HFE GENE POLYMORPHISMS IN MALAYS

FRANCIS JATTA

UNIVERSITI SAINS MALAYSIA

2020

HFE GENE POLYMORPHISMS IN MALAYS

by

FRANCIS JATTA

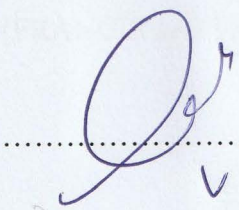
Thesis submitted in partial fulfilment of the requirements
for the degree of
Master of Science (Forensic Science)

September 2020

CERTIFICATE

This is to certify that the thesis titled HFE gene polymorphisms in Malays is the bona fide record of Francis Jatta's research work done under my supervision during the period from February 2020 to September 2020. I read the thesis and it conforms in my opinion to all accepted principles of scholarly presentation and is fully adequate, in nature and consistency, as a thesis to be submitted in partial fulfilment for the degree of Master of Science (Forensic Science).

Supervisor,



.....

(DR EDINUR HISHAM BIN ATAN)

Date: 17/9/2020

DECLARATION

I declare this study to be the outcome of my work, except where stated otherwise and adequately acknowledged. I also confirm that it has not been previously submitted to the Universiti Sains Malaysia or other institutions for the award of any other degree. Therefore, I grant Universiti Sains Malaysia the right to use the thesis for teaching, research, and promotional purposes.

.....

(FRANCIS JATTA)

Date: 17/09/2020.

ACKNOWLEDGEMENT

All praises and thanks to our creator God for His precious gift of life, eternal love, and mercy. His encouragement and security in difficult times helped me to carry out this thesis and complete it.

My sincere and deepest gratitude goes out to my supervisor, Dr Edinur Hisham Bin Atan, for his tireless advice and support during my work. My special thanks go to our dedicated and hard-working Master of Science (Forensic Science) mixed-mode management team. In the course of my study, I will not be fair to Norul Hajar Che Ghazali and Abdoulie M. Sanyang for their time and support. My warm appreciation goes to my fellow students in Forensic Science for their helpful guidance and inspiration during the course.

I would also like to thank the Government of The Gambia for the scholarship they gave me to pursue a Master of Science (Forensic Science) in their efforts to develop the capacity of the Forensic Unit of The Gambia Police Force.

Finally, I extend my heartfelt gratitude to my mother, children, and siblings for their unconditional love, encouragement, motivation, prayers, and patience they endured during the two years of my studies. I announce this thesis to them now and hope that it will put a huge smile on their faces.

TABLE OF CONTENTS

CERTIFICATE ..... **ii**

DECLARATION ..... **iii**

ACKNOWLEDGEMENT ..... **iv**

TABLE OF CONTENTS ..... **v**

LIST OF TABLES ..... **vii**

LIST OF FIGURES ..... **viii**

LIST OF SYMBOLS ..... **ix**

LIST OF ABBREVIATIONS ..... **xi**

ABSTRAK ..... **xiii**

ABSTRACT ..... **xiv**

CHAPTER 1 INTRODUCTION **1**

1.1 Background of study 1

1.2 Rationale of study 2

1.3 Objectives 3

1.1 General objective 3

1.2 Specific objectives 3

CHAPTER 2 LITERATURE REVIEW **4**

2.1 Haemochromatosis gene (HFE) 4

2.2 HFE structure and function 5

2.3 Prevalence of HFE gene SNPs 9

2.3.1 Hereditary haemochromatosis 9

2.4 Malay population in Peninsular Malaysia 18

CHAPTER 3 METHODOLOGY **20**

3.1 Study design 20

3.2 Sample collection 20

3.2.1 Inclusion criteria 21

3.2.2 Exclusion criteria 21

3.3 DNA extraction from whole blood 21

3.3.1 Agarose gel electrophoresis of extracted genomic DNA 22

3.3.2 DNA quantitation 23

3.4 HFE typing using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) 23

3.4.1 Genotyping for C282Y 23

3.4.2 Genotyping for H63D 24

3.4.3 Genotyping for S65C 25

3.5 Statistical methods 25

3.5.1 Allele and genotype frequencies 25

3.5.2 Hardy-Weinberg equilibrium (HWE) analysis 26

CHAPTER 4 RESULTS 28

4.1 HFE genotype frequencies and HWE in the studied Malay population 28

4.2 HFE allele frequencies in Malays and reference populations 29

4.3 Estimate risk of HH associated with C282Y, H63D, and S65C mutations in Malays 30

CHAPTER 5 DISCUSSION 31

CHAPTER 6 CONCLUSION 34

REFERENCES 35

APPENDICES 51

APPENDIX A: Human Ethical approval (USM/JEPeM/16050191) 51

APPENDIX B: Human Ethical approval (NMRR-16-1399-31311 (IIR) ... 52

LIST OF TABLES

Table 2.1: Genotype frequencies of C282Y, H63D, and S65C mutations in the HFE gene 11

Table 2.2: The C282Y, H63D and S65C allele frequencies in HFE characterised populations 13

Table 4.1: HFE genotype frequencies and HWE tests in Malays 28

Table 4.2: Allele frequencies of C282Y, H63D and S65C gene variants in Malays 30

Table 4.3: HFE genotypes associated with HH disease development in Malays 30

LIST OF FIGURES

Figure 2.1: Schema shows location of the HFE gene on the short arm of chromosomal 6. 7

Figure 2.2: Structure of HFE with the extracellular protein domains (α1 and α2) and the β2M associated with the immunoglobulin-like region (α3)..... 8

Figure 2.3: HFE protein and TFR2 signalling pathway..... 8

Figure 2.4: Map of Malaysia showing the thirteen states and three federal territories. 19

LIST OF SYMBOLS

α	Alpha
β	Beta
β_2M	Beta 2-microglobulin
%	Percentage
<	Less than
χ^2	Chi-square
Σ	sum of tested genotypes
µL	Microliter
µM	Micromolar
a	Number observed for a particular HFE allele
dH_2O	Distilled water
e	Number of expected genotype
g	Gram
km^2	Square kilometre
$MgCl_2$	Magnesium chloride
mL	Millilitre

n	Sample number
ng	Nanogram
o	Number of observed genotype
p	Frequency of allele X in a population
p^2	Frequency of homozygous genotype (XX)
q	Frequency of allele Y in a population
q^2	Frequency of homozygous genotype (YY)
2pq	Frequency of heterozygote genotype (XY)
s	Second
V	Volts
X	Times
x *g*	Times gravity

LIST OF ABBREVIATIONS

aka	Also known as
bp	Base pairs
Cyto	Small cytoplasmic portion
C282Y/C282Y	Homozygous C282Y mutation
C282Y/WT	Heterozygous C282Y mutation
C282Y/H63D	Compound heterozygous mutation for C282Y and H63D
C282Y/S65C	Compound heterozygous mutation for C282Y and S65C
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic acid
H63D/H63D	Heterozygous H63D mutation
H63D/WT	Heterozygous H63D mutation
H63D/S65C	Compound heterozygous mutation for H63D and S65C
HFE	Haemochromatosis gene
HH	Hereditary haemochromatosis
HLA	Human leukocyte antigens
HUGO	Human Genome Organisation

HWE	Hardy-Weinberg equilibrium
kb	Kilobases
KIR	Killer cell immunoglobulin-like receptor
MHC	Major histocompatibility complex
MICA	Major histocompatibility complex class I chain-related gene A
NT	Not tested
S65C/S65C	Homozygous S65C mutation
S65C/WT	Heterozygous S65C mutation
SEA	Southeast Asia
SNP	Single Nucleotide Polymorphism
TBE	Tris borate ethylenediaminetetraacetic
TFR	Transferrin receptor
USA	United States of America
UV	Ultraviolet
wt	Wild type

POLIMORFISME GEN HFE PADA ORANG MELAYU

ABSTRAK

Variasi gen hemokromatosis (HFE) telah dilaporkan dapat merencatkan peranan hepcidin, mengakibatkan gangguan genetik autosomal yang disebut hemokromatosis keturunan (HH). Variasi genetik ini merangkumi C282Y, H63D, dan S65C yang telah banyak dikaji di Eropah dan tidak ada untuk penduduk Malaysia. Tujuan kajian ini adalah untuk menilai variasi gen HFE pada orang Melayu. Sebanyak 35 sampel darah dikumpulkan dan genotip untuk C282Y, H63D, dan S65C menggunakan polimorfisme panjang sekatan reaksi-sekatan reaksi-polimerase (PCR-RFLP). Hasil kajian menunjukkan bahawa taburan frekuensi heterozigot H63D (5.71%) adalah dua kali lebih tinggi daripada heterozigot untuk S65C (2.86%). Walau bagaimanapun, tidak ada individu Melayu yang homozigot atau heterozigot untuk C282Y. Oleh itu, risiko HH pada orang Melayu rendah berbanding dengan orang Eropah yang mempunyai frekuensi tinggi alel dan genotip C282Y, H63D, dan S65C. Kajian masa depan harus merangkumi kumpulan populasi lain di Malaysia untuk penjelasan struktur populasi yang lebih baik dan risiko untuk mengembangkan HH di negara ini.

HFE GENE POLYMORPHISMS IN MALAYS

ABSTRACT

Variations in haemochromatosis gene (HFE) have been reported to inhibit the role of hepcidin, resulting in an autosomal genetic disorder called hereditary haemochromatosis (HH). These genetic variations include C282Y, H63D, and S65C which have been largely studied in Europeans and none for the Malaysian population. The aim of this study is to evaluate variations within the HFE gene in Malays. A total of 35 blood samples were collected and genotyped for C282Y, H63D, and S65C using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). The findings of the study showed that the frequency distribution of heterozygous H63D (5.71%) was twice higher than those heterozygous for S65C (2.86%). However, no Malay individuals were homozygous or heterozygous for C282Y. Therefore, risk for HH in Malays is low as compared with Europeans which have high frequencies of C282Y, H63D, and S65C alleles and genotypes. Future study should include other population groups in Malaysia for better elucidation of population structure and risk for developing HH in the country.

CHAPTER 1

INTRODUCTION

1.1 Background of study

The human genome consists thousands of polymorphic genes (Wain *et al.*, 2002), including cytokine, human leukocyte antigen (HLA), killer cell immunoglobulin-like receptor (KIR), major histocompatibility complex class I chain-related gene A (MICA) and homeostatic iron regulator or haemochromatosis gene (Kumánovics *et al.*, 2003; Reuben *et al.*, 2017). The haemochromatosis gene (HFE) is a major histocompatibility complex (MHC) class I protein located on the short arm of chromosome 6 (6p22.2) that regulates the amount of iron in human body (Barton *et al.*, 2015; Harrison and Bacon, 2005). In 1996, Feder *et al.* identified HFE gene and its mutation that contributed to a medical disorder known as hereditary haemochromatosis (HH). HH is a condition where iron is overloaded in internal organs such as heart, liver and pancreas (Barton and Edwards, 2018; Bellissimo and Agarwal, 2016; Milman *et al.*, 2019).

Many studies have shown the associations between HFE mutations and HH (Alves *et al.*, 2016; Canavesi *et al.*, 2012; Hollerer *et al.*, 2017; Jin *et al.*, 2010; Juzėnas *et al.*, 2016). Such studies also revealed unique frequency distributions of HFE gene variants between people of different ethnicities and their risk profiles to HH (Ali *et al.*, 2018; Barton *et al.*, 2015; Merryweather-Clarke *et al.*, 1997; Spínola *et al.*, 2011; Viprakasit *et al.*, 2004). The HH is mainly caused by the homozygous C282Y SNP (Allen *et al.*, 2008). Moreover, H63D homozygous or compound heterozygote (C282Y/H63D or C282Y/S65C) genotypes might cause mild HH for those with existing medical conditions (viral hepatitis and non-alcohol fatty liver disease) and unhealthy lifestyle (alcoholism) (Allen *et al.*, 2008; Gurrin *et al.*, 2009; Walsh *et al.*,

2006). Therefore, the present screening of HFE gene variants in Malays is essential as it will provide for the first time C282Y, H63D and S65C population data for this ethnic group and can be used for HH risk assessments.

1.2 Rationale of study

HFE gene codes the hepcidin protein that regulates iron uptake in human body (Loréal *et al.*, 2018). Mutations in the HFE gene might cause iron overload and lead to HH (Spínola *et al.*, 2011). HFE gene polymorphisms (C282Y, H63D and S65C) were uniquely distributed in unrelated ethnic populations worldwide (Barton *et al.*, 2015; Distante *et al.*, 2004), yet no such data is available for any of the Malaysian population groups. Therefore, this study aims to provide frequency distributions of C282Y, H63D and S65C HFE single nucleotide polymorphisms (SNPs) in Malays and use the generated SNP datasets for estimating risks of developing HH in this ethnic group.

1.3 Objectives

1.1 General objective

1. To study variations of HFE gene in Malays.

1.2 Specific objectives

2. To provide frequency distributions of C282Y, H63D and S65C HFE gene SNPs in Malays.
3. To compare frequencies of C282Y, H63D, and S65C in Malays and other populations characterised for these HFE gene SNPs.
4. To estimate the risk of HH associated with C282Y, H63D, and S65C mutations in Malays.

CHAPTER 2

LITERATURE REVIEW

2.1 Haemochromatosis gene (HFE)

The HFE gene code for hepcidin protein that involves in regulating iron intake in human body (Loréal *et al.*, 2018; Nemeth and Ganz, 2009). Feder *et al.* (1996) used a positional cloning technique to examine DNA specimens from a cohort of 178 Caucasian HH patients. The study identified two missense mutations in HFE gene; C282Y (rs1800562) and H63D (rs1799945) in exon 4 and 2, respectively. Feder *et al.* further observed that most of the patients were homozygous for C282Y arising from the transformation of guanine at nucleotide 845 to adenine (845G→A). Few other patients were either H63D homozygous or C282Y/H63D compound heterozygous. The H63D SNP resulted in nucleotide substitution at position 187 from cytosine to guanine (187C→G) (Feder *et al.*, 1996; Katsarou *et al.*, 2019; Le Gac and Férec, 2005). The identified gene was referred as HLA-H because it is close to the major histocompatibility complex (MHC) class I gene cluster on the short arm of chromosome 6. In 1997, the HUGO Gene Nomenclature Committee approved a new nomenclature for the HLA-H gene and later known as haemochromatosis (HFE) gene (Bodmer *et al.*, 1997).

A third mutation (S65C) within HFE gene was reported in a sample population of 711 HH patients (Mura *et al.*, 1999). The S65C (rs1800730) is located at nucleotide position 193 in exon 2 of the HFE gene and replaces adenine with thymidine (193A→T) (Mura *et al.*, 1999).

A number of studies of HFE gene SNPs were performed in Asians to determine the most prevalent mutation (C282Y, H63D, and S65C) in the population (Ali *et al.*,

2018; Khusainova *et al.*, 2013; Lin *et al.*, 2007; Merryweather-Clarke *et al.*, 1997; Pointon *et al.*, 2003). In 1997, a studied population of 457 Asians reported C282Y and H63D mutations in Asia (Merryweather-Clarke *et al.*, 1997). In addition, C282Y and H63D alleles were also reported in North American HFE gene mutation study that recruited 12,772 Asians (Adams *et al.*, 2005). The presence of S65C SNP was determined in Asians but found to absent in a studied population of 314 Vietnamese who lived in Vietnam. A similar report of no S65C mutation was reported in a study of HFE gene SNP in 395 Chinese Han in China (Lin *et al.*, 2007).

2.2 HFE structure and function

The 12 kb HFE gene is located at chromosome 6p22.2p (Figure 2.1). The HFE gene has seven exons, which code a predicted 343-amino acid protein. The protein contains a signal sequence, peptide-binding domains (α1, α2 and α3), a transmembrane domain (TM), and a small cytoplasmic (cyto) part. The α1 and α2 domains are the receptor-binding sites of extracellular transferrin that associate with transferrin receptor 1 (TFR1) while the α3 domain is the immunoglobulin-like region that binds transferrin receptor 2 (TFR2). HFE protein binds to beta 2-microglobulin (β2M) and controls hepcidin expression (Figure 2.2) (Barton *et al.*, 2015; Chua *et al.*, 2007; Griffiths, 2007; Santos *et al.*, 2010).

Hepcidin produced in the liver regulates the uptake of iron and controls the presence of ferroportin which carries iron from the duodenum into the bloodstream. Therefore, the signalling pathways should be preserved for the interaction of the HFE protein and transferrin receptor 2 (TFR2) (Utzschneider and Kowdley, 2010). Nevertheless, HFE mutation inhibits the formation of disulfide bonds in the protein's

α3 region, thus disrupting the pathway to engage with β2M (Feder *et al.*, 1997; Lebrón *et al.*, 1998).

The disruption of the pathway would prevent the adequate synthesis of hepcidin (Figure 2.3) (Corradini *et al.*, 2009; Utzschneider and Kowdley, 2010), thus favouring the accumulation of iron leading to iron overload that may aggravate into HH (Cançado and Chiattone, 2010; Murphree *et al.*, 2020; Pietrangelo, 2015; Porto *et al.*, 2016). The risk of developing severe conditions such as porphyria cutanea tarda (Barton and Edwards, 2016; Egger *et al.*, 2002), cirrhosis (Barton *et al.*, 2018), diabetes mellitus (Jennison and Wainwright, 2018; Raju and Venkataramappa, 2018), cardiac diseases (Rasmussen *et al.*, 2001; Sumi *et al.*, 2020), hepatocellular carcinoma and cancers (Abraham *et al.*, 2005; Jayachandran *et al.*, 2020; Jin *et al.*, 2011; Lv *et al.*, 2016; Robinson *et al.*, 2005) increases if HH is left untreated (Crownover and Covey, 2013; Gabriková *et al.*, 2012).

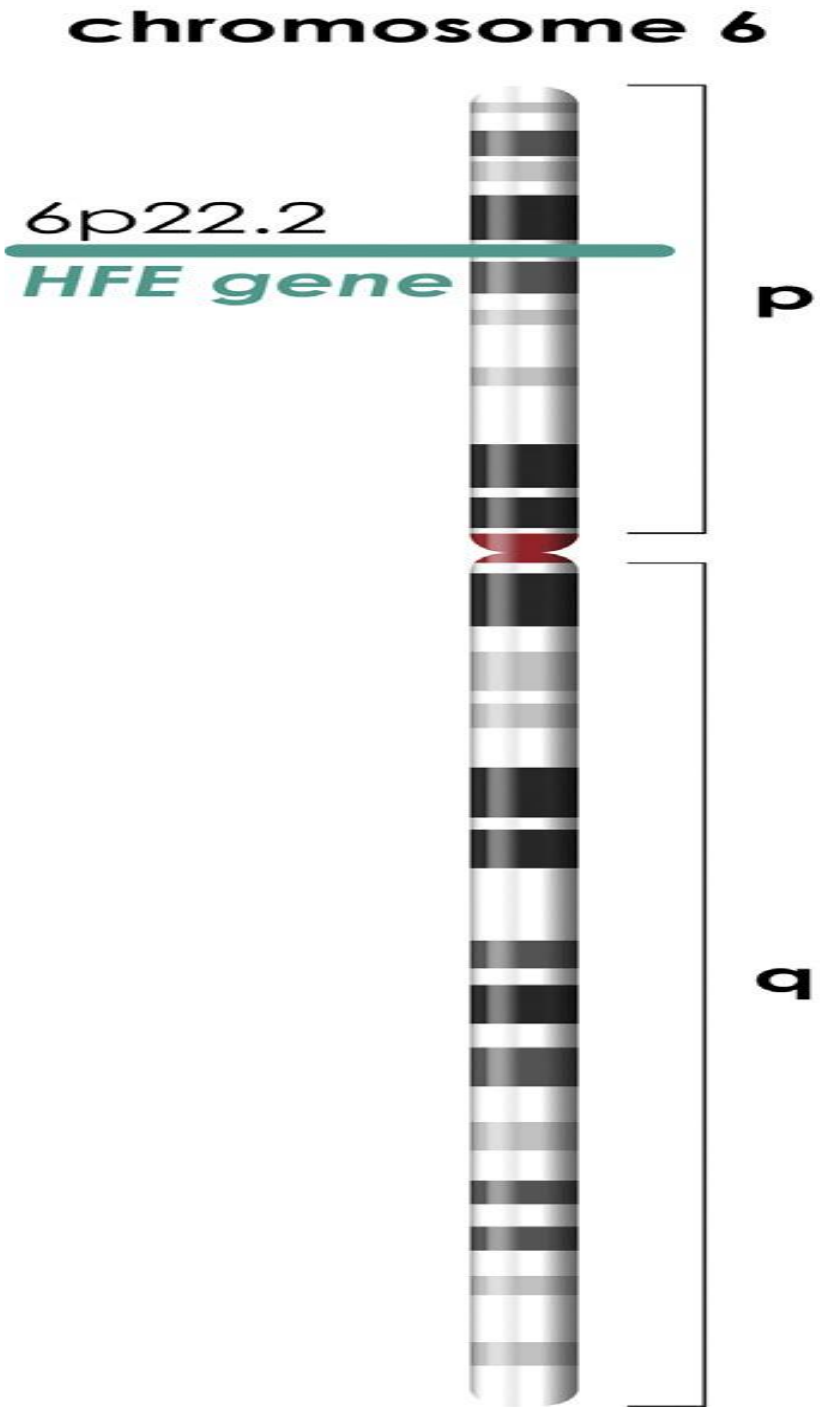


Figure 2.1: Schema shows location of the HFE gene on the short arm of chromosomal 6. The figure adapted from Katsarou *et al.* (2019).

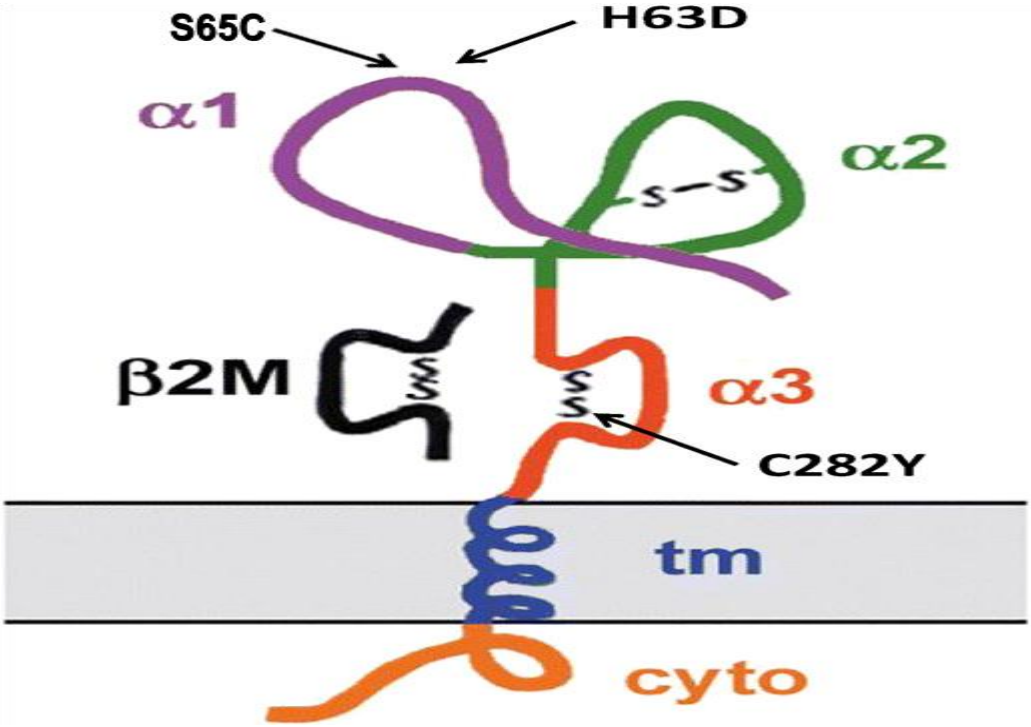


Figure 2.2: Structure of HFE with the extracellular protein domains (α1 and α2) and the β2M associated with the immunoglobulin-like region (α3). Figure modified from Barton *et al.* (2015).

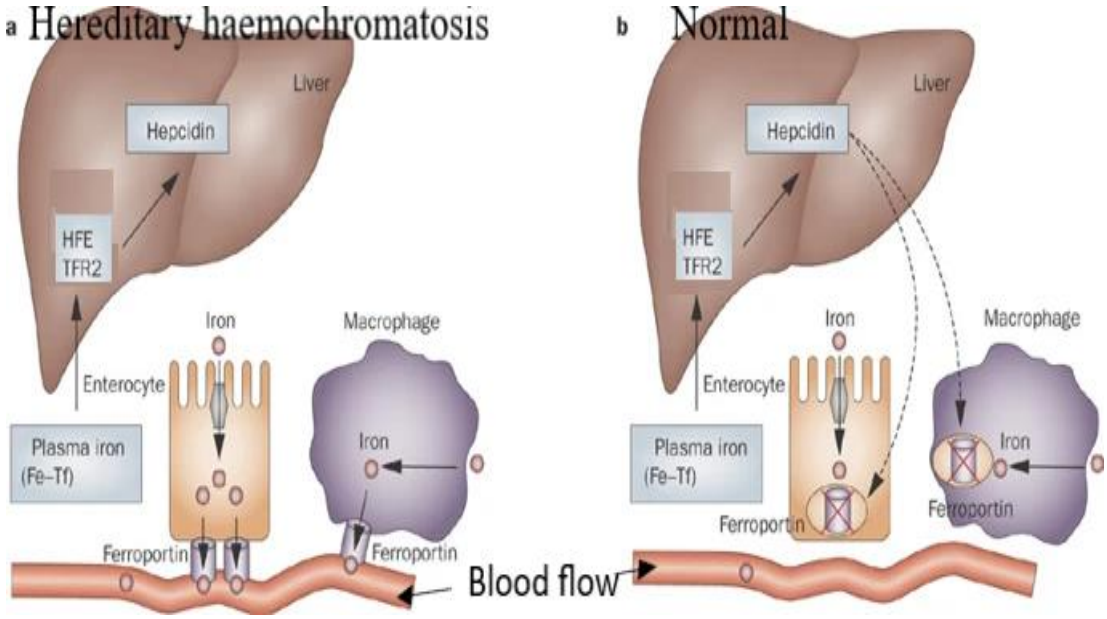


Figure 2.3: HFE protein and TFR2 signalling pathway. Figure modified from Utzschneider and Kowdley (2010).

2.3 Prevalence of HFE gene SNPs

The most common HFE SNP associated with HH is the C282Y (Gurrin *et al.*, 2009). This HFE SNP is typically high among Europeans but low or absent among non-European populations (Ali *et al.*, 2018; Merryweather-Clarke *et al.*, 1997).

A second mutation, H63D, was reported in 5% of HH patients (1% were homozygote for H63D while 4% heterozygote compound of C282Y/H63D) (Adams, 2015). The H63D mutation, by comparison, is relatively more common in worldwide populations with the highest recorded among Europeans (de Juan *et al.*, 2001) and the lowest in Southeast Asia, Africa, Central and South America populations (Khusainova *et al.*, 2013).

A third mutation, S65C, is infrequently linked to severe iron overload and has therefore not been extensively investigated (Aranda *et al.*, 2010; Brissot *et al.*, 2017; Holmström *et al.*, 2002; Mura *et al.*, 1999; Muro *et al.*, 2007). Individuals homozygous for S65C or compound heterozygote (C282Y/S65C) are associated with mild HH symptoms such as joint pain, fatigue, unintended weight loss, abnormal liver enlargement and weakness (Alexander and Kowdley, 2009; Ayonrinde *et al.*, 2008; Cheng *et al.*, 2009; Gurrin *et al.*, 2009).

2.3.1 Hereditary haemochromatosis

Hereditary haemochromatosis (HH) is a genetic blood disorder associated with iron metabolism (Pantopoulos, 2018). The hereditary disorder caused failure of the body to eliminate the excess iron obtained from food after absorption due to the inefficiency of the excretory mechanism. Consequently, this excess iron leads to a state of over-accumulation, which is harmful to cells. Usually, this clinical disease

manifestation is typically associated with C282Y mutation in HFE gene and it can be fatal if lately detected and untreated (Katsarou *et al.*, 2016). Katsarou *et al.* (2016) reported that 3-5 individuals in a 1,000 population are C282Y homozygous in North-western Europe populations.

HH affects people all over the world (Shiono *et al.*, 2001), most notably the Northern European origins (Bacon *et al.*, 2011; Juzėnas *et al.*, 2016; Tsui *et al.*, 2000; Wallace and Subramaniam, 2016). HH is commonly asymptomatic and primarily affects adult males compared to females. Menstruation is a possible explanation for this late phenomenon in women (Al Wayli *et al.*, 2011; Katsarou *et al.*, 2016). Clinical symptoms appear in men at and after the age of 40 (Puntarulo, 2005).

Genetic screening of HFE gene is thus essential because it will allow for early treatment of HH (Bassett, 2010; Brissot *et al.*, 2008). Phlebotomy therapy prevents disease development and normalises iron level (Asimakopoulou *et al.*, 2017; Koshy *et al.*, 2020; Kowdley *et al.*, 2019; Salgia and Brown, 2015). Tables 2.1 and 2.2 show the genotype and allele frequencies for C282Y, H63D and S65C mutations across the world.

Table 2.1: Genotype frequencies of C282Y, H63D, and S65C mutations in the HFE gene

Population	Genotype frequencies (%)									References
	C282Y/ wt	C282Y/ C282Y	H63D / wt	H63D/ H63D	S65C/ wt	S65C/ S65C	C282Y/ H63D	C282Y/ S65C	H63D /S65C	
Madeira Island, Portugal	0.00	0.00	32.00	3.90	1.30	0.00	0.65	0.00	0.65	Spínola *et al.* (2011)
Espírito Santo, Brazil	3.33	0.00	20.83	0.83	0.83	0.00	0.00	0.00	0.83	Alves *et al.* (2016)
Chinese Han	0.00	0.00	4.60	0.00	0.00	0.00	0.00	0.00	0.00	Lin *et al.* (2007)
Netherlands	12.70	0.30	23.60	2.20	0.00	0.00	0.00	0.00	0.00	Cobbaert *et al.* (2012)
Jordanian Arab	0.00	0.00	19.80	1.90	0.00	0.00	0.00	0.00	0.00	Alkhateeb *et al.* (2013)
Thai, Thailand	0.00	0.00	20.00	1.00	0.00	0.00	0.00	0.00	0.00	Viprakasit *et al.* (2004)
Czech Republic	6.86	0.00	26.61	1.67	2.49	0.00	1.87	0.00	0.42	Čimburová *et al.* (2005
Pakistanis, UK	0.00	0.00	4.00	0.50	0.00	0.00	0.00	0.00	0.00	Ali *et al.* (2018)
Iranian	1.20	0.00	8.50	0.00	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2006)
Turkic	0.80	0.00	15.20	1.10	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2006)
Lithuanian	5.10	0.10	26.70	2.60	3.70	0.00	1.19	0.10	0.40	Kucinskas *et al.* (2012)
Belarusian	0.00	0.00	25.00	8.30	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)

Table 2.1: Cont.

Population	Genotype frequencies (%)									References
	C282Y/ wt	**C282Y/ C282Y**	**H63D / wt**	**H63D/ H63D**	**S65C/ wt**	**S65C/ S65C**	**C282Y/ H63D**	**C282Y/ S65C**	**H63D /S65C**	
Northeastern Brazil	3.75	0.00	25.62	1.25	0.62	0.00	0.63	0.00	0.00	Leão *et al.* (2014)
Ukrainian	0.00	0.00	25.50	2.10	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Moldavian	0.00	0.00	19.10	0.00	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Karelian	0.00	0.00	19.00	1.00	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Volga–Ural	0.00	0.00	21.70	1.80	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Kazakh	0.00	0.00	15.50	0.00	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Uzbek	0.00	0.00	11.10	0.00	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Uighur	0.00	0.00	16.40	0.00	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Kirghiz	0.00	0.00	16.10	0.00	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Turkmen	0.00	0.00	16.70	5.50	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Kurd	0.00	0.00	12.20	0.00	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)
Tajik	0.00	0.00	4.80	0.00	0.00	0.00	0.00	0.00	0.00	Khusainova *et al.* (2013)

Table 2.2: The C282Y, H63D and S65C allele frequencies in HFE characterised populations

Population	Ethnicity	Allele frequencies (%)			References
		C282Y	H63D	S65D	
Myanmar	Burmese	0.00	2.90	NT	Viprakasit *et al.* (2004)
Indonesia	Indonesian	0.00	2.80	NT	Merryweather-Clarke *et al.* (1997)
Thailand	Thais	0.00	3.20	NT	Pointon *et al.* (2003)
Vietnam	Vietnamese	0.00	4.90	0.00	Pointon *et al.* (2003)
Java	Javanese	0.00	NT	NT	Cullen *et al.* (1998)
Sri Lanka	Sri Lankans	0.00	9.20	NT	Merryweather-Clarke *et al.* (1997)
Pakistan	Pakistanis	0.00	8.00	NT	Ali *et al.* (2018)
China	Chinese Han	0.00	2.30	0.00	Lin *et al.* (2007)
China	Hong Kong Chinese	0.00	2.80	NT	Merryweather-Clarke *et al.* (1997)
Taiwan	Taiwanese	0.00	1.90	NT	Merryweather-Clarke *et al.* (1997)
Saudi Arabia	Saudi Arabians	0.00	8.50	NT	Merryweather-Clarke *et al.* (1997)
Jordan	Jordanian	0.00	11.25	0.11	Alkhateeb *et al.* (2009)
India	Indians	0.50	7.50	NT	Nadkarni *et al.* (2017)
Japan	Japanese	0.00	0.99	NT	Sohda *et al.* (1999)
South Korea	Koreans	0.00	3.80	NT	Lee *et al.* (2000)
Kirghizstan	Kirghiz	NT	8.10	0.00	Khusainova *et al.* (2013)
Kurdistan	Kurds	NT	6.10	0.00	Khusainova *et al.* (2013)
Turkmenia	Turkmen	NT	13.90	0.00	Khusainova *et al.* (2013)
Kazakhstan	Kazakhs	NT	7.80	NT	Khusainova *et al.* (2013)

Table 2.2: Cont.

Population	Ethnicity	Allele frequencies (%)			References
		C282Y	H63D	S65D	
Uzbekistan	Uzbeks	NT	5.60	NT	Khusainova *et al.* (2013)
Tajikistan	Tajiks	NT	2.40	0.00	Khusainova *et al.* (2013)
Israel	Ashkenazi Jews	1.62	14.08	NT	Reish *et al.* (2010)
Central Asia	Uighurs	0.00	8.20	0.00	Khusainova *et al.* (2013)
USA	Caucasians	6.30	15.20	1.60	Beutler *et al.* (2000)
USA	Hispanics	2.70	12.40	0.60	Beutler *et al.* (2000)
USA	Asian	0.20	3.30	0.00	Beutler *et al.* (2000)
USA	Blacks	1.10	5.10	0.70	Beutler *et al.* (2000)
Brazil	Espiritos-santense	1.67	11.67	0.83	Alves *et al.* (2016)
Brazil	Brazilians	2.10	13.60	0.60	Santos *et al.* (2010)
Ecuador	Ecuadorians	0.00	3.50	4.00	Leone *et al.* (2005)
Mexico	Mexican	0.20	11.50	NT	Avila-Gomez *et al.* (2008)
Mexico	Mexicans	0.00	6.50	NT	Merryweather-Clarke *et al.* (1997)
Venezuela	Venezuelan	1.90	11.90	0.90	Avila-Gomez *et al.* (2008)
Chile	Chilean	1.28	10.58	NT	Avila-Gomez *et al.* (2008)
Columbia	Columbians	0.00	0.00	NT	Merryweather-Clarke *et al.* (1997)
Jamaica	Jamaicans	1.10	2.20	NT	Merryweather-Clarke *et al.* (1997)
Algeria	Mozabites	0.00	8.90	NT	Roth *et al.* (1997)
Ethiopia	Ethiopians	0.00	9.40	NT	Roth *et al.* (1997)
Tunisia	Tunisians	0.50	17.50	NT	Zorai *et al.* (2003)

Table 2.2: Cont.

Population	Ethnicity	Allele frequencies (%)			References
		C282Y	H63D	S65D	
Gambia	Gambians	0.00	1.30	NT	Merryweather-Clarke *et al.* (1997)
Ghana	Ghanaians	0.00	1.00	NT	Jeffery *et al.* (1999)
Kenya	Kenyans	0.00	1.30	NT	Merryweather-Clarke *et al.* (1997)
Nigeria	Nigerians	0.00	1.90	NT	Merryweather-Clarke *et al.* (1997)
Senegal	Senegalese	0.00	0.00	NT	Merryweather-Clarke *et al.* (1997)
South Africa	South Africans	0.25	0.00	NT	de Villiers *et al*. (1999)
Zambia	Zambians	0.00	0.70	NT	Merryweather-Clarke *et al.* (1997)
Australia	Aborigines	0.00	0.00	NT	Merryweather-Clarke *et al.* (1997)
Australia	Vanuatuans	0.00	0.60	NT	Merryweather-Clarke *et al.* (1997)
Sweden	Finns, Swedes, Swedish Saamis	2.00	7.90	3.00	Beckman *et al.* (2001)
Sweden	Swedish	6.10	12.40	1.60	Milman *et al.* (2005)
Turkey	Turks	0.00	13.57	NT	Merryweather-Clarke *et al.* (1997)
Spain	Catalans	3.00	20.00	1.00	Altes *et al.* (2004)
Spain	Basque	3.60	30.40	NT	Merryweather-Clarke *et al.* (1997)
Spain	Murcia	0.67	27.02	2.02	Avila-Gomez *et al.* (2008)
Ireland	Irish	14.00	17.90	NT	Ryan *et al.* (1998)

Table 2.2: Cont.

Population	Ethnicity	Allele frequencies (%)			References
		C282Y	H63D	S65D	
Hungary	Hungarian	3.40	14.40	NT	Andrikovics *et al.* (2001)
Bosnia and Herzegovina	Bosnian and Herzegovinian	2.00	7.90	3.00	Terzić *et al.* (2006)
Czech Republic	Czechs	3.40	16.20	1.25	Čimburová *et al.* (2005)
Italy	Apulians	1.50	14.00	0.50	Pietrapertosa *et al.* (2003)
Italy	Italians	3.20	13.40	1.30	Mariani *et al.* (2003)
Italy	Modena	4.71	14.93	0.74	Salvioni *et al.* (2003)
Italy	Ossola	1.60	13.30	NT	Cassanelli *et al.* (2001)
Russia	Russians	3.70	13.30	1.70	Mikhailova *et al.* (2003)
Serbia and Montenegro	Serbian and Montenegrins	1.60	15.70	1.60	Šarić *et al.* (2006)
Croatia	Croatians	3.30	14.50	1.80	Ristić *et al.* (2003)
Slovenia	Slovenians	4.00	14.50	0.50	Ristić *et al.* (2003)
Slovenia	Slovenian blood donors	3.60	12.80	1.80	Cukjati *et al.* (2007)
Portugal	Madeirenses	0.33	20.50	1.00	Spínola *et al.* (2011)
Portugal	Portuguese	3.00	23.00	NT	Porto *et al.* (1998)
Germany	Germans	3.80	13.20	NT	Merryweather-Clarke *et al.* (1997)
Germany	German blood donors	4.60	10.80	NT	Raddatz *et al.* (2003)
Bulgaria	Bulgarians	0.00	23.00	NT	Merryweather-Clarke *et al.* (1997)

Table 2.2: Cont.

Population	Ethnicity	Allele frequencies (%)			References
		C282Y	H63D	S65D	
France	North Africans	1.07	12.23	NT	Aguilar-Martinez *et al.* (2001)
France	French	7.70	14.00	1.95	Mura *et al.* (1999)
Finland	Finns	4.60	9.80	2.30	Beckman *et al.* (2001)
Poland	Polish	3.10	16.20	NT	Moczulski *et al.* (2001)
Greece	Greeks	1.30	13.50	NT	Merryweather-Clarke *et al.* (1997)
Austria	Austrians	3.70	12.90	NT	Merryweather-Clarke *et al.* (1997)
Faroe Islands	Faroese	8.00	17.50	1.00	Milman *et al.* (2005)
UK	British	8.10	15.20	NT	Merryweather-Clarke *et al.* (1997)

2.4 Malay population in Peninsular Malaysia

Malaysia is situated in Southeast Asia (SEA) and has thirteen states and three federal territories. The country is divided by the South China Sea into two regions (West Malaysia and East Malaysia) (Figure 2.4). West Malaysia (aka Peninsular Malaysia) has eleven states, and two federal territories while East Malaysia (aka Borneo) has two states and one federal territory (Mohd-Zaki *et al.*, 2014). Peninsular Malaysia has a land size of 132,090 km^2 which is approximately 40% while Borneo has a land size of 198,847 km^2 which is about 60% of the land area (Almayahi *et al.*, 2013).

Malaysia is a multi-ethnic nation that is inhabited by Malays, Chinese, Indians, Orang Asli and other minorities. The Orang Asli groups are further broken down into three subgroups such as Semang (Negrito), Senoi and Proto-Malays. Orang Asli, Malays and other indigenous people are called 'Bumiputra' (son of the soil). They are also grouped based on the languages of Aslian (Semang and Senoi) or Malayic (Malays and Proto-Malays) (Adelaar, 1993; Benjamin, 2012; Loo and Gan, 2014).

Earlier studies identified the three classes of Orang Asli who migrated to Peninsular Malaysia in separate prehistoric times. Semang was the first Orang Asli group to arrive in Peninsular Malaysia some 50,000 years ago, and their ancestry forms the Aborigines in Australia, the Philippines and Papua New Guinea (Norhalifah *et al.*, 2016; Oppenheimer, 2011). Semang people currently speak the Aslian language originating from the Senoi people who migrated from Indochina roughly 7,000 years ago to Peninsular Malaysia (Hill *et al.*, 2006; NurWaliyuddin *et al.*, 2015). Moreover, Proto-Malays, the third Orang Asli group, speaks a branch of Malayo-Polynesian subgroups of Austronesian languages called Malayic. Proto-Malays are thought to have migrated from Southern China about 5,000 years ago via Taiwan, the Philippines,

Indonesia, and Borneo and eventually reached Peninsular Malaysia around 3,000 years ago (Bellwood, 2007; Bellwood *et al.*, 2011; Diamond, 2000).

The present-day Malays known as Deutero-Malays in Peninsular Malaysia are descendants of some Proto-Malays admixed with Chinese, Indian and indigenous groups (Norhalifah *et al.*, 2016; Yahya *et al.*, 2017).

Figure 2.4: Map of Malaysia showing the thirteen states and three federal territories. Figure modified from https://www.mapsofworld.com/malaysia/malaysia-political-map.html.

CHAPTER 3

METHODOLOGY

3.1 Study design

This study analysed HFE SNP data on 35 Malay individuals. Malay populations have been the focus group of the study as they constitute the majority of the population (54.6%) in Malaysia along with Chinese (24.6%) and Indians (7.3%) as recorded in 2010 Population and Housing Census of Malaysia (Norhalifah *et al.*, 2016). A sample size of 35 has been targetted as it is within the agreed size of at least 30 needed to predict 0.05 of allele frequencies within a probability of 95 per cent in population genetic studies (Fung and Keenan, 2014; Hale *et al.*, 2012). Genomic DNA samples from these individuals were typed for C282Y, H63D and S65C by Edinur Atan's research group at USM using PCR-RFLP methodology. The study was approved by the Human Ethical Committee of Universiti Sains Malaysia (USM/JEPeM/16050191) and the Malaysia Medical Research and Ethics Committee of the Ministry of Health (NMRR-16-1399-31311 (IIR) (Appendix A and B). Details of blood sample collection, DNA extraction, PCR-RFLP typing methodology and statistical analysis are described in the following sub-sections.

3.2 Sample collection

In 2018, Norul Hajar Che Ghazali collected a total of 10 mL of blood samples from each of the Malay volunteers and kept in ethylenediaminetetraacetic acid (EDTA) tube. The volunteers were blood donors at three general hospitals in Peninsular Malaysia, namely Hospital Universiti Sains Malaysia (Kelantan), Hospital Seberang Jaya (Pulau Pinang), and Hospital Temerloh (Pahang). The following inclusion and exclusion criteria were applied for sample selection:

3.2.1 Inclusion criteria

1. Unrelated healthy Malay blood donors.
2. Adult (18 years and above).
3. Male or female.
4. Donors who voluntarily consent to participate and gave their signed written consent form.
5. Donors with no history of intermarriage for at least three generations.

3.2.2 Exclusion criteria

1. Donors with a history of diseases.
2. Donors who defer from blood donation.
3. Non-Malays.

3.3 DNA extraction from whole blood

Genomic DNA was extracted from whole blood using a commercially available DNA extraction kit (Invisorb® Spin Forensic Kit) as previously described by Hajar *et al.* (Hajar *et al.,* 2020). A minimum of 50 µL of whole blood was transferred into a 1.5 mL reaction tube and added with 10 µL and 50 µL of Proteinase K and Lysis Buffer M, respectively. The mixture was then incubated for 5 minutes at 56°C. Next, 100 µL of binding buffer B6 was added to solution and subjected to pulse vortexing and centrifugation.

Lysate was then transferred to the RTA spin filter membrane in a labelled 2 mL RTA collection tube and centrifuged for 2 minutes at 11,000 x *g*. A total of 300 µL wash buffer I was then added and again centrifuged for 1 minute at 11,000 x *g*.

Wash buffer II (750 µL) was then added to the RTA spin filter and centrifuged consecutively for 1 minute at 11,000 x *g* and for 4 minutes at full speed (18,000 x *g*) to remove both, wash buffer II and residual ethanol. The RTA spin filter was then transferred into a new 1.5 mL RTA receiver tube and added with 100 µL pre-warmed (56°C) elution buffer. This was followed by incubation for 1 minute at room temperature and centrifugation for 1 minute at 11,000 x *g*. RTA spin filter was then discarded and the extracted DNA was stored at -20°C until used.

3.3.1 Agarose gel electrophoresis of extracted genomic DNA

The extracted genomic DNA samples were fractionated using a 2% agarose gel electrophoresis. The agarose gel was prepared by mixing 4 g of agarose powder with 200 mL of 0.5X TBE buffer in a 250 mL beaker. The solution mixture was then heated for 4 minutes in a microwave oven until the agarose powder fully dissolved.

Agarose solution was then mixed with ethidium bromide (13 MgCl2 /mL) (Promega, USA) and poured onto the gel casting tray with a well-comb in place. The well-comb was removed after agarose gel solidified at room temperature. The solidified gel was placed in an electrophoresis tray and filled with 0.5X TBE buffer until it wholly submerged in the buffer solution. A mixture of 5 µL extracted genomic DNA and 1 µL DNA loading dye was loaded into the respective wells. The first well was loaded with the reference standard (3 µL of 100 bp DNA ladder) and electrophoresed for 45 minutes at 100V, and the presence of high molecular weight DNA was observed under UV light through a gel documentation system (Vilber Lourmat, Germany).

3.3.2 DNA quantitation

The Nanodrop™ 1000 spectrophotometer (Thermo Scientific, USA) was used to determine the concentration and purity of each extracted genomic DNA sample. The sampling arm and lower pedestal were first cleaned using 2 µL of distilled water (dH_2O) and wiped using Kimwipes. Then, 2 µL of elution buffer was loaded and measured as blank (0.00ng/µL). The same protocol was followed using 1 µL of DNA samples to measure DNA concentration and purity and was recorded. The absorbance of the DNA samples was calculated based on the reading of A260nm/A280nm ratio. High purity DNA samples are those within the acceptable range of 1.7 – 2.0Au.

3.4 HFE typing using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP)

HFE typing was done by using the PCR-RFLP technique. The PCR reaction mixture of 25 µL was prepared for each sample. Each of the reaction mixtures consisted of 2 µL of DNA sample with a concentration of 75 ng/µL to 125 ng/µL. Then continued with PCR amplification using thermal cycler machine. Furthermore, each of PCR products was digested with restriction enzymes. The details of genotyping for C282Y, H63D and S65C were described in the following subsection:

3.4.1 Genotyping for C282Y

Each PCR reaction mixture for amplification of a region containing C282Y SNP consists of 2 µL of DNA sample, 12.5 µL of Taq 2X Master Mix (BioLabs®, USA), 0.5 µL (0.5 µM) of forward (5′-CAA GTG CCT CCT TTG GTG AAG GTG

ACA CAT-3′) and reverse (5′-CTC AGG CAC TCC TCT CAA CC-3′) primers (1st BASE, Singapore) and 11.5 µL dH_2O.

The PCR-9700 thermal cycler machine (Applied Biosystems, USA) was used for amplification using the following thermal cycling parameters; initial denaturation at 94°C for 2 minutes, 10 cycles of denaturation at 94°C for 15 s and annealing at 65°C for 1 minute and 20 cycles of denaturation at 94°C for 15 s, annealing at 61°C for 50 s and extension at 72°C for 30 s.

The amplified PCR products were digested with *RsaI* enzymes (BioLabs®, USA) and incubated at 37°C for 15 minutes, followed by agarose gel electrophoresis. Amplification products containing C282Y SNP will produce 3 restriction fragments (203, 111, and 29 bp) while the wild type produced two restriction fragments (203 and 140 bp). Those restriction fragments were visualised under ultraviolet (UV) light using a gel documentation system (Vilber Lourmat, Germany).

3.4.2 Genotyping for H63D

Each PCR reaction mixture for amplification of a region containing H63D SNP consist of 2 µL of DNA sample, 12.5 µL of Taq 2X Master Mix (BioLabs®, USA), 0.5 µL (0.5 µM) of forward (5′-ACA TGG TTA AGG CCT GTT GC-3′) and reverse (5′-CTT GCT GTG GTT GTG ATT TTC C-3′) primers (1st BASE, Singapore) and 11.5 µL dH_2O.

PCR cycling parameters as described earlier (sub-section 3.4.1). The amplified PCR products were digested with *MboI* enzymes (BioLabs®, USA) and incubated at 37°C for 15 minutes, preceded by agarose gel electrophoresis. Amplification products containing H63D SNP will produce two restriction fragments (237 and 57 bp) while