CHARACTERISATION OF BETA GLOBIN GENE CLUSTER DELETIONS USING MULTIPLEX-GAP PCR AND MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION (MLPA)

YASMIN BT MOHAMAD REDZUWAN

SCHOOL OF HEALTH SCIENCES

UNIVERSITI SAINS MALAYSIA

2018

CHARACTERISATION OF BETA GLOBIN GENE CLUSTER DELETIONS USING MULTIPLEX-GAP PCR AND MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION (MLPA)

by

YASMIN BT MOHAMAD REDZUWAN

Thesis submitted in fulfilment of the requirements for the degree of Master of Science (Biomedicine)

DECEMBER 2018

ACKNOWLEDGEMENT

In the name of Allah, the Most Generous and the Most Merciful. All praises are devoted to Allah for His guidance and peace that give me the ingredients of success.

I wish to acknowledge several people who were involved directly or indirectly in my research project.

First, I would like to thank my supervisor, Dr Zefarina bt Zulkafli for her continuous support, caring, patience, excellent guidance and supervision throughout the experimental work and during the thesis writing. Her guidance and encouragement are greatly appreciated.

My sincere thanks also go to my co-supervisor, Dr Rozieyati bt Mohd Salleh for all her kind assistance during this study, who has been a constant source of encouragement and enthusiasm, supported me greatly and was always willing to help me. I would never be able to finish my project without her help.

I would like to extend my gratitude to other staff, master and PhD students of Molecular Laboratory, Pusat Pengajian Sains Kesihatan, Kak Nisa, Kak Aini and Kak Fatin for sharing their knowledge and experience and for their valuable helps.

My special thanks also go to my bestfriend Munira bt Hashim who patiently supported me during these stressful and difficult moments.

To all my close friends, Wan Amiera, Assyuhada, Bushra and Afifah who were always support me, thank you so much.

Finally, I wish to acknowledge the greatest support, love and encouragement for my beloved parents Mr Redzuwan Ahmad and Mrs Azrina Abd Razak for their wise counsel and unconditional support throughout my life and my studies.

……

YASMIN BT MOHD REDZUWAN (PSKM0025/17)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

%	Percentage
<	Less than
≥	Equal or more than
°C	Degree Celcius
µl	Microliter
µM	Micrometer
α	Alpha
β	Beta
β-LCR	Beta-locus control region position
β-thalassaemia	Beta thalassaemia
γ	Gamma
γδβ-thalassaemia	Gamma delta beta thalassaemia
δ	Delta
δβ-thalassaemia	Delta-beta thalassaemia
ε	Epsilon
εδβ-thalassaemias	Epsilon delta beta thalassaemias
ψ	Psi
array-CGH	Array comparative genomic hybridisation
bp	Basepair
CE	Capillary electrophoresis
CISH	Chromogenic in situ hybridisation
CNV	Copy Number Variations

DNA	Deoxyribonucleic acid
DNA	Deoxyribonucleic acid
DNAse 1	Deoxyribonuclease 1
E-box	Enhancer box
EDTA	Ethylenediamine tetraacetic acid
FBC	Full blood count
FISH	Fluorescence in situ hybridisation
Hb	Haemoglobin
Hb A_2	Haemoglobin Adult
Hb C	Haemoglobin C
Hb E	Haemoglobin E
Hb F	Haemoglobin Foetal
Hb H	Haemoglobin H
Hb Lepore	Haemoglobin Lepore
Hb S	Haemoglobin S
HBA1	Alpha globin 1
HBA2	Alpha globin 2
HBAP1	Pseudo alpha globin 1
HBAP2	Pseudo alpha globin 2
HBB	Beta globin
HBBP1	Pseudo beta globin
HBD	Delta globin
HBE	Epsilon globin
HBG1	Gamma globin 1
HBG2	Gamma globin 2
HBQ1	Theta globin gene 1
HBZ1	Zeta globin 1
HBZ2	Zeta globin 2
HPFH	Hereditary haemoglobin persistence of foetal haemoglobin

HPLC	High performance liquid chromatography
HRMA	High-resolution melting analysis
HS	Hypersensitive sites
HUSM	Hospital Universiti Sains Malaysia
IVS	Intervening sequences
JEPeM	Jawatankuasa Etika Penyelidikan (Manusia)
kb	Kilobits
MARMS	Multiplex amplification refractory mutation system
MCH	Mean cell haemoglobin
MCV	Mean cell volume
MEL	Murine erythroleukaemia
M-FISH	Multiplex fluorescence in situ hybridisation
ml	Milliliter
MLPA	Multiplex ligation-dependent probe amplification
mM	Millimolar
mRNA	Messenger ribonucleic acid
NF-E2	Nuclear factor erythroid 2
ng	Nanogram
nt	Nucleotides
OR	Olfactory receptor
PCR	Polymerase chain reaction
PTT	Protein truncation test
QF-MPCR	Quantitative fluorescent multiplex polymerase chain reaction
qPCR	Real-time quantitative PCR
RNA	Ribonucleic acid
rpm	Revolutions per minute
RQ-PCR	Real-Time Quantitative PCR
SKY	Spectral karyotyping

SNP	Single-nucleotide polymorphism
STF	Stem cell factor
STS marker	Sequence tagged site marker
USF	Upstream transcription factor
UTR	Untranslated region
UV	Ultraviolet
V	Volt

ABSTRAK

Mutasi terpadam di dalam kelompok kluster gen β-globin jarang berlaku, malahan ia sukar untuk dikesan dan merupakan satu cabaran bagi makmal diagnostik. Mutasi ini kebiasaannya berkait rapat dengan kondisi delta beta talasemia (δβ-talasemia), predisposisi hemoglobin fetus berterusan (HPFH) dan varian hemoglobin. Gangguan ini biasanya ditunjukkan oleh tahap hemoglobin F (Hb F) yang tinggi, tetapi mempunyai tahap yang normal hemoglobin A_2 (Hb A_2). Walau bagaimanapun, amat sedikit kajian yang dijalankan di Malaysia yang menumpukan pada topik ini. Oleh yang demikian, kajian ini dijalankan untuk mengisi jurang pengetahuan ini. Dalam kajian ini saringan kluster gen β-globin di kalangan pesakit yang mempunyai Hb F yang tinggi (>1%) dan Hb A_2 yang normal (<4%) dijalankan dengan menggunakan kaedah *multiplex Gap-PCR* dan *multiplex ligation-dependant probe amplification (MLPA)*. Keputusan menunjukkan bahawa tiada mutasi terpadam yang dikesan kesemua 24 pesakit apabila disaring dengan *multiplex Gap-PCR* untuk empat sasaran mutasi terpadam; δβ-talasemia, HPFH-6, Siriraj I dan Hb Lepore. Walau bagaimanapun, penemuan dari pemeriksaan MLPA ke atas 12 sampel yang dipilih secara rawak mendapati seorang pesakit adalah positif dengan dua mutasi terpadam pada kluster gen β-globin. Mutasi terpadam yang ditemukan didapati berlaku di bahagian gamma globin gen 1 (*HBGI)* dan gamma globin gen 2 (*HBG2)* pada ekson ke-3. Seterusnya kajian ini juga cuba mengaitkan sebarang pertalian mutasi terpadam yang dikesani dengan ciri-ciri hematologi yang ditunjukkan oleh pesakit. Kesimpulannya, kajian ini menunjukkan kepentingan mengesan dan memperincikan mutasi terpadam melalui teknik *multiplex Gap-PCR* dan MLPA yang mana ianya dapat membantu memberikan diagnosis definitif kepada kumpulan pesakit yang terpilih ini.

Singkatan: δβ-talasemia - delta beta talasemia; HPFH - predisposisi hemoglobin fetus berterusan; Hb F - hemoglobin F; Hb A_2 - hemoglobin A_2; MLPA - multiplex ligation-dependant probe amplification; *HBG1* - gamma globin gen 1; *HBG2* – gamma globin gen 2

Kata kunci: mutasi terpadam, kluster gen β-globin, Hb F, Hb A_2, multiplex Gap-PCR, MLPA

ABSTRACT

Deletions in the beta (β)-globin gene cluster are usually rare, yet this type of mutation is problematic to detect, and subsequently possess a challenge in the diagnostic laboratory. Deletions in this cluster are usually related to the heterozygous of the delta beta thalassaemia (δβ-thalassaemia), hereditary persistence of foetal haemoglobin (HPFH) and some of the haemoglobin variants. These disorders are typically presented by elevated levels of haemoglobin F (Hb F), but with normal haemoglobin A_2 (Hb A_2). However, there is still limited number of studies focusing on this topic that has been carried out in Malaysia. As such this study was carried out to fill this knowledge gap. In this study, screening of the selected deletional mutations in the β-globin gene cluster among patients with high Hb F (>1%) and normal Hb A_2 (<4%) was performed using a multiplex Gap-PCR and multiplex ligation-dependent probe amplification (MLPA). The results showed that no deletions were detected from all 24 patients when subjected to the multiplex Gap-PCR tested against four target deletions; delta beta (δβ) thalassaemia, hereditary persistence of foetal haemaglobin 6 (HPFH-6), Siriraj I and Hb Lepore. However, findings from the MLPA screening on 12 randomly selected samples revealed one patient was positive with double deletions within the region of the β-globin gene cluster. These deletions occur at gamma-globin gene 1 (*HBG1*) and gamma-globin gene 2 (*HBG*2) in exon 3. In tandem, this study also attempted to establish any correlations of the deletions detected with the haemotological profile presented by these patients. In conclusion, this study highlighted the importance of these deletion characterisations using multiplex Gap-PCR and MLPA, which helps in establishing a definitive diagnosis among the selected group of patients.

Abbreviation: δβ-thalassaemia - delta beta thalassaemia; HPFH - hereditary persistence of foetal haemoglobin; Hb F - haemoglobin F; Hb A_2 - haemoglobin A_2; MLPA - multiplex ligation-dependant probe amplification; *HBG1* - gamma globin gene 1; *HBG2* – gamma globin gene 2

Key words: Deletion mutation, β-globin gene cluster, Hb F, Hb A_2, multiplex Gap-PCR, MLPA

CHAPTER 1

Introduction

In nature, beta (β) globin gene cluster are arranged in a single cluster (5'- ε- Gγ- A γ-δ- β-3') at chromosome 11, band 11p15.4 as shown in Figure 1.1 (Bank, 2006). Any deletions on β-gene mutations affecting the β-globin gene cluster result in reduction or absence of synthesis of one or more of these globin chains that may give rise to β-thalassaemia and its related disorders. Examples of such disorders include delta-beta thalassaemia (δβ-thalassaemia), haemoglobin Lepore (Hb Lepore), gamma delta beta thalassaemia (γδβ-thalassaemia) and hereditary haemoglobin persistence of foetal haemoglobin (HPFH) (Ahmad *et al.*, 2017; Carrocini *et al.*, 2011; Wu *et al.*, 2017).

At present, more than 300 different mutations have been described (http://globin.cse.psu.edu). Most of these mutations are primarily point mutations, single base substitutions, small insertions, or deletions within the β-globin gene complex (Colosimo *et al.*, 2002; Elizabeth and Ann, 2010). These mutations have been shown to cause β-thalassaemia, methemoglobinemia and sickle cell disease (Andrea *et al.*, 2006).

Copious data pertaining β-globin gene mutations from different ethnic populations worldwide have been well documented (Boonyawat *et al.*, 2014; Carrocini *et al.*, 2011; Huang *et al.*, 2015). In Malaysia, the distribution and the spectrum of the β-globin gene cluster mutations from different parts of the country have been systematically delineated since mid-80's (George and Khuziah, 1984) and are progressively being studied and published (George, 2001; Hassan *et al.*, 2013; Tan *et al.*, 2001).

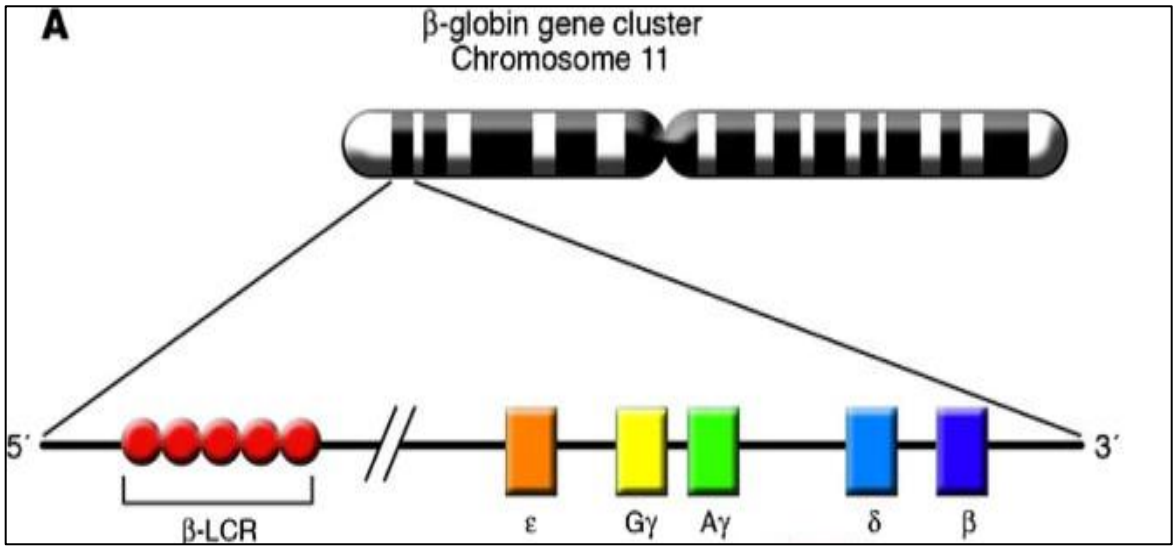

Figure 1.1: Sequence of β-globin gene cluster. Beta-globin genes are located on chromosome 11. Five functional genes: embryonic ε-gene, two foetal ^{G}γ- and ^{A}γ-genes and two adult genes called δ and β-globin gene. The order of the genes is 5' to 3' direction according to their developmental expression with respect to the beta-locus control region position (β-LCR). The β-LCR consists of five short stretches of deoxyribonucleic acid (DNA). Notes: ε – epsilon gene; ^{G}γ – gamma gene; ^{A}γ – gamma gene; δ – delta gene; β – beta gene; β-LCR – beta locus control region (Adapted from Shawky *et al*., 2012 with some modifications).

However, data on the deletional modifications related to the HPFH and δβ-thalassaemia are sparse. Only a few publications on the β-globin deletional mutations among local populations has been recorded (Hassan *et al.*, 2013; Yatim *et al.*, 2014). Given the nature of the multi-ethnic populations, the high flow of immigrants, as well as intermarriages between these groups, it is postulated that new variants of thalassaemia spectrums that are distinguishable from their origins might be observed.

Different studies have shown that the occurrence of this deletional event is not as rare as it is previously thought. With careful analysis and full molecular characterisation of their phenotypes using appropriate tools, it may help to reveal their true existence. Different

population studies have reported different types of deletions focusing on the frequency and variety of such events (Pissard *et al*., 2013 & Tritipsombut *et al*., 2012). To date, 40 different types of deletions involving the β-globin gene cluster have been described globally (Giardine *et al*., 2007 & Patrinos *et al.*, 2004).

Coinheritance of this deletion anomaly together with other β–thalassaemia mutations or haemoglobin variants can lead to mild, intermediate and severe clinical conditions (Weatherall, 2000). Moreover, in the case of other haemoglobin variants, although the heterozygous individual may not always be incapacitated by his/her conditions, but homozygous offspring inherited from these individuals may be at risk. Indeed, this is true if the deletion abnormality may involve an embryonic gene defect (Traeger-Synodinos *et al.*, 2014). As such, it is imperative to include unknown and uncharacterised deletions in the screening programme, so that thalassaemia carriers can be identified at the molecular level. Furthermore, molecular analysis of patients with thalassaemia to identify gene modifiers will improve genetic counselling and clinical management in Malaysia (Elizabeth and Ann, 2010).

For characterisation of rare and unknown β–globin gene cluster deletions, numerous techniques such as quantitative fluorescent multiplex polymerase chain reaction (QF-MPCR) (Mayuranathan *et al.*, 2014), zinc-finger nuclease (Long, 2017), high-resolution melting analysis (HRMA) (Charoenkwan *et al.*, 2017; Shih *et al.*, 2009) and real-time quantitative PCR (RT-qPCR) (Ke *et al.*, 2017) have been applied. Other techniques which are commonly described for screening of target deletion include next-generation sequencing (Wu *et al.*, 2017), Gap-PCR (Tritipsombut *et al.*, 2012) and multiplex ligation-dependent probe amplification (MLPA) (Phylipsen *et al.*, 2010; So *et al.*, 2009).

In this study, multiplex Gap-PCR and MLPA techniques were chosen to screen a selected haematology patient for the β-globin gene cluster deletions. Such mutation has been shown to be commonly detected by using PCR-based method (Cao and Galanello, 2010). Particularly for the multiplex Gap-PCR, this technique is found to be fast, simple and cost-effective in comparison to other molecular methods (Cai *et al.*, 2015; Tritipsombut *et al.*, 2012; Yatim *et al.*, 2014). The MLPA technique was used to detect the large deletion in β-thalassaemia gene clusters. The technique is based on the principle of detecting copy number variants in genomic sequences (Phylipsen *et al.*, 2010; So *et al.*, 2009). MLPA is often described as robust, quick and effective way of screening for the spectrum and frequency of β-globin gene cluster deletion mutations (Old *et al.*, 2012). It has the potential of replacing the laborious method of conventional Southern Blotting.

1.1 Significance of the study and problem statement

The main aim of this study was to detect and characterise β–globin gene cluster deletions among haematology patients at the Hospital Universiti Sains Malaysia (HUSM). The screening was specifically focused on patients who had a blood profile of haemoglobin A_2 (Hb A_2 <4%), but with an elevated level of haemoglobin F (Hb F >1%). Patient samples were subjected to molecular screening using multiplex Gap-PCR (n=24) and MLPA (n=12) for detection and characterisation of the β-globin gene cluster deletions.

It is well established that mutations, which include deletion alteration of various sizes and positions that occur within the β-globin gene cluster are responsible to certain clinical disorders. However, for some cases to achieve a final diagnosis on the β-globin gene mutation-related disorders are not always a straight cut process. This is particularly challenging for certain group of patients that display clinical parameters within the borderline area or who do not exhibit complete spectrum of the given symptoms. This often leads to misdiagnosis in these groups of patients.

Our studied cohort (patients with Hb F > 1%, Hb A_2 < 4%) is an example of such groups. Although their Hb A_2 range between low to borderline to normal levels but having an elevated Hb F at the same time hinted unclear diagnosis towards β-globin mutations trait disorders. However, definitive diagnosis can only be achieved after intensive investigations using molecular tools. Therefore, to fill this knowledge gap, this study was set to focus on the screening of deletional defect of the β-globin gene cluster in this selected group of patients using multiplex-Gap PCR and MLPA techniques.

1.2 Objective

1.1.1 General objective

- To screen for deletional defects on the β-globin gene cluster among patients with high Hb F (> 1%) and Hb A_2 (< 4%) using multiplex Gap-PCR and MLPA techniques

1.1.2 Specific Objectives

1. To optimise multiplex Gap-PCR for screening of four types of the β-globin gene cluster deletions ($^{G}\gamma(^{A}\gamma d\beta)^{0}$- thal, Thai $(\delta\beta)^{\circ}$-thal, HPFH-6, Hb Lepore)
2. To screen samples using the MLPA technique as a second-line of screening
3. To identify, characterise and compile known, unknown or rare types of deletion observed in the study.

CHAPTER 2

Literature Review

2.1 Haemoglobin: structure and genetic control

2.1.1 Structure

Haemoglobin (Hb) is a tetrameric protein with a molecular weight of 64 500 Dalton consisting of two pairs of identical alpha (α) and beta (β) subunits. It functions is to transport oxygen (O_2) from the lungs to the tissues (Figure 2.1) (Lukin and Ho, 2004).

During early foetal development, synthesis of Hb occurs initially in the yolk sac and subsequently in the liver. As the bone marrow develops, it progressively takes over this function (Sankaran *et al.*, 2010).

Newly synthesized globin chains will naturally fold into α-helical structures. This fold forms a pocket that encloses and binds a haem prosthetic group that consists of a heterocyclic porphyrin ring with an iron ion metal in the center (Fe^{2+}) (Matta-Camacho *et al.*, 2014). When the oxygen binds to haem iron, the atom adopts a ferrous (Fe^{3+}) oxidative state. It will form an oxy-ferrous intermediate, which causes the activation of oxygen (Huang and Groves, 2018).

Hb will bind to oxygen molecules when exposed to a high oxygen partial pressure (pO_2) and progressively release oxygen as the pO_2 falls. This allows Hb to perform its vital function of transporting oxygen to respiring tissues throughout the body (Brittain, 2002).

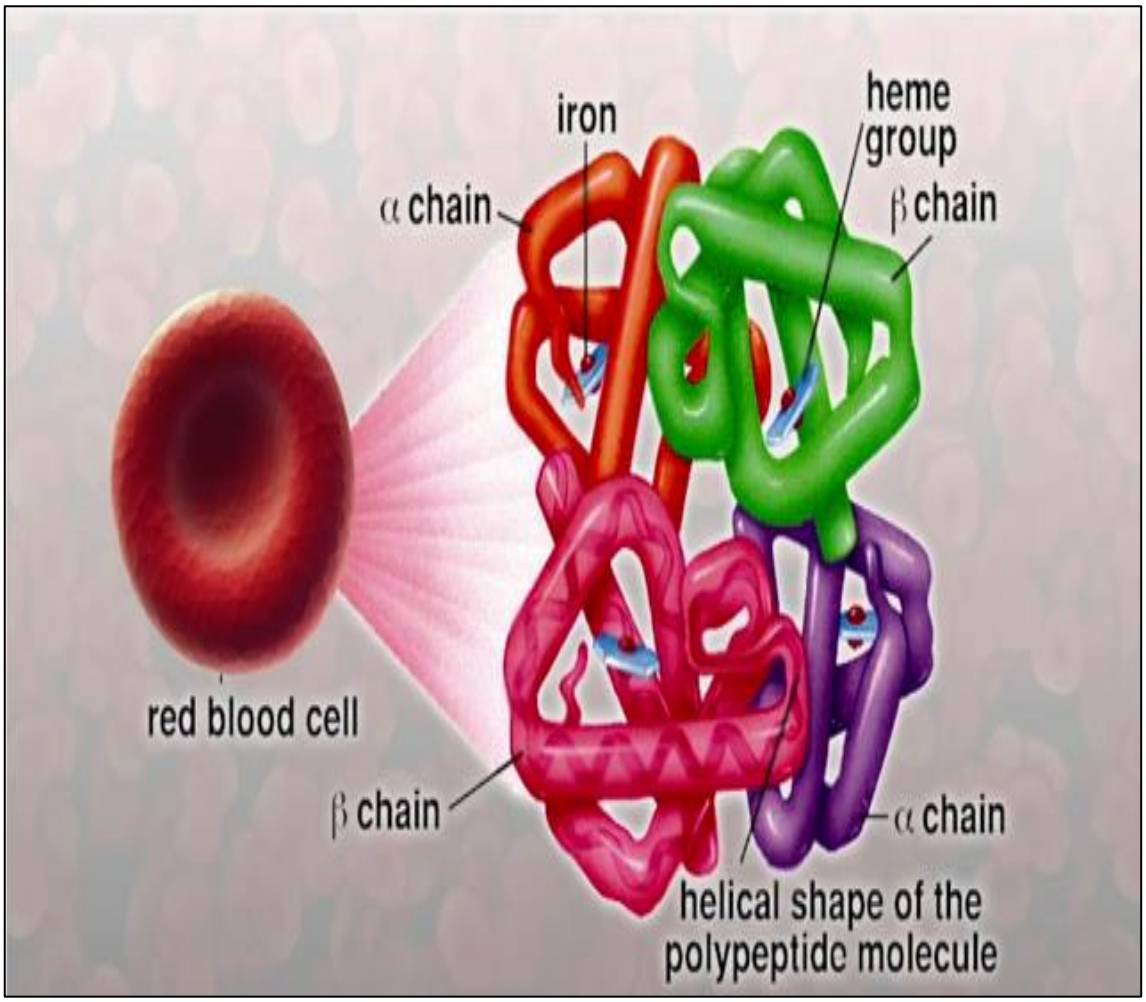

Figure 2.1: Haemoglobin structure. Haemoglobin is a tetrameric globular protein, composed of two pairs of identical α and β subunits. The helical structure of globin chains was shown forming haem pocket upon folding in each of the polypeptide chain of haemoglobin. Notes: α -alpha; β – beta (Adapted from Sylvia, 1997).

2.1.2 Haemoglobin Switching

The reason for multiple globin genes exist is that their different chemical properties are necessary at different stages of development (Table 2.1). The α and β-globin gene loci are activated approximately three weeks after conception. In early embryos, four globin genes are being activated, producing four different forms of Hb. Three of these are considered embryonic Hb which are Hb Portland (ζ2γ2), Hb Gower I (ζ2ε2) and Hb Gower II (α2ε2) and Hb F (α2γ2) (Weatherall, 2001).

Table 2.1: Haemoglobin development in human

Classification	Name	Genes	Proteins	Significance
Embryonic haemoglobins	Haemoglobin Gower I	*HBZ; HBE1*	2 ζ-globin chains 2 ε-globin chains ($\zeta_2\varepsilon_2$)	Early embryonic haemoglobin produced primarily in the yolk sac
	Haemoglobin Gower II	*HBA1* or *HBA2*; *HBE1*	2 α-globin chains 2 ε-globin chains ($\alpha_2\varepsilon_2$)	Embryonic haemoglobin produced in the liver. Not present after the 18th week of gestation
	Haemoglobin Portland	*HBZ; HBG1* or *HBG2*	2 ζ-globin chains 2 γ-globin chains ($\zeta_2\gamma_2$)	Embryonic haemoglobin produced in the liver. Not present after the 18th week of gestation
Foetal Haemoglobin	Foetal haemoglobin (Hb F)	*HBA1* or *HBA2*; *HBG1* or *HBG2*	2 α-globin chains 2 γ-globin chains ($\alpha_2\gamma_2$)	Hb F typically accounts for approximately 70% of the total haemoglobin at birth. Within the first year of life levels

Table 2.1 (continue)

Classification	Name	Genes	Proteins	Significance
				normally fall to <1%, but Hb F continues to be present in trace amounts.
Adult Haemoglobins	Haemoglobin A (Hb A)	*HBA1* or *HBA2*; *HBB*	2 α-globin chains 2 β-globin chains ($\alpha_2\beta_2$)	Hb A typically accounts for >95% of the total haemoglobin in healthy adults
	Haemoglobin A_2 (Hb A_2)	*HBA1* or *HBA2*; *HBD*	2 α-globin chains 2 δ-globin chains	Hb A_2 accounts for <3.3% of the total haemoglobin in normal adults. Hb A_2 has no clinical significance but is raised in β-thalassaemia trait, making it a useful marker for the disease

Note: α-globin - *HBA1*; ζ-globin – *HBZ*; β-globin – *HBB*; δ-globin – *HBD*; γ-globin 1 - *HBG1*; γ-globin 2 - *HBG2*; ε-globin – *HBE*; ζ – zeta; ε – epsilon; γ – gamma; δ – delta; β – beta; α – alpha. Table adapted from Sankaran and Nathan, 2010; Weatherall, 2001a.

2.1.2.1 Haemoglobin F (Hb F)

Hb F is produced at the early stage of pregnancy and is a dominant Hb expressed in the foetus development. It becomes the most abundantly produced form of Hb from approximately eight weeks post-conception until six weeks after birth. At this stage, Hb F is gradually replaced by Hb A which is mediated by a shift from γ-globin to adult β-globin production at the β-globin gene locus. Hb F continues to be produced at low levels (<1%) throughout adult life and a small amount of Hb A_2 is also produced (Sankaran and Nathan, 2010; Weatherall, 2001a).

Hb F levels can be used as one of the hallmarks to the many syndromes such as sickle-cell disease, δβ-thalassaemia and HPFH (Table 2.2).

Moreover, F-cells also contribute to the cause of certain syndromes because the survival of F-cells can increase the concentration of foetal haemoglobin in patients with the sickle-cell disease than in non-affected adult individuals. Even though the switch from γ-globin to β-globin happens shortly after birth, but to some individuals with a clinical disorder such as sickle-cell disease, β-thalassaemia, and HPFH, the patient continues to produce measurable amounts of foetal haemoglobin in adulthood (Garner *et al.*, 2000).

Size of the β-globin deletion is also being associated with the concentration of foetal haemoglobin in HPFH. Large deletions within the β-globin locus are associated with the higher foetal haemoglobin concentrations, while smaller deletions correlate with the modest increases in foetal haemoglobin concentration which can be seen in (δβ°)-thalassaemia (Bank, 2006).

Table 2.2: Conditions related to elevated Hb F

Condition	Percentage of Hb F (%)
β-thalassaemia trait	<10
δβ-thalassaemia trait	5-20
Deletional HPFH trait	20-40
Nondeletional HPFH trait	3-30
Homozygous δβ thalassaemia	100
Homozygous deletional HPFH	30-95
Haemoglobin S-HPFH	20-40
Haemoglobin E-β° thalassaemia	10-50
Haemoglobin C-β° thalassaemia	10-30

Notes; β-thalassaemia – beta thalassaemia; δβ-thalassaemia – delta beta thalassaemia; HPFH – hereditary persistence foetal of haemoglobin. Table adapted from Hoyer et al., 2002.

2.1.2.2 Haemoglobin A_2 (Hb A_2)

In addition, the determination of Hb A_2 also plays an important tool to diagnose the β-thalassaemia trait such as δβ-thalassaemia and HPFH. Moreover, this can lead to some individuals being missed during the screening programmes (Giambona *et al.*, 2009), since often there are samples that present with haematology parameters that are not consistent with atypical evaluation of β-thalassaemia trait. Therefore, in this cases interpretation plays a key role in screening programmes for thalassaemia.

2.2 General features of globin genes

2.2.1 Chromosomal organisation of the globin genes

The genes that encode for the globin proteins (or globin chains) are placed on different chromosomes which are put together in clusters. The β-globin gene cluster is located on chromosome 11, while the α-globin cluster is situated on chromosome 16 (Figure 2.2). In both genes, the globin cluster is arranged in the order in which they are expressed during the developmental stages (Johnson *et al.*, 2002).

Figure 2.2 shows that the α-globin chains are divided into two α globins (α1 and α2) and zeta (ζ), which will substitute for α in the early development. It consists of 141 amino acids. In contrast, the β-globin chains consist of epsilon (ε), gamma (γ), delta (δ) and β and with 146 amino acids.

In normal adults, about 97% of the protein molecules are Hb A (α2β2), 2% for Hb A_2 (α2δ2) and the remaining 1% is for Hb F or foetal Hb (α2γ2). In the foetus, Hb expressed are Hb Gower1, Gower 2 and Portland (Schechter, 2008).

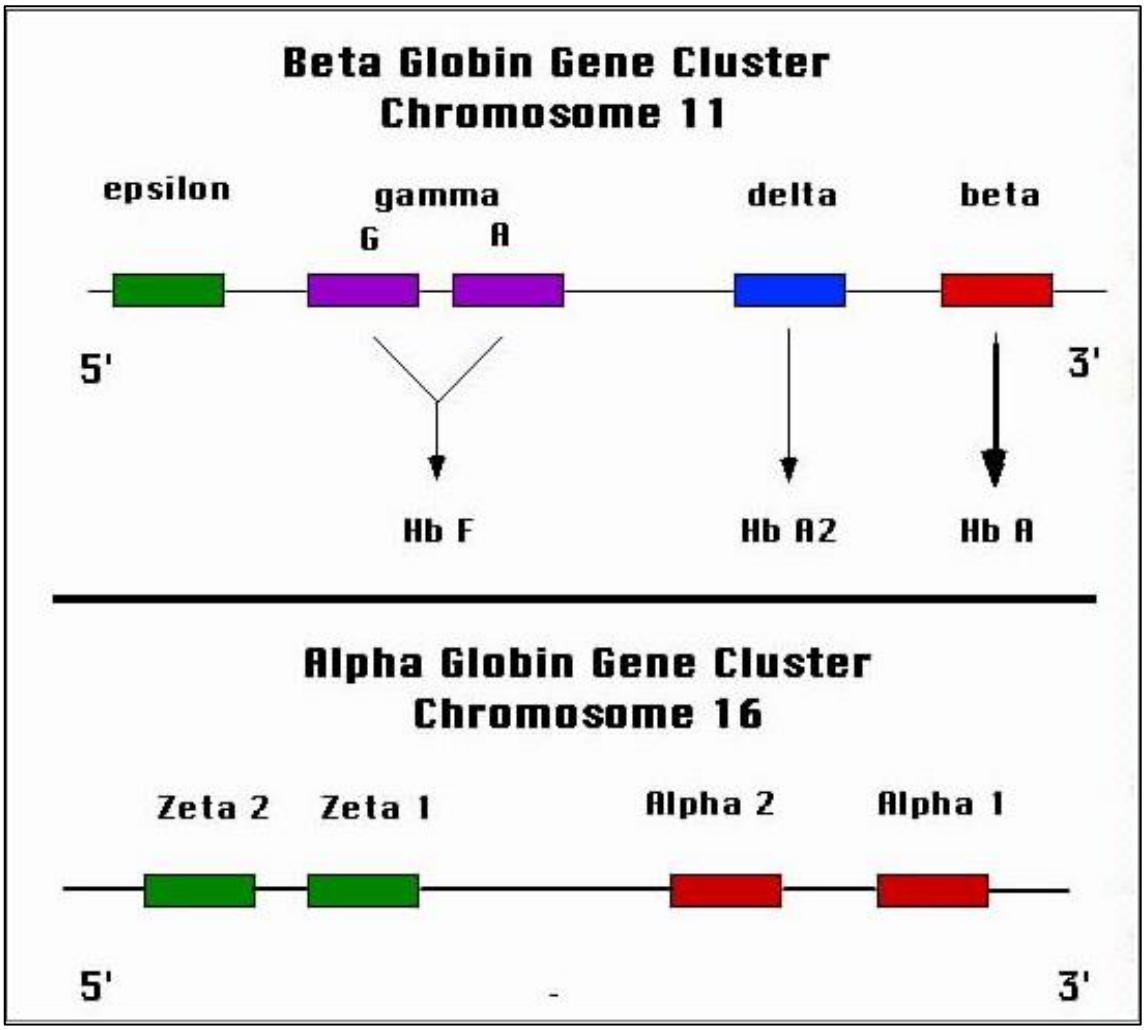

Figure 2.2: Schematic representation of the β-globin gene cluster on chromosome 11 and α-globin gene on chromosome 16. During foetal growth, the two gamma-globin genes are active and produced Hb F. The delta gene, produces in a small amount in children and adults is called Hb A_2 and with a combination of two alpha genes and two beta genes, produces a normal Hb A (Adapted from http://sickle.bwh.harvard.edu, n.d).

2.3 The β-globin genes

2.3.1 Structure of the β-globin genes

The β-globin gene encodes for 146 amino acids with 1600 basepair (bp) size. It contains five functional genes, β-globin (*HBB*), δ-globin (*HBD*), γ-globin 1 (*HBG1*), γ-globin 2 (*HBG2*), ε-globin (*HBE*) and one non-functional gene, pseudo beta globin 1 (*HBBP1*). This genes are arranged as follows: 5' to 3' as 5'-ε (*HBE*)-^{G}γ (*HBG2*)-^{A}γ(*HBG1*)-ψβ (*HBBP1*)-δ (*HBD*)-β(*HBB*)-3' in sequence from embryonic to adult genes (Figure 2.3) (Bank, 2006; Bank *et al.*, 1980). In the β-globin gene cluster, upstream of *HBE*, there is a β-locus control region (β-LCR) which has the interaction with the β-locus structural genes (Bank, 2006; Tuan *et al.*, 1989).

The β-globin LCR is required for the globin gene transcription and is the major structural component of it is murine and human β-globin loci. It contains 5 critical DNAse 1 hypersensitive (HS) sites, HS1-5. These HS1-5 are formed in regions, which are more accessible than other regions for the interactions with the transcription factors and downstream gene sequences (Figure 2.3) (Bulger *et al.*, 2003; Tolhuis *et al.*, 2002). Another HS site, 3' HS-1, 5'HS, and olfactory receptor (OR) sequences are at both located the 5' and 3' borders of the β-globin locus (Bulger *et al.*, 2000; Bulger *et al.*, 2003).

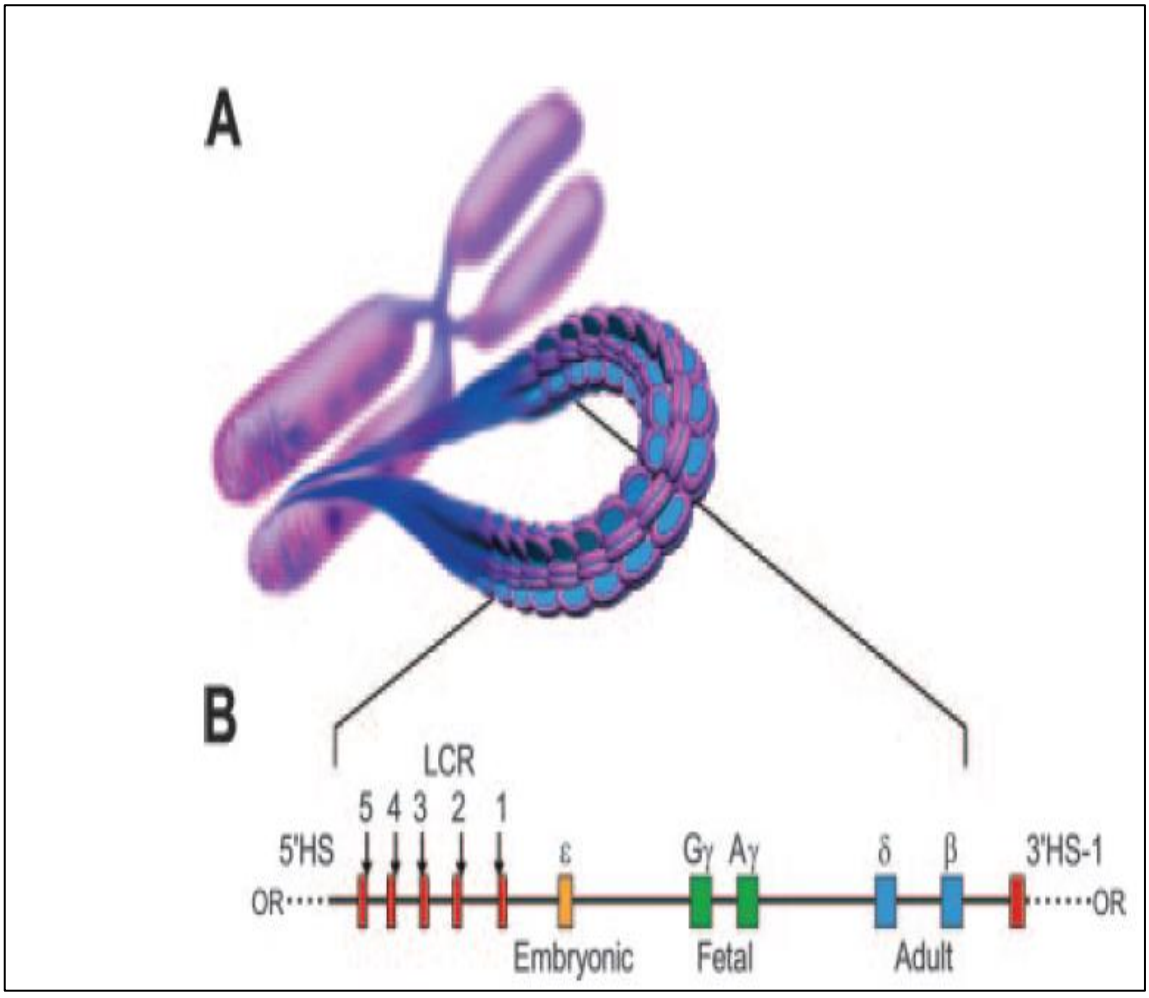

Figure 2.3: The human β-locus. (A) The β-globin locus on chromosome 11 embedded in chromatin. (B) A linear map with the globin LCR and its hypersensitive sites is shown by the vertical arrows. The structural ε,γ,δ and β-globin genes and the locations of the olfactory receptor genes are shown. Notes: OR – olfactory receptor; HS – hypersensitive site; LCR – locus control region; ε – epsilon; $^{G}\gamma$ – gamma; $^{A}\gamma$ – gamma; δ – delta; β – beta (Adapted from Bank, 2006).

The β-globin gene cluster is split into three exons (coding regions) by two non-coding intervening sequences (introns or IVS). IVS1 interferes with the coding sequence inside codon 30, and IVS2 is situated between codons 104 and 105. Exon 2 encodes the residues associated with the haem binding and αβ- dimer development, while exons 1 and 3 encode for the non-haem binding regions of the globin chain. Exon 3 also contains many amino acids that are associated with globin subunit interactions which required for the Bohr effect and 2,3-diphosphoglycerate (2,3-DPG) binding. Importantly, the conserved sequences which are critical for the gene functions are located in the 5' promoter region, at the exon-intron junctions and in the 3' untranslated region (3'UTR) at the end of the messenger ribonucleic acid (mRNA) sequences (Ho and Thein, 2000).

The promoter is a crucial region for ensuring the efficient transcription of the gene. As shown in Figure 2.4, the promoter consists of several conserved sequences that are necessary for its effective functions. Include in this region are the sequence ATAA (or TATA-box), located about -30 nucleotides from the cap, and the sequence CCAAT (or CAT-box), located between -70 and -80 bp. Additionally, the sequence CACACCC, which is duplicated in a proximal and distal sequence can also be found within this region, which has been shown to employ equal influence on the transcription process (Ghosh *et al.*, 2014).

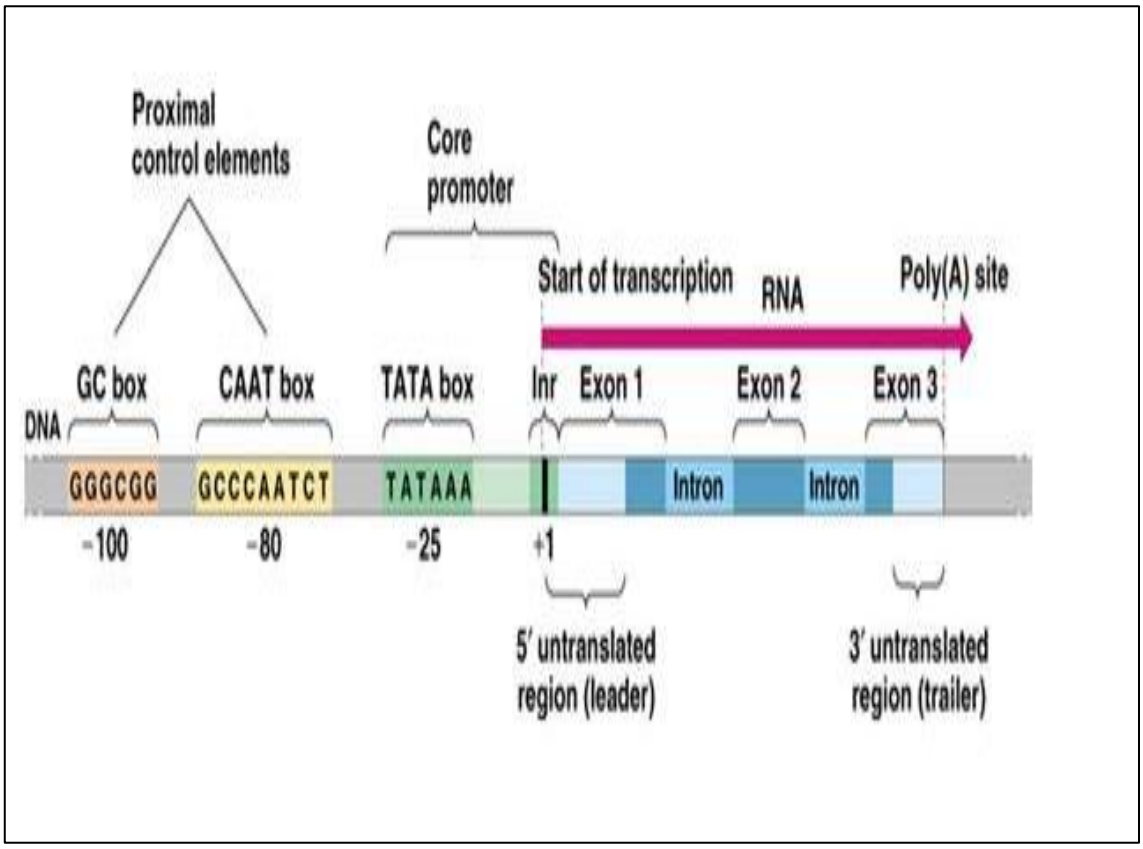

Figure 2.4: General structure of the beta-globin gene. The exons (black box) is in between introns (IVS1 and IVS2) and at the left side is the promoter region which consist of TATA box, CAT box and CACACCC boxes. Abbreviation: DNA - deoxyribonucleic acid; Inr – initiator element; RNA - ribonucleic acid (Adapted from Ghosh et al., 2014).

The 5’ untranslated region (5’ UTR) resides in a region of 50 nucleotides between 5’ terminus or ‘cap’ site of the β-globin mRNA and the initiation (ATG) codon. There are two prominently conserved sequences in the 5’UTR of the various globin genes (both α and β) which are the CTTCTG hexanucleotide. This sequence is found at 8 through 13 nucleotides downstream from the cap site and the CACCATG, in which the last three nucleotides form the initiation codon. The 3’UTR is a part of the region between the termination codon and the poly(A) tail. This region is 132 nucleotides long with one conserved sequence, TATAAA, situated 20 nucleotides upstream of the poly(A) tail. It serves as a signal for cleavage of the 3’ end of the main transcript. An addition of the poly(A) tail confers reliability on the processed mRNA, as well as enhances the translation process (Ho and Thein, 2000).

2.3.2 Alteration of β-Globin Gene Cluster

Majority cases of the β-globin alterations are due to point mutations. To date, there are numerous mutations affecting the β-globin gene cluster with the majority of cases being single nucleotide substitutions, insertions, or short deletions (Thein, 2013). Despite of the massive mutations associated with the β-globin gene which contribute to its variable phenotypes, fortunately only a few common mutations and a number of a rare variants occur in each population (Steinberg *et al.*, 2009; Weatherall, 2001b). Other type of mutations that may affect the β-globin gene are deletions, defects of transcription, defects of mRNA maturation and defects of mRNA translation (Table 2.3).

In the last 15 years, the spectrum of the β-globin mutations have been well documented in the main racial groups of Malays and Chinese in Peninsular Malaysia (Tan *et al.*, 2004). However, the occurrence of β-thalassemia disorders are found to be rare in Indian ethnic (Elizabeth and Ann, 2010). Studies have shown that 4.5% of Malaysian are heterozygous carriers for β-thalassaemia (George, 2001; George *et al.*, 2012; Tan *et al.*, 2004).

The common β-globin defects that account for 90% of the mutations in Chinese-Malaysians are CD 41/42 (-TCTT), IVS II-654 (C-T), -28 (A-G), CD 17(A-T) and CD 71/72 (+A) (George *et al.*, 1992). While, most common mutation among Malays are $\beta^{E(26Glu\text{-}Lys)}$) (HbE) (George *et al.*, 1992; Tan *et al.*, 2004; Yatim *et al.*, 2014). Despite this the complete spectrum of mutations among the Malaysian ethnics are still largely unknown. Since Malaysia is a country with a multiethnic populations, the identification of the β-globin gene cluster mutation spectrum in Malaysia is more complex (George, 2001).

Most of the examples given earlier involved mutation of nucleotide substitutions. However, in this study, the focus in on the deletional mutations. This will be reviewed in section 2.3.5.

Table 2.3: Alterations in the β-globin gene cluster and its impact

Class of mutation	Mechanism	Severity
Transcriptional mutations	Mutations in the promoter alter the binding sites for transcriptional factors.	β^{+} or β^{++}
	Mutations of the 5' untranslated region result in mild β-thalassaemia alleles. The mechanism is uncertain but may be due to reduced capping efficiency	β^{++}
RNA processing mutations	Mutations affecting the consensus splice sequence are typically less severe than those affecting the invariant nucleotides of the splice junction. However, it can still greatly reduce the amount of functional mRNA produced.	β^{+}
	Where a mutation results in the formation of a cryptic splice site both normal and aberrant mRNA is produced.	β^{+} or β^{++}
	Mutations changing the AATAAA sequence in the 3' UTR inhibit the polyadenylation process.	β^{+} or β^{++}

Class of mutation	Mechanism	Severity
RNA translation mutations	Mutations removed the initiation codon and prevent translation.	β°
	Mutations that result in a nonsense codon lead to the formation of a transcript that cannot be translated.	β°
	Mutations causing a reading frame shift result in mRNA that cannot be translated into viable globin.	β°

Note: β^{+} - alleles with some residual of β-globin production; β^{++} - the reduction in β-globin production is very mild; β° - complete absence of production of β-globin affected allele. Table are adapted from Thein, 2005; Thein and Wood, 2009.

2.3.3 The β-globin gene cluster deletion

Entries in the database of human haemoglobin variants and thalassemia mutations (HbVar) has shown that at present there are at least 228 entries associated with the deletion mutation (http://globin.bx.psu.edu/cgi-bin/hbvar/query_vars3).

Besides that, according to Harteveld *et al.* (2005), deletions affecting the β-globin gene cluster results in inactivation of the structural genes. This leads to the syndromes of β-thalassaemia, HPFH, δβ-thalassaemia, Hb Lepore and γδβ-thalassaemia (Figure 2.5). For example, HPFH, as seen in black's population, is associated with the deletions of both the β and δ-globin genes. For the δβ-thalassaemia, the β-gene is entirely removed, whereas most of the δ-gene is missing. This deletion caused the expression of both γ-globin genes at high level, to compensate for the absent of β-production (Orkin and Kazazian Jr, 1984).

Based on Figure 2.5, deletions that remove all or part of the β-globin gene are classified as β-thalassaemia deletions. Meanwhile, deletions that remove the δ and β-globin genes are termed δβ-thalassaemia and this is often associated with moderate increases in Hb F. Meanwhile, HPFH deletions are larger and remove one γ-globin gene plus the δ and β globin genes. These deletions result in elevated levels of Hb F, which is significantly higher than that observed in β or δβ-thalassaemia (Galanello, 2013). HPFH and δβ-thalassaemia are characterised by the persistent expression of γ-globin genes in adults, and also presence of hypochromic and microcytic erythrocytes (Forget, 1998; Weatherall, 2001a).