The Effects of Trans-Regulatory Loci on Globin Gene Control

Amal Mohammed Daw
February 2012

Submitted to the Faculty of Medicine and Surgery, University of Malta in Partial Fulfillment of the Requirements for the Degree of Master of Science

Supervisor: Professor Alex. E. Felice
I hereby declare that all material in this research project is my original work

Amal Mohammed Daw
To my family
Acknowledgements

I would like to thank a number of people who helped me accomplish my experiments for the successful completion of this project.

First I wish to thank deeply my supervisor Prof Dr. Alex. E. Felice MD PhD., for trusting in me to carry out work in his laboratory. His guidance and support will never go unnoticed. This experience has been extremely rewarding. Dr. Godfrey Grech was instrumental for my MSc work whose critical and important points ensured proper scientific work along the way by asking the right questions. And Dr. Joseph Borg was the strong supporter for me a long all period of my work.

I wish to express my sincere gratitude to Ms. Ruth Galdies B.Pharm., M.Sc and Ms. Wilma Cassar B.Pharm., for their constant help in laboratory practice and technical advice, and the all the staff at the Laboratory of Molecular Genetics, University of Malta, namely Dr. Christian. A. Scerri MD PhD, and others.

My ultimate credits go to my parents Mohammed and Fatama and most of all, my husband Murad. He has been indispensable for the completion of this work, very patient and good motivator. Thank you so much.

This work was partially supported by the Libyan Government Scholarship Scheme.
Abstract

Augmenting the level of foetal haemoglobin (HbF) in sickle cell disease or β-thalassemia patients would greatly ameliorate the symptoms associated with these diseases. This can only be achieved by first understanding the genetic switch from foetal to adult haemoglobin that is commonly referred to as γ to β globin gene switching. In this study a combination of clinical research coupled with basic research was carried out. One hundred seven β-thalassemia heterozygotes were identified by the screening program carried out by the Laboratory of Molecular Genetics, University of Malta. A complete blood count accompanied all blood samples to the laboratory. HbF and HbA2 measurements were conducted using a High-Pressure Liquid Chromatography. DNA sequencing of all samples was performed for the β globin gene and its promoter. The results showed the commonly encountered β thalassaemia alleles in the Maltese population; these being IVSI-6C, IVSI-110A, Codon 39 T, IVSII-1A and other rare ones as well including a deletion of 2 nts in the β globin gene coding sequence identified for the first time in Malta. The -158 C\(\rightarrow\)T 5'Gγ globin SNP genotyping was performed by XmnI restriction enzyme digest. Three polymorphisms (rs766432, rs11886868 and rs4671393) in BCL11A and another three (rs4895441, rs28384513 and rs9399137) in MYB were genotyped using Real Time PCR. There was a positive association between the MYB and HbF levels in β0 Codon 39 heterozygotes whilst not with other β thalassaemia alleles. Only one dimorphism in BCL11A was associated with higher HbF levels in β thalassaemia heterozygotes. As exemplified by EKLF in previous work carried out in the laboratory, and other transcription factors such as BCL11A and MYB in this study are thought to play a significant role in promoter-specific gene activation and warrant further investigation regarding their role in globin switching.
1.0 Introduction

1.1 Structure and Function of Haemoglobin
   1.1.0 Structure
   1.1.1 Developmental Globin Chain Synthesis
   1.1.2 Regulation
      1.1.2.1 Molecular Function

1.2 Erythropoiesis

1.3 Haemoglobin Variants and Disorders
   1.3.1 Thalassaemia

1.4 Control of Hb F

1.5 Haemoglobin F-Malta-I

1.6 Hereditary Persistence of Foetal Haemoglobin (HPFH)
   1.6.1 Deletional Hereditary Persistence of Foetal Haemoglobin
   1.6.2 Non-Deletional Hereditary Persistence of Foetal Haemoglobin

1.7 Erythroid Krüppel-Like Factor

1.8 B-cell CLL/lymphoma 11A (BCL11A)

1.9 Therapeutic Approaches

1.10 Specific Aim

2.0 Materials and Methods

2.1 Sample Collection and Preparation

2.2 Quantification of Haemoglobin and Globin Chains by HPLC

2.3 DNA extraction
   2.3.1 DNA extraction from whole blood using the BACC1 Nucleon Kit

2.4 Checking Genomic DNA Integrity

2.5 UV Spectrophotometry
2.6 PCR Primer Design
2.7 Polymerase Chain Reaction
  2.7.1 Preparation of the PCR Reaction Mixture
  2.7.2 Thermal profile for the Mutations and SNPs tested
  2.7.3 PCR product determination by agarose gel electrophoresis
2.8 Restriction enzyme digestion
  2.8.0 Preparation of reaction mixtures
2.9 Detection of the digested products using agarose gel electrophoresis
2.10 DNA Sequencing
  2.10.1 DNA Purification using centrifugal filter devices
2.11 Real-Time PCR assay
  2.11.1 Real Time PCR Analysis
2.12 Statistics and Data Analysis
3.0 Results
3.1 Molecular Epidemiology
  3.1.1 Frequency of β thalassaemia in the Maltese Population
  3.1.2 β globin gene sequencing reveals a rare deletion in the human
  β globin gene.
  3.1.3 Age and Gender distribution of individuals who are β
  thalassaemia heterozygote
  3.1.4 Histogram graphs for HbA2 and MCV
  3.1.5 The levels of Foetal Haemoglobin (g/dL) in β Thalassaemia
  Heterozygotes.
3.2 Genotyping of the Xmn I and (AT)XTY polymorphisms in the β
  Thalassaemia Heterozygotes.
3.3 Genotyping of modifier SNPs in BCL11A and MYB genes
  3.3.1 In vivo HbF expression varies depending on the number of
  minor allele frequency (MAF) SNPs in MYB but not in BCL11A
4.0 Discussion
  4.0.1 heamatology
4.1 Quantitative Trait Loci
4.2 Future Perspectives
4.3 Conclusion
References
Appendix A
Appendix B
Appendix C

Contents

74
75
89
92
93
Figure 1-1: Developmental Globin Chain synthesis 3
Figure 1-2: Developmental progression of a mature erythrocyte 6
Figure 1-3: A schematic view of the beta globin locus recombination 15
Figure 1-4: Deletions of the β-globin gene cluster associated with fusion proteins and HPFH 16
Figure 1-5: DNA map of point mutations causing non-deletional HPFH 18
Figure 1-6: Phenotypes caused by KLF1 mutations 21
Figure 1-7: Schematic diagram of the KLF1 molecule 23
Figure 1-8: A diagram of the postulated effects of HU in inhibiting HbS 25
Figure 2-1: Separation of globin chains by reverse phase HPLC on a BioRad VARIANT testing system. 29
Figure 2.2: Dissociation Curve Analysis from the ABI 7300 System Sequence Detection Software (SDS) 41
Figure 3-1: The distribution of β Thalassaemia alleles in the Maltese Population 43
Figure 3-2: DNA sequence chromatogram showing the location of the deleted 2 base pairs (-CT) and is marked with an arrow (↓) 44
Figure 3-3: The level of haemoglobin across all β Thalassaemia heterozygotes 46
Figure 3-4: Histogram graph showing MCV distribution in all β Thalassaemia heterozygotes in this study. 47
Figure 3-5: Histogram graph showing HbA₂ distribution in all β Thalassaemia heterozygotes in this study. 48
Figure 3-6: HbF g/dL plotted versus different types of β thalassaemia alleles 49
Figure 3-7: The location of XmnI and (AT)ₓTᵧ in the β globin locus. 50
Figure 3-8: Agarose separation of PCR-RFLP for Xmn1 digest 51
Figure 3-9: HbF g/dL of β thalassaemia heterozygotes stratified by their β thalassaemia mutation and XmnI genotype. 52
Figure 3-10: HbF g/dL of β thalassaemia heterozygotes stratified by their β thalassaemia mutation and (AT)ₓTᵧ genotype. 53
Figure 3-11: The 3 dimorphisms genotyped in the HBS1L-MYB locus 54
Figure 3-12: RT-PCR amplification output for a heterozygous $BCL11A$

Figure 3-13: RT-PCR amplification output for a homozygous wildtype $BCL11A$

Figure 3-14: RT-PCR amplification output for a homozygous mutant $BCL11A$

Figure 3-15: $BCL11A - rs766432$ genotype distribution in $\beta$ thalassaemia heterozygotes

Figure 3-16: $BCL11A - rs11886868$ genotype distribution in $\beta$ thalassaemia heterozygotes

Figure 3-17: $BCL11A - rs4671393$ genotype distribution in $\beta$ thalassaemia heterozygotes

Figure 3-18: $MYB - rs28384513$ genotype distribution in $\beta$ thalassaemia heterozygotes

Figure 3-19: $MYB - rs4895441$ genotype distribution in $\beta$ thalassaemia heterozygotes

Figure 3-20: $MYB - rs9399137$ genotype distribution in $\beta$ thalassaemia heterozygotes

Figure 3-21: Effects of cis-trans interplay with $MYB$ on the HbF of $\beta$ Thalassaemia heterozygotes

Figure 3-22: No effects of cis-trans interplay with $BCL11A$ on the HbF of $\beta$ Thalassaemia heterozygotes

Figure 4-1: No correlation was observed between HbF % and Hb in g/dL for $\beta^+$ IVS-I-6C and $\beta^0$ Codon 39 thalassaemia heterozygotes

Figure 4-2: The position of the C $\rightarrow$ T nucleotide substitution in the IVS-I of the $\beta$ globin gene is shown by an underlined and italicized base.
Table 1-1: A list of non-deletion HPFH causing mutations in $G_\gamma$ and $A_\gamma$ globin

Table 2-1 Primer sequences with Tm and PCR product length being amplified

Table 2-2 Represents the annealing temperatures used for PCR

Table 2-3 Allele specific primers for BCL11A and MYB polymorphisms

Table 2-4 Reaction Components required for each RT-PCR reaction

Table 2-5 Thermal Profile used for allele specific genotyping by RT-PCR

Table 3-1: Representation of the mean and standard error measurements for the four important $\beta$ Thalassaemia indicators.

Table 3-2: $XmnI$ genotypes in $\beta$ thalassaemia heterozygotes

Table 3-3: $(AT)_X T_Y$ genotypes in $\beta$ thalassaemia heterozygotes

Table 3-4: Combined $XmnI$ and $(AT)_X T_Y$ genotypes in $\beta$ thalassaemia heterozygotes

Table 3-5: Genotypes of BCL11A and MYB in $\beta$ thalassaemia Heterozygotes

Table 3-6: The mean HbF g/dL ± S.E. in all $\beta$ thalassaemia heterozygotes versus the three genotypes of each SNP is shown

Table 3.7 Haplotype frequencies for MYB dimorphisms

Table 3.8 Haplotype frequencies for BCL11A dimorphisms
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>μg</td>
<td>microgram</td>
</tr>
<tr>
<td>μl</td>
<td>microlitre</td>
</tr>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>ACH</td>
<td>Active chromatin hub</td>
</tr>
<tr>
<td>BP1</td>
<td>Beta Protein 1</td>
</tr>
<tr>
<td>BFU</td>
<td>Burst-forming unit</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>cM</td>
<td>centiMorgan</td>
</tr>
<tr>
<td>cRNA</td>
<td>Complementary ribonucleic acid</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dpc</td>
<td>Days post-coitum</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double stranded deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EPO</td>
<td>Erythropoietin</td>
</tr>
<tr>
<td>FKLF</td>
<td>Foetal Kruppel Like Factor</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>Hb</td>
<td>Haemoglobin</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone deacetylase</td>
</tr>
<tr>
<td>HEP</td>
<td>Human erythroid progenitor</td>
</tr>
<tr>
<td>HbA</td>
<td>Adult haemoglobin</td>
</tr>
<tr>
<td>HbF</td>
<td>Foetal haemoglobin</td>
</tr>
<tr>
<td>HBA</td>
<td>Alpha globin gene</td>
</tr>
<tr>
<td>HBB</td>
<td>Beta globin gene</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>HBD</td>
<td>Delta globin gene</td>
</tr>
<tr>
<td>HBE</td>
<td>Epsilon globin gene</td>
</tr>
<tr>
<td>HBG1</td>
<td>Ay globin gene</td>
</tr>
<tr>
<td>HBG2</td>
<td>Gy globin gene</td>
</tr>
<tr>
<td>HPFH</td>
<td>Hereditary Persistence of Foetal Haemoglobin</td>
</tr>
<tr>
<td>HPLC</td>
<td>High Performance Liquid Chromatography</td>
</tr>
<tr>
<td>HU</td>
<td>Hydroxyurea</td>
</tr>
<tr>
<td>HS</td>
<td>Hypersensitive site</td>
</tr>
<tr>
<td>IEF</td>
<td>Isoelectric focusing</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukins</td>
</tr>
<tr>
<td>IU</td>
<td>International units</td>
</tr>
<tr>
<td>IVS</td>
<td>Intervening sequence</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kiloDaltons</td>
</tr>
<tr>
<td>KLF1</td>
<td>Erythroid Krüppel Like Factor 1</td>
</tr>
<tr>
<td>LCR</td>
<td>Locus control region</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage disequilibrium</td>
</tr>
<tr>
<td>LOD</td>
<td>Logarithm of odds</td>
</tr>
<tr>
<td>mA</td>
<td>milliamperes</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
<tr>
<td>mM</td>
<td>milliMolar</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NPL</td>
<td>Non parametric logarithm of odds</td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PYR</td>
<td>Polypyrimidine</td>
</tr>
<tr>
<td>RE</td>
<td>Restriction enzyme</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SCD</td>
<td>Sickle cell disease</td>
</tr>
<tr>
<td>SCF</td>
<td>Stem cell factor</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>Tm</td>
<td>Melting temperature</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumour necrosis factor</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>Volts</td>
</tr>
<tr>
<td>w/v</td>
<td>Weight by volume</td>
</tr>
</tbody>
</table>
CHAPTER 1

LITERATURE REVIEW
1.0 Introduction

Reactivating the foetal haemoglobin program of development to raise the level of foetal haemoglobin in sickle cell disease or β-thalassaemia patients would greatly ameliorate the symptoms of these diseases. This can only be achieved by first understanding fully the genetic switch from Foetal to Adult haemoglobin that is commonly referred to as γ to β globin gene switching. Substantial progress has recently been made in this field, however the exact mechanism remains elusive. The combination of clinical research coupled with basic research using state-of-the-art technologies should generate the necessary data and knowledge to complete the remaining pieces of this intricate mechanism. Using such an approach, a unique Maltese family with elevated levels of Hb F was identified through the National Screening Program conducted in Malta. An extensive molecular study revealed a novel mutation in the human EKLF/KLF1 transcription factor that was responsible for an overall dominant Hb F increase in all family members with the mutation, and gene expression array analysis delivered a list of KLF1-target genes that might also be implicated in the globin biosynthesis pathway. The same study showed how BCL11A is developmentally regulated by KLF1 and that diminished levels of BCL11A as a result of KLF1 haploinsufficiency resulted in higher Hb F levels. My work shall study further KLF1 and associated partners and their role in globin gene switching.

1.1 Structure and Function of Haemoglobin

1.1.0 Structure

Haemoglobin is the oxygen transporter of erythrocytes. It is also one of the best-understood allosteric proteins. Human haemoglobins are a duplex of heterodimers composed of two α and two non-α globins each associated with a haeme group. The haeme group is formed by an iron atom bounded by a porphyrin ring (Perutz 1978; Fermi and Perutz et al. 1984)
1.1.1 Developmental Globin Chain Synthesis

There are seven normal haemoglobin types that are developmentally expressed and appear at different stages of the developing human body. These consist of embryonic haemoglobins, Hb Gower 1 (ζγε2), Hb Portland (ζγγ2), and Hb Gower 2 (αεε2), the foetal haemoglobins HbF (α2γγ2) and (α2γε2) and adult haemoglobin made of Hb A (α2β2) with a variant HbA2 (α2δε2) (Bank, 2006). This distribution reflects the patterns of expression of the α-globin gene locus on human chromosome 16 and the β-globin gene locus on human chromosome 11. The pattern of expression of these genes shifts from the more 5' genes on the DNA to more 3' genes during foetal, then neonatal, and then adult developmental stages that is also interestingly encountered with a shift in the main organ site of erythropoiesis whereby embryonic haemoglobins occur in the yolk sac, foetal haemoglobin occurs in the foetal liver and spleen, and adult haemoglobin in the bone marrow (Figure 1.1).

Figure 1.1: Developmental Globin Chain synthesis. The X-axis shows the timeline of the expression of the human globin genes from early stages of foetal development to the changes that occur at birth and in the first year of life. Also shown are the major sites of erythropoiesis and the types of haemoglobin-containing cells during these periods. (Adapted from Weatherall and Clegg, 2001).
However in the case of acute anaemia, stress induces a physiological response that includes the rapid development of new erythrocytes. This process is referred to as stress erythropoiesis, which is distinct from steady state erythropoiesis. Much of what we know about stress erythropoiesis comes from the analysis of murine models. The signals that regulate this process are Hedgehog, bone morphogenetic protein 4 (BMP4), stem cell factor and hypoxia. Recent findings show that stress erythropoiesis utilizes a population of erythroid-restricted self-renewing stress progenitors (Paulson et al., 2011).

The α-globin gene locus, spanning a region of around 30kb on the short arm of chromosome 16, contains the embryonic ζ2 gene, three pseudo genes, ψζ1, ψα2 and ψα1 and the two-α genes – α2 and α1. Their regulation is extensively described in Higgs et al., (2008). The θ1 gene has also been identified at the 3' end (Marks et al., 1986). The β-globin locus is found on the short arm of chromosome 11p15.5 and spans a region of around 90Kb. The cluster 5' to 3' is comprised in this order; the embryonic e-gene, two foetal genes, Gγ and Aγ (that code for the same protein except that at amino acid position 136, Gγ carries glycine and Aγ carries alanine), a pseudo gene ψβ, and the adult δ and β genes. The globin genes are relatively small in size being only around 2 kilobases (kb) long. The somewhat small size made the identification and characterization of mutations leading to thalassaemia and other haemoglobin variants much easier than other conditions. The structure of the β-globin gene is characteristic of all other globin genes and is made up of three coding exons and two intervening sequences together with short untranslated regions preceding the initiator AU and following the translation terminator signal. The 5' terminus of the β-RNA has a typical CAP structure and the 3' end a poly(A) tail (Klug and Cummings, 2000).

### 1.1.2 Regulation

#### 1.1.2.1 Molecular Function

It was thought that only the first 100 nucleotides (nt) 5' to the CAP site of the globins, which include a number of short DNA sequences that constitute cis regulatory motifs are important in their pattern of expression. However cis regulatory
DNA sequences even extend up to -540bp upstream of the β-globin gene as evidenced by Berg et al (1989). These cis-regulatory protein-binding sites interact with specific trans-acting factors that regulate transcription and globin gene expression. The globin gene promoters include three sets of sequences that are highly conserved and are frequent in all globin genes (Nienhuis and Maniatis, 1987; Slightom et al., 1980; Liebhaber et al., 1980; and Spritz et al., 1980). A duplicated CACCC sequence, known to attract and bind Krüpple-like factors, in the β-globin gene is located between -90 to -105 and is also present in the ε and γ genes (Nienhuis and Maniatis, 1987). Similarly the α-gene promoter is very rich in GC sequences and is known to attract Sp1 and Sp1-like proteins. The α-gene promoter (Liebhaber et al., 1980) also comprises a CACCC similar sequence box of the β-globin promoter at position -84 to -89. Two other conserved sequences include TATAA (the TATA box), located at around position -30 and a CCAAT sequence (CAT box), located at around positions -70 to -80. The δ-globin gene promoter contains three imperfect CAT boxes (Spritz et al., 1980) while each γ-gene promoter has a duplicated CAT box (Slightom et al., 1980). The δ- and β- globin gene promoters both have DNA sequences attracting a negative repressor protein beta protein 1 (BP-1) expressed very early in development (Berg et al., 1989), which may explain why they are heavily suppressed in early development. Mutations or nucleotide substitutions in such regions may result in either abolishing of transcription factor binding site or a creation of a binding site which may result in decrease or increase in the production of the respective globin as observed in a group of patients with thalassaemia and HPFH).

1.2 Erythropoiesis

Erythropoiesis is the process in which erythrocytes are produced (Figure 1.2). This generally occurs in two waves during embryogenesis, i.e. primitive and definitive. In humans, primitive erythropoiesis starts in the yolk sac by day 18 of gestation, giving rise to primitive erythroid cells that enter the embryo proper at day 21 and circulate until approximately 12 weeks of gestation. The primitive erythroid cells are nucleated, they contain embryonic haemoglobins and they proliferate and terminally differentiate within the vascular network (Peschle et al., 1985).
Figure 1.2: Developmental progression of a mature erythrocyte. In this process of red blood cell maturation, a cell undergoes a series of differentiations. In human adults, all the above steps except the last occur inside the bone marrow. At the reticulocyte stage, the nucleus is expelled and reticulocytes are released into the bloodstream where ultimately mature erythrocytes are formed through the spleen. (The BFU-E and CFU-E images were taken from http://www.copewithcytokines.de)

Burst forming unit erythroid (BFU-E), is a term used to describe the earliest known erythroid precursor cells that eventually differentiate into erythrocytes. BFU-E can be assayed and identified in a colony formation assay by the specific morphology of the cells and are referred to also as low proliferative potential colony-forming cells. BFU-E produces large colonies of erythroid cells that consist of "bursts" of smaller colonies. These colonies appear after 10 to 15 days in culture. Within the erythropoietic stem cell lineage BFU-E precedes the more mature CFU-E (colony-
forming units erythroid). Maturation of the cells is accompanied by a progressive increase of Epo receptors on BFU-E and the cells become progressively more dependent upon Epo. It has been shown that among various hematopoietic progenitors BFU-E is the most sensitive to treatment with 4-HC (4-hydroperoxycyclophosphamide).

BFU-E cells respond to a large number of cytokines that affect their growth, differentiation and survival. BFU-E entirely depends on the continuous presence of IL3. IL3 appears to be one of the priming activities in vitro and in vivo that renders progenitor cells sensitive for the action of other cytokines (Aglietta et al., 1993). The growth of BFU-E is promoted by EDF (eosinophil differentiation factor), a factor now known to be identical with IL5 (Nakamura K et al., 1993). As early as 4 weeks of gestation, definitive erythroid progenitors, i.e. burst forming units-erythroid (BFU-E), are found in the yolk sac and by 5 weeks in the foetal liver. After the 7th week of gestation there are no haematopoietic progenitors in the yolk sac. The foetal liver takes over as the main definitive haematopoietic tissue from the 7th to the 25th week of gestation. The bone marrow then takes the final lead in haematopoiesis. Definitive erythroid cells differentiate in proximity to macrophages and they extrude their nuclei before entering the bloodstream. Erythropoiesis has been thoroughly studied at the molecular level, and despite many questions remaining unanswered, the essential role of a number of transcription factors, such as MYB (Mucenski et al. 1991), KLF1 (Nuez et al., 1995), Tal1 (Shivdasani et al., 1995), Lmo2 (Warren et al., 1994) and GATA1 (Fujiwara et al., 1996), has been unraveled through genetic studies. Recently it was shown that in the mouse, primitive circulating erythroid cells enucleate in the foetal liver at the end of their maturation process, and continue circulating until the end of their cycle (Isern et al., 2008). Definitive erythropoiesis starts in the foetal liver and later on during gestation it migrates to the spleen and bone marrow. Definitive cells appear in the blood (Wong et al., 1985). They are generated from committed progenitors that go through a stable number of divisions, cell size reduction, protein production, protein degradation, cell cycle block and enucleation. A model linking the production of F erythrocytes to the pattern of erythropoiesis has been anticipated (Stamatoyannopoulos et al., 1987); F erythrocytes arise from erythroid progenitor cells that have undergone a rapid erythroid differentiation pathway. In conditions where erythroid requirement is increased, a greater proportion of erythroid cells
would trail this pathway and an increase in F erythrocytes would ensue. In conditions with chronic shortage of progenitor cells, most erythroid cells would follow this pathway resulting in a persistent increase in Hb F levels. In persistent erythrocyte destruction, the anaemia may be lessened by increase in progenitor compartments, resulting in normal HbF levels. This proposition can explain the increase in Hb F levels in many acquired anaemias and thus far, facts have been consistent with the fundamental principles of this premise (Stamatoyannopoulos, and Nienhuis, 1992).

The progression of haemoglobin gene switching is also brought about by changes in intranuclear transcription factor islands that control gene regulation. In order to achieve this, an active chromatin hub (ACH) forms 5' to the β globin locus whose presence depends on numerous transcription factors, including KLF1. The ACH is a nuclear compartment dedicated to RNA polymerase II transcription, formed by cis-regulatory elements of the β-globin locus (Palstra et al. 2003). In erythroid cells, a substructure of the ACH, consisting of 5'HS-62/-60, 3'HS1 and HS at the 5' side of the LCR, is formed independently of KLF1. Progression of this substructure to a fully functional ACH, including the HS at the 3' side of the LCR and the active β-globin gene, is dependent on the presence of KLF1.

### 1.3 Haemoglobin Variants and Disorders

Haemoglobin disorders are a broad spectrum of diseases, which can be divided in those resulting from an inherited structural alteration in one of the globin chains, like sickle cell anaemia, and those resulting from inherited defects in the rate of synthesis of one or more globin chains, the thalassaemias. The latter results in imbalanced globin chain production, ineffective erythropoiesis, haemolysis and a variable degree of anaemia.

#### 1.3.1 Thalassaemia

There are many different types of thalassaemias depending on the globin chain(s) affected (α, β, δ, and δβ) that lead to defective haemoglobin production. In fact, in these cases the damage to the red cells or their precursors comes from the globin
chains that are formed in relative excess, like in the case of α-thalassaemia (Higgs and Weatherall, 2009) in which a reduced rate of α-chain synthesis results in excess of γ chains in foetal life and γ4 tetramers, or haemoglobin Bart’s. In adult life, deficiency of α-chain synthesis results in excess of β chains and β4 tetramers, or haemoglobin H. The molecular basis of thalassaemias falls into deletional and non-deletional mutations and is extremely diverse. An open access database, the HbVar database (http://globin.cse.psu.edu/hbvar/menu.html), has been created with details of the mutations and deletions worldwide, leading to the deposition of all haemoglobin variants, all types of thalassaemia and other haemoglobinopathies (Giardine et al., 2007). In Malta, molecular studies have identified an unusual distribution of β-thalassaemia mutations. The β+ IVS I-6C mutation alone accounts for 71.4% of β+-thalassaemia chromosomes tested in Maltese homozygotes. The β⁰ Codon 39T, IVS I-110 and IVS II-1A mutations, which are most common in the Mediterranean, account for the rest of β-thalassaemia chromosomes in Maltese patients (Felice et al., 2007). The β+ IVS I-6C was also identified in two Hb Valletta heterozygote and one HbS heterozygotes (Scerri et al., 1993) and data on the two probands and their families showed objectively that the β+ IVS I-6C mutation suppresses β globin output to about 40% of normal activity (down to about 3.0 pg/gene/cell). Thus the β+ IVS I-6C mutation can be considered to result in a relatively mild type of thalassaemia. There is a wide distribution of Hb F level in Maltese β thalassaemia traits, and it is yet to be determined whether this is due to the β thalassaemia mutation itself or in conjunction with other modifier genes and SNPs. Modifier genes include BCL11A and MYB, and more recently mutations in KLF1. Since the identification of a large Maltese family carrying heterozygous mutation in KLF1 (p.K288X), it has yet to be determined whether any of the Maltese β thalassaemia trait also carry the same mutation.

1.4 Control of Hb F

Foetal haemoglobin (Hb F) is the main oxygen carrier utilized by foetuses up until birth and till approximately the age of 7 months. Hb F starts to decline after birth and gradually replaced by the adult haemoglobin (Hb A) allowing it to slowly take over and replace Hb F as the main blood-oxygen transporter in adult human beings. Hb F production, although heavily suppressed, is not completely silent, and in fact its
Chapter 1 Literature Review

synthesis still continues throughout human life. In normal healthy adults, Hb F constitutes an amount lesser than 2%. In certain individuals having what is known as Hereditary Persistence of Foetal Haemoglobin (HPFH), the normal percentage of expected Hb F level is considerably augmented. There are many genetic and possibly other external stimuli that are responsible for HPFH. The genetic factors comprise a number of single nucleotide polymorphisms (SNPs) in modifier genes, as well as mutations. The rs7482144 polymorphism in the Gγ promoter at position -158 C→T is one genetic modifier that acts to augment high levels of Hb F. A study done by Sankaran et al (2010) showed amounts of Hb F around 10% to 40% and revealed the presence of point mutations and deletions in the β globin gene cluster that hindered the globin switching from foetal to adult type. Thus the Hb F level in such individuals was observed to be exceptionally high if compared to the normal expectancy level of Hb F. Sankaran et al. (2010), suggested that common DNA polymorphisms occurring at the location of the β-globin locus could indeed be acting in cis in the regulation of Hb F levels in “normal” individuals. The rs7482144 polymorphism (also known as Xmn1) in the proximal promoter of the Gγ globin was detected to be present in these individuals with HPFH. In this study the T allele at rs7482144 was associated with the increase in Hb F, however some individuals were found to have raised Hb F even when an allele related to the decrease in Hb F (the C allele) was present somewhat similar to the MTHFR conditional mutation (C677T). In most cases, mutations for MTHFR must be present in both copies of a person’s MTHFR genes to have any detrimental effect. However not all people will develop high homocysteine levels (Varga et al., 2005). In another example of conditional mutants, HLA DQ2 DQ8 will confer a strong genetic predisposition to Coeliac Disease and whoever carries these alleles is at a high risk. However, only if gluten is ingested in the diet will one get clinical manifestations. Once gluten is removed from the diet, and even though HLA alleles DQ2 DQ8 are present, they will not manifest in the genetic disorder (Brett et al., 1999).

This indicated that the rs7482144 polymorphism alone was not sufficient to correlate the HPFH to a single genetic factor but implied that this rise in Hb F level was a more complex trait that was being manipulated by various DNA sequence variants of low effect and thus suggested that Hb F was a polygenic trait and not monogenic. This prompted the use of linkage studies for the analysis of genetic markers across the
human genome in families with members having high amounts of Hb F. Sankaran et al. (2010) also showed other key chromosome locations/regions that could have a consistent link with the Hb F regulation. Linkage peaks in chromosomes 6q23, 8q and Xp22.2 were identified. Replication data of linkage signals on the two chromosomes 6q23 and 8q were detected in other families (Thein, Menzel, Lathrop & Garner, 2009). These findings were vital in demonstrating that variants could be present and having a high or moderate impact on the levels of Hb F in individuals and strengthened the hypothesis that genetic factors controlled Hb F trait in trans with the β-globin locus on chromosome 11.

Thein et al. (2007) produced association studies on “normal” Northern Europe twins with a wide range of F cell distribution including high levels. This was done to fine-map the 6q23 Hb F linkage peak to a relatively small intergenic interval of approximately 80-kilo bases that resulted between the two genes; HBS1L and MYB genes. HBS1L gene encodes for a guanosine-5’-triphosphate (GTP)-binding protein while MYB gene encodes for a transcriptional regulation factor of human haematopoiesis. This indicated that the intergenic region between the HBS1L and the MYB genes as the second association with the Hb F locus apart from the rs7482144 polymorphism. The MYB gene is known to be an important factor in erythropoietic regulation and a lowering of γ-globin expression was related to the overexpression of the MYB gene in a strong manner.

The correlation between the two genes HBS1L and MYB was further strengthened in a study conducted by Jiang et al. (2006). In this research a comparison of the messenger RNA (mRNA) of five protein-coding genes in cultured erythroid cells between persons having a raised Hb F level and others having a normal level of Hb F. This study showed a significant relation with the high levels of Hb F and under expressed levels of HBS1L and MYB. This indicated that: 1) either both genes were somehow responsible for the regulation of Hb F or 2) that one of these genes alone was responsible for the Hb F regulation. In the same study, the MYB proto-oncogene and the HBS1L gene were reverse transcribed from total RNA by the use of reverse transcriptase III in vitro and cloned. Expression levels of the γ-globin and the two genes; HBS1L and the MYB, were evaluated by real time polymerase chain reaction (RT-PCR). The results indicated that over expression of the proto-oncogene MYB
provided an inhibitory effect on the γ-globin expression whilst overexpression of the gene \textit{HBS1L} had no relevant effect on the expression of the γ-globin. This indicated a strong possibility that the \textit{MYB} protooncogene is involved with the regulation mechanism of Hb F.

Another important locus that appears to be a quantitative modifier of Hb F is one found on chromosome 2p16, that comprises a gene called \textit{B-Cell Leukaemia/Lymphoma 11A (BCL11A)} (Menzel \textit{et al.}, 2007). The \textit{BCL11A} identified by Menzel \textit{et al.} (2007) and later shown by Sankaran \textit{et al.} (2010) that the gene plays a very important role in Hb F regulation and control. The study by Sankaran \textit{et al.} (2010) identified a strong correlation between levels of Hb F and the zinc finger protein that is encoded by the \textit{BCL11A} gene and that indeed it explained around 15\% of the phenotypic variation of Hb F quantity. It was demonstrated that this gene was responsible for the coding of an important agent that serves as a direct transcriptional regulation factor of the switching between adult haemoglobin (Hb A) and Hb F. This study also managed to point out the relation between reduced levels of \textit{BCL11A} gene expression and the raised level of Hb F. It was later discovered that another master regulator of erythropoiesis, \textit{Krüppel Like Factor (KLF1)}, was in turn regulating the \textit{BCL11A}. This study showed for the first time how mutations in \textit{KLF1} resulted in low levels of this transcription factor, that in turn lowered the Hb F repressor gene, BCL11A and increased significant levels of Hb F (Borg \textit{et al.}, 2010).

In the study conducted by Borg \textit{et al.} (2010) HPFH was detected in a group of ten individuals from the same family comprising a total of 27 members in the Maltese island. In this research, linkage analysis and genome-wide SNP scan were used as techniques that ultimately identified a candidate region on chromosome 19p13.12-13. Borg \textit{et al.} (2010) discovered that the \textit{KLF1} gene, had a nonsense mutation (p.K288X) that was heterozygous for the ten members out of the twenty-seven-member family that had HPFH. This mutation was responsible for ablating the DNA-binding domain of the \textit{KLF1} gene, an important erythroid regulatory transcription factor. In these individuals with HPFH, expression-profiling techniques where performed on the primary erythroid progenitors, showing that in these individuals there was a significant down regulation of the \textit{KLF1} target genes. Borg \textit{et al.} (2010) observed that there is indeed a stable link between HPFH condition and the
haploinsufficiency of the *KLF1* gene. Quantitative S1 nuclease protective assays performed in this study demonstrated the consistent rise in *HBG1/HBG2* expression (that encode the synthesis of \(\gamma\)-globin) occurred with the knockdown of *KLF1*. This important finding showed that the *KLF1* has an indirectly proportional correlation with levels of Hb F. The *BCL11A* gene levels also decreased (at both the mRNA and protein level). The above-mentioned studies explain how and why three genes with occurring Polymorphisms or mutations exert an important genetic influence on Hb F levels.

### 1.5 Haemoglobin F-Malta-I

An umbilical cord bloods survey in Malta (Cauchi *et al.*, 1969) revealed a new foetal haemoglobin variant, designated as Hb F Malta I \([\alpha2\gamma2 117(G19) \text{His}\rightarrow\text{Arg}]\). Hb F Malta I is a clinically silent \(\gamma\) globin variant. It has been considered a valuable marker of the levels of *in vivo* expression of the human \(G\gamma\) globin gene and the parameters that control differential gene expression. This variant was found at an incidence of about 1.6% at that time in the Maltese population and was the very first population in which a relatively high incidence of a foetal haemoglobin variant was observed. The abnormal foetal haemoglobin declined at the same rate as Hb F and was not present in the infants' parents hence suggesting that it was a foetal haemoglobin (\(\gamma\)-chain) variant. Mazza *et al* (1980) showed for the first time, examples of this \(\gamma\) chain mutant in newborns of other than Maltese origin. Nevertheless, on the basis of the geographical location of their regions of origin, they cannot exclude a common mutation for these and the previously reported cases. In the study by Mazza *et al* (1980), the % of Hb F Malta I at birth varied from 24% to 31% of the total Hb F; this amount is greater than that reported for Maltese newborns (Cauchi *et al.*, 1969) were variant Hb F was found ranging from 20.7% to 27.4% in 12 subjects and by another study (Schroeder *et al.*, 1973) who reported quantities ranging from 17% to 27% in 35 newborns. A novel and stable abnormal haemoglobin (Hb Valletta) was discovered in 1991 (Kutlar *et al.*, 1991) This Haemoglobin is characterized by a Thr→Pro substitution at position 87 of the \(\beta\) chain. This mutation was linked to that of the \(\gamma\) chain variant Hb F-Malta-I with a His→Arg mutation at position 117 of the \(G\) \(\gamma\) chain. Both variants were detected in the blood samples of 34 Maltese and two
Italian newborn babies with isoelectrofocusing and reverse phased high performance liquid chromatography. Similar analyses of cord blood from 388 additional Maltese newborns failed to identify either one of these two variants. Additional analyses of 353 Maltese adults (including 39 β - thalassaemia heterozygotes) resulted in the detection of two adult Hb Valletta heterozygotes. Dot-blot hybridization analyses of amplified DNA with a probe specific for the G γ F Malta I variant showed that both also carried that mutation. These results show close linkage of the mutant forms of the G γ and β-globin genes, 27-28 kb apart, and a failure to identify chromosomes with either the Hb F-Malta-I mutation alone or with the Hb Valletta mutation alone, indicating a low recombination frequency. Pulis et al (2007) concluded that the β globin gene cluster haplotype in Hb FMalta-I homozygotes occurs on a haplotype Va background (+ + - - - - + + - for the 5'ε Hinc II, 5'γ Apal, 5' γ XmnI, γ γ and αγ HindIII, ψβ and 3'ψβ HincII, 5'β Hinfl, β AvaII and 3'β BamHI sites respectively) The common Mediterranean haplotypes Va, IIIb, I and II accounted for 66.2% of the wildtype alleles tested in Hb F-Malta-I heterozygotes. Another six haplotypes were also commonly found in Hb F-Malta-I heterozygotes and together accounted for another 20% of wild-type alleles tested. About 13% of wild-type alleles tested in Hb F-Malta-I heterozygotes had haplotypes that appeared only once in the cohort tested. Another novel finding since 2007 was the co-inheritance of Hb F Malta I and Hb F Sardegna compound heterozygotes which make them carry 2 quantifiable β globin chains and 4 β- like globin chains [γγ, γFMaltaI, αΓI, αΓT, βΛ, and βValletta]. Hb F Sardegna (αγ75 (E19) Ile→Thr) was readily separated with a modified gradient from all other chains and appeared just before the abnormal Gγ chain of Hb F Malta I. This combination of haemoglobin variants offers the opportunity to study independent markers of all 6 β-like globin genes. Hb F Malta I was also found to be co-inherited with a rare form of HPFH (Malta Type), which is the main focus of this study. It has also been detected in a neonate of this family that co-inherited the Hb F Malta I with Hb F Sardegna.

1.6 Hereditary Persistence of Foetal Haemoglobin (HPFH)

A heterogeneous group of conditions is the Hereditary Persistence of Foetal Haemoglobin (HPFH), which is characterized by persistent foetal haemoglobin...
production (Conley 1980). In normal individuals, HbF levels decrease from less than 5% at six months to less than 2% throughout the whole of adult life (Craig et al., 1997). In adults, Hb F is limited to a subset of erythrocytes termed F-erythrocytes. F-erythrocytes are not analogous to foetal erythrocyte cells in the amount of Hb F per cell or in surface antigens and cell enzyme levels. The amount of Hb F per F cell varies, but Hb F can account for up to 25% of the total haemoglobin in the cell, averaging around 7pg/F erythrocyte (Dover et al., 1992). The Hb F in adult heterozygotes for HPFH may account for 20-30% of the total haemoglobin level. Methods of classification of HPFH are based on the type of globin chain produced (Gγ, Aγ or both), the type of cellular distribution of Hb F (pancellular or heterocellular) and the type of molecular defect present (deletional or non-deletional) (Bollekens and Forget, 1991). Early clues about γ to β globin gene switching were obtained by comparing phenotypes of Hb Kenya (HPFH) and Hb Lepore (δβ Thalassaemia). Figure 1.3 shows the fused genes in more detail.

Figure 1.3: The Aγ – δ globin intergenic region appears to harbor important silencer elements as exemplified by (a) Lepore (δβ) and anti-Lepore (βδ) fusion gene products and (b) the Hb Kenya and anti-Kenya fusion gene products. Adapted from Borg et al (2009)
1.6.1 Deletional Hereditary Persistence of Foetal Haemoglobin

All seven known types of deletional HPFH result from partial deletions involving the δ and β globin genes. Deletional HPFHs are characterized by pancellular Hb F production with Hb F levels of 20-30% in heterozygotes (Bollekens and Forget, 1991). The 5' breakpoints of the deletions lie between or within the Aγ and β globin genes while the 3' ends are variable. Possible explanations for the increase of Hb F in these conditions include: (1) elimination of repressor elements between the Aγ and δ globin genes (Huisman et al., 1975), (2) repositioning of enhancer elements downstream of the 3' breakpoint with respect to the γ globin genes, (3) removal of competition between the δ and β versus γ promoter elements for the Locus Control Region (LCR) and (4) large changes in chromatin conformation resulting from large deletions (Bollekens and Forget, 1991). Types I and II of deletional HPFH are found in Blacks and are characterized by large deletions of 105-106 Kb, type III has a 48-49 Kb long deletion and is found in Indian families, types IV and V are found in a few Italian families, type VI is found in the Vietnamese and type VII in Kenya, Africa. All have intact γ globin genes except type VII that has an intact Gγ globin gene and a neighbouring Aγβ hybrid gene (Hb Kenya) Types I and II are found in about 0.1% of Blacks in South-eastern United States while the rest are rarely encountered. Figure 1.4 depicts some of the deletions in the β globin locus (Huisman et al., 1997).

Figure 1.4: Deletions of the β-globin gene cluster associated with fusion proteins and HPFH. The circle 3' to the β-globin gene indicates the 3' β-globin gene enhancer. The filled vertical boxes at the 3' breakpoints of the HPFH-1 and HPFH-6 deletions indicate the locations of DNA sequences with homology to olfactory receptor genes. (Adapted from Huisman et al., 1997)

Two patients that have a homozygous DNA deletion upstream of the δ globin gene, known as the Corfu deletion, have 88% and 90% Hb F, only mild anaemia, and no
transfusion requirement present strong evidence in humans that intergenic \( \gamma-\delta \) sequences alone can control the activity of Hb F in adulttype cells, and possibly play a role in normal human haemoglobin switching as well (Orkin, 1990). The 7.2-kb Corfu deletion extends from the \( \gamma-\delta \) region upstream of the \( \delta \) gene to involve the 5' end of structural \( \delta \)-globin gene. It is probable that the \( \gamma-\delta \) intergenic region deleted in Corfu patients includes sequences that equally repress and can potentially activate \( \gamma \)-globin gene transcription. The occurrence of both positive-acting and negative-acting sequences and their interactions in chromatin may clarify the differences in \( \gamma \)-globin gene expression with different extents of deletion of this region, such as the case in different patients with \( \delta-\beta \) thalassaemia and in transgenic animals (Orkin, 1995 and Orkin, 1996). The previously described and characterized unique RNA transcripts from the intergenic \( \gamma-\delta \) region sequences hypothesized to play a role in regulating normal human \( \gamma \)-globin gene expression (Pevny et al., 1991) are disrupted in patients with the Corfu mutation (Orkin, 1990). The molecular mechanisms by which loss of these sequences might affect the output of the \( \gamma \)-globin genes are unknown.

1.6.2 Non-Deletional Hereditary Persistence of Foetal Haemoglobin

This class of HPFH occurs without deletions in the \( \beta \) globin locus. They are in fact characterized by intact \( \delta \) and \( \beta \) globin genes and result due to point mutations in the G and A \( \gamma \) globin gene promoters. All 16 mutations have been mapped within the -211 to -100 DNA regions of the \( \text{G\( _{\gamma} \)} \) and \( \text{A\( _{\gamma} \)} \) promoters (Table 1.1). \textit{In vitro} studies have confirmed that these changes are in fact responsible for the increased expression of the \( \gamma \) globin genes (Collins \textit{et al.}, 1984). Several protein-binding sequences have been identified in the promoter regions of the \( \gamma \) globin genes (figure 1.5).
Cis- and trans- acting elements of the human γ-globin gene promoters and point mutations associated with nondeletion forms of HPFH

Figure 1.5: DNA map of point mutations causing non-deletional HPFH (Adapted from Weatherall and Clegg, 2001)

Table 1.1: A list of non-deletion HPFH causing mutations in Gγ and Aγ globin

<table>
<thead>
<tr>
<th>Name</th>
<th>Mutation</th>
<th>Mutation, HGVS nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>-202(C&gt;G) Gγ;nd-HPFH</td>
<td>Gγ mutt at 202 C&gt;G</td>
<td>HGβ2:c.-255C&gt;G</td>
</tr>
<tr>
<td>-175(T&gt;C) Gγ;nd-HPFH</td>
<td>Gγ mutt at 175 T&gt;C</td>
<td>HGβ2:c.-228T&gt;C</td>
</tr>
<tr>
<td>-114(C&gt;G) Gγ;Australian;nd-HPFH</td>
<td>Gγ mutt at 114 C&gt;G</td>
<td>HGβ2:c.-167C&gt;G</td>
</tr>
<tr>
<td>-114(T&gt;C) Gγ;Japanese;nd-HPFH</td>
<td>Gγ mutt at 114 T&gt;C</td>
<td>HGβ2:c.-167C&gt;T</td>
</tr>
<tr>
<td>-114(C&gt;A) Gγ;Algerian;nd-HPFH</td>
<td>Gγ mutt at 114 C&gt;A</td>
<td>HGβ2:c.-167C&gt;A</td>
</tr>
<tr>
<td>-110(C&gt;T) Gγ;Ezerbaijani;nd-HPFH</td>
<td>Gγ mutt at 110 C&gt;T</td>
<td>HGβ2:c.-163A&gt;C</td>
</tr>
<tr>
<td>Venezuelan;nd-HPFH</td>
<td>Gγ mutt at 211 C&gt;T</td>
<td>HG1:c.-264C&gt;T</td>
</tr>
<tr>
<td>-202(C&gt;T) Aγ;nd-HPFH</td>
<td>Aγ mutt at 202 C&gt;T</td>
<td>HG1:c.-255C&gt;T</td>
</tr>
<tr>
<td>-198(T&gt;C) Aγ;British;nd-HPFH</td>
<td>Aγ mutt at 198 T&gt;C</td>
<td>HG1:c.-251T&gt;C</td>
</tr>
<tr>
<td>-196(T&gt;C) Aγ;Italian;nd-HPFH</td>
<td>Aγ mutt at 196 T&gt;C</td>
<td>HG1:c.-249C&gt;T</td>
</tr>
<tr>
<td>-195(C&gt;G) Aγ;Brazilian;nd-HPFH</td>
<td>Aγ mutt at 195 C&gt;G</td>
<td>HG1:c.-248C&gt;G</td>
</tr>
<tr>
<td>-175(T&gt;C) Aγ;Black;nd-HPFH</td>
<td>Aγ mutt at 175 T&gt;C</td>
<td>HG1:c.-228T&gt;C</td>
</tr>
<tr>
<td>-117(G&gt;A) Aγ;Black-Greek;nd-HPFH</td>
<td>Aγ mutt at 117 G&gt;A</td>
<td>HG1:c.-170G&gt;A</td>
</tr>
<tr>
<td>-117(G&gt;A) Aγ;Greek-Italian;nd-HPFH</td>
<td>Aγ mutt at 117 G&gt;A</td>
<td>HG1:c.-170G&gt;A</td>
</tr>
<tr>
<td>-114(T&gt;C) Aγ;Georgia;nd-HPFH</td>
<td>Aγ mutt at 114 T&gt;C</td>
<td>HG1:c.-167C&gt;T</td>
</tr>
<tr>
<td>-200(C&gt;T) Aγ;Tunisian;nd-HPFH</td>
<td>Aγ mutt at 200 C&gt;T</td>
<td>HG1:c.-160C&gt;T</td>
</tr>
</tbody>
</table>

Percentage Hb F levels in non-deletional HPFH heterozygotes can reach up to 40% (Huisman and Carver, 1998). In Swiss-type HPFH, a type of nondeletional HPFH,
there is only a minor heterocellular increase in Hb F level (Hb F level of about 1 – 5%) with an increase in number of F-erythrocytes. To date there has been no single mutation or abnormality, which describes this persistent increase in Hb F and F cell levels. Though factors such as the β globin locus haplotype, particularly the XmnI site 5’ to the Gγ globin gene -158 C→T mutation, an A→G mutation at position -161 of the Gγ promoter, and globin gene rearrangements (such as Gγ- Gγ, Kγ- Kγ or multiple γ globin genes) have all shown to effect Hb F levels and Gγ:Aγ ratios in normal individuals and/or individuals with anaemic stress in various studies (Thein et al., 1994; Craig et al., 1996; Aksoy et al., 1985; Leonova et al., 1996, Huisman et al., 1991 and Felice et al., 2007)

There are also cases of HPFH characterized by low levels of heterocellular foetal haemoglobin. A small proportion of individuals with this type of HPFH were found to have increased amounts of foetal haemoglobin and F cells. Although in the beginning it was called Swiss-type HPFH, later studies found it in every racial group and some evidence suggested X-linked genetic determination of the number of F cells, located at Xp22.2 (Hebbel 1992; Briehl and Nikolopoulou 1993), even though this was not the case for all the low foetal haemoglobin HPFH forms (Thevenin et al., 1997; Parise and Telen 2003) Strong evidence of genetic links have been reported recently in support to the notion that specific DNA polymorphism (SNPs) are related to the observed variation in foetal haemoglobin levels among individuals (Menzel et al., 2007; Thein et al., 2007; Lettre et al., 2008; and Uda et al., 2008). The genomic sequences of these SNPs were mapped to the gene BCL11A on chromosome 2, the intergenic region of HBSIL-MYB on chromosome 6 and upstream of the Gγ-globin in the β-globin locus on chromosome 11. MYB has been studied widely for its role in haematopoiesis, HBSIL is a poorly characterized gene with unknown biological functions. BCL11A is expressed in erythrocyte precursors and implicated in lymphoid malignancies (Satterwhite et al., 2001; and Liu et al., 2003) was recently shown to be a regulator of foetal haemoglobin in loss of function studies in human (Sankaran et al., 2008) and mouse (Sankaran et al., 2009) These results further support the importance of identifying mutated forms of transcription factors present in groups of individuals with specific phenotypes and haemoglobin indices in their erythrocytes (Singleton et al., 2008)
Borg et al. (2010) performed a genome-wide SNP scan followed by linkage analysis on a large family from Malta. Ten of the 27 family members had HPFH varying from 3.3% to 19.5%. He identified a candidate region on chromosome 19p13.12-13. DNA sequencing revealed a nonsense mutation in the *Erythroid Krüppel-Like Factor (KLF1)* gene that involved an A-to-T transversion resulting in a lys288-to-ter (K288X) premature stop codon. This was the first study that linked mutations in *KLF1* to HPFH and the regulation of *BCL11A* by *KLF1*. Gene expression profiles conducted on cultured primary erythroid cells from these individuals showed that mutation carriers had decreased expression of the foetal haemoglobin repressor *BCL11A* and an upregulation of the foetal haemoglobin genes *HBG1* and *HBG2*. Knockdown of *KLF1* in cultured erythroid progenitor cells acting as controls caused similar changes in gene expression. On restoration of *KLF1* activity in human erythroid progenitor cells (HEPs) from Maltese family members with HPFH the *BCL11A* levels were increased. This shows that *KLF1* has a dual role in the regulation of foetal-to-adult globin gene switching. It acts directly on the *HBB* locus as a preferential activator of the *HBB* gene as was reported in 1996 by Wijgerde et al. and secondly it acts indirectly by activating the expression of *BCL11A* which in turns represses the *HBG1/HBG2* genes. The global role exerted by KLF1 can be seen in figure 1.6.
Figure 1.6: Phenotypes caused by KLF1 mutations. The inner ring displays KLF1 mutations and potential modifiers. The number of different mutants reported is shown. The middle ring displays critical KLF1 target genes/loci whose expression is affected by the KLF1 mutation(s). The outer ring displays phenotypes. Clinical conditions are in the boxes; the colors refer to the publications shown below. HBA = \( \alpha \)-globin locus; HBB = \( \beta \)-globin locus; EMS = erythrocyte membrane skeleton. (Figure adapted from Borg et al 2011).
1.7 Erythroid Krüppel-Like Factor

A similar situation is encountered with the KLF1 transcription factor (Fig 1.7). It binds CACCC-box sequences (CCACACCCT) (Donze et al., 1995) frequently found in erythroid-specific gene promoters. Two such sites are present in the human and mouse β-globin promoters. KLF1’s promoter has a GATA1 functional binding site (Crossley et al., 1994) The fact that KLF1 is active in primitive and definitive hematopoiesis and not required for yolk sac erythropoiesis and erythroid commitment (Nuez et al., 1995; and Perkins et al., 1995), suggested that it is important for the transition from foetal to adult globin expression in humans. Additionally, single base substitutions in the KLF1 binding sites in the β-globin promoter cause β-thalassaemia (Orkin et al., 1982). Studies on transgenic mice bearing the human β-globin locus and inactivated KLF1 explained the induction of γ-globin expression, in the absence of transcription of β-globin, as a consequence of productive LCR promoter interactions of the γ-globins (Wijgerde et al., 1996). Schoenfelder et al., (2010) found that mouse Hbb and Hba associate with hundreds of active genes from nearly all chromosomes in nuclear foci that they called ‘transcription factories’. The 2-globin genes preferentially associated with a specific and partially overlapping subset of active genes. Schoenfelder et al., (2010) also noted that expression of the Hbb locus is strongly dependent upon KLF1, while expression of the Hba locus is only partially dependent on KLF1. Immunofluorescence examination of mouse erythroid cells displayed that most KLF1 concentrated to the cytoplasm and that nuclear KLF1 was present in isolated sites that overlapped with RNAII foci. Erythroid cells from KLF1 null mice specifically disrupted the association of KLF1-regulated genes within the Hbb-associated network. KLF1 knockout more insidiously disrupted interactions within the specific Hba network. Schoenfelder et al., (2010) revealed that KLF1-regulated genes share KLF1-containing transcription factories and that KLF1 is required for the clustering of these coregulated genes. It was suggested that transcriptional regulation involves a complex 3-dimensional network rather than factors acting on single genes in isolation.
Figure 1.7: Schematic diagram of the KLF1 molecule. The three C-terminal C2H2 zinc fingers are shown, each chelating a single zinc ion. The fingers are linked together by "TGERP"-like motif, which assists in binding to target DNA. The activation/repression domain is found at the N-terminus of the molecule (Adapted from Pearson et al., 2008).

1.8 B-cell CLL/lymphoma 11A (BCL11A)

The aforementioned transcriptional regulators have been shown to bind globin proximal promoters and so their effect is attributable directly to transcriptional levels of the globin genes. Recently, a gene that has not been studied for its role in the red blood cell lineage, BCL11A, was identified as a stage-specific regulator of Hb F expression (Sankaran et al., 2008) in human, and embryonic genes (Sankaran et al., 2009) εγ and βh1 in the mouse. Expression profiling of primary human erythroid cells after knockdown and subsequent differentiation showed no differences in expression of well-characterized transcriptional regulators of globin genes, GATA1, FOG-1, NF-E2 and KLF1, suggesting that the effect of BCL11A on the γ-globin gene regulation is unlikely to be mediated by these transcriptional regulators. Still, in immunoprecipitation assays (IP), GATA1, FOG-1 and components of the NuRD complex were confirmed to interact with and explain partly the repressor activity of BCL11A in erythroid cells. Surprisingly, BCL11A was not found in the proximal promoter of γ-globin, neither showed robust binding to other regions of the β-globin cluster. However, it was found to bind to three regions: the HS3 of the LCR, a region of high Hb F-associated Corfu deletion upstream of the δ-globin gene and another region downstream the Ay-globin gene that is commonly deleted in certain forms of HPFH (Bank 2006) in primary human erythroid progenitors. The current experimental approaches to decipher the molecular mechanism of the silencing of the γ-globin genes focus on transcription factors that are found mutated in individuals with higher than normal levels of HbF, as it appears to be the case for the BCL11A. Still, the
promoter of the γ-globin genes is of major importance for resolving such a complex gene regulation puzzle, even though so far single protein loss of function experiments have not assigned a critical role for any of the known factors found to be present in their promoters. The basal transcription machinery and the numerous general transcription factors are also thought to play a significant role in promoter-specific gene activation and should be investigated regarding their role in globin switching.

1.9 Therapeutic Approaches

Currently, blood transfusions and iron chelation (deferoxamine) therapy are the main therapeutic procedures for thalassaemias. Bone marrow transplantation is also a solution but the outcome depends on the age and clinical condition of the individual and there are also potential post-transplant problems such as infections or graft-versus-host disease (GVHD). There has been a considerable number of thalassaemic and sickle cell patients (Lucarelli et al., 1995; and Di Bartolomeo P 1997) who underwent bone marrow transplantation and overcame the disease but one out of three developed acute or chronic GVHD ranging first from mild then to severe grades (Gaziev et al. 1997; and Bernaudi et al., 2007). The observation that patients recovering from cytotoxic drug therapy or other periods of erythroid expansion may reactivate foetal haemoglobin synthesis has led to the use of many agents that increase the production of Hb F in order to decrease the probability of in vivo sickling and reduce ineffective erythropoiesis. Such agents include erythropoietin, cytotoxic drugs (5-Azacytidine, Hydroxyurea) and butyrate analogs. All of them have been used in clinical trials (Olivieri and Weatherall 1998; Swank and Stamatoyannopoulos 1998; Weatherall 2003) as potential inducers of Hb F as well as in combinations in order to test for optimal doses and additive or synergistic effects of the agents. Even though there has been a considerable advance in understanding how these agents work biochemically and new treatments have been used with some success in patients, overall progress has been slow due to the fact that the wide range of mutations or deletions of the β-globin gene cluster resulting in thalassaemia are not equally susceptible to this type of approach. Hydroxyurea (HU) has been used most
successfully in patients with sickle cell anaemia since it results in amelioration of vaso occlusive episodes.

Administration of HU increases Hb F (figure 1.8), but there are probably other reasons for the success of the treatment (Charache, Terrin et al. 1995), possibly the reduction of the white blood cell count, changes in red cell rheology and indirect vascular effects.

![Diagram](image)

**Figure 1.8: A diagram of the postulated effects of HU in inhibiting Hb S.** *Hb S polymerization is inhibited by augmenting Hb F levels (shown as 25%) in each sickle RBC and thus decreasing the degree of microvascular obstruction at any oxygen level. (Adapted from ASH 50th Anniversary review by Schechter, 2008).*

Many pharmacological agents apart from HU and azacytidine have been tested for their capability to reactivate foetal haemoglobin (HbF) in reporter assays in human cell lines and human primary progenitor cells (Haley et al., 2003) but the maximal γ-globin induction was always compromised by cytotoxicity. Histone deacetylase (HDAC) inhibitors have been of particular interest in respect to γ-globin induction and especially a class of short-chain fatty acids (SCFAs) derivatives that have been
implicated in the displacement of repressor complexes from the promoter of \(\gamma\)-globin and induce transcriptional activation of the gene (Mankidy et al., 2006). These are now entering clinical trials. Still, the molecular mechanisms underlying the \(\gamma\)-globin transcriptional activation and silencing have not yet been uncovered.

Bone marrow transplantation appears to be a promising therapeutic approach, especially through the major advances of ES-cell-based therapy and the establishment of genetically identical ES-like cells by somatic nuclear transfer (SCNT) using donor cells from the patient (Jaenisch 2004) Human and mouse fibroblasts were reprogrammed \textit{in vitro} into pluripotent stem cell-like cells (iPS) through retroviral transduction of combinations of transcription factors (\textit{OCT4, KLF4, SOX2, c-MYC}) (Yu et al., 2007; and Hanna et al. 2009) Using the sickle cell anaemia mouse model, iPS cells were corrected for the \(\beta\)s allele by homologous recombination and then transplanted to irradiated sickle mice. This resulted in the correction of the disease phenotype (Hanna et al., 2007) The drawback of applying ES or iPS cell based therapy to humans is mainly the possibility of developing malignancies that can not be excluded unless safer methods of gene delivery, retroviral vector-free, assure the efficient expression of the reprogramming factors in a controllable system (Hanna et al., 2008), as well as developing alternatives for the oncogenes currently required for reprogramming.

1.10 Specific Aim

The specific aims of this project were to 1. Determine and quantify the foetal haemoglobin (Hb F) levels from a collection of discordant \(\beta\) thalassaemia heterozygotes in order to exclude effects of sever anaemia as in the homozygous.

2. To correlate the effects of important Hb F modifier genes – BCL11A and MYB on Hb F levels.

BCL11A- and MYB were genotyped for three SNPs each using Real Time PCR. The genotypes were correlated with the \textit{in vivo} Hb F levels of the \(\beta\) thalassaemia heterozygotes to establish whether they have any influence on haemoglobin expression.
CHAPTER 2
MATERIALS & METHODS
2.1 Sample Collection and Preparation

Venous blood samples were collected from patients with thalassaemia trait and recruited into this study for analysis and eventual storage at the Maltese DNA Bio Bank. The blood was used for iso electric focusing (IEF), haemoglobin phenotyping and DNA analysis. The blood was collected in a tube containing K$_2$-EDTA as an anticoagulant, and stored at a temperature of -20°C until analysed. The research was conducted according to the guidelines issued by the Ministry of Health and approved by the local Research Ethics Committee. A letter of consent was handed out to all participants.

2.2 Quantification of Haemoglobin and Globin Chains by HPLC

The globin chains were carried out on patients with thalassaemia trait carrying different levels of foetal haemoglobin. The globin chains were separated and quantified by reverse phase high performance liquid chromatography (HPLC). The VARIANT Haemoglobin Testing System (Bio-Rad Laboratories, USA) was used. The Globin Chains program was modified as follows; 75% buffer B [CH$_3$CN/CH$_3$OH:NaCl 0.155M pH 2.7 25:40:35] in 40% buffer B in 90 minutes, flow rate of 0.7 mL min$^{-1}$ in order to separate $\gamma$ from the normal $\alpha$ globin chain effectively.

Following injection into the system, the globin chains were dissociated and separated on a C8 column on the basis of their hydrophobic differences during a 90 minute run. The results were given in the form of a chromatogram where the relative quantities of globin chains are represented by the peak area. Different globin chain ratios were calculated using these peak areas (figure 2.1).
2.3 DNA extraction

Genomic DNA was extracted and purified from peripheral blood leucocytes by using a kit supplied by GE Healthcare – Illustra Nucleon Genomic DNA Extraction Kits.

2.3.1 DNA extraction from whole blood using the BACC1 Nucleon Kit

Whole blood that was present in EDTA tubes and stored at $-20^\circ$C was allowed to thaw at room temperature for about one hour on a rotator machine to mix evenly. 1mL of blood were transferred into a DNA extraction tube to which 4mL of 1x erythrocyte lysing buffer (Reagent A – 10mM Tris-HCl, 320 M sucrose, 5mM MgCl$_2$, 1%(v/v) Triton X-100, pH 8.0) were added. The mixture was rotary mixed for 5 minutes at room temperature then centrifuged at 1600xg for 5 minutes. The supernatant was

Figure 2-1: Separation of globin chains by reverse phase HPLC on a BioRad VARIANT testing system.
discarded without disturbing the pellet. To the pellet 350μl of reagent B were added. The mixture was vortexed briefly and transferred to a 1.5ml microcentrifuge tube. To this tube 100μl of sodium perchlorate were added and mixed several times by inverting. 600μl of chloroform and 150μl of Nuclear resin were then added respectively to the tube. The final mixture was centrifuged at 350g for 1 minute.

The upper aqueous phase (containing DNA) was aspirated without disturbing the interphase using a sterile transfer pipette, and transferred to a clean 15mL DNA extraction tube that containing 75% of ethanol. This tube was centrifuged at 11,000 × g for 4 minutes and the resulting supernatant decanted off. One mL of 70% ethanol was added to each tube followed by centrifugation as described above. This step was repeated one to two times in order to purify the DNA.

The DNA pellet was dried using a vacuum centrifuge for around 20 minutes until the pellet was completely dry. The pellet was redissolved in 100 μL TE buffer (1 M Tris-Cl, 0.5 M EDTA; pH 8.0) and left overnight on a rotator. Each DNA sample was labeled with the assigned code and stored in a refrigerator at a temperature of 4°C until analysed.

2.4 Checking Genomic DNA Integrity

The integrity of the extracted genomic DNA was tested using agarose gel electrophoresis. Agarose is a linear polymer composed of alternating residues of D- and L- galactose joined by α (1→3) and β (1→4) glycosidic linkages that forms a gel after heating it in a buffer solution (Sambrook and Russell 2001). The rate of migration of DNA in an agarose gel is in part determined by the molecular size of DNA, agarose concentration and the applied voltage. 1% agarose gels were used to check genomic DNA integrity, made by dissolving 1g of agarose in 100ml of 1x TAE buffer (40mM Tris, 20mM acetate, 2mM EDTA; pH 8.1). The solution was heated in a microwave for approximately 5 minutes, and the dissolved agarose was then left to
cool to a temperature of around 55°C. Ethidium bromide was then added from a stock solution of 10mg/mL in water to a final concentration of 0.5μg/mL. Ethidium bromide is a fluorescent dye that binds to the bases of the DNA molecule and emits an orange fluorescence on excitation by U.V. light. Five micro litres from each DNA sample obtained in extraction were mixed with 2μl of 6x loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll in water) and pipetted into each well of the gel, using a micropipette. The loading buffer contains a dye that makes the sample easy to see. It also contains glycerol, which increases the density of DNA, forcing it to sink into the bottom of the well. A DNA marker containing DNA fragments of known sizes was also loaded onto the gel. The size marker used was generated from two different plasmids by EcoRI-restriction (New England Biolabs), and range from 100 to 1000bp.

The electrophoresis tank was closed with its respective lid and the electrical leads were attached in such a way so that the migration of the DNA was towards the anode. Electrophoresis was carried out at a voltage of 150V and a current of 400mA for 20 minutes. Once a desirable migration was observed, the gel was removed from the electrophoresis tank, and viewed on a transilluminator, and photographed for records. The intensity of the bands gives an approximate indication of DNA concentration and quality. Genomic DNA should be much larger than 100kb in size and should migrate more slowly than linear dimeric molecules of digested plasmid.

2.5 UV Spectrophotometry

UV spectrophotometry was used to determine the approximate concentration and purity of each extracted DNA sample. A NanoDropTM 2000C Spectrophotometer (Thermo Fisher Scientific) was used and the dsDNA program selected. This gives the ratio of the absorbance at 260nm and at 280nm (A260:280) and the concentration of DNA in ng/μL. Correct zero adjustment was obtained before by blanking the instrument with 2μl of 1x TE buffer. 2μL of the genomic DNA sample were then used to determine its concentration and purity. DNA samples with a ratio of 1.7 were
purified using a YM-30 filter (MicroCon® Millipore USA) since a reduced ratio indicates protein contamination in the DNA sample.

2.6 PCR Primer Design

A PCR primer is a short oligonucleotide sequence synthesized artificially to enable the selective and repeated amplification of the target region of interest during a polymerase chain reaction. There are several factors that need to be taken into consideration when designing PCR primers. The primers were chosen to be between 18-25bp long, to have a random distribution of bases and avoid nucleotide sequence repeats. Primer pairs were chosen to have similar melting temperatures (Tm) to ensure simultaneous annealing of primers. The approximate annealing temperature is calculated from \( T_m = 2 \sum (A+T) + 4 \sum (G+C) \) (Suggs 1981). The distance between the primers was chosen to be between 200 and 500bp and care was taken to ensure that restriction enzyme digestion of the PCR product resulted in fragments of unequal size, to make separation by gel electrophoresis easier. All primers used in a PCR should have similar melting temperatures and GC content. Typically primers with melting temperatures in the range of 60-62°C are chosen. The original primers were aliquoted and diluted to form working solutions at concentration of 50µM (unless specified otherwise) and stored at -20°C. The sequences of the PCR primers used in this study, along with the respective annealing temperature and amplicon size are listed in table 2.1.
### Chapter 2

**Materials and Methods**

#### 2.7 Polymerase Chain Reaction

The polymerase chain reaction (PCR) is an effective technique by which a strand of DNA or cDNA template of interest is amplified many millions of times in relatively short period of time. The number of DNA fragments obtained is thought to increase...
expONENTIALLY PRODUCING 2n STRANDS FROM THE ORIGINAL DNA TEMPLATE WITH EACH CYCLE IN WHICH, THE WHOLE AMPLIFICATION PROCESS IS MEDIATED BY A PROKARYOTIC DNA DEPENDENT DNA POLYMERASE (SAIKI ET AL., 1985).

2.7.1 Preparation of the PCR Reaction Mixture

The PCR reaction mixture consisted of an appropriate buffer having the right concentration of MgCl₂, the four exclusive nucleotides (dATP, dCTP, dTTP and dGTP), forward and reverse primers, genomic DNA, and finally Taq polymerase that catalyses the actual reaction. The final volume of the reaction mixture used was 25µl. The mixes were the same for all single nucleotide polymorphisms (SNPs) tested by DNA sequencing. To each single PCR tube, the PCR mix consisted of 12µl 2x redy mix, 11µl of sterile distilled water and 0.2µl of each forward and reverse primer (50pM) required for the specific PCR, and finally the addition of 2µl (100ng) of genomic DNA sample.

A positive and a negative control were also prepared with each batch of samples. This is important to monitor each run being made and as part of quality control standard practice. The positive control was made up of a previously known sample on which the PCR was already performed and DNA was amplified, while a negative control consisted of the same reaction mixture however without the addition of DNA. The negative control is important to monitor and point out any contamination that may result. All tubes were placed into the thermal cycler equipment (Biometra T3000 thermal cycler) and thermal cycling was initiated.

2.7.2 Thermal profile for the Mutations and SNPs tested

The thermal profiles of the PCR runs were optimized by changing the annealing temperatures. This was achieved by calculating to be 5°C lower from the Tm of the primers used. The PCR for all SNPs in the study were carried out for 30 cycles with
each cycle being optimized as follows; denaturation at a temperature of $95^\circ C$ for 50 seconds, annealing temperature of between 55 to $65^\circ C$ (depending which SNP, see table 2.2 for details) for 1 minute and elongation at a temperature of $72^\circ C$ for 1 minute. At the end of the last cycle, a final extension cycle of 10 minutes at a temperature of $72^\circ C$ was performed. Once PCR was complete, all samples were removed from the thermal cycler and stored at $-20^\circ C$ prior to analyses.

<table>
<thead>
<tr>
<th>Site</th>
<th>Annealing Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>KLF1</td>
<td>56°C</td>
</tr>
<tr>
<td>5'Gγ Xmn I</td>
<td>61°C</td>
</tr>
<tr>
<td>5'β (AT)γ Tγ</td>
<td>58°C</td>
</tr>
<tr>
<td>Betathal</td>
<td>53°C</td>
</tr>
</tbody>
</table>

Table 2.2 represents the annealing temperatures used for PCR

2.7.3 PCR product determination by agarose gel electrophoresis

PCR fragments obtained in this study were determined by an agarose gel electrophoresis. The concentration of agarose used for determining the PCR fragments this time varied from the one used in whole genomic DNA determination, as the concentration depends on the expected size of the product. The expected size of the PCR products is already given in table 2.1. In most cases a 1.5% agarose gel containing ethidium bromide was used. 5μl of each of the PCR products were pipetted directly into each separate well in the gel. 7μl of the size marker 100bp ladder were loaded into the first well of each row to act as a reference marker used for comparison with the PCR product to ensure that the length of the fragment obtained was the one expected.

Electrophoresis for all PCR products was carried out in 1x TAE (Tris-acetate EDTA; pH 7.3) or 1x TBE (Tris- Borate- EDTA; pH 8.0) at a voltage of 160Volts and a current of 400mA for 20 minutes. The gel was then illuminated using UV irradiation and photographed. A log graph of the fragment sizes of the 100bp DNA ladder
against distance travelled was produced. The size of the PCR fragment obtained was calculated from this graph.

### 2.8 Restriction enzyme digestion

The \textit{Xmn} I RFLP introduces or abolishes a restriction enzyme recognition site, which can be readily detectable on an agarose gel electrophoresis. The \textit{Xmn} I constitute the basis of elevated Hb F in certain cases. The enzyme was purchased from New England Bio Labs Inc. and are as follows:

\begin{align*}
5' & \text{G A A N N}^\text{\textasteriskcentered} \text{N N T T C} \ldots 3' \\
3' & \text{C T T N}^\text{\textasteriskcentered} \text{N N A A G} \ldots 5'
\end{align*}

\textit{Xmn} I recognition site

### 2.8.0 Preparation of reaction mixtures

A reaction mixture of \textit{Xmn}I was prepared for each tube by first pipetting 2.0\mu l of 10x NE buffer 4 (50mM NaCl, 10mM Tris-HCl, 10mM MgCl\textsubscript{2} and 1mM DTT pH 7.9) with 5.0 units (0.25\mu l) of enzyme \textit{Xmn} I in separate mix containing 0.2\mu l of 100x BSA and 14.55\mu l of sterile water. 14.3\mu l of sterile water were then added to add up the volume up to 17\mu l. 3\mu l from each PCR product were added into corresponding tubes to make up a total reaction volume of 20\mu l. Known positive and negative controls were also included with every batch of samples done. A known homozygote for the cleaved site was used as positive control while as a negative control a sample with no enzyme added to the reaction mixture containing the PCR product was used. All tubes were placed in a thermal cycler (Cyclogene, Techne Ltd., Cambridge, UK) that was pre-programmed to 37^\circ C for 16 hours for all digests. After incubation all tubes were stored at -20^\circ C until analysed.
2.9 Detection of the digested products using agarose gel electrophoresis

Considering the size of the expected fragments of the β globin locus genotypes it was decided that a 2 to 2.5% agarose gel would give good fragment length separation, and resolution. The agarose gels were prepared as described in section 2.8. Electrophoresis was carried out at a voltage of 15Volts/cm and a current of 300mA for 1 hour and 30 minutes. The electrophoresis was stopped after a desirable migration was observed and the gel was removed from the electrophoresis tank. The gel was viewed on a scanner emitting UV irradiation, and photographed for records and for genotyping results. Repeats were performed for those samples that did not yield enough digest products. The genotypes of individual DNA samples were determined, recorded and documented in order to be able to correlate with the genotypes obtained from the cord blood. The Xmn I genotypes by RFLP were determined as those having no restriction enzyme cleavage site and denoted as -/-, those homozygotes for the cleavage site as +/-, while the heterozygotes were denoted as +/-.

2.10 DNA Sequencing

Complete β-globin gene, including its promoter (and AT_xT_y), 5' Gγ -158 C→T Xmn I site and complete KLF1 gene, were analysed by DNA sequencing using a modification of the Sanger di-deoxynucleotide chain termination chemistry, in which terminators are labelled with fluorescent dyes for automated detection. DNA sequencing was used to genotype a microsatellite repeat (AT)_xT_y which is found -530bp upstream of the β globin gene and Xmn I, and to screen for point mutations in the KLF1 gene that are known to cause high Hb F.
2.10.1 DNA Purification using centrifugal filter devices

PCR products to be used for DNA sequencing were purified using microcon centrifugal filter devices. These filter devices serve simply and efficiently concentrate and desalt DNA samples for downstream processes such as sequencing. A microcon sample reservoir was inserted into a 1.5 mL vial. Without touching the membrane with the pipette tip, 20\mu l of PCR product were pipetted into the sample reservoir and sealed by attaching cap in place. The whole assembly was placed in a microcentrifuge and counterbalanced with a similar device, and centrifuged at 500 x g for 12 minutes. The filtrate was discarded and the sample reservoir retained. 20\mu l of sterile distilled water were pipette into the sample reservoir and left to stand for 5 minutes at room temperature. The sample reservoir was then inverted 1800 so as to face down into a new 1.5 mL microcentrifuge tube. The assembly was then centrifuged for 1,500 x g for 5 minutes at which point DNA is collected at the bottom, desalted and purified.

2.11 Real-Time PCR

2.11.1 Real Time PCR Analysis

RT-PCR protocol was optimised for the best possible allele specificity, using a set of samples with known genotypes obtained from a previous study (M. Bugeja personal communication). A known heterozygote sample was tested using eight replicates, since the result obtained from it was used for correction of differential amplification, for all the other samples of the same SNP.

Two RT-PCR mixes, consisting of QuantiTect SYBR Green master mix (Qiagen GmbH; Cologne, Germany) together with forward, reverse primers and sterile water were first prepared using the allele specific primers mentioned in table 2.3. One RT-PCR mix contained the wildtype allele specific primer together with the common reverse primer, while the other mix contained the mutant allele specific primer together with the reverse common primer. The two RT-PCR mixes were prepared by multiplying the volumes of the components mentioned in the table 2.4 according to the number of samples tested. Each RT-PCR mix was prepared in excess, to
compensate for any losses of volume due to pipetting. The multiplied volumes were pipetted in a microfuge tube.

<table>
<thead>
<tr>
<th>Primer set</th>
<th>Site</th>
<th>Sequence 5' → 3'</th>
<th>T_m</th>
<th>PCR product (bp)</th>
<th>SNP rs number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABCL-FW1</td>
<td>CCG TTT AGC TTT ATT AAG GTA TCA</td>
<td>64°C</td>
<td>146</td>
<td>rs766432</td>
</tr>
<tr>
<td></td>
<td>ABCL-FW2</td>
<td>CCG TTT AGC TTT ATT AAG GTA TCC</td>
<td>62°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ABCL-RV</td>
<td>GGC TTT CTA GAC TGG TGG ACG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>BBCL-FW1</td>
<td>CCA GAA TCA TTC TGC TCT G1I G</td>
<td>56°C</td>
<td>149</td>
<td>rs11886868</td>
</tr>
<tr>
<td></td>
<td>BBCL-FW2</td>
<td>CCA GAA TCA TTC TGC TCT G1I A</td>
<td>60°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BBCL-RV</td>
<td>CCT TCC CTA ACC CTC TGA CC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CBCL-FW1</td>
<td>GTG CTG TGG ACA GCA AAG CTT ΔA</td>
<td>66°C</td>
<td>115</td>
<td>rs4671393</td>
</tr>
<tr>
<td></td>
<td>CBCL-FW2</td>
<td>GTG CTG TGG ACA GCA AAG CTT ΔG</td>
<td>64°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CBCL-RV</td>
<td>CTT CCT GTT CAC CTC TGA TCC C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>AMYB-FW</td>
<td>GCT GGT TAT GGG AAT AGA GAG</td>
<td>64°C</td>
<td>105</td>
<td>rs4895441</td>
</tr>
<tr>
<td></td>
<td>AMYB-R1</td>
<td>CTC AGT TCT CTG CTC ATG TCT</td>
<td>62°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AMYB-R2</td>
<td>CTC AGT TCT CTG CTC ATG TCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CMYB-FW</td>
<td>ACT GTG TGC TTA ATG AAA GAT C</td>
<td>56°C</td>
<td>126</td>
<td>rs28384513</td>
</tr>
<tr>
<td></td>
<td>CMYB-R1</td>
<td>AGC TAC CTA CGC CAG CGT TAT</td>
<td>60°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CMYB-R2</td>
<td>AGC TAC CTA CGC CAG CGT TAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>EMYB-FW</td>
<td>GGT CTG CTT CTA CAA TAG ACA T</td>
<td>66°C</td>
<td>137</td>
<td>rs9399137</td>
</tr>
<tr>
<td></td>
<td>EMYB-R1</td>
<td>GGT TGC TGG TGA AAA AAC TGG G</td>
<td>64°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EMYB-R2</td>
<td>GGT TGC TGG TGA AAA AAC TGG A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3 Allele specific primers for BCL11A and MYB polymorphisms
Chapter 2 Materials and Methods

<table>
<thead>
<tr>
<th>Component:</th>
<th>Volume per reaction:</th>
<th>Final Concentration:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2x QuantiTect SYBR Green PCR Master Mix</td>
<td>12.5 µl</td>
<td>1x</td>
</tr>
<tr>
<td>Forward Primer (wildtype or mutant allele specific primer)</td>
<td>0.1 µl</td>
<td>0.2 µM</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>0.1 µl</td>
<td>0.2 µM</td>
</tr>
<tr>
<td>Sterile Water</td>
<td>9.8 µl</td>
<td>-</td>
</tr>
<tr>
<td>Template DNA</td>
<td>2.5 µl</td>
<td>15.6 ng/reaction</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>-</td>
<td>2.5 mM</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td>25 µl</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2.4 Reaction Components required for each RT-PCR reaction

A total volume of 25 µl was used for each RT-PCR reaction. 22.5 µl of RT-PCR mix together with 2.5 µl of DNA template, were pipetted in each well of a MicroAMp™ Optical Reaction Plate (Applied Biosystems; Foster City, USA). Every RT-PCR reaction contained one of the RT-PCR mixes. The NTC contained 2.5 µl sterile water instead of the DNA template. Two negative controls (no template controls-NTCs) (one for the wildtype mix and another for the mutant mix) were included within each run to verify the quality of each real time PCR run.

The MicroAMp™ Optical 96-Well Reaction Plate (Applied Biosystems; Foster City, USA) was then covered with a MicroAMp™ 96 Well Optical Adhesive Film (Applied Biosystems; Foster City, USA) and sealed using a MicroAmpTM Adhesive Film Applicator (Applied Biosystems; Foster City, USA). The plate was then placed in the Applied Biosystems 7300 Real time PCR System (Applied Biosystems; Foster City, USA).
Chapter 2

City, USA). The thermal profile used was the one shown in Table 2.5. A dissociation stage (figure 2.2) was included within each RT-PCR run to determine any unspecific product formed. This stage consisted of 15 seconds at 95°C, 30 seconds at 60°C and 15 seconds at 95°C and these conditions were automatically set by the 7300 System Sequence Detection Software (Version 1.3.1). The amplification plot threshold line was always set to 0.2, while the baseline was always set from cycles 3-15.

<table>
<thead>
<tr>
<th>Step</th>
<th>Time</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR initial activation step</td>
<td>15 min</td>
<td>95°C</td>
</tr>
<tr>
<td>Cycling Steps: 40 cycles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>20 seconds</td>
<td>94°C</td>
</tr>
<tr>
<td>Annealing/Extension (detection)</td>
<td>1 minute</td>
<td>64°C</td>
</tr>
</tbody>
</table>

Table 2.5 Thermal Profile used for allele specific genotyping by RT-PCR

Figure 2.2: An output of the ABI 7300 System Sequence Detection Software (SDS) showing the dissociation curve analysis carried out for every run with narrow peaks indicating a pure PCR product reaction.
2.12 Statistics and Data Analysis

Statistical and data analysis to obtain genotypes for the \textit{BCL11A} and \textit{MYB} was as follows; each RT-PCR reaction for wildtype and mutant allele genotyping was carried out in triplicate. At the end of the RT-PCR run, the product formed was also checked on an agarose gel electrophoresis to verify its size. The RT-PCR run gave satisfactory results because the wildtype and mutant primers amplified with optimal specificity. This can be shown by the differential amplification of the primers, that was very close to zero (0.15). In addition, the $\Delta$C$_T$s for the homozygous wildtype and mutant samples were almost the same (8.0 and -7.3 for the homozygous wildtype and mutant samples respectively, see Section 3.0 – Results Chapter for amplification RT-PCR output. This indicated that the thermal profile chosen and reagent preparation were ideal to conduct genotyping by RT-PCR.

All \textit{Xmn1} genotypes were obtained by PCR – RFLP using the \textit{Xmn1} restriction enzyme digest method and recorded genotypes in a database till further studied using the PASW 18 statistical package suite for Windows (SPSS, Chicago, IL). The chi-square test was used to test for any significant difference in the distribution of the genotype frequencies and Fisher’s exact test for any significant difference in the distribution of alleles, for all SNPs genotyped in this study. The Hardy-Weinberg equilibrium was also calculated and a chi-square test was performed between the observed and the expected genotype frequencies. Haplotype frequencies were also constructed using the estimated haplotypes (EH) program and a chi-square test was performed to compare the haplotype frequencies across the different patient subgroups of $\beta$ thalassaemia. All statistical tests were considered two tailed at a level of significance 0.05, where the null hypothesis (no association between alleles and of genotypes/haplotypes with the Hb F expression) was rejected at p<0.05. The \textit{BCL11A} and \textit{MYB} genotype distribution according to $\beta$ thalassaemia mutation were plotted by scatter plot using PASW18.
CHAPTER 3

RESULTS
3.1 Molecular Epidemiology

3.1.1 Frequency of $\beta$ thalassaemia in the Maltese Population

One hundred and sixty seven (167) $\beta$ Thalassaemia heterozygotes were identified as part of the screening and referral program from the Thalassaemia and Molecular Genetics Clinic including the Ante Natal Clinic, Mater Dei Hospital, Malta. A full blood count, HbF and HbA$_2$ quantification were conducted as described in Chapter 2. Figure 3-1 shows the seven-$\beta$ Thalassaemia alleles identified together with their %. The most commonly identified mutation in Malta is the $\beta^+$ IVS-I-6C (59.3%), followed by $\beta^0$ Codon 39 (22.2%) and IVS-I-110A (7.8%) and IVS-II-1A (6.4%).

![Figure 3-1: The distribution of $\beta$ Thalassaemia alleles in the Maltese Population. The $\beta^+$ IVS1 nt 6C represents the largest proportion from all the rest and is depicted in light brown color. This is followed by Codon 39, IVS 1 nt 110A and IVS2 nt1 A mutations. Rare occurrence of other thalassaemia mutations is also shown.](image-url)
3.1.2 β globin gene sequencing reveals a rare deletion in the human β globin gene.

A 43 year old male was identified to carry a 2bp deletion in the HBB gene. His HbA2 was of 5.1 and HbF of 0.7%. His total Hb was 12.7g/dL and a very low MCV count of 55.5fl. The individual carried both an XmnI negative (-/-) and a common (AT)7T7 homozygote genotype. The mutation is at position HBB:c.17_18delCT Codon 5 (-CT) resulting in a frame shift mutation CCT(Pro)->C-- β0. Figure 3.2 represents the location of the deleted 2 nucleotides causing an overlapping DNA sequence downstream in the electrophoretogram.

Figure 3-2: DNA sequence chromatogram showing the location of the deleted 2 base pairs (-CT) and is marked with an arrow (↓)
Chapter 3 Results

3.1.3 Age and Gender distribution of individuals who are \( \beta \) thalassaemia heterozygote

A large portion of the individuals were females (65%) that included referred and Ante Natal Clinic visits. The rest (35%) were referred males. More females appear to have a \( \beta \) thalassaemia heterozygous allele than males. Any referred male subjects were independent and unrelated from the female subjects in this study.

3.1.4 Histogram graphs for HbA\(_2\) and MCV

The haematological indicators for \( \beta \) Thalassaemia have long been a raised HbA\(_2\) and a lower MCV with variable total haemoglobin. Figure 3-3, 3-4 and 3-5 shows a histogram distribution of these indices in the \( \beta \) Thalassaemia population studied. Table 3-1 shows the mean and standard error of the haematological indices measured in this study (HbA\(_2\) and MCV) together with total haemoglobin and foetal haemoglobin. One-way ANOVA testing conducted between the IVS1nt6C, COD39, IVS1nt110A and IVS2nt1A to determine whether there are any significant differences in HbA\(_2\), HbF, MCV and total Hb measurements yielded strong positive results. HbA\(_2\), MCV and total Hb had a p value of 0.03 or less whilst that of HbF was of 0.053, marginally significant.

<table>
<thead>
<tr>
<th>( \beta ) Thalassaemia allele</th>
<th>N</th>
<th>HbA(_2)</th>
<th>HbF g/dl</th>
<th>MCV</th>
<th>Hb</th>
</tr>
</thead>
<tbody>
<tr>
<td>IVS1nt1A</td>
<td>1</td>
<td>4.9 ±0.1</td>
<td>0.02</td>
<td>59.7</td>
<td>10.7</td>
</tr>
<tr>
<td>IVS1nt2A</td>
<td>2</td>
<td>6.6 ±0.1</td>
<td>0.026 ±0.004</td>
<td>60.9 ±0.4</td>
<td>10.3 ±0.1</td>
</tr>
<tr>
<td>IVS1nt6C</td>
<td>99</td>
<td>4.3 ±0.07</td>
<td>0.08 ±0.017</td>
<td>69.0 ±0.5</td>
<td>12.04 ±0.2</td>
</tr>
<tr>
<td>IVS1nt110A</td>
<td>14</td>
<td>4.96 ±0.13</td>
<td>0.14 ±0.06</td>
<td>63.7 ±0.9</td>
<td>11.7 ±0.31</td>
</tr>
<tr>
<td>COD 39</td>
<td>37</td>
<td>5.46 ±0.14</td>
<td>0.16 ±0.02</td>
<td>62.0 ±1.1</td>
<td>11.25 ±0.2</td>
</tr>
<tr>
<td>IVS2nt1A</td>
<td>13</td>
<td>5.45 ±0.24</td>
<td>0.23 ±0.06</td>
<td>65.2 ±2.0</td>
<td>10.6 ±0.32</td>
</tr>
<tr>
<td>HBB:c.17_18delCT Codon 5 (-CT)</td>
<td>5.1</td>
<td>0.09</td>
<td>55.5</td>
<td>12.7</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3-1: Representation of the mean and standard error measurements for the four important \( \beta \) Thalassaemia indicators. The mutations are represented in hierarchical order according to their position in the \( \beta \) globin gene.
Chapter 3

The histogram below shows the level of total Hb measured in all the β thalassaemia heterozygotes. These measurements overlap between normal and below normal accepted range (i.e., < 11 g/dL). The low levels encountered in the β thalassaemia heterozygotes can be accounted due to pregnant females at the time of haemoglobin measurement.

Figure 3-3: The level of haemoglobin across all β Thalassaemia heterozygotes
Figure 3-4: Histogram graph showing MCV distribution in all β Thalassaemia heterozygotes in this study. Reference range for normal healthy adults without thalassaemia is between 80 and 87. One individual is shown as an outlier from the normal distribution histogram and carries normal levels of MCV with a total Hb of 9.6g/dL (1).
Figure 3-5: Histogram graph showing HbA2 distribution in all β Thalassaemia heterozygotes in this study.

HbA2 for normal healthy adults without thalassaemia should be < 3.
3.1.5 The levels of Foetal Haemoglobin (g/dL) in β Thalassaemia Heterozygotes.

The HbF levels were quantified for all patients and the value in g/dL was calculated by using the formula: total Hb g/dL * HbF % / 100. The HbF levels were plotted and stratified according to different β Thalassaemia alleles observed. This is shown in figure 3-6. From a total of 167 β thalassaemia heterozygotes, 31 (19%) had an HbF level of > 0.14 g/dL.

Figure 3-6: HbF g/dL plotted versus different types of β thalassaemia alleles.
3.2 Genotyping of the Xmn I and (AT)$_x$T$_Y$ polymorphisms in the β Thalassaemia Heterozygotes.

Genotyping was carried out for two important sites in the β globin locus, the XmnI site and the (AT)$_x$T$_Y$ repeat polymorphism. Figure 3-7 shows the location of these sites that have been genotyped.

![Figure 3-7: The location of XmnI and (AT)$_x$T$_Y$ in the β globin locus.](image)

The genotyping results for XmnI and (AT)$_x$T$_Y$ are shown in table 3-2 and 3-3. Table 3-4 represents the combined genotypes together, whilst figure 3-8 represents an agarose gel electrophoresis for the digested XmnI fragments.

![Figure 3.8: Agarose separation of PCR-RFLP for XmnI digest. Well positions (1) mutant [++] (2) wildtype [-/-] (3) heterozygote [+/-] and (4) to (6) are wildtype [-/-]. A 100bp ladder is also represented in well position (7).](image)
### Chapter 3 Results

<table>
<thead>
<tr>
<th>Allele Frequency</th>
<th>N</th>
<th>HbF g/dL</th>
<th>Allele Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>XmnI Homozygous Wildtype</td>
<td>130</td>
<td>0.1 ±0.01</td>
<td>C allele 0.88</td>
</tr>
<tr>
<td>Heterozygote</td>
<td>35</td>
<td>0.13 ±0.03</td>
<td>T allele 0.17</td>
</tr>
<tr>
<td>Homozygous Mutant</td>
<td>2</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>167</td>
<td>334</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-2: XmnI genotypes in β thalassaemia heterozygotes

<table>
<thead>
<tr>
<th>Haplotypes (Alleles)</th>
<th>N</th>
<th>HbF g/dL</th>
<th>Allele Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(AT)77T7 (AT)77T7</td>
<td>128</td>
<td>0.1 ±0.01</td>
<td>(AT)7T7 0.87</td>
</tr>
<tr>
<td>(AT)77T7 (AT)9T5</td>
<td>34</td>
<td>0.14 ±0.03</td>
<td>(AT)9T5 0.13</td>
</tr>
<tr>
<td>(AT)9T5 (AT)9T5</td>
<td>5</td>
<td>0.13 ±0.08</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>167</td>
<td>334</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-3: (AT)7T7 genotypes in β thalassaemia heterozygotes

<table>
<thead>
<tr>
<th>Haplotypes (Alleles)</th>
<th>N</th>
<th>HbF g/dL</th>
<th>Allele Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>XmnI -ve (AT)77T7</td>
<td>102</td>
<td>0.09 ±0.01</td>
<td>XmnI-ve (AT)7T7 0.76</td>
</tr>
<tr>
<td>XmnI-ve (AT)9T5</td>
<td>24</td>
<td>0.13 ±0.04</td>
<td>XmnI-ve (AT)9T5 0.09</td>
</tr>
<tr>
<td>XmnI +ve (AT)77T7</td>
<td>4</td>
<td>0.05 ±0.02</td>
<td>XmnI+ve (AT)7T7 0.08</td>
</tr>
<tr>
<td>XmnI +ve (AT)77T7</td>
<td>25</td>
<td>0.12 ±0.3</td>
<td>XmnI+ve (AT)9T5 0.006</td>
</tr>
<tr>
<td>XmnI +/- (AT)77T7 (AT)9T5</td>
<td>10</td>
<td>0.17 ±0.06</td>
<td>XmnI+/- (AT)7T7 (AT)9T5 0.06</td>
</tr>
<tr>
<td>XmnI +/ (AT)77T7 (AT)7T7</td>
<td>1</td>
<td>0.05*</td>
<td></td>
</tr>
<tr>
<td>XmnI +/+ (AT)9T5 (AT)9T5</td>
<td>1</td>
<td>0.4*</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>167</td>
<td>334</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-4: Combined XmnI and (AT)7T7 genotypes in β thalassaemia heterozygotes. (*) Values marked with an asterix come from single individuals hence mean values cannot be compared.
In figure 3-9 the HbF g/dL has been plotted versus β thalassaemia alleles and labelled by XmnI genotyping results. Only two XmnI homozygotes were identified. One in IVS-I nt6C and one in IVS-II nt1A. Figure 3-10 shows the same stratification of HbF g/dL but labelled by (AT)\textsubscript{x}T\textsubscript{Y} genotypes.

Figure 3-9: HbF g/dL of β thalassaemia heterozygotes stratified by their β thalassaemia mutation and XmnI genotype.
Figure 3-10: HbF g/dL of β thalassaemia heterozygotes stratified by their β thalassaemia mutation and (AT)$_x$Ty genotype.
3.3 Genotyping of modifier SNPs in BCL11A and MYB genes

Three polymorphisms in BCL11A and another three polymorphisms in MYB were selectively genotyped using Allele specific real time PCR by Sybr Green chemistry. The genotype results are shown in table 3-5.

<table>
<thead>
<tr>
<th></th>
<th>BCL11A</th>
<th>MYB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rs766432</td>
<td>rs28386513</td>
</tr>
<tr>
<td>Homozygous Wildtype</td>
<td>99</td>
<td>74</td>
</tr>
<tr>
<td>Heterozygotes</td>
<td>36</td>
<td>67</td>
</tr>
<tr>
<td>Homozygous Mutant</td>
<td>31</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 3-5: Genotypes of BCL11A and MYB in β thalassaemia heterozygotes

A map of the intergenic polymorphisms between MYB and HBS1L is shown in figure 3-11. The RT-PCR genotyping output for BCL11A rs11886868 homozygous wildtype, heterozygous and homozygous mutant is shown in figures 3-12 to 3-14 respectively.

Figure 3-11: The 3 dimorphisms genotyped on chromosome 6q22.3 fall between MYB and HBS1L in an intergenic region that is comprised of introns of the same genes.
Figure 3-12 RT-PCR amplification output for a heterozygous *BCL11A* rs11886868 (reactions carried out in triplicate)

Figure 3-13 RT-PCR amplification output for a homozygous wildtype *BCL11A* rs11886868 (reactions carried out in triplicate)

Figure 3-14 RT-PCR amplification output for a homozygous mutant *BCL11A* rs11886868 (reactions carried out in triplicate)
Chapter 3

Results

All BCL11A and MYB SNPs were in HWE except rs766432 ($\chi^2 = 38.1$) and rs4895441 ($\chi^2 = 7.96$). This may be due to possible bias in the sample collection since the cohort does not represent a random collection of the Maltese population. The HbF g/dL of all $\beta$ thalassaemia patients was correlated with both BCL11A and MYB genotypes (table 3-6)

<table>
<thead>
<tr>
<th></th>
<th>Genotype</th>
<th>Mean ± S.E. HbF g/dL</th>
<th>p value</th>
<th>F</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCL1A</td>
<td>rs766432</td>
<td>-/-</td>
<td>0.09 ±0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-/+</td>
<td>0.12 ±0.01</td>
<td>p = 0.458</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+/+</td>
<td>0.13 ±0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs11886868</td>
<td>-/-</td>
<td>0.09 ±0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-/+</td>
<td>0.12 ±0.02</td>
<td>p = 0.299</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+/+</td>
<td>0.14 ±0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs4671393</td>
<td>-/-</td>
<td>0.09 ±0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-/+</td>
<td>0.13 ±0.02</td>
<td>p = 0.004</td>
<td>5.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+/+</td>
<td>0.30 ±0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYB</td>
<td>rs4895441</td>
<td>-/-</td>
<td>0.09 ±0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-/+</td>
<td>0.08 ±0.01</td>
<td>p = 0.001</td>
<td>12.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+/+</td>
<td>0.30 ±0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs28384513</td>
<td>-/-</td>
<td>0.10 ±0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-/+</td>
<td>0.09 ±0.01</td>
<td>p = 0.173</td>
<td>1.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+/+</td>
<td>0.16 ±0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs9399137</td>
<td>-/-</td>
<td>0.10 ±0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-/+</td>
<td>0.10 ±0.02</td>
<td>p = 0.186</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+/+</td>
<td>0.18 ±0.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3-6: One Way ANOVA for the mean HbF g/dL ± S.E. in all $\beta$ thalassaemia heterozygotes versus the three genotypes of each SNP is shown in the table. Significant p values are highlighted in red and represent statistical significant associations between the dimorphism tested and a higher mean level of HbF in the $\beta$ thalassaemia heterozygotes with homozygous mutant dimorphisms in rs4671393 (BCL11A) and rs4895441 (MYB)

Six scatter plots (figures 3-15 to 3-20) have been obtained for the distribution of HbF g/dL in all the different $\beta$ thalassaemia heterozygote categories and labelled according to their genotype status of both BCL11A and MYB genes. It is of notable interest that the $\beta^0$ thalassaemia codon39 heterozygotes exhibit higher HbF levels than the rest of $\beta$ thalassaemia heterozygote types when one takes into account the MYB SNPs and excluding the three outliers in IVS-I-6C, this is better visualized in figs 3-18 to 3-20.
Chapter 3

Results

Figure 3-15: BCL11A - rs766432 genotype distribution in β thalassaemia heterozygotes.

○ AA ○ AC ○ CC

Figure 3-16: BCL11A - rs11886868 genotype distribution in β thalassaemia heterozygotes.

○ TT ○ TC ○ CC

57
Figure 3-17: BCL11A – rs4671393 genotype distribution in β thalassaemia heterozygotes.

Figure 3-18: MYB – rs28384513 genotype distribution in β thalassaemia heterozygotes.
Figure 3-19: MYB – rs4895441 genotype distribution in β thalassaemia heterozygotes.

rs4895441
- Homozygous Wildtype
- Heterozygote
- Homozygous Mutant

Figure 3-20: MYB – rs9399137 genotype distribution in β thalassaemia heterozygotes.

rs9399137
- Homozygous Wildtype
- Heterozygote
- Homozygous Mutant

59
3.3.1 *In vivo* HbF expression varies depending on the number of minor allele frequency (MAF) SNPs in MYB but not in BCL11A

The levels of HbF g/dL of two groups of β Thalassaemia (IVS1-6C and COD39) were correlated together and stratified according to the number of rare polymorphisms in the MYB region. The two sample groups were both XmnI negative and carried an \((AT)_7\) homozygous repeat polymorphism. From figure 3-21 it can be seen that the levels of HbF g/dL in COD39 β Thalassaemia heterozygotes depends and varies on the number of MYB rare polymorphisms present in their genotype. This study was based on the three most prominent SNPs known in MYB that were previously associated with HbF increase.

The same cannot be said for rare polymorphisms in BCL11A. The HbF levels between the two groups did not correlate well between themselves, and the distribution was independent of the genotypes observed. Figure 3-22 shows the independent distribution of HbF g/dL in both groups.
Figure 3-21: Effects of *cis-trans* interplay with MYB on the HbF of β Thalassaemia heterozygotes
Chapter 3 Results

0.400
0.300
0.200
0.100
0.000
-0.100

Mean +/- 2 SE
HbF g/dL

N = 20
6
12
9
13
3
10
2
7
2
2
1

Cumulative number of BCL11A SNPs on both alleles
rs766432 rs11886868 rs4671393

Figure 3-22: No effects of *cis-trans* interplay with BCL11A on the HbF of β Thalassaemia heterozygotes
Tables 3.7 and 3.8 show the constructed haplotypes for the MYB and BCL11A dimorphisms. These were constructed using EH software.

<table>
<thead>
<tr>
<th>Number of MYB SNPs</th>
<th>N = cases</th>
<th>Haplotype Frequency</th>
<th>rs28384513</th>
<th>rs4895441</th>
<th>rs9399137</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>45</td>
<td>27.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>27</td>
<td>16.5</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>15</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>13</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>9.1</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>7.7</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>7</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>4.2</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 3.7 Haplotype frequencies for MYB dimorphisms. (-) indicates the presence of the wildtype allele whilst (+) indicates the presence of the mutant allele.

<table>
<thead>
<tr>
<th>Number of BCL11A SNPs</th>
<th>N = cases</th>
<th>Haplotype Frequency</th>
<th>rs766432</th>
<th>rs4671393</th>
<th>rs11886868</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>63</td>
<td>37.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>34</td>
<td>20.5</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>15.8</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>7.9</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>8.5</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>4.3</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>3.3</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1.8</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 3.8 Haplotype frequencies for BCL11A dimorphisms. (-) indicates the presence of the wildtype allele whilst (+) indicates the presence of the mutant allele.

The prominent haplotypes for the MYB gene that appear responsible for a higher HbF in $\beta^0$ COD39 vs. $\beta^+$ IVS-I-6C includes; (1) two copies of the rare haplotype with mutant alleles in every dimorphism (i.e., ++ +) or a combination of one rare (+ + +) haplotype and at least one of the (- + +) or (+ + -) that contain two mutated dimorphisms from a total of three. It is noteworthy that the second dimorphism rs4895441 is commonly mutated across all
significant haplotypes and is also the only significant one from the three analyzed alone (refer to table 3.6). In BCL11A, no specific haplotype was more noticeable in $\beta^0$ COD39 vs. $\beta^+$ IVS-I-6C and therefore this gene does not appear to have specific haplotypes (or frameworks) that contribute to higher HbF but rather specific dimorphisms.
CHAPTER 4
DISCUSSION
4.0 Discussion

Extensive genotyping of 7 SNPs connected with γ / β globin gene expression among 167 β thalassaemia heterozygotes from Malta revealed further cis-trans interplay acting on the levels of Hb F without inter-current effects of severe anaemia as in the homozygotes. The distribution of the common and rare β-thalassaemia alleles among the population of Malta has been confirmed, but another rare allele (Kollia et al 1989) has also been found for the first time among Maltese β thalassaemia heterozygotes. In particular, however, an incremental effect of MYB alleles on the expression of Hb F among the β^0 COD39 but not the β^+ IVS-I-6C heterozygotes has been observed and raises important questions concerning trans acting regulators such as KLF1, BCL11A, MYB and possibly others that could be further documented in vivo.

4.0.1 Haematology

The identification of all β thalassaemia alleles in the test cohort of this study, included all the commonly encountered β thalassaemia mutations as well as a newly identified β^0 thalassaemia mutation. All β thalassaemia heterozygotes had a lower MCV than normal (66fl ±5.7) and a higher HbA2 (4.8% ±0.8). The predominant β thalassaemia mutation in Malta is the IVS-I-6C mutation and has remained largely the same during the past years. The next three mutations that are also identified and found on the Maltese islands include Codon 39T, IVS-I-110A, and IVS-II-1A (Scerri 1998). The remaining mutations are very rare, and only amount 1 to 2 individuals being detected with these mutations; and include IVS-I-1A and IVS-I-2A. Interestingly, one other rare observation was for a 2bp-deletion in codon 5 of the β globin gene, that resulted in a frameshift mutation and causing a β^0 allele. This was observed for the first time
on the Maltese islands, and it remains yet to be revealed whether additional carriers are known in Malta or not. In cases where β thalassaemia mutations are not identified yet; mRNA and/or β globin biosynthesis can be carried out to determine the α/non-α ratio. The ratio should be equivalent to 1 in cases without β thalassaemia and higher in the event that β thalassaemia is present. If the ratio is 1, and haematological indices still indicate that haemoglobin levels (including foetal haemoglobin) are not normal or perturbed, these are useful in identifying new regulators of haemoglobin.

The pathophysiological effects that determine elevated Hb F % among β thalassaemia homozygotes are complex and difficult to define in terms of genetic effects (Weatherall and Clegg, 2001). The preferential survival of F-erythrocytes and the effect of the anaemia and elevated erythropoietin or other cytokines on conditional SNPs / mutations such as the XMN1 dimorphism are thought to be the main determinants of Hb F in homozygotes. Most β thalassaemia heterozygotes are not severely anaemic (Hb < 10 g/dL) though a mild anaemia could occur during the mid-trimester in pregnant women. Anaemia among β thalassaemia heterozygotes is more likely due to inter-current conditions such as iron deficiency / pregnancy / inflammation and others. Few (< 10%) of the heterozygotes reported here were clinically anaemic (Hb < 10 g/dL) and, in any case, there was no correlation between the circulating Hb mass and Hb F expression independent of the β thalassaemia allele (Fig 4-1). Nevertheless, for the sake of accuracy the Hb F expression in vivo was quantified as the (absolute) Hb F in g / dL such as carried out previously on Maltese β thalassaemia homozygotes (Felice et al., 2007).
Figure 4-1: No correlation was observed between HbF % and Hb in g/dL for \( \beta^+ \) IVS-I-6C and \( \beta^0 \) Codon 39 thalassaemia heterozygotes

The Maltese \( \beta \) thalassaemia heterozygotes had typical haematological and Hb profiles as that of other populations with one exception (see figure 3-3). One \( \beta \) thalassaemia heterozygote carried normal levels of MCV but a total Hb value of 9.6 g/dL. The patient may be carrying normal levels of MCV due to double heterozygote with \( \alpha \) thalassaemia mutation thereby resulting in a normal MCV and MCH. However, the alpha thalassaemia mutation still needs to be confirmed in the further studies that are planned. Surprisingly, 3 \( \beta^+ \) IVS-I-6C heterozygotes and 1 \( \beta \) IVS-I-110A heterozygote had the highest levels of Hb F (see figure 3-6) from all \( \beta \) thalassaemia heterozygotes recruited. This is surprising considering that all the rest of all \( \beta \) thalassaemia heterozygotes follow a consistent pattern of higher HbF in the more severe type of \( \beta^0 \)
thalassaemia versus other milder forms of $\beta^+$ thalassaemia. Nineteen (19%) of the total $\beta$ thalassaemia heterozygotes analyzed had an Hb F > 0.140 g/dL. The $\beta$ thalassaemia allele frequencies in Malta are still unique and quite different from neighboring countries such as Italy, Greece and Tunisia. The $\beta^+$ IVS-I-6C in Malta accounts for almost 60% of the total thalassaemia alleles, whilst only 1.7, 8.4 and 10.1% for Tunisia, Greece and Italy respectively. The highest $\beta$ thalassaemia alleles in Italy, Tunisia, France, Spain and Portugal are $\beta^0$ Codon 39 (41%, 45%, 42%, 35.5% and 37.3% respectively) and in Greece are $\beta$ IVS-I-110A (42.5%). These latter mutations in Malta fall within much smaller numbers making our population quite distinct. The frequencies were obtained from the ITHANET online database of $\beta$ thalassaemia entries (http://www.ithanet.eu/base/freq) In another Mediterranean country to the south of Malta – Libya, the $\beta$ thalassaemia allele frequency ranged from 50% for the $\beta$ Codon 39, 25% for the IVS-I, 110A, 16.7% for the IVS-I, 6C and other rare mutations account for the rest (Marwan, 1998).

Although recent data from Maltese families had indicated the potential of KLF1 to act as master regulator of developmental globin gene expression, KLF1 mutations or polymorphisms in fact are very rare. KLF1 mutations that have been published so far from a Maltese family study have tested negative in all 167 $\beta$ thalassaemia heterozygotes. The KLF1 p.K288X mutation has only been found in one other family from Malta and none of the $\beta$ thalassaemia heterozygotes studied here were carriers. There are still further DNA sequences of the KLF1 gene that need to be studied in the 167 $\beta$ thalassaemia heterozygotes to fully exclude whether other mutations might be present or not. Therefore the underlying hypothesis of this study was the study of interplay between XMN1, $(AT)_xT_Y$, BCL11A and MYB in absence of heterogeneity in erythropoietic stress and KLF1.
On the other hand, Sequence variations in XMN1, \((AT)_X T_Y\), BCL11A and MYB were common. The XMN1 mutant allele (T) in the \(\beta\) thalassaemia heterozygotes is of around 11%. Perhaps lower than most of other studies conducted elsewhere, since the lowest recorded is of 18% and highest of around 32% (Thein \textit{et al} 2009). The XMN1 dimorphism is one of only three well documented "conditional mutations" in man. The others include the MTHFR C677T mutation and HLA-DQ2 DQ8 in Coeliac Disease.

The MTHFR C677T raises homocysteine levels when folate is deficient. Individuals who carry the HLA DQ2 and DQ8 have a strong genetic predisposition to develop coeliac disease. In general, individuals can lead a normal life except when they ingest gluten-containing food; this will precipitate a condition known as Coeliac Disease (Brett \textit{et al}., 1999). Once the gluten is removed from the diet, the individuals will revert back to normal. Hence, HLA class II genes, DQ2 and DQ8 are classical examples of conditional mutations. And as suggested above, an important conditional mutation is the \(-158 C\rightarrow T 5' Gy\) globin gene and known as XMN1. The XMN1 C\(\rightarrow\)T increases \(Gy\) in \(cis\) when erythropoiesis is stressed. It is the major determinant of \(Gy\) and Hb F levels in adult, anaemic thalassaemia homozygotes. In the neonate, it acts in concert with variability in the \((AT)_X T_Y\) site.

The \((AT)_X T_Y\) repeat polymorphism was already studied before in a subset of the Maltese population – Hb F Malta I as well as in a random number of cord DNA samples that represent the Maltese population. The same repeat polymorphism was also studied in the \(\beta\) thalassaemia heterozygotes and had similar frequencies to those encountered before. As evidenced in previous studies, the long \((AT)_X T_5\) was tightly linked with the Hb F Malta I allele whilst the remaining 'normal' allele \(in trans\) to the Hb F Malta I heterozygotes) had three different configurations in the order of
most common being \((AT)_7T_7\), \((AT)_9T_5\) and \((AT)_{11}T_3\). The longer \((AT)_xT_y\) morphs are more suppressive of Hb F, since beta protein - 1 (BP1) binds tighter to the longer \((AT)_xT_y\) types than the shorter and more common form of \((AT)_xT_y\). In the \(\beta\) thalassaemia heterozygotes, the \(XMN1(\text{AT})_xT_y\) haplotypes were constructed and correlated with HbF expression. The results showed diverse roles of the constructed haplotypes on Hb F expression however there is no direct or strong evidence that one \((XMN1)\) acts more or less than the other \([(AT)_xT_y]\). It is observed, that they are in fact working in concert together – i.e., interplay between the two alleles. There is also an observation of a higher mean in Hb F g/dL in the double heterozygote for \(XMN1\) and \((AT)_xT_y\) (mean = 0.17g/dL ±0.06).

4.1 Quantitative Trait Loci

\(BCL11A\) was extensively genotyped for three sites polymorphisms that have been gaining increasing interest in the field of haemoglobin switching and control. The data observed was consistent with published frequencies in the NCBI repositories. When comparing the genotypes of \(BCL11A\) with HbF expression data from the \(\beta\) thalassaemia heterozygotes, it was very encouraging to note a higher HbF expression with at least one of the mutant type dimorphisms (rs4671393). This SNP in \(BCL11A\) has been extensively studied in sickle cell anaemia patients but not so much in large studies in \(\beta\) thalassaemia patients. The exact mechanism or role of how this SNP functions is still poorly understood. The \(BCL11A\) dimorphisms genotyped in the Maltese \(\beta\) thalassaemia patients included rs766432, rs4671393 and rs11886868. They are all SNPs located in the intron of \(BCL11A\) and the exact function of how they are capable in augmenting HbF is not known. The results in this study are in perfect agreement with previous studies conducted also on more than a thousand sickle cell
anaemia patients by Galarneau et al (2010) and showed that rs4671393 is the most significant SNP in BCL11A.

The MYB was also studied for three well-known SNPs that are known to influence the Hb F levels in vivo under conditions with erythropoietic stress. In the MYB gene, there is perhaps more conflicting discussion between different studies. The most significant SNP in the Maltese β thalassaemia heterozygotes was rs4895441. This was already identified by Lettre et al (2008) as one of the top candidate dimorphisms in a large cohort of 1275 patients with sickle cell anaemia. However, a study by Coelho et al (2011) could not replicate this association albeit in a much smaller population (n=79). When looked in more detail, the MYB effect was combined across all three SNPs genotyped, and the effect was more evident in the β Codon 39 heterozygotes and not so much across all others. Hence, the MYB effect appears to be exclusive for the β Codon 39 in Maltese β thalassaemia heterozygotes. Why is this so? The exact answer is not known, yet. However it can be suggested that codon 39 and IVS-I-6C do have different genetic backgrounds and leading to different end-points. This T→C mutation at nt 6 of IVS-I in the β globin reduces the efficiency of splicing at the 5' site (figure 4-2).

<table>
<thead>
<tr>
<th>Codon #</th>
<th>IVS-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>AG^</td>
</tr>
<tr>
<td></td>
<td>GTTGGTATCAAGGTT</td>
</tr>
<tr>
<td>IVS-I-6 (T-&gt;C)</td>
<td>AG^</td>
</tr>
<tr>
<td></td>
<td>GTTGGCATCAAGGTT</td>
</tr>
</tbody>
</table>

Figure 4-2: The position of the T→C nucleotide substitution in the IVS-I of the β globin gene is shown by an underlined and italicized base.
β Codon 39 is the result of a pre-mature stop codon mutation from Glycine (CAG) → stop codon (TAG) and resulting in a complete absence of the protein peptide leading to a β0 allele and in the homozygote state a severe-type of β thalassaemia that is transfusion dependent. Hence MYB appears to function and augment HbF in β thalassaemia heterozygotes when there is absence of the β globin peptide (such as β0 Codon 39) more than simply an altered β globin transcript (deficiency in globin transcript) such as the case of IVS-I-6C. This shall be validated in further research planned on a larger scale in the Maltese patients. It is also plausible that a yet undetermined linked DNA change to the β0 Codon 39 remains to be identified that is capable of facilitating interaction with the mutant forms of MYB gene.

It appears increasingly evident due to the conflicting reports and results of these modifier genes and their implicated SNPs within their intron regions may have additional linked mutations or polymorphisms that harbor the true influence on HbF. Of course, different population types and genetic backgrounds will sometimes dilute or remove completely these ‘putative’ mutations or polymorphisms that are linked with the established ones. It can also be the case that haplotypes and allele scoring that includes quantitative SNPlotyploting across different genes, such as the case of MYB in the Maltese β thalassaemia heterozygotes, may be more meaningful and play a direct role in HbF control rather than individual SNPs. This has been exemplified by figure 3-17 showing that a higher number of MYB SNPs that consist of rare haplotypes with all dimorphisms mutated (i.e., 6 or at least 5 mutated allele) in β0 Codon 39 has a much higher HbF level than the β+ IVS-I-6C.
Some important limitations of the study include the lack of measurement of Ferritin in all the patients in the study group. Ferritin measurement is a helpful test for diagnosis of iron deficiency. Increased values are generally detectable in thalassaemia syndromes due to increased erythropoiesis, including beta thalassaemia heterozygotes. The test would also be important to evaluate the degree of iron deficiency – if present – in the beta thalassaemia heterozygotes.

### 4.2 Future Perspectives

The total number of beta thalassaemia heterozygotes ($n = 167$) is expected to increase for further studies on the same research approach. This may have been a limitation for the current study, and the positive associations identified with the BCL11A and MYB need to be interpreted and analyzed in the context of additional material. An important experiment would be to measure the quantity of MYB RNA/protein from patients with different MYB haplotypes or frameworks to assess the degree of impairment of MYB. This can be done by qPCR in the case of RNA and by Western blotting in the case of protein assays. Another interesting opportunity is to study the discrepancy in HbF levels between Thalassaemia IVS-I nt6 C patients who appear to have a rather heterogeneous distribution of HbF. It remains yet to be identified whether combined genotypes or extended haplotypes are responsible for this phenomenon or perhaps additional QTLs that are as yet unidentified.
4.3 Conclusion

The new observation that MYB gene frameworks in the context of $\beta^{0}$9 but not in the $\beta^{+}$IVS-I-6C raises new questions about mechanisms underlying the high Hb F of $\beta$ thalassaemia homozygotes. In conclusion, understanding the mechanism behind $\gamma$-globin gene switching and all the players involved in the process is critical to design new therapeutic approaches for $\beta$-thalassaemia and sickle cell disease. Such insights will also be important for revealing the basic features of developmental gene regulation of the $\beta$ globin locus. The development of different methods and approaches are necessary to overcome technical obstacles in understanding mechanisms of gene regulation. From our current observations, MYB and KLF1 are the most likely targets that require further investigation.

At the same time while we are trying to understand factors that directly regulate the $\gamma$-globin gene regulation, the molecular connections to phenomena such as signaling pathways, stress erythropoiesis and alterations in cell cycle kinetics as well as polymorphic genetic alternation are also of great importance. To shed light on indirect $\gamma$-globin regulatory mechanisms we studied several thalassaemia heterozygotes with high and low HbF levels. Future studies will elucidate the complex signaling and regulatory networks that control gene expression in this important developmental locus and will have important implications for the treatment of diseases such $\beta$-thalassaemia and sickle cell anaemia.
References


References


References


References


References


References


References


References


References


References


Appendix A

Statistical formulae

i. Mean ($\bar{x}$) and standard deviation (sd)

$$\bar{x} = \frac{\sum x_i}{n}$$

$$sd = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

Where $\bar{x} = \text{average or mean}$

$sd = \text{standard deviation}$

$x_i = \text{values}$

$n = \text{total number of values}$

ii. Chi square test

The chi-square test is based on the chi square distribution. It analyses together the observed data ($O$) with the expected data ($E$) in contingency table ($h \times k$) and determines whether the differences between the observed and expected scores can be attributed to some degree of significant difference between the two or if this difference between the data is caused by chance. The test is performed with the following formula:

$$\chi^2 = \sum_{i=1}^{hk} \frac{(O_i - E_i)^2}{E_i}$$
The degrees of freedom (v) of an estimate is calculated by \((h-1)(k-1)\).

### iii. Independent sample t-test

The independent sample t-test is a quantitative test that was used to test for any significant difference between the average age at diagnosis of male and female coeliac individuals (i.e., two independent normally distributed groups), assuming equal variances. The formula is as shown below;

\[
t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

and a mutual estimate of;

\[
S_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}
\]

where \(S_1\) and \(S_2\) = sample standard deviations

\(\bar{x}_1\) and \(\bar{x}_2\) = sample means

\(n_1\) and \(n_2\) = sample sizes

\(\mu_1\) and \(\mu_2\) = population means

degrees of freedom = \(v = n_1 + n_2 - 2\)
iv. **Fisher exact test**

Fisher's exact test is employed to test for independence in a 2 by 2 contingency table. This test was used to test the allele frequencies of both SNPs in the cord blood controls and coeliac individuals, testing the hypothesis that the two column percentages are equal. The end probability of the observed outcome of the test is given by the following formula:

\[
P = \frac{a! \cdot b! \cdot c! \cdot d!}{(a+b)! \cdot (c+d)! \cdot (a+c)! \cdot (b+d)!} N!
\]

Where \(a, b, c, d\) stand for the frequencies in the four cells and \(N = \text{total number of cases}\).
Appendices

Appendix B

Preparation of buffers for DNA extraction

10 X SE buffer

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>750mM</td>
<td>43.9 g/L</td>
</tr>
<tr>
<td>Na₂EDTA</td>
<td>250mM</td>
<td>93.1 g/L</td>
</tr>
</tbody>
</table>

Dissolved completely in deionised water and adjust to pH 8.0

10 X Erythrocyte lysing buffer

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH₄Cl</td>
<td>1.55 M</td>
<td>84.5 g/L</td>
</tr>
<tr>
<td>KHCO₃</td>
<td>100mM</td>
<td>10.0 g/L</td>
</tr>
<tr>
<td>Na₂EDTA</td>
<td>1mM</td>
<td>0.3722 g/L</td>
</tr>
</tbody>
</table>

Dissolve completely in deionised water and adjust to pH 7.4

N.B. To prepare working solution dilute to 1 X concentration with deionised water

Preparation of electrophoresis buffer

50 X Tris-acetate (TAE) buffer (per litre)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris base</td>
<td>242g</td>
</tr>
<tr>
<td>Glacial acetic acid</td>
<td>57.1ml</td>
</tr>
<tr>
<td>0.5 M EDTA (pH 8.0)</td>
<td>100ml</td>
</tr>
</tbody>
</table>

After dissolving contents top up to 1 litre with deionised water.

N.B. To prepare working solution dilute to 1 X concentration with deionised water
# Appendix C

The genetic code (RNA to Amino Acids)

<table>
<thead>
<tr>
<th>First position (5' end)</th>
<th>Second Position</th>
<th>Third position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>C</td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Tyr</td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Tyr</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Stop</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Stop</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td>Met (start)</td>
<td>Thr</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>Val (Met)</td>
<td>Ala</td>
</tr>
</tbody>
</table>