CONTRIBUTION OF HBS1L GENE ON Hb F LEVEL IN β -THALASSEMIA / Hb E

RIYAZ AHMAD PANDIT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(MOLECULAR GENETICS AND GENETIC ENGINEERING)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2006

ISBN 974-04-7386-5

Thesis
Entitled

CONTRIBUTION OF HBS1L GENE ON Hb F LEVEL IN β -THALASSEMIA / Hb E

Riyaz Ahmad Pandit
.....
Mr. Riyaz Ahmad Pandit
Candidate

.....
Asst. Prof. Kanokporn Triwitayakorn, Ph.D.
Major-Advisor

Chayanon Peerapittayamongkol
.....
Lect. Chayanon Peerapittayamongkol, M.D, Ph.D.
Co-Advisor

Pranee Fucharoen
.....
Lect. Pranee Fucharoen, Ph.D.
Co-advisor

Saovaros Svasti
.....
M. L. Saovaros Svasti, Ph.D.
Co-advisor

.....
Prof. M. R. Jisnuson Svasti, Ph.D.
Dean
Faculty of Graduate Studies
Mahidol University

V. Akkarapatumwong
.....
Asst. Prof. Varaporn Akkarapatumwong, Ph.D.
Chair
Master of Science Programme in
Molecular Genetics and Genetic Engineering
Institute of Molecular Biology and Genetics

Thesis
Entitled

CONTRIBUTION OF HBS1L GENE ON Hb F LEVEL IN β -THALASSEMIA / Hb E

was submitted to the Faculty of Graduate Studies, Mahidol University
For the degree of Master of Science (Molecular Genetics and Genetic Engineering)
on
May 23, 2006

Mr. Riyaz Ahmad Pandit
Candidate

Asst. Prof. Kanokporn Triwitayakorn, Ph.D.
Chair

Lect. Chayanon Peerapittayamongkol, MD. Ph.D.
Member

Lect. Pranee Fucharoen , Ph.D.
Member

Prof. Suthat Fuchroen, MD.
Member

M. L. Saovaros Svasti, Ph.D.
Member

Prof. M. R. Jisnuson Svasti, Ph.D.
Dean
Faculty of Graduate Studies
Mahidol University

Asst. Prof. Chartchai Krittanai, Ph.D.
Acting Director
Institute of Molecular Biology and Genetics
Mahidol University

ACKNOWLEDGEMENTS

This work was carried out in Thalassemia Research Center (TRC), Institute of Science and Technology for Research and Development, Mahidol University, Thailand.

I am extremely thankful to Prof. Suthat Fucharoen for helping me financially in part, for his valuable advice and suggestion in data interpretation and providing excellent research facilities.

I would like to thanks my advisor Dr. Kanokporn Triwitayakorn for providing me an opportunity to work on this thesis and my co-advisor Dr. Chayanon Peerapittayamongkol for helping me in statistical part of this work.

The success of this thesis is attributed to the extensive support and assistance from Dr. Pranee Fucharoen and Dr. M. L. Saovaros Savasti. I would like to express my deepest gratitude to them for invaluable guidance, benevolence, support and good consultation, enabling me to achieve this thesis.

I would like to thank National Center of Genetic Engineering and Biotechnology (BIOTEC), and Thailand Research Fund for project support.

This thesis was supported in part by the Thesis Grant, Faculty of Graduate Studies, Mahidol University.

I tender my thanks to Miss Thongperm Munkongdee and all TRC colleagues for being kind, helpful and co-operative.

I am grateful to Miss Petcharin Pareong for her help during sequencing process and Mr Weeradej Khonsuntia for his help in polyacrylamide gel electrophoresis and silver staining.

I am deeply thankful to all teachers of Institute of Molecular Biology and Genetics specially to Dr. Apinunt udomkit, Dr. Chanan Angsuthanasombat, Dr. Varaporn Akkarapatumwong for their knowledge, kindness, help and care throughout my studying period.

My special thanks go to all of my friends, staff, and technicians of the Institute of Molecular Biology, Genetics, and Thalassemia Research Center for their generosity and understanding.

Finally, I would like to express my deepest appreciation to my father, mother, brother and elder sister for their eternal love, support and encouragement throughout my life.

Riyaz Ahmad Pandit

CONTRIBUTION OF HBS1L GENE ON Hb F LEVEL IN β-THALASSEMIA/Hb E

RIYAZ AHMAD PANDIT 4736002 MBMG/M

M.Sc. (MOLECULAR GENETICS AND GENETIC ENGINEERING)

THESIS ADVISORS: KANOKPORN TRIWITAYAKORN, Ph.D., CHAYANON PEERAPITTAYAMONGKOL, Ph.D., MD. PRANEE FUCHROEN, Ph.D., M. L. SAOVAROS SVASTI, Ph.D.

ABSTRACT

β-thalassemia is a monogenic disease that occurs from the defective β-globin chain synthesis. Excessive unpaired α-globin chains precipitate in erythroid precursor results in premature cell death. An increase in fetal hemoglobin (HbF) reduces the globin chain imbalance, consequently improving thalassemic symptoms. Genome-wide SNP search revealed 5 significant SNPs in *HBS1L* gene associated with severity of β-thalassemia / Hb E patients. Previous quantitative trait locus study showed that *HBS1L* gene and the nearby genes may be involved with the high Hb F production.

This work aims to identify the SNPs in *HBS1L* gene located on the chromosome 6q23 that may be responsible for the elevated Hb F in β thalassemia. Direct sequencing of all exons including exon-intron junctions of 16 mild and 14 severe β-thalassemia/Hb E cases was performed. A Total of 22 SNPs and a 5 bp deletion near the 4th intron-exon junction were discovered. Haplotype analysis showed that SNPs 3 to 12 were in linkage disequilibrium. Because of the presence of *ApaL*1 restriction site, SNP 7 was selected for genotyping in 295 mild and 180 severe cases using PCR-RFLP method. The results revealed that C/C genotype of SNP 7 in mild cases was higher than that in severe cases and the allele frequency between mild and severe cases was significantly different ($p = 0.002$). Although *Xmn*I polymorphism on the position of -158 $^{G}\gamma$-globin gene seems to have a stronger effect on Hb F level than SNP 7 polymorphism of *HBS1L* gene, the latter has been shown to have a modulating effect on Hb F level when *Xmn*1 genotype is -/-.

KEY WORDS: - β-THALASSEMIA/Hb E / Hb F / HBS1L GENE / SNP.

107 pp. ISBN 974-04-7386-5

ผลของยีน HBS1L ต่อปริมาณฮีโมโกลบินเอฟในผู้ป่วยเบต้าธาลัสซีเมียฮีโมโกลบินอี
(CONTRIBUTION OF HBS1L ON Hb F LEVEL IN β-THALASSEMIA/Hb E)

RIYAZ AHMAD PANDIT 4736002 MBMG/M

วท.ม. (อณูพันธุศาสตร์และพันธุวิศวกรรมศาสตร์)

คณะกรรมการควบคุมวิทยานิพนธ์: กนกพร ไตรวิทยากร, Ph.D., ชยานนท์ พีระพิทยมงคล, Ph.D, M.D., ปราณี ฟู่เจริญ, Ph.D., ม.ล. เสาวรส สวัสดิวัฒน์, Ph.D.

บทคัดย่อ

โรคเบต้าธาลัสซีเมียเป็นโรคทางพันธุกรรมที่เกิดจากความผิดปกติของการสังเคราะห์สายเบต้าโกลบิน สายอัลฟ่าโกลบินที่เกินจะตกตะกอนในเม็ดเลือดแดงตัวอ่อนทำให้เกิดการตายของเซลล์ การที่ผู้ป่วยธาลัสซีเมียมีปริมาณฮีโมโกลบินเอ๊ฟสูงขึ้นจะช่วยลดความไม่สมดุลของสายโปรตีนโกลบิน ยังผลให้อาการรุนแรงที่พบในผู้ป่วยลดลง จากการศึกษา SNP ในจีโนมพบว่า มี SNP 5 ตำแหน่งในยีน HBS1L ที่มีความสัมพันธ์กับความรุนแรงของโรคในผู้ป่วยเบต้าธาลัสซีเมียฮีโมโกลบินอี การวิเคราะห์ quantitative trait พบว่า ยีน HBS1L และยีนบริเวณใกล้เคียงอาจมีผลต่อปริมาณฮีโมโกลบินเอ๊ฟ

การศึกษานี้ต้องการหา SNP ในยีน HBS1L บนโครโมโซม 6q23 ที่มีผลต่อการเพิ่มปริมาณฮีโมโกลบินเอ๊ฟ ในผู้ป่วยเบต้าธาลัสซีเมียฮีโมโกลบินอี โดยการตรวจหาลำดับเบสของ DNA ใน exon รวมทั้งบริเวณ intron-exon junction ของทุก exon บนยีน HBS1L ในผู้ป่วยเบต้าธาลัสซีเมียฮีโมโกลบินอีที่มีอาการน้อย 16 คน และมีอาการรุนแรง 14 คน พบ SNP 22 ตำแหน่ง และ พบนิวคลีโอไทด์ขาดหายไป 5 ตัว (5 bp deletion) ในบริเวณ intron-exon junction ลำดับที่ 4 เมื่อทำการวิเคราะห์ haplotype พบว่า SNP ที่ 3 ถึง 12 มี linkage disequilibrium อยู่ เนื่องจาก SNP ตำแหน่งที่ 7 มีจุดตัดของเอ็นไซม์ตัดจำเพาะ *Apa*LI จึงทำการตรวจ genotype ของ SNP 7 โดยวิธี PCR-RFLP ในผู้ป่วยที่มีอาการน้อย 295 คน และอาการรุนแรง 180 คน พบว่าความถี่ของ genotype ชนิด C/C ในผู้ป่วยที่มีอาการน้อยจะสูงกว่าในผู้ป่วยที่มีอาการรุนแรงอย่างมีนัยสำคัญทางสถิติ (p=0.002) polymorphism ที่ตำแหน่ง -158 $^{G}\gamma$ *Xmn*I ซึ่งมีรายงานว่ามีผลต่อปริมาณฮีโมโกลบินเอ๊ฟ และ SNP 7 จะเป็นอิสระต่อกันในการควบคุมปริมาณฮีโมโกลบินเอ๊ฟ แม้ว่า *Xmn*I polymorphism จะมีอิทธิพลมากกว่า แต่ polymorphism ของ SNP 7 จะมีผลต่อปริมาณฮีโมโกลบินเอ๊ฟ เมื่อผู้ป่วยมี *Xmn*I genotype เป็น -/-

107 หน้า ISBN 974-04-7386-5

CONTENTS

CONTENTS (Cont'd)

Page

CONTENTS (Cont'd)

Page

CONTENTS (Cont'd)

Page

LIST OF TABLES

LIST OF TABLES (Cont'd)

LIST OF FIGURES

LIST OF FIGURES (Cont'd)

LIST OF FIGURES (Cont'd)

LIST OF ABBREVIATIONS

%	=	per cent
^{o}C	=	degree Celsius
μg	=	microgram(s)
μM	=	micromolar
μl	=	microliter(s)
aa	=	amino acid(s)
Abs F	=	absolute fetal hemoglobin
bp	=	base pair
cM	=	centiMorgans
cm	=	centimeter (s)
Cor% F	=	correct fetal hemoglobin.
dATP	=	deoxyadenosine 5'-triphosphate
dCTP	=	deoxycytidine 5'-triphosphate
dGTP	=	deoxyguanosine 5'-triphosphate
DNA	=	deoxyribonucleic acid
dNTPs	=	dATP, dCTP, dGTP and dTTP
dTTP	=	deoxythymidine 5'-triphosphate
e.g.	=	*exempli gratia* (for example)
FC	=	fetal cells
g	=	gram(s)
g/dl	=	gram per deci liter
HS	=	hypersensitive site
IVS	=	intervening space
kb	=	kilobase (s)
kg	=	kilogram (s)
LCR	=	locus controlling region
LD	=	linkage disequilibrium

LIST OF ABBREVIATIONS (cont.)

M	=	molar
mA	=	miliampere(s)
mg	=	milligram(s)
min	=	minute(s)
ml	=	milliliter(s)
mM	=	millimolar
mm	=	millimeter(s)
mmHg	=	millimeter(s) of mercury
$MgCl_2$	=	magnesium chloride
ng	=	nanogram(s)
OD	=	optical density
PCR	=	polymerase chain reaction
pM	=	picomolar
QTL	=	quantitative trait
rpm	=	revolutions per minute
RFLP	=	restriction fragment length polymorphism
mRNA	=	messenger ribonucleic acid
sec	=	second(s)
SNP	=	Single nucleotide polymorphism
TBE	=	tris-borate EDTA
U	=	unit(s)
V	=	volt(s)
W	=	watt(s)

CHAPTER 1
INTRODUCTION

Thalassemia is a monogenic disease, occurs due to the synthesis of globins from the α- and β- globin gene clusters. These globin chains assemble to produce different subtypes of hemoglobins in response to the changing oxygen requirements during human development [1]. In severe β-thalassemia, the accumulation of excess unpaired α hemoglobin chain forms insoluble membrane damaging precipitates in the red cell precursor and causes premature cell death. [2]. Furthermore, deletions of the non-alpha globin gene cluster encompassing the β- and δ-globin genes are associated with a genetic variant of β-thalassemia, which is characterized by a marked increase in the synthesis of Hb F. In these patients the γ-globin genes remain intact and are expressed in the postnatal life, thus reducing the ineffective erythropoiesis and ameliorating the severity and symptoms of β-thalassemia [3].

Substantial efforts have been invested in studying mechanism of switching of human γ-globin genes and mutations that alter the pattern of their expression in adult life culminating in the condition known as hereditary persistent fetal hemoglobin (HPFH). In twin study [4], the proportion of F-cell variation due to all genetic factors was estimated to be 89%. Among genetic factors, *Xmn*I -$^{G}\gamma$ polymorphism is accounting for 19% of total variance and the remaining due to other genetic loci that contribute to continuous phenotype (quantitative trait loci, QTLs). Linkage studies in Indian Kindred have identified two more regions, 6q23 and 8q [5, 6]. This study also suggests an interaction between *Xmn*I -$^{G}\gamma$ site and chromosome 8q in influencing the HbF levels. Detailed annotation of 6q23 region encompassing 1.5 Mb has revealed four known genes (*MYB, ALDH8AI, HBS1L* and *PDE7B*) and one uncharacterized *AHII* gene [5]. Linkage analysis in sib pair with sickle cell disease has localized the QTL to chromosome X22.2. Alleles on X22.2 appeared to be codominant and not affected by X inactivation [7].

On the other hand, Steinberg et al, who studied sickle cell anemia patients, also observed that Hb F expression in sickle cell patients inhibit the polymerisation of sickle hemoglobin resulting in mild or benign clinical course. From their SNP work in sickle cell patients they concluded that genetic elements adjoining the 6q22.3-q23.2 QTL may harbor trans-acting elements that help modulate the baseline Hb F level in sickle cell anemia [8].

In Thalassemic Research Center (TRC), genome-wide search for single nucleotide polymorphism association with severity of β-thalassemia / Hb E patients revealed five significant SNPs in *HBS1L* (Hsp70 subfamily B suppressor 1-like) gene. However all 5 SNPs were located in intron regions (unpublished data)

Due to the significant association of genetic elements on chromosome 6q23 including *HBS1L* gene with high Hb F level, this study therefore aimed to identify the SNP(s) in all exons including exon-intron junctions and promotor region upto 1 kb from transcription start site, in the *HBS1L* gene. Direct sequencing was performed in 30 selected cases with mild and severe β-thalassemia / Hb E disease who have different genotype of *XmnI* polymorphism and different levels of Hb F. When potential the SNP(s) was found in sequencing analysis, the SNP was genotyped in 295 severe and 180 mild cases by PCR-RFLP techniques followed by association studies using statistical analysis.

CHAPTER 2
OBJECTIVE

The objective of this thesis is:

1. To identify SNPs in the *HBS1L* gene spanning in the region of chromosomes 6q23 that may be responsible for elevated HbF in 30 selected cases of β thalassemia / Hb E patients.
2. To genotype, the potential SNP(s) in 295 severe and 180 mild cases of β-thalassemia / Hb E cases and evaluate the significance of SNP(s) to the severity and Hb F level of patients.

CHAPTER 3

LITERATURE REVIEW

Blood, which is composed of different types of cell, performs variety of tasks in our body, one among which is to pick up oxygen in the lungs and delivers it to the every tissue of our body to maintain the viability of cells [2]. The portion of the blood that performs this work is the red blood cells, which contain hundreds molecule of hemoglobin. Hemoglobin is a quaternary structure, comprised of two alpha (α) globin and two beta (β) globin chains. Each globin chain contains rectangular heme. In the center of each heme group is an iron (Fe) atom. For normal hemoglobin formation, two gene clusters are very important: α-gene cluster which is located on chromosome 16 and β-gene cluster on chromosome 11.

3.1 Chromosome Organization of Globin Gene Clusters

The human α-globin cluster spans about 80 kb on the short arm of chromosome 16, in a band p13.3, in the order 5′-ζ2-ψζ1-ψα2-ψα1-α2-α1-θ-3′ [9]. It includes an embryonic gene (ζ2), two fetal/adult genes (α2 and α1), three pseudo genes (ψζ1, ψα2, ψα1) and a gene (θ) of unidentified function [10]. Whereas the β-like genes form a cluster, which spans 70 kb on the short arm of chromosome 11, in a band p15.5, and include an embryonic gene (ε), two fetal genes ($^{G}\gamma$ and $^{A}\gamma$), two adult genes (δ and β) and a pseudogene (ψβ) [11]. The α- and β-globin loci have important upstream regulatory regions. In the β-globin locus, this is called the locus control region (LCR) [12, 13, 14]. Five DNaseI-hypersensitive sites have been identified upstream of the β-globin locus. The most 5' site (HS5) does not show tissue specificity while HS1-4 and the LCR are largely erythroid specific and establish a transcriptionally active domain that spans the entire β-globin gene cluster. In the α-globin locus, HS-40, the major control element of α-globin gene family, locates 40 kb upstream of ζ-globin gene [15]. There are various erythroid-specific and ubiquitous DNA-binding protein binding sites within the 300 bp core region similar to the locus control region of the human β-globin gene cluster [15]. Their functions, however, are

Somewhat different from each other because HS-40 is designated as a positive control element of the human α-globin gene cluster [16][Fig 1].

3.2 Globin Gene Structure.

The β-like genes on chromosome 11 contain two non-coding regions known as intervening sequences (IVS) or introns of 122–130 and 850–900 base pairs (bp), between codons 30 and 31, and 104 and 105 respectively. For the α-like globin genes on chromosome 16 the intervening sequences interrupt the coding sequences between codons 31 and 32 and codons 99 and 100 [11]. Although the precise codon position numbers at which the interruption occurs differ between the α and β-like globin genes, the introns occur precisely at the same position relative to the regions of the primary structure of the α- and β-globin chains, which are homologous. The first intervening sequence (IVSI) is shorter than the second intervening sequence (IVSII) in both α- and β-globin genes but the IVSII of the human β-globin genes is much larger than that of the α- globin genes [11].

3.3 Normal Human Hemoglobin

The composition of human hemoglobin (Hb) changes in order to meet the changing oxygen requirements during developmental processes [17, 18]. All the normal hemoglobins are tetramers of two pairs of unlike globin chains. Adult (HbA) and fetal (Hb F) hemoglobins have α-globin chains that are combined with β-globin chain (HbA, $\alpha_2\beta_2$),) or γ-globin chains (Hb F, $\alpha_2\gamma_2$) whereas in the embryo, ζ-globin chains combine with γ-globin chains (Hb Portland, $\zeta_2\gamma_2$) or ε-globin chains (Hb Gower 1, $\zeta_2\varepsilon_2$), and α-globins combine with ε-chains to form Hb Gower 2 ($\alpha_2\varepsilon_2$) [19]. The developmental regulation of encoding globin genes reflects their sequential activation in a 5′–3′ direction; the way in which these developmental switches are controlled in globin gene expression is still not fully understood [20].

3.3.1 Hemoglobin switching.

Embryonic erythropoiesis occurs in the blood islands of the yolk sac and is associated with predominant ζ- and ε-globin expression (Hb Gower1 $\zeta_2\varepsilon_2$). The onset of fetal erythropoiesis in the liver at 6 to 8 weeks gestation coincides with upregulation of the α- and γ-globin gene expression to form fetal hemoglobin (Hb F;

$\alpha_2\gamma_2$), which becomes the predominant subtype. Shortly before birth, a second switch in hemoglobin occurs, coincident with migration of erythropoiesis to the bone marrow. Although the α-globin gene expression remains constant, the γ-globin gene expression is downregulated and β-globin synthesis increases with the formation of adult hemoglobin, (Hb A, $\alpha_2\beta_2$). This event, termed as γ- to β-globin switch is completed by 6 months after birth with adult Hb F being less than 1% of total hemoglobin [21]. It is very important to keep in mind that the switches in hemoglobin subtype are determined by the developmental maturity of the fetus and not by the site of erythropoiesis [22] [Fig 1].

3.3.2 Molecular basis of hemoglobin switching.

A lot of work has been done to understand the molecular mechanism of hemoglobin switching. Experiment on transgenic mice has revealed important sequences with and immediately flanking the globin gene that confers the appropriate developmental and tissue selectivity of expression. However, the expression was low in this context and subject to positional effects [23]. Restoration of high level positional-independent expression requires the presence of the locus control region (LCR) 5' to the ε-globin gene, which is a major regulatory sequence consisting of four erythroid specific DNase 1 hypersensitive sites (HS1-4) [12, 24]. Current model has suggested that the LCR sequentially exposed to the ε-, γ-, δ- and β globin gene promoters to activate expression during development. The LCR appears to influence globin gene expression by acting predominantly as a tissue specific transcriptional enhance. In condition such as HPFH, there appears to be a reciprocal relationship between γ- and β-globin chains production. This observation supports the notion of competition for transcription factors involved in interaction between the LCR and the globin gene promoters during embryonic, fetal and adult stage development. Similar events are thought to occur during erythroid maturation in the bone marrow.

The *cis* acting sequences in the individual globin gene and the LCR serve as templates for binding transcription factors. The function of these proteins is diverse including transcription activation, mediation of the interaction between the LCR and promoters and alteration of chromatin structure [25, 26]. Many of the factors that bind to the γ-globin regulatory sequences have been identified, including the hematopoitic-

specific proteins, NF-E2 and GATA-1 that bind to the LCR [27, 28]. GATA-1 also binds to the γ-globin promoter where it can function as either repressor or activator. Ubiquitous transcription factors such as Sp-1 and YY-1 bind to the LCR and globin promoter sequence. Three transcription factors are believed to be important for the switch from the γ- to β-globin gene expression during development. The first factor, erythroid Kruppel-like factor (EKLF), binds to the β-globin promoter CACCC box and is essential for high-level expression [29]. The second factor, FKLF, a novel Kruppel-like factor, was found to activate human ε- and γ-globin genes of K562 cells [30]. Similar to EKLF that interacts with the β-globin gene CACCC box, FKLF showed preferential interaction with the γ-promoter CACCC box. Over expression of FKLF enhanced γ-mRNA synthesis suggesting the *in vivo* role of FKLF in the control of γ-gene expression. The third transcription factor is the stage selector protein that binds to the stage selector element into the γ-promoter and the LCR [31]. Comprehensive analysis has revealed that the HPFH disorder arises by point mutation or deletion, which alters *cis* acting protein interactions with key regulatory elements in the γ-gene promoter. These changes may directly modify interaction between stage specific transcription factors and the LCR; the net result is persistent γ-gene expression during adult stage development.

In normal adult hemoglobin, the α- and the β- globin proteins are made in precisely equal amounts, despite the differing number of genes and are closely balanced. Any alteration or disruption in this balance results in a hemoglobin disorder. The genetic disorders of hemoglobin can be broadly classified into three groups:

1) The thalassemias, in which there is an imbalanced synthesis of normal globin chains.
2) Hemoglobin variants, in which there is a structural alteration of the amino acid residues of the globin chain.
3) Hereditary persistence of fetal hemoglobin (HPFH), in which there is a defect in the development progression from fetal to adult hemoglobin production.

3.4 Molecular Basis of Thalassemia and Abnormal Hemoglobin.

3.4.1 Alpha-Thalassemia

The α-thalassemia is characterized by the decrease or absence of α-globin chains production. It occurs mainly from a large deletion in the α-globin gene cluster involving one or both of the linked α-globin genes [32]. Less frequent α-thalassemia can occur from nondeletion defects such as point mutation or small deletion and insertion within either α_1 or α_2 globin gene [33]. These mutations may affect splicing, polyadenylation, translation or post-translational stability of the α-globin product [34]. α-Thalassemias can be divided into two main forms; the severe form, α-thalassemia 1 (α^0-thalassemia) and a mild form α-thalassemia 2 (α^+-thalassemia). The cause of α-thalassemia 1 is due to gene deletions involved different lengths of the α-globin gene cluster, which removes the two linked α-globin genes on the chromosome. The α-thalassemia 2 is due to either deletion of one α-globin gene or mutations that partially or completely inactivate one of the linked pair of the α-globin genes [35, 36]. Homozygous α-thalassemia 1, which results from the loss of all four α-globin genes in both chromosomes leads to no α-globin chain synthesis and results in the most severe form of thalassemia called Hb Bart's hydrops fetalis. Compound heterozygous for α-thalassemia 1 and α-thalassemia 2 is caused by the loss of three α-globin genes and results in the intermediate form of thalassemia called Hb H disease [37, 38].

3.4.2 Beta-Thalassemia.

The β-thalassemia is characterized by decreased or absent β-globin chain synthesis due to various abnormalities of the β-globin gene. Point mutations and small deletions or insertions in the β-globin gene nucleotide sequence are mainly responsible for the molecular defect of β-thalassemia [39]. β^+-thalassemia reduces the level of functional mRNA or protein while β^0-thalassemia causes the production of non-functional mRNA or produces no mRNA [40]. More than 200 mutations in the functionally important regions of the β-globin gene, which result in a deficit of β-globin chain production have been described [41]. The defect ranges from minimal (mild β^+-thalassemia alleles) to a complete absence (β^0-thalassemia alleles). With rare

exceptions heterozygotes from β-thalassemia, including β^0 thalassemia are clinically asymptomatic with minor hematological abnormalities whereas homozygotes or compound heterozygotes for β^0 thalassemia have severe disease and are transfusion dependent [42]. These mutations can be classified according to the type or category of defect in gene expression that they cause.

3.4.2.1 Frameshift mutation.

The reading frame of a mRNA is usually invariant. Translation starts at an AUG codon and continues in triplets to the termination codon UAA, UAG or UGA. The synthesis of normal globin chain requires that the triplet codon of the mRNA encoding each amino acid of the peptide chain can be set in the proper uninterrupted alignment. Nucleotide insertion or deletions other than those of multiple of three will lead to a shift in the reading frame (frameshift) of mRNA translation and the synthesis of a novel peptide sequence until a new in-phase termination codon is encountered, which usually occurs within a short distance. The mutant globin chain usually accumulates in markedly reduced amount or degraded rapidly because the mutant globin chain, either itself or its mRNA is unstable, resulting in a thalassemia phenotype. The frameshift resulting from a 4bp deletion (-TTCT) at codon positions 41 and 42 is a particularly common mutation among patients in Southern China and Thailand accounting for 40% of β-thalassemia in some regions [43, 44, 45, 46]. The deletion causes a frameshift and creates a stop codon at the new codon position 59 resulting in β^0-thalassemia [47, 48]. Although most frameshift mutations are associated with typical β^0-thalassemia, those that occur relatively far into coding sequence in exon 3 are associated with the phenotype of dominant β^0-thalassemia, including moderately severe hemolytic anemia, splenomegaly and inclusion body formation. This result occurs presumably because the mutant β-globin chains are able to bind heme and produce aggregates that are relatively resistant to proteolytic degradation [49, 50].

3.4.2.2 Nonsense mutation.

Nonsense mutation occurs when a single base substitution or point mutation is introduced inside the coding region of gene. This results in the creation of one of the termination codon, which inturn results in premature cessation of mRNA translation and may lead to the formation of truncated proteins. However, these abnormal proteins are not usually detected, presumably because they are very unstable and rapidly degraded. A number of nonsense mutations have been reported to cause β^0-thalassemia. The most common nonsense mutations are codon 17 AAG → TAG [51], which is common in Southern Asian region and codon 39 CAG → TAG [52] which is common in patients of Mediterranean ancestry.

3.4.2.3 Splicing mutation.

The coding region of human globin gene is interrupted at two positions by intervening sequences or introns, which are transcribed into globin precursor mRNA molecule, but subsequently excised, and the proper ends of the coding sequences are religated to yield the mature mRNA. This posttranscriptional processing of mRNA precursors is termed as splicing. A crucial prerequisite for the proper splicing of globin precursor mRNA molecules is the presence of specific nucleotide sequence at the junctions between coding sequences (exons) and intervening sequences (introns). The dinucleotides, GT and AG, present at the 5' and 3' splice junction respectively, are necessary for accurate splicing. Mutations of these residues result in block of normal mRNA splicing. Severity of mutation varies according to the site of mutation i.e. inside or outside conserved sequence. As a group, these mutations are generally associated with a phenotype of β^+-thalassemia. The G → C and G → T mutations at position 5 of the IVSI reduce splicing at the mutated donor site. They appear to activate cryptic donor sites, two in exon 1 and one in IVSI, which are utilized preferentially to the normal donor site [53]. Mutations within the second intron have also been identified [54, 55]. An interesting but unexplained feature of these mutations is the activation of alternative splicing by the C → T mutation at IVSII position 654. All mRNA are processed by splicing from the normal intron 2 donor to the activated cryptic receptor and from the mutated new donor to the normal intron 2 acceptor [55]. This transcript retains 73 nucleotides from the second intron,

rendering it functionless. In the mutations involving IVSII positions 705 and 745, some normal RNA are made in addition to the abnormally spliced products [53]. Accordingly, the mutation at the IVSII 654 position is characterized by β^0-thalassemia and the latter two mutations by β^+-thalassemia. However, the exact mechanism of these mutations by which they activate a cryptic acceptor site upstream from the mutation is unclear and also why a cryptic acceptor is used in preference to an unaltered acceptor splice site farther downstream. The IVS II-654 mutation is frequent among patients in China and Thailand, accounting for 20% of β-thalassemia in some regions [56].

3.5 Hemoglobin variants

Numeral independent factors influence the relative production of human hemoglobin variants. Most α-chain variants comprise about 25 percent of the total hemoglobin, whereas most stable β-chain variants are accounting for about half of the total hemoglobin. Many α- and β-chain variants constitute smaller proportion of the total hemoglobin than functioning globin chain. Usually, structural alterations of hemoglobin include single and double amino acid substitution, amino acid addition, deletions and fusions formed of δ- and β- globin chains or γ-and β- globin chains. About 90% are single amino acid substitution in the α-, β-, γ-, or δ-globin chains, no embryonic variants are known. The most common hemoglobin variants are Hb E, C and S [57]. The carrier (heterozygous) states for all three are, for the most part, symptomless. However, there are a few structural hemoglobin variants that are synthesized ineffectively and do have a thalassemia phenotype [58].

3.5.1 Hemoglobin E

Hemoglobin E is the second most prevalent hemoglobin variants worldwide. It is most concentrated at the borders of Thailand, Laos and Cambodia, an area dubbed as Hb E triangle [59]. Hb E is also encountered with increasing frequency in the immigrant populations of Europe, North America and Australia [60]. Hb E was first discovered in 1954 [61], and it became the fourth abnormal hemoglobin identified

by electrophoresis [62]. The substitution of lysine to glutamic acid at position 26 of the β-globin chain was identified in 1961 [63].

Individuals heterozygous for Hb E are clinically asymptomatic. Hemoglobin level is normal but red blood cells are microcytic with an average mean corpuscular volume (MCV) 84 ± 5fl [64].

Hb E homozygotes usually have normal hemoglobin level but some may be mildly anemic with prominent microcytosis and significant morphologic changes. Clinical symptoms are rare. No physical abnormalities are observed. However, hemoglobin analysis showed about 85-90 % Hb E with the remainder Hb F [68]. Red cell survival is slightly reduced and osmotic fragility is decreased. There is defective β^E–globin chain synthesis in all Hb E homozygotes with an average α/non-α ratio equivalent to the ratio in β^+-thalassemia heterozygote [66, 67]. Shortened cell survival may result, in part, from the instability of Hb E, a property attributed to the tendency of β^E dimers to dissociates into monomers, thereby exposing reactive SH group.

The thalassemic phenotype of the β^E gene emerges due to activation of a cryptic donor splice site by the codon 26 G→A mutation [68]. The new cryptic donor splice site at codon 25 competes with the normal donor splice site at the beginning of the first intron, thereby reducing β^E-mRNA generation [69, 70]. The abnormal spliced mRNA is non-functional since part of exon 1 is missing and a new stop codon is generated. Thus, the defective β^E chain synthesis is due to abnormal RNA splicing and decreased mRNA production, which results in a mild thalassemia phenotype [68]. Deceased synthesis of β^E-globin may be compounded by β^E–mRNA instability. In addition, *in vitro* experiments have shown that Hb E is mildly unstable and may be susceptible to oxidant damage [71].

3.6 Hereditary persistence of fetal hemoglobin (HPFH)

The physiologic switch from the production of fetal hemoglobin (Hb F) to the adult form of Hb (Hb A) is usually accomplished by 2 years of age. Hereditary persistence of fetal hemoglobin (HPFH) is a condition in which some defect occurs in normal switch from fetal to adult hemoglobin production resulting in continuous production of fetal hemoglobin in adult life. Co-inheritance of some forms of HPFH

can modify the phenotypes associated with thalassemia [19]. Usually normal adults continue to synthesize small quantities (<1%) of Hb F; the residual amounts of Hb F are restricted to a subset of erythrocytes termed fetal cells (F cells) [72]. Surveys of healthy blood donor in several population group show that the distribution of Hb F in adults is continuous and varies considerably (>20 fold) between individuals [73]. The heritability of fetal cell levels has been estimated to be 0.89 in the European population [4]. The genetic influences include DNA sequence variants *in-cis* to the β-globin complex such as the C → T single base substitution at position -158 in the promoter of the $^{G}\gamma$-globin gene (referred to as the *Xmn*I-$^{G}\gamma$ site).

Pancellular HPFH is inherited in a Mendelian fashion as alleles of the β-globin complex, caused either by extensive deletions of the β-globin cluster or point mutations in the γ-globin promoter [74]. However, there is another group characterized by modest elevations of Hb F levels (1% to 4%), distributed in an uneven fashion among the F cells. In this group of HPFH cases (heterocellular HPFH), mutations may not be identifiable within the β-globin cluster, and in many cases, the determinant is not linked to the β-complex, implicating the presence of *trans*-acting factor(s) [75, 76]. The importance of this condition is clearly demonstrated by the striking amelioration of the phenotype in individuals homozygous for β-thalassemia or sickle cell disease who also co-inherit a HPFH determinant [77, 78]. Furthermore, the eventual characterization of the genetic basis for this form of heterocellular HPFH will provide important insights into developmentally regulated gene expression, and may lead to new therapeutic strategies for the hemoglobinopathies.

3.7 Clinical Classification of β-Thalassemia

Based on clinical manifestation, the thalassemias can be divided into three groups.

3.7.1 Thalassemia minor

This term is used interchangeably for people who have small red cells and mild or no anemia due to thalassemia. These individuals are clinically well and are usually detected by routine blood testing. Affected people in minor thalassemia group usually represent the carrier states or traits.

3.7.2 Thalassemia intermedia.

Patients with thalassemia intermedia have significant anemia but are able to survive without blood transfusion. The sign and symptoms are comparable to those of thalassemia major but of a lesser magnitude. The factors that go into the diagnosis are (1) The degree to which the patient tolerate anemia, (2) threshold of the physician to transfuse patients with thalassemia in association with intercurrent illness. Growth and development during childhood is relatively uncompromised, pubescence takes place normally and fertility is preserved [79, 80].

3.7.3 Thalassemia major

Thalassemia major is a condition of severe thalassemia in which chronic blood transfusion is required for patients. Some patients could survive without blood transfusion for a while, but would have distressing deformities. Patients develop all signs and symptoms associated with severe anemia such as growth retardation, hepatosplenomegaly and thalassemic faces. The disease is usually fatal early in life. β-thalassemia major usually results either from the compound heterozygous state for two different β-globin mutations or homozygous state of the same mutation [81, 82].

3.8 β-Thalassemia/ Hb E diseases.

Thalassemias and structural hemoglobin variants occurs together at a high frequency in many populations [18]. There are higher chances that an individual could inherit genes for both types of conditions.. The compound heterozygous state of β-thalassemia and Hb E leads to β-thalassemia/Hb E disease with the phenotype ranging from mild anemia to the most severe form of β-thalassemia major [84, 85, 86]. However, β-thalassemia/Hb E disease is generally a syndrome of intermediate severity with hemoglobin levels at 7.7 to 8 g/dl although the clinical spectrum can vary considerably [83]. β-thalassemia/Hb E patients have hemoglobin levels varying from 3 to 11 g/dl [87, 88]. The presence of β^{+}-thalassemia, α-thalassemia and homozygosity for the *Xmn*I cleavage site in the $^{G}\gamma$-globin gene are major disease severity modifying factors. However, the extreme variation remains unexplained in many cases [87].

3.9 Clinical Diversity of β-Thalassemia

The clinical manifestations of β-thalassaemia are extremely diverse, spanning a broad spectrum from the transfusion dependent state of thalassemia major to the asymptomatic state of thalassemia trait. Figure 2 demonstrates the pictorial representation of bell shaped distribution of hemoglobin levels in a large number of β-thalassemia patients where x-axis represents hemoglobin level in g/dl and y-axis represents number of patients [83]. Most patients who have hemoglobin levels in the middle range from 6g/dl to 9g/dl are categorized in intermediate cases and their clinical symptoms range from a condition that is only slightly less severe than transfusion dependent β-thalassemia. However, there are two extreme ends; the left hand side represents the severe group having hemoglobin less than 6g/dl. These patients present thalassemic symptoms within 6 months of life. If not being treated with regular blood transfusions, they can die within the first 2 years. Meanwhile, the right hand side represents the group of patients whose hemoglobin level is more than 9g/dl, consequently categorized into mild cases. Patients in this group show little or no clinical symptoms of β-thalassemia. In summary, clinical manifestation of β-thalassemia is not homogenous.

The basic pathophysiology of β-thalassemia relates to a quantifiable deficiency of functional β-globin chains, which leads to an imbalanced globin chain production and excess of α-hemoglobin chains [42, 89]. The excessive α-hemoglobin chains aggregate early in red cell precursors, forming inclusion bodies that cause mechanical damage and premature destruction of red blood cells in the bone marrow i.e. ineffective erythropoiesis. Red blood cells that survive to reach the peripheral circulation are prematurely destroyed in the spleen. Anemia in β-thalassemia thus results from the combination of ineffective erythropoiesis and peripheral hemolysis. The imbalanced α-/non α-globin synthesis is the major factor in determining the severity of the disease (Fig 3) [19]. Secondary complications of bone disease, splenomegaly, endocrine, cardiac damage and the iron loading that results from the increased gastrointestinal absorption in the blood can be related to the severity of anemia.

3.10. Factors affecting severity of β-thalassemia

3.10.1 β-Thalassemia mutations: β-Thal, β^{+}-Thal.

In a large number of patients, the reduced disease severity can be explained by the inheritance of the milder β-thalassemia alleles (β^{++} and *silent*) that allows the production of a significant proportion of β-globin chains. A substantial number, however, have β^{0} thalassemia and in such cases, the clinical manifestation of the patients is usually heterogeneous. The absence of β-globin chains in a number of patients is compensated by an inherent ability to produce fetal hemoglobin (Hb F, $\alpha_2\gamma_2$) resulting in a mild clinical symptoms.

3.10.2 Co-inheritance of α-thalassaemia.

Various populations in which β-thalassemia is prevalent, α-thalassemia also occurs at a high frequency and hence it is not uncommon to co-inherit both conditions. This interaction alone provides the basis for considerable clinical heterogeneity. The degree of amelioration depends on the severity of the β-thalassemia alleles and the number of functional α-globin genes [45, 90, 91, 92]. Co-inheritance of a single α-globin gene deletion has very little effect on β^{0}-thalassemia. While individuals with two α-globin gene deletions and β^{+}thalassemia may have milder anemia [74]. In β-thalassemia heterozygotes, co-inheritance of α-thalassemia normalizes the hypochromic microcytosis [93].

3.10.3 *Xmn*I $^{G}\gamma$ polymorphism.

Genetic determinant influencing Hb F response can be within the β-globin gene complex or trans acting. The C→ T substitution at position -158 of the γ-globin gene, referred as the *Xmn*I-$^{G}\gamma$ polymorphism, is a common sequence variant in all population groups, presenting at a frequency of 0.32 to 0.35 [94]. This genetic variant accounts for about one third of the variation in Hb F in normal adults. It has been shown that healthy individuals with the *Xmn*I-$^{G}\gamma$ +/– or +/+ genotype are more likely to have a high percentage of F cells [95]. During the hemopoietic stress conditions like homozygous β-thalassemia and β-thalassemia/Hb E the presence of the *Xmn*l-$^{G}\gamma$ site favors a higher Hb F response [96, 97], which inturn reduces α- chain

imbalance, increases the overall hemoglobin level and decrease the severity of both homozygous β-thalassemia and β-thalassemia/Hb E patients [87]. However, there are some patients who despite being *Xmn*I $^{G}\gamma$-/- show enhanced Hb F indicates other genetic factors responsible for high Hb F response. [90, 98]

3.11 Stimulation of Hb F production.

Globin chain imbalance can also be reduced if there is an inherited propensity for producing γ globin chain which combines with the excess α-globin to form fetal hemoglobin (Hb F, $\alpha_2\gamma_2$). Although the production of Hb F is almost switched off at birth, all adults continue to produce residual amount of Hb F. In β-thalassemia, Hb F level is relatively increased due to the selective survival of the erythroid precursors, which synthesize relatively more γ globin chains. However, patients with β-thalassemia differ considerably in their ability to synthesize γ-globin chains and their Hb F response. This becomes evident in the group of homozygous β^0-thalassemia patients who have a mild disease despite the absence of Hb A [99]. These patients appear to have an inherited ability to produce Hb F and are able to maintain a reasonable level of hemoglobin, all of which is Hb F. Hence, against this background of an increased Hb F from the expanded erythroid mass and the selective survival of fetal cells (FC) are genetic factors which account for the individual Hb F response to the stress of β-thalassemia.

Induction of Hb F using drugs and the inherited condition of HPFH clearly demonstrated that the stringent development program governing globin gene expression can be altered. The ability to enhance *in-vivo* fetal globin synthesis by pharmacological manipulation was initially demonstrated in baboon treated with 5-azacytidine (5-AzaC) [100]. Subsequent studies confirmed the ability of 5-AzaC to increase Hb F in patients with β-thalassemia and sickle cell disease [101, 102].

Demonstration of a temporal relationship between developmental repression of γ-globin gene expression and methylation of 5' flanking regions of the γ-globin genes [103] suggested the strategy for reactivation of the genes. Numerous attempts have been made to increase the synthesis of fetal γ-globin chains in an effort to improve the disease severity of β-thalassemia. The underlying principle for this approach is based

on observations that patients with the severe forms of β-thalassemias who produce unusually high levels of Hb F tend to have milder disease since postnatal fetal hemoglobin ($\alpha_2\gamma_2$, Hb F) compensates for the unbalanced α/β globin chain ratio.

In addition of 5-azacytidine several other drugs like hydroxyurea (HU) [104], butyrate analogues [105], cytosine arabinoside [106], Myleran [107] and Vinblastine [108] have been shown to stimulate Hb F production in some primates and patients with sickle cell disease or β-thalassemia. However, many of these drugs have low efficacy and specificity while some are potentially cytotoxic and carcinogenic [109, 110]. There is therefore a clear need for additional drug discovery program and the development of alternative treatment strategies, which can stimulate Hb F production with better efficiency and less toxicity.

3.12 Genetic determinants for fetal Hb production.

Besides *cis* modulating factors in and around the β-globin gene cluster including *Xmn*I-$^{G}\gamma$ polymorphism, there are some thalassemic patients who have an enhanced Hb F response despite being *Xmn*I-$^{G}\gamma$-/- [90, 99].

Family studies have shown that the inherent capacity of Hb F production is due to genetic determinant, which might not link to the β-globin cluster. As evident from Swee Ley Thein's group, who performed sib-pair studies in normal adults and showed that more than 50% of the fetal cells variance in the general population is accounted for by *trans*-acting factors [94]. Indeed, analysis of a group of thalassemic intermedia patients revealed seven sib-ships with discordant phenotype despite identical α-globin and β-globin genotypes. The steady state Hb F value between the siblings ranged from 1g/dl to as much as 8-9g/dl and was attributed to genetic determinant not liked to the β-globin locus [90].

Trans-acting loci controlling Hb F and F cell levels have now been mapped to three regions of the genome chromosomes 6q23, Xp22 and 8q. [106, 7, 5.]

Chromosome Xp22.2-p22.3 has been linked with variation in fetal cell (Hb F) level in sickle cell disease by linkage analysis [7]. This study suggested that fetal cell production (FCP) locus on the short arm of X chromosome partially control the percentage of F reticulocytes.

Chromosome 8q was revealed by the linkage studies in Asian Indian kindred, which showed the linkage of fetal cells level (Hb F) to chromosome 8q [5]. The influence of the quantitative trait locus was shown to be conditional on *Xmn*I-$^{G}\gamma$ polymorphism genotype suggesting a genetic interaction.

Chromosome 6q23 was discovered by genomewide linkage analysis in large Indian family. This study showed a significant linkage between a *trans*-acting locus on chromosomal 6q23 and increased Hb F expression in Indian Kindred including individuals with HPFH associated with β-thalassemia. Subsequent detailed annotation of the 6q23 region encompassing ~1.5 Mb revealed four known genes (*MYB, ALDH8AI*, *HBS1L* and *PDE7B*), one uncharacterized *AHI1* gene and a high level of alternate splicing Figure 4. This candidate interval on chromosome 6q23 comprising 1571770 bp (~1.5 Mb) of DNA was defined by the markers D6S270 (Z16636) and DbAD6 (AJ606363) [6]. From these 5 protein coding genes, more attention has been paid to *MYB* and *AHI1* gene because these genes had the functional significance, while they do argue that *MYB* or *AH1* in 6q23 region may not be a candidate gene responsible for high Hb F in HPFH Indian family

On the other hand, Steinberg et al, who studied sickle cell anemia patients, also observed that Hb F expression in sickle cell patients can inhibit the polymerisation of sickle hemoglobin and result in a mild or benign clinical course. From their SNP work in sickle cell patients they concluded that genetic elements adjoining the 6q22.3-q23.2 QTL may harbor trans-acting elements that help modulate baseline Hb F level in sickle cell anemia [8].

HBS1L *(HBS1-like (S. cerevisiae))* gene is 94.5 kb long. The major product is 7.1 kb mRNA encoded by 18 exons, which is translated into a polypeptide of 648 amino acids with a predicted molecular mass of around 75 kDa. It utilizes three poly A signals, out of which the major transcripts consume second poly A signal. It also gives rise to alternate transcript, which utilizes the same first four exons of the major transcript and terminates at fifth alternate exon “4A” (missing the subsequent exons 5–18 of the primary transcript) [Fig 6]. The exon 4A sequence contains an open reading frame resulting in an additional 489 amino acids which are unique to this splice variant [6].

HBS1L was originally identified during a comparative genome hybridization study searching for chromosomal imbalances in pancreatic adenocarcinoma [112]. It was also discovered among the co-amplified genes in the chromosomal region of 6q24.

Phylogenetic studies suggested that *HBS1L* may be associated with translating ribosomes and aid in the passage of the nascent polypeptide through the ribosome channel. Alternatively, it may bring the amino-acyl-tRNA to the ribosome [113, 114]. Another product of *HBS1L* gene, which is alternate exon "4A", is the unique splice variant. The protein sequence of this alternate transcript is novel with no significant homologies, and whose function is entirely unknown. [106].

3.13 Single Nucleotide Polymorphism. (SNP)

Single nucleotide polymorphism (SNP) corresponds to differences of the two alleles at a single nucleotide position. In a strict sense, a single base change, in a population at a frequency of >1% is termed as a SNP. When a single base change occurs at <1%, it is generally considered to be a mutation. SNPs are usually considered widespread in the population and mutations are usually rare. However, there are some examples where this explanation do not stand correct. For example, mutations in the breast cancer susceptibility gene *BRCA1*, have been found in 1-2% of Jewish populations [115] and mutations in the *CFTR* gene causing cystic fibrosis disease in the Caucasian population with a carrier frequency of around 2% [116] are still called mutation not polymorphism.

Although most SNPs reside within non-coding genomic regions, an important subset corresponding to mutations in the genes are associated with diseases or other phenotypes. SNPs occur frequently in most genomes and have a low mutation rate, feature that makes them desirable for use in building comprehensive genetic maps. Identification of SNP requires PCR-based techniques following by automated DNA sequencing and restriction fragment length polymorphism (RFLP).

3.14 Statistical Analysis Methods.

3.14.1 Haplotype Analysis.

The Human Genome Project and other large-scale efforts have identified millions of genetic markers that can be used in genetic studies. Although each marker can be analyzed independently of other markers, it is much more informative to analyze markers in a region of interest simultaneously. The combination of marker alleles on a single chromosome is called a haplotype (*Haplo*id Geno*type*). There is great interest in understanding haplotype structures in the human genome or a particular gene using identified genetic markers. Haplotype structures may provide critical information on human evolutionary history and the identification of genetic variants underlying various human traits. For haplotype including markers that are tightly linked with each other, for example markers (SNPs) within the same gene, alleles at these markers often display statistical dependence, a phenomenon called linkage disequilibrium (LD) or allelic association. One major aspect of haplotype analysis is to identify LD patterns in different regions and different populations because the existence of LD among markers makes it possible to infer population histories and localize genetic variants underlying complex traits. LD is affected by many factors including the age of the variants, population history, recombination rates, gene conversion, natural selection, and other factors.

The haplotype-block model has important implications for association studies, since it indicates a simple rationale for how to select SNP markers. The main haplotypes could be labeled with a small number of haplotype-tagging SNPs (htSNPs) sufficient to distinguish between the main haplotypes within a block [117]; thus, most of the variation in the genome could be traced with a limited number of SNP markers

3.14.2 Linkage disequilibrium

Linkage disequilibrium is a critical factor in identifying disease associated genetic variants and designing efficient studies to detect disease gene associations. There are many calculations to measure the degree of association between two polymorphisms [118]. The most commonly used ones are D' and r^2. Both |D'| and r^2 range between 0 and 1. Simulation studies based on simple population genetics models suggested that useful LD extends only a few Kilobase (kb) around

common single nucleotide polymorphisms (SNPs) [119]. LD is both locus and population oriented. The magnitude of LD in a local region tends to be small if the estimated local recombination rate is high [120]. Since the recombination rate is associated with many characteristics of the DNA sequences [121], it is not surprising that LD is also significantly associated with sequence properties [122].

3.14.3 Hardy-Weinberg Equilibrium(HWE)

The Hardy-Weinberg Equilibrium relies on the law of joint probabilities that the probability of two independent events is precisely equal to the product of the probabilities of the events. The mathematical model developed by Hardy and Weinberg is based on a number of assumptions. (1) The population is large enough so that error in measuring allele frequencies are negligible. (2) There is no selective advantage for any genotype. (3) Mating within the population is random and (4) factors such as mutation and migration are absent or rare events and can be ignored [123]. Expected genotype frequencies are calculated from allele frequencies under the assumption $p^2 + 2pq + q^2 = 1$, where p and q are the allele frequencies and p^2, q^2 and 2pq correspond to the frequencies of the three possible genotypic states.

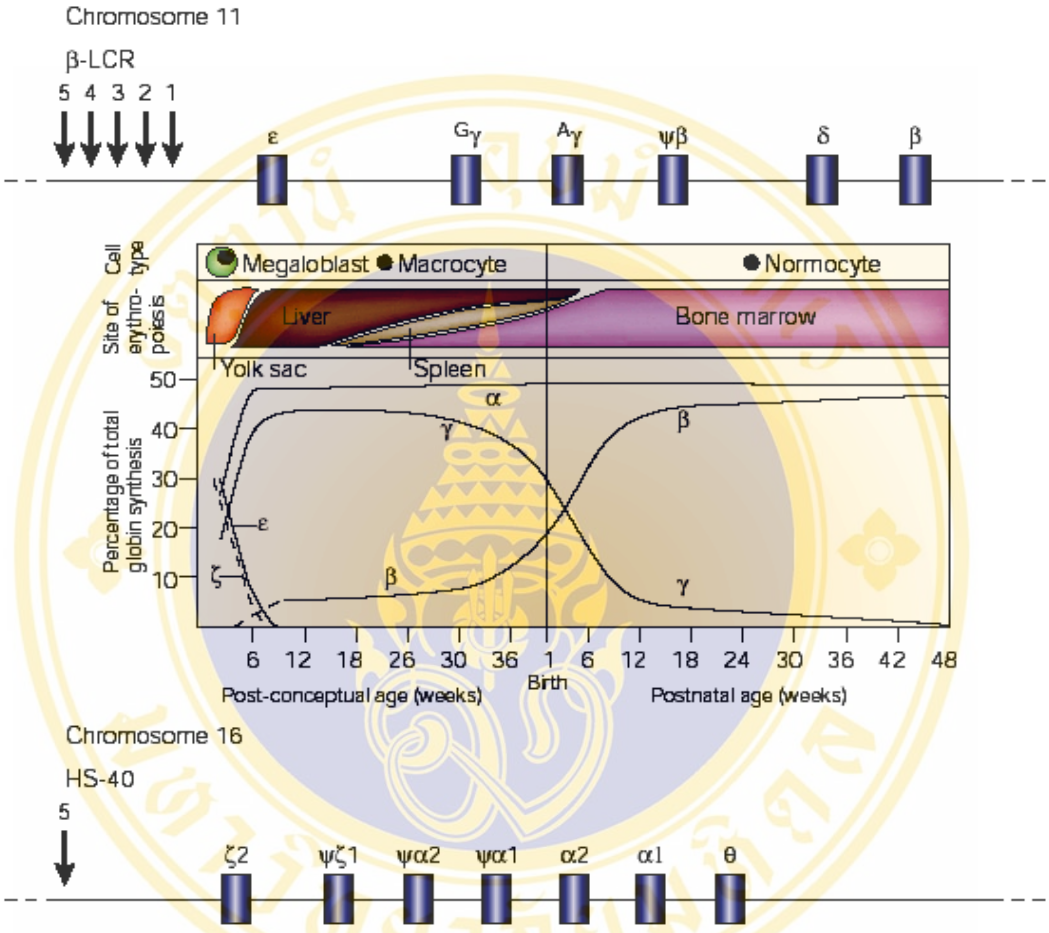


Figure 1. **The α-globin gene cluster on chromosome 16 and the β-globin gene cluster on chromosome 11.** Vertical arrows indicate the location of DNaseI hypersensitive sites that are thought to be involved in globin gene regulation. The products of the Gγ- and Aγ- genes are γ-chains with either glycine (Gγ) or alanine (Aγ) at position 136. The insert shows the sequential activation of the embryonic, fetal and adult globins. (LCR, locus control region.) **[13]**

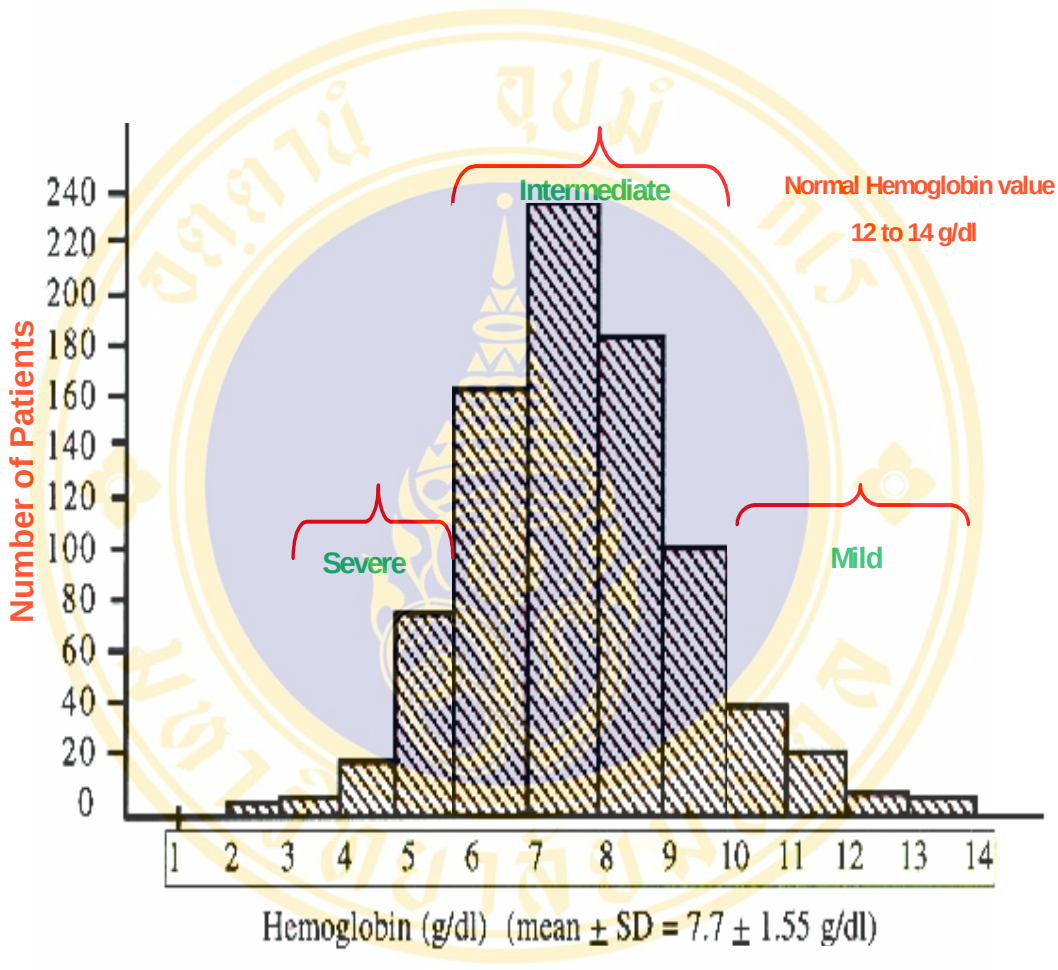


Figure 2. Varied hemoglobin values in mild, intermediate and severe β-thalassemic patients. [91]

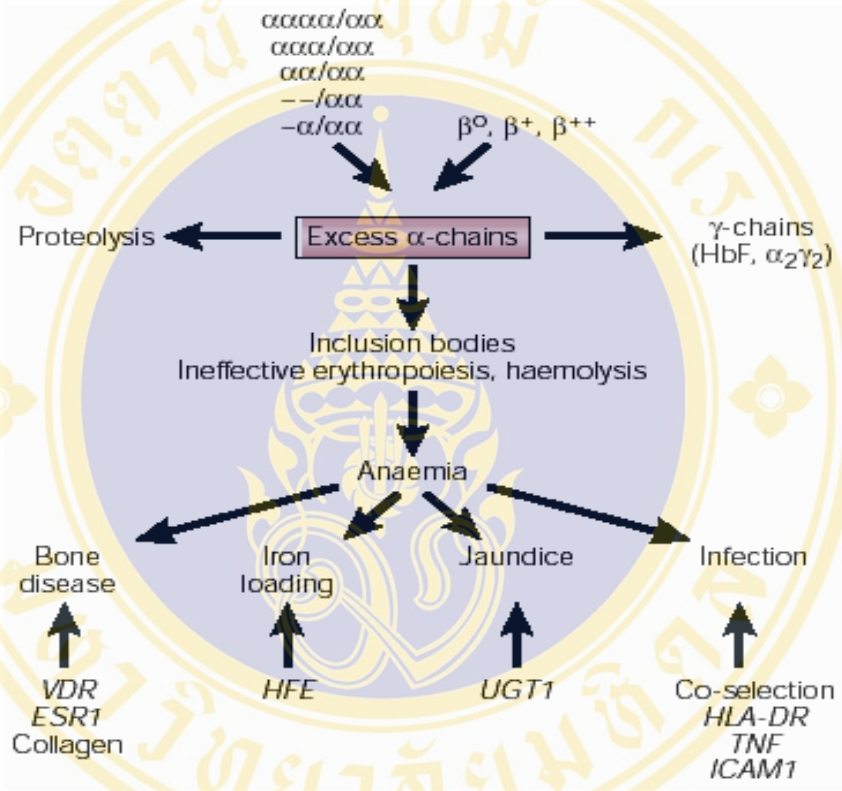


Figure 3 **Pathophysiology of β thalassaemia.** The precipitation of excessive α-globin giving rise to number of pathophysiological conditions in β-thalassemic patients.[13]

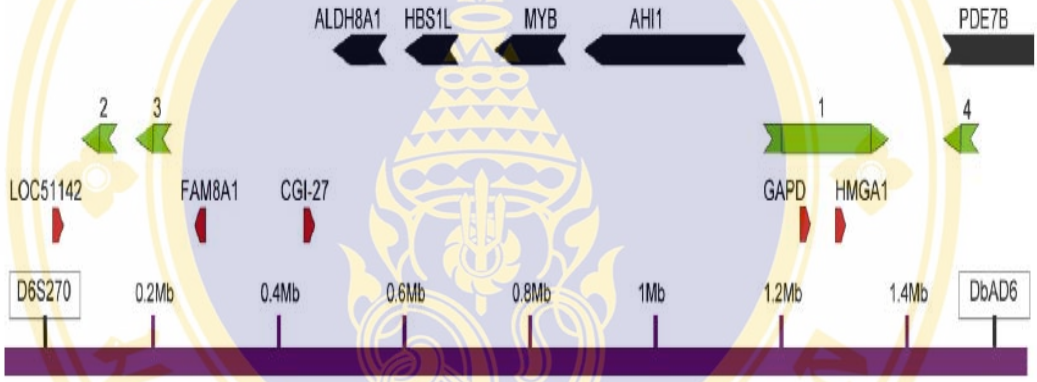
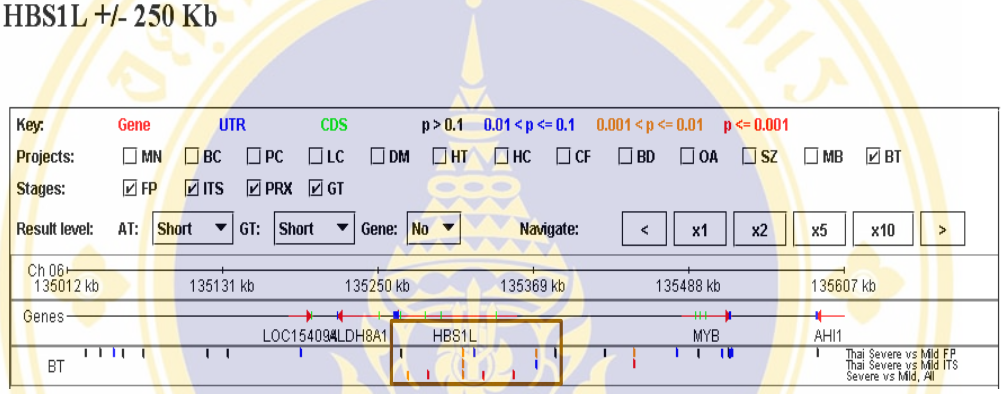


Figure 4. **1.5 Mb candidate region**. which contain 5 protein coding genes on top, non-coding RNA genes in middle and pseudogenes below between two makers. Direction of arrows specify transcriptional direction in 5' to 3' of the genes.[132]



Chromosome	Gene	Assay	P-Value (AF_P)	Signif
6q23-q24	HBS1L	1014021	2.31e-03	**
	GTP binding	9376074	2.07e-03	**
		1041480	1.65e-03	**
		987690	9.92e-04	***
		4376364	8.37e-05	***

Figure 5. Scan view of five significant SNPs associated with high Hb F in *HBS1L* gene, located on chromosome 6 in β-thalassemic patients, discovered during Genome Wide SNP search by Thalassemia Research Center (TRC) Thailand

Unpublished Data.

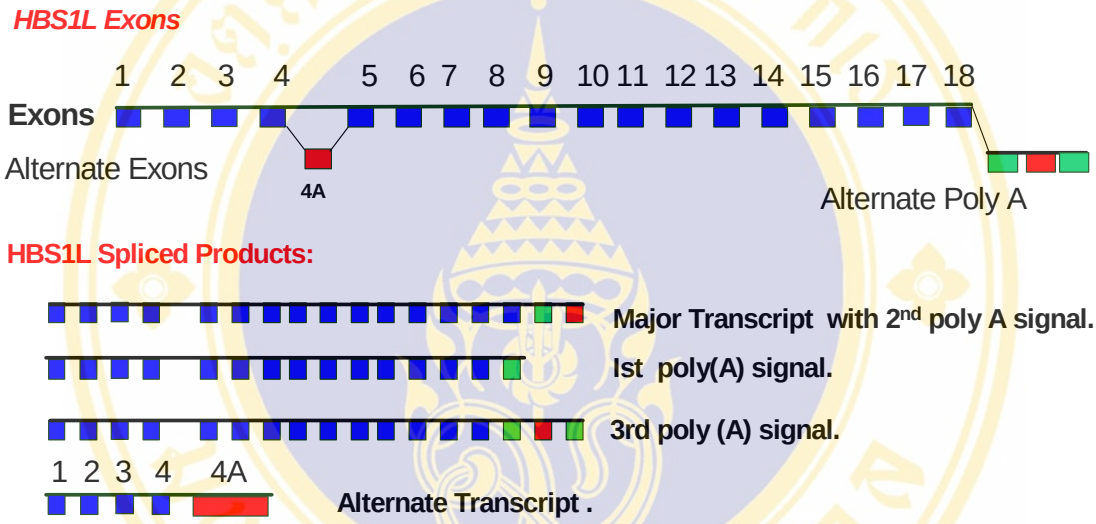


Figure 6. Structure of *HBS1L* gene, major and alternate transcript 4A [132].

CHAPTER 4
MATERIALS AND METHODS

4.1 Materials

4.1.1 Subjects and DNA samples

A total of 575 DNA samples, 196 mild, 309 severe with normal α-globin genes and 70 β-thalassemia / Hb E with α-thalassemia were recruited in this study. The race of all subjects was Thai-Chinese. The patients taken for this study were diagnosed by hematological data and DNA analysis. They were categorized into mild, intermediate and severe groups according to the severity score, which relied on hemoglobin level, age onset of thalassemic symptom, age at which patient receives his or her first blood transfusion, size of spleen and degree of growth retardation (Table 1) [124]. In general, patients whose total severity score was less than 3.5 were categorized in mild cases, patients whose total severity score ranged from 3.5 to 7.5 were considered to be the intermediate case and the patients whose total severity score was more than 7.5 were positioned in severe group of β-thalassemia.

From the above entries, DNA samples of 30 patients were taken for *HBS1L* gene sequencing, out of which 14 were severe and 16 were mild cases with different values of Hb F and different genotypes of *Xmn*I (Table 2, Appendix A). After DNA sequencing, the significant polymorphic site of *HBS1L* gene was further genotyped in 545 patient samples.

4.1.2 Oligonucleotide primers for PCR amplification

Primers for PCR amplification were designed for all 18 exons, alternative 4A exon and promoter region of *HBS1L* using Vector NTI, primer premier version 5 software program's. Primers for exon 1 to exon 18 were designed in such a fashion that they cover all exons including exon intron junction, (Fig 7a) and three poly A signals for exon 18. Primer for promoter region was designed in such a way that its PCR product will encompass 1kb of promoter region. For alternate exon 4A, whose

exon length was more than 2.1 kb, three primer pairs were designed in overlapping fashion to perform PCR amplification in three different reactions namely E4A1, E4A2 and E4A3 with PCR products ranging from 968 to 1090 bp. These PCR products cover the whole exon including intron-exon junction and poly A signals (Fig 7 b). All primers were purchased from Bio Basic Inc., Canada. The primer sequences and the expected size of PCR products are listed in Table 3**.**

4.1.3 Oligonucleotide primers for DNA sequencing.

Sequencing primers for PCR products sequencing for exons 1-8 and 10-17 were either forward or reverse primers used for PCR amplification. However, additional primers were designed for exons 9, 18, 4A3 and 1 Kb promoter region sequencing. Sequences and alignment of primers used for sequencing are listed in Table 4.

4.1.4 Chemicals.

Chemical name	Formula	Company
Absolute ethanol	C_2H_5OH	MERCK
Acetic acid	CH_3COOH	MERCK
Acrylamide	C_3H_5NO	BIO BASIC
Agarose		Gibco BRL
Ammonium persulfate	$(NH_4)_2S_2O_8$	AMRESCO
Bromophenol blue	$C_{19}H_9Br_4O_5SNa$	SIGMA
Formaldehyde solution	CH_2O	MERCK
Silver nitrate	$AgNO_3$	MERCK
Sodium carbonate	Na_2CO_3	BDH
Urea	$CO(NH_2)_2$	BDH
Sodium thiosulfate	$Na_2S_2O_3.5H_2O$	Fisher Scientific
Sodium iodide	NaI	Bio vista
Glass milk		Bio vista

4.1.5 Enzymes

Taq DNA polymerase	Promega.
ApaL1 restriction enzyme	Biolabs, Fermentus.

4.1.6 Miscellaneous materials

DNA size marker 100 bp + 1.5 k bp

dATP, dCTP, dGTP and dTTP Fermentus

Rain shield Blue Power

4.1 7. Reagents

4.1.7.1 Reagents for agarose gel electrophoresis

0.5 µg/ml Ethidium bromide solution

Loading dye solution; 25% glycerol, 60 mM EDTA and 0.25% bromophenol blue

10X AGB buffer pH 8.0 ; 890 mM Tris-HCl pH 8.0, 890 mM boric acid and 25 mM EDTA

4.1.7.2 Reagents for amplification of Exons

2 mM dNTPs (dATP, dCTP, dGTP and dTTP)

25 mM $MgCl_2$

10X PCR buffer ; Tris·Cl, KCl, $(NH_4)2SO_4$, 15 mM $MgCl_2$; pH 8.7

4.1.7.3 Reagents for PCR product purification.

Sodium iodide

Glass milk, (Silica matrix) Bio vista.

Washing buffer, 10 mM Tris-HCl (pH 7.4), 0.5 mM EDTA, 50 mM NaCl, 50% ethanol

4.1.7.4 Reagents for DNA sequencing by MegaBACE 500 DNA Analysis System

DYEnamic ET Dye Terminator cycle Sequencing Kit containing

MegaBACE Loading Solution (70% formamide, 1 mM EDTA)

Ammonium acetate.

4.1.7.5 Reagents for RFLP.

ApaL1 restriction enzyme (Bio labs)

10 X NEbuffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9@ 25°C)

100 X BSA.

4.1.7.6 Reagents for denaturing polyacrylamide gel electrophoresis and silver staining

10 X TBE buffer pH 8.0 ; 890 mM Tris-HCl pH 8.0, 890 mM boric acid and 25 mM EDTA.

40% Acrylamide:bis-acrylamide (48:2)

10% Ammonium persulfate.

Bind Silane ; 0.3% silane, 0.5% acetic acid and 94% ethanol.

98% Formamide loading dye; 98% formamide, 10 mM EDTA, pH 8.0, 10% bromophenol blue and 10% xylene cyanol FF .

10 mg/ml Sodium thiosulfate.

15% Acetic acid.

4.1.8 Instruments

Avanti^{TM} 30 Centrifuge, Bechman, USA

Centrifuge 5417C, Eppendrof, Germany

Electrophoresis set, Wealter Corp, Taiwan.

Gene Amp PCR System 2400, Perkin Elemer Cetus, USA.

Gene Amp PCR system 9700, Applied Biosystem, USA.

MegaBACE 500 DNA Analysis System, Amersham Biosciences,UK

U-2000 double-beam spectrophotometer, HITASHI, Japan

Vortex-2 GENIE, Science Industries, Inc., USA

Water bath, Julabo SW21, Germany

White/Ultraviolet Transilluminator, Bio Doc IT^{TM} System, USA

Gel doc Bio-Rad Laboratories segrate, Milan, Italy.

4.1.9 Computer softwares

Chromas.MFC. Version 2.22 , Technelysium Pty Ltd, USA.

Primer premier. Version 5. Premier Biosoft Corp, CA, USA.

Vector NTI. Version 5.1 InforMax, Inc, USA.

Haploview. version Daly lab, Cambridge, USA.

SPSS 14.0 version LEAD Technology, Inc, USA

4.2 Methods.

4.2.1 Detection of concentration, purity and quality of genomic DNA

Although stock DNA samples from previous genome project were used in this study, purity of DNA was determined by calculating the ratio of OD_{260} to OD_{280}. Pure DNA has OD_{260}/OD_{280} ratio of 1.7-1.9.

Genomic DNA concentration was determined by measuring the absorbance at wavelength 260 nm using spectrophotometer. Diluted DNA samples were prepared by diluting 0.5 µl of genomic DNA with 49.5 µl of distilled water (1:100 dilution). The optical density (OD) at wavelength 260 and 280 nm of diluted DNA sample was measured and the concentration of DNA was calculated by using formula shown below.

DNA concentration (µg/ml) = OD_{260} x dilution x 50 µg/ml

The quality of genomic DNA was also detected by agarose gel electrophoresis using 0.8% agarose gel. Agarose gel was prepared by completely melting agarose powder in 1X AGB buffer and allowed to cool to 50-60^{o}C before pouring into the electrophoresis chamber set with comb inserted. The gel was kept at room temperature approximately 20 min to set into semisolid gel. The 1 µl of DNA sample was mixed with 1/3 volume of loading dye and loaded into gel slots in a submarine condition. Electrophoresis was set at 80 volts for 90 min. After that, the gel was stained in 0.5 µg/ml ethidium bromide solution for 5 min and destained with distilled water for 5 min. The DNA pattern was visualized under UV light by UV transilluminator and photographed. The DNA with high quality and no fragment shearing should have the pattern of the length of DNA higher than 23 kb.

4.2.2 DNA amplification of all exons and 1 kb promoter region of HBS1L gene.

The PCR reactions of each exon were optimized on normal DNA template sample by varying $MgCl_2$ from 1.5 mM to 2.5 mM and annealing temperature from 47 ºC to 56 ºC . After getting optimized PCR profile (Table 3), all 18 exons, alternate exon 4A and 1 kb promoter region was amplified using genomic DNA

of 30 β-Thal/Hb E patients as a template by PCR method. PCR reaction for all exon was performed in total volume of 50µl. The reactions composed of 100 ng of genomic DNA, 0.2 mM dNTPs, 0.2 µM each primer for each marker and 1 Unit of *Taq* DNA polymerase in 1X PCR buffer (Tris-HCl, KCl, $(NH_4)_2SO_4$, 15 mM $MgCl_2$; pH 8.7) with the varied concentration of $MgCl_2$ (Table 3). Then 1 µl of PCR product was analyzed by agarose gel electrophoresis using 1.5% agarose gel. The concentration of amplified PCR products were estimated from agarose gel by comparison with the intensity of the bands of standard size marker, 100 bp DNA ladder. Total amount of 100 bp DNA ladder in one lane loading was 200 ng, which contained some fragments of 500 bp, 400 bp, 300 bp, 200 bp and 100 bp with approximately 33 ng, 8.8 ng, 6.6 ng, 4.4 ng and 4.4 ng, respectively.

4.2.3 Purification of PCR product by Gene Clean Method.

The PCR product of every sample of each exon was purified before performing DNA sequencing. PCR product was transferred from PCR tube to the 1.5 ml eppendrof tube. Then 3 volumes of 6 M NaI and 5 µl of GLASSMILK was added, followed by incubation at room temperature for 10 min (vortex every 1 min) and centrifuged at 12,000 rpm for 10 sec. The GLASSMILK pellet was washed 3 times with 500 µl of new washing buffer [10mM Tris-HCl (pH 7.4), 0.5 mM EDTA, 50 mM NaCl, 50% ethanol] centrifuged at 12,000 rpm for 10 sec, then supernatant was discarded. The pellet was air dried for 10 min and resuspened with 15 to 20 µl of sterile distilled water. The suspension was then incubated at 55°C for 5 min and centrifuged at 12,000 rpm for 30 sec. Finally, eluted DNA was transferred to a new microcentrifuge tube.

4.2.4 DNA Sequencing.

After purification by Gene Clean method, the concentration of PCR products was estimated from agarose gel by comparison with the intensity of the bands of 100 bp DNA ladder. PCR product concentration was adjusted between 20 to 50 ng by adding sterile distilled water for every sample before pass on for sequencing step.

PCR products of 30 sample amplified from each exon were sequenced using DYEnamic ET Dye Terminator Cycle Sequencing Kit, in which terminators are labeled with florescent dye for automated detection (Amersham Bioscience). The sequencing was performed in 10 µl final volume containing 20 to 50 ng DNA, 4 µl of DYEnamic ET terminator reagent premix and 1 µl of 5 pM of sequencing primer. The reaction was undertaken for 25 to 30 cycles of denaturation at 95 °C for 20 sec, annealing at 50 °C for 15 sec and extension for 60 °C for 1 min. The PCR product was mixed with 10 µl of steriled water, 2 µl of 7.5 M ammonium acetate and 60 µl of 95% ethanol. After mixing and centrifugation at 12,000 rpm for 15 min, the supernatant was removed and the pellet was washed with 200 µl of 70% ethanol followed by centrifugation at 12,000 rpm for 15 min. The pellet was collected and air dried at room temperature for 5 min. The pellet was resuspened with 10 µl of MegaBACE loading solution (70% formamide, 1 mM EDTA), vortexed vigorously for 10 sec and centrifuged briefly. Finally, the suspension was analyzed using the MegaBACE 500 sequencer [125, 126].

4.2.5 5 bp insertion/deletion determination by denaturing polyacrylamide gel electrophoresis and silver staining.

For the 5 bp in/del determination in intron 4 of major transcript, primers for PCR amplification reaction were designed around the insertion/deletion region with expected size of 296 bp (Table 5). PCR reaction was composed of 100 ng of genomic DNA, 0.2 mM dNTPs, 0.2 µM each primer for each marker and 1 Unit of *Taq* DNA polymerase in 1X PCR buffer (Tris-HCl, KCl, $(NH_4)_2SO_4$, 15 mM $MgCl_2$; pH 8.7) and 1.5 mM $MgCl_2$.. After optimization, PCR profile was set at denaturation 94ºC for 45 sec, annealing 50 ºC for 30 sec, extension 72 ºC for 45 sec.

Amplified PCR product was run on 1.5 % agarose gel electrophoresis to confirm the amplification reaction and to screen the size of expected PCR product in all 30 samples. To determine the 5 bp insertion/deletion polymorphism, 5% denaturing polyacrylamide gel followed by silver staining was used. Two glass plates, short and long, were used to prepare the gel. The shorter plate was covered with Rain Shield and the longer plate was shielded with Bind Silane for allowing the gel to fix with the

glass, then 0.5 mm spacers was placed between two plates. The 5% denaturing polyacrylamide gel were prepared by mixing of 5 ml of 40% acrylamide stock solution, 18 g of urea, 4 ml of 10X TBE buffer and sterile distilled water to obtain the total volume of 40 ml followed by the adding of 250 µl of 10% ammonium persulfate and 30 µl of TEMED. The solution was mixed and immediately poured in assembled plates. After gel polymerization for 2 hours, the gel plate was attached to the electrophoresis status and 1X TBE buffer was filled in the upper and the lower reservoirs. Before running samples, the gel was pre-run at 30 mA, 50 W, 1,500 V for 30 min.

The amplified PCR product of 30 samples were prepared by mixing the equal volume of 3 ng of PCR product and 98% formamide loading dye. The samples were denatured at 95^{o}C for 2 min. and loaded immediately. The electrophoresis running was set at 30 mA, 50 W, 1,500 V for 2 h 30 min.

After electrophoresis, the shorter plate was removed out from the longer one. The gel which bound to the longer plate was fixed with 1.5 L of 15% acetic acid for 20 min. In the same period, two liters of silver solution was freshly prepared by mixing 0.1% silver nitrate with 3 ml formaldehyde. After fixing with acetic acid, the gel was washed three times with 1 L of distilled water. The gel was stained for 30 min with gentle shaking. After staining, the gel was rinsed with 1 L of distilled water. Finally, the band was visualized by the developer solution, which already prepared and stored at –20^{o}C. Before adding the developer solution to the gel, the solution was mixed with formaldehyde and of sodium thiosulfate. The gel was soaked with developer. When the first band was developed, the gel was transferred to another container, submerged in the other developer solution.. Until every band of all sample including marker was visualized, 15% acetic acid was added to stop the reaction.

4.2.6 Hardy-Weinberg Equilibrium (HWE)

This test is useful to ensure that there is no (or little) population stratification or different arrangements and that each SNP is giving the expected genotype distribution for the observed allele frequencies. Expected genotype frequencies are calculated from allele frequencies under the assumption $p^2 + 2pq + q^2$

= 1, where p and q are the allele frequencies and p^2, q^2 and 2pq correspond to the frequencies of the three possible genotypic states. Any deviation from HWE is warrant for genotype rechecking because it may indicate inbreeding, population stratification, genotyping errors and functional SNPs under selection [127].

4.2.7 Haplotype analysis and linkage disequilibrium.

Haplotype analyses and linkage disequilibrium of discovered SNPs were performed by using Haploview software Daly lab, cambridge USA.

4.2.8 Transcription factor binding site analysis.

Transcription factor binding site in 1 kb 5' of HbS1L gene was analyzed by Gene2promoter programme on www.genomatix.com .

4.2.9 Genotyping of SNP 7 by PCR-RFLP method.

Single nucleotide polymorphism C→T (SNP 7) was discovered at 5' untranslated region located within the CREB (cAMP response element binding protein) transcription factor binding site, which also had sequence of *ApaL1* restriction enzyme site (G↓TGCAC). Transversion of C to T abolishes *ApaL1* restriction site. Primers of exon one HBS1L 1F and HBS1L 1R, which cover the site of SNP 7 were used for PCR amplification reaction in 545 DNA samples including 180 mild, 295 severe and 70 cases of β-thal / Hb E coinherited α-thal. PCR reaction of each sample was performed in a total volume of 25µl. The reactions composed of 100 to 150 ng of genomic DNA, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 0.2 µM each primer and 1 Unit of *Taq* DNA polymerase in 1X PCR buffer (Tris-HCl, KCl, $(NH_4)_2SO_4$, 15 mM $MgCl_2$; pH 8.7). The reaction was amplified for 35 cycles with denaturation at 95 °C for 45 sec, annealing at 54 °C for 30 sec and extension for 72 °C for 45sec. The 456 bp product was subsequently subjected to restriction enzyme digestion for RFLP study.

Total volume of digestion reaction per sample was composed of 1 µl ApaL1 restriction enzyme, 1 µl of 1X NE buffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1mM dithiothreitol, pH 7.9, 0.1 µl of 100x BSA, 1 µl PCR product and 6.9 µl of sterile distilled water. Digestion reaction was incubated

in water bath at 37°C for two hours. Restriction fragment length polymorphism (RFLPs) was detected on 1.5 % agarose gel electrophoresis.

4.2.10 Statistical Analyses of SNP 7.

Comparison of genotypic results with the level of HbF was conducted with the implementation of the Statistical Package for the Social Science (SPSS) version 14 (SPSS inc.). Allele and genotype frequencies distribution of SNP 7 variants was compared with Hb F distribution in severe and mild cases and β-thalassemia co-inherited with α-thalassemia. Scattergrams were created by using SPSS program to observe the Hb, Hb F and Cor% F distribution in SNP 7 genotype and *Xmn*1 genotype individually and in combination.

A commonly employed statistical test that can assist in measuring any correlation that may exist is the Chi- Square (χ^2) [128]. This procedure ultimately assesses difference in proportions of categories for cross table (usually two), specifically comparing observed frequencies of each category in the study against what is expected to occur. The χ^2 statistic formula shown below is a measure of the deviation of observed values from expected, with a large chi-square score, indicating a large deviation in frequencies.

$$\chi^2 = \Sigma \frac{(\text{observed value} - \text{Expected value})^2}{(\text{Expected Value})}$$

If the χ^2 is significant at a pre-defined alpha level (usually 0.05) it is said that the two variables are not independent from each other or that they are dependent (associated) with each other [128].

Statistically used standard contingency table analysis incorporating the chi-square was employed to measure any correlation. The level of significance was set at $P < 0.05$.

Table 1. Criteria and scoring system for classification of β-thalassemia patients.

	Mild	Score	Intermedia	Score	Severe	Score
- Hb* (g/dl)	>7	0	6-7	1	<6	2
- Age onset (yr)	>10	0	2-10	0.5	<2	1
- Age at 1st BlTx (yr)	>10	0	4-10	1	<4	2
- Requirement BlTx	None/Rare	0	Occasional	1	Regular	2
- Size of spleen (cm)	<4	0	4-10	1	>10, s/p	2
- Growth retardation	-	0	+/-	0.5	+	1

*Hb (g/dl) = Hemoglobin at steady-state

Mild case	Severity Score	<3.5
Intermediate case	Severity Score	3.5-7.5
Severe case	Severity Score	>7.5

Table 2. *Xmn*I and Hb F status of 30 β-thalassemia/HbE cases used for sequencing purpose.

Phenotype	***Xmn*I**	**Hb F status**	**Number of Samples**
Severe.	+/-	Low	6
Mild.	+/-	High	6
Severe.	-/-	Low	4
Mild.	-/-	Low	4
Severe.	-/-	High	4
Mild.	-/-	High	4
Mild.	+/+	High	2

Total number of 30 samples for DNA sequencing.14 Severe, 16 Mild.

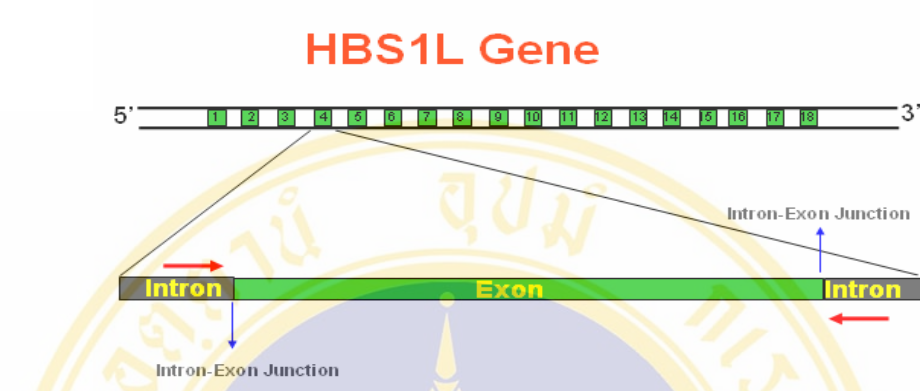


Figure 7a. Primer design method for each exon of major transcript of *HBS1L* gene encompass intron-exon junction of 18 exons.

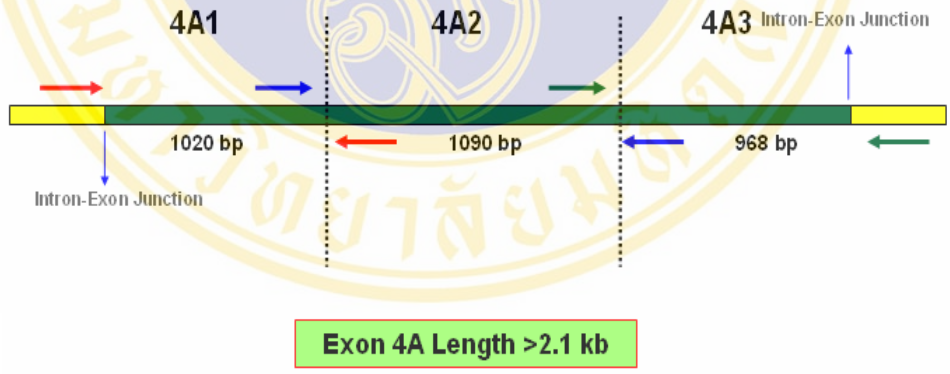


Figure 7b.Location of primers and size of PCR product of alternate exon 4A of *HBS1L* gene**.**

Table 3.Primer sequences and $MgCl_2$ conc. used for individual PCR reaction.

Exon	Primer sequence (5'- 3')		Length nt	PCR Product Size (bp)	Annealing Tempt.ºC	$MgCl_2$ (mM)
HBS1L**1**F	5' ATTCTAGGCGGCGGATTTC	3'	19	465	54	1.5
HBS1L**1**R	5' GGCAACACTGACGAGAAAC	3'	19			
HBS1L**2**F	5' ACTGAAGAACTGTGTATGTC	3	20	526	53	2.0
HBS1L**2**R	5' GCAAAGTATGGTGGAAAAGA	3'	20			
HBS1L**3**F	5' TAGGCTGAGTGGAATGGTT	3'	19	554	52	1.5
HBS1L**3**R	5' ACACTCCTCACAAAACATAC	3'	20			
HBS1L**4**F	5' TGCCAAAGAGAAGGGAGAG	3'	19	354	52	2.0
HBS1L**4**R	5' CTCAAGCAAACGAAAGCCT	3'	19			
HBS1L**5**F	5' AGGAAGAGATTTGGTGGA	3'	18	532	49	1.5
HBS1L**5**R	5' GCCAGAATCAAGCCTAAT	3'	18			
HBS1L**6**F	5' TCTCTGCCCTTCTGTCATTG	3'	20	415	56	1.5
HBS1L**6**R	5' TGCCTCCTTGCTCTCTGTC	3'	19			
HBS1L**7**F	5' ATTGGCGTGTCTTGTTG	3'	17	545	47	1.5
HBS1L**7**R	5' AGTCAAAATAGTTCCCTG	3'	18			
HBS1L**8**F	5' TCTTACTTTGTCCTCCCTT	3'	19	612	51	2.5
HBS1L**8**R	5' CTCATCAAGTCCTTTACCAT	3'	20			
HBS1L**9**F	5' CTGGTTGCTGTTTGTTTATG	3'	20	530	53	2.0
HBS1L**9**R	5' GGCAACCGACTAATAAAATG	3'	20			
HBS1L**10**F	5' GCAAGGTCTGATATTCCAT	3'	19	497	50	1.5
HBS1L**10**R	5' ACACTCCTCTATGACCTC	3'	18			
HBS1L**11**F	5' GGAGGTCATAGAGGAGTGT	3'	19	502	54	1.5
HBS1L**11**R	5' TCCGACTACAAATCCTATGC	3'	20			
HBS1L**12**F	5' TGAGCATTTCCTTTGAGCAT	3'	20	532	52	1.5
HBS1L**12**R	5' GAAAGATGAGAAATGGTGAG	3'	20			
HBS1L**13**F	5' AAGTAAAGTGTGGTATGCC	3'	19	472	51	2.0
HBS1L**13**R	5' CAGACACAATCACAACAGA	3'	19			
HBS1L**14**F	5' GAAGCACAAATCAGCCAAAC	3'	20	201	54	1.5
HBS1L**14**R	5' ATGTGTTTAGAGGTGAGGTC	3'	20			

Table 3. Primer sequences and $MgCl_2$ Conc. used for individual PCR reaction

Exon	Primer sequence (5'- 3')		Length nt	PCR Product Size (bp)	Annealing Tempt.ºC	$MgCl_2$ (mM)
HBS1L**15**F	5' TCAGCACACATACTTGGCAT	3'	20	532	54	1.5
HBS1L**15**R	5' AAGTCCTCAGGCTCAAAGAT	3'	20			
HBS1L**16**F	5'TAGAGGAAGCAACTGGAGAT	3'	20	461	55	1.5
HBS1L**16**R	5'CACCACCACACATAACAGAA	3'	20			
HBS1L**17**F	5'CCGTAAGGGATAGAGTT	3'	17	527	47	1.5
HBS1L**17**R	5'GGCTTCTCCTAATCAATG	3'	18			
HBS1L**18**F	5'GGAAAGGAAGAAAGTTGC	3'	18	1076	49	1.5
HBS1L**18**R	5'CTCTTACTTACTCACAGG	3'	18			
Exon 4A						
HBS1L**4.1**F	5'GAGACAATCCTGACTAC	3'	17	1020	47	1.5
HBS1L**4.1**R	5'GATGGACTTTCACACTG	3'	17			
HBS1L**4.2**F	5'TACAGGACAGTTTAGGAAGT	3'	20	1090	52	1.5
HBS1L**4.2**R	5'CTAATCAACTGACCTACGAC	3'	20			
HBS1L**4.3**F	5'CTGTGTCTTCGTTACCCACT	3'	20	968	56	1.5
HBS1L**4.3**R	5'AGCAGCAGCAGCAGCAGAA	3'	19			
Primer for 5' 1kb.						
F	5'CATTCAGCATTACCACCTA	3'	19	1114	51	2.0
R	5'AACTCCTTCCAAAACACTC	3'	19			

Table 4 Primer sequences and alignment used for sequencing exons of major transcript of *HBS1L* gene

Exon	Primer alignment	Primer sequence	Length "nt"
HBS1L**1**	R	5'GGCAACACTGACGAGAAAC 3'	19
HBS1L**2**	F	5'ACTGAAGAACTGTGTATGTC 3	20
HBS1L**3**	R	5'ACACTCCTCACAAAACATAC 3'	20
HBS1L**4**	F	5'TGCCAAAGAGAAGGGAGAG 3'	19
HBS1L**5**	R	5'GCCAGAATCAAGCCTAAT 3'	18
HBS1L**6**	R	5' TGCCTCCTTGCTCTCTGTC 3'	19
HBS1L**7**	R	5'AGTCAAAATAGTTCCCTG 3'	18
HBS1L**8**	R	5'CTCATCAAGTCCTTTACCAT 3'	20
HBS1L**9**	F	5'GGGTAGTTAGGGTCTTCC 3'	18
HBS1L**10**	F	5'GCAAGGTCTGATATTCCAT 3'	19
HBS1L**11**	F	5'GGAGGTCATAGAGGAGTGT 3'	19
HBS1L**12**	F	5'TGAGCATTTCCTTTGAGCAT 3'	20
HBS1L**13**	F	5'AAGTAAAGTGTGGTATGCC 3'	19
HBS1L**14**	F	5' GAAGCACAAATCAGCCAAAC3'	20
HBS1L**15**	F	5'TCAGCACACATACTTGGCAT 3'	20
HBS1L**16**	R	5'CACCACCACACATAACAGAA 3'	20
HBS1L**17**	F	5'CCGTAAGGGATAGAGTT 3'	17
HBS1L**18**	1.F	5'CGTTTCTGGATACAGTGA 3'	18
	2.R	5'TGGGAAAAGCAAGTTATGG 3'	19
	3.R	5'AGGTCAAATATCAACTGCC 3'	19

Table 4. Primer sequences and alignment used for sequencing 4A alternate exon and 5' I Kb region of *HBS1L* gene.

Exon	Primer alignment	Primer sequence	Length "nt"
Exon 4A.			
HBS1L**4.1**	F	5'GAGACAATCCTGACTAC 3'	17
HBS1L**4.1**	R	5'GATGGACTTTCACACTG 3'	17
HBS1L**4.2**	F	5'TACAGGACAGTTTAGGAAGT 3'	20
HBS1L**4.2**	R	5'CTAATCAACTGACCTACGAC 3'	20
HBS1L**4.3**	F	5'CTGTGTCTTCGTTACCCACT 3'	20
HBS1L**4.3**	R	5'AGCAGCAGCAGCAGCAGAA 3'	19
HBS1L**4.3**	F	5'CTTCTTTCAGTGATAACTGGC3'	21
Primer for 5' 1kb.			
	HBS1LF	5'CATTCAGCATTACCACCTA 3'	19
	HBS1LR	5'AACTCCTTCCAAAACACTC' 3'	19
	HBS1LR	5'TGGCTCCTTCCTGTAATCC 3'	19

Table 5. Primer sequence and PCR condition for determination of 5 bp insertion/deletion polymorphism in intron 4.

Primer sequence		nt	PCR Product Size (bp)	Annealing temp (°C)	$MgCl_2$ (mM)
F	5'CCTAGATTAAAGTGACAGA3'	19	296	50	1.5
R	5'CAGGCACTTCAAATCCCA3'	18			

CHAPTER 5
RESULTS

5.1. DNA concentration and purification.

Genomic DNA integrity, concentration and purification of 30 samples were determined by agarose gel electrophoresis and UV spectrophotometry (Method section 4.2.1). The results showed that DNA concentration of these samples ranged from 122.5 ng/µl to 1227.5 ng/µl, which was adequate for process in the next steps. The purity of DNA was also determined by OD_{260}/OD_{280} ratio and it was found that the ratio in all DNA preparations was in the range of 1.50–1.92. Agarose gel electrophoresis also demonstrated that these DNAs had high molecular weight without degradation as demonstrated by (Fig 8).

5.2. Amplification of all exons and 1 kb promoter region by PCR.

Total 30 genomic DNA samples, 14 from severe and 16 from mild cases, were used as DNA templates for amplification of *HBS1L* gene specially for all exons, exon-intron junctions and 1 kb promoter region. The PCR reactions and conditions used for amplification of these fragments were described in Method section**.** After amplification 1 to 2 µl of PCR product was analyzed by the 1.5% agarose gel electrophoresis. The amplified products were visualized under UV exposure after the gel was stained with ethidium bromide (Method section 4.2.2). Various PCR profiles were tried to optimize the amplification of the fragment of exon 4A3. Despite many attempts none of the PCR profile worked. Primers were redesigned but the problem persists. After observation of DNA sequence of 4A3 fragment, it was noticed that the amplifying region contained an approximately 40 “A” sequence region in the intron close to exon 4A, which could have been the reason for PCR failure. Primers closer to the exon and excluding the poly A region were redesigned, and finally specific PCR product band was successfully visualized on agarose gel. The results showed that all exons and 1 kb promoter region were successfully amplified from genomic DNA

samples, generating specific PCR products matched to the expected sizes and were in the sufficient amount for sequencing process after purification (Fig 9)

5.3 Sequence determination of PCR products of all exons and 1 kb promoter region of HBS1L gene.

The amount of PCR products obtained from each exon and 1kb promoter region was estimated after the agarose gel electrophoresis by comparing with the band intensities of the PCR products loaded on each lane with the known amount of the DNA markers (Method section 4.2.4). It was found that the concentration of PCR products varied from 15 ng/µl to 80 ng/µl depending on the amplified sample. Prior to determination of DNA sequence, PCR products were adjusted to have the final concentration between 20 to 50 ng/µl by dilution with distil water.

The sequences of PCR products derived from 30 β-thalassemic individuals (14 severe and 16 mild) were analyzed by MegaBACE 500 DNA Analysis System. Prior to analysis, the samples were subjected for one more PCR using DYEnamic ET Dye Terminator Cycle Sequencing Kit (Method Section 4.2.4). The raw data was obtained and processed by Genetic Profiler program to give the DNA sequence of each sample as electrophoregram. The results revealed 22 SNPs in *HBS1L* exons including exon -intron junction, out of which, 5 SNPs were new. Physical position of 22 SNP is demonstrated in Fig 10. Chromatogram example of each SNP are shown in Fig 11. Genotype of all SNPs discovered during sequencing process in 30 cases is shown in Table 6.

Moreover, frequency of SNPs was calculated and NCBI database was used to compare the frequency of discovered SNPs against the frequency of discovered SNPs in other population. Results shows that out of 22 SNPs, only 11 SNPs have been studied in other populations and frequency of Thai SNPs was very close to frequency of Chinese SNPs **(**Table 7).

Furthermore, out of 22 SNPs, 3 SNPs were discovered on exon regions i.e. exon 4A and exon 17. Verification of these SNPs status was performed and it was found that one out of the two SNPs on exon 4A at the codon position of 200 was nonsynonimous, changing amino acid glutamate to glycine, while the other SNPs on exon 4A and 17 were synonymous (Table 8).

5.4. Functional significance of SNPs in the promoter region

Out of 22 discovered SNPs, 7 SNPs were located on promoter region and 5' untranslated region. Some of these SNPs were positioned on important transcription factor binding sites such as VDR (vitamin D receptor), ELF2, GATA and ETSF, Hox 1, CREB (cAMP response element binding protein) as shown in Table 9.

5.5. 5 bp insertion/deletion in intron 4.

Beside 22 SNPs, 5 bp insertion/deletion (in/del) polymorphism was discovered in the intron 4 of the major transcript close to the exon 5 of HBS1L gene. A 5 bp "AATTA" (insertion/deletion) polymorphism was flanked by SNP 11 and SNP 12. An example of the chromatogram of 5 bp in/del polymorphism is shown in Fig 12.

However, it was difficult to differentiate between heterozygous and homozygous deletions from the chromatogram in many samples. In order to confirm the heterozygosity and homozygosity of in/del polymorphism in all samples, the amplified DNAs which encompass the 5 bp deletion area with the expected PCR size of 296 bp were performed and PCR products were run on denaturing polyacrylamide gel electrophoresis (5%) followed by silver staining (Fig 13). The result showed that, out of 30 samples, 13 samples had heterozygous 5 bp in/del polymorphism and 7 samples had homozygous in/del polymorphism while 10 samples had no insertion/ deletion polymorphism as shown in Table 10. Although this polymorphism does not seem to correlate with the amount of Abs Hb F. However, it appears that the heterozygous polymorphism is associated with XmnI -/- (69% of cases).

5.6. Hardy-Weinberg Equilibrium (HWE)

Test for Hardy-Weinberg equilibrium (WHE) in all discovered SNPs during sequencing was performed by p^2 2pq q^2 conventional method in order to determine whether our discovered SNPs in *HBS1L* exhibit HWE. The results showed that value of HWE ranged from 0.6 to 1.3 giving genotype frequencies mostly within expected range signifying that there was a little deviation in the frequencies of the expected genotype in discovered SNPs

The use of SNP combined with relatively small sample population size for association studies presented difficulties for testing HWE. An important assumption in

testing HWE is that the population is large enough to allow the accurate estimation of allele frequency. This result indicated that when a significantly small population size is used, the likelihood of producing unreliable or incurring a false rejection of HWE is significantly increased.

5.7. Haplotype Analysis and Linkage Disequilibrium

A haplotype is a string of consecutive SNPs lying on the same chromosome. After confirming all 22 SNPs discovered by direct sequencing step, haplotype analysis was performed in order to observe the haplotype blocks and the SNPs in linkage disequilibrium. Haplotype analysis shows two haplotype blocks, first comprised of 10 SNPs starting from SNP 3 to SNP 12 and the other block started from SNP 17 and SNP 19 (Fig 14). Out of these two SNP tags, SNPs 3, 4, 6, 7, 8, 11 and 12 SNPS were in strong linkage disequilibrium, having D' value = 1.0 , r^2 = 1.0 and D' confidence interval ranging between 0.9 -1.0. Besides, SNPs 5, 9, 10, 17, 18 and 19 were loosely in linkage disequilibrium having D' value =1, ranging from 0.11 to 6.77 and D' confidence values from 0.38 to 1.0. Moreover, SNPs 1, 2, 13, 14, 15, 16, 21 and 22 provided the least values and were not in linkage disequilibrium.

5.8. SNP 7 Genotyping

Although SNPs finding can be discovered basically by DNA sequencing, the technique is costly and time-consuming for genotyping of every samples. Among the SNPs in linkages disequilibrium, the C to T substitution in SNP 7 in 5' untranslated region abolishes the restriction site for *Apa*LI enzyme. Thus all, 547 samples, 297 samples were severe β-thalassemia cases, 180 samples were mild β-thalassemia cases and 70 samples of β-thalassemia cases co inherited with α-thalassemia were included for genotyping of SNP 7 in this study. The samples were amplified and digest with *Apa*L1 to distinguish the alleles C (cleavable by *Apa*LI enzyme into 158 bp and 326 bp) and T (not cleavable by *Apa*LI enzyme). Firstly, 1 µl of PCR products of 10 previously sequenced samples were subjected for restriction enzyme digestion reaction to verify the specificity of restriction enzyme. The result revealed that the PCR products of these 10 sequenced samples, whose genotypes were already known from sequencing step, were in the expected sizes on agarose gel electrophoresis, signifying

that the enzyme digested product was specific without nonspecificity (Fig 15). After confirming specificity of enzyme with control samples, genotyping of 547 samples was performed as mention in Method section 4.2.9. Hardy-Weinberg Equilibrium (HWE) for severe, mild and β-thalassemia / Hb E co-inherited with α thalassemia (B/A/T) was calculated and the ranged from 1.02 to 1.09 giving p value of 0.685 in mild and 0.780 demonstrated that genotype frequencies were not within the expected ranges (Table 11).

The results also showed that the frequency of homozygous C/C genotype was higher in mild cases as compared to severe cases whereas the frequency of T/T genotype was higher in severe cases as compared to mild cases with the p-value 0.002 (Table 11).

The frequency of C and T alleles was also calculated, which showed that the frequency of C allele was considerably higher in mild cases than that of severe cases. While frequency of T allele was significantly higher in severe cases as compared to mild cases (p value 0.002) (Table 11).

5.9 Statistical Analysis for the effect of *Xmn*I and SNP 7 polymorphism to the levels of Hb, Abs F and Cor% F.

All together 486 severe and mild cases with β-thalassemia/Hb E disease without co-inherited α-thalassemia were subjects for associated study of SNP 7 and Hb F level. Scattergram charts were created using SPSS 14 statistic program to observe the distribution of Hb, Hb F and correct percentage of Hb F (Cor% F) levels in each genotype of SNP 7 and *Xmn*I polymorphisms and sex (Figure 16-19). The results showed that C/C genotype of SNP 7 had a slightly higher Hb, HbF and Cor% F levels, while C/T and T/T genotype did not show any significant difference in their values (Fig 16). Statistical analysis revealed the significant effect of SNP 7 to the levels of Hb, Abs F and Cor% F with p values of 0.014, 0.007 and 0.007 respectively.

The results also showed that gender do not contribute to the effect of SNP 7 (Fig 17-18). The effect of the level of Hb, Hb F and Cor% F was more pronounced with the *Xmn*I polymorphism, p values were less than 0.00002 in all parameters (Fig 19).

Comparative scattergram charts were also created to analyze the distribution of Hb, Abs.F and Cor% F in *Xmn*I +/+, +/-, -/- with different genotypes of SNP 7 (Fig 20-22**)**. The results suggested that patients who possess *Xmn*I -/- genotype with SNP 7 T/T genotype have significantly lower level of Abs.F and Cor% F (p Value 0.017) while patients with *Xmn*I +/+ genotype with SNP 7 T/T, compared to SNP 7 C/C do not show any change on Abs F and Cor% F level (p-value 0.810) (Table 12). The results signified that *Xmn*I is a stronger modulator factor than SNP 7 T/T genotype.

The amounts of average Abs.F was compared in individuals who have different genotypes of SNP 7 and *Xmn*I polymorphism (Table 12). It was found that those subjects who possess *Xmn*I +/+ genotype in combination with SNP 7 C/C genotype had a highest average Abs. F level (3.28 ± 1.35 g/dl) while cases who acquired *Xmn*I -/- genotype with SNP 7 T/T genotype had the lower level of Abs.F (1.56 ± 0.86 g/dl).

Furthermore the data confirmed that the effect of *Xmn*I polymorphism on the level of Abs F was more pronounced than that of SNP 7, as observed in patients with SNP T/T, there are higher Abs F level in cases with *Xmn*I +/+ and lower in cases with *Xmn*I -/- (3.12 ± 1.40 vs 1.56 ± 0.86, p value = $1^* 10^{-5}$), and in cases with *Xmn*I +/+ with SNP 7 CC or TT were not significantly different in Abs F level (3.28 ± 1.35 vs 3.12 ± 1.40, p value = 0.810). However, SNP 7 C/C genotype may have some effects on the expression of Hb F as the increased amount of Abs F were observed in cases who were CC/-- vs TT/-- (2.24 ± 1.39 g/dl vs 1.56 ± 0.86 g/dl, p value = 0.017). Moreover, SNP 7 C/C genotype showed a non-significant different level of Abs F in cases who had *Xmn*I ++ and -- (3.28 ± 1.35 g/dl vs 2.24 ± 1.39 g/dl, p value = 0.096) (Table 12).

Co-inheritance of *Xmn*I +/+, +/-,-/- with SNP 7 C/C, C/T ,T/T genotypes in mild and severe cases of β-thalassemia / Hb E patients was analysed to classify their contribution to the severity of the patients (Table 13). The results showed that the number of patients having SNP 7 C/C and *Xmn*I +/+ genotype was higher in mild cases as compared to severe cases (ratio of mild and severe cases is 3.5). While those having SNP 7 T/T with *Xmn*I-/- genotype were significantly higher in severe cases as

compared to mild cases (ratio of mild and severe cases is 0.3). The results also confirmed the more pronounced effect of *Xmn*I to the severity of the patients.

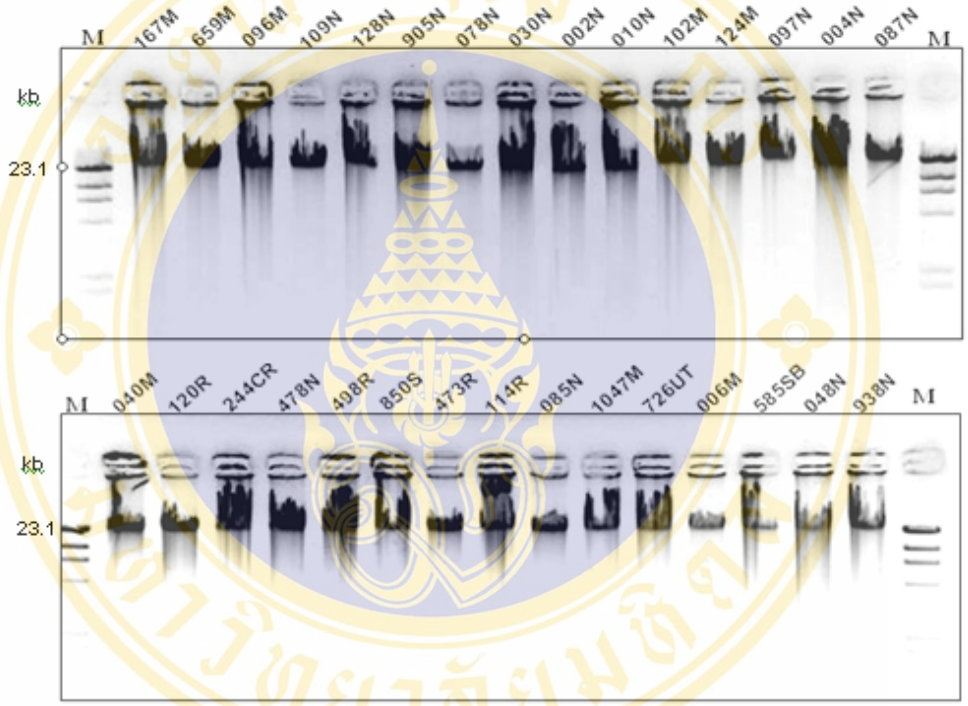


Figure 8. Agarose gel electrophoresis showing the high molecular weight DNA of 30 genomic DNA samples.

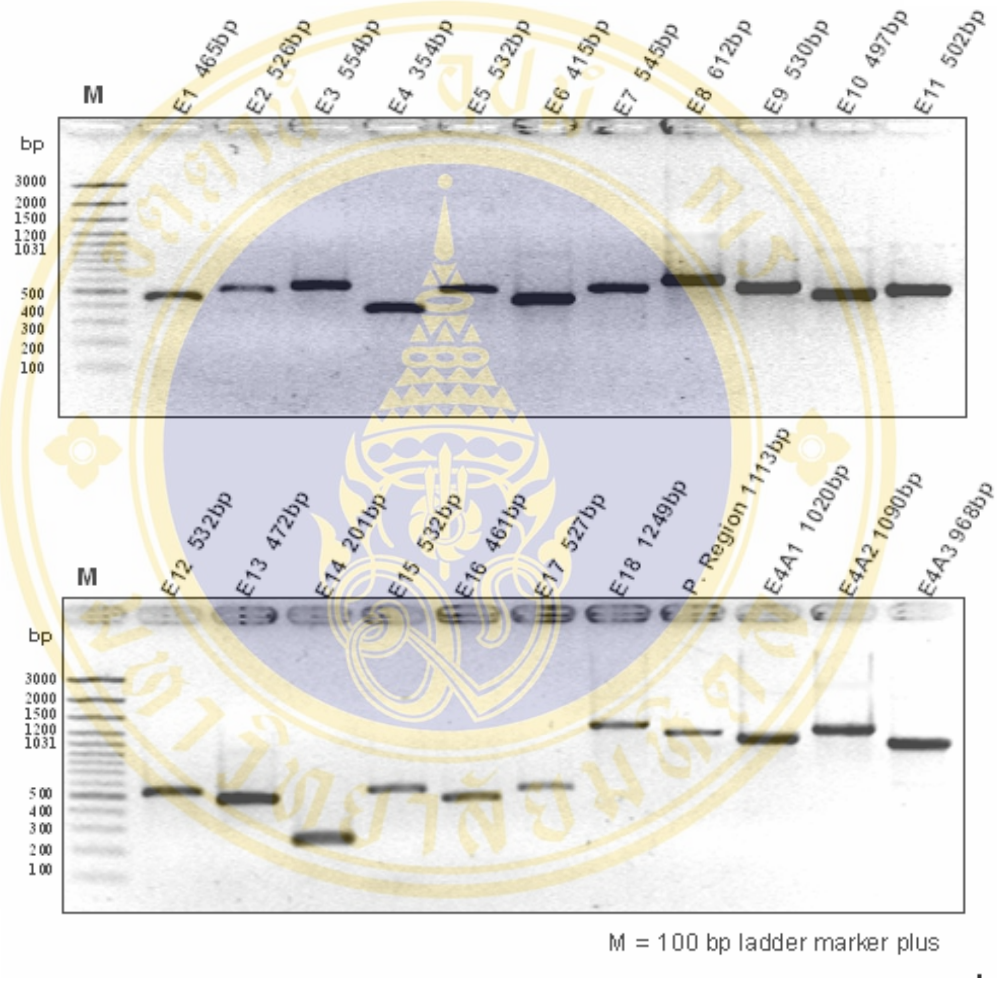


Figure 9. An example of agarose gel electrophoresis of all PCR amplified fragments and purified from exons,1 kb promoter region and all three reactions of alternate exon 4A.

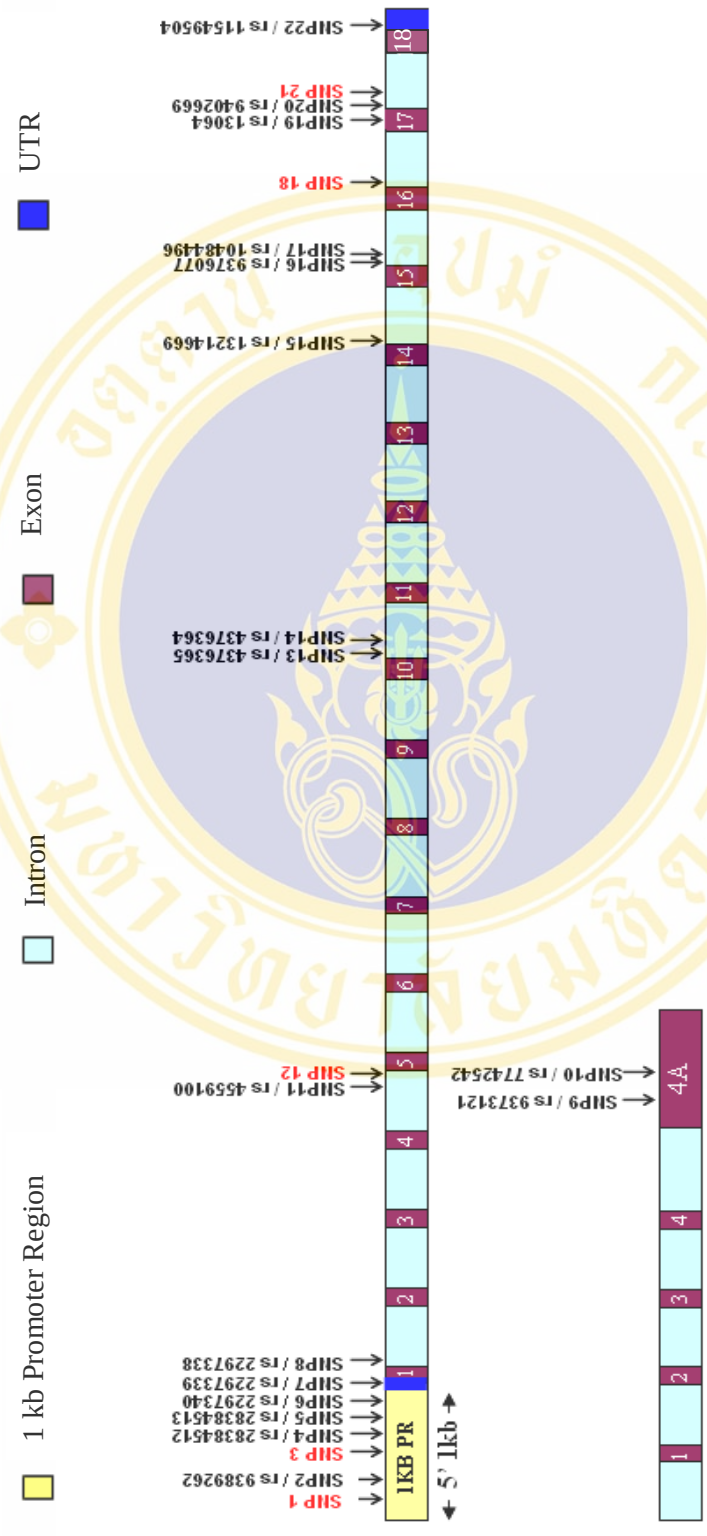


Figure 10. Physical position of SNPs on *HBS1L* gene revealed by resequencing.

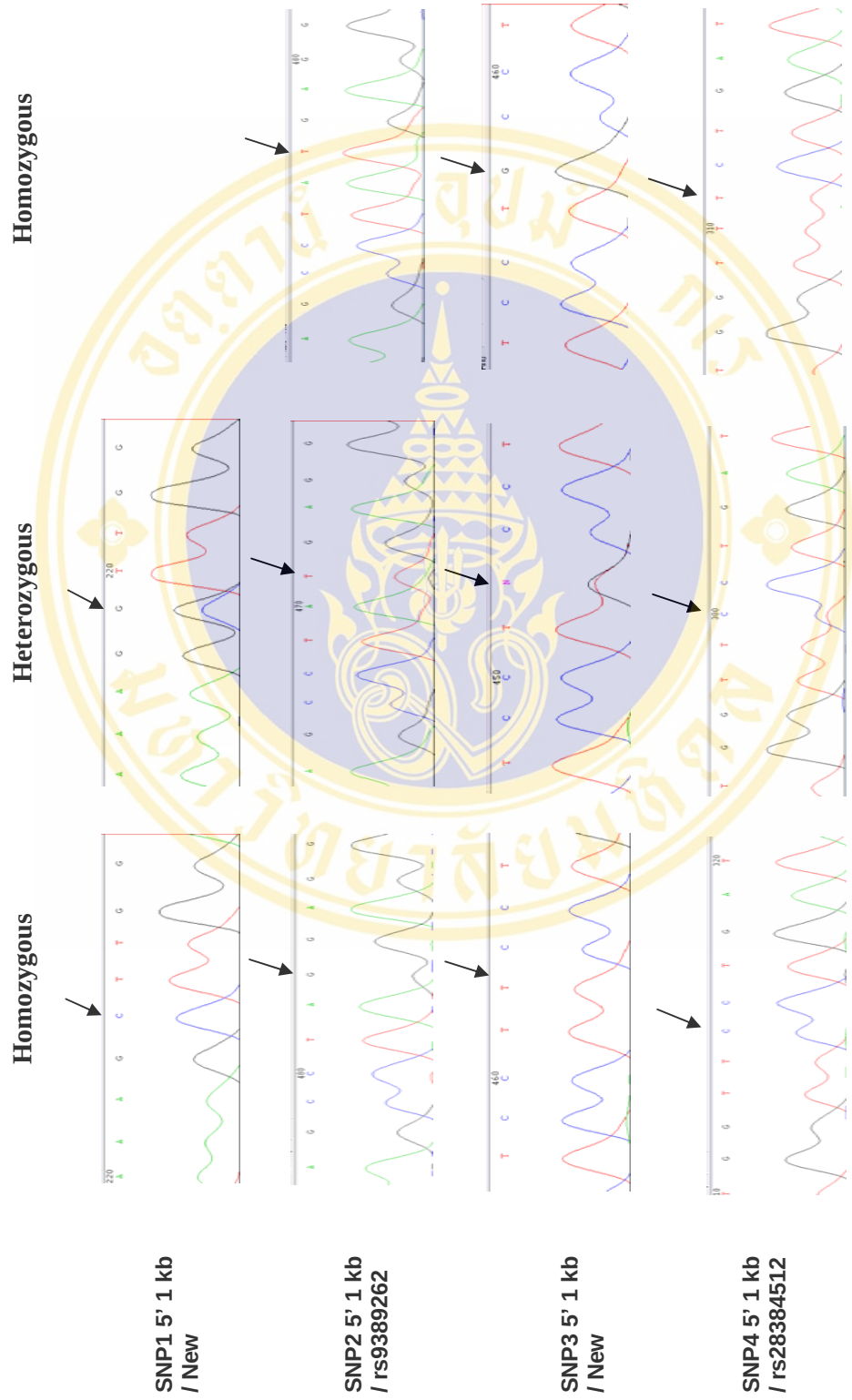


Figure 11. Chromatogram of SNP discovered by resequencing.

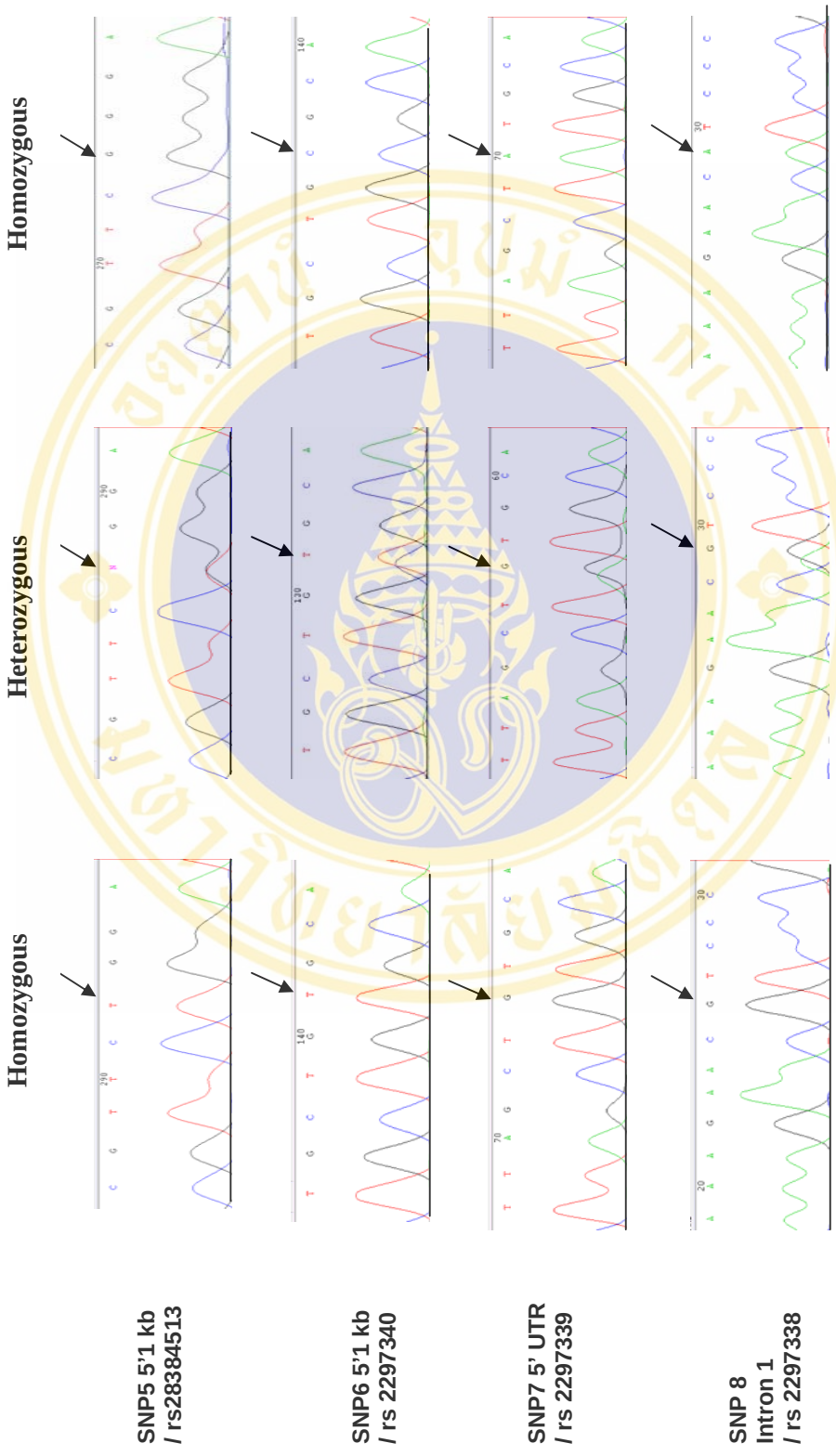


Figure 11. Chromatogram of SNP discovered by resequencing. (Cont'd)

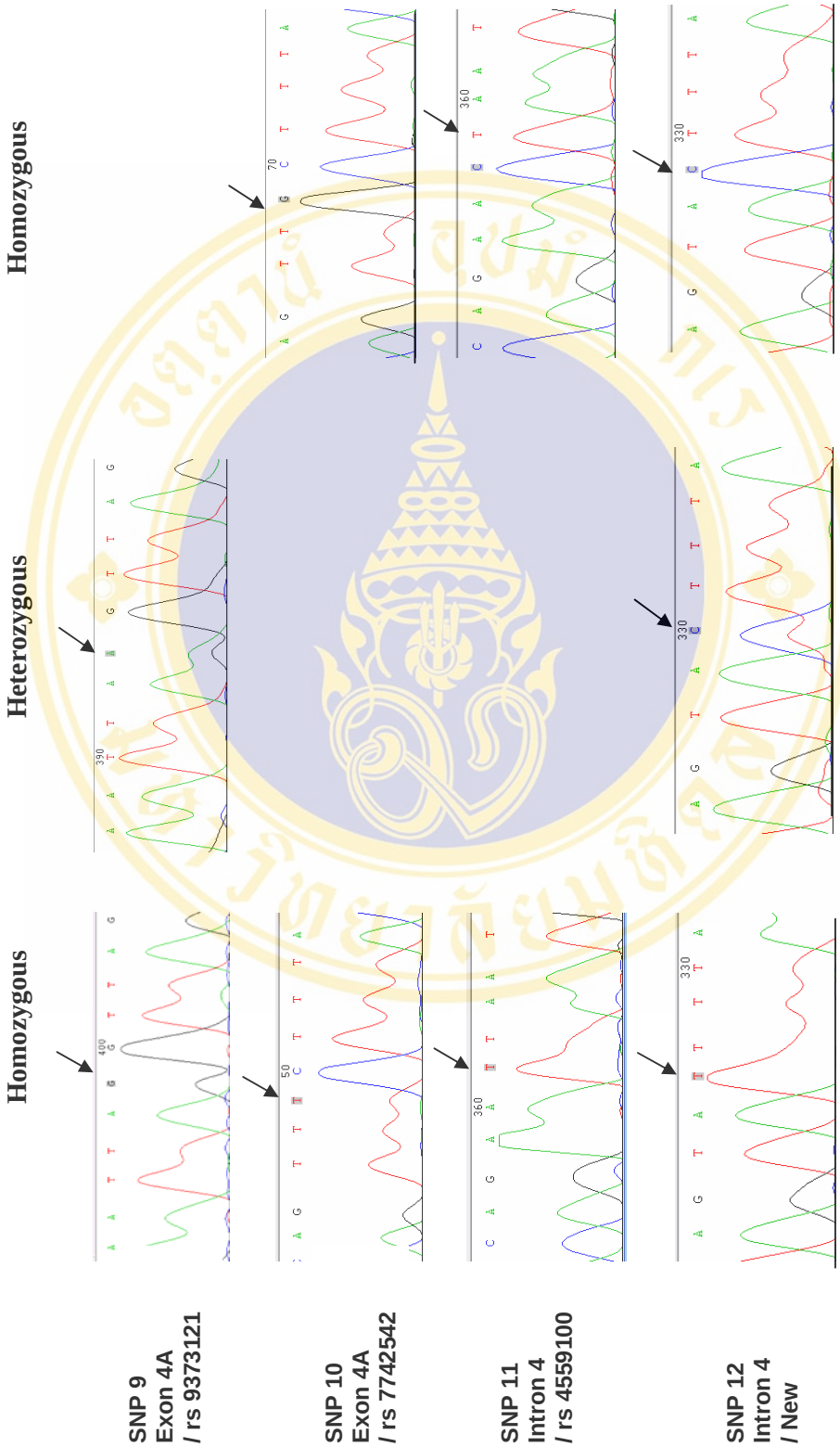


Figure 11. Chromatogram of SNP discovered by resequencing. (Cont'd)

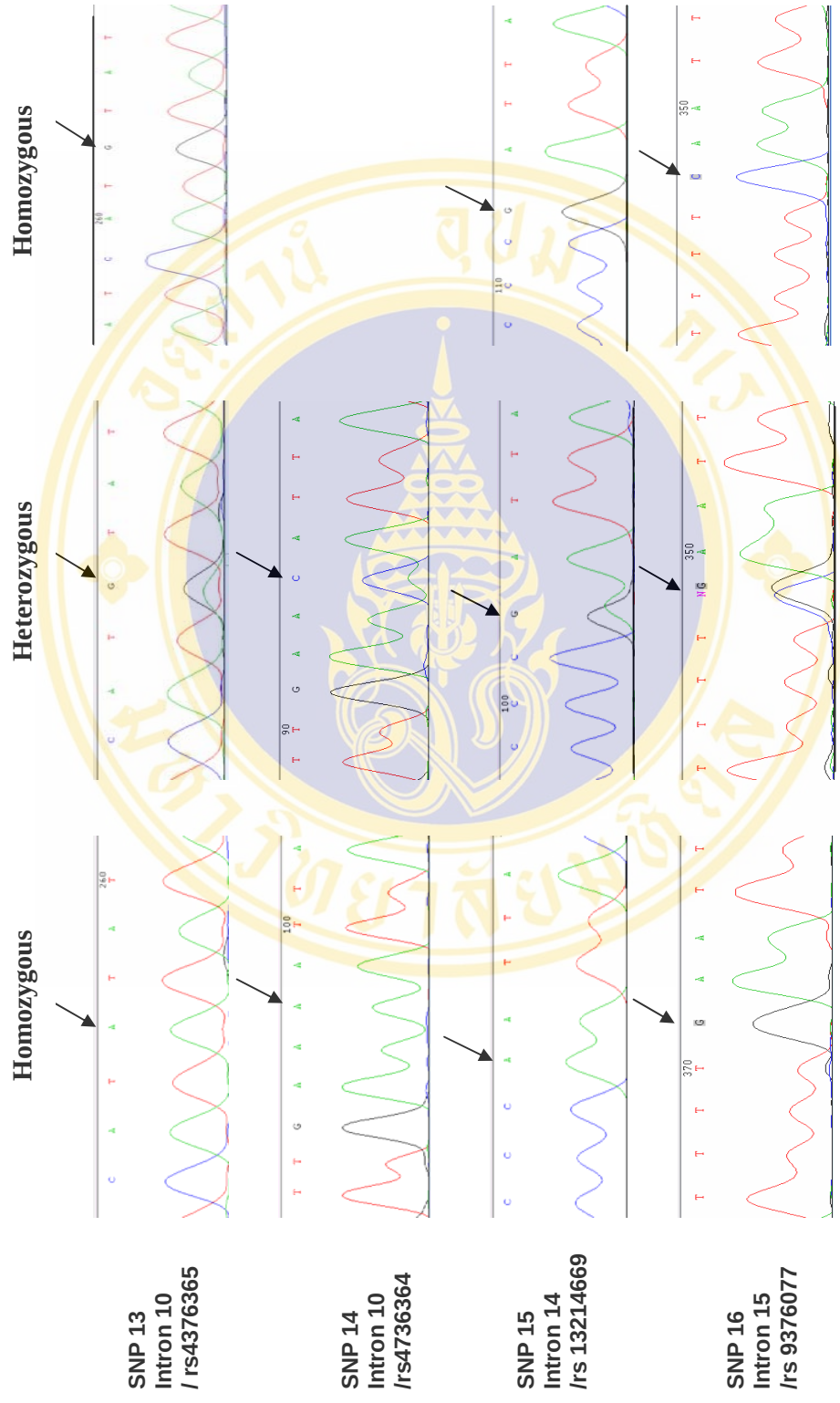


Figure 11. Chromatogram of SNP discovered by resequencing. (Cont'd)

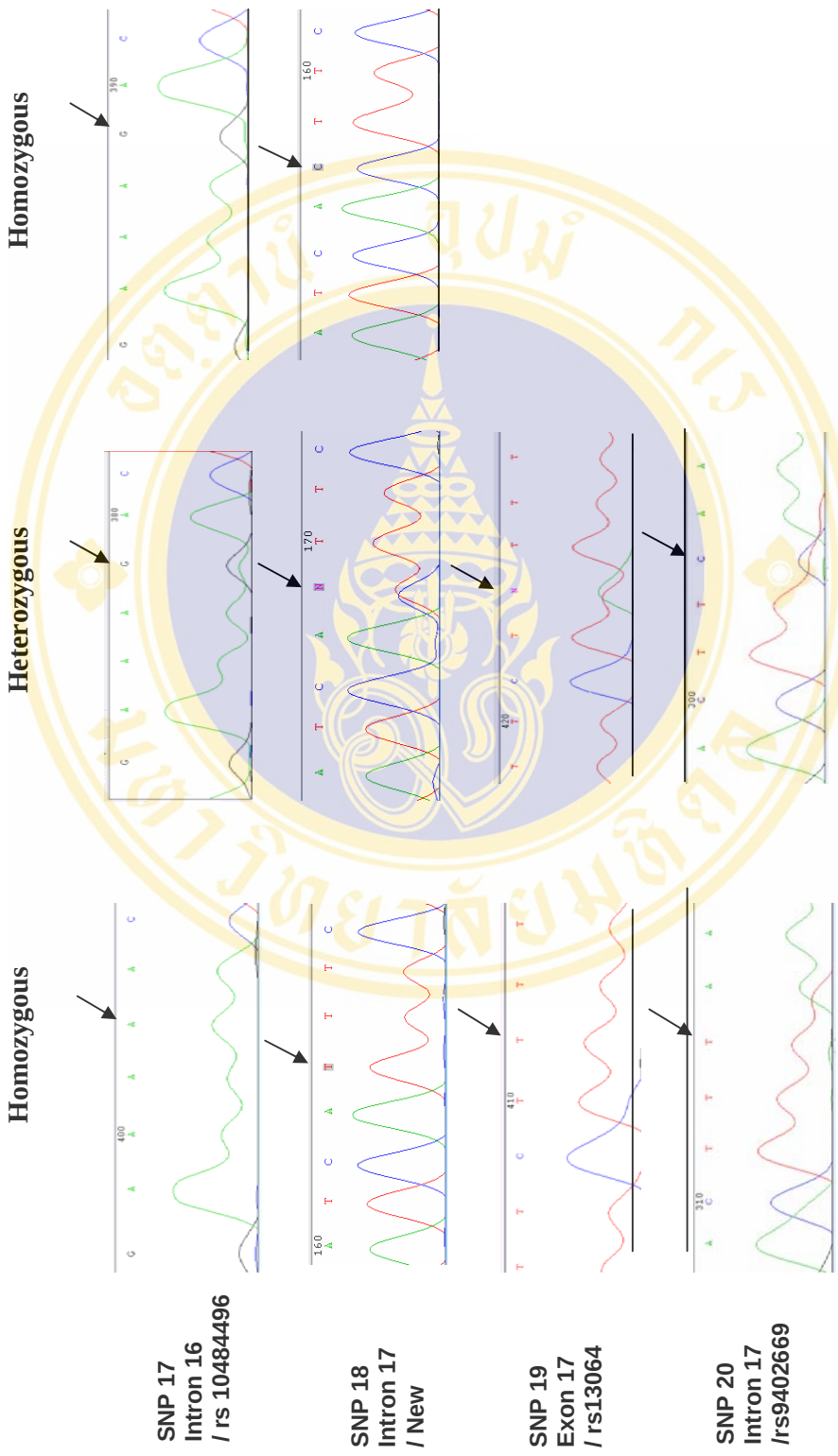


Figure 11. Chromatogram of SNP discovered by resequencing. (Cont'd)

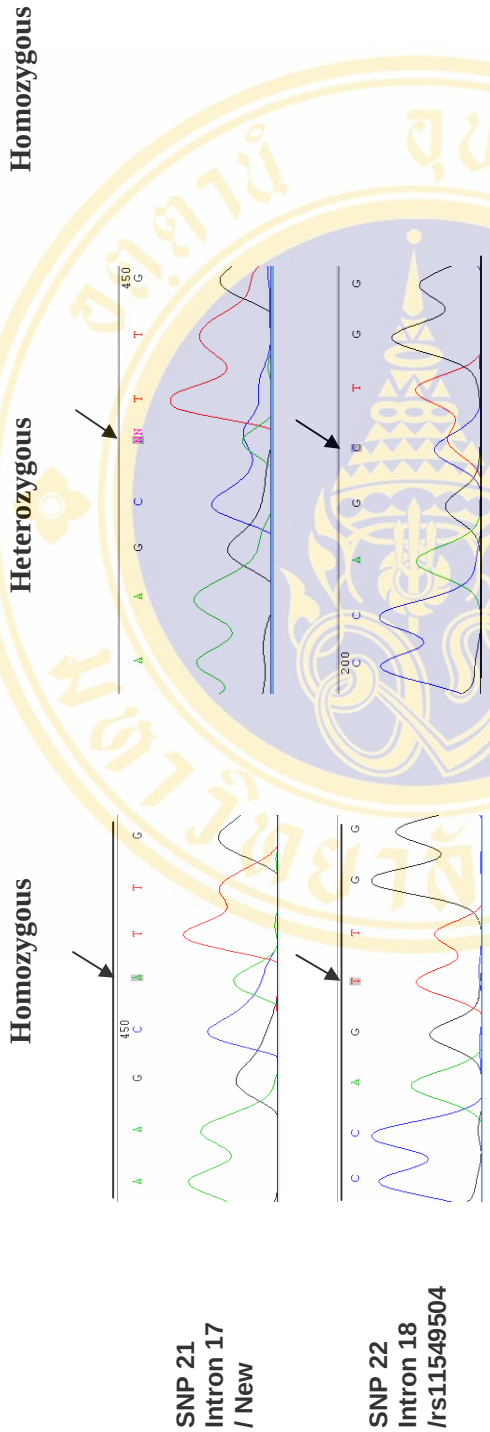


Figure 11. Chromatogram of SNP discovered by resequencing. (Cont'd)

Table 6 Genotype of 22 SNPs, discovered from sequencing

Sample ID	Gender	Phenotype	-158GγXmnI	Abs HbF (g/dL)	SNP_ID 1	SNP_ID 2	SNP_ID 3	SNP_ID 4	SNP_ID 5	SNP_ID 6	SNP_ID 7
					Ref.seq C	C	A	G	A	A	C
167M	Female	Severe	AG +/-	0.8	C	C	A	G	A	A	C
659M	Female	Severe	AG +/-	0.8	C	A	C	A	A/C	G	T
096M	Male	Severe	AG +/-	1.4	C	C/A	A/C	G/A	A	A/G	C/T
109N	Female	Severe	AG +/-	1.4	C	C	A	G	A	A	C
128N	Male	Severe	AG +/-	1.4	C	A	C	A	A/C	G	T
905N	Female	Severe	AG +/-	1.4	C	C	A	G	A	A	C
078N	Female	Mild	AG +/-	4.7	C	C	A	G	A	A	C
030N	Female	Mild	AG +/-	4.5	C	C	A	G	A	A	C
002N	Female	Mild	AG +/-	3.7	C	C/A	A/C	G/A	A	A/G	C/T
010N	Female	Mild	AG +/-	3.7	C	C	A	G	A	A	C
102M	Male	Mild	AG +/-	3.6	C	C	A	G	A	A	C
124M	Male	Mild	AG +/-	3.6	C	C/A	A/C	G/A	A/C	A/G	C/T
097N	Female	Severe	G -/-	0.7	C	C/A	A/C	G/A	A/C	A/G	C/T
004N	Female	Severe	G -/-	1.2	C	C/A	A/C	G/A	A/C	A/G	C/T
087N	Female	Severe	G -/-	1.4	C	C/A	A/C	G/A	A	A/G	C/T
040M	Female	Severe	G -/-	1.7	C	C/A	A/C	G/A	A	A/G	C/T
120R	Female	Mild	G -/-	0.9	C	A	C	A	A/C	G	T
244CR	Female	Mild	G -/-	1.9	C	A	A/C	G/A	A	A/G	C/T
478N	Male	Mild	G -/-	1.5	C	C	C	A	A/C	G	T
498R	Female	Mild	G -/-	2.0	C	A	A	G	A	A	C
850S	Female	Severe	G -/-	3.1	C	A	A/C	G/A	A/C	A/G	C/T
473R	Female	Severe	G -/-	3.4	C	A	A/C	G/A	A/C	A/G	C/T
114R	Female	Severe	G -/-	2.6	C	A	C	A	C	G	T
085N	Female	Severe	G -/-	2.4	G/C	A	A/C	G/A	A	A/G	C/T
1047M	Female	Mild	G -/-	3.2	G/C	C	A/C	G/A	A	A/G	C/T
726UT	Male	Mild	G -/-	3.8	C	A	A	G	A	A	C
006M	Male	Mild	G -/-	2.9	G/C	A	A/C	G/A	A/C	A/G	C/T
585SB	Female	Mild	G -/-	4.4	C	C/A	A/C	G/A	A/C	A/G	C/T
048N	Male	Mild	A +/+	4.0	C	A	C	A	A/C	G	T
938N	Male	Mild	A +/+	2.8	C	C	A	G	A	A	C

Table 6 Genotype of 22 SNPs, discovered from sequencing (Cont'd)

SNP_ID 8	SNP_ID 9	SNP_ID 10	SNP_ID 11	SNP_ID 12	SNP_ID 13	SNP_ID 14	SNP_ID 15	SNP_ID 16	SNP_ID 17	SNP_ID 18	SNP_ID 19	SNP_ID 20	SNP_ID 21	SNP_ID 22
Ref.seqC	C	A	A	A	A	A	A	G	A	A	T	T	A	A
C	C	A	A	A	G	A	A	G	A	A	T	T	C/A	A
T	C/T	C	G	G	A	A/C	G/A	C	A/G	A/G	T/C	T/A	C/A	A/G
C/T	C/T	C	A/G	A/G	G/A	A/C	A	G/C	A	A	T	T/A	C/A	A/G
C	C	A	A	A	A	A	G	G	A	A	T	T	C/A	A
T	C/T	C	G	G	G	A/C	G/A	C	A/G	A/G	T/C	T/A	C/A	A/G
C	C	A	A	A	A	A	A	G	A	A	T	T	C/A	A
C	C	A	A	A	A	A	A	G	A	A	T	T	A	A
C	C	A	A	A	A	A	A	G	A	A	T	T	C/A	A
C/T	C/T	C	A/G	A/G	G/A	A/C	A	G/C	A	A	T	T/A	C/A	A/G
C	C	A	A	A	A	A	A	G	A	A	T	T	C/A	A
C	C	A	A	A	A	A	A	G	A	A	T	T	A	A
C/T	C	C	A/G	A/G	G/A	A	G/A	G/C	A/G	A/G	T/C	T	A	A
C/T	C	C	A/G	A/G	***	A	G/A	G/C	A/G	A/G	T/C	T	A	A
C/T	C	C	A/G	A/G	G/A	A	G/A	G/C	A/G	A/G	T/C	T	A	A
C/T	C/T	***	A/G	A/G	G/A	A/C	***	***	***	A	T	T/A	A	A/G
C/T	C/T	C	A/G	A/G	G/A	A/C	A	G/C	A	A	T	T/A	A	A/G
T	C/T	C	G	G	G	A/C	G/A	C	A/G	A/G	T	T/A	A	A/G
C/T	C/T	C	A/G	A/G	G/A	A/C	A	G/C	A	A	T	T/A	C/A	A/G
T	C	C	G	G	G	A/C	G/A	C	A/G	A/G	T/C	T/A	C/A	A/G
C	C	A	A	A	A	A	A	G	A	A	T	T	C/A	A
C/T	C	C	A/G	A/G	G/A	A	G/A	G/C	A/G	A/G	T/C	T	C/A	A
C/T	C	C	A/G	A/G	G/A	A	G/A	G/C	A/G	A/G	***	***	***	A
T	C	C	G	G	G	A	G/A	C	G	G	***	T	***	A
C/T	C/T	C	A/G	A/G	G/A	A/C	A	G/C	A	A	T	T/A	C/A	A/G
C/T	C/T	C	A/G	A/G	G/A	A/C	A	G/C	A	A	T	T/A	C/A	A/G
C	C	A	A	G	G	A	A	C	A	A	T	T	C/A	A
C/T	C	C	A/G	A/G	G/A	A	G/A	G/C	A/G	A/G	T/C	T	A	A
C/T	C	C	A/G	A/G	***	A	G/A	G/C	A/G	A/G	T/C	T	A	A
T	C/T	C	G	G	G	A/C	G/A	C	A/G	A/G	T/C	T/A	A	A/G
C	C	A	A	A	A	A	A	G	A	A	T	T	A	***

Table 7 Comparison of HBS1L SNPs in Thai and other populations

SNP_ID	RefSNP_ID	Thai	Caucasian	Japanese	European	African American	China	Sub-Sharan African
1	New SNP	C[0.95] / G[0.05]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
2	rs9389262	C[0.466] / A[0.533]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
3	New SNP	A[0.566] / C[0.433]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
4	rs28384512	A[0.433] / G[0.566]	A[0.470] / G[0.530]	A[0.550] / G[0.450]	Unknown	Unknown	Unknown	Unknown
5	rs28384513	A[0.766] / C[0.233]	A[0.660] / C[0.340]	A[0.700] / C[0.300]	Unknown	Unknown	Unknown	Unknown
6	rs2297340	C[0.433] / T[0.566]	C[0.460] / T[0.540]	C[0.590] / T[0.410]	Unknown	Unknown	Unknown	Unknown
7	rs2297339	A[0.433] / G[0.566]	A[0.360] / G[0.640]	A[0.520] / G[0.420]	A[0.417] / G[0.583]	A[0.391] / G[0.609]	A[0.417] / G[0.583]	Unknown
8	rs2297338	A[0.433] / G[0.566]	Unknown	A[0.500] / G[0.500]	A[0.483] / G[0.517]	Unknown	A[0.567] / G[0.433]	A[0.280] / G[0.720]
9	rs9373121	C[0.816] / T[0.183]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
10	rs7742542	G[0.655] / T[0.344]	G[0.483] / T[0.517]	G[0.500] / T[0.500]	Unknown	Unknown	G[0.567] / T[0.433]	G[0.308] / T[0.692]
11	rs4559100	A[0.566] / G[0.433]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
12	New SNP	A[0.533] / G[0.466]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
13	rs4376365	A[0.535] / G[0.464]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
14	rs4376364	G[0.2] / T[0.8]	G[0.890] / T[0.110]	G[0.216] / T[0.784]	Unknown	Unknown	G[0.233] / T[0.767]	G[0.033] / T[0.967]
15	rs13214669	A[0.741] / G[0.258]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
16	rs9376077	G[0.465] / C[0.534]	Unknown	Unknown	G[0.497] / C[0.521]	G[0.435] / C[0.565]	G[0.417] / C[0.583]	Unknown
17	rs10484496	A[0.758] / G[0.241]	A[0.750] / G[0.250]	Unknown	Unknown	A[0.750] / G[0.250]	A[0.771] / G[0.229]	Unknown
18	New SNP	A[0.766] / G[0.233]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
19	rs13064/	A[0.821] / G[0.178]	A[0.614] / G[0.386]	Unknown	Unknown	Unknown	Unknown	Unknown
20	rs9402669	T[0.214] / A[0.785]	Unknown	T[0.216] / A[0.784]	T[0.103] / A[0.897]	Unknown	T[0.233] / A[0.767]	T[0.033] / A[0.967]
21	New SNP	A[0.714] / C[0.285]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
22	rs11549504	A[0.793] / G[0.206]	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown

Table 8 Amino acid status of codon at the SNP position in exon regions.

SNP	Exon	Codon no.	Codon change	Amino acid residue
9	4A	84	**C**TA → **T**TA	Leucine → Leucine
10	4A	200	G**A**A → G**G**A	Glutamate → Glycine.
19	17	26	TT**T** → TT**C**	Phenylalanine → Phenylalanine

Table 9 SNPs found in transcription factor binding sequence. Capital bold letters show the position of SNPs.

SNP ID	Position	Transcription factor	sequence
2	1Kb promoter region	VDR (Vitamin D receptor)	gcgatccgccctc**C**taggcttccca
3	1Kb promoter region	ELF2	gattacagg**A**aggagcc
4	1Kb promoter region	GATA ETSF	t**C**ctgataatctg attatcag**G**aaccaaat
5	1Kb promoter region	Hox 1 (Homeobox)	aaaataatcc**A**gaacgc
7	5' UTR	CREB (cAMP response element binding site)	caagcgcgacgtcttagct**G**t

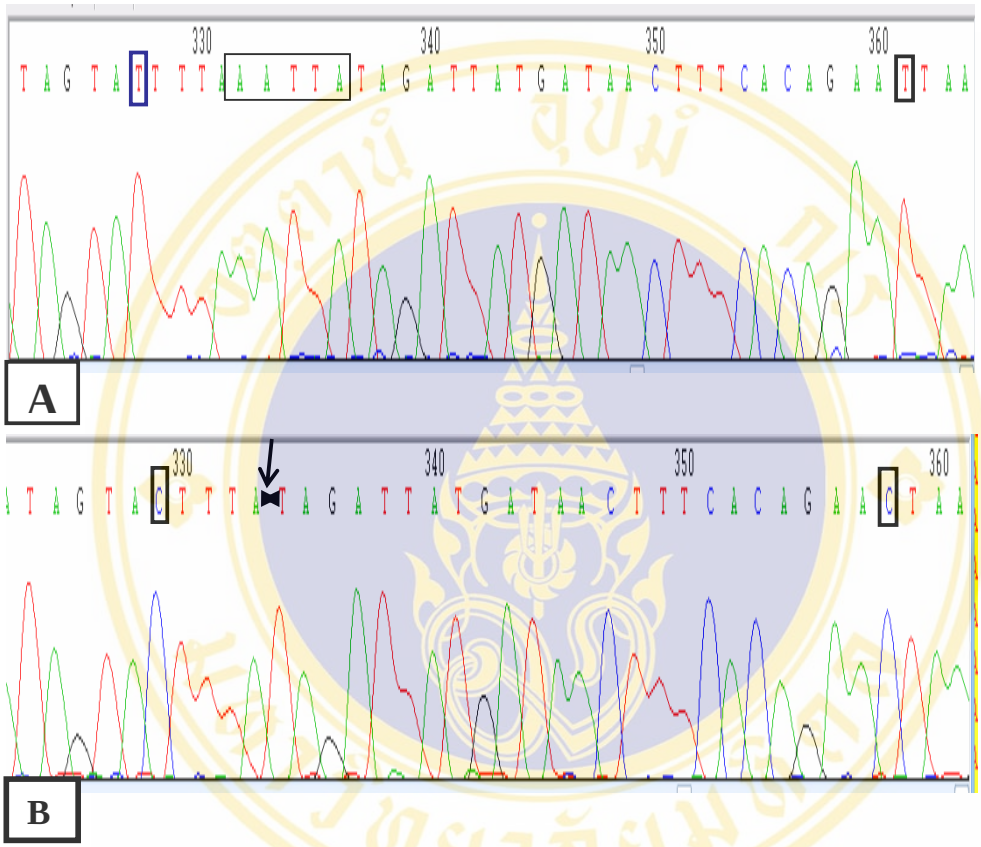


Figure 12. The example chromatogram showing the 5 bp in/del polymorphism in the intron 4 of the major transcript.

A. Chromatogram of homozygous non-deletion in the intron 4 closed to the exon 5 of the major transcript.

B. Chromatogram of homozygous deletion in the intron 4 closed to the exon 5 of the major transcript showing of 5 bp in/del polymorphism flanked by SNPs 11 and 12.

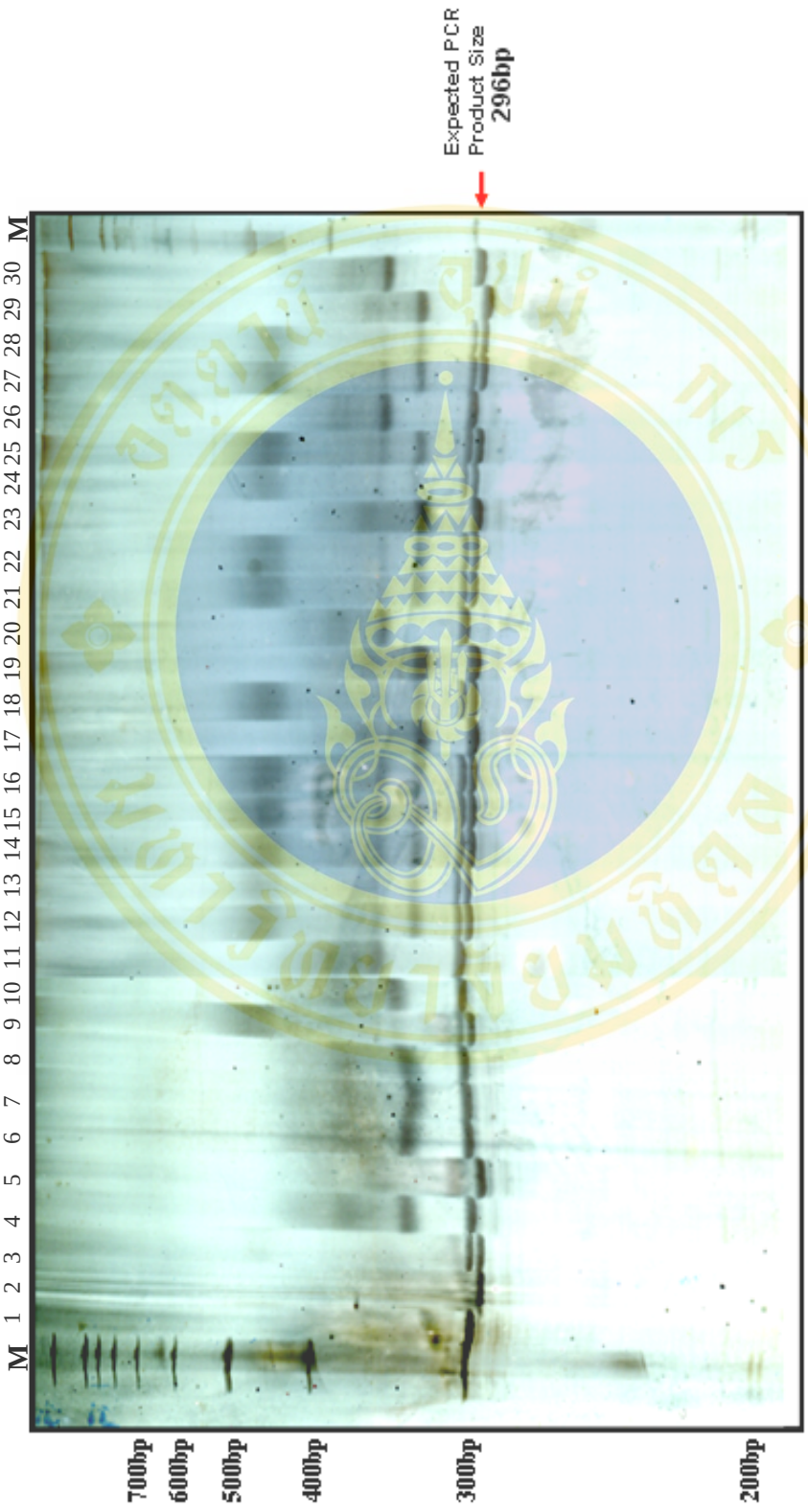


Figure 13. 5% denaturing polyacrylamide gel electrophoresis of PCR product showing the heterozygosity and homozygosity of 5 bp in/del polymorphism of intron 4 close to the exon 5 of the major transcript of *HBS1L* gene .

M = 100 bp Marker.

Table 10. Comparison of the homozygous and heterozygous 5 bp in/del polymorphisms in 30 β-thalassemia / Hb E patients with *Xmn*I polymorphism and the amount of Abs Hb F.

Patient ID	Code	Phenotype	Abs HbF (g/dl)	*Xmn*I	5 bp Deletion		
					Homozygous(N)	Heterozygous	Homozygous (Del)
167M	1	Severe	0.8	+/-	***		
659M	2	Severe	0.8	+/-			***
096M	3	Severe	1.4	+/-			***
109N	4	Severe	1.4	+/-	***		
128N	5	Severe	1.4	+/-			***
905N	6	Severe	1.4	+/-	***		
078N	7	Mild	4.7	+/-	***		
030N	8	Mild	4.5	+/-	***		
002N	9	Mild	3.7	+/-		***	
010N	10	Mild	3.7	+/-	***		
102M	11	Mild	3.6	+/-	***		
124M	12	Mild	3.6	+/-		***	
097N	13	Severe	0.7	-/-		***	
004N	14	Severe	1.2	-/-		***	
087N	15	Severe	1.4	-/-		***	
040M	16	Severe	1.7	-/-		***	
120R	17	Mild	0.9	-/-			***
244CR	18	Mild	1.9	-/-		***	
478N	19	Mild	1.5	-/-			***
498R	20	Mild	2	-/-	***		
850S	21	Severe	3.1	-/-		***	
473R	22	Severe	3.4	-/-		***	
114R	23	Severe	2.6	-/-			***
085N	24	Severe	2.4	-/-		***	
1047M	25	Mild	3.2	-/-		***	
726UT	26	Mild	3.8	-/-	***		
006M	27	Mild	2.9	-/-		***	
585SB	28	Mild	4.4	-/-		***	
048N	29	Mild	4	+/+			***
938N	30	Mild	2.8	+/+	***		

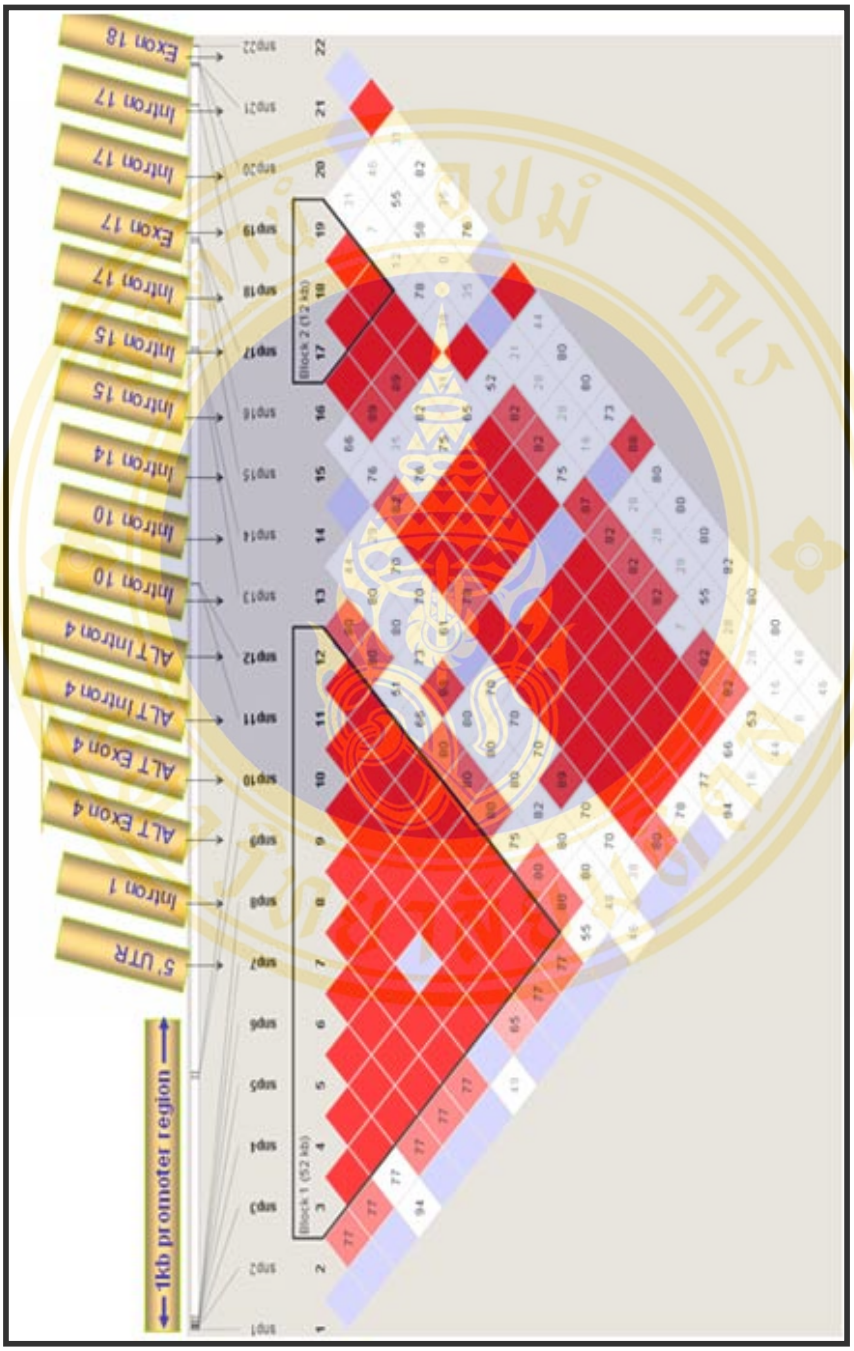


Figure 14 Haplotype of 22 SNPs shows two haplotype blocks. Small colored blocks indicate the SNPs in linkage disequilibrium.

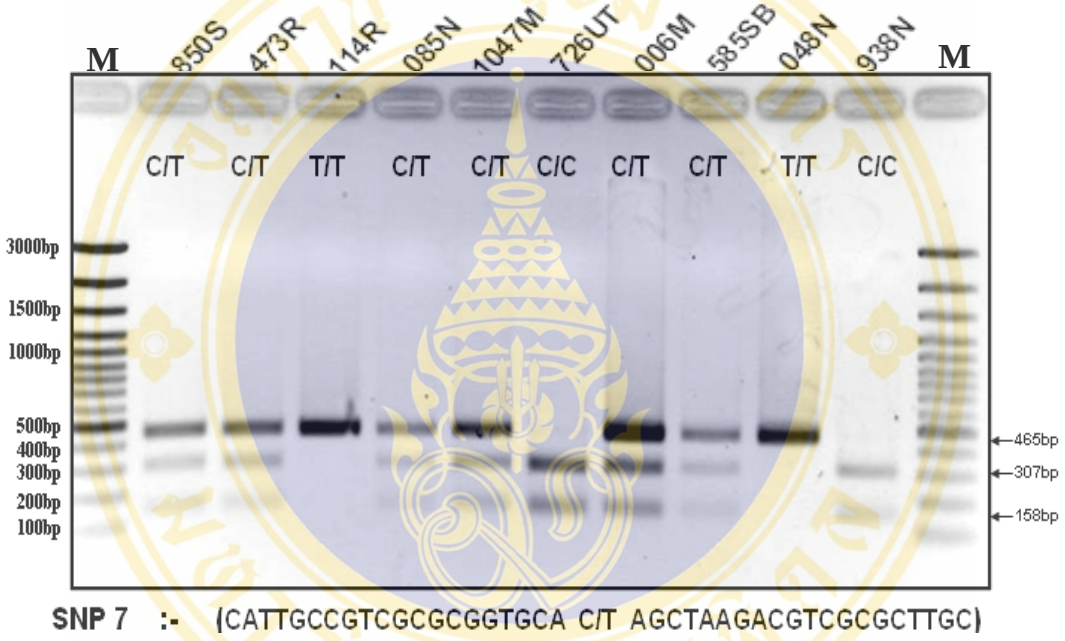


Figure 15 Agarose gel electrophoresis of 10 sequenced samples digested with the *Apa*L1 restriction enzyme and used as positive control for genotyping.

M = 100 bp ladder marker plus.

Table 11. Distribution of SNP 7, C and T, (genotype and allele) in mild, severe and β-thalassemia co-inherited with α-thalassemia (B/A/T) cases

	Genotype					Alleles	
Group	CC	CT	TT	N Genotypes	HWE p = value	C	T
Mild case	52 (29%)	86 (48%)	42 (23%)	180	0.685	190 (53%)	170 (47%)
B/A/T	17 (24%)	30 (43%)	23 (33%)	70		64 (46%)	76 (54%)
Severe cases	51 (17%)	139 (47%)	107 (36%)	297	0.780	241(41%)	353 (59%)
Total	120 (22%)	255 (47%)	172 (31%)	547		495 (45%)	599 (55%)

p = 0.245

p = 0.345

SNP 7 mild vs severe, excluding cases co inherited with α- thalassemia, p = 0.002

SNP 7 allele C vs allele T p = 0.002

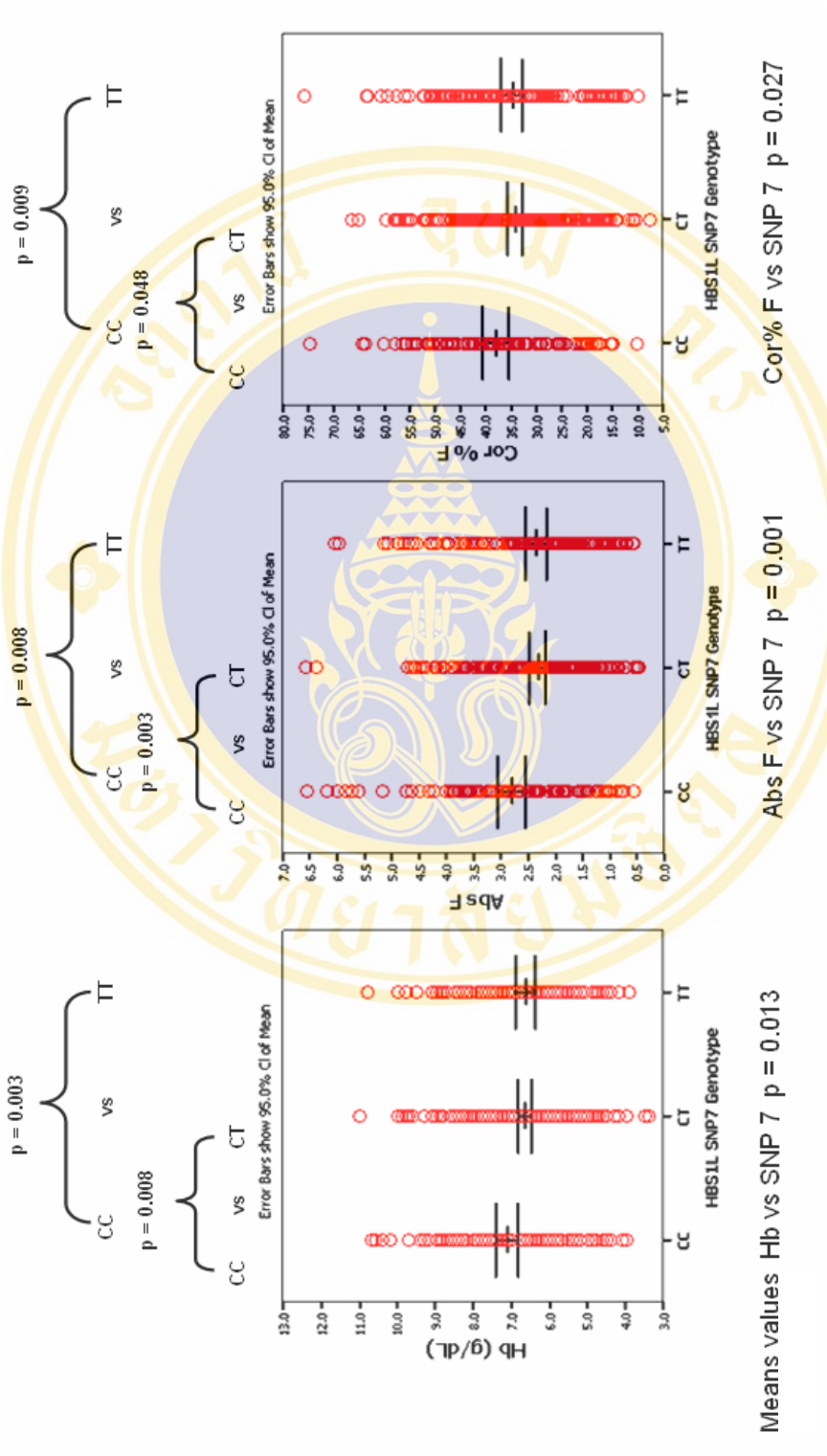


Figure 16 Scattergram of Hb, Abs F, Cor% F distribution in CC, CT, TT genotype of SNP 7 of *HBS1L* gene.

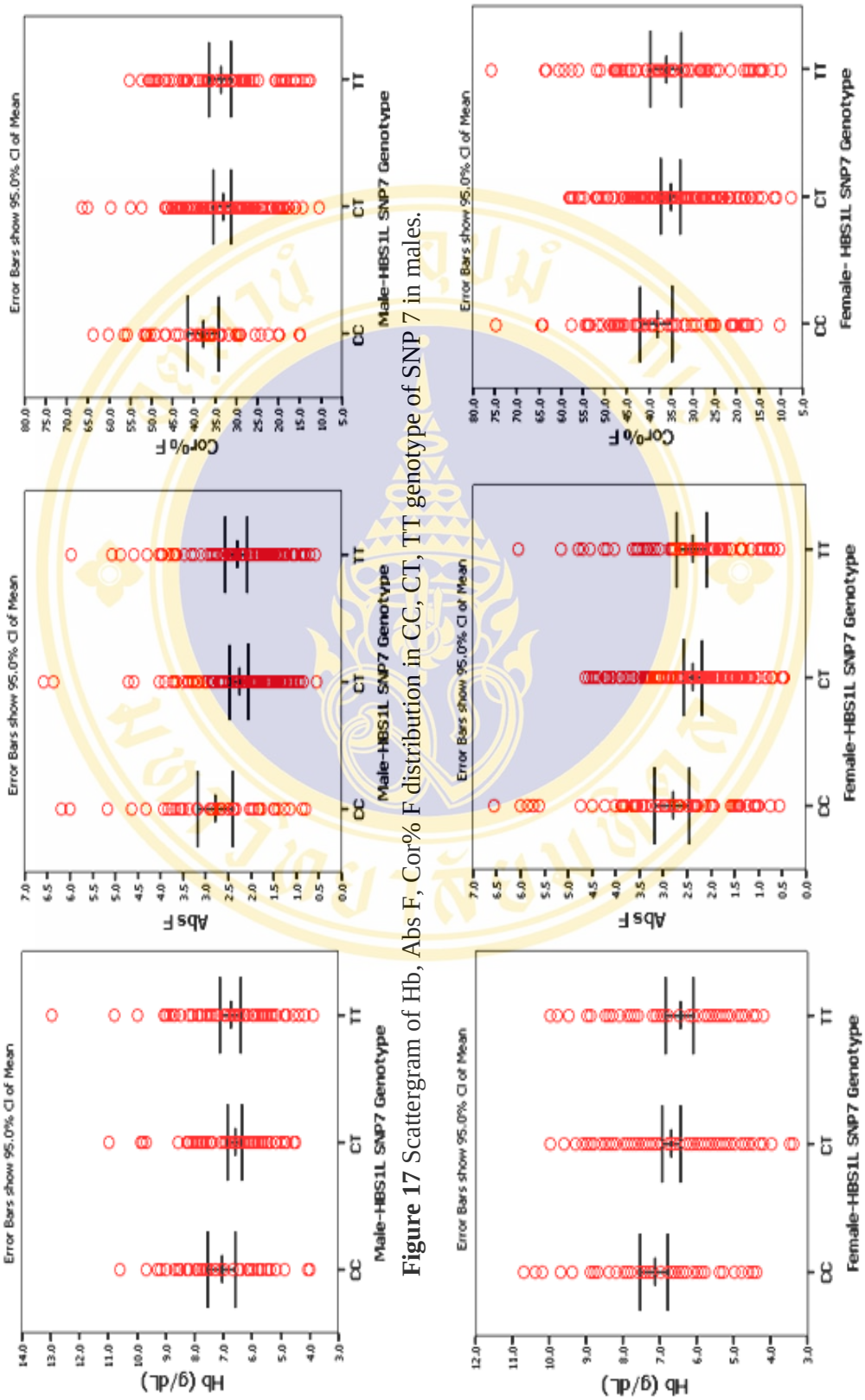


Figure 17 Scattergram of Hb, Abs F, Cor% F distribution in CC, CT, TT genotype of SNP 7 in males.

Figure 18 Scattergram of Hb, Abs F, Cor% F distribution in CC, CT, TT genotype of SNP 7 in females.

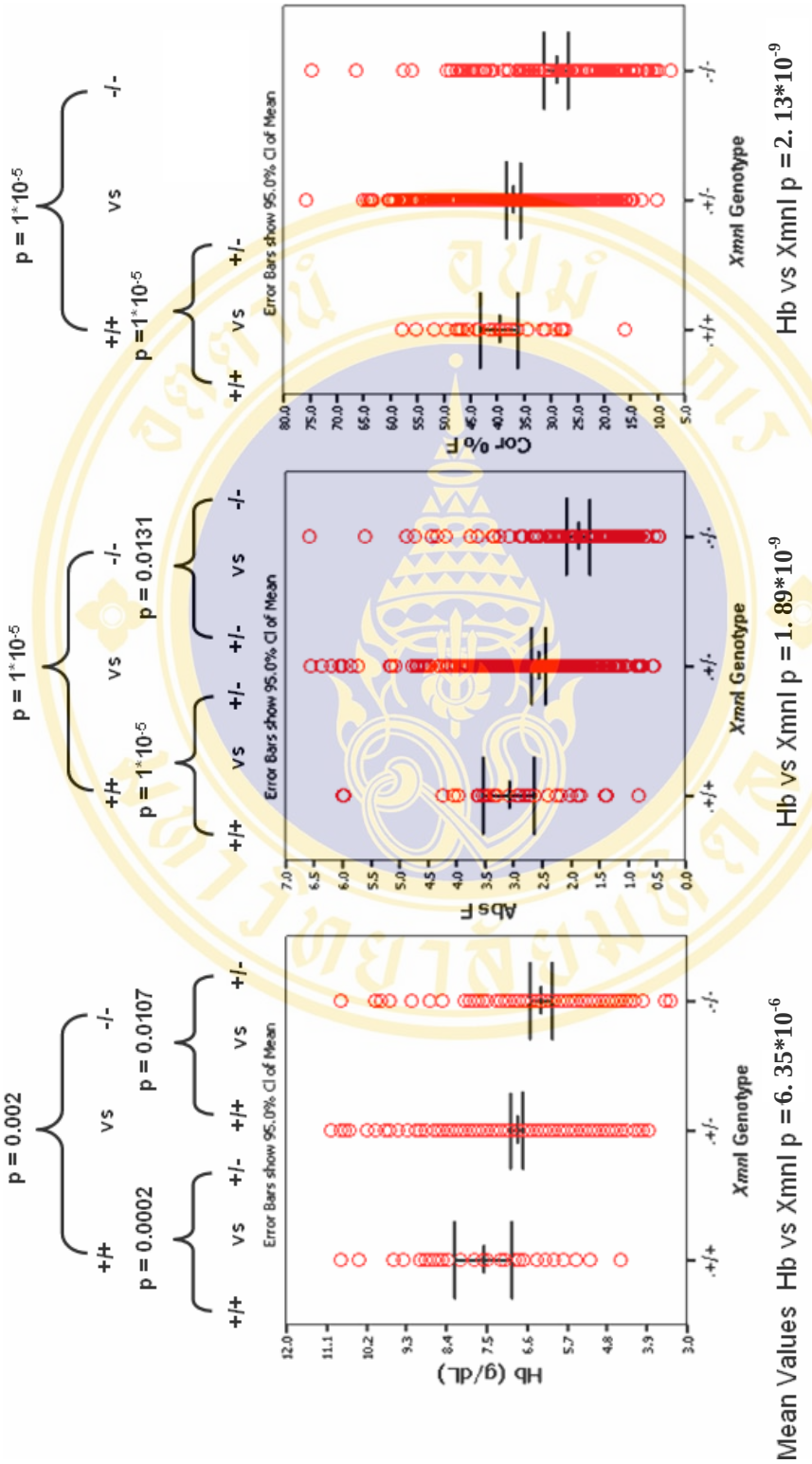


Figure 19 Scattergram of Hb, Abs F, Cor% F distribution in all cases according to +/+,+/-,-/- *Xmn*I genotype.

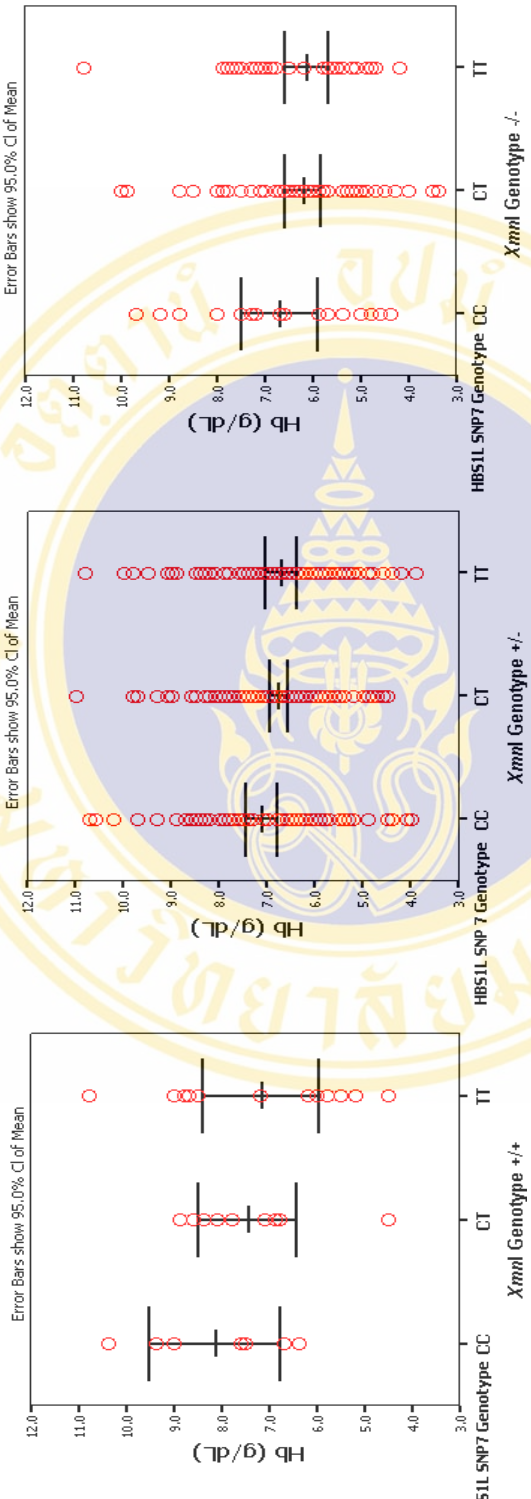
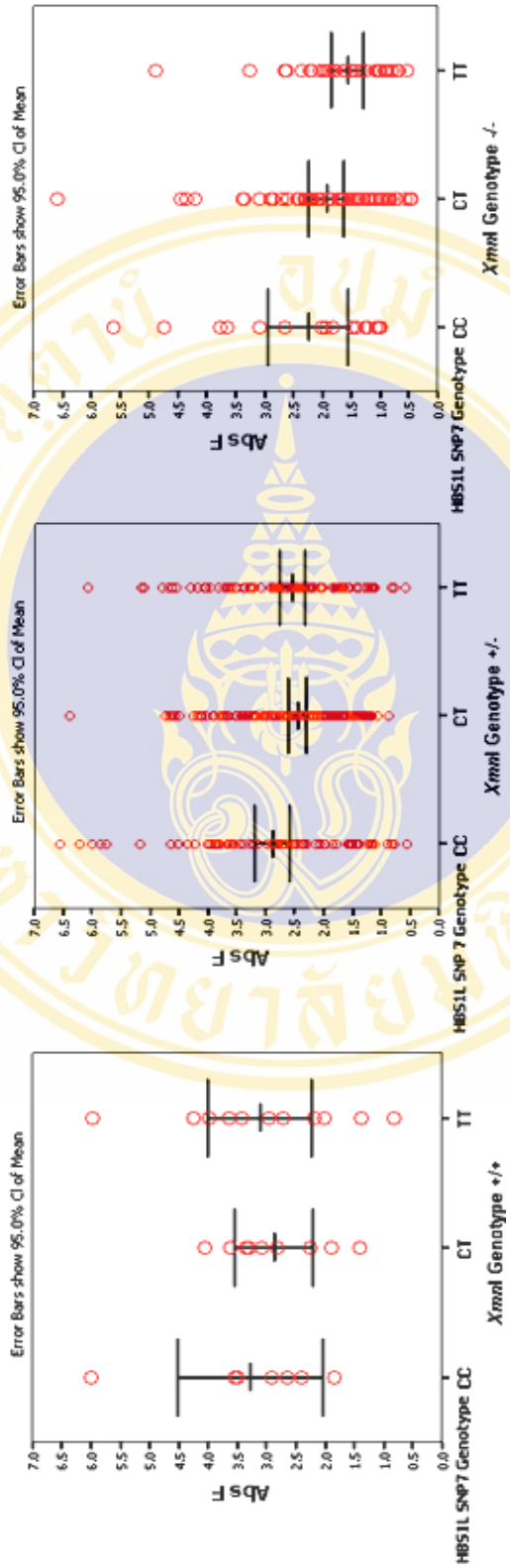


Figure 20 Comparison of Hb distribution in *XmnI* +/+, +/-,-/- genotype with *HBS1L* SNP 7 CC, CT, TT genotype



ANOVA p value 0.8312 0.0096 0.1119

Figure 21 Comparison of Abs F distribution in *XmnI* +/+, +/-,-/- genotype with *HBS1L* SNP 7 CC, CT, TT genotype

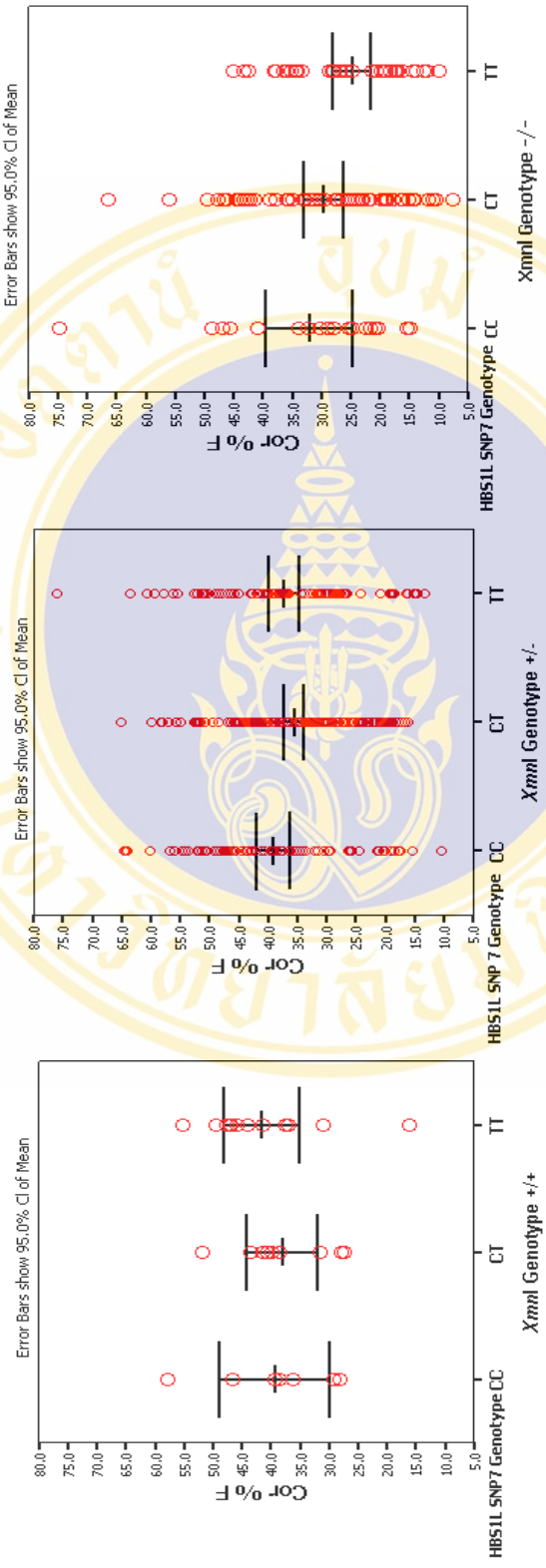


Figure 22 Comparison of Cor % F distribution in *XmnI* +/+, +/-,-/- genotype with *HBS1L* SNP 7 CC, CT, TT genotype

Table 12. The amount of absolute F in individuals who have different genotypes of SNP 7 and *Xmn*I polymorphism.

SNP 7 Genotype	XmnI Genotype		
	+/+	+/-	-/-
CC	3.28 ±1.35	2.89 ± 1.33	2.24 ± 1.39
CT	2.88 ±0.86	2.45 ± 0.98	1.92 ± 1.15
TT	3.12 ±1.40	2.54 ± 1.1	1.56 ± 0.86

CC/++ TT/++ p = 0.810 CC/- - TT/- - p = 0.017

CC/++ CC/- - p = 0.096 TT/++ TT/- - p = $1*10^{-5}$

. **Table 13**. Genotype distribution of *Xmn*I and SNP 7 in mild and severe cases excluding β-thalassemia with α-thalassemia..

SNP 7 Genotype	*Xmn*I Genotype		
	+/+ (Mild / Severe)	+/- (Mild / Severe)	-/- (Mild / Severe)
CC	6 (5 / 1)	73 (42 / 31)	21 (5 / 16)
CT	10 (8 / 2)	152 (66 / 86)	57 (12 / 45)
TT	12 (7 / 5)	93 (31 / 62)	37 (4 / 33)

CHAPTER 6
DISCUSSION

6.1 Sample selection for HBS1L - SNP study.

DNA samples taken for this study were selected from the stock DNAs previously extracted for the large cohort genome-wide study. The patients are β-thalassemia / Hb E. Who were classified as the mild and severe groups according to the designated criteria (Table 1) [124]. Previous studies looking for modifying factors that determine the severity difference among thalassemic patients have showed that the patients who appear to have an inherited ability to produce high Hb F level exhibit mild clinical symptoms [140]. *Xmn*I polymorphism, C → T, at the position -158 of the γ-globin gene is one of the determinant associated with high Hb F production [140]. Recently, a number of genes were found on chromosome 6q23, including *MYB, ALDH8A1, HBS1L, PED7B* and an uncharacterized *AHI1* gene, which may harbour trans-acting elements that help modulate Hb F level [5]. However, *HBS1L* gene was selected for this study because from the genome-wide SNP study, searching for β-thalassemia / Hb E disease severity modifier gene showed that there are 5 significant SNPs in *HBS1L* gene associated with severity of the patients. Genotype of *Xmn*I polymorphism and the level of Hb F were thus considered for samples selected for this study (Table 2)

6.2 SNP analysis by DNA sequencing and PCR based technique.

A total of 30 genomic DNA samples were taken for DNA sequencing after DNA integrity, concentration and purification were determined. PCR amplification was performed to amplify all exons including intron-exon junction of *HBS1L* gene. All primers resulted in the expected PCR product with good quality for sequencing except primers for exon 4A3. This region was finally found to contain an approximately 40 “A” tract in intron close to exon 4A. From this result, it was

concluded that Taq DNA polymerase enzyme was unable to incorporate the repeated A nucleotides in PCR reaction and results in non amplification of PCR product.

During sequencing step, either forward or reverse primers were used to sequence PCR products, whose sizes ranges from 500 to 600 bp. PCR products having size longer than 600 bp, sequencing was performed by using both the forward and reverse primers. However, due to the presence of poly A or poly T tracts, electrophoregrams of some sequencing reactions were cut short and were unreadable. The examples of such sequencing reactions were the 1 kb promoter region, exon 4A and exon 18 electrophoregrams. The expected PCR size was more than 600 bp and the nucleotide sequence were not clear after multi poly A or poly T tracts (more than 9) (Figure 23). Such electrophoregrams were initially interpreted as gene deletion sequence. However, to clarify and read the full sequence of amplified PCR product, new sequencing primers were designed to avoid poly A or T sequence and sequencing was performed from opposite direction. The results showed clear electrophoregram upto the repeated T or A region and was unreadable after repeated region.

Sequencing results also revealed 5 bp deletion in intron 4 near closed to the exon 5 and flanked by SNP 11 on the 5' end and SNP 12 on the 3' end. All chromatograms for exon 5 fragment were aligned by using vector NTI software, and the results suggests that 5 bp deletion polymorphism was associated with the C allele of SNP 11 and 12 (Fig. 12). Due to the 5 bp deletion, chromatograms of the heterozygous state were not clear and the heterozygous genotype for SNP 11 and 12 had been inferred from 5% polyacrylamide gel electrophoresis.

6.3 Selection of SNP 7 for genotyping.

Haplotype analysis showed that SNP 3, 4, 6, 7, 8, 11 and 12 were in the strong linkage disequilibrium. Gene promoter transcription factor binding site TFBS prediction program MatInspector available via in www.genomatrix.com used to analyze the possible function of SNP found on 1 kb 5' of HBS1L gene. which showed that 5 SNPs were located within the different transcription factor binding site sequences, (Table 9). Moreover, two SNPs, SNPs 4 and 7 were located on GATA and CREB transcription factor binding sites respectively. The two transcription factors were previously demonstrated to play role in erythroid developmental system [136].

GATA is a multigene family of transcription factors, which binds to DNA consensus sequence (T/A) GATA (A/G). These are an erythroid specific transcription factors, which are required for survival, maturation and proliferation of erythroid precursor cells. GATA-1 plays a critical role during megakaryocytic proliferation and differentiation [129][130] while GATA-2 levels are high in progenitor cells and decline during erythroid maturation [131].

CBP was originally discovered based on its ability to interact with the cAMP response element binding protein (CREB) [132]. It functions in combination with p300 and acts as cofactors for broad number of transcription factors both within and outside the hematopoietic system [133]. It has been shown that CBP and p300 possess intrinsic histone acetyltransferase (HAT) activity [134][135]. Histone acetylation has been correlated with transcriptionally active DNA because of relaxed chromatin configuration, which facilitates transcription factor access to DNA.

However, only SNP 7 located in restriction cleavage site of enzyme *Apa*L1. As SNP 7 was also in linkage disequilibrium of SNP 4 and genotyping of SNP 7 was relatively easier by restriction enzyme digestion than other methods, so SNP 7 was preferred over SNP 4 for genotyping process.

6.4 SNP 7 Genotyping, correlation of SNP 7 and Hb F level.

SNP 7 was genotyped in mild and severe cases to analyze whether this SNP has any effect on Hb, Absolute F and Cor% F regulation**.** Comparison of allele frequency showed that C/C genotype frequency was significantly higher in mild cases as compared to severe cases and T/T genotype frequency was higher in severe than mild cases (Table 13) and there was a significant difference in allele difference in genotype frequency between mild and severe cases ($p = 0.002$).

Scattergrams charts for SNP 7, sex and *Xmn*I genotype for Hb, HbF and Cor% F distribution (Fig 19 to 22) suggested that T/T genotype was related with lower average levels of Hb, AbsF and Cor % F as compared to C/C genotype of SNP 7. Furthermore, SNP 7 C/C and T/T did not seem to have any effect on Hb, AbsF and Cor% F levels in different gender. Moreover *Xmn*I scattergram chart showed that *Xmn*I -/- genotype is significantly associated with low level on Hb , AbsF and Cor% F

while *Xmn*I +/+ was associated with higher levels of Hb , AbsF and Cor% F corresponding to the previous reports of *Xmn*I ($p = 1.4*10^{-8}$).

Comparison of average Abs F distribution in *Xmn*1 and SNP 7 genotype in different combination (Table 12) suggested that cases who possess *Xmn*I +/+ genotype combination with SNP 7 C/C, T/C or T/T tended to have a higher level of AbsF indicating that *Xmn*I +/+ genotype is a stronger modulating factor than SNP 7. On the contrary, patients who have *Xmn*I -/- genotype and C/C genotype of SNP 7 tended to have a higher AbsF level indicating that C/C genotype may also play a role with AbsF level compared to the *Xmn*I +/+ and SNP 7 C/C genotype. This assumption is further strengthened by the fact that patients whose *Xmn*I genotype is -/- with SNP 7 T/T genotype tend be more severe, possessing very low AbsF cor% F and Hb level (Fig 23-25). This observation is further supported by the fact that the ratio of mild to severe cases who had *Xmn*I +/+ genotype with SNP 7 C/C, C/T or T/T is higher as compared with the low ratio in those who had *Xmn*I -/- genotype with SNP 7 C/C, C/T or T/T (Table 13). From the statistic analysis no evidence supported the gene-gene interaction effects of *Xmn*I and SNP 7 (SNP 7 C/C with *Xmn*I +/+ vs severity p = 0.4 and SNP 7 T/T with *Xmn*I -/- vs severity p = 0.37). These results suggested that SNP 7 of HBS1L might have some roles in AbsF regulation either alone or have some synergistic effect to *Xmn*I.

Previous studies have shown that CREB in combination with p300 possess intrinsic histone acetyltransferase (HAT) activity [133][134]. Histone acetylation has been correlated with transcriptionally active DNA because of relaxed chromatin configuration, which facilitates transcription factors access to DNA.

CREB binding protein and p300 has been shown to acetylates zinc finger domain of GATA-1 protein, which markedly stimulate the transcriptional activity of GATA-1 and is required for erythroid differentiation. [136]. CBP and p300 has also been shown to acetylate nonhistone nuclear protein including p53 [136], EKLF [138], TFIIF [139] etc.

SNP 4, which lies within the GATA transcription binding sequence is only 176 bp away from SNP 7, which lies within the CREB transcription binding site. Based on previous studies and our data, it may be assumed three possible ways by which these SNPs affect the HBS1L activity. Viz:-

1. SNPs 7 affects CREB binding protein with p300 (CBP/p300), which unable to acetylate DNA in 5'UTR region thus preventing binding of other transcription factors resulting in inefficient activity of HBS1L gene
2. SNP 4 affects binding of GATA transcription factor, which hinders the binding of GATA, either alone or in combination with SNP 7 and influences normal expression of HBS1L gene.
3. Both of the transcription factors bind loosely to their respective sites resulting in the inability of CBP/p300 to acetylation GATA, which is required for its normal activity [135], thus resulting in improper *HBS1L* gene expression.

These assumptions are further supported by a speculative model, which proposed that erythroid DNA binding proteins interact with each other as well as the general co-activators like CBP/p300. This inturn modifies histone and individual transcription factors through acetylation and might link promoter/enhancer bound proteins with components of the basal transcription machinery (Fig. 24) [132].

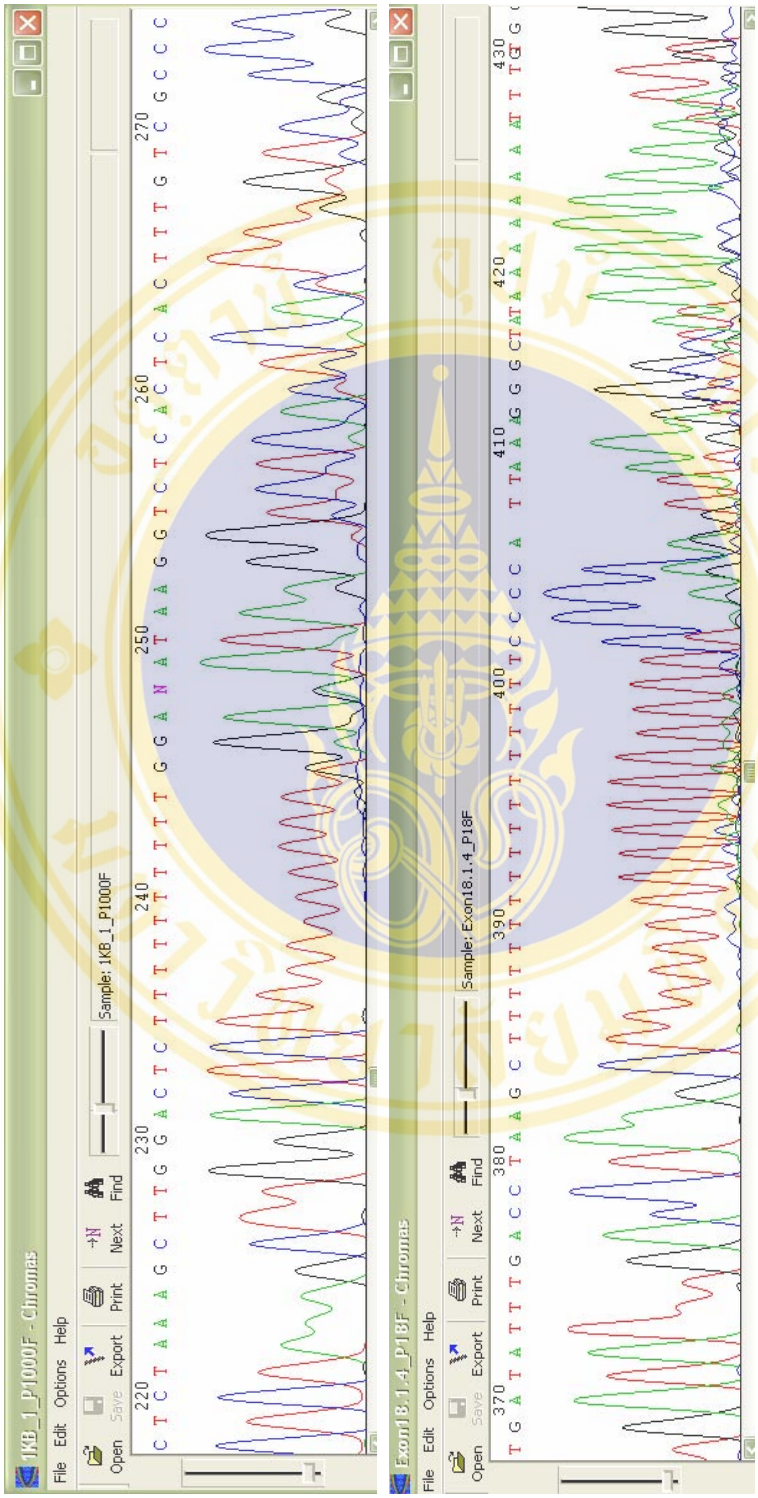


Figure 23. 5 ‘1 kb promoter region and exon 18 electrophoregram before and after poly T sequence.

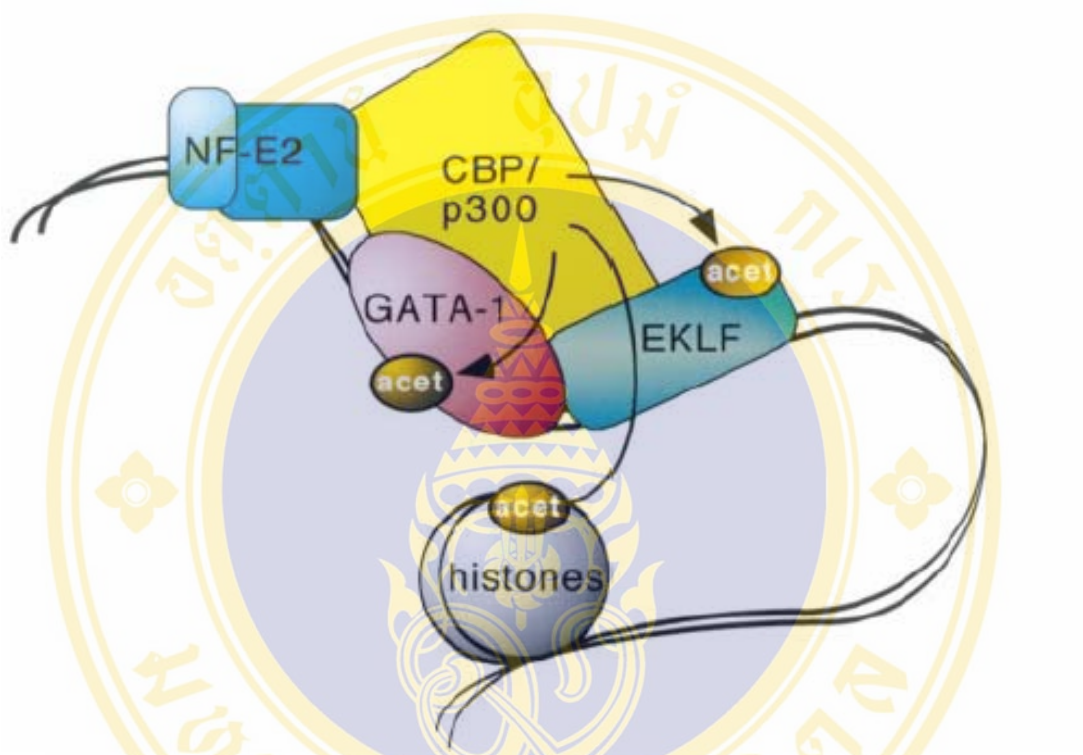


Figure 24. Hypothetical model in which NF-E2, GATA-1 and EKLF cooperates to recruit CBP and p300 to the locus control region of the β-globin gene cluster [132].

CHAPTER 7
CONCLUSION

β-thalassemia, one of the most common genetic disease, caused by defective β-globin synthesis. The major cause of anemia in β-thalassemia, secondary to imbalance globin chain synthesis, is its deleterious effect of the excess α-globin chain on erythroid maturation. Co-inheritance of α-thalassemia or increased Hb F expression can ameliorate the severity of patients due to reduced degree of globin chain imbalance. The study of factors effecting Hb F expression may provide a future novel therapeutic intervention.

Linkage analysis showed that the gene(s) on chromosome 6q23 may effect Hb F level. This was confirmed by genome-wide SNP analysis, which indicated that HBS1L gene was associated with severity of β-thalassemia patients.

In this study, the search for SNPs in HBS1L gene that may affect the Hb F level, genomic DNA from 14 mild and 16 severe β-thalassemia / Hb E patients who had different genotype of *Xmn*1 polymorphism and Hb F level were resequenced. All exons including exon-intron junctions, promotor region upto 1 kb from start site, and 3' untranslated region of *HBS1L* gene spanning in the region of chromosomes 6q23 were amplified. The results showed that all exons, 3' UTR and 1 kb promoter region could be amplified. After amplification, the DNA sequencing of all amplified fragments was performed by didioxynucleotide chain terminator method. Results revealed 22 SNPs, 5' bp deletion in intron 4 near exon 5. Out of 22 SPS, 5 SNPs were new and 6 SNPs were found prompter region. Results also revealed 1 nonsynonimous SNP in exon 4A which changes amino acid residue Glutamate to Glycine. 22 SNPs were used to create haplotype and explore the SNPs, which are in linkage disequilibrium. Results revealed that SNP 3 to 12 were in the same haplotype block. Out of which SNP 2, 4, 6, 7, 8, 11 and 12 were in strong linkage disequilibrium.
Analysis of 7 SNPs located on 1 kb promoter region and 5' UTR revealed that 5 SNPs were in putative transcription factor binding sites. The CREB transcription factor binding site, on which SNP 7 was located has been previously demonstrated to play

role in erythroid maturation. SNP 7 was thus genotyped in 295 severe, 180 mild and 70 β-thalassemia / Hb E co-inherited with α-thalassemia by PCR-RFLP method. The results revealed that allele frequency of SNP 7 in β-thalassemia / Hb E with co-inherited with α-thalassemia were not statistically significant from mild and severe cases. C/C genotype of SNP 7 was higher in mild cases and allele frequency between mild and severe cases was significantly different. Comparison of average absolute F distribution suggested that C/C genotype was associated with the higher Hb F while T/T genotype was associated with the lower Hb F. Furthermore, data suggested that *Xmn*I, a well known factor associated with Hb F level, and HBS1L SNP 7 are independent to each other in regulating of Hb F. The *Xmn*I had stronger modulating effect on Hb F level. However, SNP 7 showed significant association with high Hb F level in *Xmn*I -/- genotype.

REFERENCES

1. Jane S. M. and Cunningham J. M. Understanding fetal globin gene expression: A step towards effective Hb F reactivation in hemoglobinopathies. Br. J. of haematology.1998;102:415-422
2. Donald V. and Judith G. V. Biochemistry , 3rd edition , Wiley publishers, 2004.
3. Mastropietro F., Mediano G., Cappabianca M. P., Foglietta E., D'Asero C, Mezzabotta M., Ponzini D., Maffei L., Amato A., Lerone M.,Grisanti P., Biagio1 P. D., Rinaldi S and Bianco I. Factors regulating Hb F synthesis in thalassemic diseases BMC Blood disorders.2002;2:2.
4. Garner C., Tatu T., Thein S.L. Genetics influence on F cells and other hematologic variables: a twin heritability study . Blood 2000.95 (1): 342-6.
5. Close J., Game L., Clark B., Bergounioux J., Gerovassili A., Thein S. L. Genome annotation of a 1.5 Mb region of human chromosome 6q23 encompassing a quantitative trait locus for fetal hemoglobin expression in adults BMC Genomics 2004, 5:33
6. Garner, C., Tatu T., Thein S. L.. Evidence of genetic interaction between the beta-Globin Complex and Chromosome 8q in the expression of fetal hemoglobin. Am J Hum Genet 2002:70 (3) 793-9.
7. Dover G. J., Smith K. D., Chang Y. C., Purvis S, Mays A, Mayers D.A. Fetal hemoglobin level in sickle cell disease and normal individuals are partially controlled by an X-linked gene located at Xp22.2. Blood 1992;80:816-24.
8. Wyszynski D.F., Baldwin C.T, Cleves M.A., Amirault Y., Nolan V.G., Farrell J. J., Bisbee A., Kutlar A., Farrer L. A., Steinberg M. H. Polymorphisms near a chromosome 6q QTL area are associated with modulation of fetal hemoglobin level in sickle cell anemia. Cellular and Molecular Biology 2004; 50 (1):23-33.
9. Turcinov D., Krishnamoorthy R., Janicijevic B., Merkovic I., Mustac M., Lapoumeroulie C., Chaventre A., Rudan P. Anthropogenetical analysis of abnormal human alpha-globin gene cluster arrangement on chromosomes 16. Coll Antropol 2000;24:295-301

10. Higgs D. R., Vickers M. A., Wilkie A O., Pretorius I M., Jarman A P., Weatherall D. J. A review of the molecular genetics of the human alpha-globin gene cluster Blood 1989;73:1081-1104.
11. Collins F. S., Weissman S.M., The molecular genetics of human hemoglobin. Prog Nucleic Acid Res Mol. Biol 1984; 31:315-462
12. Tuan D, Solomon W, li Q, London I. M., The "beta-like-globin" gene domain in human erythroid cells. Proc Natl Acad Sci USA 1985;82:6384-6388
13. Forrester W. C, Takegawa S, Papayannopoulos T, Stamatoyannopoulos G, Groudine M. Evidence for a locus activation region: the formation of developmentally stable hypersensitive site in globin-expression hybrids. Nucleic Acids Res 1987;15:10159-10177
14. Groveld F, Antoniou M, Van Assendelft G. B, de Boer E, Hurst J , Kollias G, MacFarlane F, Wrighton N. The regulation of expression of human beta-globin genes. Prog Clin Biol Res 1987;251:133-144
15. Vyas P, Vickers M. A., Simmons D. L., Ayyub H., Craddock C. F., Higgs D.R. Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. Cell 1992;69:781-793.
16. Craddock C. F., Vyas P, Sharpe J. A., Ayyub H, Higgs D.R. Contrasting effects of alpha- and beta-globin regulatory elements on chromatin structure may be related to their different chromosomal environment. Embo J 1995;14:1718-1726
17. Weatherall D. J, Clegg, J. B., Higgs, D. R., Wood, W. G. The Metabolic and Molecular Basis of Inherited Disease (ed. Sciver, C. R.). McGraw-Hill, New York. 2001:4571 - 636.
18. Weatherall D. J., Clegg J. B. Thalassaemia - a global public health problem. Nature Med. 1996;2:847- 849.
19. Weatherall D. J. Phenotype - Genotype Relationships In Monogenic Disease: Lessons From The Thalassaemias Macmillan Magazines Ltd. 2001 April 2001;2:245.
20. Stamatoyannopoulos G., Grosveld F. The Molecular Basis of Blood Disease (eds Stamatoyannopoulos, G., Majerus, P. W., Perimutter, R. M. & Varmus, H.) Saunders, New York. 2001:135 - 182.

21. Stamatoyannopoulos G., Nienhuis A. W. Hemoglobin Switching, in Stamatoyannopoulos G, Nienhuis AW, Majerus PJ, et al (eds): The Molecular Basis of Blood Disease (ed. 2). Philadelphia, Saunders. 1994;107-56.

22. Poncz M. P., Hentorn P., Stoekert C. Globin gene expression in hereditary persistence of fetal hemoglobin and $\delta\beta^0$-thalassemia.McLean N (ed) Oxford surveys of Eukaryotic Genes, Oxford, UK, Oxford University Press. 1989.

23. Starck J, Sarkar R, Romana M, Bhargava A, Scarpa A. L,, Tanaka M, Chaberlain JW, Weissman SM, Forget BG. Developmental regulation of human γ- and β-globin gene in the absence of the locus control region. Blood 1994;84:1656-1665.

24. Grosveld F, Van Assenddelft G. B, Greeves D. R, Kollias G. Position-independent, high level expression of the human β-globin gene in transgenic mice. Cell 1987;51:957-985.

25. Orkin S. M. Regulation of globin gene expression in erythroid cells. Eur J Biochem 1995;231:271-281.

26. Stamatoyannopoulos J. A, Goodwin A, Joyce T, Lowry C. H. NP-E2 and GATA bining motifs are required for the formation of DNase I hypersensitive site 4 of the human β-globin locus control region EMBO J 1995;14:106-116.

27. Martin I. K., Tsai S. F., Orkin S. H. Increase γ-globin expression in a nondeletion HPFH mediated by an erythroid specific DNA-binidng factor. Nature 1989;338:435-438.

28. Andrew N, Erdjument-Bromage H, Davidson M. B, Tempest P, Orkin S. H. Erythriod transcription factor NF-E2 is a haematopoietic-specific basic leucine zipper protein. Nature 1993;362:722-728.

29. Bieker J. J., Southwood C. M. The erythroid Kruppel-like factor Transactivation domain is a critical component for cell specific inducibility of a β-globin promoter. Mol Cell Biol 1995;15:852-860.

30. Asnano H, Li XS, Stamatoyannopoulos G. FKLF, a novel Kruppel-like factor that activates human embryonic and fetal β-like globin genes. Mol Cell Biol 1999;19:3571-3579.

31. Jane S. M., Ney P. A., Vanin E. F., Gumucio D. L., Nienhuis A. W. Identification of a stage selector element in the human γ-globin gene promoter that fosters preferential interaction with 5' HS2 enhance when in competeion with the β-promoter. EMBO J 1992;11:2961-2969.

32. Higgs D. R., Wood W G., Jarman A. P., Vickers M. A., Wilkie A. O., Lamb J, Vyas P, Bennett J. P. The alpha-thalassemias. Ann N Y Acad Sci 1990; 612:15-22

33. Winichagoon P, Thonglairoam V, Fucharoen S, Tanphaichitr VS, Wasi P. α-Thalassemia in Thailand. Hemoglobin 1988;12:485-498.

34. Weatherall D. J., Higgs D. R., Clegg J. B., Hill A. S., Nicholls R. Heterogenicity and origins of the α-Thalassemias in Thalassemias: Pathophysiology and Management , Part A, edited by S Fucharoen, PJ Rowley, NW Paul, page 3-14, Alan R. Liss Inc., New York 1988.

35. Higgs D.R, α-Thalassemia. Baillier's Clinical Hematology 1993;6:117-150.

36. Leihber S.A. α-Thalassemia. Hemoglobin 1989;13:685-731.

37. Weatherall D.J, Clegg J.B, Boon W.H. The hemoglobin constitution of infants with the hemoglobin bart's hydrops foetolis syndrome. Br J Hematol 1970;18:357-367.

38. Todd D, Lai MC, Beaven G. H., Huehns E R. The abnormal hemoglobins in homozygous alpha-thalassemia. Br J Hematol 1970;19:27-31.

39. Neinhuis A. W., Anagnou N. P., Ley T. J. Advances in Thalassemia reseach. Blood 1984;63:738-758.

40. Orkin S. H., Antonarakis S. R. E, Kazazian H. H. Polymorphism and molecular pathology of the human β-globin gene. Prog Hematol 1983; 13:49-73.

41. Hardison R.C., Chui D.H., Giardine B, Riemer C, Patrinos G.P, Anagnou N, Miller W, Wajcman H. Hb Var: A relational database of human hemoglobin gene server. Hum Mutat 2002;19:225-223.

42. Weatherall D.J., Clegg, J. B. The Thalassemia syndrome , 4th edition , Blackwell Science , Oxford 2001.

43. Chan V., Chan T. K., Chebab F. F., Todd D. Distribution of beta- thalassemia mutations in south china and their association with haplotype. Am J Hum Genet 1987;41:678-685.

44. Fucharoen S., Fucharoen G., Sriroongrueng W, Laosombat V, Jetsrisuparb A, Prasatkaew S, Tanphaichitr VS, Suvatte V, Tuchinda S, Fukumaki Y. Molecular basis of beta-thalassemia in Thailand : analysis of beta-thalassemia mutations using the polymerase chain reaction. Hum Genet 1989;84:41-46.

45. Thein S. L., Winichagoon P., Hesketh C., Best S., Fucharoen S., Wasi P., Weatherall DJ. The molecular basis of beta-thalassemia in Thailand: application to prenatal diagnosis. Am J Hum Genet 1990;47:369-375.

46. Fucharoen S., Winichagoon P. Hemoglobinopathies in Southeast Asia: Molecular biology and clinical medicine. Hemoglobin 1997;21:299-319.

47. Kimura A, Matsunaga E, Takihara Y, Nakamura T, Takagi Y, Lin S, Lee H .Structural analysis of a beta-thalassemia gene found in Taiwan. J Biol Chem 1983;258:2748-2749.

48. Kazazian H. H., Jr.,Orkin S H., Antonarakis S E., Sexton J P., Boehm C. D., Goff S. C., Waber P. G. Molecular characterization of seven beta-thalassemia mutation in Asian Indians. EMBO J 1984;3:593-596.

49. Thein S. L., Kesketh C, Taylor P, Temperley I. J., Hutchinson R. M., Old J. M., Wood W. G., Clegg J. B., Weatherall D. J., Molecular basis for dominantly inherited inclusion body beta–thalassemia Proc Natl Acad Sci USA 1990;80:273-277.

50. Thein S. L. Dominant beta-thalassemia : molecular basis and Pathophysiology. Br J Hematol 1992;80;273-277.

51. Chang J. C., Kan Y. W. Beta^0-thalassmia, a nonsense mutation in man Proc.Natl. Acad Sci USA 1979; 76:2886.

52. Trecartin R. F., Liebhaber S. A., Chang J. C., Lee K. Y., Kan Y. W., Furbetta M, Angius A, Cao A. Beta zero-thalassemia in Sardinia is caused by a nonsense mutation J Clin Invest 1981;68:1012-1017.

53. Treisman R, Orkin S. H., Maniatis T. Specific transcription and RNA splicing defect in mice cloned beta- thalassemia genes. Nature. 1983; 302:591-596.

54. Dobkin C, Pergolozzi R. G., Bahre P, Bank A. Abnormal splice in a mutatant human beta-globin gene not at the site of a mutation. Proc Natl Acad Sci USA 1983; 80:1184-1188.

55. Cheng T. C., Orkin S. H., Antonarakis S. E., Potter M., Sexton J. P., Markham A. F., Giardina P. J., Li A., Kazazian H. H. J. Beta-thalassemia in Chinese: use of in vivo RNA analysis and Oligonucleotides hybridization in systemic characterization of molecular defects. Proc Natl Acad Sci USA 1984;81:2821-2825.

56. Takihara Y ,Matsunaga E, Nakamura T, Lin S, Lee H, Fukumaki Y, Takagi Y. One base substitution is IVS-2 causes a beta+-thalassemia phonotype in a Chinese patient. Biochem Biophys Red Commun 1984;121:324-330.

57. Eastaman J. W., Lorey F, Arnopp J, Currier R. J., Sherwin J, Cunnigham G. Distribution of hemoglobin F, A, S, C, E and D quantities in 4 million newborn screening specimens Clin Chem 1999;45:683-685.

58. Huehns E. R., Bellingham A. J. Diseases function and stability of hemoglobin. Br J Hematol 1969;17:1-10.

59. Wasi P. Geographic distribution of hemoglobin variants in South East Asia. In winter, WP, editors. Hemoglobin variants in human populations, Vol. 2. Florida: CRC Press. 1986. pp 111-127.

60. Kazazian H. H., Jr., Waber P. G., Boehm C. D., Lee J. I., Antonarakis S. E., Fairbanks V. F. Hemoglobin E in Europeans: further evidence for multiple origins of the beta E-globin gene. Am J Hum Genet 1984;36:212-217.

61. Frischer H, Bowman J. Hemoglobin E, and oxidatively unstable mutation. J lab Clin Med 1975; 85:531-539.

62. Itano H. A., Bergren W. R., Sturgeon P. Identification of fourth abnormal hemoglobin. J. Am. Chem. Soc. 1954;76:2278.

63. Hunt J. A., and Ingram V. M. Abnormal human hemoglobins. VI. The clinical difference between Hb A and E. Biochem, Biophys. Acta 1961;49: 520-546.

64. Keller P. and Kohne, E. Hypochromie der Erythrozyten bei heterozygotem Hämoglobin E (β26 Glu→lys). Acta Haematol. 1976;56:276.

65. Na-Nakorn S. and Minnich, V. Studies on hemoglobin E. III. Homozygous hemoglobin E and variants of thalassemia and hemoglobin E. A family study. Blood XII:1957;529-538.

66. Pootrakul S., Assawamunkong S., Na-Nakorn S. Thalassemic trait: Hematologic and hemoglobin synthesis studies. Hemoglobin 1976;1:75-83

67. Wasi P., Winichagoon P., Baramee T., Fucharoen S. Globin chain synthesis in heterozygous and homozygous hemoglobin E. Hemoglobin 1982a;6:75-78.

68. Orkin S. H., Kazazian H. H., Antonarakis S. E., Ostrer H, Goff S. C., Sexton J. P. Abnormal RNA processing due to the exon mutation of beta E-globin gene. Nature 1982;300:768-769.

69. Traeger J, Wood W. G., Clegg J. B., Weatherall D. J. Defective synthesis of HbE is due to reduced levels of beta E mRNA. Nature 1980 ;288:497-499.

70. Benz E. J., Berman B. W., Tonkonow B. L., Coupal E, Coates T, Boxer L. A., Altman A, Adams J. G. Molecular analysis of the beta-thalassemia phenotype associated with inheritance of hemoglobin E (alpha2 beta2 (26) Glu leads to Lys). J clin Invest 1981;68:18-126.

71. Traeger J, Winichagoon P, Pootrakul P, Piankijagum A, Wasi P. Variable severity of Sothereast Asain beta^0-thalassemia/HbE disease. Birth Defects Orig Artic Ser 1987;23:241-248.

72. Boyer S. H., Belding T K., Margolet L, Noyes A. N: Fetal hemoglobin restriction to a few erythrocytes (F cells) in normal human adults. Science. 1975, 188:361-363.

73. Zago M. A, Wood W. G, Clegg J. B, Weatherall D. J, O'Sullivan M, GunsonH: Genetic control of F cells in human adults. Blood 1979, 53:977-986.

74. Wood W. G. Hereditary Persistence of Fetal Hemoglobin and Delta Beta Thalassemia. Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management Edited by: SteinbergMH, ForgetBG, HiggsDR and Nagel RL. Cambridge, UK, Cambridge University Press;2001:356-388.

75. Thein S. L., Sampietro M, Rohde K, Rochette J, Weatherall D. J., Lathrop G M., Demenais F: Detection of a major gene for heterocellular hereditary

persistence of fetal hemoglobin after accounting for genetic modifiers. Am J Hum Genet 1994;54:214.

76. Giampaolo A, Mavilio F, Sposi NM, Care A, Massa A, Cianetti L, Petrini M, Russo R, Capellini MD, Marinucci M: Heterocellular HPFH: Molecular mechanisms of abnormal γ-gene expression in association with β-thalassemia and linkage relationship with the β-globin gene cluster. Hum Genet, 1984; 66:151.
77. Thein S. L., Weatherall D. J. A non-deletion hereditary persistence of fetal hemoglobin (HPFH) determinant not linked to the β-globin gene complex, in Stamatoyannopoulos G, Nienhuis AW, (eds): Hemoglobin Switching, Part B: Cellular and Molecular Mechanisms. New York, NY, Liss, 1989, p 97.
78. Cappellini M. D., Fiorelli G, Bernini L. F: Interaction between homozygous β^0 thalassaemia and the Swiss type of hereditary persistence of fetal haemoglobin. Br J Haematol 1981;48:561.
79. Aksoy M. Thalassemia intermedia: a genetic study in 11 patients. J Med Genet 1970;7:47-51.
80. Thein S. L., Al-Hakim I, Hoffbrand A. V. Thalassemia intermedia : a new molecular basis. Br J hematol 1984;56:333-337.
81. Anand R, Boehm C. D., Kaazian H. H. Jr., Vanin E. F. Molecular characterization of beta zero-thalassemia resulting from a 1.4 Kilobase deletion. Blood 1988;72:636-641.
82. Raiola G, Galati M. C., De Sanctis V, Caruso Nicoletti M, Pintor C, De simone M, Arcuri V. M., Anastasi S. Growth and puberty in thalassemia major. J Pediatr Endocrinol Metab 2003; 16:259-266.
83. Fucharoen S, Winichagoon P, Pootrakul P., Variability severity of Southern Asian β-Thalassemia/HbE disease. Birth Defects 1988b 2(5A):241-248.
84. Fucharoen S,Winichagoon P, Pootrakul P, Piankijagum A, Wasi P. Variable severity of Southeast Asain beta^0-β-thalassemia/HbE disease. Birth defects Orig Artic Ser 1987; 23:241-248.
85. Yang K. G., Kutlar F George E, Wilson J. B., Kutlar A, Stoming T. A., Gonzalez-Redono J. M., Huisman T. H. Molecular characterization of beta-globin

gene mutation in Malay patients with HbE-beta-thalassemia and thalassemia major. Br J Hematol 1989;72:73-80.

86. Thein S. L., Winichagoon P, Hesketh C, Best S , Fucharoen S, Wasi P ,Weatherall D. J. The molecular basis of beta-thalassemia in Thailand: application to prenatal diagnosis. Am J Hum Genet 1990;47:369-375.
87. Winichagoon P, Thonglairoam V. Fucharoen S, Wilairat P, Fukumaki Y, Wasi P. Severity difference in beta- thalassemia/ hemoglobin E syndromes: implication of genetic factors. Br J Hemotal 1993;83: 632-639.
88. Kalpravidh, Komolvanich S, Wilairat P, Fucharoen S. Globin chain turnover in reticulocytes from patients with β^0-thalassemia/HbE disease. Eur j Hematol 1995;55:322-326.
89. Schrier, S. L. Pathophysiology of thalassemia. Current Opinion in Hematology 2002:9;123-126.
90. Ho P. J., Hall G. W., Luo L. Y., Weatherall D. J., Thein S. L. β-thalassemia intermedia: is it possible to consistently predict phenotype from genotype. Br J Haematol 1998;100:70-78.
91. Camaaschella C, Maza U, Roetto A, Gottardi E, Parziale A, Travi M,. Genetic interactions in thalassemia intermedia: analysis of β-mutations, α-genotype, γ-promoters, and β-LCR hypersensitive sites 2 and 4 in Italian patients. Am J Haematol 1995;48:82-87.
92. Winichagoon P, Fucharoen S, Chen P, Wasi P. Genetic factors affecting clinical severity in beta-thalassemia syndrome. J, Pediatric Hematol/Oncology 2000;22:573-80.
93. Kanavakis E, Wainscoat J. S., Wood W. G., Weatherall D. J. Cao A, Furbetta M. The interaction of α-thalassemia with heterozygous β-thalassemia. Br J Haemotol 1982;52:465-73.
94. Garner C, Tatu T., Reittie J.E., Littlewood T., Darley J., Cervino S., Farrall M., Kelly P., Spector T.D and Thein S.L. Genetics influence on F cells and other hematologic variables : a twin heritability study . Blood 2000;95 (1): 342-6.

95. Sampietro, M., Thein, S.L., Conontrereas, M. and Pazmany, L. Variation of HbF and F-cell number with the Gγ Xmn 1 (C-T) polymorphism in normal individuals. Blood, 1992;79:832-833.

96. Thein S. L., Wainscoat J. S., Sampietro M, Old J. M., Cappellini D, Fiorello G, et al. Association of thalassemia intermedia with a β-globin gene haplotype, Br J Haematol 1987;65:367-73.

97. Labie D, Pagnier J, Lapoumeroulie C, Rouabhi F, Dunda-Belkhadia O, Chardin P, et al. Common haplotype dependency of high Gγ-globin gene expression and high Hb F levels in β-thalassemia and sickle cell anemia patients, Proc Natl Acad Sci USA 1985;82:2111-4.

98. Rund D, Oron-Karni V, Filon D, Goldfarb A, Rachmilewitz E, Oppenheim A. Genetic analysis of β-thalassemia intermedia in Israel: Diversity of mechanisms and unpredictability phenotype. Am J Hematol 1997;54:16-22.

99. Galanello R, Dessi E, Melis M. A., Addis M, Sanna M. A., Rosatelli C, et al. Molecular analysis of β^0-thalassemia intermedia in Sardinia. Blood 1989;74:823-827.

100. DeSimone J., Heller P., Hall L., Zweiers, D. 5-Azacytidine stimulates fetal hemoglobin synthesis in anemic baboons. Pro natl Acad Sci, USA. 1982;79:4428-4431.

101. Ley T J., Desimone J., Anagnou N. P., Keller G. H., Humphries, R. K., Turner, P. H., Young, N. S., Heller, P. and Nienhuis, A.W. 5-Acacytidine selectively increases γ-globin synthesis in a patient with β^+-thalassemia. New England of Medicine.1982;307:1469-1475.

102. Ley T J., Desimone J., Noguchi C. T., Turner P. H., Schechter, A. N., Heller, P. & Nienhuis, A.W. 5-Azacytidine increase γ-globin synthesis and reduces the proportion of dense cells in patients with sickle cell anemia .Blood. 1983;62:370-380.

103. Felsenfeld G, Mcghee J. Methylation and gene control. Nature 1982;10:925-934.

104. Cokic V P., Smith R D., Beleslin-Cokic B. B., Njoroge J. M., Miller J. L., Gladwin M. T., Schecter A. N. Hydroxyurea induces fetal hemoglobin by

the nitric oxide-dependent activation of soluble guanylyl cyclase. J Clin Invest 2003;111:231-239.

105. Liakopoulou E, Li Q, Stamatoyannopoulos G. Induction of Fetal Hemoglobin by Propionic and Butyric Acid Derivaties: Correlations between Chemical Structure and Potency of Hb F Induction . Blood Cells Mol Dis 2002;29:48-56.
106. Veith R, Galanello R, Papayannopoulou, T., Stamatoyannopoulos G. Stimulation of F-cell production in patients with sickle cell anemia treated with cytarabine or hydroxyurea. N Engl J Med 1985;313:1571-1575.
107. Liu D. P., Liang C. C, Ao Z. H, Jia P. C, Chen S. S. Treatment of severe beta-thalassemia patients with Myleran. Am J Hematol 1990;33:50-55.
108. Veith R, Papayannopoulou, T.,Kurachi S, Stamatoyannopoulos G. Treatment of baboon with Vinblastine: Insight into the mechanism of pharmocologic stimulation of HB F in the adult. Blood. 1985;66:456-459.
109. Nagai T, Tarumota T, Miyoshi T, Ohmine K, Muroi K , Komatsu N, Sassa S, Ozawa K. Oxidatives stress is involved in hydroxyurea-induced erythroid differentiation. Br J Hematol 2003;121:657-661.
110. Editorial: 5-azacytidine for beta-thalassemia. Lancet 1983;1:36.
111. Craig J. E., Rochette J, Fisher C. A., Weatherall D. J., Marc S, Lathrop G. M., Demenais F, Thein S. L: Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. Nature Genetics 1996, 12:58-64.
112. Wallrapp C., Miller-Pillasch F, Solinas-Toldo S, Lichter P, Friess H, Buchler M, Fink T, Adler G, Gress T. M: Characterization of a high copy number amplification at 6q24 in pancreatic cancer identifies c-myb as a candidate oncogene. Cancer Research 1997, 57:3135-3139.
113. Nelson R. J., Ziegelhoffer T, Nicolet C, Werner-Washburne M, Craig E. A: The translation machinery and 70 kD heat shock protein cooperate in protein synthesis. Cell 1992, 71:97-105.
114. Wallrapp C, Verrier S-B, Zhouravleva G, Philippe H, Philippe M, Gress T. M., Jean-Jean O. The product of the mammalian orthologue of the *Saccharomyces cerevisiae* HBS1 gene is phylogenetically related to

eukaryotic release factor 3 (eRF3) but does not carry eRF3-like activity. FEBS Letters 1998, 440:387-392.

115. Bahar A. Y., Taylor P. J., Andrews L, Proos A, Burnett L, Tucker K. The frequency of founder mutations in the BRCA1, BRCA2, and APC gene in Australian Ashkenazi jews: implications for the generality of U.S. population data. Cancer 2001;92:440-445.
116. Roque M, Godoy C. P., Castellanos M, Pusiol E, Mayorga L. S. Population screening of F508 del (DeltaF508), the most frequent mutation in the CFTR gene associated with cystic fibrosis in Argentina. Hum Mut. 2001;61:1454-1456.
117. Johnson G. C., Esposito L, Barratt B. J. Haplotype tagging for the identification of common disease genes. Nat. Genet. 2000;29, 233-237.
118. Devlin B, Risch N: A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 2002;29, 311-322.
119. Kruglyak L: Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet 1999;. 22, 139-144.
120. Reich D. E., Cargill M, Bolk S. Linkage disequilibrium in the human genome. Nature 2002; 411, 199-204.
121. Kong A, Gudbjartsson DF, Sainz J: A high-resolution recombination map of the human genome. Nat. Genet. 2002; 31, 241-247.
122. Dawson E, Abecasis GR, Bumpstead S. A first-generation linkage disequilibrium map of human chromosome 22. Nature 2002; 418, 544-548.
123. Cummings, M. R. Human Heredity. Principles and issues, 4th edition. Wadsworth Publishing Company: Belmount, USA 1997.
124. Sipichai O., Whitacre J., Munkongdee T., Kumkhaek C., Makarasara W., Winichagoon P., abel K., Braun A and Fucharoen S.Genetic Analysis of Candidate Modifier Polymorphisms in Hb E^0-Thalassemia Patients. Ann. N.Y. Acad. Sci. 2005;1054: 433–438.
125. Sanger, F. Nicklen, S and Coulson, R. A. DNA sequencing with chain terminating inhibitors. Proc. Nat. Acad. Sci. USA.1977;74:5463-567.
126. Prober, J. M., Trainor G. L., Dam R.J., Hobbs F. W., Robertson C. W., Zagursky R. J., Cocuzza A. J., Jensen M. A., Baumeister K. A system for repid DNA

Sequencing with fluorescent chain- terminating dideoxynucleotides. Science. 1987;238,336-341.

127. Merten, Thomas R. Introducing students to population genetics genetics and the Hardy-Weinberg Principal. The American Biology Teacher. 1992;52(2):103-107.
128. Zar J. Biostatistical Analysis.Prentince-Hall: London.1996.
129. Weiss M. J., Orkinn S. H. Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. Proc Natl Acad Sci USA 1995;92:9623-9627.
130. Pevny L, Lin C.S., D' Agati V, Simon M. C., Orkin S. H., Costantini F. Development of hematopoietic cells lacking transcription factor GATA-1. Development. 1995;121:163-172.
131. Leonard M, Brice M, Engel J. D., Papayannopoulou T. Dynamics of GATA transcription factor expression during erythroid differentiation factor expression during erythroid differentiation. Blood 1993;82:1071-1079.
132. Chrivia J. C., Kwok R. P. S, Lamb N, Hagiwara M, Montminy MR, Goodman RG. Phosphroylated CREB binds specifically to the nuclear protein CBP. Nature.1993;365:855-859.
133. Gerd A. Blobel. CREB-binding protein and p300: molecular integrators of hematopoietic transcription. Blood. 2000;95:745-755.
134. Ogryzko V. V., Schiltz L. R., Russanova V, Howard B. H., Nakatani Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell. 1996;87:953-959.
135. Bannister A. J., Kouzarides T. The CBP co-activator is a histone acetyltransferase. Nature.1996;384:641-643.
136. Hung H. L, lau J., Kim J. A., Weiss M. J., Blobel G. A. CREB-binding protein acetylates haematopoitic transcription factor GATA-1 at functional important site. Molecular and Cellular Biology. 1999; 19(5):3496-3505.
137. Gu W., Roeder R. G. Activation of p53 sequence specific DNA binding by acetylation of the p53 C-terminal domain. Cell. 1997;90:595-606.

138. Zhang W, Bieker J. J. Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. Proc Natl Acad Sci U S A. 1998;95:9855-9860.

139. Imhof A., Yang X. J., Ogryzko V. V., Nakatani Y, Wolfe A. P., Gen H. Acetylation of general transcription factors by histone acetyltransferases. Current Biol.1997; 7:689-692.

140. Winichgoon P., Thonglairoam V., Fuchroen S., Wilairat P., Fukumaki Y., Wasi P. Severity differences in β-thalassemia / Haemoglobin E Syndrome: Implication of genetic factors. Br. J of Haematology. 1993; 83: 633-639.

APPENDIX

Clinical Data of Mild and Severe cases take for resequencing purpose

Subject ID	Severity	Age	Gender	Hb (g/dL)	Age on set	Age at lst BITx (yr)	Requirement BITx	Size of Spleen (cm)	Growth retardation	Abs HbF (g/dL)
167M	Severe	27	Female	4.4	2m	2y	occasional	small	160	0.8
659M	Severe	24	Female	5.6	3m	1y	frequently	small	150	0.8
096M	Severe	23	Male	5.0	7m	2y	occasional	12	170	1.4
109N	Severe	16	Female	5.3	3y	4y	occasional	small	145	1.4
128N	Severe	31	Male	5.3	3y	3y	frequently	small	173	1.4
905N	Severe	30	Female	6.7	1-2y	1-2y	frequently	small	158	1.4
078N	Mild	34	Female	8.1	26y	-	no	7	160	4.7
030N	Mild	14	Female	8.4	5y	9y	no	4	148	4.5
002N	Mild	19	Female	8.5	12y	-	no	3	168	3.7
010N	Mild	20	Female	8.0	7y	-	no	4	155	3.7
102M	Mild	32	Male	9.3	6-7y	-	no	5	170	3.6
124M	Mild	22	Male	9.1	6y	12y	rarely	4	173	3.6
097N	Severe	18	Female	3.4	8m	1y	frequently	small	134	0.7
004N	Severe	17	Female	4.7	8m	1y	frequently	small	166	1.2
087N	Severe	28	Female	7.0	1y	6y	frequently	small	149	1.4
040M	Severe	33	Female	6.2	6m	6y	occasional	small	156	1.7
120R	Mild	29	Female	6.8	9y	10y	rarely	7	165	0.9
244CR	Mild	17	Female	7.5	2y	16y	rarely	small	161	1.9
478N	Mild	22	Male	7.8	10y	-	no	4	169	1.5
498R	Mild	17	Female	7.2	1.5y	-	no	1	160	2.0
850S	Severe	19	Female	6.7	3m	3m	frequently	small	151	3.1
473R	Severe	14	Female	6.8	11m	2y	frequently	small	144	3.4
114R	Severe	36	Female	6.5	2y	2y	frequently	small	149	2.6
085N	Severe	11	Female	4.2	1y	1y	frequently	small	120	2.4
1047M	Mild	41	Female	8.5	1-10y	37y	rarely	10	158	3.2
726UT	Mild	19	Male	9.2	3y	3y	occasional	6	167	3.8
006M	Mild	22	Male	8.0	11y	18y	rarely	10	172	2.9
585SB	Mild	4	Female	7.8	2y	-	no	6	98	4.4
048N	Mild	7	Male	8.7	4y	-	no	8	113	4.0
938N	Mild	26	Male	8.7	26y	-	no	6	172	2.8

Clinical Data of Mild and Severe cases take for resequencing purpose

beta-Allele 1	beta-Allele 2	alpha-Allele 1	alpha-Allele 2	b_thal_XmnI	1041480	9402669	987690	9376074	4376364	1014021
Cod 17 (A>T)	Cod 26 (G>A)	Normal	Normal	AG	G	.	A	C	T	A
Cod 17 (A>T)	Cod 26 (G>A)	Normal	Normal	AG	T	.	G	T	.	G
Cod 17 (A>T)	Cod 26 (G>A)	Normal	Normal	AG	GT	AT	GA	TC	TG	AG
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	AG	G	.	A	C	T	A
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	AG	T	AT	G	T	TG	G
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	AG	G	A	.	C	T	A
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	AG	G	A	A	C	T	A
Cod 17 (A>T)	Cod 26 (G>A)	Normal	Normal	AG	G	A	A	C	T	A
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	AG	GT	AT	GA	TC	TG	AG
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	AG	G	.	A	C	T	A
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	AG	G	A	A	C	T	.
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	AG	GT	.	GA	TC	T	AG
IVS II-654 (C>T)	Cod 26 (G>A)	Normal	Normal	G	GT	A	GA	TC	T	AG
Cod 17 (A>T)	Cod 26 (G>A)	Normal	Normal	G	GT	.	GA	TC	T	AG
Cod 17 (A>T)	Cod 26 (G>A)	Normal	Normal	G	T	AT	G	T	TG	G
IVS II-654 (C>T)	Cod 26 (G>A)	Normal	Normal	G	GT	AT	GA	TC	TG	AG
IVS I-5 (G>C)	Cod 26 (G>A)	Normal	Normal	G	T	AT	G	T	TG	G
Cod 17 (A>T)	Cod 26 (G>A)	Normal	Normal	G	GT	AT	GA	TC	TG	AG
Cod 17 (A>T)	Cod 26 (G>A)	Normal	Normal	G	T	AT	G	T	TG	G
Cod 35 (-C)	Cod 26 (G>A)	Normal	Normal	G	G	A	A	C	T	A
IVS II-654 (C>T)	Cod 26 (G>A)	Normal	Normal	G	GT	A	GA	TC	T	AG
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	G	GT	.	GA	TC	T	AG
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	G	T	.	G	T	T	G
IVS I-5 (G>C)	Cod 26 (G>A)	Normal	Normal	G	GT	AT	GA	TC	TG	AG
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	G	GT	AT	GA	TC	TG	AG
Cod 17 (A>T)	Cod 26 (G>A)	Normal	Normal	G	G	.	A	C	T	A
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	G	GT	A	GA	TC	T	AG
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	G	GT	.	GA	TC	T	AG
IVS I-1 (G>T)	Cod 26 (G>A)	Normal	Normal	A	T	AT	G	T	TG	G
Cod 41/42 (-TTCT)	Cod 26 (G>A)	Normal	Normal	A	G	A	A	C	T	A

BIOGRAPHY

NAME Mr. Riyaz Ahmad Pandit.

DATE OF BIRTH 22 June 1975

PLACE OF BIRTH Srinagar, Kashmir, India.

INSTITUTIONS ATTENDED

Kashmir University, India. 1994-2003:
Bachelor of Medical Science.

Symbiosis Institute, Pune, India. 2003-2004:
Post Graduate Diploma of Emergency Medical Service (PGDEMS).

Mahidol University, 2004-2006:
Master of Science (Molecular Genetics and Genetic Engineering).

RESEARCH GRANT Partial Financial Support (TRC).
Support in Part by the Thesis Grant, Faculty of Graduate studies, Mahidol University.

HOME ADDRESS Tenga Pore Nawa Kadal Srinagar Kashmir India 190002.

E-mail : riyazpandit@yahoo.com
riyazpandit@rediffmail.com