the β–globin gene (inactive), TFII–D is not bound and the initiator sequence is occupied by protein complexes consisting of TFII–I and USF2.

Another E–box, located 60 bp 3' to the transcription initiation site, interacts with USF1 and USF2. Mutations of either the initiator or the downstream E–box impair transcription of the β–globin gene in vitro.

Mutation of sequences between the initiator and the E–box at +60 had no effect nor did a mutation introduced into the species non–conserved E–box at +20 (Leach *et al.*, 2003).

Distal enhancers are needed for high–level expression of any gene in the linked globin gene clusters. These are the β locus control region (β–LCR) for the β–globin gene cluster and HS–40 or major regulatory element for the α–globin gene cluster.

The **β–LCR** is a very large regulatory region, containing five DNase hypersensitive sites (5'– HS5 to HS1 –3') in humans, extending over 17 kb between an OR gene and HBE (Fig. 4 B). The β–LCR is necessary for opening a chromosomal domain and chromatin repositioning. Detailed structure and mechanism of Hb enhancement mediated by β–LCR can be found in Paragraph 2.2.

The distal enhancer for the α–globin gene, **HS–40**, is much smaller (250 bp) than the β–LCR, and it's located in a widely expressed gene called C16orf35 (Fig. 4 A). HS–40 is sufficient for strong enhancement and high activity in erythroid cells. Motifs in distal positive regulator HS–40 can be found in Fig. 8 (Steinberg *et al.*, 2009).

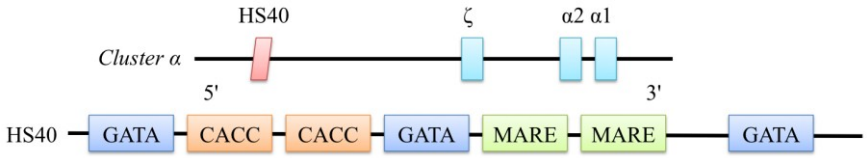


Figure 8
Motifs in distal positive regulator HS–40 for the α–globin gene cluster.

Both the β–LCR and the α–globin HS–40 are very strong, erythroid–specific enhancers needed for the expression of any of the linked globin genes.

Intergenic regulatory elements

Studies of *cis*–linked intergenic regulatory elements that control β–like globin gene expression have revealed a variety of motifs that activate or repress the globin genes at the appropriate stage of development. These elements include silencers, insulators, MARs/SARs, and boundary elements:

Silencers. Silencers bind repressor protein complexes that interfere with promoter activity, thereby downregulating gene expression. A silencer located in the distal promoter of the ε–globin gene controls autonomous repression of ε–globin gene expression during the fetal and adult stages of development (Harju *et al.*, 2002). Other silencers can be found in the distal promoter of both $^{G}\gamma$– and $^{A}\gamma$–globin gene (Fig. 9) (Steinberg *et al.*, 2009).

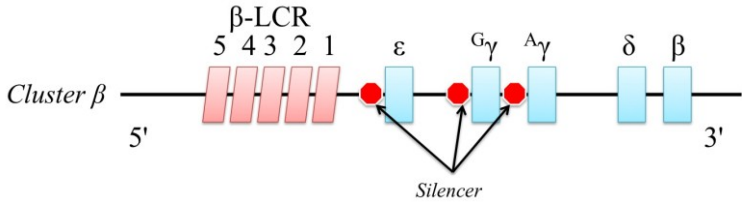


Figure 9
Schematic representation of the position of Silencer elements on the β–globin gene cluster.

GATA1 and YY1 proteins constitute at least two of the components of the repressor complex.

Transgenic mouse experiments have localized a potential **silencer element** in the **–382 to –730** region (Stamatoyannopoulos *et al.*, 1993). That the −382 to −730 region may contain a silencer has also been shown by the finding of an HPFH mutation at position −567 that alters GATA1 binding (Chen *et al.*, 2008; Steinberg *et al.*, 2009).

Insulators. Insulator elements protect against the negative effects of neighboring heterochromatin and may serve as boundary elements (See Below) that flank or demarcate an open, transcriptionally active chromatin domain. Insulator elements may block histone deacetylase activity. The human β–LCR has itself insulator properties. Insulators facilitate the activity of enhancers located within an open chromatin domain.

MARs/SARs. Matrix attachment regions (MARs) or scaffold attachment regions (SARs) are elements that promote binding to the nuclear matrix, resulting in the formation of contiguous DNA sequence loops. These elements may provide a barrier

by shielding the locus from the effects of surrounding negative chromatin or provide a structural restraint to chromatin remodeling.

Boundary Elements. Boundary elements may be located at various positions within a locus and may assume a restrictive role regarding gene expression when associated with binding proteins. Three properties may be characteristic of boundary elements: 1) possible association with insulators, 2) maintenance of a steady state between open and closed chromatin, and 3) presence of terminal domain sequence elements and binding proteins. Boundary elements have been defined as sequences that isolate specific chromatin domains. They contain *cis*–acting elements that have a positive influence within the domain and prevent chromatin influence from outside the domain. (Harju *et al.*, 2002)

2.2 The β–Globin Locus Control Region (β–LCR) role in globin expression

β–LCR is located 6–25 kb upstream from the ε–globin gene and contains a series of developmentally stable DNase I hypersensitive sites (5'– HS5 to HS1 – 3').

A large number of studies indicates that the activities of the β–LCR are mostly localized to the core elements of the HS, which are approximately 300 bp long, where transcriptional factors bind. Nonetheless, the regions flanking the HS elements of the β–LCR are also important for function.

Three transcription factor binding motifs are present in almost all HS: Maf–response elements (**MAREs**) to which transcriptional activator proteins of the basic leucine zipper class can bind (e.g. NFE2, LCRF1/Nrf1, and Bach1); **GATA** motifs, to which GATA1 and its relatives bind; **CACC**, to which a family of Zinc–finger (EKLF, Sp1) can bind (Steinberg *et al.*, 2009).

HS2, the strongest enhancer of the β–globin β–LCR, like HS–40, have two MAREs without which enhancement is strongly decreased (Caterina *et al.*, 1994).

HS2 also has three E–boxes, where TAL–1 and its partners bind. This protein involved in regulation of hematopoiesis, appears to be a key for enhancement by HS2 (Elnitski *et al.*, 1997).

Position of these elements can be found in Fig. 10.

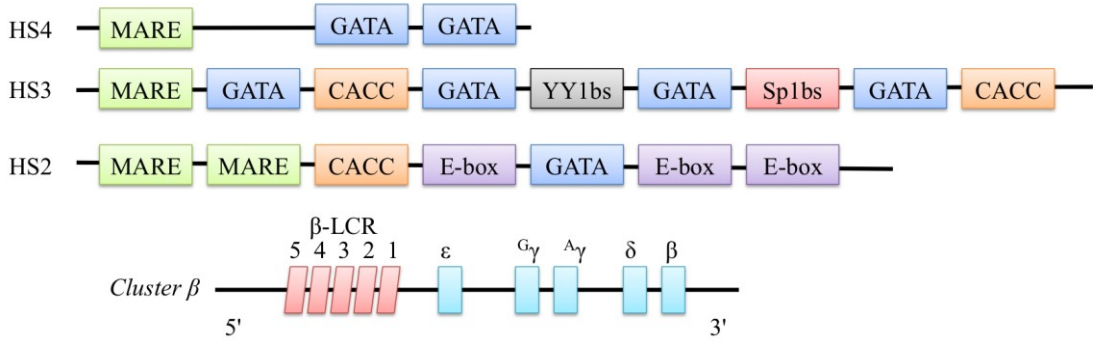


Figure 10
Motifs in distal positive regulator: HSs of the β–LCR.

A study by Fraser *et al.*, 1993 showed that the HS of the β–LCR have developmental specificity. Other experiments on transgenic mice confirmed this hypothesis. For example deletion of the core element of HS3 results in total absence of ε–globin gene expression in day–9 embryonic cells, while γ–gene expression in embryonic cells is normal, suggesting that HS3 is necessary for activation of ε–globin gene transcription. However, γ–globin gene expression, is totally absent in fetal liver cells, indicating that the core of HS3 is necessary for γ–gene transcription in the fetal stage of definitive erythropoiesis (Steinberg *et al.*, 2009).

Individual function of each HS linked to reporter gene is summarized in Table I (Harju *et al.*, 2002).

Some uncertainty still exists regarding the direct effects of the β–LCR on chromatin conformation. For example experiments have been recently indicated that the β–LCR is not required for opening the chromatin domain in knockout mice, in contrast to what happens in humans. While in (εγδβ)–thalassemia mutants due to β–LCR deletions there is total inactivation of the β–locus chromatin and total absence of transcription of the β–cluster genes *in cis*, when the β–LCR is deleted from the endogenous murine β–LCR, the globin genes continue to show some low

levels of expression, and the chromatin of the locus remains in the open configuration. Why β–LCR deletions in humans and the β–LCR knockouts in mice phenotype is different is still unknown.

A possible reason may be due to the differences in the composition and organization of the murine and the human β–LCRs. Alternatively, the total silencing of the locus in the human β–LCR deletions might not be caused by the deletion of the β–LCR *per se*, but the juxtaposition to the locus of heterochromatic regions, located upstream, that silence the genes of the locus (Steinberg *et al.*, 2009).

One of the main questions in transcriptional regulation is how *cis*–regulatory elements communicate with the promoter of a gene over large distances.

Four different models (Fig. 11) were proposed to explain the functions of β–LCR and provide evidences that the HSs of β–LCR act either independently or synergistically. Other than these models, chromatin remodeling, topologic alterations and modifications of proteins associated with the chromatin are also involved in the β–LCR's function (Kukreti *et al.*, 2010).

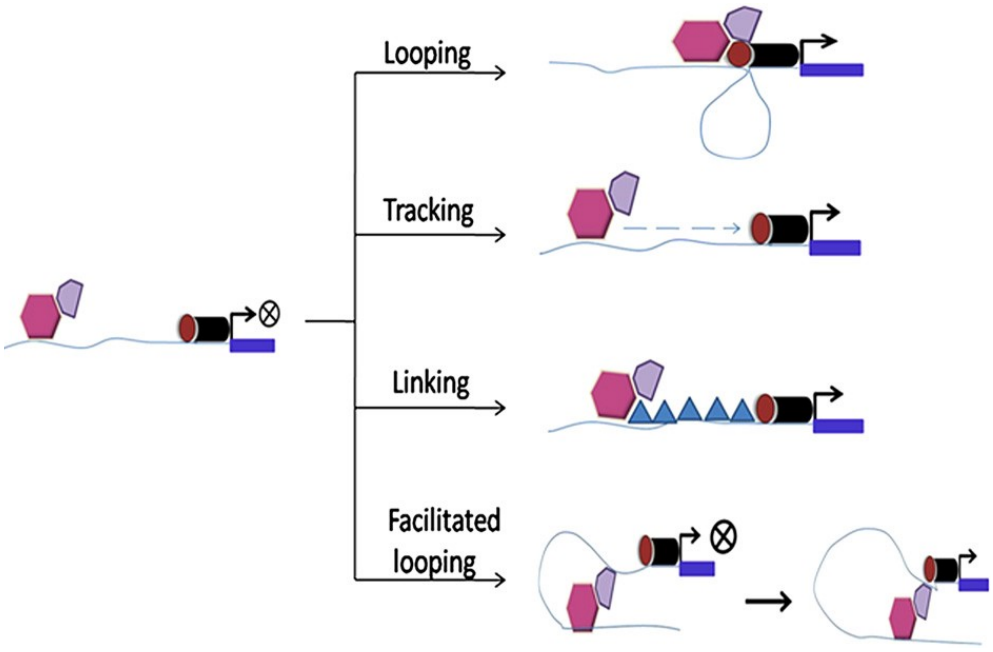


Figure 11
Models for β–LCR mechanism of enhancement.

(Kukreti *et al.*, 2010)

The "**looping model**" was first proposed by Choi and Engel, 1988, who suggested that the distal and proximal regulatory elements could interact directly by the formation of a loop. Each HS of the β–LCR does not function alone. The factors bound to each of the β–LCR HSs contribute cooperatively to form a larger physical entity termed the "holocomplex".

The HS core elements in fact bind transcription factors and the core–flanking sequences constraining the holocomplex in the proper conformation (Tang *et al.*, 2002). This structure loops so that the β–LCR comes in close proximity and associate with proximal promoter and enhancer elements of the gene. This allows the delivery of β–LCR–bound transcription proteins to interact with the basal transcription apparatus, already bound at the promoter to form a stable transcription complex, thus enhancing globin gene expression.

A variation of this model sees β–LCR as a multiple element receptor, where chromatin remodeling factors bind. After completion of chromatin–remodeling activity, the β–LCR can directly interact with downstream genes and facilitate their expression (Kukreti *et al.*, 2010).

HS sites at the β–LCR are required to form a unique 3–dimensional structure at the chromosomal location of the β–globin locus on chromosome 11, with "looping out" of the locus in a unique spatial configuration, an active chromatin loop at the β–globin locus, termed an "*active chromatin hub*" (ACH). The β–ACH includes the β–LCR sequences, interacting with downstream globin genes and intergenic sequences, transcriptional factors, and chromatin remodeling elements necessary for globin gene expression and Hb switching in a special compact configuration (Fig. 12).

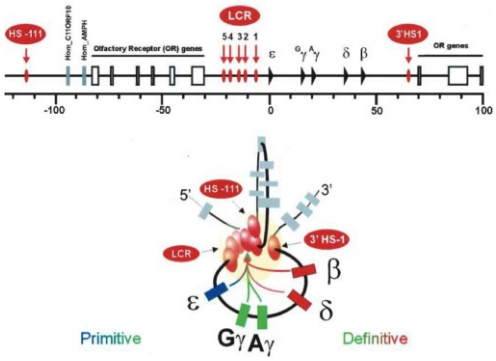


Figure 12
Two–dimensional representation of the 3–D interactions between the HSs and the globin genes in erythroid cells, forming the so called Active Chromatin Hub (ACH).
(Patrinos *et al.*, 2004)

The β–ACH structure is not found in non–erythroid cells (Bank, 2006).

The "**tracking model**", also called "scanning model", suggests that transcription factors and cofactors bind to recognition sequences in the β–LCR sequences, forming an activation complex that migrates, or tracks, linearly along the DNA helix of the locus. When this transcription complex encounters the basal transcription machinery located at the correct promoter according to the developmental stage, the complete transcriptional apparatus is assembled and initiation of transcription of that gene takes place. The choice of which gene has to be transcribed is mediated by histon deacetylase and methylase proteins within the complex, after the transcription is activated: it is supposed that this mechanism can limit activation to a particular developmental stage.

In "**facilitated–tracking model**" aspects of both the looping and tracking models are incorporated. Some transcription factors bind 5'HS of β–LCR, which loops to contact in promoter–distal regions of downstream DNA. There, the transcription factor complex is released and tracks along the chromatin until a stable loop structure is established and gene expression proceeds (Kukreti *et al.*, 2010).

The "**linking model**" proposes that gene activation is the product of sequential stage–specific binding of transcription factors and chromatin "*facilitators*" throughout the locus to form an array of chromatin elements that initially define the domain to be transcribed.

For example, when the chain reaches the γ–globin promoter in human fetal liver, the facilitator enhances binding of the transcription factors that activate γ–globin gene expression. The regulatory protein also acts with the boundary elements as a boundary to inhibit further chain extension. Therefore, the downstream β–globin gene is inactive. When the boundary function of the promoter is absent, the complex at this proximal promoter may actually function as another link in the chain. The current gene is inactivated and the following one is activated.

In this view, the interaction between β–LCR and individual promoters is mediated by a chain of complexes containing transcription factors and chromatin facilitators.

Moreover, non–DNA–binding facilitator proteins form a continuous protein chain from the β–LCR to the globin gene to be transcribed (Tang *et al.*, 2002).

There has been a debate in recent years whether this communication between β–LCR and β–cluster takes place via a non–contact mechanism (linking, tracking) or via a contact mechanism (looping). Although it is now firmly established that looping takes place within the β–globin locus (and other loci), it is still not clear how these long–range contacts are established and what the precise role of the β–LCR is. (Palstra *et al.*, 2008)

2.3 Chromatin role in β–like globin gene expression: the PYR role

Studies have not only focused on the interactions of *trans*–acting factors and *cis*–acting elements including the β–LCR, but also on remodeling of chromatin. Moreover, topologic alterations and modifications of proteins associated with the chromatin are concomitant with the β–LCR's function. These modifications directly or indirectly imply the ability of β–LCRs to alter chromatin configuration and conformation.

The exact molecular mechanism of gene expression during human development is still not fully understood. It is unclear the mechanism by which the β–LCR changes the chromatin conformation throughout the entire β–globin gene cluster (Kukreti *et al.*, 2010).

Acetylation of histones in chromatin leads to increased expression of specific genes at structural gene loci, including the β–globin locus, by making gene sequences more available to transcription factors. Specific acetylations of histones H1–H8 in chromatin can be achieved either by histone acetylase enzymes (HAs)

alone or by ATP–driven molecular machines providing this function, such as SWI/SNF complexes.

By contrast, deacetylation of histones in chromatin by histone deacetylases (HDACs), can repress and silence specific gene function at many loci and may be involved in repressing human γ–globin gene expression.

Other enzymatic complexes can also modify histone N–terminal tails by methylation, phosphorylation, and ubiquitination as well (Bank, 2006).

Bank group has described a chromatin remodeling complex called **PYR complex** found only in adult hematopoietic cells. It is a single complex with Ikaros (its DNA–binding subunit) and contains both positive–acting protein subunits of the SWI/SNF complex that activate gene transcription and repressive protein subunits of the NURD complex, an HDAC–containing complex that can repress transcription.

PYR complex binds to a 250–bp polypyrimidine (PYR)–rich DNA sequence 1 kb upstream of the human δ–globin gene; the PYR complex DNA–binding site is deleted in the HPFH–2, but not in the Sicilian δβ–thalassemia (Para. 4.1, Fig. 18). Deletion of 511 bp of DNA including the PYR binding site upstream of the human δ–globin gene delays human γ to β switching in mice.

PYR complex has a role in normal human γ to β switching, and that perhaps inhibition of PYR complex action, for example, by siRNA inhibition of Ikaros activity, may be an additional therapeutic approach to reactivation of γ–globin in adult erythroid cells.

PYR role in Hb switching has been proposed: during fetal life, the γ–globin genes associate with the β–LCR through erythroid and fetal stage–specific transcription factors, and with putative embryonic–fetal stage–specific chromatin remodeling complexes, to favor γ–globin transcription.

In adult cells, the β–ACH configuration is changed so that the β–LCR now preferentially associates with and activates human β–globin gene expression.

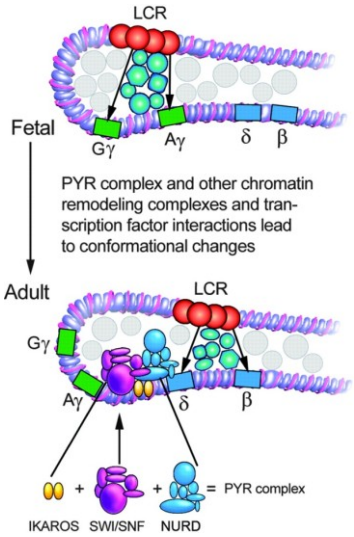


Figure 13
The proposed role of the PYR complex in human hemoglobin switching.
(Bank, 2006)

To induce conformational changes at the β–globin locus in adult–type cells it has been supposed an adult stage–specific formation and action of the PYR complex, by binding to the intergenic γ–δ sequences at the PYR–binding site and repressing γ–globin transcription by its HDACs. The SWI/SNF subunits of PYR complex may be involved as well. All of these proposed roles for PYR complex must await confirmation by further studies of the role of Ikaros and the PYR complex (Fig. 13) (Bank, 2006).

2.4 Summary on the fetal to adult switch

One of the most interesting characteristics of human Fetal to Adult Hb switching is its perfect control and the continuation of synthesis of small amounts of HbF in the adult.

The most accepted explanation comes from the observations made in transgenic mice carrying either the γ– or the β–globin gene or both genes linked to the β–LCR. When the genes were alone, developmental control was lost. When the genes were linked together, developmental control was restored. Such findings led to the proposal that the γ–globin gene is regulated through competition with the β–globin gene and vice versa.

The hypothesis is that in the embryonic stage, the β–LCR interacts with the ε–globin gene; the downstream genes are being turned off competitively. In the fetus, the ε gene is silenced due to its Silencer element, and the β–LCR interacts with the $^{G}\gamma$ and $^{A}\gamma$ genes. In the adult, the γ genes are silenced, and the β–LCR can now interact with the β–globin gene, the last gene of the locus. The closer the gene, the higher is the advantage. Its placement at the 3' end of the locus might explain why

the adult genes are totally shut off in embryonic cells when it is located in its normal chromosomal position, whereas it is always expressed in the embryo if it is placed next to the β–LCR (Fig. 14).

In addition to the *cis*–acting elements already described, also the *trans*–acting factors such as EKLF, GATA1, its cofactor FOG1, and perhaps other factors of the KLF/SP1 family are likely to facilitate the formation of the loop between the β–LCR and the γ– or β–gene promoters (Steinberg *et al.*, 2009).

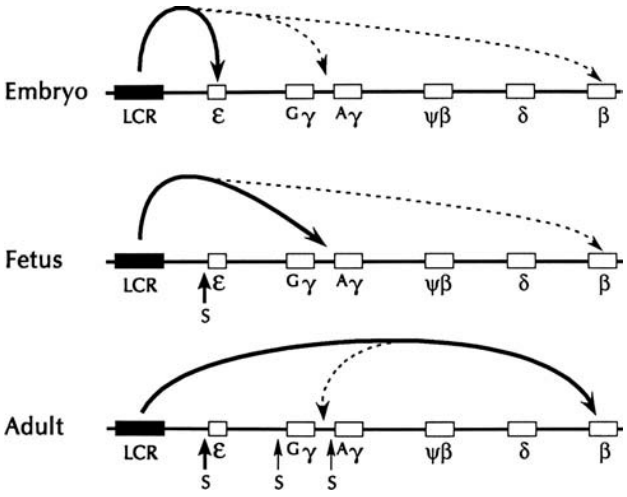


Figure 14
Model of the competitive control of hemoglobin switching. "S" indicates the activity of a silencer element.

(Steinberg *et al.*, 2009)

3. CLASSIFICATION OF THE DISORDERS OF HB

To 2013, more than 1000 disorders of Hb synthesis and/or structure have been identified and characterized: each one can lead to a distinct clinical phenotype. Genotype/phenotype correlations have provided important understanding of pathophysiological mechanisms of disease in the disorders of Hb.

Hb disorders can be classified into two general categories (listed in detail in Table II A and B):

A. **Thalassemia Syndromes** or **Quantitative Hemoglobinopathies** in which there is a quantitative defect in the production of one of the globin subunits, either total absence or marked reduction.
B. **Qualitative Hemoglobinopathies** in which a **structural defect** on the gene alters the primary structure of the globin (Forget and Bunn, 2013). The amino acid substitution may be phenotypically silent or change the properties of the tetramer, resulting in differences in stability, O_2 affinity and other functions of the Hb.

3.1 The Thalassemia syndromes

The thalassemia syndromes are inherited disorders characterized by absence or markedly decreased accumulation of one of the globin subunits of Hb. Thalassemia was first defined by Dr. Thomas B. Cooley in 1925 at the annual meeting of the American Pediatric Society where he presented five young children with severe anemia, splenomegaly and bone abnormalities (Weatherall, 2010).

In the α–thalassemias, there is absent or reduced production of α–globin whereas in the β–thalassemias, there is absent or decreased production of β–globin. Rare thalassemias affecting the production of δ– or γ–globin subunits have also been described but are not clinically significant disorders (Forget and Bunn, 2013).

3.1.1 α–thalassemias

In contrast to the β–thalassemias, which are usually caused by point mutations of HBB (Para. 3.1.2), the α–thalassemia syndromes are usually due to the deletion of one or more α–globin genes. Several different point mutations cause the less common non–deletional forms of α–thalassemia ($\alpha^T\alpha/\alpha\alpha$).

α–thalassemias are subclassified according to the number of α–globin genes that are deleted (or mutated):

- α^+–thalassemia: one gene deleted (–α/αα);
- α^0–thalassemia: two genes deleted (–α/–α or – –/αα);
- HbH disease: three genes deleted (– –/–α);
- Hydrops fetalis with Hb Bart's: four genes deleted (– –/– –)

  (Forget and Bunn, 2013).

In α–thalassemia, the synthesis of α–globin chains is downregulated so that in fetal life there is anemia and the excess γ–globin chains form soluble tetramers (γ_4) called Hb Bart's. In adult life, α–thalassemia also causes anemia but, because by this time the γ to β switch is complete, the excess non–α chains assemble into β_4 tetramers, called HbH. The degree of anemia and the amounts of the abnormal Hbs produced (Bart's and H) reflect the degree to which α–globin synthesis has been downregulated (Steinberg *et al.*, 2009).

In contrast to β–thalassemia minor, HbA_2 levels do not raise, so its identification is based mainly on hematological parameters (Para. 3.1.2).

Clinically, the deletion of only one of the four α–globin genes is not associated with significant hematologic abnormalities and is sometimes called the "silent carrier".

The deletion of two α–globin genes can occur on the same chromosomes (*in cis*) or on opposite chromosomes (*in trans*). The clinical phenotype is similar with both genotypes and consists of mild hypochromic, microcytic anemia, without

hemolysis, somewhat analogous to that of β–thalasssemia trait (Para. 3.1.2), but somewhat less severe (Forget and Bunn, 2013).

These genotypes (–α/αα and – –/αα or –α/–α) may produce detectable amounts of Hb Bart's at birth. In addition, occasional cells containing HbH may be detected in adults with the – –/αα genotype (Steinberg *et al.*, 2009).

The deletion (or significantly decreased expression) of three α–globin genes is associated with the syndrome of HbH disease, a compensated hemolytic anemia that usually does not require treatment by RBC transfusion. However, HbH is relatively unstable and does precipitate as RBCs age, forming inclusion bodies that damage RBCs and shorten their lifespan.

The deletion of all four α–globin genes is usually fatal during late pregnancy or shortly after birth. This condition is called hydrops fetalis with Hb Bart's: this homotetramer has a very high O_2 affinity, similar to that of myoglobin, and does not release O_2 to the tissues under physiologic conditions. Consequently, the infant, whose RBCs lack HbF or HbA and contain mostly Hb Bart's, suffers severe hypoxia resulting in hydrops fetalis. Rare cases have been rescued by intrauterine transfusions, but these children subsequently require lifelong transfusion support similar to that required by children with β–thalassemia major (Forget and Bunn, 2013).

The molecular cause of deleted alleles is explained by DNA sequence analysis: the α–globin genes are embedded within two highly homologous, 4–kb duplication units whose sequence identity appears to have been maintained throughout evolution by gene conversion and unequal crossover events. These regions are divided into homologous subsegments (X, Y, and Z) by non–homologous elements (I, II, and III).

Reciprocal recombination between Z segments (Fig. 15 A), which are 3.7 kb apart, produces chromosomes with only one fusion α–gene ($\alpha^{-3.7}$, rightward deletion) and chromosomes with three α–genes ($\alpha\alpha\alpha^{anti\ 3.7}$).

Recombination between homologous X boxes, which are 4.2 kb apart, also gives rise to an α–thalassemia determinant ($\alpha^{-4.2}$, leftward deletion) (Fig. 15 B) and a $\alpha\alpha\alpha^{anti\ 4.2}$ chromosome.

A) Rightward Crossover **B) Leftward Crossover**

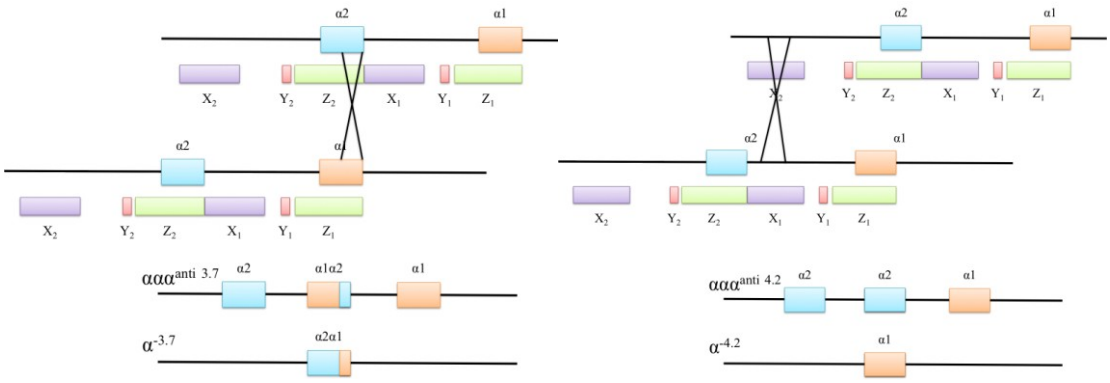


Figure 15
The mechanism by which the common deletions underlying α^{+}–thalassemia occur.

A) Rightward Crossover between misaligned Z boxes give rise to the $\alpha^{-3.7}$ and $\alpha\alpha\alpha^{anti\ 3.7}$ chromosomes
B) Leftward Crossovers between misaligned X boxes give rise to $\alpha^{-4.2}$ and $\alpha\alpha\alpha^{anti\ 4.2}$ chromosomes

Further recombination events between the resulting chromosomes may give rise to quadruplicated and quintuplicated α–genes ($\alpha\alpha\alpha\alpha^{anti\ 3.7}$, $\alpha\alpha\alpha\alpha^{anti\ 4.2}$).
Normal individuals with 4, 5 or 6 α–genes have been described: although there appears to be an excess of α–mRNA and α–globin produced from the ααα arrangement homozygotes (ααα/ααα) they are still hematologically normal (Hematological reference values can be found in Table III). While it has not been formally analyzed in human, in sheep it has been shown that the efficiency with which multiple α genes are expressed decreases from the 5' to 3' position on the chromosome. In this case, as more α genes are added to the cluster (αα to ααα to αααα), any increase in α–globin expression may be less than expected from the addition of fully expressed extra genes (Steinberg *et al.*, 2009).

In addition to these forms of α–thalassemia inherited with Mendelian genetics, there are two α–thalassemia syndromes that are caused by de novo or acquired mutations affecting expression of the α–globin genes: 1) the α–thalassemia with

mental retardation syndrome (ATR); and 2) acquired α–thalassemia (HbH disease) associated with myelodysplastic syndromes (ATMDS) (Forget and Bunn, 2013).

3.1.2 β–thalassemias

The β–thalassemias can be subclassified into those in which there is *total absence* of normal β–globin subunit synthesis or accumulation, the **β^0–thalassemias**, and those in which some structurally normal β–globin subunits are synthesized, but in markedly *decreased amounts*, the **β^+–thalassemias** (Forget and Bunn, 2013), with the mildest forms sometimes referred to as β^{++} or "silent" β–thalassemia. These common forms of β–thalassemias are inherited as mendelian recessives (Steinberg *et al.*, 2009).

The molecular basis of the β–thalassemias is very heterogeneous (Thein, 2013): more than 200 β–thalassemia alleles have now been characterized (http://globin.cse.psu.edu) involving mutations that affect the transcription of the β–globin gene, post–transcriptional processing of its pre–mRNA, or the translation of its mRNA into protein (Forget and Bunn, 2013).

Some structurally abnormal β–chain variants are also associated with quantitative deficiencies of β–globin chain production and have a phenotype of β–thalassemia, in which case they are referred to as "**thalassemic Hemoglobinopathies**" (e.g. HbE β(26) Glu→Lys).

In other structural variants, the mutated β–globins are so unstable that they undergo very rapid post–synthetic degradation. These hyperunstable β–chain variants act in a dominant negative way, causing a disease phenotype even when present in the heterozygous state, and hence have been referred to as "**dominantly inherited β–thalassemia**" (Steinberg *et al.*, 2009).

Usually, the mutations causing β–thalassemia are *point mutations* affecting a single nucleotide, or a small number of nucleotides, in the β–globin gene (Forget and Bunn, 2013).

One of the first mutations to be characterized and extensively studied is the nonsense mutation at codon 39 (**C**AG→**T**AG). This mutation is one of the most common cause of β–thalassemia in the Mediterranean population and accounts for most of the cases of β–thalassemia in Sardinia (See *"Aim of this study"*) (Steinberg *et al.*, 2009).

Rare *deletion forms* of β–thalassemia have also been described. These deletions are usually caused by "unequal" crossing over between the linked and partially homologous genes like δ– and β–globin genes, resulting in either the loss of genes or the formation of a fusion δβ–globin gene (e.g. δβ hybrid Lepore).

Large deletions involving part or all of the β–globin gene cluster are responsible for the δβ–thalassemias, the εγδβ–thalassemias, and HPFH syndromes (Para. 4.1).

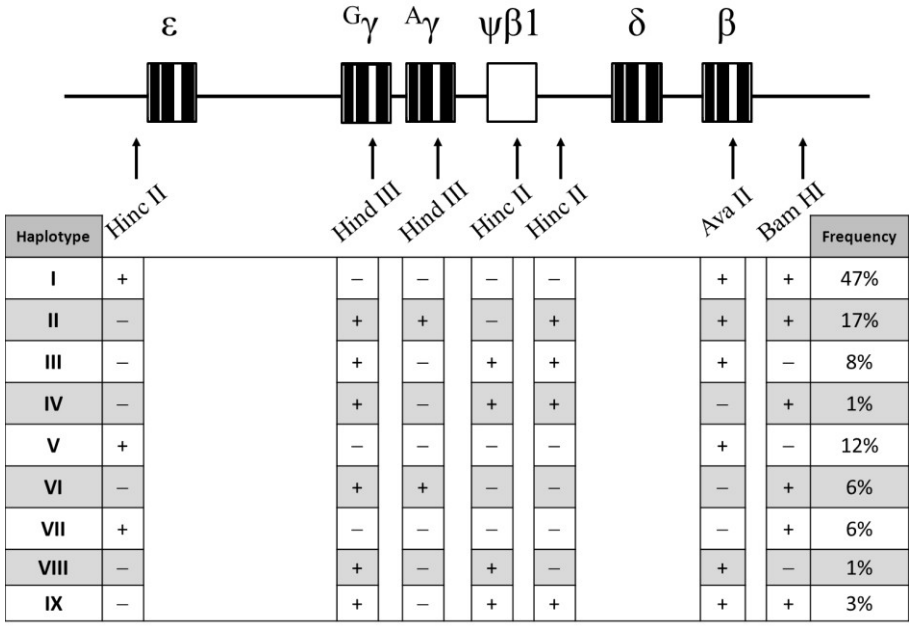
The εγδβ–thalassemias are associated with neonatal hemolytic anemia, but this resolves during the first few months of life and the associated phenotype in adults is that of β–thalassemia trait (See Para. 3.2) (Forget and Bunn, 2013).

β–thalassemias caused by point mutations were among the first human genetic diseases to be examined through recombinant DNA analysis: it showed the existence of multiple restriction site polymorphisms within the β–globin gene cluster and it allowed the detailed characterization of chromosome regions in which mutant β–globin genes reside.

The coupling of specific mutations and haplotypes is an example of *linkage disequilibrium* in which specific alleles become associated with neutral changes in the DNA (SNPs): this means that haplotypes can be coupled with a series of specific mutations on β–globin gene (Orkin *et al.*, 1982).

One of the first observations on the β–cluster was made by Orkin who performed combined analysis of DNA polymorphisms in the human β–globin gene cluster and in cloned β–gene, revealing the association of specific β–thalassemia mutations and β–gene SNPs with particular flanking polymorphism.

Nine different cleavage patterns were observed and designated haplotypes I to IX: in this Thesis they will be referred as Orkin haplotypes. Polymorphic restriction enzyme sites in the β–globin gene cluster and relative haplotypes are shown in Fig. 16. Their position on the β–cluster can be found in Paragraph 8.1.



Haplotype	Hinc II	Hind III	Hind III	Hinc II	Hinc II	Ava II	Bam HI	Frequency
I	+	–	–	–	–	+	+	47%
II	–	+	+	–	+	+	+	17%
III	–	+	–	+	+	+	–	8%
IV	–	+	–	+	+	–	+	1%
V	+	–	–	–	–	+	–	12%
VI	–	+	+	–	–	–	+	6%
VII	+	–	–	–	–	–	+	6%
VIII	–	+	–	+	–	+	–	1%
IX	–	+	–	+	+	+	+	3%

Figure 16
Polymorphic restriction enzyme sites in the β–globin gene cluster.
+ indicates the presence of cleavage at a particular site, and – indicates absence of cleavage. Haplotypes indicate the patterns of cleavage along a chromosome and are numbered I to IX.

(Orkin *et al.*, 1982)

These haplotypes are coupled with specific β–thalassemia defects: this indicates the presence of a specific mutant alleles carrying a mutation within β–globin gene. For example, the codon 39 mutation was identified on haplotype II (24% of Italians β^0(39) chromosomes) as well as haplotype IX (5%) (Orkin *et al.*, 1982).

Additionally, sequencing of β–globin gene revealed five SNPs: the synonymous mutation C→T at codon 2 (His→His) (See Para. 2.1 "Enhancers" for more information), and four SNPs in IVS2 (+16C→G or *Ava* II − → +), +74G→T, +81C→T and +666T→C). These SNPs indicated the existence of three β–globin gene frameworks:

- **Type 1** is characteristic of β–genes of haplotypes I, II and IX and show the same nucleotides as reference sequence (CCGCT), in the same order described above
- **Type 2** β–gene present in haplotypes III, V and VIII and have a single substitution at IVS2 +74 (CC**T**CT)
- **Type 3** genes encompass haplotypes IV, VI and VII and display all five polymorphisms (**TGTTC**) (Orkin *et al.*, 1982).

3.2 Pathophysiology and phenotypes of β–thalassemia

To reach the right corpuscular Hb concentration it is essential that α–globin as well as β–globin (or γ–globin) mRNA are expressed at very high levels and very closely matched: in fact free α–globin subunits are particularly toxic to erythroid cells. This problem is physiologically overcome by the presence of α–Hb stabilizing protein (AHSP), a chaperone that is expressed at high levels in erythroid cells and binds tightly to heme–intact α–globin subunits. The AHSP protects the cell from potentially toxic oxidized (Fe^{3+}) heme until it is reduced to functional Fe^{2+} heme. When α/AHSP dimer encounters a β–globin subunit, the α–globin dissociates from AHSP to form the extremely stable αβ dimer. This process is facilitated by electrostatic attraction between positively charged α–globin subunits and negatively charged β–globin subunits.

Mutations of globin genes, that impair synthesis or some that alter the structure of globin subunits, give rise to thalassemia and anemia of varying degree (Forget and Bunn, 2013). Also, altered AHSP expression levels or AHSP variants could modulate the severity of β–thalassemia in humans (Kong *et al.*, 2004).

According to the severity of the phenotype, β–thalassemias can be subclassified in three different groups having different degrees of biosynthetic imbalance between α– and non–α–globins:

- **β–thalassemia trait** or **β–thalassemia minor (Tm)** (heterozygotes for a defective β–gene). Heterozygotes have minimal anemia, but hypochromic (mean corpuscular Hb [MCH] 18–24 pg), microcytic (mean corpuscular volume [MCV] 65–80 fl) red blood cells (Hematological reference values can be found in Table III). They are characterized by an increased proportion of HbA_2 (normal <3.2%, β–thalassemia trait 3.5%–6.0%), and HbF levels that vary from normal (<1.0%) to slightly raised (1.0%–3.0%) (Forget and Bunn, 2013). Most cases of normal HbA_2 in β–thalassemia trait result from coinheritance of δ–thalassemia *in cis* or *trans* to a β–thalassemic gene (Steinberg *et al.*, 2009), or for the presence of the −196 C→T ndHPFH mutation on HBG1 gene, which may compete for transcriptional factors shared with HBD gene (Loudianos *et al.*, 1992).
- **β–thalassemia major** or **Cooley anemia (TM)** (homozygotes or compound heterozygotes), are characterized by a severe transfusion–dependent hemolytic anemia associated with marked ineffective erythropoiesis resulting in destruction of erythroid precursor cells in the bone marrow (Forget and Bunn, 2013).
- **β–thalassemia intermedia (TI)** is a less common clinical phenotype. It was first termed Thalassemia Intermedia by Sturgeon to describe the patients who had "clinical manifestations that are too severe to be termed minor and too mild to be termed major" (Sturgeon *et al.*, 1955). TI is characterized by Hb levels maintained around 7–10 g/dl, by more severe red blood cell (RBC) abnormalities than Tm, by a varying degree of spleen enlargement, and by skeletal changes such as expansion of facial bones and obliteration of maxillary sinuses which causes protrusion of the upper jaw. Also TI accumulate iron due to increased intestinal iron absorption and reduced expression of hepcidin, a hepatic peptide that plays a central role in iron homeostasis (Taher *et al.*, 2009). TI patients show a moderate to severe, partially compensated, hemolytic anemia that usually does not require chronic transfusion therapy to maintain a satisfactory circulating Hb level in the affected patient, although occasional transfusions may be required if the anemia worsens because of associated complications (Forget and Bunn, 2013).

Such TI patients have a milder disease because there is less severe α– to non–α–globin subunit imbalance than in typical β–TM patients, resulting in less accumulation of free α–subunits that cause the ineffective erythropoiesis.

There are different possible causes for such a lowered α– to non–α–globin subunit imbalance, including:

- inheritance of milder *β^{+}–thalassemia* mutations with less severe than usual deficiency of β–globin subunit production;
- coinheritance of a form of *α–thalassemia* (contrary ααα/αα or αααα/αα genotype exacerbates TI);
- coinheritance of determinants associated with *increased production of the γ–subunit of HbF* (Chap. 4).

Most patients with TI carry two mutant β–globin genes: they have a genotype typical of TM, but the phenotype is modified by one of the factors listed above.

However, rare cases of TI are caused by heterozygosity for a single mutant β–globin gene associated with the production of a highly unstable β–globin subunit (dominant β–thalassemic allele) (Forget and Bunn, 2013).

4. PERSISTENCE OF FETAL HEMOGLOBIN

Sickle cell disease (SCD) and β–thalassemia are among the most common genetic diseases worldwide and have a major impact on global health and mortality. Both these Hbpathies display a remarkable diversity in their disease severity. A major ameliorating factor is an innate ability to produce HbF.

HbF levels vary considerably, not only in patients with β–Hb disorders, but also in healthy normal adults.

Although the majority of adults have HbF of less the 0.6% of total Hb, 10–15% of individuals have increases ranging from 0.8% to 5%. The latter individuals are considered to have heterocellular hereditary persistence of fetal Hb (hHPFH), which refers to the unequal distribution of HbF among the erythrocytes.

When coinherited with β–thalassemia or SCD, hHPFH can increase HbF output to levels that are clinically beneficial (Thein *et al.*, 2007).

Some other molecular conditions lead to pancellular HPFH: heterozygotes show essentially normal red cell indices, a normal level of HbA_2 and higher levels of HbF (15%–30%) with a more homogeneous, pancellular distribution (Steinberg *et al.*, 2009).

HPFH homozygotes can express up to 100% of HbF (Thein and Craig, 1998) in RBC and are clinically normal, although with reduced MCV and MCH indices.

cis–acting variants explain some of the variability in HPFH and include mutations on the promoter of γ genes, large deletions on the β–globin gene cluster and polymorphic associations within the β–cluster.

Many *cis*–acting forms of persistence of HbF are due to gene deletions and are classified as either deletional HPFH (dHPFH) or δβ–thalassemia (Para. 4.1).

In other *cis*–acting forms HbF levels are influenced by a SNP in the promoter region of HBG1 and HBG2 genes, which affect gene expression of either the single gene or both γ–globin genes (non–deletional HPFH, ndHPFH) (Para. 4.2) (Steinberg *et al.*, 2009). Both deletional and non–deletional conditions are also referred in

OMIM database as Fetal Hb Quantitative Trait Locus 1 (**HBFQTL1**, entry #141749).

Furthermore, polymorphic configurations within the β–cluster (*in cis*) have been associated with variability in HbF levels in sickle cell anemia and β–thalassemia: they include an $(AT)_x(T)_y$ polymorphism upstream of the HBB promoter, SNPs upstream of the $^{G}\gamma$ gene (pre–$^{G}\gamma$ framework) and in $^{G}\gamma$ IVS2, and a $(AT)_xN_{12}(AT)_y$ polymorphism within the HS2 element of the β–LCR (Steinberg *et al.*, 2009).

There are many studies addressing the association of these SNPs or microsatellites with a specific β–thalassemia gene. It was shown as, sometimes, a mutation can be located on two or more different configurations and how these can act differently on the phenotype of the patient.

An example has been shown in a work by De Angioletti *et al.*, 2004: the study was carried out in 17 patients, compound heterozygotes for the β–thalassaemia mutation $\beta^{+}(-87)$ C→G, who showed two different phenotypes and two distinct haplotypes linked to the β–thalassemic mutation. In fact, the $\beta^{+}(-87)$ was found associated with both Orkin haplotype VIII ($\beta^{+}(-87)$/VIII) and V ($\beta^{+}(-87)$/V). The patients with the $\beta^{+}(-87)$/VIII showed milder clinical conditions, with significantly higher levels of Hb and HbF and higher synthesis of $^{G}\gamma$ than the patients with the $\beta^{+}(-87)$/V.

The two groups differed not only in the Orkin haplotype associated with the $\beta^{+}(-87)$ mutation, but also in the polymorphic configuration of SNPs and microsatellites on β–globin gene cluster. A schematic representation of these associations can be found in Fig. 17.

Two mechanisms could account for the increased HbF synthesis. One is present in all patients: the mutation $\beta^{+}(-87)$ *per se* reduces the binding of transcriptional factors to the β–globin promoter and makes them available for the γ–globin promoter.

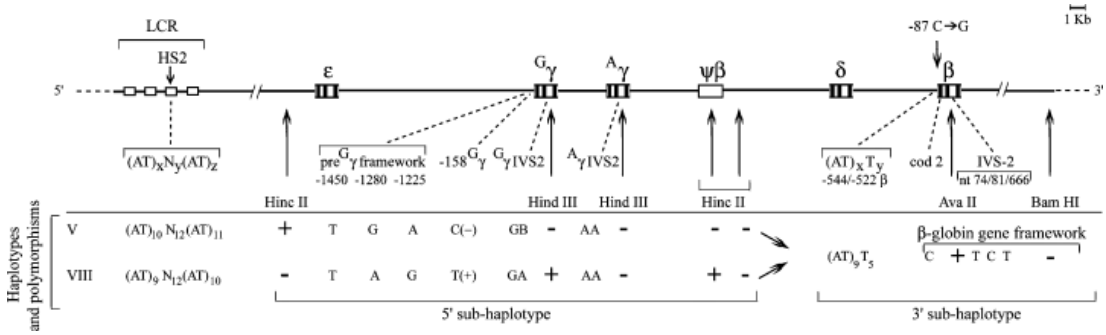


Figure 17
Schema of the cluster of the β–globin genes and of the two haplotypes and polymorphic configurations (lower line) found associated with the β^{+}(–87) C→G mutation.
The polymorphic microsatellite in the $^{G}\gamma$ and $^{A}\gamma$ IVS2 were reported as in Lapoumeroulie *et al.*, 1999: *GA = $(TG)_{11}(CG)_3$, GB = $(TG)_{19}(CG)_2CACG$, AA = $(TG)_{13}$.*

(De Angioletti *et al.*, 2004)

The other mechanism is active only in patients with $\beta^{+}(-87)$/VIII and could be ascribed to the *rare configuration of polymorphisms* typical of this chromosome. It is still unclear whether only some polymorphisms or the entire configuration are involved in the HbF increase. The high increase of the HbF, mostly of the $^{G}\gamma$–type, strongly suggest the hypothesis that the −158 $^{G}\gamma$ *Xmn* I + plays a principal role and that the other polymorphisms could exert a cooperative role in the modulation of HbF in patients with erythropoietic stress (De Angioletti *et al.*, 2004).

By the way, in none of known similar cases there is direct evidence from functional studies that different alleles at these sites bring to altered γ gene expression; it is possible that these are markers in *linkage disequilibrium* with as yet unidentified elements (Steinberg *et al.*, 2009).

It is noteworthy that the *cis*–acting variants can explain less than half of the cases of persistence of HbF: >50% of the variance is in fact unlinked to the β–globin cluster (Thein *et al.*, 2007).

The loci that increase γ gene expression are clearly not linked to the β–globin cluster as they segregate independently of the β–globin gene cluster, with genetic heterogeneity and incomplete penetrance. Examples of autosomal, sex–linked, dominant, codominant, or recessive patterns have all been reported. These are due to quantitative trait loci (QTL) localized to chromosomes 2, 6, and 8 and another that

appears to be X–linked. Additional genes are known to occur but have yet to be localized. The mechanism by which the chromosome 2–, 6–, 8–, or X–linked genes affect HbF levels and F cells remains mostly unknown. It is possible that these loci encode for *trans*–acting factors that bind within the β–globin gene cluster and affect gene transcription (Steinberg *et al.*, 2009).

4.1 dHPFH and δβ–thalassemias (HBFQTL1)

In both the $^{G}\gamma^{A}\gamma(\delta\beta)^{0}$–HPFHs (or simply deletional HPFH, dHPFH) and the $^{G}\gamma^{A}\gamma(\delta\beta)^{0}$–thalassemias (shortly δβ–thalassemias) the deletions remove both the δ and β genes and extend a variable degree 3' to the cluster, in some cases for up to 100 kb. The $^{G}\gamma(^{A}\gamma\delta\beta)^{0}$–thalassemias differ in that the 5' end of the deletion partially or totally removes the $^{A}\gamma$ gene in addition to the δ and β genes.

The broad classification of the deletion conditions into three groups, although useful, is rather arbitrary: in fact there is a continuum between δβ–thalassemia and the HPFHs and that not all conditions fit neatly into one of these groups.

dHPFH and δβ–thalassemia were originally distinguished on hematological and clinical grounds. Heterozygous δβ–thalassemia had a similar red cell picture to β–thalassemia, with hypochromic and microcytic erythrocytes, but a normal level of HbA_2 (<3.0%). In addition, there was a raised level of HbF (5%–15%) that had a heterogeneous intercellular distribution.

On the contrary, dHPFH heterozygotes, showed normal RBC indices, with an homogenous intracellular distribution.

As more and more cases and different molecular types of the two conditions were described, these differences became indistinct and it is now clear that there is considerable overlap in many of the parameters that were initially used to differentiate them.

Most deletions are due to illegitimate recombination between minimal homology of a few nucleotides at the breakpoints. Most show clean breakage and reunion, although in some cases a few "orphan" nucleotides of unknown origin have

been inserted. Several of the deletions start or end within repetitive sequence elements (Alu or LINES).

Principal δβ–thalassemia and dHPFH are summarized in the Table IV, while the summary Table V shows the hematological results and Hb analysis on the deletion conditions listed in Table IV.

It is very difficult to clarify the mechanism by which the deletions alter the expression of HbF. Several explanations have been proposed:

- **Loss of Regulatory Regions**: a regulatory region responsible for repressing γ gene expression in adult life lie between the ^{A}γ– and δ–globin. Its loss due to a deletion may cause the incapability to switch off γ genes in favor of adult genes.
- **Newly Apposed Enhancer Sequences**: there may be a positive influence of the newly apposed enhancer sequences (e.g. 3'HS1) translocated from the 3' end of the cluster to 3' of ^{A}γ gene, activating the promoter of the ^{A}γ– and also the ^{G}γ–gene. Besides, some deletions fuse the remaining γ genes with sequences close to an OR gene the cluster is surrounded by (Para. 1.3). This brings positive or negative regulatory elements normally associated with OR genes into proximity of the β–like globin genes.
- **Competition between ^{G}γ ^{A}γ and δ β**: gene expression involves interaction of the β–LCR with gene promoters; the γ genes may be in competition with the δ and β genes for expression during adult life. In normal circumstances, this competition would favor the δ and β genes as a result of the chromatin conformation or stage–specific *trans*–acting factors or both. When both the δ and β gene promoters are deleted, the γ promoters may be free to interact with the β–LCR, resulting in the persistent HbF (Steinberg *et al.*, 2009).

Some of the possible mechanisms and the involved *trans*–acting factors were discovered using a subtractive hybridization (SSH) method that identified differently expressed transcripts in reticulocytes from a normal and a dHPFH subject.

A recent study by de Andrade *et al.*, 2006 selected both HPFH–2 and Sicilian δβ–thalassemia (Fig. 18) and compared their transcriptome to normal samples through SSH.

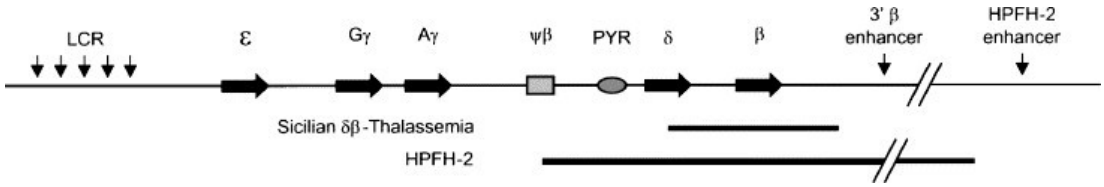


Figure 18
Representation of the human β–globin gene cluster. Black bars represent the extent of the described deletions. PYR binding region, 3' β enhancer and HPFH–2 enhancer are also indicated.
(de Andrade *et al.*, 2006)

This work found transcripts differentially expressed in both normal and in HPFH–2 or Sicilian δβ–thalassemia groups.

Among these, several chromatin remodeling factors (ARID1B, TSPYL1) and transcription factors (ZHX2) were found and analyzed:

ARID1B[2] (also known as BAF250b) was found increased in the subjects heterozygous for HPFH–2 (approximately a 5 and 2 fold change); also **TSPYL** (a member of Nucleosome Assembly Protein family involved in chromatin remodeling) was found increased.

The increased expression of ARID1B and TSPYL1 could maintain an open chromatin structure in the region comprising the fetal globin, leading to high levels of expression in the adult stage. However, their expression in Sicilian δβ–thalassemia is decreased approximately 0.15 fold, which may be a consequence of the fragments that are removed in HPFH–2 and maintained in Sicilian thalassemia, such as the PYR region, for example.

[2] ARID1B is an SWI–like protein, involved in ATP–dependent chromatin modification, and is potentially a subunit of the PYR complex, which associates with a γ–δ intergenic region in the β–like globin cluster.
This region has been implicated in γ–globin silencing during the adult stage of development; its removal in dHPFH may contribute to the reactivation of fetal globin.
The PYR complex contains Ikaros as its DNA binding subunit, proteins from the SWI/SNF transcription activator chromatin remodeling family and a member of the NuRD transcription repressor family, Mi–2. These features suggest that PYR may be an important complex involved in globin gene switching.

ZHX2 has zinc finger and homeobox domains and is categorized as a transcriptional repressor. Interestingly, it is located in a genomic region (8q24.13) in which a quantitative trait locus (QTL) associated with a heterocellular HPFH was recognized (Para. 4.6).

ZHX2 is downregulated in both HPFH and δβ–thalassemia and its expression seems to be inversely correlated to γ–globin in the HPFH samples analyzed.

The presence of transcription and chromatin remodeling factors with altered expression in HPFH may indicate that the involvement of *trans*–acting factors has a decisive role in the consequent phenotype.

A model is proposed by the authors to explain the effect of different deletions within the β–globin cluster (Fig. 19).

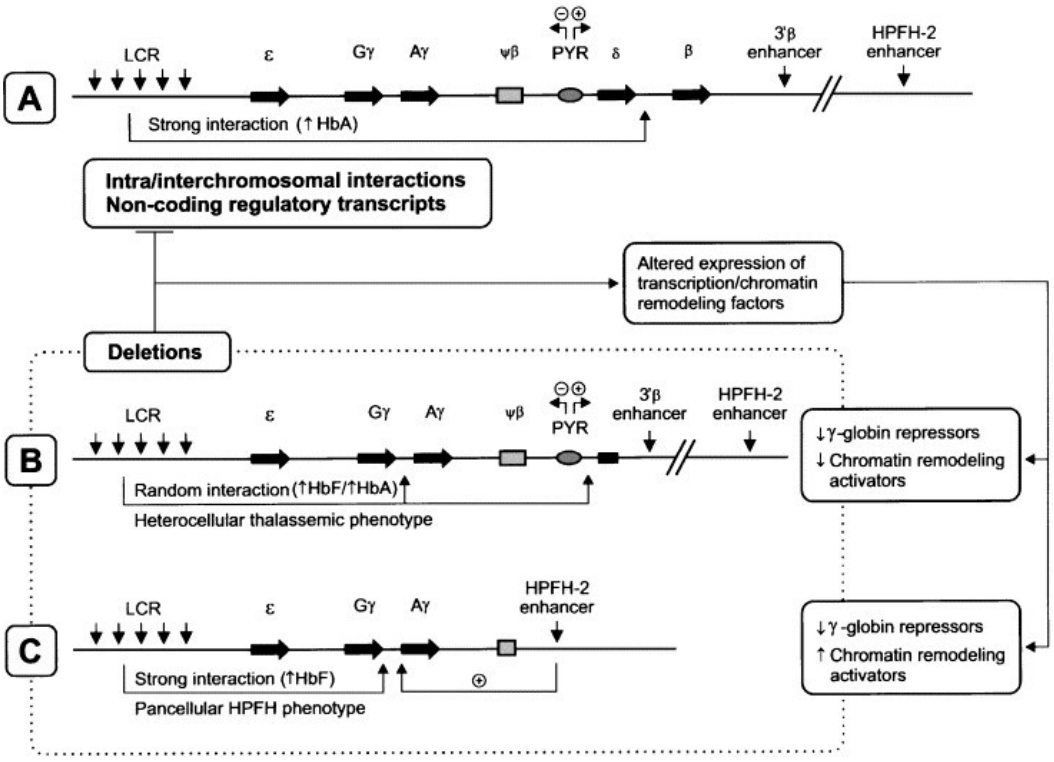


Figure 19
Model to explain the effect of different deletions within the β–globin cluster.
(de Andrade *et al.*, 2006)

Under normal conditions (Fig. 19 A), the β–LCR interacts with the δ– and β–globin genes during the adult stage of development, involving intra– and interchromosomal interactions mediated by transcription factors and non–coding regulatory transcripts, perhaps also involving the PYR complex.

Various HPFH and δβ–thalassemia deletions may interfere with the expression of *trans*–acting transcription and chromatin remodeling factors in an unknown way, leading to a disruption of interchromosomal interactions or decreased expression of regulatory non–coding RNAs.

These *cis* and *trans* alterations may alter the pattern of β–LCR, promoter, enhancer and transcription factor interactions that regulate gene expression in the β–globin cluster.

In addition, different transcription and/or chromatin remodeling factors may synergistically interact with the remaining regions in the loci to produce a random (δβ–thalassemia) (Fig. 19 B) or stable (HPFH) (Fig. 19 C) patterns of γ–globin expression (de Andrade *et al.*, 2006).

4.2 Non–Deletional Hereditary Persistence of Fetal Hb (ndHPFH) (HBFQTL1)

A lot of conditions with persistent HbF are caused by mutations on the γ genes promoters; HbF levels in these conditions vary from 3% to 35% with a preponderance of either $^{G}\gamma$– or $^{A}\gamma$–chains according to the gene promoter containing the mutation.

Analyses of the proteins that bind to the promoters have demonstrated the complexity of the region, in which binding sites for different proteins frequently overlap (Fig. 20).

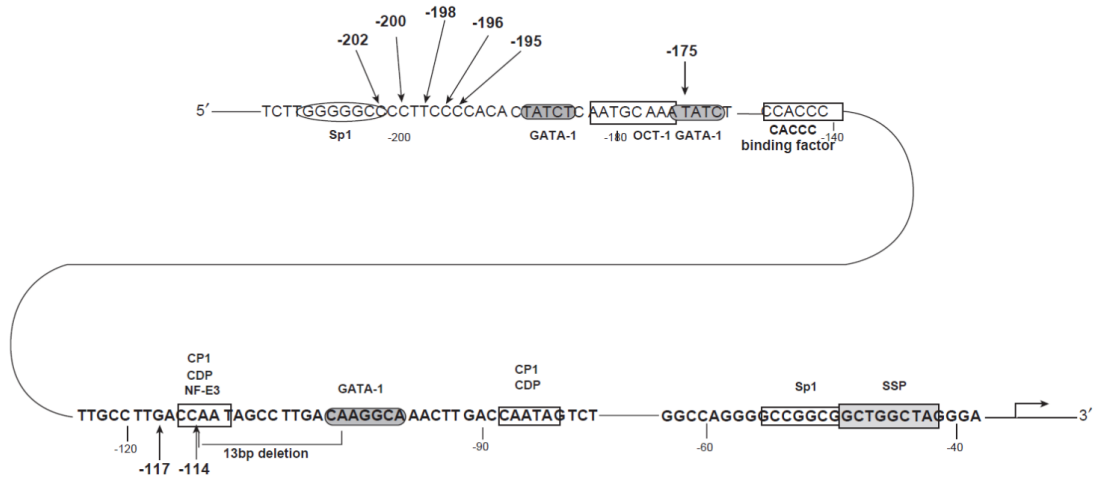


Figure 20
Transcription factor binding sites in the promoters of the γ–globin genes; $^{A}\gamma$ and $^{G}\gamma$ genes are identical in this region. The positions of mutations that result in ndHPFH are indicated.
(Steinberg *et al.*, 2009)

Altered binding of GATA1, OCT1, CP1, CDP1, NF–E3, and the stage selector protein SSP have been reported and several mutants have also been studied in transient transfection assays or *in vitro* transcription assays. Several of these mutations have been tested *in vivo* on transgenic mice and recapitulate the human phenotype to varying degrees (Steinberg *et al.*, 2009).

A lot of mutations have been described and are listed at http://globin.bx.psu.edu ($^{G}\gamma$ and $^{A}\gamma$ entries) named with the position of mutation followed by ndHPFH or simply HPFH. Principal entries are summarized in the Table VI.

Individuals with γ promoter mutations have balanced globin chain synthesis and a normal MCH (Table VI); on the other hand the mild anemia seen in compound heterozygotes for β–thalassemia and the $^{A}\gamma$ −196 and −117 mutations may indicate that the total γ + β chain output from the HPFH chromosome in these cases is slightly less than normal. Therefore it is clear that haploinsufficiency of the β gene is not completely compensated by the expression of γ genes.

Increased γ chain production is matched by a decrease in the expression of the β gene *in cis* to the γ mutation. This reciprocal relationship could be explained by the competition between these two genes for β–LCR. Altered binding of transcription

factors to the γ gene promoter as a result of the mutation might alter the association with the β–LCR, resulting in a more stable binding to γ genes than would normally occur in adult red cells.

Decreased binding of a repressor protein or increased binding of an activator to the γ gene promoter could therefore increase the gene's competitiveness. Different mutants would result in different probabilities of an interaction with the β–LCR: this could explain also the variability in the amount of HbF produced (Steinberg *et al.*, 2009).

Some ndHPFH will be discussed individually.

4.2.1 $^{A}\gamma$ −117 G→A (rs35378915)

It was first found in a Greek HPFH allele; it is also referred as the Black or Greek ndHPFH. It has been found linked to haplotype VII in Greeks and Italians (Pistidda *et al.*, 1995). Heterozygotes produce 14.2 ± 1.1 g/dl of Hb, with normal RBC indices and HbA_2. This mutation rise HbF up to 12.1 ± 2.8 % on total Hb, with the prevalence of $^{A}\gamma$ globin (93.4 ± 4.7 % of HbF) (Table VI). α/non–α ratio nearly close to 1 indicates that this mutation does not shift the balance between α and γ + β + δ globins (Steinberg *et al.*, 2009).

$^{A}\gamma$ −117 G→A locates between the two distal CCAAT elements (2 bases upstream the distal CCAAT box of HBG1) and results in increased CP–1 (NF–Y) (2–3 fold) and CDP binding (3 fold), while interfering with the binding of an erythroid cell specific factor, NF–E3 (Superti–Furga *et al.*, 1988) and GATA1 (Berry *et al.*, 1992).

Also the DR–1 binding site is disrupted by the HPFH −117 mutation reducing its γ–gene silencing function (Para. 2.1) (Steinberg *et al.*, 2009).

Two mechanisms have been proposed to explain the HPFH phenotype: the $^{A}\gamma$ −117 HPFH mutation either increases the transcription by increasing promoter strength or it allows γ gene expression in the adult by interfering with γ gene

silencing. At least in the system used *in vivo* on mice transfected with a silenced construct containing the –117 HPFH mutation, the most likely mechanism of action of this HPFH is the increase in γ gene promoter strength (Li *et al.*, 2001).

Through transient expression assay it has been defined that Greek HPFH allele produces about 1.4 times as much RNA as the wild type $^{A}\gamma$ gene when standardized against RNA transcribed from the $^{G}\gamma$ genes *in cis* (Rixon and Gelinas, 1988).

It is of interest that the activity of the δ–globin gene is consistently decreased in all types of HPFH: it can reach up to 70% decrease in homozygous Sardinian –117 HPFH patients (Ottolenghi *et al.*, 1989).

4.2.2 $^{A}\gamma$ −196 C→T (rs35983258)

It is usually referred as Italian ndHPFH; this mutation has been found in unrelated persons with the same phenotype and occurs mostly in Italy; this ndHPFH apparently arose independently since it is associated with different haplotypes (OMIM entry: 142200.0027).

It is associated with higher levels of HbF (13.7 ± 2.9 %), mostly $\alpha_2{}^{A}\gamma_2$, and normal levels of HbA_2 (Table VI).

It is possible that this mutation alters the affinity of a transacting factor. It is surrounded by a GATA1 and Sp1 binding site.

In Sardinia, the $^{A}\gamma$ −196 C→T ndHPFH mutation can be found associated with the $\beta^{0}(39)$ **C**AG→**T**AG stop codon of the HBB gene: this association is named Sardinian δβ–thalassemia due to the presence of a β–thalassemia allele and the reduced amount of δ globin (HbA_2), which would be expected at higher levels in a β–thalassemia trait. It is supposed that the reason why HBD gene is under–expressed, although the δ gene is entirely normal, is probably due to the suppressive effect of the *cis* ndHPFH mutation, which may compete for the transcription factors and β–LCR (Loudianos *et al.*, 1992).

This two closely linked mutations are expected to give selective advantage because the $\beta^{0}(39)$ mutation protects against malaria (Ayi *et al.*, 2004) while the

increased γ–globin production ameliorates the severity of the β–thalassemia (Taher *et al.*, 2009).

Two mutations close to $^{A}\gamma$ –196 have been studied *in vivo*: the British $^{A}\gamma$ –198 T→C and the Brasilian $^{A}\gamma$ –195 C→G. The two mutations are described below to show how two very adjacent mutations can act oppositely. It is unclear whether the $^{A}\gamma$ –196 acts similar to –198 or –195. However the –196 mutation is associated with high levels of HbF (15–20%), expressed in a pancellular fashion, whereas the –198 and –195 mutations are associated with lower levels of HbF (3–10%) expressed in a heterocellular fashion (Forget, 1998).

$^{A}\gamma$ –198 T→C creates a motif that can play a role similar to that of the γCACCC box, increases Sp1 binding and creates novel binding for another ubiquitous protein. The –198 HPFH mutation increases γ gene expression by 4 to 5 fold in stable or transient transfections of erythroid cell lines, while this enhancement is not observed in non–erythroid cell lines. The results of *in vivo* studies indicate that the –198 T→C HPFH mutation up–regulates γ gene transcription only in presence of a γCACCC box deleted construct, while the level of γ mRNA is similar to the controls when the mice are transfected with a non–deleted construct.

The same mechanism may account for the HPFH phenotype in humans: normally, in the adult, the CACCC box may become inactivated, contributing to the down–regulation of γ gene expression, but the –198 T→C mutation produces an alternative CACCC box that allows the interaction of the γ promoter with the LCR, resulting in the phenotype of HPFH (Li *et al.*, 2001).

On the contrary, the Brasilian **$^{A}\gamma$ –195 C→G HPFH** does not create a CACCC motif, but decreases the affinity at the Sp1 protein. In an *in vitro* study it was shown that this ndHPFH cause upregulation of the luciferase activity in erythroid cells and that the presence of HS2 is necessary for the increase in the expression of the vector,

for effective promoter action. However the enhancement produced *in vitro* is several times less than that seen *in vivo* (Takahashi *et al.*, 2003).

4.2.3 $^{G}\gamma$ −175 T→C (rs63750654)

This mutation has been found in Black, Sardinian and British persons. Data show a slightly reduced Hb (12.7 ± 1.1 g/dl) and normal RBC indices. HbF levels are among the highest for $^{G}\gamma$ ndHPFH (20.3 ± 2.8%), mostly $\alpha_2{}^{G}\gamma_2$ (Table VI) (Steinberg *et al.*, 2009).

The region where the −175 mutation is located in is noteworthy because it contains an octanucleotide sequence that is present in the promoter region of a number of genes and is the binding site for the *trans*–acting factor OCT1; in addition this octamer is flanked on either side by a consensus sequence for GATA1.

The −175 HPFH mutation affects the one nucleotide that is present in the partially overlapping binding sites of both factors (Forget, 1998), altering the interaction with GATA1 and removing the binding site for OCT1**.** However**,** the relevance of these *in vitro* effects to the HPFH phenotype remains unknown (Steinberg *et al.*, 2009). In transient and stable transfection assays, the −175 HPFH mutation augments the strength of the γ promoter ~4 fold. The −175 HPFH mutation also decreases binding of HMG1, a ubiquitous chromatin structural protein, but the physiological meaning of this decline is unclear (Li *et al.*, 2001).

4.2.4 $^{G}\gamma$ −158 C→T (rs7482144)

Within the β–cluster there are numerous polymorphisms affecting HbF production in many different ways. The best known of these is the $^{G}\gamma$ −158 C→T, which creates an *Xmn* I restriction site. This variant, formerly grouped as ndHPFH, have now to be reclassified based on the new findings.

Normal individuals who are homozygous for T (*Xmn* I +/+) may have a slight increase in F cells, and it has been associated with increased HbF production under conditions of erythropoietic stress. In fact, among SCD, HbE and β–thalassemia

patients, those with −158 T tend to have higher HbF levels and a higher proportion of $^{G}\gamma$ chains.

The cause of this differences in phenotype is unknown as no transcription factors seems to bind there, but it may be related to the different chromosomal backgrounds and mutations on other genes involved in Hb switch which are in *linkage disequilibrium* with this SNP (Steinberg *et al.*, 2009).

To understand whether *Xmn* I SNP is a causal variant at the β–globin locus or not, the association between *Xmn* I SNP and HbF levels was replicated through GWAS in African–Americans with SCD.

However, rs10128556 (the *Hinc* II G→A (–/+) polymorphism on HBPB1) was found more strongly associated with HbF levels than *Xmn* I by two orders of magnitude. So, when conditioned on *Hinc* II SNP, the HbF association result for rs7482144 is not significant, indicating that *Xmn* I is not a causal variant for HbF levels in the population analyzed (Galarneau *et al.*, 2010).

4.3 Modern approaches for the identification of genetic modifiers of β–thalassemia phenotype

Elucidation of the just described *cis*–acting genetic modifiers has not been too difficult as these loci have a major clinical effect and the genetic variants are common; however, these modifiers do not explain all the clinical heterogeneity and many important questions remain regarding the role of other genetic modifiers and their interactions with the β–globin gene cluster.

Due to the complex interaction of the multiple factors, a genetic approach might be the most efficacious way of dissecting the molecular mechanism and identifying the modifier genes.

The modern approach involves association studies, which looks for differences in the frequencies of genetic variants between ethnically matched cases and controls to find variants that are strongly associated with the disease or trait (in this case, HbF production). If a variant is more common in cases than controls, then association can be inferred.

Until recently, association studies have been based on candidate genes selected according to researchers understanding of the disease pathophysiology; however guessing the causal genes could well leave out some of the key genetic players.

In contrast, **genome wide association studies** (**GWAS**) scans the whole human genome and is able to reveal the unsuspected interactions. It is generally accepted that any two genomes are 99.9% identical and GWAS looks for differences between different genomes to identify the variants that are more common in the cases when compared to case controls.

The genome is thus portrayed as a series of high *linkage disequilibrium* (LD) regions named **haplotype blocks**, separated by short segments of very low LD referred to as **recombination hotspots**.

Nevertheless GWAS may tell if a locus is important in the pathogenesis of a condition but it does not identify the causal variants and the mechanism that is responsible (Thein, 2008).

4.4 HBFQTL2: Chromosome 6–linked

(HMIP 6q22.3–q23.1) (OMIM #142470)

One of the first reports of HPFH unlinked to the β–globin gene cluster was done by Thein *et al.*, 1994: she reported a large Indian family with β–thalassemia and/or HPFH in which HbF segregated as an independent trait from β–thalassemia. The locus linked to the HbF production was found two years later and mapped on chromosome 6q22.3–q23.1 (Craig *et al.*, 1996), which was further refined to a 1.5–Mb region containing five coding genes (AH11, MYB, ALDH8, HBS1L, and PDE7B) and four non–coding RNAs (Close *et al.*, 2004; Steinberg *et al.*, 2009).

A systematic investigation of genetic variants between HBS1L–MYB was then made by resequencing this 125–kb region by using DNA from 32 European control subjects. It identified the principal genetic variants that account for the chromosome 6q QTL for HbF. These variants are located in HBS1L–MYB intergenic polymorphism (**HMIP**) block (Thein *et al.*, 2007).

GWAS and sequence analysis of this region allowed for the identification of many SNPs correlated with HbF production: rs9376092, rs9399137 (Nuinoon *et al.*, 2010), rs4895441 (Lettre *et al.*, 2008), rs9389268 (Galanello *et al.*, 2009), and so on.

A recent study provides the functional link between genetic association of HMIP with control of fetal Hb and other hematologic parameters. It was suggested that HMIP contains regulatory sequences that could be important in hematopoiesis by controlling MYB expression: in fact a study identified several potential *cis*–regulatory elements as strong GATA1 signals that coincided with the DNase I hypersensitive sites present in MYB–expressing erythroid cells. Therefore, this study provides the functional link between genetic association of HMIP with control of fetal Hb and other hematologic parameters (Wahlberg *et al.*, 2009).

These information were confirmed by another study in which it is shown that MYB gene is causally involved in controlling HbF production at the HBFQTL2 locus (Galarneau *et al.*, 2010)

4.5 HBFQTL3: Chromosome X–linked

(Xp22.2) (OMIM #305435)

Analysis of HbF levels in healthy Japanese adults led to the suggestion that an X–linked gene was involved in increasing HbF levels with a frequency of 11% of males affected and 21% of females being carriers. Further studies confirmed a linkage between F cell production locus (mapped at Xp22.2–22.3) and HbF production (Steinberg *et al.*, 2009), mediated through a not yet defined *trans*–acting mechanism.

4.6 HBFQTL4: Chromosome 8–linked

(8q) (OMIM #606789)

In the same large Indian kindred in which it was mapped a gene modifying HbF production to chromosome 6q23 (Craig *et al.*, 1996; Thein *et al.*, 1994), it was identified another quantitative trait locus (QTL) on 8q conditional on the $^{G}\gamma$ *Xmn* I

polymorphism. The results indicated that an interaction between the $^{G}\gamma$ *Xmn* I site and a QTL on 8q influence the production of fetal Hb (Garner *et al.*, 2002).

4.7 HBFQTL5: Chromosome 2–linked and Role of BCL11A

(2p5) (OMIM #142335)

GWAS mapping of individuals with HbF made it possible to find a quantitative trait locus influencing F cell production on chromosome 2. It maps to a zinc–finger protein, BCL11A, previously implicated in leukemias (Steinberg *et al.*, 2009). The strongest association was with the SNP **rs11886868** (C allele) in intron 2 of the BCL11A gene. The C allele was also associated with an ameliorated phenotype in patients with β–thalassemia and SCA: this may indicate that this SNP may modify these phenotypes by increasing HbF levels (Galanello *et al.*, 2009; Uda *et al.*, 2008).

Other SNPs associated with HbF were detected: rs766432 (Nuinoon *et al.*, 2010), rs4671393, rs7557939 (Lettre *et al.*, 2008).

The variants characteristic of higher levels of HbF are associated with lower levels of BCL11A mRNA expression. This variation in expression appears to have an effect on the function of BCL11A at the human β–globin locus, where it is thought to act with the erythroid transcription factors GATA1 and FOG1, as well as the NuRD remodeling and repressor complex to silence the γ–globin genes (Fig. 21) (Sankaran *et al.*, 2010).

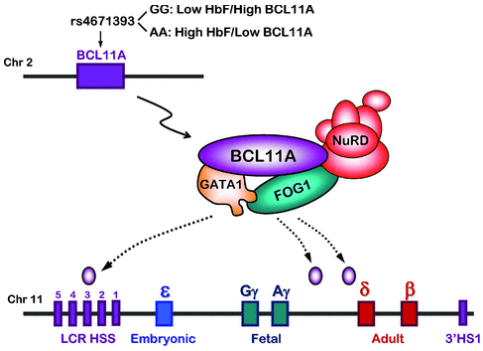


Figure 21
BCL11A role on the human β–globin locus.
(Sankaran *et al.*, 2010)

Chromatin immunoprecipitation (ChIP) experiments showed that BCL11A directly binds to several locations of the β–globin locus in adult erythroid progenitor cells; one of these is the GGCCGG motif at position −56 to −51 of the $^{G}\gamma$ gene promoter (Steinberg *et al.*, 2009).

Some interacting partners of BCL11A have been already discovered: the lysine–specific demethylase 1 and repressor element–1 silencing transcription factor corepressor 1 (**LSD1/CoREST**) histone demethylase complex interacts with BCL11A and is required for full developmental silencing of mouse embryonic β–like globin genes and human γ–globin genes in adult erythroid cells *in vivo*.

Furthermore, the DNA methyltransferase 1 (DNMT1) is identified as a BCL11A–associated protein in the proteomic screen and it is required to maintain HbF silencing in primary human adult erythroid cells (Xu *et al.*, 2013).

4.8 HBFQTL6: Chromosome 19–linked and Role of KLF1

(19p13.2) (OMIM #613566)

HPFH was found in 10 of 27 members from a Maltese family in which HbF ($\alpha_2^G\gamma_2$ and $\alpha_2^A\gamma_2$) ranged from 3.3 to 19.5% of total Hb, displaying mild hypochromic microcytic indices displayed by the HPFH individuals. Through a genome–wide SNP scan followed by linkage analysis a candidate region on chromosome 19p13.12–13 was found. Sequencing revealed a nonsense mutation in the KLF1 gene, which ablated the DNA–binding domain of this key erythroid transcriptional regulator. Deregulation of KLF1 could explain the altered erythroid indices.

Expression profiling on primary erythroid progenitors showed that KLF1 target genes were downregulated in samples from individuals with HPFH. Functional assays suggested that, in addition to its established role in regulating adult globin expression, KLF1 is a key activator of the BCL11A gene, which encodes a suppressor of HbF expression (Borg *et al.*, 2010).

The new studies link HPFH, KLF1 and BCL11A together in a unified model. In fact KLF1 indirectly regulates γ–globin expression by directly regulating BCL11A and globin switching. During embryonic and fetal development or in KLF1 haploinsufficient adults, KLF1 levels are low, resulting in low levels of adult β–globin and BCL11A and high levels of γ–globin.

In adults with two functional copies of KLF1, increased expression of KLF1 in definitive red blood cells promotes high levels of β–globin and BCL11A expression, which in turn represses γ–globin expression (Fig. 22) (Bieker, 2010).

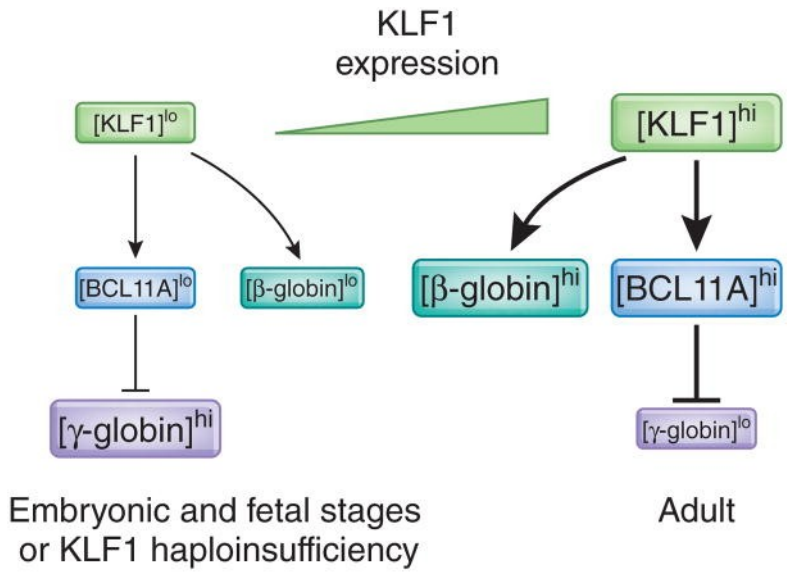


Figure 22
KLF1 role in globin switching.

(Bieker, 2010)

AIM OF THIS STUDY

β–thalassemia affects millions of individuals worldwide. These Hb disorders are characterized by extreme clinical heterogeneity, complicating patient management and treatment. A better understanding of this clinical variability would dramatically improve care and might also guide the development of novel therapies (Lettre, 2012).

HbF has been shown as a common and major modifier of disease severity in individuals with β–thalassemia. γ–globins in fact reduce the relative excess of α–globin and provide a potent selective survival advantage for cells making HbF in the context of the ineffective erythropoiesis characteristic of the most severe forms of β–thalassemia.

Certain β–globin promoter mutations (e.g. $\beta^{+}(-87)$) are associated with increased γ–chain synthesis from the same chromosome; also coinheritance of an HPFH gene is associated with higher HbF and gives advantage in presence of β–thalassemia. In addition to SNPs within the β–globin locus, genome–wide association studies (GWAS) have identified additional loci that can contribute significantly to the variation in HbF levels (Nienhuis and Nathan, 2012).

In the Sardinian population, the incidence of β–thalassemia trait is very high (approximately 12.5%); among these, 95% carry the $\beta^{0}(39)$ nonsense mutation and about 2% the $\beta^{0}(-A)6$ frameshift mutation (Masala *et al.*, 1988; Rosatelli *et al.*, 1987; Rosatelli *et al.*, 1992). In the Sardinian population $\beta^{0}(39)$ mutation was found associated with five different Orkin haplotypes: II (64%), I (22%) and IX (4%) are the most common (Masala *et al.*, 1988; Piras *et al.*, 2005; Pirastu *et al.*, 1987).

In Sardinia, homozygous $\beta^{0}(-A)6$ or compound heterozygous $\beta^{0}(-A)6/\beta^{0}(39)$ show TI forms more frequently than TM.

In homozygous $\beta^{0}(39)$–thalassemia TM is the most frequent phenotype, although rare TI can be found (10%) (Galanello *et al.*, 1989).

HbF, when present, is the only useful Hb tetramer synthesized and this is mostly the principal cause of the intermediate phenotype.

This research intends to increase knowledge about the regulation of γ–globin genes, with particular attention to the switch fetus→adult.

It may give support to the current pre– and post–natal diagnosis and therapeutic approach of β–thalassemia in order to prevent unnecessary and harmful transfusion regimes.

The molecular mechanisms undergoing the fetus→adult switch were investigated in healthy individuals showing persistence of HbF, both in the presence and absence of β–thalassemia trait. In fact, these individuals are an excellent natural model for the study of the mechanisms that lead to a failure in the fetal genes switching off.

Samples were analyzed to: define the qualitative and quantitative composition of the Hb; identify potential β–thalassemia mutations, distinguish ndHPFH from dHPFH; study SNPs and microsatellite within the β–gene or the entire cluster and possibly relate the persistence of HbF with the presence of specific polymorphic configurations.

ndHPFH promoters were also evaluated in *in vitro* reporter assays to quantify the enhancement of each ndHPFH mutation with respect to the wild type sequence.

MATERIALS AND METHODS

5. CHOICE OF THE SAMPLES

Whole–blood samples were collected using EDTA as anticoagulant at the Blood Transfusion Centre of ASL N. 2 of Olbia.

All subjects have released informed consent.

The samples were selected on the basis of the results obtained by means of CE–HPLC and hematological data collected by the Blood Transfusion Centre of Olbia.

Only samples with HbF levels above 5% were included in this research, in order to keep out heterocellular HPFH (hHPFH) (Chap. 4).

6. PREPARATION OF THE SAMPLES

One aliquot of the whole blood (Para. 6.1) was used for biochemical analysis (Chap. 7), the other one (Para. 6.2) for DNA extraction and consequent molecular analysis (Chap. 8).

6.1 Preparation of Hemolysates

The first blood aliquot was washed through centrifugation at 1300 g for 5' in physiological solution (NaCl 0.9%) until the supernatant appeared clean and clear. The addition of one volume of ipotonic solution (tridistilled H_2O, mQ) causes the hemolysis and the consequent suspension of Hb (hemolysate), directly used for electrophoretic analysis.

Chromatography required an additional step to remove cell membranes: half of the volume of CCl_4 was added to the solution. After centrifugation (20900 g for 2'), the supernatant was recollected and contained Hb in water could be readily used for HPLC analysis or stored at –20°C for a long time.

Both Hb solutions were quantified by spectrophotometric analysis at 576 nm.

6.2 DNA extraction and quantification

DNA extraction was performed either with commercial kits (GenElute™ Blood Genomic DNA Kit, Sigma) or Salting out (Miller *et al.*, 1988). The salting out protocol, based on Miller *et al.*, 1998, can be found at the end of this Thesis (Protocol 1).

DNA was quantified with NanoDrop 8000 Spectrophotometer (Thermo Scientific).

7. BIOCHEMICAL ANALYSIS OF THE SAMPLES

At the biochemical level, unambiguous Hb phenotype determination requires the combination of electrophoretic and chromatographic techniques: cation exchange high performance liquid chromatography (**CE–HPLC**) and isoelectric focusing (**IEF**) were used to separate Hb tetramers, while denaturing polyacrylamide gel electrophoresis (**AUT–PAGE**) and Reverse Phase–HPLC (**RP–HPLC**) made it possible to distinguish the constituent globin chains.

Because of its powerful resolution, IEF can resolve proteins that differ in pI by as little as 0.01. Globin chains with a neutral–to–neutral amino acid substitution can be separated by AUT–PAGE and, especially, RP–HPLC.

7.1 CE–HPLC

Cation–Exchange High Performance Liquid Chromatography (CE–HPLC) was performed by the staff of Blood Transfusion Centre of Olbia on the Bio–Rad Variant automated analyzer with the "β–thalassemia short" program.

It has been designed to separate and determine in 5–6 min the area percentages for HbA_2 and HbF and to provide *qualitative* and *quantitative* determinations of normal and abnormal Hbs (Riou *et al.*, 1997).

7.2 IEF

Isoelectric focusing was performed on a polyacrylamide slab gel (%T:%C = 5.2%:3%) where a pH gradient is formed through ampholytes; proteins move until the position in the gel at which the pH is equal to the pI is reached and there "focus".

The gel was prepared in a sandwich of treated glass: the mold was brushed with a "Repel solution" (5% dimethyldichlorosilane, in chloroform); the support instead was treated with a "Bind solution" (0.2% Silane A174 in acetone).

The mixture for the gel was prepared by adding ampholytes at pH 6.7 to 7.7 (6.2%), ampholytes at pH 3–10 (0.5%), TEMED (0.15%) and ammonium persulfate (APS, 0.56 mg/ml), and it was casted into the prepared sandwich for 1 h at room temperature.

At this point a pre–run was necessary to form the pH gradient: strips of absorbent paper were wetted with the specific electrolyte (40 mM glutamic acid for anode, 0.1 M NaOH for cathode) and placed onto the gel. A voltage of 400 V at 4°C was applied to the gel until the amperage went below 4 mA.

20 µg of Hb were loaded and run for 2 h at 1600 V at 4°C.

At the end of the run, the Hbs were fixed in Tricloracetic Acid (20% TCA) for 10'; the gel was then colored for 30' in a "colorant solution" (0.1% BBF, 50% EtOH, 5% Acetic Acid) and decolored in the same solvents (30% EtOH, 6% Acetic Acid) (Masala and Manca, 1991).

Yellow colored bands were turned into blue by soaking the gel into water: this allowed a semi–quantitative densitometric analysis at 600 nm.

7.3 Preparative IEF: HbF enrichment

In adults, HbF concentrations are sometimes too low for the quantification of $^{G}\gamma/^{A}\gamma$ ratio by means of RP–HPLC. It is therefore necessary to perform a preliminary step to enrich the sample in HbF.

Minimum 4 µg of HbF need to be injected in RP–HPLC to be detectable. To be sure to recover a sufficient quantity of HbF, at least 30 µg of HbF were loaded on IEF gel.

45 µl of concentrated hemolysate (depending on the concentration of %HbF obtained by CE–HPLC) were placed on a strip of absorbent paper (30x5 mm) lying on the focusing gel.

After the run the HbF band was excided from the gel and suspended in 200 µl circa of mQ.

Subsequent cycles of freezing and thawing allowed all the HbF to pass from gel to water; centrifugation at top speed for 30' and filtration of the supernatant produced a RP–HPLC injectable sample enriched in the HbF fraction (Manca and Masala, 1990).

7.4 AUT–PAGE

Polyacrylamide Gel Electrophoresis in presence of Acetic Acid, Urea and the non–ionic detergent Triton X–100 (AUT–PAGE) separates denatured globin chains by their net positive charge and hydrophobicity.

The polyacrylamide gel (%T:%C = 12%:0.6%) mixture was composed of 6M Urea, 4.5% Acetic Acid, 1.8% Triton X–100, 0.36% TEMED, 0.4 mg/ml Riboflavin. The gel (160x200x0.75 mm) was casted under UV for 1h.

Each sample was denatured in a "denaturing solution" (8M Urea, 10% Acetic Acid, 10% β–mercaptoethanol) for at least 30' at room temperature and 15 µg of denatured sample were loaded into the well. The tracking dye (pyronin–G) was added to the denaturing solution to follow the run.

The run was done on PROTEAN II xi Cell (Bio–Rad) in 5% Acetic Acid, for 19 hours at 4°C, with constant amperage (6 mA for one slab, 10 mA for two).

After the run, the gel was colored and fixed with a "colorant solution" (0.5% Coomassie Brilliant Blue R–250, in 30% methanol and 7% Acetic Acid) and decolored in the same solvents (Manca *et al.*, 1986).

7.5 RP–HPLC

Reversed Phase–High Performance Liquid Chromatography (RP–HPLC) was performed on a 10°C thermostated Vydac C4 column (4.6x250 mm, ⌀ 300 Å) with a linear gradient of acetonitrile:water; the more the proteins are hydrophobic, the more they are retained by the column.

The mobile phase was composed of two buffers: A) 20% acetonitrile CH_3CN, 0.1% trifluoroacetic acid TFA; B) 60% CH_3CN, 0.1 TFA (Masala and Manca, 1994).

Hemolysates, deprived of cell membranes (Para. 6.1) were diluted to 0.1% Hb with mQ and 20 µl were injected into the column.

The run conditions were adjusted according to the needs, in terms of linear gradient of %B and time ($\Delta = \frac{\%B_2 - \%B_1}{t} \approx 0.17$) and flow (1 to 1.2 ml/min).

By means of RP–HPLC, relative concentrations of β–globins respect to the non–α globins and the $^{G}\gamma/^{A}\gamma$ ratio were calculated.

8. MOLECULAR ANALYSIS OF THE SAMPLES

Through the information obtained from electrophoresis and chromatography, direct DNA sequencing of specific PCR products was performed in order to identify the abnormality involved, which may consist in single point mutations or microinsertions/deletions. The search for large deletions (e.g. dHPFH) is facilitated by the **GAP–PCR** and multiplex ligation–dependent probe amplification (**MLPA**).

8.1 PCR

Polymerase Chain Reaction (PCR) was used to amplify DNA fragments for sequence reaction.

Except for some oligonucleotide primers already published, the others were designed using the online software Web Primer (http://www.yeastgenome.org/cgi–bin/web–primer) (See Table VII). All the position, when not explicit, are referred to CAP site of the interested gene.

Each reaction was performed with *Termus aquaticus* (*Taq*) DNA Polymerase (Sigma) for the subsequent sequence analysis or *Pyrococcus furiosus* (*Pfu*) Polymerase (Life Technologies) for cloning.

Reactions were carried out on a Thermal Cycler (Gene Amp PCR System 2700 and 2400, Applied BioSystem). All reactions were commenced by a Hot Start.

To improve throughput and specificity, Betaine 1 M (Sigma) was sometimes added to the reaction. Betaine acts as a PCR enhancer by reducing the formation of secondary structure caused by GC–rich regions (Henke *et al.*, 1997) and eliminating the base pair composition dependence of DNA melting (Rees *et al.*, 1993).

Complete $^{G}\gamma$ and $^{A}\gamma$ gene

Complete $^{G}\gamma$ and $^{A}\gamma$ genes were amplified with the same Chemical and Thermal protocol, except for the primer pairs.

- R160–Gγ2 primers for $^{G}\gamma$ gene
- R159–Aγ10 primers for $^{A}\gamma$ gene

Chemical protocol: 1X Buffer, 2 mM $MgCl_2$, 375 µM dNTPs, 0.3 µM of each primer, 1M Betaine, 0.2 µg of genomic DNA, 2.5 U of polymerase.

Thermal file: 3’ at 95°C, 35 cycles of 1’ at 95°C, 1’ at 52°C and 3’ at 72°C.

***Eminested and nested PCR for γ genes**:*

The larger and low–throughput fragments obtained by amplification of complete γ genes (See Above) were used for eminested and nested PCR to increase the quantity of DNA for sequence analyses.

Chemical Protocols and Thermal files were optimized to be identical for all the fragments, except for the primer pairs:

Chemical Protocol: 1X Buffer, 2 mM $MgCl_2$, 250 µM dNTPs, 0.3 µM of each primer, 0.3 µg of genomic DNA, 2.5 U of polymerase.

Thermal file: 3’ at 95°C; 35 cycles of 1’ at 95°C, 45’’ at 50°C and 3’ at 72°C; additional final extension was added (4’ at 74°C).

***$^{G}\gamma$ gene eminested** from $^{G}\gamma$ R160–Gγ2 fragment:*

- **R160–R161** primers amplify $^{G}\gamma$ promoter (from –623 to +53, in 5'UTR) and were used to identify $^{G}\gamma$–ndHPFH mutations. Through this fragment the $^{G}\gamma$ −158 C→T (*Xmn* I polymorphism) was also determined.
- **Gγ1–Gγ2** primers amplify 780 bp (from +945 in IVS2 to +1726 in 3'UTR). This fragment was used to define the *Hind* II SNP of Orkin haplotype with sequencing primer AGγ8.

***$^{A}\gamma$ gene eminested** from $^{A}\gamma$ R159–Aγ10 fragment*

- **R159–R161** primer pair amplifies $^{A}\gamma$ promoter (from –622 to +53, in 5'UTR) and was used to identify $^{A}\gamma$–ndHPFH mutations.
- **Gγ1–Aγ10** primers amplify 989 bp (from +941 in IVS2 to +1929 in 3'UTR). This fragment is also used to define the *Hind* II SNP of Orkin haplotype with sequencing primer AGγ8.

Complete β gene

The complete β gene was directly amplified from genomic DNA with 3 different PCRs, each with its own Chemical and Thermal protocol.

- **P1–P2** primers amplify a 706 bp region (–140 to +566), encompassing a region from the proximal promoter to the beginning of IVS2.

Chemical Protocol: 1X Buffer, 3.5 mM $MgCl_2$, 250 μM dNTPs, 0.3 μM of each primer, 0.3 μg of genomic DNA, 2.5 U of polymerase.

Thermal file: 3' at 95°C; 35 cycles of 1' at 95°C, 45'' at 65°C and 1' at 72°C; additional final extension was added (4' at 74°C).

- **β7–β8** primers amplify a 923 bp region (+333 to +1255), including part of exon 2 and IVS2. This fragment contains also the *Ava* II polymorphism for Orkin haplotypes.

Chemical Protocol: 1X Buffer, 3 mM $MgCl_2$, 250 μM dNTPs, 0.3 μM of each primer, 0.3 μg of genomic DNA, 2.5 U of polymerase.

Thermal file: 3' at 95°C; 35 cycles of 1' at 95°C, 1' at 60°C and 1' at 72°C; additional final extension was added (4' at 74°C).

- **β9–β10** primers amplify a 956 bp region (+799 to +1764), from IVS2 to 5'UTR of β–gene.

Chemical Protocol: 1X Buffer, 1.5 mM $MgCl_2$, 250 μM dNTPs, 0.3 μM of each primer, 0.3 μg of genomic DNA, 2.5 U of polymerase.

Thermal file: 3' at 95°C; 35 cycles of 1' at 95°C, 1' at 55°C and 1' at 72°C; additional final extension was added (4' at 74°C).

SNPs and microsatellites on the β cluster

To define both Orkin haplotypes (Para. 3.1.2) and De Angioletti configurations (Chap. 4), different PCRs were set for each SNP or microsatellite.

All the PCRs contained 1X Buffer, 250 μM dNTPs, 0.3 μM of each primer, 0.3 μg of genomic DNA, 2.5 U of polymerase. $MgCl_2$ was used at different concentrations (from 1.5 to 4 mM).

Thermal file was identical for all the PCR except for the annealing temperature (T_a): 5' at 95°C; 35 cycles of 1' at 95°C, 1' at T_a and 1' at 72°C; additional final extension was added (4' at 74°C).

$MgCl_2$ and annealing temperature can be found below.

- **HS2_F–HS2_R** [3.5 mM $MgCl_2$, T_a=57°C]: these primers amplify a fragment that contains the microsatellite $(AT)_xN_y(AT)_z$ in β–LCR–HS2.
- **5'ε_F–5'ε_R** [2.5 mM $MgCl_2$, T_a=51°C]: this pair isolate a region upstream to the ε gene in which a *Hinc* II restriction (GTC**AA**C) site is present (rs3834466, AA/A +/−).
- **pre_Gγ_F–pre_Gγ_R** [2.5 mM $MgCl_2$, T_a=55°C] primers amplify a fragment in the distal promoter of $^{G}\gamma$ gene, containing the polymorphisms at −1450 T/G, −1280 G/A and −1225 A/G of the so called pre–$^{G}\gamma$ framework.
- $^{G}\gamma$ −158 C→T (*Xmn* I polymorphism), $^{G}\gamma$ and $^{A}\gamma$ IVS2 microsatellites, $^{G}\gamma$ and $^{A}\gamma$ Hind III: see Para. 8.1, "*Eminested and nested PCR for γ genes*"

- **pseudo_β_F–pseudo_β_R** [1.5 mM $MgCl_2$, T_a=55°C] primers identify the *Hinc* II restriction site (GTTG**A**C) on the ψβ1 gene (rs10128556 G/A (−/+)).
- **3'pseudo_β_F–3'pseudo_β_R** [3.5 − 4 mM $MgCl_2$, T_a=55°C] isolate a region ~2.5 kb downstream the ψβ1 gene in which a *Hinc* II (GTTA**A**C) restriction site can be present: the polymorphism (T/A −/+) is located at −5003 bp to the CAP of δ gene.
- **pro_β_F– pro_β_R** [3.5 mM $MgCl_2$, T_a=51°C] primer pair amplifies a portion of the distal promoter region of β–globin gene where $(AT)_xT_y$ motif is present.
- **3'β_Bam_F–3'β_Bam_R** [2.5 mM $MgCl_2$, T_a=51°C] fragment locate 8.2 kb circa downstream the β–globin gene, in which a *Bam* HI restriction site (**G**GATCC) is present (G/A +/−).

8.2 MLPA

MLPA (Multiplex Ligation–dependent Probe Amplification) is a multiplex PCR method detecting abnormal copy numbers of up to 50 different genomic DNA or RNA sequences, which is able to distinguish sequences differing in only one nucleotide (Schouten *et al.*, 2002).

MLPA probes hybridize to the target, are legated together to form a ssDNA and are amplified by a single PCR primer pair. The resulting amplification produces marked fragments from 130 and 480 nt in length and can be analyzed by capillary electrophoresis.

Comparing the peak pattern obtained to that of reference samples indicates which sequences show aberrant copy numbers (http://www.mrc–holland.com).

MLPA P102 HBB probemix (MRC–Holland) was used for copy number quantification when a unknown deletion/duplication was suspected on the β–cluster.

MLPA reaction was performed according to the provided protocol, by using 100 ng of DNA and a 20 h hybridization step.

Data were analyzed with the **Coffalyser.Net** MLPA data analysis software available at the manufacturer's website.

8.3 GAP–PCR

When the breakpoint sequences are known, GAP–PCR can readily identify large deletions. It consist in a multiplex PCR with a common forward primer and two different reverse primers or vice versa: one pair flanks the deletion and produces a PCR product only when the deletion is present; the other one is used as control and amplifies the non–deleted chromosome.

Sicilian δβ–thalassemia was confirmed thanks to the GAP–PCR method using F1, F2 and F3 primers. F1–F2 fragment (1585 bp) is produced when a non–deleted allele is present; F1–F3 fragment (1150 bp) is only amplified when the Sicilian δβ–thalassemia allele is present.

PCR was performed in a 25 µl reaction volume containing 1X Buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 10 pmol of each primer, 100 ng of DNA, 0.5 U of *Taq* Polymerase (Sigma).

Thermal file: 4' at 94°C; 30 cycles of 1' at 94°C, 1' at 60°C and 2' at 72°C; additional final extension was added (4' at 72°C) (Craig *et al.*, 1994).

8.4 PCR fragment purification and quantification

After the amplification, all the reaction volume was electrophoresed through a 1–1.2% agarose, 1X TAE, etidium bromurated gel at 7.5 volts/cm for 45 hour in presence of a molecular weight marker (Marker III, VI and 100 bp, Roche).

Bands were visualized on an UV transilluminator, quantified by comparison with the marker and excised with sterile scalpel.

DNA was recovered from agarose by means of the Montage Gel Extraction Kit (Millipore).

8.5 Sequence Reaction

The sequence reaction was performed by using the BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies). This kit is based on Sanger chain–terminating method (Sanger *et al.*, 1977) and each terminator or

dideoxyribonucleotide (ddNTP) is associated with a specific fluorochrome. Given that all 4 ddNTPs have different fluorochrome, the reaction can be performed in a single tube.

The 20 µl reaction volume mixture consisted of: 2 µl of 5X Buffer, 2 µl of Mix, 3.2 pmol of primer, 1–2 ng of PCR product per 100 bp.

Sequence primers were almost always the same ones used in PCR; sometimes, because of the length of the fragment or other necessities, specific sequence primers were designed.

Thermal file: 1' at 96°C, 25 cycles of 10'' at 96°C, 5'' at 50°C and 4' at 60°.

Excess of unincorporated dye terminators and other contaminants (unincorporated primers, salts, and so on) were completely removed prior to capillary electrophoresis through the Sigma Spin Post–Reaction Clean–Up Columns (Sigma).

Purified sample was mixed with 2 volumes of formamide and injected on a 3130 Genetic Analyzer (Life Technologies) capillary electrophoresis.

8.6 Sequence Analysis

Sequence analysis was performed with the help of two software: Geospiza FinchTV (http://www.geospiza.com) and Another Plasmid Editor or APE (http://biologylabs.utah.edu/jorgensen/wayned/ape/).

FinchTV allowed for the reading and editing sequence files, while APE was used to align the obtained sequences with the reference β–globin cluster (AC #: U01317).

9. EXPRESSION STUDIES ON γ PROMOTERS

The strength of each identified ndHPFH promoter was compared to the wild type promoter. To do so, both wild type and mutated promoters were cloned into a reporter plasmid whose *in vitro* expression was measured. γ promoter cloning and transfection experiments were based on the work by Chen *et al.*, 2008.

9.1 Promoter isolation from genomic DNA

Both $^{G}\gamma$ and $^{A}\gamma$ wild type and ndHPFH promoters were isolated from genomic DNA of the samples who carried the ndHPFH mutations.

To do so a PCR with the *Pyrococcus furiosus* polymerase (**Pfu**) was performed: it possess the 3'→5' exonuclease (proofreading) activity which gives a much more accurate and high–fidelity synthesis of DNA compared to the *Taq* polymerase.

Primers were designed by adding restriction sites to the 5' termini of the oligonucleotides used for PCR. These primer–specific restriction sites are transferred to the termini of the target DNA during amplification and can then be cleaved with appropriate restriction enzymes to generate amplified segments of DNA with cohesive termini ready to be cloned in a plasmid.

Moreover primers were designed to create a different restriction site on each end of the amplified DNA to obtain a directional cloning of the fragment.

To provide an adequate toehold for the restriction enzymes a tail of 3 or 4 Adenosine (A) was added (Sambrook and Russel, 2001).

Restriction sites for the chosen enzymes were excluded in the amplified fragment through *in silico* digestion with APE.

Amplification of γ promoters

γ truncated promoters amplification produced a 1431 bp fragment for $^{G}\gamma$ promoter (–1382 to +49) and a 1429 bp fragment for $^{A}\gamma$ (–1380 to +49).

$^{G}\gamma$ and $^{A}\gamma$ promoters were amplified with a different forward primer (Gγ_*Kpn* I and Aγ_*Kpn* I, respectively). Both the forward primers were designed by adding three A and the *Kpn* I restriction site to the specific primer sequence (5'–AAA–GGTAC▼C–sequence–3'). A common reverse primer (R161_*Xho* I) was engineered in the same way (5'–AAA–C▼TCGAG–sequence–3').

PCR mix contained 1X *Pfu* Buffer, 2.5 and 3 mM Mg^{2+} for $^{A}\gamma$ and $^{G}\gamma$, respectively, 375 μM dNTPs, 0.3 μM of each primer, 0.3 μg of DNA, 3U of *Pfu*

(Life Technologies) and was performed through a touch down PCR technique (3' at 94°C; 15 cycles of 45'' at 94°C, 1' at 65°C and 4' at 72°; 20 cycles of 45'' at 94°C, 45'' at 63°C, 4' at 72°C. A final elongation step 10' at 74°C was added).

2 µl of PCR product were electrophoresed and quantified on agarose gel, while the remaining volume was immediately purified with phenol:chloroform protocol (Protocol 2) in order to inactivate the *Pfu*[3] and recover a ready to be digested DNA.

9.2 Cloning of PCR fragments

γ promoters were first cloned in pBluescript II SK (shortly pSK) to use them in cotransfection experiments and then transferred in pGL2–Basic plasmid for luciferase assays (Para. 9.3). Both plasmids confer ampicillin resistance to the transformed bacteria (Fig. 23).

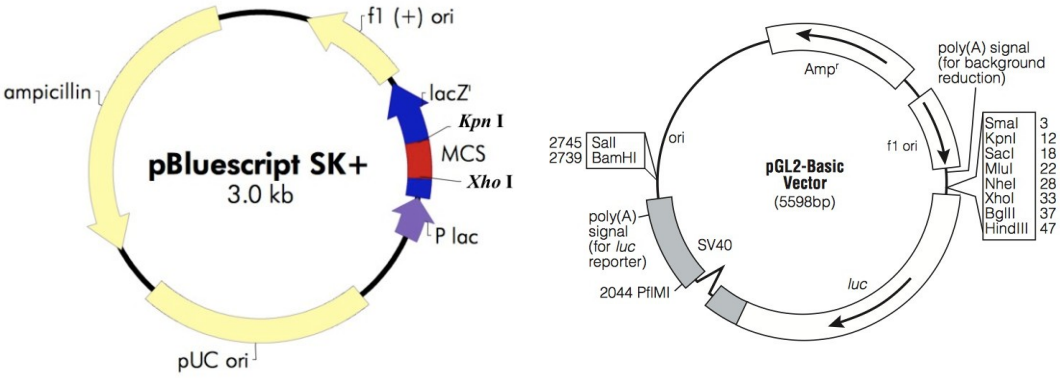


Figure 23
Plasmids used in this thesis: pBluescript (pSK) on the left; pGL2 Basic on the right.

Two digestions were necessary to prepare the fragment and the plasmid for the ligase reaction. Both the entire quantity of purified PCR product and 3 µg of plasmid were digested with 3 U of the restriction enzyme (*Kpn* I) for 1 h at 37°C, then purified with phenol:chloroform precipitation (Protocol 2) and digested again with the second enzyme (*Xho* I).

[3] The Pfu, when in the deficiency of dNTPs, favors the proofreading activity and could digest the ends of the amplified fragment which contain the restriction sites.

The enzyme with the most alkaline pH was chosen for the second digestion in order to have a pH at which the alkaline phosphatase works.

The double digested plasmid was incubated 1 h at 37°C with Calf Intestinal **Alkaline Phosphatase** (CIAP or CIP) simply adding 1 unit of CIAP directly to the second digestion tube. CIAP removes phosphate groups from the 5' end of the vector so that it cannot ligate, preventing the recircularization of the linearized vector and improving the possibility to transform only with the vector that contains the appropriate insert.

Both double digested insert and dephosphorylated vector were electrophoresed for 30' in a 1% agarose TAE gel. Bands were excided with sterile scalpel and DNA was purified with Wizard SV Gel and PCR Clean–Up System (Promega).

Purified DNA was now quantified by means of NanoDrop and a **ligation** was set in a 10 µl volume: 25 ng of vector were ligated to the insert with 3 units of T4 DNA Ligase (vector:insert 1:3) by incubating at 16°C overnight. Also a negative ligase control was set, at the same conditions described above except for the absence of insert.

$CaCl_2$ competent bacteria (DH5α strand of *E. coli*, obtained through Protocol 3) were **transformed** with half a volume of ligase reaction (Protocol 4), plated on a LB Agar (Sigma) in presence of ampicillin and incubated overnight at 37°C. Also a negative ligase control was used to transform DH5α, and it produced colonies only if the recircularized vector was present.

Grown monoclonal colonies were inoculated in 2 ml of LB Broth (**mini–inoculum**) in presence of ampicillin and incubated overnight at 37°C, in agitation.

1 ml of the mini–inoculum was used to recover circular plasmid DNA with a alkaline lysis combined with SDS technique (**mini–preparation**) (Protocol 5).

Exposure of bacterial suspension to a strong anionic detergent at high pH lyses the cells and releases plasmid DNA into the supernatant.

During lysis all the proteins, broken cell walls and denatured chromosomal DNA are caught in large complexes coated with SDS. The denatured material is removed by centrifugation, while plasmid DNA is recovered from the supernatant.

Sodium acetate is added and it neutralizes the pH so that the plasmid can renature. In fact, although the alkaline solution completely disrupts base pairing, the strands of closed circular plasmid DNA are unable to separate from each other because they are supercoiled. As long as the intensity and duration of exposure to OH^- is not too great, the two strands of plasmid DNA renature when the pH is returned to neutral (Sambrook and Russel, 2001).

Given that the plasmid DNA extracted with Protocol 5 is not pure, it could not be used for sequence analysis, but only for digestion with restriction enzymes.

Recovered plasmid DNA was double digested and electrophoresed to confirm the positive transformation: 2 bands are expected and their length should be the same as vector and insert, respectively.

When the mini–inoculum had been confirmed to be positive, a **midi–inoculum** was performed: 5 µl of mini–inoculum were inoculated in 50 ml of LB Broth (Sigma) in presence of ampicillin and incubated overnight at 37°C, in agitation.

1 ml of midi–inoculum was used for a pure mini–preparation with the commercial kit Wizard Plus SV Miniprep DNA Pur System (Promega) and 450 ng of insert were **sequenced** (with the same primers used in PCR and the sequence primer "γ_seq") (Table VII) to confirm that no mutations were inserted during the cloning process.

A **midi–preparation** was performed on the midi–inoculum with PureLink HiPure Plasmid Midiprep Kit (Life Technologies) to recover a large number of copies of construct.

9.3 Reporter assay

Reporter systems are used to measure transcriptional activity, in particular to study promoters and enhancers and their interactions with *cis*– and *trans*–acting elements as well as their responses to environmental changes.

Firefly luciferase is largely used as a reporter gene: it produces light from luciferin in the presence of ATP (Sambrook and Russel, 2001); the luciferase activity, measured through a luminometer (Viktor X5, PerkinElmer), is directly correlated to the levels of the enzyme and accordingly to promoter strength.

Transient transfection was chosen for the luciferase assay: K562 cells (Lozzio and Lozzio, 1975), an erythroleukemic non–adherent cell line largely utilized in expression studies of γ–genes, were used.

K562 were grown and maintained approximately from 1 to $5 \cdot 10^5$ cells/ml in RPMI 1640 GlutaMAX™ medium (Life Technologies), plus 10% fetal bovine serum, at 37°C, 5% CO_2.

Cationic lipids were used to convey the DNA inside the cells: they create artificial membrane vesicles (liposomes) that bind plasmid DNA molecules; these complexes adhere to and fuse with the negatively charged cell membrane, bringing the DNA within the cell. This process is called **lipofection** (Sambrook and Russel, 2001). Every transfection was carried out in triple and twice.

Transfection was performed with the Lipofectamine LTX reagent (Life Technologies) according to the manufacturer's protocol; briefly 500 ng of total plasmid DNA were added to 0.5 µl of Plus in 100 µl of Opti–MEM serum and antibiotic free medium (Life Technologies) and incubated for 5' at room temperature; 1.25 µl of LTX was then added. After a 30' incubation, the mixture was added to $1 \cdot 10^5$ cells/500 µl in a 24–well plate and incubated for 36–48 h at 37°C.

Basically the inserted γ promoter upstream the *Firefly* luciferase gene can drive the expression of the enzyme in K562. Mutations on the cloned promoter were tested

to understand their ability to increase or decrease luciferase's expression and hence luminescence.

The *Renilla* luciferase expression vector pRL–TK (Promega) (Fig. 24) was cotransfected with the proband pGL2 construct (1:20 pRL:pGL2) for normalization of transfection efficiency and to provide an internal control for lipofection. Also a negative control (with no pGL2 construct) and a positive control (pGL2–Control Vector, Promega) were used in the experiments.

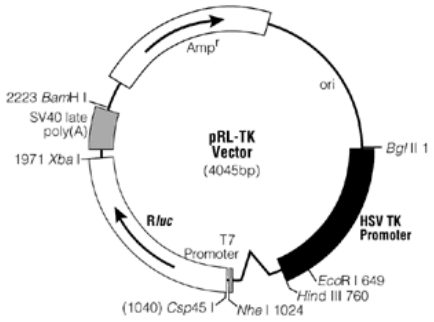


Figure 24
pRL–TK used as coreporter.

Cells were then lysed and *Firefly* (from pGL2) and *Renilla* (from pRL) luciferase gene activities were measured using the Dual–Luciferase Reporter Assay (Promega) (Fig. 25).

The luminescence from the *Firefly* luciferase of each well was normalized to the relative luminescence from the *Renilla* one, in order to obtain comparable values among experiments. The value of the wild type promoter was set as 100% of luminescence.

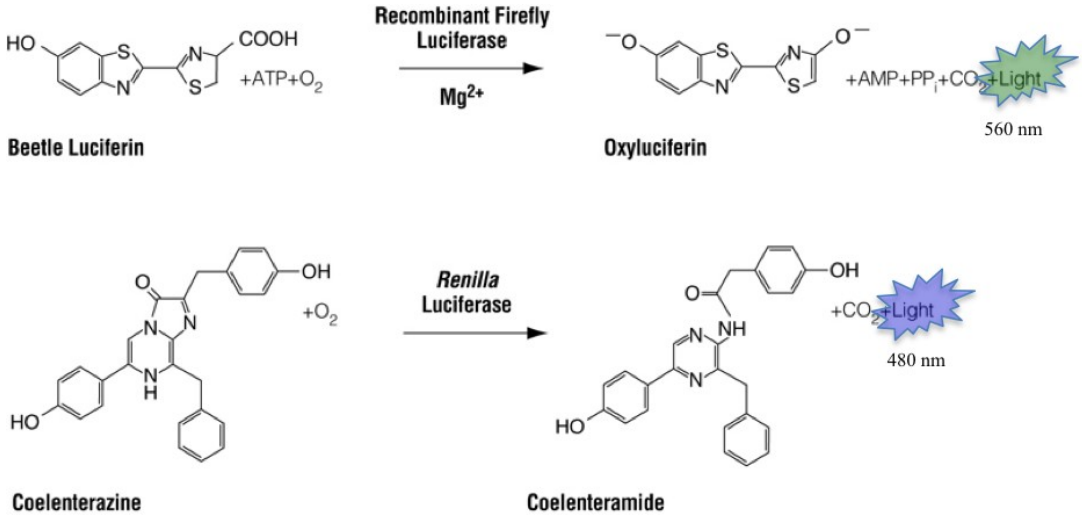


Figure 25
Dual Luciferase Reporter Assay and its bioluminescent reactions catalyzed by firefly and Renilla luciferases.

RESULTS

First level analysis

The samples were analyzed as described in Materials and Methods.

Table VIII shows the hematological data and Hb quantity and quality of the studied samples.

The parameters were normal, except for a few patients (BGy, BGi, OM, PF) that had MCH values on the lower limit compared to normal (MCH 22.4 ± 2.0 vs 25 to 35 pg). Moreover some samples showed a decreased MCV: BL, BGy, BGi, OM (females) (72.5 ± 7.1 vs 78 fl) and PF, CI (males) (72.3 ± 13.2 vs 82 fl). Slight reductions of [Hb] and increments of RBC are also visible. These data are compatible with a β–thalassemic trait (Table VIII).

CE–HPLC allowed for the quantification HbF and HbA_2 tetramers, while through RP–HPLC the globins relative percentages were calculated, defining $^{G}\gamma$ and $^{A}\gamma$ relative ratios ($^{G}\gamma$ or $^{A}\gamma$ / total of γ globins) (Table VIII). Samples with HbF levels lower than 15% needed an HbF enrichment by Preparative IEF (Para. 7.3) before injection on RP–HPLC, according to the method described by Manca and Masala, 1990.

Figure 26 shows illustrative chromatographic patterns: an increase of the HbF (Panel B and C) or γ globin fraction (Panel D, E, F) is notable.

As shown in Table VIII, all 17 samples presented very high levels of HbF compared to normal subjects. With the exception of OM (8.6%) and PF (9.5%), the values were always higher than 11% of circulating Hb, with peaks of 28% (FVR). HbF levels higher than 5% showed that the increase was not caused by the mere presence of a β–thalassemic trait or a hHPFH, but rather, by a pHPFH condition.

Two samples (BGi and OM) showed higher than normal levels of HbA_2 (>5% against a normal value of ≤ 3.2) (Table VIII).

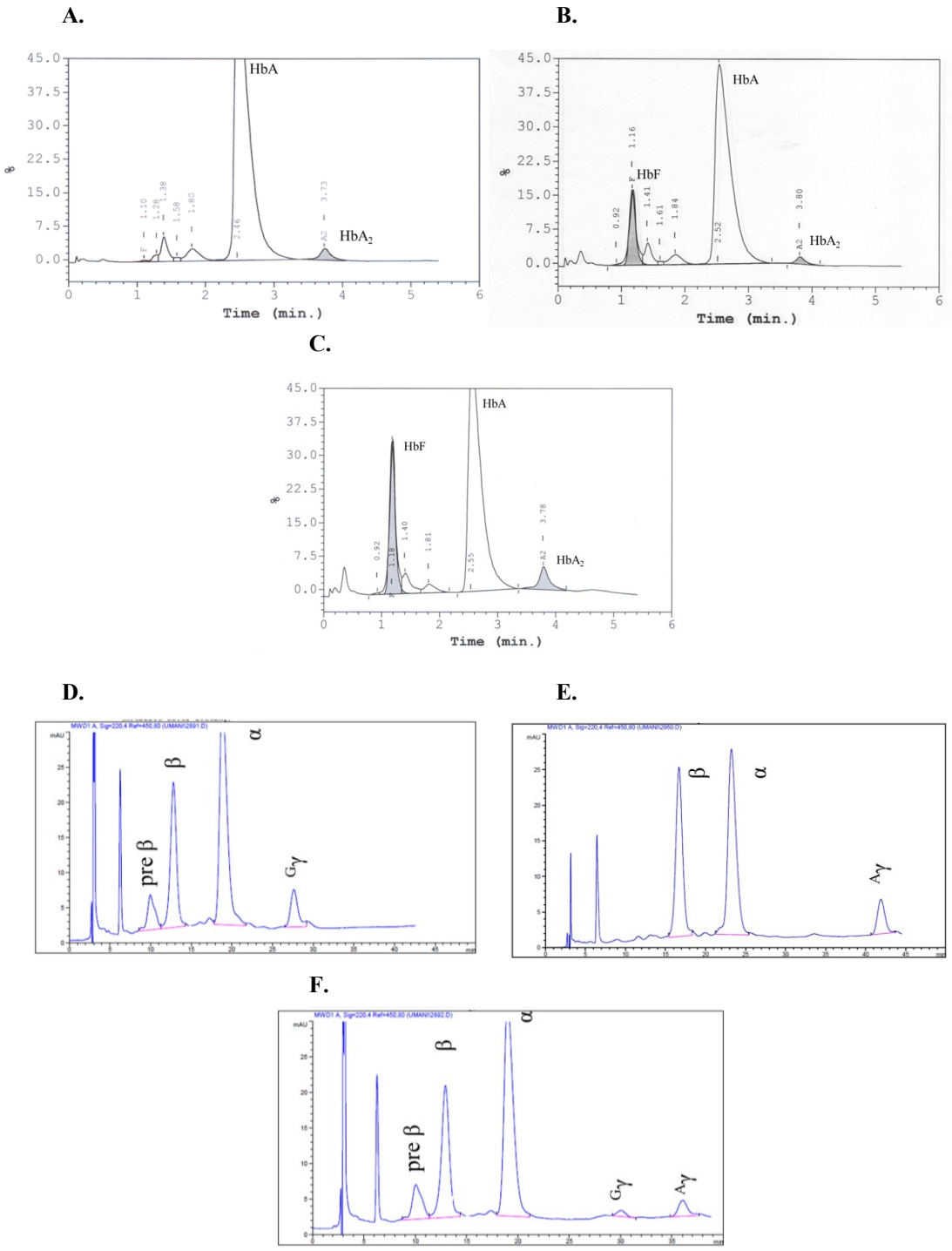


Figure 26

CE–HPLchromatograms: A) Normal sample; B) High HbF; C) Increased HbF and HbA_2
RP–HPLchromatograms: D) Persistence of $^{G}\gamma$ globin; E) Persistence of $^{A}\gamma$ globin; F) Persistence of both $^{G}\gamma$ and $^{A}\gamma$ globins

By means of chromatographic and electrophoretic techniques (AUT–PAGE, IEF) some samples were excluded from this research, when they presented extra–bands attributable to qualitative variants of hemoglobin.

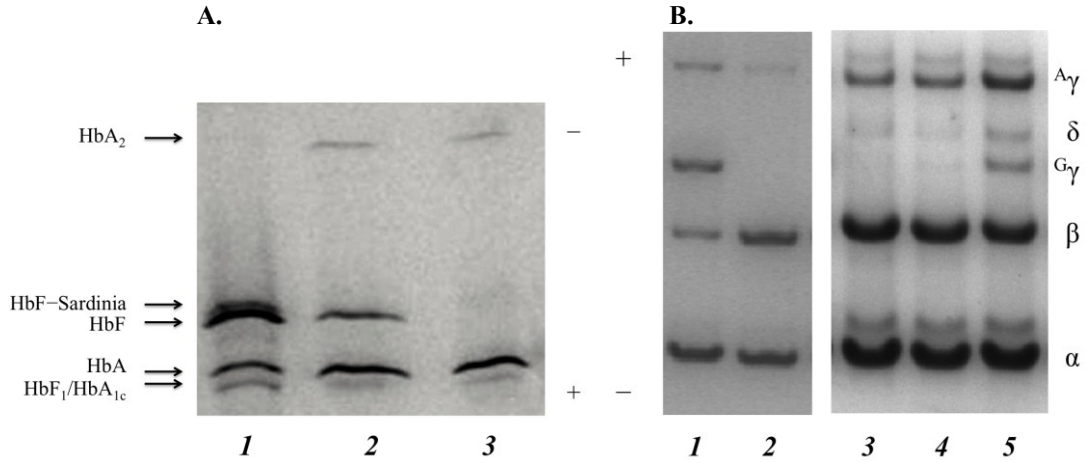


Figure 27
A) Isoelectric focusing (1: Heterozygous HbF–Sardinia newborn, 2: HPFH adult; 3: normal adult)
B) AUT–PAGE: (1: normal newborn; 2: $^{A}\gamma$–HPFH; 3, 4, 5: $^{G}\gamma+^{A}\gamma$–HPFH)

IEF and AUT–PAGE confirmed the presence of HbF (Fig. 27 A and B). In only 2 subjects the HbF was solely composed by $^{G}\gamma$ globin (CA and FVR); 7 presented only $^{A}\gamma$ (CI, GM, CMA, BA, BGy, CR, CV); some other samples showed both globins (Table VIII).

It is noteworthy that none of these samples showed HbF–Sardinia tetramer (by IEF) or $^{A}\gamma^{T}$ globin (by RP–HPLC), although this is the most common variant of HbF (30% of Sardinian population) (Manca and Masala, 2008).

No sample showed quantifiable level of HbH (β_4).

Second level analysis

Samples were then analyzed at molecular level: first of all, when possible, the complete β–globin gene sequence was obtained. It revealed heterozygous β–thalassemic alleles in 7 samples: 6 were β^{0}(39)/β, 1 was β^{+}(–87)/β (Table IX).

Additionally, β–globin gene sequences revealed five SNPs: the synonymous mutation C→T at codon 2 (His→His), and four SNPs in IVS2 (+16 C→G or *Ava* II − → +), +74 G→T, +81 C→T and +666 T→C). These SNPs indicate the existence of three β–globin gene frameworks: Type 1 (CCGCT), Type 2 (CCTCT), Type 3

(TGTTC) (Orkin *et al.*, 1982). Type 1 was found in 11/26, Type 2 in 5/26 and Type 3 in 9/26 alleles. Only one case revealed a recombinant β–gene Type (GM: CCTTT).

Promoters of fetal genes were sequenced to identify ndHPFH mutations. The choice of amplifying either $^{G}\gamma$ or $^{A}\gamma$ or both globin genes loci was driven by the quality of fetal globin expressed by each sample.

Four ndHPFH were found: $^{A}\gamma$ –196 C→T, $^{A}\gamma$ –117 G→A, $^{G}\gamma$ –175 T→C and $^{G}\gamma$ –158 C→T (or $^{G}\gamma$ *Xmn* I, as presented in Table X). These mutations were already known to be present in Sardinia.

All $^{A}\gamma$ –196 C→T samples (5/17) showed a β–trait ($\beta^{0}(39)/\beta$), normal levels of HbA_2 (2.5 ± 0.1%) and HbF levels near to 20% (21.5 ± 2.2%). Hematological parameters in this group were normal (GM, CMA, MGB) or slightly varied (BL, CI). CI, GM and CMA showed the exclusive production of $^{A}\gamma$ globin, while MGB and BL presented also $^{G}\gamma$ globin. MGB was confirmed to be heterozygous for $^{G}\gamma$ *Xmn* I polymorphism, while BL showed two substitutions (–474 C→T and –309 A→G) on the $^{G}\gamma$ promoter. Cloning of this promoter in pSK established that these two mutations were *in cis*.

$^{A}\gamma$ –117 G→A was found on 7/17 samples, with normal β–globin genes. All these samples carried the Type 3 β–gene framework, where the synonymous mutation C→T at codon 2 (His→His) is present. HbF levels in this group were the less variable among all ndHPFH found in this study, settling at 14.4 ± 1.6%. 4/7 produced only the $^{A}\gamma$ globin; the others showed persistence of both γ globin, with peaks of ~40% of $^{G}\gamma$ globin (PG). Among the latest group, MS and PG showed a normal $^{G}\gamma$ promoter, while DGI was heterozygous for $^{G}\gamma$ *Xmn* I polymorphism.

Only two samples were heterozygous for $^{G}\gamma$ –175 T→C (CA, FVR), showing different levels of HbF (19.4% and 28.3%, respectively). Only these samples presented exclusive synthesis of $^{G}\gamma$ globin. No β–thalassemic alleles were found in this group.

Only 3 samples were solely heterozygous for $^{G}\gamma$ –158 C→T ($^{G}\gamma$ *Xmn* I): in fact no other ndHPFH mutation was found on either $^{G}\gamma$ or $^{A}\gamma$ promoter. This group is the

most heterogeneous with regard to the production of HbF in terms of quantity and quality. Also a different β–thalassemic allele was found on each $^{G}\gamma$ –158 sample (Table X).

CI, CR, CV, PG and OM samples were heterozygous for AGCA deletion (–225/–222) in the $^{A}\gamma$ promoter: this deletion is known to be always present on the promoter of the gene encoding the $^{A}\gamma^{T}$ globin (Manca and Masala, 2008).

All the $^{G}\gamma$ promoters showed a homozygous GA→AG inversion at –535/–534, while all the $^{A}\gamma$ promoters revealed a homozygous C→G transversion at –369. These differences could be due to rare SNPs in the reference sequence.

MLPA for β–globin cluster was performed on samples which were thought to carry a dHPFH: only one sample (PF) showed the presence of a deletion (Sicilian δβ–thalassemia) (Fig. 28). GAP–PCR and subsequent sequencing analysis confirmed the same break–point of this deletion as already described in literature.

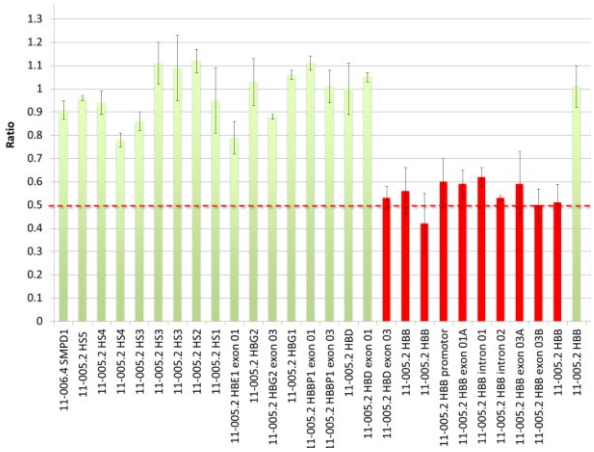


Figure 28
Bar chart for PF sample analyzed with SALSA MLPA kit P102–B1 HBB. In red the heterozygous deleted region.

Due to a recurrence of the $\beta^{0}(39)$ allele in $^{A}\gamma$ –196 C→T samples and the synonymous C→T mutation at codon 2 on β gene in $^{A}\gamma$ –117 G→A samples, Orkin haplotypes were analyzed to assess if an association between ndHPFH mutation and β–globin cluster was present. The $^{A}\gamma$ –196 group showed haplotype I while $^{A}\gamma$ –117 haplotype VII; both showed also a recombinant haplotype (Table XI).

Furthermore, Orkin haplotypes were studied in BGi sample, heterozygous for β^{+}(–87)/β: sequence analysis showed the presence of haplotype I and VIII (Table XI); in addition some other polymorphisms described by De Angioletti *et al.*, 2004 were identified. BGi revealed $(AT)_9N_{12}(AT)_{10}$ motif in the β–LCR HS2 region,

TAG at −1450, −1280, −1225 $^{G}\gamma$–globin gene (pre $^{G}\gamma$ framework), T at −158 $^{G}\gamma$–globin gene, $(TG)_{11}(CG)_3$ motif in the $^{G}\gamma$–IVS2, $(TG)_{13}$ motif in the $^{A}\gamma$–IVS2, $(AT)_9T_5$ motif in β promoter.

Reporter assays

Luciferase assay performed on γ globin ndHPFH promoters compared to wild type promoters gave the following results: $^{A}\gamma$ –196 C→T increased luminescence of ~1.4 fold, $^{A}\gamma$ –117 G→A of ~1.3 fold, $^{G}\gamma$ –175 T→C of ~3.5 fold (Fig. 29 A, B and C respectively).

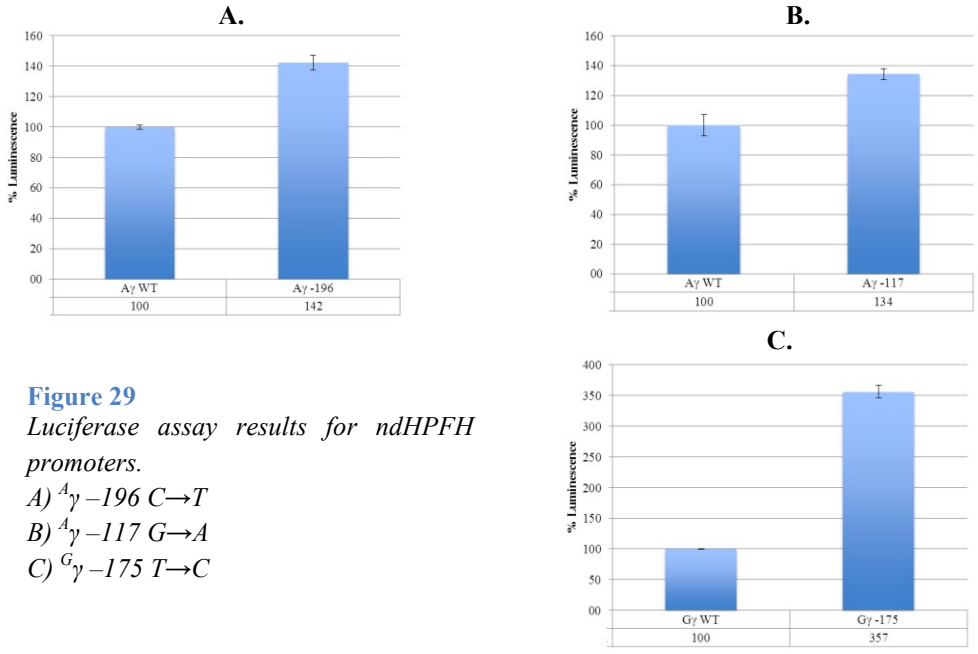


Figure 29
Luciferase assay results for ndHPFH promoters.
A) $^{A}\gamma$ –196 C→T
B) $^{A}\gamma$ –117 G→A
C) $^{G}\gamma$ –175 T→C

It was also determined whether or not the two mutations found in BL (–474 C→T and –309 A→G) were able to alter the expression of $^{G}\gamma$ globin. Experiments showed a decrease of luminescence of 10% circa compared to the wild type promoter (Fig. 30 A).

Moreover, since BL also carried $^{A}\gamma$ –196 ndHPFH allele, the $^{A}\gamma$ wild type and $^{A}\gamma$ –196 promoter were cotransfected ($^{A}\gamma$ pSK: $^{G}\gamma$ pGL2 1:1): the decrease of expression of $^{G}\gamma$ was not reverted and appeared higher in the presence of the mutated $^{A}\gamma$ allele (a 10% decrease of luminescence compared to the one from the cotransfection with wild type $^{A}\gamma$) (Fig. 30 B).

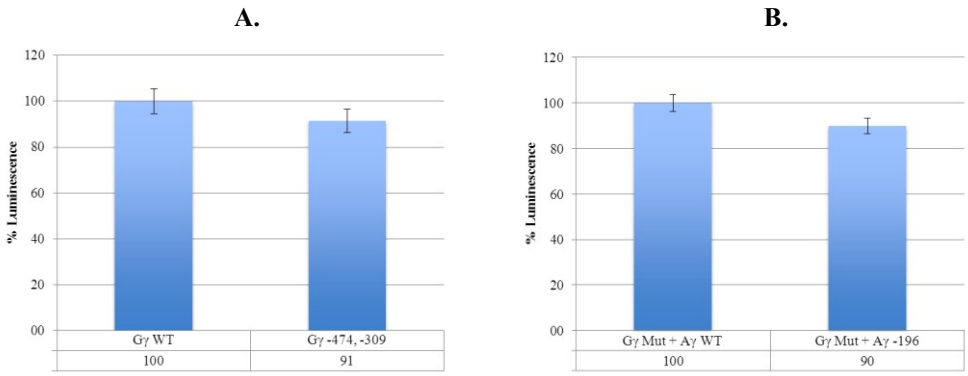


Figure 30

Luciferase assay results for BL mutated promoter.
A) $^{G}\gamma$ muted promoter (–474 C→T and –309 A→G)
B) Cotransfection of $^{G}\gamma$ muted promoter (–474 C→T and –309 A→G) + $^{A}\gamma$ WT or $^{A}\gamma$ –196 C→T

DISCUSSION

The β–thalassemias occur with a very high clinical variability, even in presence of the same genotype, suggesting that different genetic determinants can modulate the clinical course of the disease. Unfortunately, γ–globin genes are switched off during the development even in presence of one or two defective β–genes. Either the absence or the significant reduction of β–globin leads to the accumulation of α–globin within the cell. Free α–globins precipitate and damage the membranes of RBC precursors causing ineffective erythropoiesis.

In general, any factor that reduces the imbalance between α/non–α globin can improve the clinical phenotype (β^{+}–alleles, α–thalassemia, higher expression of AHSP, persistence of HbF).

Among them, Hereditary Persistence of Fetal Hemoglobin (HPFH) syndrome, which leads to a non–silenced γ–gene in adulthood, appears to be the most effective: patients with a TM genotype ($\beta^{0}(39)/\beta^{0}(39)$), although they are expected to suffer of a severe transfusion dependent anemia, show a TI phenotype, mostly because of the persistence of HbF. In fact, the γ–globin genes activity causes an increase in HbF that is able to reduce the α/non–α globin imbalance and compensate for the biosynthetic lack of HbA, typical of these pathologies.

Unfortunately, this phenomenon occurs in a limited number of people, so it is very rare that the persistence of HbF is coinherited in presence of β–thalassemia (Para. 3.2).

Recently, given such evidences, research interest has been directed to the design of new strategies aimed at reactivation of γ–globin genes. The development of alternative therapies based on the modulation of the expression of target genes and the control of Hb fetus→adult switch, however, requires a deep knowledge and complete understanding of the complex molecular mechanisms that regulate globin gene expression, which still remain mostly unclear.

This research intends to increase knowledge about the regulation of γ–globin genes, with particular attention to the switch fetus→adult. It may also give support to the current pre– and post–natal diagnosis and therapeutic approach of β–thalassemia, in order to prevent unnecessary and harmful transfusion regimes.

First level analysis

Samples were first screened for hematological data and hemoglobin profiles (Table VIII). Blood samples from donors were analyzed by means of hematocrit and CE–HPLC, so it was possible to evaluate the hematological parameters and profile of Hb tetramers.

Through CE–HPLC samples with HbF levels higher than 5% were selected: in this way hHPFH samples were excluded from the analysis (Chap. 4). Moreover, a first screening of extra–peaks (in addition to HbF, HbA and HbA_2) was carried out.

Given that CE–HPLC is not able to discriminate some variants like the HbJ–Sardegna (α^J–Sardegna, cod. 50 His→Asp), which coelutes with HbF, the Hb tetramers were also separated by IEF. The presence of HbF was therefore confirmed and samples that showed HbJ or other extra–bands were excluded.

For this purpose, also globin monomers were separated by AUT–PAGE and RP–HPLC.

Moreover RP–HPLC gave a relative quantification of individual globins and defined the quality of HbF ($^G\gamma/^A\gamma$ ratio) with greater sensitivity than AUT–PAGE.

No sample showed, by means of CE–HPLC or IEF, the presence of HbH (β_4) homotetramer. It appears as a narrow peak which elutes immediately in CE–HPLC or as a very anodic band in IEF. The presence of HbH is associated with the deletion or malfunction of at least two of the four α genes: β–globins, without enough α counterparty, homotetramerize and form a Hb which is not useful to the transport of O_2 (Para. 3.1.1).

The lack of HbH in the analyzed samples made it possible to establish, in good confidence, that at least 3/4 α genes were functioning. Since a single not functional α

gene is not able, alone, to improve the thalassemic phenotype, it was decided not to analyze the α–cluster at the molecular level.

HbA_2 was found increased in 2 samples (BGi, OM), later explained by the presence of a β–trait (Table VIII).

The HbF levels were very heterogeneous (Table VIII), ranging from a minimum of 8.6% (OM) to a maximum of 28.3% (FVR), with 2 recurring mean values (~15% and ~22% of HbF). This data emphasizes the extreme variability in the multifactorial and not yet fully explained persistence of HbF.

Hematological data of the samples made it possible to distinguish normal samples from carriers of α– and β–thalassemia.

The parameters were normal, except for a few patients (BGy, BGi, BL, CI, OM, PF). Two samples (BGi and OM) showed higher than normal levels of HbA_2 (>5% against a normal value of ≤ 3.2). Molecular results explained these hematological alterations except for BGy (See Below).

Second level analysis

After the initial screening by means of electrophoretic and chromatographic techniques, the molecular investigation was carried out on the 17 samples that showed no structural variants.

The entire β–gene was sequenced in all samples: 6 were $\beta^0(39)/\beta$, 1 was $\beta^+(-87)/\beta$ (Table IX). In addition, PF showed a large deletion which removes the β–globin gene (Sicilian δβ–thalassemia, See Below).

In total 8/17 samples were carrier of a β–thalassemia allele: only 2 of them (OM and BGi) showed the HbA_2 increase typical of β–trait, while for the other samples it was impossible to predict the mutated allele from the hematological parameters (Table VIII). The slight decrease or normal values of MCV and MCH, in addition to normal levels of HbA_2 in samples BL, CMA, CI, GM and MGB was later explained by the presence of $^{A}\gamma$ –196 C→T ndHPFH mutation linked to the $\beta^0(39)$ allele (See below).

PF instead, heterozygous for Sicilian δβ–thalassemia, lacks a functional δ gene and that explains normal levels of HbA_2.

Only BGy's lower MCV and MCH were not explained by her β–globin genes, which turned out to be normal. Therefore, for this sample, an α–thalassemia trait is plausible, although the reductions in hematological parameters are so mild (in addition to the absence of HbH) that at best a single dysfunctional α–gene would be suspected.

Promoters of fetal genes were sequenced on the basis of either $^{G}\gamma$ or $^{A}\gamma$ or both globin produced by each sample.

Four ndHPFH were found: $^{A}\gamma$ –196 C→T, $^{A}\gamma$ –117 G→A, $^{G}\gamma$ –175 T→C and $^{G}\gamma$ –158 C→T (or $^{G}\gamma$ *Xmn* I, as presented in Table X). These mutations were already known to be present in Sardinia.

$^{G}\gamma$ –175 T→C

2/17 samples (CA and FVR) showed persistence of only $^{G}\gamma$ and turned out to be heterozygous for $^{G}\gamma$ –175 T→C mutation (Para. 4.2.3). This mutation is one of the strongest ndHPFH, capable of pushing HbF up to 20.3 ± 2.8%. $^{G}\gamma$ –175 T→C is besides linked to normal hematological values and a HbF, mostly $\alpha_2{}^{G}\gamma_2$ (94.0 ± 5.2%) (Steinberg *et al.*, 2009) (Table VI).

CA showed HbF=19.4%, 100% $\alpha_2{}^{G}\gamma_2$, in accordance with literature, while FVR showed a much higher level of HbF (28.3%). The sequence analysis performed on the β–gene highlighted different β–gene framework (Type 3) in CA respect to FVR. Type 3 framework (TGTTC) (Orkin *et al.*, 1982) possesses the synonymous mutation at codon 2: it falls on one of the three E–box (Enhancer box) present within the β–gene (Leach *et al.*, 2003). Leach and colleagues demonstrate that mutations at the level of E–box can reduce the binding affinity for activators of transcription (Para. 2.1); these factors could be therefore free to bind γ–globin promoters.

This hypothesis is in disagreement with data obtained for samples CA and FVR; FVR in fact expresses higher levels of HbF, although CV showed the mutation

on the E–box. The difference may be due to other *cis*– and *trans*–acting factors, not least the other IVS2 SNPs which characterize the Type 3 framework that, even if described as polymorphic sites, have never been associated with the modulation of the expression of fetal genes. Up to now, in fact, two of these polymorphisms (+81 and +666) have been studied only in association with β–thalassemia and β^S alleles (Bilgen *et al.*, 2011).

Further studies are necessary to verify the possible involvement of these SNPs on gene expression in $^G\gamma$ –175 T→C ndHPFH subjects.

$^G\gamma$ –158 C→T

The $^G\gamma$ –158 C→T or $^G\gamma$ *Xmn* I polymorphism (Para. 4.2.4) was found in 5 samples, two of which (DGI and MGB) were also carriers of mutations $^A\gamma$ ndHPFH. Given the high incidence of $^G\gamma$ –158 (33–35%) in normal population, it is not surprising to observe it in association with $^A\gamma$ –117 (DGI) or $^A\gamma$ –196 (MGB) ndHPFH mutations. These two samples will be discussed below.

Regarding BGi, OM and PF, characterized by the sole presence of the *Xmn* I polymorphism, they were very different in terms of the amount of HbF and its quality (Table X). The cause of this differences in phenotype is unknown as no transcription factors seem to bind there, but it may be related to the different chromosomal backgrounds and mutations on other genes involved in Hb switch, which are in *linkage disequilibrium* with this SNP (Steinberg *et al.*, 2009). Moreover, by hematological data in all three patients and higher HbA_2 in BGi and OM, the presence of a β–trait was assumed.

BGi, within this group, showed the greatest proportion of HbF. The analysis of the β–genes on this sample verified the presence of the β^+(–87) mutation in heterozygosis, on a Type 2/2 β–framework.

Given that β^+(–87) is known to be present on two different haplotypes of β–cluster, one overexpressing HbF ($^G\gamma$/Total γ = 69.4 ± 2.6%) (Fig. 17)

(De Angioletti *et al.*, 2004), Orkin haplotypes, microsatellite and polymorphisms described in De Angioletti's work were characterized.

All results are comparable to those found in the overexpressing HbF haplotype; thus, the proportion of HbF in BGi seems to be explainable with the haplotype described by De Angioletti.

Furthermore, although haplotype VIII had already been described in association with Type 2 β framework (Orkin *et al.*, 1982), this framework was found in Haplotype I in this study; this observation was confirmed also in BL.

OM showed levels of HbF equal to 8.6%, with a ratio of about 80:20 in favor of $^{G}\gamma$. The sequence of the β–gene showed $\beta^{0}(39)$ allele in homozygosis. It could be surprising to see the $^{A}\gamma$–globin rise up to a 20% in the presence of a normal $^{A}\gamma$ promoter (Table X). However *Xmn* I polymorphism had already been associated with concomitant activation of $^{G}\gamma$ and $^{A}\gamma$ globin gene transcription on a β^{0}–thalassemia chromosome. Nevertheless, whether or not the −158 *Xmn* I polymorphism was responsible alone for the activation of both γ genes, was not elucidated so it is possible a coinheritance of other genetic modifiers, such as QTLs for HbF that are not linked to the β globin complex (Haj Khelil *et al.*, 2011).

Although PF showed $^{G}\gamma$ –158 C→T and HbF levels similar to OM, he showed inverted $^{G}\gamma$:$^{A}\gamma$ ratios (20:80 for PF vs 80:20 for OM). Moreover, even in this case, no $^{A}\gamma$–ndHPFH mutation which could explain the increase of $^{A}\gamma$–globin was found. However, even though hematological parameters were profoundly altered (suggesting a β–trait), the HbA_2 turned out to be normal (Table X).

A deletion was therefore suspected, causing the increase of HbF and involving also the δ–gene (dHPFH or δβ–thalassemia). The entire β–cluster was analyzed by means of MLPA, revealing a heterozygous deletion from the third exon of δ gene up to a non–well–defined region (for lack of probes), approximately 9 kb downstream β–gene, in which the 3' breakpoint could be found.

Basing our study on work published by Craig *et al.*, 1994, the most probable deletion, compatible with the information obtained by MLPA, was the Sicilian δβ–thalassemia which was subsequently confirmed by GAP–PCR and sequencing. This deletion involves a fragment that starts 188 bases before the beginning of the third exon of the δ–gene delta and ends at 5,710 bp downstream of the termination codon of the β–gene, removing 13,378 bp (Table IV). Heterozygotes for Sicilian δβ–thalassemia show mild anemia, typical thalassemic RBC morphology, hypochromia, basophilic stippling, increased HbF (4.0–18.6%), low/normal HbA_2 (2.3%), and 25–56% $^{G}\gamma$ levels (http://globin.bx.psu.edu). Hematological results of PF are in agreement with the ones already described in literature.

$^{A}\gamma$ –117 G→A

The $^{A}\gamma$ –117 G→A (Para. 4.2.1) is known to be linked to normal hematological data except for the production of HbF (12.1 ± 2.8%), mostly $^{A}\gamma$ (93.4 ± 4.7%) (Table VI) (Steinberg *et al.*, 2009). This mutation was found in 7 samples, all expressing similar HbF levels (14.4 ± 1.6%): 4 of them (BG, BGy, CR, CV) expressed only the $^{A}\gamma$ globin, as expected from literature. MS, PG and DGI, on the contrary, also expressed the $^{G}\gamma$–globin. DGI was found heterozygous for $^{G}\gamma$ *Xmn* I, and this could explain the persistence of $^{G}\gamma$–globin.

However MS and PG revealed a normal $^{G}\gamma$–gene promoter and, therefore, it still remains unclear the cause of the synthesis of $^{G}\gamma$ globin in these two patients.

Complete β–gene sequence showed normal β–alleles in all these samples and a recurrent Type 3 β–framework (TGTTC). $^{A}\gamma$ –117 samples were the only ones in this study in which Type 3 framework can be found, suggesting that the mutation originated only once on the same β–cluster. CV and DGI were therefore analyzed for Orkin haplotype and revealed the presence of haplotype VII, as expected from literature (http://globin.bx.psu.edu and Pistidda *et al.*, 1995). This is the only ndHPFH mutation, among those analyzed in this study, that falls always on the same haplotype.

$^{A}\gamma$ –196 C→T

The $^{A}\gamma$ –196 C→T (Para. 4.2.2) is associated with normal hematological data and 13.7 ± 2.0% of HbF, mostly $^{A}\gamma$ type (95.0 ± 0.2%) (Table VI) (Steinberg *et al.*, 2009). In this study this mutation was observed in 5 samples: HbF levels were higher than the ones just described (21.5 ± 2.2%). All samples were heterozygous for the β^{0}(39) allele, and this difference from the reference data could be the cause of higher level of HbF in the samples analyzed in this study. β–gene frameworks and haplotypes (Table X and XI) were very variable, meaning that this mutation either originated on different chromosomes or is very old and recombined on different haplotypes.

It is important to emphasize that in all of these samples no/slight changes in hematological values and HbA_2 levels were found, despite the β–trait.

The lack of increase of HbA_2 in presence of heterozygous β^{0}(39) allele has already been described and this association is named Sardinian δβ–thalassemia because of the presence of a β–thalassemia allele and the reduced amount of δ globin (HbA_2). It is supposed that the reason why δ–globin gene is under–expressed, although the δ gene is entirely normal, is probably imputable to the suppressive effect of the *cis* $^{A}\gamma$ –196 ndHPFH mutation, which may compete for the transcription factors and β–LCR (Loudianos *et al.*, 1992). Moreover, among ndHPFH found in this study, $^{A}\gamma$ –196 is the only one that could compensate MCV and MCH up to normal values, in presence of β–thalassemia allele.

CI, GM, and CMA showed only $^{A}\gamma$ globin, while MGB and BL produced also $^{G}\gamma$. MGB was carrier of $^{G}\gamma$ *Xmn* I and this could cause the persistence of $^{G}\gamma$, while BL was –/– for this polymorphism. However BL presented two heterozygous substitutions *in cis* on the $^{G}\gamma$ promoter: the –474 C→T mutation, never described, and the –309 A→G, a known polymorphism, never associated with the persistence of HbF, but only used to define peculiar haplotypes associated with β^{S} (Barbosa *et al.*, 2010) or 5' sub–haplotype in De Angioletti's work (De Angioletti *et al.*, 2004). Given that BL showed persistence of $^{G}\gamma$, it was supposed that the two substitutions

could act in synergy or individually, increasing the synthesis of this globin. The reporter assay results, discussed below, disproved this hypothesis.

The nucleotide sequencing of the $^{A}\gamma$ promoter revealed a deletion of 4 bp in heterozygosis, from –225 to –222 (AGCA), in samples CI, CR, CV, PG and OM. It is known that this deletion characterizes the $^{A}\gamma^{T}$ locus, whose globin tetramerizes in HbF–Sardegna ($\alpha_2{}^{A}\gamma^{T}{}_2$), the most common HbF variant in the world [$^{A}\gamma$75 (E19) Ile→Thr] (Manca and Masala, 2008). The incidence of this variant is so high in the Sardinian population (33%), that it could be considered a polymorphism.

Manca and Masala, 2008 believe that the presence of the –225/–222 deletion on the promoter reduces the binding efficiency of some transcription factors, resulting in a slightly reduced synthesis of $^{A}\gamma^{T}$ globin compared to the normal $^{A}\gamma^{I}$.

Although the samples mentioned above were heterozygous for the $^{A}\gamma^{T}$ allele, none of them showed persistence of the corresponding globin, perfectly recognizable from $^{A}\gamma^{I}$ by the screening techniques used in this study. There is no data on ndHPFH mutations which leads to the expression of $^{A}\gamma^{T}$ globin. This could mean that 1) ndHPFH mutations occurred only on the $^{A}\gamma^{I}$ locus; 2) if a ndHPFH mutation affects a $^{A}\gamma^{T}$ gene, it is not capable of altering the expression, perhaps because of the deletion.

Reporter assays

To assess whether a single ndHPFH mutation in the $^{G}\gamma$ and $^{A}\gamma$ genes was able to modify the strength of the associated promoter, expression assays were performed by means of luciferase as a reporter gene.

The wild type and mutated promoters of $^{G}\gamma$ and $^{A}\gamma$ were inserted upstream of the luciferase gene (in the plasmid pGL2–Basic) and their luminescence was then evaluated (Chap. 9).

To validate this system, $^{G}\gamma$ –175 promoter, already known to increase the luminescence of 4 fold circa (Li *et al.*, 2001), was constructed and transfected in K562 cells. In our experiments the luminescence was increased by 3.6 fold circa,

confirming that the system gives reproducible results. The differences were statistically significant (p=0.01).

$^{A}\gamma$ –117 RNA synthesis had already been evaluated through transient expression assay and produced about 1.4 fold as much RNA as the wild type $^{A}\gamma$ gene when standardized against RNA transcribed from the $^{G}\gamma$ genes *in cis* (Rixon and Gelinas, 1988). In this study, the strength of $^{A}\gamma$ –117 was evaluated by means of luciferase assay, showing an increase of 1.34 fold ($p<0.01$), perfectly consistent with previous findings.

The $^{A}\gamma$ –196 promoter expression had been already analyzed and did not show any increase in gene expression *in vitro* (Ronchi *et al.*, 1989). In this study, when transfected in K562 for the luciferase assay, the $^{A}\gamma$ –196 showed an increase of 1.42 fold, compared to the wild type $^{A}\gamma$ promoter ($p <0.01$). Therefore, it is plausible that the luciferase assay is more sensitive than the expression assay chosen by Ronchi *et al.*, 1989.

Two mutations had been already described in the proximity of $^{A}\gamma$ –196: the first, $^{A}\gamma$ –198, is able to increase the expression of the associated gene of 4–5 fold (Li *et al.*, 2001), while the second, $^{A}\gamma$ –195, shows a very slight increase of $^{A}\gamma$ if HS2 enhancer is not present; consequently, HS2 is necessary for an effective action of the $^{A}\gamma$ –195 promoter (Takahashi *et al.*, 2003). Although the $^{A}\gamma$ –196 mutation does not show an increase as high as $^{A}\gamma$ –198, it still proves to be able to determine a visible increase of luminescence even in the absence of HS2. Hence, $^{A}\gamma$ –196 turns out to be a stronger ndHPFH mutation compared to the $^{A}\gamma$ –195.

It was also determined whether or not the two mutations found in BL $^{G}\gamma$ promoter (–474 C→T and –309 A→G) were able to alter the expression of $^{G}\gamma$ globin. Luciferase assay showed a decrease of luminescence of 10% circa compared to the wild type promoter ($p<0.05$).

This result invalidated the hypothesis that the two mutations could be responsible for the $^{G}\gamma$ persistence in BL sample; however it should not be surprising: in fact –474 C→T mutation falls on the "silencing" region of the γ gene (Para. 2.1)

and could create a binding site for repressors, reducing the synthesis of the gene associated with the mutated promoter. Hence, it is possible that the mechanisms that increases $^{G}\gamma$ resides in other *cis*– and *trans*–agent factors.

Since BL also carried $^{A}\gamma$ –196 ndHPFH allele, to exclude *trans*–activation of $^{G}\gamma$ gene, the $^{A}\gamma$ wild type and $^{A}\gamma$ –196 promoter were cotransfected. The decrease of expression of $^{G}\gamma$ was not reverted and appeared higher in the presence of the mutated $^{A}\gamma$ allele (a 10% decrease), meaning that $^{A}\gamma$ –196 is not able to *trans*–activate $^{G}\gamma$ gene in presence of –474T and –309G mutations. The decrease in the presence of $^{A}\gamma$ –196 is easily explainable: the $^{A}\gamma$ –196 promoter is stronger than the $^{A}\gamma$ wild type and this means that it has a better ability to bind transcription activators. It is clear that in the presence of the $^{A}\gamma$ –196, more transcription factors are "captured" by the mutated $^{A}\gamma$ promoter, thereby reducing the efficiency of transcription of the $^{G}\gamma$ promoter.

Unfortunately, these results did not explain the reason for the persistence of $^{G}\gamma$ globin in the BL sample.

The enhancement produced in these luciferase assays was demonstrated to be several times less than that seen *in vivo*. However, the results are similar to those described with other mutations and refer to a promoter which is decontextualized from the β–cluster: it is not possible to predict a different behavior in terms of *in vivo* expression. For example, the presence of β–LCR (or HS2 alone) could change the final result of the assay, by increasing the difference in γ promoter activation as it was proved with the $^{A}\gamma$ –195 mutation (Takahashi *et al.*, 2003).

The data presented in this thesis confirm the complexity of the globin switch and the heterogeneity of the molecular mechanisms underlying the persistence of HbF, emphasizing the importance of the molecular context of a mutation.

The ability to associate a given mutation or a given combination of polymorphisms to higher HbF levels in adulthood could clarify the likelihood of a β–thalassemic fetus to develop a mild disease during extra–uterine life, and consequently, to prevent unnecessary and harmful transfusion regimes.

TABLES

Table I

Role of β–LCR 5' Hypersensitive Sites (HSs)

(Harju *et al.*, 2002)

5' HS4
Transcription activation activity
Position–independent expression
Copy number–dependent expression
5' HS3
Transcription activation activity
Position–independent expression when linked to β–globin gene
Copy number–dependent expression when linked to β–globin gene
Position–dependent expression when linked to $^{A}\gamma$–globin gene
Chromatin opening activity in single copy transgenes
5' HS2
Enhancer activity
Position–independent expression
Copy number–dependent expression
Binding sites for GATA1, NF–E2, Sp1, Ap–1, USF; all necessary for enhancer function
Deletion of single GATA1–NF–E2, Sp1, or USF site does not impair position–independent expression
373 bp HS core confers position–independent expression
5' HS1
No transcriptional activation activity

Table II
Classification of Hemoglobin Disorders
(Forget and Bunn, 2013)

A. QUANTITATIVE DISORDERS OF GLOBIN CHAIN SYNTHESIS/ACCUMULATION
The thalassemia syndromes

β–Thalassemia
Clinical classification:
β–Thalassemia minor or trait β–Thalassemia major β–Thalassemia intermedia
Biochemical/genetic classification:
β^0–Thalassemia β^+–Thalassemia δ–Thalassemia γ–Thalassemia Lepore fusion gene δβ–Thalassemia εγδβ–Thalassemia HPFH "Dominant" β–thalassemia (structural variants with β–thalassemia phenotype)
β–Thalassemia with other variants:
HbS/β–thalassemia HbE/β–thalassemia Other
α–Thalassemia
Deletions of α–globin genes:
One gene: α^+–thalassemia Two genes in *cis*: α^0–thalassemia Two genes in *trans*: homozygous α^+–thalassemia (phenotype of α^0–thalassemia) Three genes: HbH disease Four genes: Hydrops fetalis with Hb Bart's
Nondeletion mutants:
Hb Constant Spring Other
De novo and acquired α–thalassemia
α–Thalassemia with mental retardation syndrome (ATR):
Due to large deletions on chromosome 16 involving the α–globin genes Due to mutations of the ATRX transcription factor gene on chromosome X
α–Thalassemia associated with myelodysplastic syndromes (ATMDS):
Due to mutations of the ATRX gene

(continued)

B. QUALITATIVE DISORDERS OF GLOBIN STRUCTURE: STRUCTURAL VARIANTS OF HEMOGLOBIN

Sickle cell disorders
SA, sickle cell trait SS, sickle cell anemia/disease SC, HbSC disease S/β thal, sickle β–thalassemia disease S with other Hb variants: D, O–Arab, other SF, Hb S/HPFH
Hemoglobins with decreased stability (unstable hemoglobin variants)
Mutants causing congenital Heinz body hemolytic anemia Acquired instability—oxidant hemolysis: Drug–induced, G6PD deficiency
Hemoglobins with altered oxygen affinity
High/increased oxygen affinity states:
Fetal red cells Decreased RBC 2,3–BPG Carboxyhemoglobinemia, HbCO Structural variants
Low/decreased oxygen affinity states:
Increased RBC 2,3–BPG Structural variants
Methemoglobinemia
Congenital methemoglobinemia:
Structural variants Cytochrome b5 reductase deficiency
Acquired (toxic) methemoglobinemia
Post–translational modifications
Nonenzymatic glycosylation Amino–terminal acetylation Amino–terminal carbamylation Deamidation

Table III
Hematological reference values of Red Blood Cells

Parameter	Patient	Lower limit	Upper limit	Unit
Hemoglobin (Hb)[1]	M	13.2	16.2	g/dl
	F	12	15.2	g/dl
Glycosylated hemoglobin (HbA_{1c})[2]	< 50 y	3.6	5	% of Hb
	> 50 y	3.9	5.3	% of Hb
Mean cell volume (MCV)[1]	M	82	102	fl
	F	78	101	fl
Red blood cell distribution width (RDW)[1]		11.5	14.5	%
Mean cell hemoglobin (MCH)[3]		25	35	pg/cell
Mean corpuscular hemoglobin concentration (MCHC)[1]		31	35	g/dl
Erythrocytes/Red blood cells (RBC)[1]	M	4.3	6.2	mln/µl
	F	3.8	5.5	mln/ µl
Reticulocytes[1]	Adult	0.5	1.5	% of RBC
	Newborn	1.1	4.5	% of RBC
	Infant	0.5	3.1	% of RBC

[1] Normal Reference Range Table from The University of Texas Southwestern Medical Center at Dallas. Used in Interactive Case Study Companion to Pathologic basis of disease.
[2] Reference range list from Uppsala University Hospital ("Laborationslista"). Artnr 40284 Sj74a. Issued on April 22, 2008.
[3] Deepak A. Rao; Le, Tao; Bhushan, Vikas (2007). First Aid for the USMLE Step 1 2008 (First Aid for the Usmle Step 1). McGraw–Hill Medical. ISBN 0–07–149868–0.

Table IV
Some of the principal deletions and their breakpoints in δβ–thalassemia and dHPFH
(Steinberg *et al.*, 2009)

Type	5' Breakpoint	3' Breakpoint	Del kb
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$ HPFH			
HPFH–1 (Black)	~4 kb 5' to the δ gene	~95 kb from the 3' end of the β gene	84.918
HPFH–2 (Ghanaian)	IVS2 of the pseudo β gene	~90 kb downstream from the 3' end of the pseudo β gene	84.281
HPFH–3 (Indian)	5' end of the pseudo β gene	30 kb downstream of the β gene	49.733
HPFH–4 (Italian)	5 kb from the 5' to the δ gene	~30 kb 3' to the β gene	~40
HPFH–5 (Italian)	3.2 kb 5' to the δ gene	0.7 kb 3' to the β gene	12.91
HPFH–7 (Vietnamese)	~3.5 kb 5' to the δ gene	~8 kb 3' to the breakpoint f ound in HPFH–3	30
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$ thalassemia			
Sicilian	δ–IVS2, 188 bp from the beginning of the third exon	5,710 bp 3' to the termination codon of the β gene	13.378
East European	1.7 kb 5' to the δ gene	within the 5' end of the β gene	~9.5
Black	~2.4 kb 5' to the δ gene to 200 bp at the	3' end of the β gene, before Alu Repeats	11.803
Indian	1 kb 3' of the $^{A}\gamma$–globin gene	>10 kb 3' of the β gene	32.624
Spanish	Midpoint between the pseudo β and δ genes into the inverted Alu I repeat	~100 kb 3' to the β gene	~93
Japanese	2134 and 2137 bp 3' to the poly A site of the $^{A}\gamma$ gene	β gene (165148)	113.63
Turkish	1.5 kb 5' to the enhancer–like sequence of $^{A}\gamma$	L1 repeat sequence 3' to the β gene	~30
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$ thalassemia			
Thai (HPFH–6)	3' to the $^{G}\gamma$ gene	More than 45 kb from the 3' end of the β gene	79.278
Black	IVS–II in the $^{A}\gamma$ gene to 200 bp	3' end of the L1 repetitive sequence	35.811
Chinese	$^{A}\gamma$–IVS–II and extends	~75 kb downstream from the β gene	78.847
Italian	exon 2 of the $^{A}\gamma$ gene	~30 kb 3' to the β gene	~52
Belgian	3' end of the $^{A}\gamma$	Same as 3' brkpt of HPFH–3, HPFH–4	~50
German	1.5 to 1.9 kb from the 3' end of the $^{G}\gamma$ gene	27 kb 3' to the β gene	~52
Turkish	1,075 bp 3' to the termination codon of the $^{G}\gamma$ gene	9,657 bp 3' to the termination codon of the β gene	36.211
Malay 2	0.2 to 0.4 kb 5' to the $^{A}\gamma$ gene	beyond sequences 17 to 18 kb 3' to the β gene.	~42

Table V

Hematological data (mean ± SD) on δβ–thalassemia and dHPFH heterozygotes and homozygotes

(Steinberg *et al.*, 2009)

Condition	Hb g/dl	MCV fl	MCH pg	HbA_2 %	HbF %	$^{G}\gamma$ %	$^{A}\gamma$ %	
HETEROZYGOTES								
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$ thal	12.3 ± 0.9	68.3 ± 4.7	23.6 ± 1.8	2.6 ± 0.3	10.5 ± 3.7	35–72	65–28	H
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$ thal	13.1 ± 0.6	75.1 ± 3.6	24.4 ± 1.5	2.3 ± 0.3	12.2 ± 1.5	83–100	17–0	H
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$ HPFH	13–2 ± 0.6	82.6 ± 4.1	26.3 ± 1.4	2.2 ± 0.4	24.1 ± 2.7	16–71	84–29	P
HOMOZYGOTES								
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$ thal	10.9 ± 0.9	78.4 ± 2.7	24.2 ± 2.4	0	100	50–61	50–39	P
$^{G}\gamma(^{A}\gamma\delta\beta)^{0}$ thal	10.5 ± 1.2	71.2 ± 5.8	21.6 ± 1.6	0	100	100	0	P
$^{G}\gamma^{A}\gamma(\delta\beta)^{0}$ HPFH	16.6 ± 1.1	71.9 ± 3.4	24.5 ± 1.3	0	100	54–65	46–35	P

H: Heterocellular distribution.
P: Pancellular distribution.

Table VI

Summary of hematological data (mean ± SD) on ndHPFH heterozygotes defined by molecular analysis

(Steinberg *et al.*, 2009)

$^{G}\gamma$ ndHPFH

Condition	Mut	Hb g/dl	MCV fl	MCH pg	HbA_2 %	HbF %	$^{G}\gamma$ %	*Xmn* I +/–
Black	−202 C→G	13.3 ± 1.3	82.0 ± 6.5	27.7 ± 2.0	1.7 ± 0.4	15.6 ± 1.2	99.0 ± 0.5	–
Tunisian	−200 +C	13.1 ± 0.9	95.9 ± 1.8	31.9 ± 0.4	1.5 ± 0.3	25.2 ± 4.1	100 ± 0	
Black/ Sardinian/ British	−175 T→C	12.7 ± 1.1	85.2 ± 3.1	28.4 ± 1.8	1.3 ± 0.2	20.3 ± 2.8	94.0 ± 5.2	–
Japanese	−114 C→T						12.5 ± 2.1	+
Australian	−114 C→G	14.2	92		2.3	8.6	90	–
Algerian	−114 C→A				2.6 ± 0.3	1.5 ± 0.8	91.9 ± 3.5	–
Belgian	−137 C→A				1.7 ± 0.2	5.6 ± 2.9		–

$^{A}\gamma$ ndHPFH

Condition	Mut	Hb g/dl	MCV fl	MCH pg	HbA_2 %	HbF %	$^{G}\gamma$ %	*Xmn* I +/–
Black	−202 C→T	12.9 ± 0.9	84.4 ± 5.9	30.0 ± 2.3	2.7 ± 1.0	2.5 ± 0.9	92.8 ± 2.8	–
British	−198 T→C	14.2 ± 1.2	83.1 ± 3.3	29.0 ± 1.0	2.5 ± 0.4	6.9 ± 2.2	92.2 ± 1.8	–
Italian/ Chinese	−196 C→T	normal	normal	30.0 ± 0.9	1.8 ± 0.6	13.7 ± 2.0	95.0 ± 0.2	–
Brazilian	−195 C→G	normal	normal	normal	2.1	5.4 ± 1.4	89.0 ± 2.5	
Black	−175 T→C	normal	normal	normal	1.5 ± 0.2	37.4 ± 1.0	78.0 ± 5.5	+
Cretan†	−158 C→T	13.9 ± 0.7	84.5 ± 6.0	27.3 ± 1.0	2.8 ± 0.2	3.7 ± 1.2	43.3 ± 8.6	+
Greek/ Black	−117 G→A	14.2 ± 1.1	85.9 ± 4.5	28.4 ± 2.4	2.0 ± 0.3	12.1 ± 2.8	93.4 ± 4.7	–
Black ††	−114/−102del	11.4 ± 2.9	75.5 ± 2.1	26.8 ± 0.3	2.1 ± 0.0	31.0 ± 1.2	85.7 ± 5.7	–
Georgian	−114 C→T		normal	normal	normal	2.5 ± 0.4 3.	8 ± 1.3	

Table VII
Primers used for PCR, Sequencing and Cloning

PRIMER		SEQUENCE (5'→3')	Position	Locus
β–gene locus				
P1	(Fw):	GCCAAGGACAGGTACGGCTGTCATC	61997–62021	β
P2	(Rev):	CCCTTCCTATGACATGAACTTAACCAT	62676–62702	β
β7	(Fw):	TCCTGATGCTGTTATGGGCAA	62469–62489	β
β8	(Rev):	AAAAGCAGAATGGTAGCTGGA	63371–63391	β
β9	(Fw):	AAAAACTTTACACAGTCTGCC	62935–62955	β
β10	(Rev):	ATTAGCTGTTTGCAGCCTCA	63881–63900	β
γ–gene locus				
R159	(Fw):	TGAAACTGTGGTCTTTATGAAAATTG	38792–38817	^{A}γ
R160	(Fw):	GCACTGAAACTGTTGCTTTATAGGAT	33855–33880	^{G}γ
R161	(Rev):	GGCGTCTGGACTAGGAGCTTATTG	34507–34530 39443–39466	^{G}γ ^{A}γ
Gγ1	(Fw):	TGCTGCTAATGCTTCATTACAA	35422–35443 40358–40379	^{G}γ ^{A}γ
Gγ2	(Rev):	AAGTGTGGAGTGTGCACATGA	36183–36203	^{G}γ
AGγ8	(Rev):	AATGGATTGCCAAAACGGTC	35874–35893 40790–40809	^{G}γ ^{A}γ
Aγ10	(Rev):	ACTCAGCTGCAATCAATCCA	41309–41328	^{A}γ
SNPs and Microsatellites				
HS2_F	(Fw):	CCATAGTCCAAGCATGAGCA	8544–8563	β–LCR
HS2_R	(Rev):	ATAGAGGCCACCTGCAAGATA	8986–9006	β–LCR
5'ε _F[1]	(Fw):	TCTCTGTTTGATGACAAATTC	18652–18672	5' ε
5'ε _R[1]	(Rev):	AGTCATTGGTCAAGGCTGACC	19391–19411	5' ε
pre_Gγ_F	(Fw):	GAGTAACTGGAAGATACTGATAA	32965–32987	5' ^{G}γ
pre_Gγ_R	(Rev):	AACTCGATCCATGACCTTGG	33264–33283	5' ^{G}γ
pseudo_β_F	(Fw):	GAACAGAAGTTGAGATAGAGA	46426–46446	Ψβ
pseudo_β_R	(Rev):	ACTCAGTGGTCTTGTGGGCT	47107–47126	Ψβ
3'pseudo β_F	(Fw):	TCTGCATTTGACTCTGTTAGC	49476–49496	3' Ψβ
3'pseudo β_R	(Rev):	GGACCCTAACTGATATAACTA	50069–50089	3' Ψβ
pro_β_F	(Fw):	TAAGAGGTCTCTAGTTTTTTATCTC	61413–61437	5' β
pro_β_R	(Rev):	CTACCATAATTCAGCTTTGGGAT	61745–61767	5' β
3'β_Bam_F	(Fw):	GCTACGGTAACCAAAACAGCA	71809–71829	3' β
3'β_Bam_R	(Rev):	GGGGTTGTTTGTTTTTTCTTG	72350–72370	3' β
Sicilian δβ–Thalassemia (dHPFH)				
F1[2]	(Fw):	TTGGGTTTCTGATAGGCACTG	54954–54974	δ
F2[2]	(Rev):	GTGTCACCCATTAATGCCTTGTAC	56515–56538	3' δ
F3[2]	(Rev):	TAGATCCCTTTGCCATTATG	69463–69482	3' β

[1] (Falchi *et al.*, 2005)
[2] (Craig *et al.*, 1994)

(continued)

PRIMER		SEQUENCE (5'→ 3')	Position	Locus
γ promoter [Cloning]				
Gγ_*Kpn* I	(Fw):	aaa*ggtacc*–TCTTTTAGCCGCCTAACA	33096–33113	$^{G}\gamma$
Aγ_*Kpn* I	(Fw):	aaa*ggtacc*CCTTTTAGCCATCTGTAT	38034–38051	$^{A}\gamma$
γ_Seq	(Fw):	GAAGTGAACCTAGCATTTATAC	33552– 33573 38490– 38511	$^{G}\gamma$ $^{A}\gamma$
R161_*Xho* I	(Rev):	aaa*ctcgag*–TCTGGACTAGGAGCTTATTG	34507–34526 39443–39462	$^{G}\gamma$ $^{A}\gamma$
Gγ_*Kpn* I	(Fw):	aaa*ggtacc*–TCTTTTAGCCGCCTAACA	33096–33113	$^{G}\gamma$

Primer positions are referred to NCBI Reference Sequence U01317(β–cluster).

Table VIII
Hematological Results and Hb composition of analyzed samples

SAMPLE	SEX	HbA_2 %	HbF %	$^{G}\gamma$ %	$^{A}\gamma^{1}$ %	RBC 10^6 /µl	Hb g/dl	MCV fl	MCH pg
BA	F	nd	14.5	–	100	nd	nd	nd	nd
BL*	F	2.6	23.5	22.9	77.1	4.59	11.8↓	80.3↓	25.7
BGy	F	2.1	11.9	–	100	4.92	12.1	75.1↓	24.7↓
BGi*	F	5.7↑	21.4	63.6	36.4	6.14↑	14.3	71.2↓	23.3↓
CA	M	nd	19.4	100	–	nd	nd	nd	nd
CMA*	F	2.6	20.7	–	100	4.52	12.6	82.3	27.8
CV	M	1.8	16.5	traces	100	5.08	13.9	86.6	27.3
CI*	M	2.5	18.8	traces	100	6.37↑	16.1	81.6↓	25.3
CR	M	2.0	13.2	traces	100	5.29	15.6	87.6	29.5
DGI	F	1.7	13.5	25.2	74.8	4.69	11.8↓	80.9	25.2
FVR	F	1.7	28.3	100	–	nd	nd	nd	nd
GM*	M	2.4	20.5	traces	100	nd	nd	nd	nd
MS	F	2.2	15.9	20.4	79.6	5.04	13.6	82	27
MGB*	M	2.3	24.0	23.5	76.5	5.52	14.6	82.6	26.5
OM*	F	5.3↑	8.6	78.9	21.6	5.42	11↓	63.5↓↓	20.3↓↓
PF*	M	2.5	9.5	18.2	81.8	6.11	12.9↓	62.9↓↓	21.1↓↓
PG	F	1.8	15.1	41.6	58.4	nd	nd	nd	nd

* *These samples were confirmed for a β–thalassemic trait by molecular analysis.*
Hematological reference values can be found in Table III.
The up/down arrow indicates an increase/decrease compared to normal values.

Table IX
β–gene framework of analyzed samples

SAMPLE	β–alleles	cod 2 His→His C→T	+16 *Ava* II (C+, G–)	+74 G→T	+81 C→T	+666 T→C	β–gene Type
BA	nd	nd nd	nd nd	nd nd	nd nd	nd nd	nd nd
BL	β^0(39)/β	C C	C C	T T	C C	T T	2 2
BGy	β/β	T T	G G	T T	T T	C C	3 3
BGi	β^+(–87)/β	C C	C C	T T	C C	T T	2 2
CA	β/β	C T	C G	G T	C T	T C	1 3
CMA	β^0(39)/β	C C	C C	nd nd	nd nd	T T	nd nd
CV	β/β	C T	C G	G T	C T	T C	1 3
CI	β^0(39)/β	C C	C C	G G	C C	T T	1 1
CR	β/β	T T	G G	T T	T T	C C	3 3
DGI	β/β	C T	C G	G T	C T	T C	1 3
FVR	β/β	C C	C C	nd nd	nd nd	T T	nd nd
GM	β^0(39)/β	C C	C C	G T	C T	T T	1 Rec.
MS	β/β	C T	C G	G T	C T	T C	1 3
MGB	β^0(39)/β	C C	C C	G G	C C	T T	1 1
OM	β^0(39)/β	C C	C C	G T	C C	T T	1 2
PF*	–/β	nd nd	nd nd	nd nd	nd nd	nd nd	nd nd
PG	β/β	C T	C G	G T	C T	T C	1 3

+16, +74, +81, +666 nucleotides are numbered from the first nucleotide in IVS2.
* *PF is deleted on β–cluster (Sicilian δβ–thalassemia).*

Table X
Hemoglobin profiles and molecular data of HPFH samples

$^{G}\gamma$ –175 T→C

Sample	%HbF	%$^{G}\gamma$	%$^{A}\gamma$	ndHPFH		β–gene	β–gene Type	%HbA_2
				$^{G}\gamma$ *Xmn* I	$^{G}\gamma$ –175 T→C			
CA	19.4	100		–/–	T/C	β/β	1/3	nd
FVR	28.3	100		–/–	T/C	β/β	nd	1.7

$^{A}\gamma$ –196 C→T

Sample	%HbF	%$^{G}\gamma$	%$^{A}\gamma$	ndHPFH		β–gene	β–gene Type	%HbA_2
				$^{G}\gamma$ *Xmn* I	$^{A}\gamma$ –196 C→T			
CI*	18.8	traces	100	–/–	C/T	β^0(39)/β	1/1	2.5
GM	20.5	traces	100	nd	C/T	β^0(39)/β	1/Rec.	2.4
CMA	20.7		100	nd	C/T	β^0(39)/β	nd/nd	2.6
MGB	24	23.5	76.5	–/+	C/T	β^0(39)/β	1/1	2.3
BL**	23.5	22.9	77.1	–/–	C/T	β^0(39)/β	2/2	2.6

$^{A}\gamma$ –117 G→A

Sample	%HbF	%$^{G}\gamma$	%$^{A}\gamma$	ndHPFH		β–gene	β–gene Type	%HbA_2
				$^{G}\gamma$ *Xmn* I	$^{A}\gamma$ –117 G→A			
BA	14.5	traces	100	nd	G/A	nd	nd	nd
BGy	11.9		100	nd	G/A	β/β	3/3	2.1
CR*	13.2	traces	100	nd	G/A	β/β	3/3	2.0
CV*	16.5	traces	100	nd	G/A	β/β	1/3	1.8
MS	15.9	20.4	79.6	–/–	G/A	β/β	1/3	2.2
PG*	15.1	41.6	58.4	–/–	G/A	β/β	1/3	1.8
DGI	13.5	25.2	74.8	–/+	G/A	β/β	1/3	1.7

(continued)

$^{G}\gamma$ –158 ($^{G}\gamma$ *Xmn* I) C→T

Sample	%HbF	%$^{G}\gamma$	%$^{A}\gamma$	ndHPFH		β–gene	β–gene Type	%HbA_2
				$^{G}\gamma$ *Xmn* I	Other ndHPFH			
BGi	21.4	63.6	36.4	–/+	None	β^{+}(–87)/β	2/2	5.7
OM*	8.6	78.9	21.6	–/+	None	β^{0}(39)/β	1/2	5.3
PF***	9.5	18.2	81.8	–/+	None	–/β	nd	2.5

* *The samples are heterozygous for AGCA deletion (–225/–222) in the $^{A}\gamma$ promoter.*
** *The sample BL presents two substitutions (–474 C→T and –309 A→G) on the $^{G}\gamma$ promoter*
*** *PF is deleted on β–cluster (Sicilian δβ–thalassemia)*

Table XI
Haplotypes of β–globin gene cluster in the analyzed samples

SMP	ndHPFH	β–alleles	*Hinc* II 5' ε	*Hind* III $^{G}\gamma$	*Hind* III $^{A}\gamma$	*Hinc* II Ψβ	*Hinc* II 3' Ψβ	*Ava* II β	*Bam* HI 3' β	Orkin Haplotype
BGi	$^{G}\gamma$ *Xmn* I +/–	$\beta^{+}(-87)/\beta$	+ –	– +	– –	– +	– –	+ +	+ –	I VIII
BL	$^{A}\gamma$ –196 C/T	$\beta^{0}(39)/\beta$	+ +	– +	– –	– –	– –	+ +	+ +	I I Rec.
CI	$^{A}\gamma$ –196 C/T	$\beta^{0}(39)/\beta$	+ +	– +	– –	– –	– –	+ +	+ +	I I Rec.
CV	$^{A}\gamma$ –117 G/A	β/β	+ +	– –	– –	– –	– –	– +	+ +	VII I Rec.
DGI	$^{G}\gamma$ *Xmn* I +/– $^{A}\gamma$ –117 G/A	β/β	+ –	– +	– –	– +	– –	– +	+ +	VII Rec.

The differences from original reference haplotype (Orkin *et al.*, 1982) *are colored in red.*

REFERENCES

Albitar M, Care A, Peschle C, Liebhaber SA (1992) Developmental switching of messenger RNA expression from the human alpha–globin cluster: fetal/adult pattern of theta–globin gene expression. *Blood* **80:** 1586–1591

Ayi K, Turrini F, Piga A, Arese P (2004) Enhanced phagocytosis of ring–parasitized mutant erythrocytes: a common mechanism that may explain protection against falciparum malaria in sickle trait and beta–thalassemia trait. *Blood* **104:** 3364–3371

Bank A (2006) Regulation of human fetal hemoglobin: new players, new complexities. *Blood* **107:** 435–443

Barbosa CG, Goncalves–Santos NJ, Souza–Ribeiro SB, Moura–Neto JP, Takahashi D, Silva DO, Hurtado–Guerrero AF, Reis MG, Goncalves MS (2010) Promoter region sequence differences in the A and G gamma globin genes of Brazilian sickle cell anemia patients. *Brazilian journal of medical and biological research* **43:** 705–711

Berry M, Grosveld F, Dillon N (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature* **358:** 499–502

Bharadwaj RR, Trainor CD, Pasceri P, Ellis J (2003) LCR–regulated transgene expression levels depend on the Oct–1 site in the AT–rich region of beta–globin intron–2. *Blood* **101:** 1603–1610

Bieker JJ (2010) Putting a finger on the switch. *Nature genetics* **42:** 733–734

Bilgen T, Arikan Y, Canatan D, Yesilipek A, Keser I (2011) The association between intragenic SNP haplotypes and mutations of the beta globin gene in a Turkish population. *Blood cells, molecules & diseases* **46:** 226–229

Borg J, Papadopoulos P, Georgitsi M, Gutierrez L, Grech G, Fanis P, Phylactides M, Verkerk AJ, van der Spek PJ, Scerri CA, Cassar W, Galdies R, van Ijcken W, Ozgur Z, Gillemans N, Hou J, Bugeja M, Grosveld FG, von Lindern M, Felice AE, Patrinos GP, Philipsen S (2010) Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nature genetics* **42:** 801–805

Caterina JJ, Ciavatta DJ, Donze D, Behringer RR, Townes TM (1994) Multiple elements in human beta–globin locus control region 5' HS 2 are involved in enhancer activity and position–independent, transgene expression. *Nucleic acids research* **22:** 1006–1011

Chen Z, Luo HY, Basran RK, Hsu TH, Mang DW, Nuntakarn L, Rosenfield CG, Patrinos GP, Hardison RC, Steinberg MH, Chui DH (2008) A T–to–G transversion at nucleotide –567 upstream of HBG2 in a GATA–1 binding motif is associated with elevated hemoglobin F. *Molecular and cellular biology* **28:** 4386–4393

Choi OR, Engel JD (1988) Developmental regulation of beta–globin gene switching. *Cell* **55:** 17–26

Close J, Game L, Clark B, Bergounioux J, Gerovassili A, Thein SL (2004) Genome annotation of a 1.5 Mb region of human chromosome 6q23 encompassing a quantitative trait locus for fetal hemoglobin expression in adults. *BMC genomics* **5:** 33

Collins FS, Weissman SM (1984) The molecular genetics of human hemoglobin. *Progress in nucleic acid research and molecular biology* **31:** 315–462

Craig JE, Barnetson RA, Prior J, Raven JL, Thein SL (1994) Rapid detection of deletions causing delta beta thalassemia and hereditary persistence of fetal hemoglobin by enzymatic amplification. *Blood* **83:** 1673–1682

Craig JE, Rochette J, Fisher CA, Weatherall DJ, Marc S, Lathrop GM, Demenais F, Thein S (1996) Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nature genetics* **12:** 58–64

de Andrade TG, Peterson KR, Cunha AF, Moreira LS, Fattori A, Saad ST, Costa FF (2006) Identification of novel candidate genes for globin regulation in erythroid cells containing large deletions of the human beta–globin gene cluster. *Blood cells, molecules & diseases* **37:** 82–90

De Angioletti M, Lacerra G, Pagano L, Alessi M, D'Avino R, Manca L, Carestia C (2004) Beta–thalassaemia–87 C—>G: relationship of the Hb F modulation and polymorphisms in compound heterozygous patients. *British journal of haematology* **126:** 743–749

Elnitski L, Miller W, Hardison R (1997) Conserved E boxes function as part of the enhancer in hypersensitive site 2 of the beta–globin locus control region. Role of basic helix–loop–helix proteins. *The Journal of biological chemistry* **272:** 369–378

Falchi A, Giovannoni L, Vacca L, Latini V, Vona G, Varesi L (2005) Beta–globin gene cluster haplotypes associated with beta–thalassemia on Corsica island. *American journal of hematology* **78:** 27–32

Fei YJ, Fujita S, Huisman TH (1988) Two different theta globin gene deletions observed among black newborn babies. *British journal of haematology* **68:** 249–253

Fischel–Ghodsian N, Nicholls RD, Higgs DR (1987) Unusual features of CpG–rich (HTF) islands in the human alpha globin complex: association with non–functional pseudogenes and presence within the 3' portion of the zeta gene. *Nucleic acids research* **15:** 9215–9225

Forget BG (1998) Molecular basis of hereditary persistence of fetal hemoglobin. *Annals of the New York Academy of Sciences* **850:** 38–44

Forget BG, Bunn HF (2013) Classification of the disorders of hemoglobin. *Cold Spring Harbor perspectives in medicine* **3:** a011684

Fraser P, Pruzina S, Antoniou M, Grosveld F (1993) Each hypersensitive site of the human beta–globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes & development* **7:** 106–113

Galanello R, Dessi E, Melis MA, Addis M, Sanna MA, Rosatelli C, Argiolu F, Giagu N, Turco MP, Cacace E, et al. (1989) Molecular analysis of beta0–thalassemia intermedia in Sardinia. *Blood* **74:** 823–827

Galanello R, Sanna S, Perseu L, Sollaino MC, Satta S, Lai ME, Barella S, Uda M, Usala G, Abecasis GR, Cao A (2009) Amelioration of Sardinian beta0–thalassemia by genetic modifiers. *Blood* **114:** 3935–3937

Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G (2010) Fine–mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nature genetics* **42:** 1049–1051

Garner CP, Tatu T, Best S, Creary L, Thein SL (2002) Evidence of genetic interaction between the beta–globin complex and chromosome 8q in the expression of fetal hemoglobin. *American journal of human genetics* **70:** 793–799

Goh SH, Lee YT, Bhanu NV, Cam MC, Desper R, Martin BM, Moharram R, Gherman RB, Miller JL (2005) A newly discovered human alpha–globin gene. *Blood* **106:** 1466–1472

Haj Khelil A, Moriniere M, Laradi S, Khelif A, Perrin P, Ben Chibani J, Baklouti F (2011) Xmn I polymorphism associated with concomitant activation of Ggamma and Agamma globin gene transcription on a beta0–thalassemia chromosome. *Blood cells, molecules & diseases* **46:** 133–138

Harju S, McQueen KJ, Peterson KR (2002) Chromatin structure and control of beta–like globin gene switching. *Experimental biology and medicine* **227:** 683–700

Henke W, Herdel K, Jung K, Schnorr D, Loening SA (1997) Betaine improves the PCR amplification of GC–rich DNA sequences. *Nucleic acids research* **25:** 3957–3958

Hsu SL, Marks J, Shaw JP, Tam M, Higgs DR, Shen CC, Shen CK (1988) Structure and expression of the human theta1 globin gene. *Nature* **331:** 94–96

Kong Y, Zhou S, Kihm AJ, Katein AM, Yu X, Gell DA, Mackay JP, Adachi K, Foster–Brown L, Louden CS, Gow AJ, Weiss MJ (2004) Loss of alpha–hemoglobin–stabilizing protein impairs erythropoiesis and exacerbates beta–thalassemia. *The Journal of clinical investigation* **114:** 1457–1466

Kukreti S, Kaur H, Kaushik M, Bansal A, Saxena S, Kaushik S, Kukreti R (2010) Structural polymorphism at LCR and its role in beta–globin gene regulation. *Biochimie* **92:** 1199–1206

Lapoumeroulie C, Castiglia L, Ruberto C, Fichera M, Amata S, Labie D, Ragusa A (1999) Genetic variations in human fetal globin gene microsatellites and their functional relevance. *Human genetics* **104:** 307–314

Leach KM, Vieira KF, Kang SH, Aslanian A, Teichmann M, Roeder RG, Bungert J (2003) Characterization of the human beta–globin downstream promoter region. *Nucleic acids research* **31:** 1292–1301

Lettre G (2012) The search for genetic modifiers of disease severity in the beta–hemoglobinopathies. *Cold Spring Harbor perspectives in medicine* **2**

Lettre G, Sankaran VG, Bezerra MA, Araujo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN, Orkin SH (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:** 11869–11874

Li Q, Duan ZJ, Stamatoyannopoulos G (2001) Analysis of the mechanism of action of non–deletion hereditary persistence of fetal hemoglobin mutants in transgenic mice. *The EMBO journal* **20:** 157–164

Loudianos G, Moi P, Lavinha J, Galanello R, Cao A, Pirastu M (1992) Normal delta–globin gene sequences in Sardinian nondeletional delta beta–thalassemia. *Hemoglobin* **16:** 503–509

Lozzio CB, Lozzio BB (1975) Human chronic myelogenous leukemia cell–line with positive Philadelphia chromosome. *Blood* **45:** 321–334

Manca L, Formato M, Demuro P, Pilo G, Gallisai D, Orzalesi M, Masala B (1986) The gamma globin chain heterogeneity of the Sardinian newborn baby. *Hemoglobin* **10:** 519–528

Manca L, Masala B (1990) A simple approach to the determination of the gamma chain composition of HbF in adult human blood samples. *Hemoglobin* **14:** 517–527

Manca L, Masala B (2008) Disorders of the synthesis of human fetal hemoglobin. *IUBMB life* **60:** 94–111

Masala B, Manca L (1991) Detection of the common Hb F Sardinia [Agamma(E19)Ile—>Thr] variant by isoelectric focusing in normal newborns and in adults affected by elevated fetal hemoglobin syndromes. *Clinica chimica acta; international journal of clinical chemistry* **198:** 195–202

Masala B, Manca L (1994) Detection of globin chains by reversed–phase high–performance liquid chromatography. *Methods in enzymology* **231:** 21–44

Masala B, Manca L, Gallisai D, Stangoni A, Lanclos KD, Kutlar F, Yang KG, Huisman TH (1988) Biochemical and molecular aspects of beta–thalassemia types in northern Sardinia. *Hemoglobin* **12:** 661–671

Maston GA, Evans SK, Green MR (2006) Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics* **7:** 29–59

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids research* **16:** 1215

Nienhuis AW, Nathan DG (2012) Pathophysiology and Clinical Manifestations of the beta–Thalassemias. *Cold Spring Harbor perspectives in medicine* **2:** a011726

Nuinoon M, Makarasara W, Mushiroda T, Setianingsih I, Wahidiyat PA, Sripichai O, Kumasaka N, Takahashi A, Svasti S, Munkongdee T, Mahasirimongkol S, Peerapittayamongkol C, Viprakasit V, Kamatani N, Winichagoon P, Kubo M, Nakamura Y, Fucharoen S (2010) A genome–wide association identified the common genetic variants influence disease severity in beta0–thalassemia/hemoglobin E. *Human genetics* **127:** 303–314

Orkin SH, Kazazian HH, Jr., Antonarakis SE, Goff SC, Boehm CD, Sexton JP, Waber PG, Giardina PJ (1982) Linkage of beta–thalassaemia mutations and beta–globin gene polymorphisms with DNA polymorphisms in human beta–globin gene cluster. *Nature* **296:** 627–631

Ottolenghi S, Mantovani R, Nicolis S, Ronchi A, Giglioni B (1989) DNA sequences regulating human globin gene transcription in non–deletional hereditary persistence of fetal hemoglobin. *Hemoglobin* **13:** 523–541

Palstra RJ, de Laat W, Grosveld F (2008) Beta–globin regulation and long–range interactions. *Advances in genetics* **61:** 107–142

Patrinos GP, de Krom M, de Boer E, Langeveld A, Imam AM, Strouboulis J, de Laat W, Grosveld FG (2004) Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes & development* **18:** 1495–1509

Piras I, Vona G, Falchi A, Latini V, Ristaldi S, Vacca L, Varesi L, Calo CM (2005) Beta–globin cluster haplotypes in normal individuals and beta(0)39–thalassemia carriers from Sardinia, Italy. *American journal of human biology: the official journal of the Human Biology Council* **17:** 765–772

Pirastu M, Galanello R, Doherty MA, Tuveri T, Cao A, Kan YW (1987) The same beta–globin gene mutation is present on nine different beta–thalassemia chromosomes in a Sardinian population. *Proceedings of the National Academy of Sciences of the United States of America* **84:** 2882–2885

Pistidda P, Frogheri L, Oggiano L, Guiso L, Manca L, Dore F, Masala B, Gilman JG, Longinotti M (1995) Fetal hemoglobin expression in compound heterozygotes for –117 (G—>A) Agamma HPFH and beta zero 39 nonsense thalassemia. *American journal of hematology* **49:** 267–270

Rees WA, Yager TD, Korte J, von Hippel PH (1993) Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* **32:** 137–144

Riou J, Godart C, Hurtrel D, Mathis M, Bimet C, Bardakdjian–Michau J, Prehu C, Wajcman H, Galacteros F (1997) Cation–exchange HPLC evaluated for presumptive identification of hemoglobin variants. *Clinical chemistry* **43:** 34–39

Rixon MW, Gelinas RE (1988) A fetal globin gene mutation in Agamma nondeletion hereditary persistence of fetal hemoglobin increases promoter strength in a nonerythroid cell. *Molecular and cellular biology* **8:** 713–721

Ronchi A, Nicolis S, Santoro C, Ottolenghi S (1989) Increased Sp1 binding mediates erythroid–specific overexpression of a mutated (HPFH) gamma–globulin promoter. *Nucleic acids research* **17:** 10231–10241

Rosatelli C, Leoni GB, Tuveri T, Scalas MT, Di Tucci A, Cao A (1987) Beta thalassaemia mutations in Sardinians: implications for prenatal diagnosis. *Journal of medical genetics* **24:** 97–100

Rosatelli MC, Tuveri T, Scalas MT, Leoni GB, Sardu R, Faa V, Meloni A, Pischedda MA, Demurtas M, Monni G, et al. (1992) Molecular screening and fetal diagnosis of beta–thalassemia in the Italian population. *Human genetics* **89:** 585–589

Sambrook J, Russel DW (2001) *Molecular cloning: a laboratory manual*, 3rd ed. edn. Cold Spring Harbor, N.Y. *Cold Spring Harbor Laboratory Press*.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain–terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74:** 5463–5467

Sankaran VG, Xu J, Orkin SH (2010) Transcriptional silencing of fetal hemoglobin by BCL11A. *Annals of the New York Academy of Sciences* **1202:** 64–68

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation–dependent probe amplification. *Nucleic acids research* **30:** e57

Stamatoyannopoulos G, Josephson B, Zhang JW, Li Q (1993) Developmental regulation of human gamma–globin genes in transgenic mice. *Molecular and cellular biology* **13:** 7636–7644

Steinberg MH, Forget B (2009) Disorders of hemoglobin: genetics, pathophysiology, and clinical management, 2nd Edition. *Cambridge University Press.*

Sturgeon P, Itano HA, Bergren WR (1955) Genetic and biochemical studies of intermediate types of Cooley's anaemia. *British journal of haematology* **1:** 264–277

Superti–Furga G, Barberis A, Schaffner G, Busslinger M (1988) The –117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the gamma–globin gene. *The EMBO journal* **7:** 3099–3107

Taher AT, Musallam KM, Cappellini MD (2009) Thalassaemia intermedia: an update. *Mediterranean journal of hematology and infectious diseases* **1:** e2009004

Takahashi T, Schreiber R, Krieger JE, Saad ST, Costa FF (2003) Analysis of the mechanism of action of the Brazilian type (Agamma –195 C—>G) of hereditary persistence of fetal hemoglobin. *European journal of haematology* **71:** 418–424

Tang Y, Liu DP, Liang CC (2002) Further understanding of the beta–globin locus regulation at the molecular level: looping or linking models? *Genes to cells: devoted to molecular & cellular mechanisms* **7:** 889–900

Thein SL (2008) Genetic modifiers of the beta–haemoglobinopathies. *British journal of haematology* **141:** 357–366

Thein SL (2013) The molecular basis of beta–thalassemia. *Cold Spring Harbor perspectives in medicine* **3:** a011700

Thein SL, Craig JE (1998) Genetics of HbF/F cell variance in adults and heterocellular hereditary persistence of fetal hemoglobin. *Hemoglobin* **22:** 401–414

Thein SL, Menzel S, Peng X, Best S, Jiang J, Close J, Silver N, Gerovasilli A, Ping C, Yamaguchi M, Wahlberg K, Ulug P, Spector TD, Garner C, Matsuda F, Farrall M, Lathrop M (2007) Intergenic variants of HBS1L–MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 11346–11351

Thein SL, Sampietro M, Rohde K, Rochette J, Weatherall DJ, Lathrop GM, Demenais F (1994) Detection of a major gene for heterocellular hereditary persistence of fetal hemoglobin after accounting for genetic modifiers. *American journal of human genetics* **54:** 214–228

Trudel M, Costantini F (1987) A 3' enhancer contributes to the stage–specific expression of the human beta–globin gene. *Genes & development* **1:** 954–961

Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, Piras MG, Sestu N, Lai S, Dei M, Mulas A, Crisponi L, Naitza S, Asunis I, Deiana M, Nagaraja R, Perseu L, Satta S, Cipollina MD, Sollaino C, Moi P, Hirschhorn JN, Orkin SH, Abecasis GR, Schlessinger D, Cao A (2008) Genome–wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta–thalassemia. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 1620–1625

Wahlberg K, Jiang J, Rooks H, Jawaid K, Matsuda F, Yamaguchi M, Lathrop M, Thein SL, Best S (2009) The HBS1L–MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells. *Blood* **114:** 1254–1262

Watson HC, Kendrew JC (1961) The amino–acid sequence of sperm whale myoglobin. Comparison between the amino–acid sequences of sperm whale myoglobin and of human hemoglobin. *Nature* **190:** 670–672

Weatherall D (2010) Thalassemia: the long road from the bedside through the laboratory to the community. *Nature medicine* **16:** 1112–1115

Xu J, Bauer DE, Kerenyi MA, Vo TD, Hou S, Hsu YJ, Yao H, Trowbridge JJ, Mandel G, Orkin SH (2013) Corepressor–dependent silencing of fetal hemoglobin expression by BCL11A. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 6518–6523

PROTOCOLS

Protocol 1

Salting Out

DNA extraction from Whole Blood

SOLUTIONS FOR SALTING OUT

- **LYSIS BUFFER:** 0.32 M Surcose, 10 mM Tris/Cl pH 7.5, 5 mM $MgCl_2$, 1 % Triton X–100
- **PHYSIO BUFFER:** 0.075 M NaCl, 0.025 M EDTA
- **BUFFER A:** 10 mM Tris/Cl pH 8.0, 2 mM EDTA
- **Proteinase K:** 10 mg/ml in 10 % SDS, 20 mM EDTA
- **NaCl** 6 M

LYSIS OF RBC

- Add cold **Lysis Buffer** (4°C) to **2 ml** of **whole blood** in EDTA, up to a **volume of 10 ml**
- Mix gently by inversion
- Incubate at **4°C** for **15'–30'**
- Centrifuge **20'** at **1500 g**
- Remove the supernatant

WASH AND LYSIS OF WBC

- Resuspend the pellet in **Physio Buffer** up to a **volume of 10 ml**
- Centrifuge **10'** at **1500 g**
- Repeat the wash 2 more times with Physio Buffer
- Discard the supernatant and lyse the pellet in **3 ml** of **Buffer A**
- Vortex to suspend and break up the pellet.

DENATURATION AND REMOVAL OF PROTEINS

- Add **100 ml** of **10% SDS**
- Add **400–1000 µl** of **proteinase K in SDS**
- Incubate **37°C overnight**
- When the sample reaches the room temperature and add **1 ml** of **NaCl 6 M**
- Shake vigorously for 15''
- Centrifuge **15'** at **1500 g**
- Transfer the supernatant into a clean tube

PRECIPITATION AND RESUSPENSION OF DNA

- Add 1 volume of **isopropanol 100%**
- Centrifuge **15'** at **2200 g** and discard the supernatant
- Add **1 ml** of ethanol **70%**
- Centrifuge **15'** at **2200 g** and discard the supernatant
- Wait for the evaporation of all EtOH
- Add **TE** pH = 8 (**500 µl – 1 ml**)

Protocol 2
Phenol:chloroform DNA purification

- Add **mQ water up to 200 µl** to the solution containing DNA to be precipitated
- Add **200 µl** of **phenol:chloroform:isoamylalcohol** (25:24:1) (pH = 8)
- Vortex for a few seconds to obtain a homogeneous emulsion
- Centrifuge at top speed (**20900 g**) for **5'**
- Recover the aqueous phase and put it into a new 1.5 ml tube
- Add **CH_3COONa** (3M, pH 5.5): **1/10** of the recovered aqueous solution
- Add **2–2.5** volumes of **100% EtOH** at room temperature
- Add **1 µl** of **glycogen 10 mg/ml**
- Incubate at **–80° C** for 30' or **–20°C** overnight
- Centrifuge at **top speed** for **10'**
- Remove the supernatant
- Add **500 µl** of **70% EtOH**
- Resuspend the pellet
- Centrifuge at **top speed** for **5'**
- Remove all the ethanol
- Resuspend in **mQ water**

Protocol 3

Preparation of $CaCl_2$ competent cells

Competence should be **$5 \cdot 10^5$ to $1 \cdot 10^6$** Cloning forming Units **per 1 µg** of supercoiled plasmid DNA

- Plate DH5α cells on LB agar without antibiotic and grow them overnight
- Inoculate a colony of DH5α and incubate in **2 ml** of **LB**, 37°C overnight, in agitation
- Inoculate **500 µl** of mini–inoculum in **100 m**l of **LB**
- Growth at 37°C in agitation
- Stop the growth by putting the flask on ice (5–10') when the Abs (OD) at 600 nm is 0.4

- Pellet the bacteria by centrifugation at **1700 g** for 1**0'**, at **4° C**
- Remove the supernatant and resuspend with **20 ml** ice cold **0.1 M $CaCl_2$**
- Centrifuge again and remove the supernatant
- Resuspend in **4 ml** of ice cold **$CaCl_2$** + **10%** sterile **glycerol**
- Aliquot and store at –80°C

Protocol 4
Transformation of DH5α

- Add **2–10 ng** of supercoiled **plasmid** or **½ volume** of the **ligase** reaction to a cold tube
- Add **100 µl** of **competent cells** to the tube
- Incubate on ice for 30'
- Incubate 3' at 37°C (thermal shock)
- Incubate for 10' at room temperature

- In sterility, add **500 µl** of **LB broth** (without antibiotic)
- Incubate at 37°C for 50'
- Centrifuge for **5'** at **1900 g**
- Remove 500 µl of supernatant
- Resuspend the pellet with the remaining volume of LB
- Plate the cells on LB agar plate with specific antibiotic
- Incubate at 37°C overnight

Protocol 5

Mini–preparation

Alkaline lysis in presence of SDS

SOLUTIONS FOR MINI–PREPARATION

Solution I: 50 mM glucose, EDTA10 mM Tris/Cl pH 8 25 mM
Solution II: 200 mM NaOH, 1% SDS
Solution III: 7.5 M NH_4–acetate

- Transfer 1 ml of mini–inoculation in a 2 ml tube
- Centrifuge at **20900 g** for **1'** and remove the supernatant
- Resuspend the pellet in **200 µl** of **Solution I** (Isotonic)
- Add **200 µl** of **Solution II** (alkaline lysis)
- Mix gently by inversion
- Incubate no more than 5' at room temperature
- Add **200 µl** of cold **Solution III**
- Incubate on ice 10'
- Centrifuge at **20900 g** for **10'**
- Recover the supernatant containing plasmid DNA and low molecular weight contaminants
- Add **2 volumes** of **100% EtOH** and incubate 10' at room temperature
- Centrifuge at **20900 g** for **15'**
- Resuspend the pellet with **500 µl** of **80% EtOH**
- Centrifuge at **20900 g** for **5'**
- Resuspend in **50 µ**l of **TE** containing **RNase A** (20 mg/ml)
- Incubate at 37 ° C for 10'
- Use the extracted DNA for restriction digestion

LIST OF ABBREVIATIONS

ACH:	Active Chromatin Hub
AHSP:	α–hemoglobin stabilizing protein
AUT–PAGE:	Acetic Acid, Urea, Triton X–100 PAGE
CE–HPLC:	Cation Exchange HPLC
GWAS:	Genome–Wide Association Studies
dHPFH:	Deletional HPFH
Hb:	Hemoglobin
HbA:	Adult Hemoglobin
HBA1:	α1–globin gene
HbA_2:	Minor Adult Hemoglobin
HBA2:	α2–globin gene
HBB:	β–globin gene
HBD:	δ–globin gene
HbF:	Fetal Hemoglobin
HBG1:	$^{A}\gamma$–globin gene
HBG2:	$^{G}\gamma$–globin gene
hHPFH:	Heterocellular HPFH
HPFH:	Hereditary Persistence of Fetal Hemoglobin
HPLC:	High Performance Liquid Chromatography
HS1–5:	DNase hypersensitive sites
IEF:	Isoelectric focusing
IVS:	Intervening sequence, Intron
LCR:	Locus Control Region
LD:	Linkage disequilibrium
MCH:	Mean Corpuscular Hemoglobin
MCV:	Mean Corpuscular Volume
ndHPFH:	non–deletional HPFH
OR:	Olfactory Receptor
ORF:	Open reading frame
pHPFH:	Pancellular HPFH
QTL:	Quantitative trait locus
RBC:	Red blood Cells
RP–HPLC:	Reversed Phase HPLC
SCD:	Sickle cell disease
UTR:	Untranslated region