MOLECULAR CHARACTERISATION OF ALPHA-THALASSAEMIA IN PATIENTS INVESTIGATED FOR HYPOCHROMIC MICROCYTIC INDICES IN HOSPITAL UNIVERSITI SAINS MALAYSIA

DIVASHINI A/P VIJIAN

UNIVERSITI SAINS MALAYSIA

2023

MOLECULAR CHARACTERISATION OF ALPHA-THALASSAEMIA IN PATIENTS INVESTIGATED FOR HYPOCHROMIC MICROCYTIC INDICES IN HOSPITAL UNIVERSITI SAINS MALAYSIA

by

DIVASHINI A/P VIJIAN

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

March 2023

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere appreciation and gratitude to my main supervisor, Dr Wan Suriana Wan Ab Rahman, for the continuous guidance, support, and motivation throughout this study. Her patience, immense knowledge, and enthusiasm were very helpful in completing this study. My sincere thanks to my co-supervisors Associate Professor Dr Thirumulu Ponnuraj Kannan and Dr Zefarina Zulkafli, for the constant support, insightful comments, and encouraging words that make this whole journey easier. I'm very grateful to Universiti Sains Malaysia for providing a Research University Individual Grant (1001/PPSG/8012366) for this project. This is a collaborative study between Universiti Sains Malaysia, USM, and the Institute for Medical Research, IMR. At this opportunity, I would like to express my thanks and gratitude to the Unit of Haematology, IMR, for the opportunity and willingness to be a part of this study. Not forgetting my field supervisor, Dr Ezalia Esa, from IMR for her constant support and help in this study. A special mention goes to Mr Syazuwan Hassan, Madam Faidatul Syazlin, and all the staff in the Unit of Haematology, IMR, for their knowledge and skilful training. I would like to take this opportunity to thank all the staffs in Department of Haematology, Hospital Universiti Sains Malaysia, special mention to Madam Norazlina, for the support and guidance. I would like to acknowledge all the staff in the School of Dental Sciences USM for their help and assistance in finishing my laboratory work involved in this study. I am extremely grateful to my parents for their love, prayers, and sacrifices for educating me for my future. I would like to say thanks to my friends for their constant encouragement. Finally, my thanks go to all the people who have supported me in completing the research work, directly or indirectly.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVATIONS

RBCs	Red Blood Cells
Hb	Haemoglobin
Hb A	Haemoglobin Adult
Hb F	Haemoglobin Fetal
Hb A_2	Haemoglobin A_2
Hb CS	Haemoglobin Constant Spring
Hb S	Haemoglobin S
SNP	Single nucleotide polymorphism
HBA	Alpha globin gene
kbp	Kilo base pairs
IDA	Iron deficiency anaemia
FBC	Full blood count
FBP	Full blood picture
MCV	Mean corpuscular volume
MCH	Mean corpuscular haemoglobin
MCHC	Mean corpuscular haemoglobin concentration
RDW	Red cell distribution width
Hct	Haematocrit
HPLC	High performance liquid chromatography
CE	Capillary electrophoresis
PCR	Polymerase chain reaction
MARMS	Multiplex arms refractory mutation system
HUSM	Hospital Universiti Sains Malaysia
IMR	Institute for Medical Research
NGS	Next generation sequencing

LAMP	Loop mediated isothermal amplification
MLPA	Multiplex ligation dependent probe amplification
DNA	Deoxyribonucleic acid
Min	Minutes
Sec	Seconds
UV	Ultraviolet ray
TBE	Tris-borate-EDTA
kbp	Kilo base pairs
bp	Base pairs
RNA	Ribonucleic acid
F	Forward
R	Reverse
RR	Reverse complement
WT	Wild type
WBCs	White blood cells

LIST OF UNITS AND SYMBOLS

α	Alpha
β	Beta
δ	Delta
γ	Gamma
ζ	Embryonic gene
g/dl	Gram/decilitre
%	Percentage
fL	Femtolitre
pg	picograms
′	prime
ml	Millilitre
µl	Microliter
°C	Degree Celsius
µM	Micromolar
ng	Nanogram
ng/µl	Nanogram per microliter
<	Less than
g	gram
x g	Gravitational force
=	Equals

LIST OF APPENDICES

PERINCIAN MOLEKULAR ALFA-TALASEMIA BAGI PESAKIT YANG MEMPUNYAI INDISIS HIPOKROMIK MIKROSITIK DI HOSPITAL UNIVERSITI SAINS MALAYSIA

ABSTRAK

Alfa (α) talasemia merupakan penyakit yang disebabkan oleh gangguan genetik yang telah menjejaskan 5% daripada populasi dunia. Mutasi penghapusan atau bukan penghapusan melibatkan satu atau kedua-dua *HBA1* dan *HBA2* pada kromosom 16 yang menyebabkan pengurangan dalam penghasilan rantai globin α, iaitukomponen hemoglobin (Hb) yang diperlukan untuk pembentukan sel darah merah. Oleh itu, pengurangan Hb akan membawa kepada anemia. Pelbagai jenis mutasi dalam α-talasemia telah ditemui yang mana menghasilkan spektrum manifestasi klinikal yang meluas, daripada tiada gejala hingga membawa maut. Kajian ini berfokuskan kepada parameter sel darah merah dan mutasi molekul α-talasemia untuk menentukan kelaziman dan menerangkan parameter hematologi berdasarkan mutasi yang dikesan. Kajian keratan rentas yang melibatkan 136 pesakit yang disyaki α-talasemia telah dikumpulkan. DNA yang diekstrak daripada sampel darah telah dilakukan ujian "multiplex GAP-polymerase chain reaction", "multiplex amplification refractory mutation system" (MARMS-PCR) dan "duplex-PCR" untuk mengesan mutasi α penghapusan dan bukan penghapusan yang sering dilaporkan. "Multiplex ligation dependent probe amplification" dan penjujukan Sanger dilakukan untuk mengesan mutasi yang jarang berlaku pada pesakit yang tidak membawa sebarang mutasi biasa. Kelaziman α-talasemia dalam kajian ini ialah 47.1%. Tiga puluh sembilan dan 1.4 peratus pesakit masing-masing didapati mempunyai mutasi α-talasemia heterozigot dan homozigot, dengan 6.6 peratus adalah heterozigot majmuk. Dalam kalangan

pesakit, genotip berikut ditemui: -$\alpha^{3.7}$/αα (15.4 %), -$\alpha^{4.2}$/αα (3.7 %), --^{SEA}/αα (7.4 %), α^{CS}α/αα (10.3 %), α^{Adana}α/αα (0.7 %), $\alpha^{Quong\ Szeα}$/αα (1.5 %), -$\alpha^{3.7}$/-$\alpha^{3.7}$ (0.7 %), α^{CS}α/α^{CS}α (0.7 %), -$\alpha^{4.2}$/α^{CS}α (0.7 %), –^{SEA}/α^{CS}α (1.5 %), –^{SEA}/$\alpha^{Quong\ Sze}$α (0.7 %), -$\alpha^{3.7}$/α^{Adana}α (0.7 %), --^{SEA}/-$\alpha^{3.7}$ (2.2 %) dan α^{CS}α/α^{Adana}α (0.7 %). Berdasarkan mutasi yang dijumpai analisis statistik telah dilakukan dan beberapa parameter seperti Hb (p=0.020), min isi padu korpuskel (p=0.008), min hemoglobin korpuskel (p=0.018), sel darah merah (p=0.029) dan hematokrit (p=0.049) merupakan petunjuk kepada perubahan ketara dalam kalangan pesakit yang mengalami mutasi penghapusan, tetapi tiada perbezaan yang ketara antara pesakit dengan mutasi bukan penghapusan. Pesakit dengan mutasi penghapusan gen tunggal dan bukan penghapusan hanya menunjukkan perbezaan yang ketara untuk HbA_2 (p=0.028). Pelbagai parameter hematologi diperhatikan dalam kalangan pesakit termasuk mereka yang mempunyai genotip yang sama. Oleh itu, berdasarkan parameter hematologi sahaja, ia tidak mencukupi untuk menerangkan mutasi khusus α-talasemia.

MOLECULAR CHARACTERISATION OF ALPHA-THALASSAEMIA IN PATIENTS INVESTIGATED FOR HYPOCHROMIC MICROCYTIC INDICES IN HOSPITAL UNIVERSITI SAINS MALAYSIA

ABSTRACT

Alpha (α)-thalassaemia is a common genetic disorder that affects 5 % of the worldwide population. Deletional or non-deletional mutations of one or both *HBA1* and *HBA2* on chromosome 16 cause a reduction/abnormal in the production of α globin chains, a component of haemoglobin (Hb) which are required for the formation of red blood cells (RBC). Thus, reduced Hb leads to anaemia. Many genetic mutations in α-thalassaemia have been discovered, which produced wide spectrum of clinical manifestation, ranging from asymptomatic to lethal. This study focused on the characterisation of RBC parameters and molecular based on α-thalassaemia mutations to determine the prevalence and to describe the RBC parameters based on the mutations detected. A cross-sectional study involving 136 suspected α-thalassaemia patients was collected. The DNA extracted from blood samples was subjected to the multiplex GAP-polymerase chain reaction, multiplex amplification refractory mutation system polymerase chain reaction (MARMS-PCR), and duplex-polymerase chain reaction to detect common deletional and non-deletional α mutations, respectively. Multiplex ligation dependent probe amplification and Sanger sequencing were performed to detect rare mutations in patients that do not carry any common mutations. The prevalence of α-thalassaemia in this study was 47.1 %. Thirty-nine and 1.4 percent of patients were found to have heterozygous and homozygous α-thalassaemia mutations, respectively, with 6.6 percent being compound heterozygous. Among the patients, the following genotypes were found: $-\alpha^{3.7}/\alpha\alpha$ (15.4 %), $-\alpha^{4.2}/\alpha\alpha$

(3.7 %), --^{SEA}/αα (7.4 %), α^{CS}α/αα (10.3 %), α^{Adana}α/αα (0.7 %), α$^{Quong\ Szeα}$/αα (1.5 %), -α$^{3.7}$/-α$^{3.7}$ (0.7 %), α^{CS}α/α^{CS}α (0.7 %), -α$^{4.2}$/α^{CS}α (0.7 %), –^{SEA}/α^{CS}α (1.5 %), –^{SEA}/α$^{Quong\ Sze}$α (0.7 %), -α$^{3.7}$/α^{Adana}α (0.7 %), --^{SEA}/-α$^{3.7}$ (2.2 %) and α^{CS}α/α^{Adana}α (0.7 %). Statistical analysis of RBC parameters according to the mutations was performed. A few indicators, such as Hb (p=0.020), mean corpuscular volume (p=0.008), mean corpuscular haemoglobin (p=0.018), RBC (p=0.029) and haematocrit (p=0.049), showed significant changes among patients with deletional mutations, but there were no significant differences between patients with non-deletional mutations. patients with single gene deletional and non-deletional mutations shows only significant differences for HbA_2 (p=0.028). A wide range of RBC parameters were observed among the patients, including those with the same genotypes. Thus, based on RBC parameters alone, they are not sufficient to describe the specific mutations of α-thalassaemia.

CHAPTER 1

INTRODUCTION

1.1 Red blood cells

Haemopoiesis starts with pluripotent stem cells in the bone marrow which undergo differentiation and proliferation into progenitor committed cells and finally mature to red blood cells (RBCs), white blood cells (WBCs) and platelets (Rifkind and Nagababu, 2013). The shape of healthy RBCs is biconcave disc shaped when not subjected to external stress. The primary function of RBCs is to transport oxygen from the lung to the rest of the body and carbon dioxide from the organs to the lung (Diez-Silva *et al.*, 2010). The life span of healthy RBCs is 120 days (Rifkind and Nagababu, 2013). A flexible membrane of RBCs with a high surface to volume ratio will ease the reversible elastic deformation of the RBC as it passes through the small capillaries repeatedly in microcirculation (Diez-Silva *et al.*, 2010). The size of RBCs is approximately 7.5 to 8.7 μm in diameter, and 1.7 to 2.2 μm in thickness. Furthermore, RBCs cytosol contain haemoglobin (Hb) molecules which are essential for gas transport within the circulation (Fung, 2013). The membrane is made up of a phospholipid bilayer and two-dimensional network of spectrin molecules underlying it. The elasticity of the membrane and biconcave shape of RBCs is contributed by the phospholipid bilayer and spectrin network (Tse and Lux, 1999). However, changes at genetic and molecular levels may alter the RBCs deformability and lead to morphological changes in cell shape (Diez-Silva *et al.*, 2010).

1.2 Haemoglobin

Hb is a component of RBCs, comprised of heme and globin (Khoshouei *et al.*, 2017). Globin is a tetramer, small globular metalloproteins which contain 150 amino

acids and combine with a heme group to produce haemoglobin (Ghosh, 2008). As shown in figure 1.1 each heme group comprises an atom of ferrous iron, Fe^{2+} and surrounded by a porphyrin ring, four nitrogen-containing pyrroles. Therefore, each adult Hb molecules (Hb A) carries four heme groups with a polypeptide globin chain of 141 (α) and 146 (β) amino acid residues (Marengo-Rowe, 2006). Oxygen molecule will bind to an iron atom thus, each Hb will carry four oxygen molecules (Kanias and Acker, 2010). Normal Hb in men is 13.5-18.0 g/dl while 11.5-16.0 g/dl in women (Thomas and Lumb, 2012).

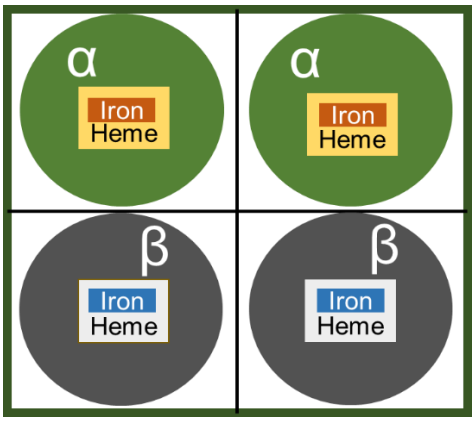

Figure 1.1 Illustration of Hb structure

Hb A consists of pairs of α and β globin chains ($\alpha_2\beta_2$) while HbA_2 consists of two α and two delta (δ) chain ($\alpha_2\delta_2$), and foetal Hb (Hb F) consists of pairs of α and gamma (γ) globin chains ($\alpha_2\gamma_2$). More than 95 % of adult Hb is Hb A while 2.2-3.5 % is HbA_2 and less than 1 % is Hb F. Hb F is the main Hb during foetal life and is found around 50-95 % at birth. However, the level of Hb F will decline 6 months after birth as the production of Hb A increases. Hb F has a greater oxygen affinity compared to Hb A to transport oxygen between the maternal and foetal circulations in the placenta (Thomas and Lumb, 2012).

1.3 Globin gene

The globin gene clusters encode the protein subunits of Hb and have a considerable amount of variability in their base composition, thus a single nucleotide change produces or removes a cutting site for a restriction enzyme, which is called single nucleotide polymorphism (SNP). These changes forms a series of pattern which occur among different populations of the world at varying frequencies (Brenner and Miller, 2014). Globin genes are organized into two clusters and located on a different chromosome. The α-like and β-like genes are located on chromosome 16 and chromosome 11 respectively (Thein, 2013). The arrangement of the genes is found in the order as shown in Figure 1.2. Alpha like gene cluster is located close to the telomere of chromosome 16 (16p13.3) that includes an embryonic gene, zeta (ζ) and two foetal/adult genes arranged along the chromosome surrounded by widely expressed genes (Higgs, 2013). The genes for the α chain and γ chain are duplicated while the gene for β chain is encoded by single gene locus. The two duplicated α chain genes are α2 and α1 (*HBA1* and *HBA2*) and the product of the pair α genes is identical. Two duplicated γ chain genes are Gγ and Aγ which are located on chromosome 11. The G and A amino acids refer to glycine and alanine respectively that differentiate the two types of genes (Maloy and Hughes, 2013). The genetic defect that occurs in these globin genes will lead to either reduction or absence of α chain or β chain production (Higgs, 2013). A balanced expression of these genes is required for the efficient production of Hb. Hence, a genetic defect in the globin genes leads to the pathological phenotypes of inherited anaemia called thalassaemia (Hardison, 2012).

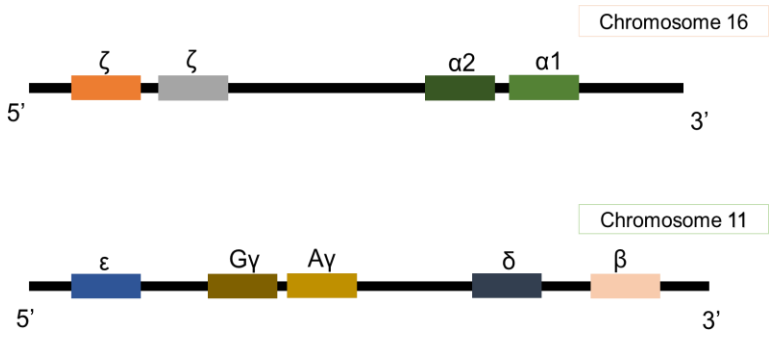

Figure 1.2 α and β globin gene cluster on chromosome 16 and chromosome 11

1.4 Thalassaemia

Thalassaemia is an inherited autosomal recessive disorder due to reduction or absence in the production of Hb subunits (Tan *et al.*, 2010). Every year more than 300,000 children are born with inherited Hb disorders, especially thalassaemia (Weatherall, 2012). Approximately 80 % of these cases occur in the low or middle income countries of the tropical belt stretching from sub-Saharan Africa, through the Mediterranean region and the Middle East, to South and Southeast Asia (Weatherall, 2012). Inherited Hb related disorders can be subdivided into two groups, the first group is composed of structural Hb variants, commonly reported are Hb S, C and E, whereas the second group is quantitative Hb disorder such as thalassaemia, which includes α-thalassaemia and β-thalassaemia (Weatherall *et al.*, 2006). Thalassaemia is inherited in a Mendelian-recessive manner where the carrier or heterozygous parents, pass on one copy of a gene for Hb to their children. Hence, structural Hb variants and thalassaemia often occur at high frequencies in the same population (Weatherall, 2012). The disorder results in large numbers of RBCs being destroyed, which leads to anaemia. Alpha-thalassaemia and β-thalassaemia are caused by reduced or absence in the synthesis of α globin chains and β globin chains respectively (Muncie Jr and Campbell, 2009).

1.4.1 α-thalassaemia

Alpha-thalassaemia is caused by deletional or non-deletional mutation of α globin gene. Deletion of one or both α genes (*HBA1* and *HBA2*) from the chromosome 16 occurs in 95 % of cases which result in the reduction of α globin chain production (Harteveld and Higgs, 2010). Inheritance of α genes is complex as each individual inherits two α alleles from each parent. Therefore, there are four functional α globin genes in healthy individuals. Hence, four functional α globin genes (αα/αα) may be decreased by 1, 2, 3 or all 4 copies of the genes due to deletion or point mutation. This abnormality explains the variation in clinical diagnosis and disease severity (Harteveld and Higgs, 2010). The milder form of α-thalassaemia results with reduction in α globin chain production is termed as α^+ or α-thalassaemia 2. While, mutation in two α globin chains with the absence of globin chains synthesis is termed α^0 or α-thalassaemia 1 (Munkongdee *et al.*, 2020). The α-thalassaemia 2 is the most common type of α-thalassaemia throughout the world with an estimated frequency between 30-50 % (Rosnah *et al.*, 2012).

The α-thalassaemia carriers are usually asymptomatic and may be identified after routine haematological analyses or during antenatal screening. They usually have mild anaemia or normal Hb. Most of thalassaemia carriers have single α globin gene defect with borderline or normal Hb level. A drop in Hb level among the carriers can be due to blood loss, poor nutrition, infection or other underlying disease (Farashi and Harteveld, 2018). The presence of only one functional α globin gene or inactivation of three α globin genes due to compound heterozygosity for two different mutations or homozygous leads to Hb H disease. In this condition, expression of the α globin chain is reduced to less than 30 % of the normal levels (Farashi and Harteveld, 2018). In adults, when the production of α chain is decreased, there will be excess β chains which

form β_4 tetramer, known as Hb H. Hb H has high oxygen affinity but its instability leads to the formation of inclusion bodies in RBCs and a variable degree of haemolytic anaemia (Piel and Weatherall, 2014). Patients may have splenomegaly, jaundice, and other complications such as gallstones, leg ulcers or infections (Farashi and Harteveld, 2018). The deletion of all four functional α globin genes (--/--) resulting in absence of α globin chain production leads to Hb Barts hydrops foetalis syndrome. It is the most severe and fatal clinical phenotype of α-thalassaemia (Al-Allawi *et al.,* 2010). Since there is no α globin chain production, Hb F is not able to form but instead, Hb Barts (γ_4) form from γ chain tetramer in foetus (Farashi and Harteveld, 2018). This condition leads to intra-uterine hypoxia and intrauterine death (Al-Allawi *et al.*, 2010). Hb H disease and Hb Barts hydrops foetalis are predominantly reported in Southeast Asia, the Mediterranean area and the Middle East (Farashi and Harteveld, 2018).

Specifically, 5 % of world's population is affected with α-thalassaemia (Vichinsky, 2010). Single α gene deletion has been reported in broad tropical regions from sub-Saharan Africa through the Mediterranean regions and the Middle East to the Indian subcontinent and the whole of east and Southeast Asia at the frequency of 10 %-25 % (Williams and Weatherall, 2012). While in few localized areas such as North India and Papua New Guinea, up to 80 % of the populations carry this mild form of α-thalassaemia. Furthermore, severe forms of α-thalassaemia occur due to the loss of both α chains, and this condition was observed among southeast Asia and in some of the Mediterranean island populations (Williams and Weatherall, 2012). The gene frequency of α-thalassaemia in northern and southern Thailand was reported to be 30 % and 16 % while in Brunei was 4.3 %, where 65 % of the population was Malay, and 20 % was Chinese (Winichagoon and Fucharoen, 2000). The prevalence of α-thalassaemia trait in Malaysia was 15.8 % (Azma *et al.*, 2014).

1.4.2 Deletional α-thalassaemia

The mutation in α globin genes could be either deletion or non-deletional mutations. However, deletion of α globin genes is commonly reported compared to non-deletional or point mutations which lead to α-thalassaemia. The α globin genes are duplicated and embedded into two highly homologous 4 kilo base pair (kbp) units. The region is divided into X, Y and Z homology boxes (Higgs, 2013). The rearrangement in this region is single α globin gene deletion due to unequal homologous recombination during meiosis between the mispaired X and Z boxes giving rise to the deletion (Farashi and Harteveld, 2018). A chromosome with 3.7 kbp deletion with one α gene deletion is due to reciprocal recombination between highly homologous regions, Z boxes. Next, 4.2 kbp deletion occurred due to recombination between mispaired homologous X boxes (Farashi and Harteveld, 2018). The mechanism involved in the rearrangements of the α globin gene resulting in the deletion of double α globin gene is not always clear. Furthermore, breakpoints involving double gene deletions do not always show homology such as in the $\alpha^{3.7}$ and $\alpha^{4.2}$ deletions. Thus, the mechanisms involved in this condition is more complex (Farashi and Harteveld, 2018). A study identified more than 40 deletion mutations in α-thalassaemia (Karakas *et al.*, 2015). Hence, the most common deletion type involving a single gene deletion are $\alpha^{3.7}$ and $\alpha^{4.2}$. Apart from that, common double gene deletion reported in the world are $\alpha^{20.5}$, α^{SEA}, α^{MED}, α^{THAI}, and α^{FIL} (Karakas *et al.*, 2015). The SEA deletion is about 20.5 kbp long and involves both functional α globin genes while leaving the ζ2 globin gene alone. THAI deletion is approximately 33.45 kbp, including both functional *HBA* genes and the ζ2 globin gene (Pornprasert *et al.*, 2011). In FIL deletions both α and ζ2 globin gene deleted (Eng *et al.*, 2000).

1.4.3 Non-deletional α-thalassaemia

The non-deletional mutations are rare and mostly occur in α2 gene. The expression of α2 gene is higher compared to α1 gene with a ratio of 3:1 (Fucharoen and Winichagoon, 2011). Non-deletional mutations cause abnormalities in the structure and this lead to unstable Hb variants (Karakas *et al.*, 2015). The mutations involve single nucleotide substitutions, deletions or insertions in the region required for α globin chain synthesis. Due to the mutation, several molecular mechanisms have been described including abnormalities in the RNA splicing and mRNA translation initiation, frameshift and nonsense mutations, in-frame deletions, and chain termination mutations. Most of the mutations result in α+ thalassaemia (Farashi and Harteveld, 2018). The common non-deletional type of α-thalassaemia reported in Southeast Asia is described in Table 1.1. Hb Constant Spring (Hb CS), Hb Quong Sze and Hb Adana are commonly reported mutations whereas codon 30 and codon 35 mutation is rarely identified (Eng *et al.*, 2001).

Table 1.1 Common and severe non-deletion of α globin gene mutation found among Southeast Asian population

Non-deletion mutation	Nucleotides	Reference
Initiation codon	(ATG→A-G)	(Eng *et al.*, 2001)
Codon 30	(ΔGAG)	
Codon 35	(TCC→ CCC)	
Codon 59 (Hb Adana)	(GGC→ GAC)	
Codon 125 (Hb Quong Sze)	(CTG→ CCG)	
Termination codon (Hb Constant Spring)	(TAA→ CAA)	

A mutation was identified at the initiation codon of α2 gene leading to the conversion of ATG→A-G. This results in the reduction of mRNA expression and instability of mRNA thus producing unstable α globin chains (Lei *et al.*, 2019). Next, the mutation in codon 30 of α2 gene results in the production of delta (Δ) GAG instead of glutamic acid. This mutation will result in the production of unstable α globin chain which leads to RBCs destruction (Ma *et al.*, 2001). Hb CS is due to a stop codon

mutation in *HBA2* (Komvilaisak *et al.*, 2018) and it is the most common non-deletion mutation found in Southeast Asia (Jomoui *et al.*, 2015). This mutation produces elongated α globin chain with additional 31 amino acids. The level of Hb CS is significantly low and difficult to be detected in peripheral blood. This is due to the disruption of the untranslated region leading to mRNA instability (Jomoui *et al.*, 2015).

Hb Quong Sze results from codon 125 mutation in α2 globin gene which leads to the production of proline amino acid instead of leucine. This Hb variant is commonly reported in Southern China and more prevalent among Malaysian Chinese compared to Malays (Ahmad *et al.*, 2013). This is a highly unstable Hb variant (Wee *et al.*, 2009). Mutation in α globin gene produces unstable α globin chain and is rapidly degraded, which may result in the clinical phenotype of α-thalassaemia (Karakas *et al.*, 2015). Thus, the unstable proteins may precipitate in the RBC, causing insoluble and damage to the RBCs membrane (Farashi and Harteveld, 2018). Hb Adana is high in Indonesia which is reported at a frequency of 16 %. It is due to point mutation in codon 59 affecting either *HBA1* or *HBA2*. The mutation results in substitution of small non-charged glycine to large charge aspartic acid. This leads to α globin instability and results in haemolysis and precipitation on RBCs membrane (Singh *et al.*, 2018). Hence, a study suggests that the Hb Adana identified among Malaysians might be due to gene flow from Indonesia as a result of regular travel of the citizens (Ahmad *et al.*, 2013).

1.5 Diagnosis of α-thalassaemia

The pathologic defects on globin chain are characterized by quantitative or qualitative abnormality. The quantitative defect of the globin chains includes α-

thalassaemia and β-thalassaemia whereas, qualitative defect or haemoglobinopathy are structural Hb variants such as sickle cell anaemia, Hb S. These types of globin defects' interaction results in a wide variety of clinical presentation and diseases (Weatherall and Clegg, 2008). Clinically, α-thalassaemia is classified into thalassaemia minor, intermedia and major. Patient with thalassaemia minor (or thalassaemia trait) is usually asymptomatic or presents with mild anaemia. Individual with thalassaemia intermedia described as highly diverse group of patients presents with various clinical severities, from mild to severe anaemia. Some of these patients may not require blood transfusion, while others may need frequent blood transfusion. Patients with thalassaemia major presents with severe anaemia early in life and requires frequent and lifelong blood transfusion and iron chelation (Viprakasit and Ekwattanakit, 2018). In α-thalassaemia, reduced protein synthesis will cause accumulation of excess polypeptide (β globin chain) encoded by the unaffected gene as the expression of α globin chain is reduced. Thus, imbalance in α and β globin chains cause abnormal maturation of RBCs, resulting in microcytosis that can be observed as a laboratory characteristic (Sabath, 2017).

Clinical phenotype of α-thalassaemia depends on the number of genes affected as normal individual carry four functional α genes. Hence, inactivation or deletion of one α gene results in silent carrier of α-thalassaemia showing insignificant haematological findings. Two affected α globin genes result in microcytic-hypochromic anaemia. Hb H disease is found when three α globin genes are deleted or mutated resulting in moderate anaemia and marked microcytosis. Hb Bart or hydrops foetalis is the most severe form of α thalassaemia in which all four α globin genes are affected and often cause death *in utero* (Gilad *et al.*, 2017). Infants with Hb Bart's hydrops foetalis have variable amounts of non-functional Hb H (β_4) and Hb

Bart's (γ_4), and variable amounts of functional Hb Portland ($\zeta_2\gamma_2$) (Farashi and Harteveld, 2018). They will not be able to produce α globin chain, therefore there will be excess in Hb Bart which results in severe anaemia (Hb level 3-8 g/dl), marked hepatosplenomegaly, and cardiac failure. Foetus affected with this are either stillborn or die soon after birth and usually not compatible with postnatal life. Study has reported maternal complications during pregnancy which includes preeclampsia (hypertension and fluid retention with or without proteinuria), poly-oligohydramnios (increased or reduced accumulation of amniotic fluid, respectively) haemorrhage, anaemia, and sepsis (Weatherall *et al.*, 1970). Due to severe clinical presentation, early termination of these pregnancies is recommended. Intensive and lifelong blood transfusion and supportive care are necessary but could not prevent irreversible abnormalities (Farashi and Harteveld, 2018).

Clinical presentation of individuals with α-thalassaemia may vary as it has high variability in genetic defect. Generally, the level of Hb in α-thalassaemia varies from normal to severe anaemia (Hb 7.5-15.5 g/dl), meanwhile there is reduction of mean corpuscular volume (MCV < 79 fL), mean corpuscular haemoglobin (MCH < 27 pg), and normal to slight decrease in HbA_2 level depending on the number of functional α globin genes (Farashi and Harteveld, 2018). The level of MCV and MCH between patients with one defected α globin gene and patient having two functional α globin genes is significantly different. Thus, the level of MCV and MCH is lower in patients with more defected α globin genes. Therefore, MCV and MCH can be used for the prediction of the genotype (Akhavan-Niaki *et al.*, 2012). Haematocrit (Hct) measures the proportion of RBCs in the blood. Theoretically, the level of Hct decrease with a greater number of affected α genes (Fasola *et al.*, 2022). Next, the red cell distribution width (RDW) is a routine full blood count parameter that measures the width of RBC

size distribution. The RDW level elevated with the presence of anisocytosis of red cells in the blood. The increase in the level of RDW is reflecting the severity of anisocytosis. This parameter able to differentiate between iron deficiency anaemia (IDA) and thalassaemia trait (Matos *et al.*, 2015). Higher value of RDW was reported among IDA compared to thalassaemia patients (Nusrat *et al.*, 2020). The normal value of RDW-coefficient variation (CV) is less than 14.5 (Nusrat *et al.*, 2020).

1.6 Routine screening of α-thalassaemia

Screening of α-thalassaemia is important to inform couples at high risk, especially severe forms of the disease occurring in high frequency areas. Apart from that, it also helps in preventing severe maternal complications in the case of Hb Bart's and able to provide accurate diagnosis in cases of IDA and α-thalassaemia coinherited with Hb S or β-thalassaemia (Piel and Weatherall, 2014). Diagnosis of thalassaemia requires combination of laboratory tests including full blood count (FBC), full blood picture (FBP) and also quantification of Hb by high performance liquid chromatography (HPLC) and capillary/gel electrophoresis (Munkongdee *et al.*, 2020). Most of α-thalassaemia traits are asymptomatic and identified only by chance after routine haematological analyses.

1.6.1 Full blood count and full blood picture

Full blood count is done by an automated blood cell analyser where a blood sample is aspirated and separated into different fluidic streams which contain different solutions of buffer to obtain the specific purpose of analysis (Viprakasit and Ekwattanakit, 2018). RBC indices such as MCV, MCH, and mean corpuscular haemoglobin concentration (MCHC) levels are calculated from Hb, haematocrit, and RBC count. In thalassaemia screening, MCV level less than 80 fL and MCH less than

27 pg are used as cut-off levels for a positive screening result (Cappellini *et al.*, 2014). Thus, MCV and MCH values might be also low due to IDA and anaemia of chronic disease, as found in thalassaemia trait (Viprakasit and Ekwattanakit, 2018).

In thalassaemia patients, RBCs shows microcytic, hypochromic, apparent anisocytosis and poikilocytosis. Hypochromic RBCs is indicated by increase in diameter of central pallor, whereas variation in shape and size is due to presence of schistocytes, microspherocytes, target cells and nucleated RBCs. Microcytic anaemia among thalassaemia patients is due to impaired globin chain synthesis, which leads to decreased Hb synthesis (Urrechaga *et al.*, 2011). Among thalassaemia carrier individuals, the RBCs morphology changes are less severe compared to patient with thalassaemia major and intermedia. However, interpretation of blood smear based on RBCs morphology solely is not possible to define thalassaemia disease as those interpretations can be detected in IDA as well (Viprakasit and Ekwattanakit, 2018). Next, patients with Hb H disease will show a significant difference in blood count with low MCV (MCV range between 55-60 fL), chronic haemolysis and splenomegaly, while, patients may have normal life span without significant complications (Sabath, 2017). Blood smear stained with brilliant cresol blue shows presence of inclusion bodies in the RBCs which indicates presence of Hb H (Khera *et al.*, 2015). These red cells are unstable, destroyed prematurely in the spleen resulting in moderate to severe haemolysis (Galanello and Cao, 2011). Whereas, foetus with Hb Barts will show anisopoikilocytosis with hypochromic RBCs, and presence of nucleated RBCs in the blood smear (Aiempanakit and Apinantriyo, 2018). Hence, HPLC and capillary electrophoresis (CE) or gel electrophoresis is used to distinguish between thalassaemia disease and trait by quantification of Hb level (Munkongdee *et al.*, 2020).

1.6.2 High performance liquid chromatography

High performance liquid chromatography is one of the techniques which separates a net positive charged molecules (normal and variant Hbs) by adsorption onto a negatively charged stationary phase in a chromatography column and elution by a mobile phase (Bain *et al.*, 2011). The sensitivity and precise method for quantitative analyses of Hb components in RBCs produce excellent resolution and reproducibility are the main advantage of this technique (Khera *et al.*, 2015). HPLC is able to measure HbA_2, Hb F and Hb variants which helps in the identification of normal and abnormal levels of Hb in a blood sample (Old *et al.*, 2012). HbA_2 level plays an important role in β-thalassaemia screening program wherein if the value is more than 4 %, the individual is presumed as β-thalassaemia carrier. However, there were significant reports from previous study with HbA_2 between 3.0 and 3.9 % (borderline) and confirmed as β-thalassaemia carriers by molecular analysis (Rosnah *et al.*, 2017). Conditions that result in a borderline HbA_2 level, on the other hand, should be investigated further for silent mutations, α-thalassaemia, and co-existing nutritional deficiencies (Khera *et al.*, 2015). Hb H is characterised by a reduced rate of α chain synthesis. As a result, excess β chains result in the formation of Hb H and a decrease in HbA_2 levels. Hence, Hb H and Hb Bart's are important in the diagnosis of the disease (Viprakasit and Ekwattanakit, 2018) and the main limitation of the HPLC system is to quantify the level of Hb H and Hb Bart's (Munkongdee *et al.*, 2020). Despite the fact that HPLC is automated and capable of distinguishing distinct Hb with high precision, it has limitations in detecting and quantifying the fraction of Hb H and Hb Bart's, as well as having low sensitivity in detecting Hb CS, α globin variation (Viprakasit and Ekwattanakit, 2018).

1.6.3 Capillary electrophoresis

Capillary electrophoresis method is highly used in areas with high prevalence of Hb E and α-thalassaemia. This is a diagnostic tool used to separate Hb fractions and to quantify the level of each fraction (Old *et al.*, 2012). The technique can separate and quantify HbA_2, Hb F, Hb E, Hb H, and Hb Bart's where these parameters are mainly required for the diagnosis of α and β-thalassaemia as well as hemoglobinopathies (Sangkitporn *et al.*, 2011). Briefly, in CE techniques, charged molecules are separated in an alkaline buffer by their electrophoretic mobility (You-Qiong *et al.*, 2016). Advantages of CE include the turn-around time, high throughput, small sample volume requirement and reasonable cost. Apart from that, CE system can quantify the level of Hb H and Hb Bart's. In addition, CE system can separate Hb E and HbA_2 which can distinguish between homozygous Hb E and Hb E/β thalassaemia, and this separation is limited in HPLC system (Sangkitporn *et al.*, 2011).

On the other hand, level of Hb H, Hb Bart's and Hb CS is found to be decreased in blood samples stored for a long time especially at high temperature as these Hbs are unstable (Munkongdee *et al.*, 2020). The lifespan of Hb H is 28 to 37 days (Harewood and Azevedo, 2017), but some studies stated that Hb H may has a shorter life span which is 12 to 19 days (Harewood and Azevedo, 2017; Schrier, 2002). Factors that lead to unstable Hb H are abnormal RBC membrane with increased inclusion bodies and increased rigidity (Harewood and Azevedo, 2017). This limitation can be overcome with the usage of freshly collected blood sample. However, molecular analysis of α globin genes is essential to fully understand the clinical manifestations as well as for a definitive diagnosis (Traeger-Synodinos *et al.*, 2015).

1.7 Molecular diagnosis

In the last decade, molecular methods in disease diagnosis have undergone rapid growth and development. The diagnosis can identify genomic variation including detection, sub-classification, prognosis, and even in monitoring the response to therapy (Hsiao, 2019). The rapid development of these molecular techniques is due to the increased demand for the results involving genetic and genomic variations. Hence, molecular diagnosis serves as an important role in identifying individuals with α-thalassaemia. Furthermore, the molecular findings are the most useful and informative tools for accurate diagnosis (Sabath, 2017).

1.7.1 Gap-polymerase chain reaction

Gap-polymerase chain reaction (Gap-PCR) is a common method used to detect α-thalassaemia due to the deletion in the α globin genes (Brancaleoni *et al.*, 2016). Hence, this method provides a quick diagnostic test involving both α^0 and α^+-thalassaemia common deletion mutations (Old *et al.*, 2012) as listed in Table 1.2. Gap-PCR method works based on the inability of the primers to generate PCR amplification product unless a deletion joined the flanking sequences together. However, the amplication will only occur if deletion is present and the product will be examined by electrophoresis (Lam *et al.*, 2013). Previously, Southern blotting was a method commonly used to detect gene deletion. However, this method is time consuming, not suitable for large scale screening and labour intensive (Ou-Yang *et al.*, 2004). Multiplex Gap-PCR is used for the detection of common deletion α-thalassaemia as this technique is a rapid, simple and a non-radioactive method (Old *et al.*, 2012). The limitation of this method is that the deletion endpoints must be known to design the primers.

Apart from that, next generation sequencing (NGS), loop-mediated isothermal amplification (LAMP) and multiplex ligation dependent probe amplification (MLPA) also can detect deletion α-thalassaemia. MLPA detects both common and uncommon deletion mutations in the α globin gene and it is a quantitative analysis and does not require specific primers as Gap-PCR. However, MLPA has a lower resolution and difficulty in detecting *trans HBA* deletions (Sabath, 2017). Next, the LAMP technique has a high sensitivity, specificity, simplicity and cost effective as well as eliminates the need for highly sophisticated thermocycling equipment in detecting the deletion mutation (Notomi *et al.*, 2000). However, there was false positive result reported for $\alpha^{4.2}$ and $\alpha^{3.7}$ deletion when using LAMP technique. This might be due to the fact that deleted genes were closely related to the wild type gene sequence before mutation, which was more likely to have false amplification (Wang *et al.*, 2020). NGS reduces the possibility for obtaining false-negative results and it helps to eliminate the need for repeating blood sample and further referral tests in detecting α-thalassaemia as it is an accurate technique (Zhuang *et al.*, 2019). However, the steps or techniques required in NGS is tough, time consuming and expensive (Sabath, 2017). Therefore, Gap-PCR is used for the identification of common deleted mutation in *HBA.*

Table 1.2 Common deletion types of mutation in thalassaemia detected by Gap-PCR

Disorder	Deletion mutation	Reference
α^0	$--^{SEA}$	(Old *et al.*, 2012)
	$--^{MED}$	
	$-(\alpha)^{20.5}$	
	$--^{FIL}$	
	$--^{THAI}$	
α^+	$-\alpha^{3.7}$	
	$-\alpha^{4.2}$	

1.7.2 Single tube multiplex amplification refractory mutation system-PCR

Multiplex amplification refractory mutation system (MARMS) is based on the PCR system where specific primers with a nucleotide at their 3′ end corresponding to

normal or mutant sequence and distinguish the normal and mutant alleles. Amplification of the DNA fragment occurred only if there was a perfect match of the primer with the genomic sequence (Old *et al.*, 2012). Non-deletional α-thalassaemia results from point mutations, deletions or insertion involving α globin genes. Therefore, the mutation can be detected by MARMS-PCR as the specific primer leads to selective amplification of mutated genes (Old *et al.*, 2012). Single tube MARMS-PCR was developed for the detection of common and severe non-deletional α-thalassaemia found in Southeast Asia as listed in Table 1.1. This is a rapid, simple and non-radioactive method to identify the mutations (Eng *et al.*, 2001). Previously, the Southern blot technique was used but it is time consuming, expensive, and labour intensive. Therefore, Southern blot application in a larger scale population screening was challenging despite the fact it was useful for individual screening. This method is relatively simple and cost-effective for the detection of α-thalassaemia mutations (Shaji *et al.*, 2000). Apart from that, DNA sequencing is able to identify both novel and uncommon non-deletion mutations (Sabath, 2017). DNA sequencing is known as an accurate method as it is applied to the whole genome to detect any presence of single point mutations, insertions, translocations, or deletion. However, MARMS-PCR is used for the detection of non-deletional mutation on a large scale to increase efficiency and involves lower cost compared to the sequencing method (Farashi and Harteveld, 2018).

1.7.3 Multiplex ligation dependent probe amplification

Multiplex ligation dependent probe amplification is a technique proven to detect common and rare deletions that could not be detected in cases after performing routine diagnostic techniques. Therefore, the limitation of Gap-PCR to detect rare mutations is a privilege of the MLPA technique (Old *et al.*, 2012). The MLPA is a

rapid quantitative and simple technique used to detect α globin gene cluster deletion (Yuregir *et al.*, 2016). This technique-initiated DNA denaturation and incubated the denatured DNA with a mixture of MLPA probes. The MLPA probes consist of left probe oligonucleotides (LPO) and right probe oligonucleotides (RPO). Next, is the hybridization of both probes to adjacent target sequences, and in the ligation reaction, the probes will be ligated. The ligated probes will be amplified in the PCR process. The quantity of target sequences in the sample is based on the amount produced from the amplification of ligated probes (Sabath, 2017). A study suggested that MLPA is an effective method. However, it is known to be more suitable as a complementary method rather than a stand-alone technique (Yuregir *et al.*, 2016) because it is unable to detect small deletions (Sabath, 2017). Hence, analysing the deletion mutations in the *HBA* by Gap-PCR and MLPA results in more accurate findings, and increases the detection of carriers. However, MLPA is unable to detect non-deletional mutations. Thus, this technique alone is not sufficient to diagnose α-thalassaemia patients. Therefore, a combination of a few techniques, including PCR and sequencing, will produce more accurate results. Furthermore, the techniques has been published previously by using a commercially available kit, SALSA MLPA Kit, HBA140-B4 (MRC-Holland), specifically to detect α-thalassaemia (Yuregir *et al.*, 2016).

1.7.4 Sanger sequencing

Sanger sequencing is a technique that sequences specific or target DNA regions by using oligonucleotide primers. The technique begins with double stranded DNA (dsDNA) denaturation and produces single stranded DNA (ssDNA). Then the ssDNA was ligated to the primers and elongated with a mixture of deoxynucleotide primers (dNTPs), consisting of Adenine (A), Guanine (G), Tyrosine (T) and Cytosine (C) to build a new strand. In addition, each dNTP has a fluorescent marker, so the attachment

of each dNTPs will fluoresce based on the nucleotide. By convention, each dNTPs including A, G, T and C, will fluoresce green, black, red, and blue respectively. Next, the fluorescent intensity will be detected by the machine and translated into a peak. Sanger sequencing is a technique that is capable of differentiating homozygous and heterozygous variants based on the fluorescence intensity (Gomes and Korf, 2018). Thalassaemia is caused by single nucleotide mutations and copy number variants, while Sanger sequencing is commonly employed to diagnose the disease. However, Sanger sequencing is known to be consistent and a confirmatory test to detect single nucleotide variants in the *HBA* (Fan *et al.*, 2019). The main limitation of the technique is that it is not useful in detecting deletion mutations that cause α-thalassaemia. Meanwhile, Sanger sequencing is the most comprehensive method to detect the non-deletional mutations. Therefore, the mutations that were identified by MARMS-PCR are detectable in the sequencing technique (Sabath, 2017).

1.8 Problem statement

The prevalence of α-thalassaemia is higher than β-thalassaemia worldwide. However, the detection of α-thalassaemia by Southern Blot analysis was cumbersome and time-consuming previously. Since the national thalassaemia screening program was initiated in various countries for many years (Mat *et al.*, 2020), the molecular diagnosis became important. Molecular characterization of α-thalassaemia in any given population is a prerequisite for the establishment of prenatal diagnosis and genetic counselling services, especially in those cases which cannot be confirmed by haematological parameters. Though haematological parameters play an important role in α-thalassaemia, yet there is dearth of information available to describe between the haematological parameters in different α globin gene mutations. Hence, this research

will be oriented towards detection of deletion and non-deletion mutations of α-thalassaemia by multiplex Gap-PCR and MARMS-PCR respectively. Since both these methods are designed for detection of only the common mutations, some unresolved cases will be proceeded with direct DNA sequencing and MLPA. Even though deletion mutation of α genes is mostly reported and detected than non-deletional mutations, the latter determinants may give rise to a more severe reduction in α chains thus giving a more severe clinical presentation, which needs to be identified early. Next, the interactions of deletion and non-deletion mutation lead to more severe clinical symptoms. Therefore, the detection of non-deletion α globin genes are principal as it is also identified in our community. Furthermore, MARMS-PCR for non-deletional α-thalassaemia is not established yet in our laboratory. Therefore, a single tube MARMS assay will be established in the haematology laboratory of Hospital Universiti Sains Malaysia to detect the non-deletional type of α-thalassaemia.

1.9 Justification of study

Molecular analysis is a definitive method for the diagnosis of α-thalassaemia carriers. The MARMS-PCR and Gap-PCR used in this study are less time consuming, rapid, and less laborious compared to the Southern blot technique used before. Nevertheless, some mutations cannot be detected by these techniques, as they are designed to detect only the common ones. DNA sequencing able to detect the uncommon mutations while MLPA able to detect the uncommon deletional mutations. Thus, the combination of DNA sequencing and MLPA will increase the diagnostic accuracy in this disorder. Besides, this research study will pave way in establishing the MARMS-PCR and MLPA techniques in the haematology laboratory of Hospital Universiti Sains Malaysia. Hence, the study of the molecular result with the routine

RBC parameters will help in complementing each other to arrive at an accurate diagnosis. Thus, these parameters can be used as a preliminary screening tool for detecting α gene mutations in thalassaemia patients.

1.10 Objectives

1.10.1 General objective

To investigate the red blood cells parameters and molecular characterisation of α globin gene mutations and its variants in α-thalassaemia patients diagnosed in Hospital Universiti Sains Malaysia.

1.10.2 Specific objectives

1. To determine the prevalence of α globin gene mutations among the cases sent for Hb analysis in Hospital Universiti Sains Malaysia using single tube MARMS-PCR, multiplex Gap-PCR assays, MLPA and DNA sequencing.
2. To describe the RBC parameters (Hb, MCV, MCH, MCHC, RBC, RDW, Hct, Hb A, HbA_2 and Hb F) based on the α-thalassaemia mutations in Hospital Universiti Sains Malaysia.

1.11 Research questions

1. What is the prevalence of α globin gene mutations among α-thalassaemia patients in Hospital Universiti Sains Malaysia?
2. Do the RBC parameters (Hb, MCV, MCH, MCHC, RBC, RDW, Hct, Hb A, HbA_2 and Hb F) change based on the α-thalassaemia mutation in Hospital Universiti Sains Malaysia?

1.12 Research hypotheses

1. There is high prevalence of α globin gene mutations in Hospital Universiti Sains Malaysia.
2. There are changes in the RBC parameters (Hb, MCV, MCH, MCHC, RBC, RDW, Hct, Hb A, HbA_2 and Hb F) based on the α-thalassaemia mutation in Hospital Universiti Sains Malaysia.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Study design

This is a cross-sectional study that was carried out in Craniofacial Science laboratory, (School of Dental Sciences), Centre Research Laboratory (School of Medical Sciences Universiti Sains Malaysia), Haematology Laboratory (Hospital Universiti Sains Malaysia) and Haematology Unit, Institute for Medical Research. The study was conducted from October 2020 to June 2022.

2.1.2 Data collection

This study was carried out on patients who has sent for Hb analysis and suspected of having α-thalassaemiawith hypochromic microcytic indices. Patients that show Hb analysis suspected of β-thalassaemia were excluded from the study. The blood samples were sent to haematology laboratory of Hospital Universiti Sains Malaysia for Hb analysis. Clinical and laboratory data consisting of age, race, gender, full blood count and clinical diagnosis were recorded.

Molecular analyses were performed in Craniofacial Science laboratory, (School of Dental Sciences), Centre Research Laboratory (School of Medical Sciences Universiti Sains Malaysia) and Haematology Unit, Institute for Medical Research. This study was approved by the Medical Research & Ethics Committee (NMMR-21-606-58737) and Human Research Ethics Committee, Universiti Sains Malaysia (USM/JEPeM/20020104). The ethical approvals are attached in the appendices A and B.