DETECTION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN ACQUIRED ANEMIA PATIENTS WITH HIGH HEMOGLOBIN F

SITI NUR NABEELA A'IFAH BINTI MOHAMMAD

UNIVERSITI SAINS MALAYSIA

2024

DETECTION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN ACQUIRED ANEMIA PATIENTS WITH HIGH HEMOGLOBIN F

by

SITI NUR NABEELA A'IFAH BINTI MOHAMMAD

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

March 2024

ACKNOWLEDGEMENT

In the name of Allah, the most gracious and the most merciful. All praises to Allah for the strength and blessings throughout my tough journey.

It is a great pleasure to acknowledge my deepest thanks and gratitude to my supervisors; Dr. Zefarina Zulkafli, Assoc. Prof. Dr. Edinur Hisham Atan and Dr. Maryam Azlan for all the guidance, and advice during my journey for the master's degree. I would like to say thanks to my friends for their constant encouragement. Thank you to Hematology Department, PPSP, and laboratories staff PPSP, PPSK and IPS for providing equipment and guidance in my study. I am deeply grateful to my parents (Mohammad Md Deris and Che Hanisah), my father and mother-in-law (Nik Azhar Ahmad and Rozni Rapandi), my sisters (Jalilah, Mardila, and Saufiah) my brothers (Shahir, Hakiki, Sukur, and Nailul), my brothers in law (Kamarulzaman, Asri and Fazly), my sisters in law (Syeqah and Atirah), my youngest brothers and sister-in-law (Irfan, Asyraf, Aidil and Anis), nieces and nephews for support, prayers and encouragement. In my academic path, Allah tested me by losing my father and having a vehicle crash. He also gave me strength by presenting me with a good husband and a beautiful daughter. Special thanks to my beloved husband, Nik Mohamad Aiman Arif who continuously supported me, sacrificed his time, and patience, and always believed in me throughout my years of study and through the process of research and writing this thesis. To my beloved daughter, Nik Aleena Maryam, thanks for giving me infinite happiness and pleasure when I was writing my thesis, viva voce, and correction thesis. Finally, my thanks go to all the people who have supported me in completing the research work, directly or indirectly. May Allah SWT bless all of us.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATION, SYMBOLS AND UNITS

α	Alpha
β	Beta
δ	Delta
γ	Gamma
°C	Degree celsius
°C/sec	Degree celsius per second
°C/min	Degree celsius per minute
>	Greater than/ modifier letter right arrowhead
∞	Infinity
<	Less than
≤	Less than or equal to
±	Plus-minus
µg	Microgram
µL	Microliter
µM	Micromolar
%	Percent
/	Solidus
~	Tilde
X	Times
A	Adenine
bp	Base pair
C	Cytosine
CE	Cation exchange
Cq	Cycle of quantification
Cd	Codon
dNTPs	Deoxynucleoside triphosphates
HPLC	High performance liquid chromatography
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FBC	Full blood count
fL	femtolitre
g	Gram
g	Gravity
G	Guanine
Hb	Hemoglobin
HbA	Adult hemoglobin
HbA2	Hemoglobin A2
HbE	Hemoglobin E
HbF	Fetal hemoglobin
HPFH	Hereditary persistence fetal hemoglobin
USM	Universiti Sains Malaysia
PPSP	Pusat Pengajian Sains Perubatan
PPSK	Pusat Pengajian Sains Kesihatan
CRL	Central Research Laboratory
IVS	Intervening sequence

k	Kilo
kbp	Kilobase pair
LCR	Locus control region
mg	Miligram
min	Minute
mL	Mililitre
mM	Milimolar
M	Molar
MCV	Mean corpuscular volume
MCH	Mean corpuscular hemoglobin
MCHC	Mean corpuscular hemoglobin concentration
$MgCl_2$	Magnesium chloride
Tm	Melting temperature
n	Number of subjects
ng	Nanogram
ng/mL	Nanogram per microlitre
NTC	Non-template control
pg	petagram
pH	Power of hydrogen
PCR	Polymerase chain reaction
QTLs	Quantitative trait loci
RCF	Relative centrifugal force
RFU	Relative fluorescence unit
SEM	Standard error mean
SNPs	Single nucleotide polymorphisms
SCA	Sickle cell anemia
SCD	Sickle cell disease
T	Thymine
TBE	Tris/Borate/EDTA buffer
V	Volts

LIST OF APPENDICES

PENGESANAN POLIMORFISME NUKLEOTIDA TUNGGAL PADA PESAKIT ANEMIA PEROLEHAN DENGAN HEMOGLOBIN F YANG TINGGI

ABSTRAK

Anemia adalah keadaan klinikal biasa yang boleh diperoleh atau diwarisi. Kebiasaanya, orang dewasa yang sihat menunjukkan paras hemoglobin fetus (HbF) <1%, tetapi kawasan genomik kebolehubahan dan faktor genetik boleh menyebabkan tahap HbF yang lebih tinggi (>1%). Walau bagaimanapun, data tahap HbF yang dikaitkan dengan punca anemia yang diperoleh dalam kalangan penduduk Malaysia adalah terhad. Oleh itu, kajian ini bertujuan untuk menentukan hubungan antara tahap HbF dan polimorfisme nukleotida tunggal (SNPs) dalam pesakit anemia yang diperoleh. Sejumlah 106 daripada 223 pesakit anemia dikesan mempunyai tahap HbF yang tinggi menggunakan cecair kromatografi berprestasi tinggi (HPLC). Terdapat 79 (74.5%) sampel pesakit daripada 106 pesakit mempunyai HbA2 yang tinggi (=/>3.2%) dan telah diuji dengan sistem mutasi penguatan refraktori multipleks-reaksi rantai polimerase (ARMS-PCR) untuk mutasi gen β-globin manakala baki 27 sampel pesakit dengan HbF tinggi yang mempunyai HbA2 lebih rendah (<3.2%) telah diuji menggunakan ruang multipleks-PCR untuk deletan empat gen β-globin (Siriraj J Gγ(Aγδβ)o-thal, Thai (δβ)°-thalassaemia, keturunan berterusan hemoglobin-6 fetus, dan Hb Lepore). Mutasi gen β-globin dikesan dalam 50 pesakit menggunakan MARMS-PCR, 37 heterozigot Cd26, 6 heterozigot IVS 1-5, 3 heterozigot Cd 41/42, 1 heterozigot IVS 1-1, 2 sebatian heterozigot Cd26 bersama Cd8/9, dan 1 sebatian heterozigot Cd26 bersama Cd41/42. Walau bagaimanapun, tiada deletan gen β-globin dikesan dalam kesemua 27 pesakit. Selain itu, tiada perbezaan yang signifikan antara

tahap HbF pesakit yang diperoleh dengan diwarisi. 36 sampel pesakit dengan tahap HbF tinggi dan 5 sampel pesakit dengan tahap HbF normal telah dipilih untuk analisis pengenotipan menggunakan platform mikrotatasusun Infinium Asian Screening untuk mengenal pasti SNPs. Dua SNP yang paling ketara ditemui dalam kajian ini ialah rs73170684 *GSTK1* dan rs2893863 *CDK1* diperhatikan sebagai varian paling ketara yang mencapai ambang yang ketara GWA ($p<10^{-8}$). Walau bagaimanapun, SNP daripada lokus utama biasa, seperti gen *HBS1L-MYB* dan *BCL11A*, tidak dikesan dalam kajian ini. Oleh itu, penemuan ini boleh digunakan sebagai peramal genetik baru dan garis panduan untuk kajian pada masa hadapan dalam tahap HbF yang tinggi dalam kalangan pesakit anemia untuk rawatan yang lebih baik.

DETECTION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN ACQUIRED ANEMIA PATIENTS WITH HIGH HEMOGLOBIN F

ABSTRACT

Anemia is a common clinical condition that can be either acquired or inherited. Normally, adults display fetal hemoglobin (HbF) levels of <1%, but variability in genomic regions can cause higher HbF levels (>1%). Unlike inherited anemia, clinical and genetic data on HbF levels associated with acquired causes of anemia among the Malaysian population are still scarce. Therefore, this study aims to determine the association between HbF level and single nucleotide polymorphisms (SNPs) in acquired anemia patients. A total of 106 out of 223 anaemic patients were detected to have high HbF levels using high performance liquid chromatography (HPLC). From 106 patients with high HbF, 79 (74.5%) samples were found to have high HbA2 (=/>3.2%) and were tested with multiplex amplification refractory mutations-system polymerase chain reactions (ARMS-PCR) for β-globin gene mutation while the remaining 27 anemic patients with high HbF has lower HbA2 (<3.2%) were tested using multiplex gap-PCR for four β-globin gene cluster deletion (Siriraj J Gγ(Aγδβ)^{o}-thalassemia, Thai (δβ)°-thalassaemia, HPFH-6, and Hb Lepore). β-globin gene mutations were detected in 50 patients using multiplex ARMS-PCR, 37 heterozygous Cd26, 6 heterozygous IVS 1-5, 3 heterozygous Cd 41/42, 1 heterozygous IVS 1-1, 2 compound heterozygous Cd26 with Cd8/9, and 1 compound heterozygous Cd26 with Cd41/42. However, no β-globin gene cluster deletion detected in all 27 patients. Besides, there was no significant difference between the HbF levels of acquired and inherited anemic patients. 36 genomic DNA of samples with high HbF and no mutation and deletion together with 5 DNA samples with normal HbF level were chosen for

analysis using the Infinium Asian Screening SNPs microarray platform. Two SNPs, rs73170684 and rs2893863 in *GSTK1* and *CDK1* gene were observed as the most significant variants that achieved GWA significant threshold ($p<10^{-8}$). However, other SNPs from common major loci associated with high HbF, such as those in *HBS1L-MYB* and *BCL11A* genes, were not significant. Thus, these findings can be used as a new genetic predictor and guideline for future studies in high HbF levels among acquired anemic patients for better treatment.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Acquired anemia is caused by factors outside and not inherited from parent gene. The acquired anemia can be caused by drug intake, chronic disease, infection, and micronutrient deficiency (lack of iron), and pregnancy which all together can reduce erythroid precursors (Sankaran & Weiss, 2015; Sokolova, Mararenko, Rozin, Podrumar, & Gotlieb, 2017). In Malaysia, 19.3% to 57.4% of pregnant women associated with anemia, while, 34.6% of them caused by iron deficiency anaemia (Abd Rahman, Idris, Isa, Rahman, & Mahdy, 2022). Additionally, 75.8% of Malaysian patients with chronic kidney disease who do not receive dialysis were found to have anemia (Salman et al., 2016). Consequently, the prevalence of acquired anaemia may influence the progression of the medical condition and decrease the anemia patients chance of survival (Goodnough & Nissenson, 2004).

Fetal hemoglobin (HbF) consists of two α-globin and two γ-globin chains. The γ-globin chains of HbF are encoded by *HBG1* and *HBG2*, which are part of the β-gene cluster (Steinberg, 2020). During fetal development, the switching process from HbF to adult hemoglobin (HbA) occurs during the first year of life until the normal range (<1%) of HbF is reached in children and adults (Thomas & Lumb, 2012). High HbF levels in children and adult are caused by the reactivation of the *HBG2* gene, and usually observed in inherited or acquired anemia (Sankaran & Orkin, 2013). Acquired anemia with high HbF can be caused by cancer, pregnancy, and drug induced. The high HbF can be indicator to patient who have cancer such as leukemia, and breast cancer (Wolk, Martin, Reinus, & St, 2006). High HbF also found during pregnancy due to modification of erythropoiesis. Besides, certain drugs such as hydroxyurea,

thalidomide and butyrate that used in hematological disorder, autoimmune disease and cancer can ameliorate severity of anemia by increase HbF level (Canani, Di Costanzo, & Leone, 2012).

In contrast, inherited anemia cases associated with high HbF levels among adults, are due to mutations or deletions in the β-globin gene cluster located at chromosome 11. These molecular event leading to several disorders, such as β-thalassemia, sickle cell anemia (SCA), hemoglobin (Hb) E disorders, hereditary persistence of fetal hemoglobin (HPFH), and δβ-thalassemia (Carrocini et al., 2011; Galanello and Origa, 2010).

Other polymorphism including single nucleotide polymorphisms (SNPs) in the *BCL11A* gene at chromosome 2 (rs11886868, rs1427407, rs766432, or rs6545816), *HBS1L-MYB* intergenic region (*HMIP)* at chromosome 6 (rs9399137), and *XMN1-HBG2* at chromosome 11 (rs7482144) were also reported to be associated with high HbF and influence severity of disease in thalassemia and sickle anemia patients (Akinsheye et al., 2011; Fong, Menzel, Lizarralde, & Barreto, 2015). A previous study found that acquired aplastic anemia with elevated HbF was correlated with *XMN1-HBG2* polymorphisms (Shimmoto, Vicari, Fernandes, Guimarães, & Figueiredo, 2006). The *BCL11A* is gene that related with high HbF and breast cancer but there is no any reported SNPs of *BCL11A* that related with acquired anemia and high HbF patient. However, the data of SNPs related with high HbF in acquired anemia is limited.

1.2 Justification of study

Anemia remains one of the major public health problems affecting communities. Unlike inherited anemia, genetic basis of HbF levels associated with acquired anemia are scarce, particularly among the Malaysian population. Therefore, this study was conducted to identify SNPs associated with acquired anemia. Hence, finding from this study can be used as a potential disease biomarker for acquired anemia and subsequently used for diagnosis and treatment.

1.3 General objectives

To detect SNPs in acquired anemia with high HbF in Hospital Universiti Sains Malaysia.

1.3.1 Specific objectives

1. To correlate hematological parameter with acquired anemia.
2. To determine the prevalence of acquired anemia
3. To identify SNPs in acquired anemia patients using Infinium Asian Screening 24-array assay.

1.4 Hypothesis

1. The increase level of HbF associated with inherited and acquired anemic patients.
2. The single nucleotide polymorphisms (SNPs) are associated with high HbF level in acquired anemia patients.

CHAPTER 2

LITERATURE REVIEW

2.1 Anemia

Anemia is a hematological condition that decreases Hb levels, resulting in a lower oxygen carrying capacity. It is caused by blood loss, lower red blood cell (RBC) production, and hemolysis, and can lead to ineffective erythropoiesis (Brandow, 2017). According to the World Health Organization (WHO), the Hb levels in anemia patients are less than 13 g/dL for men, less than 12 g/dL for women, and less than 11 g/dL for pregnant women, as shown in Table 2.1. Pregnant women and children aged 6 to 59 months may have severe anemia if their Hb levels are less than 7 g/dL, while the others may have severe anemia if their Hb levels are less than 8 g/dL (Table 2.1). Approximately 1.62 billion people among the global population suffers from anemia, and it is highly prevalent in children, the elderly and young women (Chaparro & Suchdev, 2019; McLean, Cogswell, Egli, Wojdyla, & De Benoist, 2009). The high prevalence of anemia may be due to acquired or inherited causes.

Table 2.1 Hemoglobin levels in different age groups (g/dL) ± (WHO, 2011)

Population	Non-anemic	Mild	Moderate	Severe
Children 6 - 59 months of age	>11	10.0-10.9	7.0-9.9	< 7.0
Children 5 - 11 years of age	>11.5	11.0-11.4	8.0-10.9	< 8.0
Children 12 - 14 years of age	>12.0	11.0-11.9	8.0-10.9	< 8.0
Non-pregnant women (>15)	>12.0	11.0-11.9	8.0-10.9	< 8.0
Pregnant women	>11.0	10.0-10.9	7.0-9.9	< 7.0
Men (15 years of age and above)	>13.0	11.0-12.9	8.0-10.9	< 8.0

Adapted from the World Health Organization (2011)

2.1.1 Inherited anemia

Hb genetic disorders, including thalassemia and sickle cell disease (SCD), are significantly associated with inherited anemia (Galanello & Origa, 2010). Thalassemia

is an autosomal recessive disorder that causes the reduction or absence of globin chains, which leads to a decrease in Hb levels and the formation of abnormal Hb (Cao & Galanello, 2010). The defective syntheses of alpha (α) and beta (β) globin chains result α-thalassemia and β-thalassemia, respectively. SCD, meanwhile, is caused by an abnormal Hb structure due to a single amino acid substitution in the β-globin chain, which changes the biconcave shape of RBCs into a sickle shape structure (Habara & Steinberg, 2016; Sankaran & Weiss, 2015). These inherited diseases result in a lower oxygen carrying capacity due to the abnormal structures of Hb and reduction of Hb levels (Cao & Galanello, 2010).

2.1.2 Acquired anemia

There are several acquired factors that can cause acquired anemia such as drug intake, chronic diseases, infections, cancer, and nutrition deficiency (Balarajan, Ramakrishnan, Özaltin, Shankar, & Subramanian, 2011; Chulilla, Colás, & Martín, 2009; Sankaran & Weiss, 2015). Some drugs can destroy blood cells and reduce the production of blood cells, causing hemolytic anemia and aplastic anemia, respectively (Sukhal & Gupta, 2014). For example, cephalosporins is one of the antibiotic drugs that bind covalently to the membrane protein of RBCs and form drug-coated RBCs (Garratty, 2012). This formation will induce antibodies production. The antibodies can be IgG or IgM that will bind to drug-coated RBC causing acute intravascular hemolysis and a sudden drop in Hb levels (Guleria, Sharma, Amitabh, & Nair, 2013) (Garratty, 2009).

Chronic diseases, such as kidney disease, can affect erythropoiesis and contribute to mild or moderate anemia (Salman et al., 2016). Parasitic infections, such as hookworm (*Necator americanus* and *Ancylostoma duodenale*), malaria (*Plasmodium*

falciparum and *Plasmodium vivax*) or chronic infections, such as tuberculosis (TB) and human immunodeficiency virus (HIV) infection can lead to anemia as well. TB and HIV infections are associated with high risk of anemia, and have been linked to morbidity and mortality (Saathoff et al., 2011).

Anemia is also caused by cancer, such as gastrointestinal cancer, hematological cancer, and solid tumor (Busti, Marchi, Ugolini, Castagna, & Girelli, 2018). The exact causes is multifactorial such as tumor related bleeding, chemotherapy, iron insufficiency in gastrointestinal, urogenital, and gynaecological cancers, whereas bone marrow infiltration by cancer cells is frequent in breast and prostate cancer (Gaspar, Sharma, & Das, 2015).

Nutritional deficiency, such as iron deficiency, can also lead to anemia, which can occur due to blood loss during heavy menstruation, surgery, and pregnancy. The prevalence of iron deficiency anemia has increased significantly in the central Asia region (64.7%), Latin America (62.3%) and South Asia (54.8%) (Kassebaum et al., 2014). Inadequate vitamin B12 and folate are the most frequent factor of nutrition deficiency anemia (Chaparro & Suchdev, 2019; Zimmermann & Hurrell, 2007). Anemia due to inadequate erythropoiesis is driven by vitamin B12 and folate deficiency, which impedes the production of purine and thymidylate, thus inhibiting DNA synthesis and promoting apoptosis in the erythroblast. This results in a lower quality of life, including fatigue and weight loss. Erythropoiesis requires proper dietary intake of iron, folate, and vitamin B12 to sustain a better lifestyle (Chaparro & Suchdev, 2019).

2.2 Hemoglobin

Hb is a protein complex in erythrocytes that carries oxygen throughout the body to different tissues and return carbon dioxide to the lungs. Hb consists of four covalently linked protein structures, two α-globin chains and two non-α-globin chains. The α-globin chains are expressed by the genes on chromosome 16 (α_1, α_2, and ζ), while non α-globin chains are expressed by β-globin gene locus on chromosome 11 (ε, Gγ, Aγ, β, and δ) (Figure 2.1 A and B) (Schechter et al., 2008). The expression of both genes occurs during the embryonic, fetal and adult stages and they are involved in the production of several types of hemoglobin (Table 2.2) (Kaufman & Lappin, 2019). Two different processes occur in the β-globin locus at chromosome 11p, which are the ε-globin chains switching into γ-globin chains at six to eight weeks of gestation in fetal development, and γ-globin genes switching into δ-globin and β-globin genes shortly after birth (Schechter et al., 2008).

In normal adults, there are three types of Hb, which are adult Hb (HbA and HbA2) and HbF, as shown in Table 2.2 and Figure 2.1 (A). During the fetal and infant stages, HbF is the predominant type of Hb, consisting of two α-globin and two γ-globin chains (α2γ2) that are completely active. The main or normal globin chain in adults is HbA, which consists of two α-globin and two β-globin chains (α2β2) (Figure 2.1 C). The minor constituent of adult hemoglobin is HbA2, which comprises two α-globin and two δ-globin chains (α2δ2). Hence, among adults, 92% of the total Hb concentration is made up of HbA, followed by 2.5% of HbA2 and less than 1% of HbF (Manning et al., 2007).

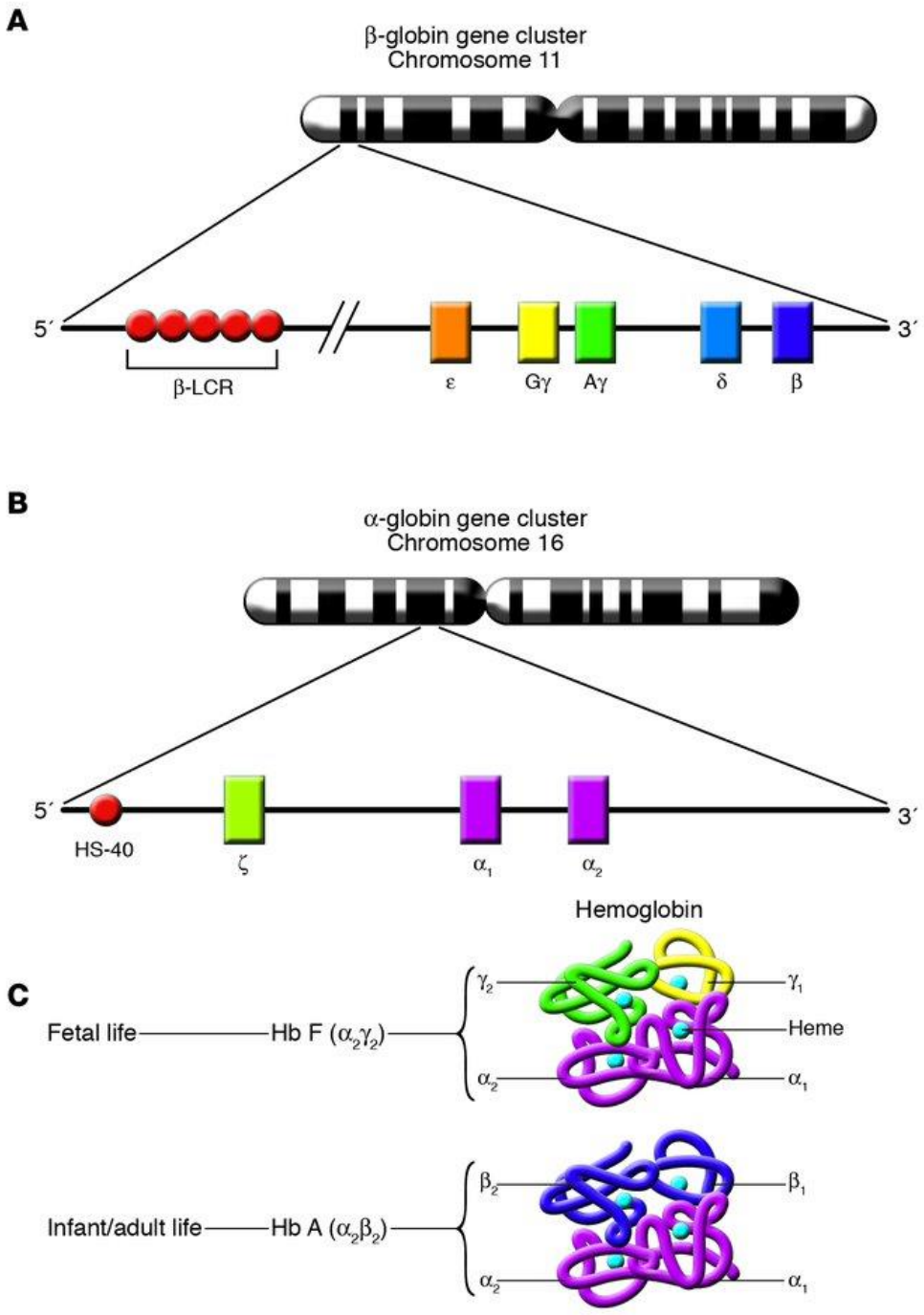

Figure 2.1 (A) The gene of β-globin gene cluster on chromosome 11. (B) The gene of α-globin gene cluster are present on chromosome 16. (C) Organizational structure at fetal and adult hemolgobin. Adapted from Frenette and Atweh (2007).

Table 2.2 The developmental stage of hemoglobin in humans

Hemoglobin (Hb)	Phenotype of Hb	Development stage
HbA	α2β2	Adult
HbA2	α2δ2	Adult
HbF	α2γ2	Fetal
Gower 1	ζ2ε2	Embryonic
Gower 2	α2ε2	Embryonic
Portland I	ζ2γ2	Embryonic

2.3 Fetal hemoglobin

The production of HbF occurs during the early stages of embryonic development throughout the gestation period to replace embryonic hemoglobin, such as Gower I, Gower II and Portland Hb, with the syntheses of two α-globin chains and two γ-globin chains in HbF (Table 2.2) (Curcio, E, M, & V, 2019; Kaufman & Lappin, 2019). The γ-globin chains are expressed by the *HBG2* and *HBG1* genes on chromosome 11p (Akinsheye et al., 2011; Sankaran & Orkin, 2013).

The presence of glycine and alanine, which have a high oxygen affinity, in the γ-globin chain subunit in those with high HbF levels is beneficial for the transportation of oxygen from the maternal to fetal circulation during fetal development (Kaufman & Lappin, 2019). In fetuses and infants, 80% of the total Hb are HbF, while the rest are HbA, and almost all HbF will be replaced with HbA after six months of age. The switching continues until the HbF is lower than 1% (Akinsheye et al., 2011). However, if the HbF levels remain high (>1%) in adults, it may be attributed to several causes such as inherited or acquired causes.

2.3.1 Increase in fetal hemoglobin due to inherited causes

The β-globin gene cluster located at chromosome 11 consists of Epsilon (ε), G-gamma (Gγ), A-gamma (Aγ), delta (δ), and β-globin genes (Cao & Galanello, 2010). More than 200 variations in the β-globin gene cluster have been detected, from the absence of β-globin chain production to the reduction of β-globin chain production (Thein, 2005). Some of these mutations affect the β-globin expression and functions (deletion, insertion, or substitutions) and cause δβ-thalassemia, HPFH, and β-thalassemia and its variants (Cao & Galanello, 2010).

2.3.1(a) β-globin gene cluster mutation

In β-thalassemia, β-globins production are affected and unstable α-globin chains will be increased (Cao & Galanello, 2010). The excess α-globin chains disturb the differentiation of erythroid and leads to ineffective erythropoiesis that will cause anemia. Thus, the free α-globin will bind with γ-globin chains and increase HbF level, which ameliorate the clinical severity of β-thalassemia and its variants (Sankaran & Orkin, 2013).

β-thalassemia can be passed down from the parent to the child, and is classified into three groups, which are β-thalassemia minor, intermedia, and major. The β-thalassemia minor, or β-thalassemia trait is a condition in a person with heterozygous β-thalassemia. Silent mutation has also been identified in β-thalassemia trait patient (Thein, 2013). Individuals with β-thalassemia minor can be asymptomatic to mild anemia (8-11g/dL) (Galanello & Origa, 2010). Clinically the β-thalassemia intermedia is more serious than minor thalassemia but less severe than major β-thalassemia. Most of patients with β-thalassemia intermedia are compound heterozygote or homozygote β-globin gene mutation and results in moderate anemia (6-10g/dL) due to ineffective

erythropoiesis (Cao & Galanello, 2010). The β-thalassemia major is the most severe transfusion-dependent anemia and associated with severe microcytic hypochromic anemia (less than 6g/dL) which requires regular blood transfusion (Cao & Galanello, 2010). Data from the Malaysian Thalassemia Registry showed 7,984 people are registered as β-thalassemia patients in 2018, with 33.53% of them having thalassemia major, 9.37% thalassemia intermedia and 57.1% thalassemia trait.

Generally, HbF levels are slightly increased in patients with heterozygous β-thalassemia due to mutations in the β-globin gene (Steinberg & Thein, 2016). HbF levels become higher in patients with heterozygous β-thalassemia due to mutations in the promoter and *HBG* gene (Steinberg & Thein, 2016). The HbF level ranges from 1% to 10% in minor β-thalassemia to 100% in β-thalassemia major (Mosca, Paleari, Leone, & Ivaldi, 2009; Steinberg & Thein, 2016).

High HbF levels are found in patients with compound heterozygous HbE/β-thalassemia, which is caused by a co-inheritance of a β-thalassemia allele and a structural variant of HbE (Rujito et al., 2016). The abnormal splicing site of the HbE variant occurs at codon 26 of the β-globin chain mutation (substitution of glutamic acid with lysine) and results in globin chain imbalance and ineffective erythropoiesis (Das et al., 2021). In Southeast Asia, including Malaysia, Thailand, Laos and Cambodia have high prevalence of HbE/β-thalassemia cases (50-70%), which may spread through interracial marriage (Fucharoen & Winichagoon, 2011; George E, 2013; Lim et al., 2015; Thachil, Owusu-Ofori, & Bates, 2014). Five percent of Malaysians have the HbE gene mutation, and Orang Asli in peninsular Malaysia have been discovered to have a

high prevalence of the mutation, with the Senoi tribe of Orang Asli having an 18.6% of HbE mutation (George E, 2013; Koh et al., 2017).

2.3.1(b) β-globin gene cluster deletion

High HbF levels in adults are also found to be associated with large deletion in β-globin gene such as δβ-thalassemia, HPFH, and Hb Lepore. This causes an increase production of γ-globin when partial or complete deletions occur in the β-globin gene cluster (Cao & Moi, 2000).

The δβ-thalassemia produced by the deletions of the β and/or δ genes in the β-globin gene cluster result in the elevation of HbF levels, but normal levels of HbA2, among adults (Mansoori, Asad, Rashid, & Karim, 2016). Cases of δβ-thalassemia with high HbF levels is highly present in Southeast Asia and divided by the levels of HbF; homozygous δβ-thalassemia with 100% HbF and heterozygous δβ-thalassemia with 17% to 30% HbF (Chalaow, Thein, & Viprakasit, 2013; Mansoori et al., 2016). β-thalassemia intermedia may be present in patients with homozygous δβ-thalassemia due to overexpression of HbF level (Mansoori et al., 2016). Heterozygous δβ-thalassemia with high HbF levels has parallel clinical symptoms with thalassemia trait in terms of normal or low HbA2 levels (Jain et al., 2022). Therefore, the severity of anaemia was alleviated by the expression of the γ-globin gene leading to an increase in HbF concentrations due to the absence of δ and β chains (Sokolova et al., 2017).

The deletions of β and δ genes also lead to HPFH resulting in an increase in HbF levels ranging around 5% to 100% depends on the type of the mutation (Galanello, 2013; Shaukat, Pudal, Yassin, Höti, & Mustafa, 2018). The HPFH in heterozygotes HbF levels may rise to between 5% and 30% with normal Hb while HPFH homozygotes

have mild hypochromia and microcytosis red cells; 100% of the hemoglobin is F, and there is no anemia (Sharma et al., 2020).

Another deletion is Hb Lepore due to crossover between δ-globin and β-globin genes, which produces partial δ-globin and β-globin chains. This causes the elevation of HbF levels at ranges of between 2% and 57%, with low or normal HbA2 levels (Josephine, Elizabeth, & Menaka, 2005; Pirastru, Manca, Trova, & Mereu, 2017). Patients with Hb Lepore display β-thalassemia phenotypes, such as mild or moderate Hb levels, and microcytic hypochromic anemia (Pirastru et al., 2017).

2.3.2 Increase in fetal hemoglobin by acquired causes

There are several acquired causes associated with the increase in HbF levels, such as leukemia, pregnancy, fetomaternal hemorrhage, autoimmune hemolytic anemia, and drugs (Mandal & Kartthik, 2019; Mendek-Czajkowska et al., 2003; Mosca et al., 2009).

An increase in HbF levels in leukemia patients can act as an indicator for carcinogenesis hazard through the reactivation of the γ-globin gene (Wolk, Martin, & Nowicki, 2007). High HbF levels act as a marker in the detection of hematopoietic malignancy, such as juvenile myelomonocytic leukemia, myeloblastic leukemia, and erythroleukemia. Hence, high HbF levels can be a tumour marker for cancer patient management (Wolk et al., 2007).

Additionally, high HbF levels is also present during pregnancy. HbF is elevated during the first or second trimester of pregnancy (Yamada et al., 2013). The increase in

HbF levels in the earlier stages of pregnancy might be stimulated by the modification of erythropoiesis (Chambers, Davies, Evans, Birchall, & Kumpel, 2012). The rapid expansion of erythroid and high F cell levels also result in slightly higher HbF levels during the first trimester of pregnancy, and may be related to hemoglobinopathies, SCD or genetic modifiers, and slowly decreases in the third trimester (Yamada et al., 2013). Fetomaternal hemorrhage involves the blood of the fetus enters the maternal circulation after the placental barrier is disrupted and increase the Hb F level of the mother (Krywko, Yarrarapu, & Shunkwiler, 2022). It may occur during the first trimester and can lead to fetal anemia, with clinical symptoms such as irregular heart rate patterns or reduced fetus weight (Murji et al., 2012). However, fetomaternal hemorrhage is a rare pregnancy complication and can be detected by the presence of high HbF levels (Sá & Moura, 2021).

Autoimmune hemolytic anemia (AIHA) is a rare acquired disorder. AIHA occurs when the antibodies act against RBCs, resulting in the destruction of premature RBCs. This leads to erythropoiesis stress. This condition may cause an increase in HbF levels of between 16% and 53.3% through the reactivation of γ-globin expression (Mandal & Kartthik, 2019; William, Rusmawatiningtyas, Makrufardi, & Widjajanto, 2021).

HbF levels can also increase by inducing the γ-globin gene using drugs, such as hydroxyurea, thalidomide, and butyrate. Hydroxyurea is an antimetabolite drug that inhibits ribonucleotide diphosphate reductase through cell stress signalling pathway, transcription factor, and expression of miRNA in post-transcription modification, that induce HbF production (Wang et al., 2002; Yasara, Premawardhena, & Mettananda,

2021). Thalidomide is a synthetic glutamic acid that can elevate the HbF levels in patients with hematological disorders, autoimmune diseases and multiple myeloma. Butyrate, meanwhile, is known as a histone deacetylase inhibitor, and translates the γ-globin mRNA and increases the HbF levels (Akinsheye et al., 2011). The enhancing of HbF levels by butyrate is used in the treatment of cancer, hemoglobinopathies and sickle cell anemia, and promotes anti-inflammatory activity (Akinsheye et al., 2011; Canani et al., 2012). These therapeutic agents have different ways of increasing HbF synthesis. Predictably, the amelioration of the severity of anemia in terms of Hb concentration was observed after reliable treatment is used by increasing the HbF levels (Canani et al., 2012; Crona et al., 2016).

2.3.3 Increase in fetal hemoglobin by genetic modifiers

Genetic modifiers are genetic variants which can influence the disease severity by enhancing or suppressing the target gene that causes an alteration of phenotypic output (Rahit & Tarailo-Graovac, 2020). A modifier variant can influence the target gene's phenotype by interacting genetically, biochemically, or functionally with one or more genes. Genetic modifiers are involved in the modification of pleiotropic (one gene affecting multiple traits) phenotypes, resulting in different phenotype and disease characterisation becomes more challenging due to individuals carrying two or more monogenic disorders (Rahit & Tarailo-Graovac, 2020).

An increase in HbF levels is also associated with genetic modifiers and acts as a secondary modifier in the β-globin gene. The genes involved in the γ-globin expression are *HBS1L-MYB, BCL11A*, and *XMN1-HBG2* (Thein, 2017). These genes have 20% to 50% of variations in HbF levels through the regulation of fetal globin

expression, increasing in HbF levels and reducing the imbalance of α-globin chains (Jaing et al., 2021; Thein, Menzel, Lathrop, & Garner, 2009). Genetic modifiers are mainly assessed using genome-wide association studies (GWAS) (Green & Barral, 2011). GWAS have been used to identify genetic modifiers in patients with inherited anemia, but studies involving those with acquired anemia with high HbF levels are still scarce. In order to find factors that regulate HbF levels, thousands of SNPs are used to screen for regions of interest with GWAS, which may function as a modifier of high HbF levels (Mohammad et al., 2022). This information can be used to develop a therapeutic agent in ameliorate the clinical severity of diseases (Munshi, Dadeech, Babu, & Khetarpal, 2015).

2.3.3(a) SNPs in patients with high HbF levels

SNPs are single-base alterations in the DNA sequence and common genetic variation at specific regions in genome (Figure 2.2). Human DNA is 99% identical across the population, but one percent is thought to represent a variation that makes unique and different from one another (Sripichai & Fucharoen, 2007). Several millions stable SNPs are distributed in the human (Sripichai & Fucharoen, 2007).

SNPs can exist in the coding, non-coding and intergenic regions (regions between gene). The SNPs in coding site change the amino acid sequence and it can encode around three to five percent of DNA sequences for the protein production (Robert & Pelletier, 2018; Sripichai & Fucharoen, 2007). However, most of SNPs commonly found in non-coding site (Sripichai & Fucharoen, 2007). The SNPs serves as important markers for comparatives or genomic studies due to they do not alter encoded proteins in noncoding regions.

Several factors contribute to the significance of SNPs in genetic studies; first, the distribution of SNPs can be used to trace the population of study. Most of SNPs are related with inherited from one to next generation. SNPs are also responsible for genetic disease after the alteration of regulatory region and sequence of gene. SNPs may not necessarily cause a disorder but some of SNPs are responsible for genetic effect that produces susceptibility to most disease. SNPs can be mapped and used as genetic marker for association studies (Jian & Li, 2021; Sripichai & Fucharoen, 2007). For example, SNPs association studies in disease can be used by comparing the case and control group (Sripichai & Fucharoen, 2007). SNP markers are useful for medical applications, including hematology research, early tumor diagnosis, antibiotics, and anticancer therapy (Govindarajan, Duraiyan, Kaliyappan, & Palanisamy, 2012).

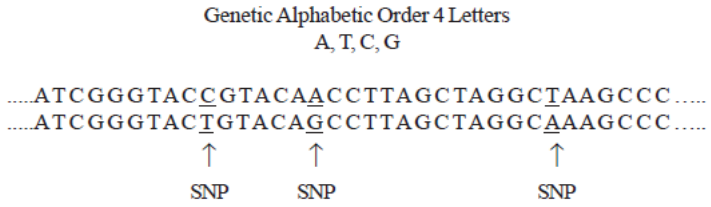

Figure 2.2 Example of SNPs in the genome sequence. Adapted from Sripichai and Fucharoen, (2007).

SNPs is the most common type of genetic modifier in high HbF levels patients (Table 2.4). SNPs may be a useful tool as a biomarker in cancerous and non-cancerous diseases, such as breast cancer, diabetes, obesity, Alzheimer's disease, bipolar disorder, and hypertension, as well as anemia. The detection of SNPs is also used to determine the specific region in the gene that is associated with high HbF levels, which may be convenient as an important biomarker signature for therapeutic purposes, especially in the management of anemia.

Genetic studies have shown the association of SNPs at the three major quantitative trait loci (QTLs); *BCL11A*, *HBS1L-MYB* and *XMNI* in the production of high HbF levels (Table 2.4) (Lettre et al., 2008; Rujito et al., 2016; Thein et al., 2009; Uda et al., 2008).

Table 2.3 Example of QTLs that are associated with HbF levels

Genes	SNPs
BCL11A	rs4671393
	rs11886868
	rs676432
	rs7557939
	rs7599488
	rs6545816
	rs1018987
	rs1427407
	rs6706648
	rs6738440
HBS1L MYB intergenic polymorphism (HMIP)	rs28384513
	rs9399137
	rs4895441
XMN1-HBG2	rs7482144

The BCL11A gene (B-cell lymphoma/leukemia 11A) located at chromosome 2p16.1 is highly expressed in hematopoietic stem cells and the brain, and involved in fetal-to-adult hemoglobin switching (Yin, Xie, Ye, Wang, & Che, 2019). The *BCL11A* gene is a direct regulator of HbF production (Sankaran, Xu, & Orkin, 2010). Several common SNPs of *BCL11A* associated with high HbF levels are found in the African-American SCD, Brazil SCD, Northern European, Chinese, Thailand, and Sardinian populations with β-thalassemia or compound heterozygous β-thalassemia and HbE disorders. These include as rs11886868, rs4671393, rs766432, rs7557939, rs7599488, rs6545816, rs1018987, rs1427407, rs6706648, rs6738440 and rs7606173 (Danjou et al., 2012; Fanis, Kousiappa, Phylactides, & Kleanthous, 2014; Rujito et al., 2016; Sebastiani et al., 2015).

The *HBS1L-MYB* intergenic region (*HMIP)* at chromosome 6q23.3; *HBS1L* is a G-protein or elongation factor, while *MYB* is an erythroid transcription factor. *HBS1L-MYB* significantly increases HbF levels in the presence of rs9399137 (T→C) and rs11759553 (A→T) SNPs among the Sabahan population with Filipino β^0-deletion (Teh et al., 2017). However, most studies show that SNPs within intergenic region of *HBS1L-MYB* are highly associated with the elevation of HbF in those with β-thalassemia in the African, Chinese, Northern European and Sardinian populations.

Another common genetic variant is *XmnI* polymorphism at position -158 of the $^{G}\gamma$ globin (HBG2) gene (C→T). The most common SNPs related with *XmnI* polymorphism is rs7482144 by showing mild β-thalassemia have *XmnI* +/+ genotype while *XmnI* -/- genotype in severe phenotypes. *XmnI*γ^G polymorphism reactivates the γ-globin gene in erythropoietic stress conditions in adults by the presence of T alleles. The *XmnI* polymorphic site is more persistent in individuals with HbF (Carrocini et al., 2011). In Malaysian Malays, heterozygous *XmnI* polymorphism (-/+) was prevalent in 63.3% of the population whereas homozygous *XmnI* polymorphic (+/+) was found in 8.2% of the population and absent in Chinese-Malaysians (Ching et al., 2006). Besides, the homozygous *XmnI* polymorphic (-/-) was 89.7% in Chinese-Malaysian patients (Elizabeth & Ann, 2010).

2.4 Hematological analysis 1

Hematological assessment is devoted to the study of the cellular processes of blood, such as the measures used to analyze the red blood cells, leukocytes and platelets. Hematological analysis, such as for MCV, MCH, MCH concentration (MCHC) and Hb, is useful in anemia screening (Lee et al., 2019).

2.4.1 Complete blood count (CBC)

The complete blood count (CBC) test, or known as full blood count (FBC), provides the particulars details on RBCs as a basic assessment for the diagnosis or/and monitoring of a patient's disease. There are several indices in FBC obtained using hematology analyzers, such as RBC count, Hb value, and red cell indices, which are MCV, MCH and MCHC.

The red cell indices were first proposed by Wintrobe in 1929, which are used to elucidate etiology of anemia. MCV represents the size of RBCs and is expressed in femtoliters, with a normal range of 87 ± 7 fl. MCH represents the amount of hemoglobin per red blood cell, and is expressed in picograms, with the normal range being 29 ± 2 pg per cell. Patients can be diagnosed as anemic when there is a reduction in Hb levels, and it can be classified into several types of anemia based on the MCV value, in which normal MCV is classified as normocytic, increased MCV as macrocytic, and decreased MCV as microcytic. However, the anemia sample can be further investigated following the CBC test if thalassemia is suspected.

2.4.2 High-performance liquid chromatography (HPLC)

High-performance liquid chromatography (HPLC) is a common separation and direct technique that has been used in hematological laboratories to determine HbA2, HbF and Hb variants. Cation exchange HPLC (CE-HPLC) is highly recommended to detect hemoglobinopathies and thalassemia with good resolution, reproducibility and quantification of normal and abnormal hemoglobin (Khera, Singh, Khuana, Gupta, & Dubey, 2015). The Hb fractions are recorded by the retention times, proportion of Hb (%) and peak shape. The CE-HPLC system (Bio-Rad Variant II HPLC system) can quantify Hb fractions, including the levels of HbA, HbA2, HbF and other Hb variants,

based on both peak resolution and peak integration quantifications. The key elements in CE-HPLC for the further investigation of β-thalassemia are HbA2 and HbF levels. Generally, high levels of HbA2 would be present in β-thalassemia with increased levels of HbF (Wajcman & Moradkhani, 2011).

2.5 Molecular diagnosis using polymerase chain reaction

Polymerase chain reaction (PCR) is a molecular method used to detect variation in DNA. There are several types of PCR that are extensively used such as reverse-transcription PCR (rt-PCR), real-time quantitative PCR (rtq-PCR), Gap-PCR, amplification refractory mutation system PCR (ARMS-PCR), and inverse shifting PCR (IS-PCR). The PCR forms that were chosen for the detection of thalassemia mutations in this study were ARMS-PCR and Gap-PCR.

2.5.1 Multiplex ARMS-PCR

The ARMS-PCR is a technique that was introduced by Newton et al. (1989). ARMS-PCR is a widely used technique involving allele-specific amplification and relies on the specificity of the 3' terminal. The wild and mutant ARMS primes are used to amplify the targeted DNA segment. The ARMS-PCR technique is simple, safe, quick, and consistent in providing accurate results of pre- and post-natal diagnoses in the detection of thalassemia using negligible quantity of DNA.

More than 200 mutations in β-thalassemia have been detected, including common and rare mutations in various ethnic groups (George, The, Rosli, Lai, & Tan, 2012). More than one mutations can be detected in a multiplex ARMS-PCR reaction set up. Table 2.5 shows two multiplex ARMS-PCR tests (ARMS-A, and ARMS-B) that

can be used to detect approximately eight known mutations of the β-globin gene cluster (Hanafi et al., 2014; Tan et al., 2004).

Table 2.4 List of β-globin thalassemia mutations detected using multiplex ARMS-PCR

Multiplex	Mutations	Amplicon size (bp)
ARMS-A	Cd 41/42 (–TTCT)	476
	IVS 1–5 (G>C)	319
	Cd 26 (G>A)	301
	Cd 17 (A>T)	275
ARMS-B	Cd 71/72 (+A)	569
	IVS1–1 (G>T)	315
	Cd 8/9 (+G)	250
	–28 (A>G)	145

bp= base pair, Cd= codon, IVS= intervening sequencing

2.5.2 Multiplex gap-PCR

Gap-PCR is used to detect β-globin gene cluster deletion. This study uses multiplex gap-PCR to detect common β-globin gene cluster deletions in Malaysia, such as δβ-thalassemia (Thai), HPFH, and Hb Lepore.

Table 2.5 List of β-globin gene deletions detected with gap-PCR

Deletions	Amplicon size (bp)
HPFH-6	974
Hb Lepore	1159
Internal control	304
Siriraj J ~ 118 kb deletion	620
Thai ($\delta\beta$) ~ 12.5 kb deletion	1447

bp= base pair, HPFH= hereditary persistence fetal hemoglobin

2.6 Single nucleotide polymorphism genotyping

There are several methods that are well-established in detecting SNPs. Allele discrimination technologies are widely used in the detection of SNPs (Kim & Misra, 2007). This technique consists of a few methods, such as primer extension,

hybridization, ligation and enzymatic cleavage, which allow the genotyping of genetic variants at SNP sites in specific templates (Kim & Misra, 2007). The detection of SNPs can be random or targeted by using advances technologies with acceptable genotyping method such as robust, safe, inexpensive, and automated (Kim & Misra, 2007).

One of the techniques used to determine SNPs is restriction fragment length polymorphism (RFLPs). RFLPs need restriction enzymes to recognize specific sequence on double strand of DNA and it is easy to form product size on SNPs genotype but not suitable for high-throughput analysis (Kim & Misra, 2007).

The Taqman assay genotyping could determine the known polymorphism. The Taqman uses two allele specific, a pair PCR primer flanking SNPs containing region (Kim & Misra, 2007). This assay is difficult to be performed since it needs high cost to use multiple probes for target specificity.

Another method can be used for SNPs genotyping is microarray. Microarray can scan thousand number of SNPs across genome (Kim & Misra, 2007). Microarray is a fast method and used for the high-throughput analysis with large number of sample (Walker, Flower, & Rigley, 2002).

Different methods exist have diverse of advantage and limitation to detect known and unknown polymorphism. The method selections will be conditioned on specific experimental design, type of samples, sensitivity and specificity.

CHAPTER 3

METHODOLOGY

3.1 Materials

The general consumables, commercial kits, chemical and reagents, instruments, apparatus and software used in this study are listed in Table 3.1 to Table 3.5 while flowchart of the study is shown in Figure 3.1.

Table 3.1 Consumables

Consumables	Manufacturer & country
Microcentrifuge tubes (1.5 mL)	Gene Era Biotech, China
0.2 mL 8-tube PCR strips with flat PCR tube 8-cap strips	Bio Rad Laboratories, US
Blood collection tube containing EDTA	BD Microtrainer, US
Falcon tube	Biologix, US
PCR tubes (0.2 mL)	Apical Scientific, Malaysia
Pipette tips (10,100,1000 μL)	Apical Scientific, Malaysia

Table 3.2 Commercial kits

Kits	Manufacturers & country
MN Blood Extraction	Macherey Nagel, Germany
HotStarTaq master mix	Qiagen, Netherlands
Taqpath ProAmp master mix	Thermo Fisher Scientific, US