GENETIC DIVERSITY OF BLOOD GROUP ANTIGEN, HNA, HPA, HFE, CYTOKINE AND HLA-G VARIANTS AMONG MALAYS, CHINESE AND INDIANS IN PENINSULAR MALAYSIA

NORUL HAJAR BINTI CHE GHAZALI

UNIVERSITI SAINS MALAYSIA

2025

GENETIC DIVERSITY OF BLOOD GROUP ANTIGEN, HNA, HPA, HFE, CYTOKINE AND HLA-G VARIANTS AMONG MALAYS, CHINESE AND INDIANS IN PENINSULAR MALAYSIA

by

NORUL HAJAR BINTI CHE GHAZALI

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

March 2025

ACKNOWLEDGEMENT

I am deeply grateful and praise **Allah**, the almighty for providing me this opportunity and granting me the capability to proceed successfully. This thesis would not have been possible without the guidance and the help of the several individuals in one way or another contributed and extended their valuable assistance in the preparation and completion of this study, it is a pleasure to thank those who made it a possibility. First and foremost, I am extremely grateful to my main supervisors, Assoc. Prof. Dr. Edinur Hisham Bin Atan for his invaluable advice, continuous support, and patience during my PhD study. His immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. My sincere thanks also goes to Dr. Nor Fazilah Che Mat, my co-supervisor for their help and consultations. I would also like to thank to Dr. Zefarina Zulkifli team members from blood bank as well as Dr. Geoffrey K. Chambers for their technical support during my sample collections and manuscripts preparations. Next, I would like to thank all my fellow lab mates in the School of Health Sciences, especially Siti Nor Assyuhada and Dr. Edward Kofi Abban. It is their kind help and support that have made my study and life in the USM a wonderful time. Not forgotten to my best friend Dr. Saffanah for her expertise in English writing and thesis formatting. Finally, I would like to express my gratitude to my parents, Che Ghazali Bin Che Kah and Sumiyati Binti Hussain, my partner Mohd Khairul Naim Bin Hashim and all my siblings, for their prayers, emotional and financial support. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study.

Thank you.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

α	Alpha
β	Beta
γ	Gamma
ε	Epsilon
λ	Lambda
ω	Omega
~	approximately
%	Percentage
§	No further test was performed
µL	Microliter
^{0}C	Degree celsius
χ^2	Chi square
<	Less than
Σ	Sum of tested genotypes

LIST OF ABBREVIATIONS

A	Adenine – abbreviations
Ach	Acheh
AE	Elution buffer
AL	Lysis buffer
AW	Wash buffer
Bg	Bugis
Bj	Banjar
Bk	Batek
bp	Base pair
C	Cytosine
Ch	Champa
Cw	Che Wong
DNA	Deoxyribonucleic acid
E	Expected value
EDTA	Ethylene diamine tetra-acetic acid
EM	Expectation maximization
EtOH	Ethanol
G	Guanine
g	Gram
gp	Glycoprotein
HDFN	haemolytic diseases of the fetus and newborn
HWP	Hardy-Weinberg proportion
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
ISEA	Island of Southeast Asia
Jw	Jawa
kDa	Kilo dalton
Kdh	Kedah
Kq	Orang Kanaq
Ks	Kensiu

Ktn	Kelantan
kya	Kilo years ago
Lh	Lanoh
LPS	Lipopolysaccharide
Md	Mandailing
mg	Milligram
MHC	Major histocompatibility complex
mL	Milliliter
Mng	Minangkabau
na	Not applicable
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
N-J	Neighbor-Joining
NK	Natural killer
O	Observed value
p	Significant value
PCA	Principal component
PCO	Principal coordinate
PCR-SSP	Polymerase chain reaction-sequence specific primer
Pt	Patani
PM	Peninsular Malaysia
PyPop	Python for Population Genetic
rpm	Rotation per minute
Sm	Semai
SNPs	Single nucleotide polymorphisms
STAT6	Signal transducer and activator of transcription 6
T	Thymine
Taq	Thermus aquaticus
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
TGF	Transforming growth factor
Th	T helper
TNF	Tumor necrosis factor
Treg	T regulatory
V	Volt
ya	Years ago

LIST OF APPENDICES

KEPELBAGAIAN VARIASI GENETIK ANTIGEN KUMPULAN DARAH, HNA, HPA, HFE, SITOKIN DAN HLA-G DALAM KALANGAN MELAYU, CINA DAN INDIA DI SEMENANJUNG MALAYSIA

ABSTRAK

Variasi gentik dalam *kumpulan darah*, *platelet antigen manusia (HPA), neutrofil antigen manusia (HNA)*, *sitokin*, *hemokromatosis (HFE)* dan antigen *leukosit manusia (HLA)* bertaburan secara unik di kalangan kumpulan populasi manusia. Sebahagian daripada gen-gen ini adalah penanda-penanda penting untuk keserasian tisu dan mempunyai peranan-peranan penting dalam kerentanan penyakit. Oleh itu, kajian ini dilakukan dengan tujuan untuk menilai polimorfisme kumpulan darah, HPA, HNA, sitokin, dan HLA-G dalam masyarakat Melayu, Cina dan India di Semenanjung Malaysia (PM) dan digunakan sebagai set data populasi bagi kajian keturunan dan kesihatan. Beberapa kumpulan darah (contohnya, *ABO* jenis *O*, *DCCee*, *MNS*, dan *Fya+b-),* HPA (contohnya, *HPA-1a/a, HPA-3a/b*, dan *-4a/a*), HNA (contohnya, *HNA-1a/a*, *HNA-3a/b*, dan *-4a/a*), dan HFE (contohnya, *H63D-S65C-C282Y*) secara umumnya adalah genotip yang paling biasa ditemui di kalangan Melayu, Cina dan India. Walau bagaimanapun, perbezaan yang jelas dalam taburan frekuensi genotip antara Melayu, Cina dan India telah direkodkan untuk sistem kumpulan darah lain (contohnya, Kidd dan Dombrock), polimorfisma nukleotida tunggal (SNPs) sitokin (contohnya, *IL-1Rpst11970C/T*) dan gen alel *HLA-G* (contohnya, *HLA-G*01:01:01:01/01:01:01:01* dan *HLA-G*01:01:03:01/01:01:01:01*). Ujian-ujian keseragaman yang dilakukan menunjukkan beberapa perbezaan yang signifikan antara Melayu, Cina, India dan set

data untuk populasi Malaysia lain yang dilaporkan sebelum ini. Perbezaan antara kumpulan populasi yang tidak berkaitan secara keturunan di Malaysia juga dapat dilihat dalam plot komponen utama/ koordinat utama dan pohon jejari leluhur. . Pemerhatian-pemerhatian ini berkaitan dengan asal-usul dan sejarah populasi mereka yang berbeza. Data-data genetik populasi yang dikumpulkan dari kajian ini dan kajian sebelumnya bukan sahaja memperkaya pandangan kontemporari mengenai kepelbagaian genetik di PM, tetapi juga menawarkan faedah potensi yang signifikan bagi pengamal kesihatan, penyelidik-penyelidik dan pembuat dasar kesihatan. Di PM, risiko aloimunisasi disebabkan oleh ketidakserasian transfusi dan kehamilan secara berpotensi dikaitkan dengan antigen-antigen ABO, MNS, Kidd, Rhesus, Kell, Duffy, HPA-3, HPA-15, HNA-1, HNA-3, dan HNA-15. Oleh itu, penaipan HPA, HNA dan kumpulan darah selain keatas ABO dan RhD boleh dilaksanakan untuk mengurangkan risiko potensi aloimunisasi diakibatkan oleh transfusi dan kehamilan di PM. Beberapa SNPs untuk gen-gen sitokin yang telah dikenalpasti sebagai faktor risiko untuk kecenderungan kepada penyakit dan didapati bertaburan berbeza di antara kumpulan populasi di PM. Ini termasuk untuk *IL1RAmspa111100C/T* dan *IL-1Rpst11970T/C* yang mana boleh menyumbang kepada kejadian penyakit yang berbeza di antara kumpulan populasi. Walaubagaimanapun, variasi-variasi patogenik lain dalam *HFE* (contohnya, *C282Y* dan *HLA-G* (*HLA-G*01:01:01:01*, *01:05N* dan *G*01:06*) yang tidak dikesan di kalangan Melayu, Cina dan India. Kesimpulannya, kajian ini menyediakan set-set data populasi yang luas dan komprehensif untuk orang Melayu, Cina dan India di. Penemuan-penemuan ini menyumbang kepada pemahaman yang lebih baik tentang kepelbagaian genetik dan profil-profil risiko kesihatan di PM serta menawarkan data berharga untuk analisis keturunan dan kesihatan pada masa depan.

GENETIC DIVERSITY OF BLOOD GROUP ANTIGEN, HNA, HPA, HFE, CYTOKINE AND HLA-G VARIANTS AMONG MALAYS, CHINESE AND INDIANS IN PENINSULAR MALAYSIA

ABSTRACT

Genetic variations in *blood group*, *human platelet antigen (HPA)*, *human neutrophil antigen (HNA)*, *cytokine, hemochromatosis (HFE)* and *human leukocyte antigen (HLA)* are uniquely distributed across human population groups. Several of these are important markers for tissue compatibility and diseases susceptibility. Thus, the present study was conducted with the aim of assessing blood group antigens, HPA, HNA, cytokine and HLA-G polymorphisms in Malays, Chinese and Indians in Peninsular Malaysia (PM) and use as population datasets reference for ancestry and health assessments. Several blood group (e.g., *ABO* type *O*, *DCCee*, *MNS* and *Fya+b-*), *HPA* (e.g., *HPA-1a/a*, *-3a/b* and *-4a/a*), *HNA* (e.g., *HNA-1a/a*, *-3a/b* and *-4a/a*) and *HFE* (e.g., *H63D-S65C-C282Y*) were generally found to be the most common in Malays, Chinese and Indians. However, clear differences on genotype frequency distributions between Malays, Chinese and Indians were recorded for other minor blood groups systems (e.g., Kidd and Dombrock), cytokine single nucleotide polymorphisms (SNPs) (e.g., *IL-1Rpst11970C/T*) and *HLA-G* alleles (e.g., *G*01:01:01:01/01:01:01:01* and *G*01:01:03:01/01:01:01:01*). Homogeneity tests showed several significant differences between Malays, Chinese, Indians and datasets for other Malaysian populations that were reported earlier. Distinctions between ancestrally unrelated population groups in Malaysia can also be seen in the principal component/coordinate analyses plots and neighbor-joining trees.

These observations are associated with their different origins and population histories. The population genetic datasets collected from the present and earlier studies not only enrich the contemporary view of genetic diversity in PM, but also offer significant potential benefits for health practitioners, researchers and policy makers. In PM, alloimmunization risks due to transfusion and gestation incompatibility are potentially associated with ABO, MNS, Kidd, Rhesus, Kell, Duffy, HPA-3, HPA-15, HNA-1, HNA-3 and HNA-15 antigens. Therefore, HPA, HNA and together with extended blood group typing (i.e., beyond the regular screening of ABO and RhD) typing could be implemented to reduce potential risks of transfusion and gestation alloimmunization in PM. Several cytokine gene SNPs have been identified as risk factors for disease susceptibility and were observed to distribute differently between population groups in PM. These include for *IL1RA mspa1 11100C/T* and *IL-1R pst1 1970T/C* which could contribute to different disease incidence between population groups. However, other pathogenic variants within *HFE* (e.g., C282Y) and *HLA-G* (*G*01:01:01:01, 01:05N* and *G*01:06*) were not detected either in Malays, Chinese and Indians. In conclusion, this study provides extensive and comprehensive population datasets for the Malays, Chinese and Indians. These findings contribute to a better understanding of genetic diversity and health risk profiles in PM as well as offering value data for future ancestral and disease analyses.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Malaysia is a country located in the heart of Southeast Asia. The country is divided into two noncontiguous lands and separated by China South Sea; the Peninsular Malaysia and East Malaysia (Figure 1.1). Peninsular Malaysia also known as West Malaysia consists of 11 states and is neighbouring with Thailand on the north and Singapore to the south (Metin, 2022). The other two states in Malaysia (Sabah and Sarawak) are parts of East Malaysia and sharing a land border with Indonesia (Kalimantan) and Brunei (Thiessen, 2012).

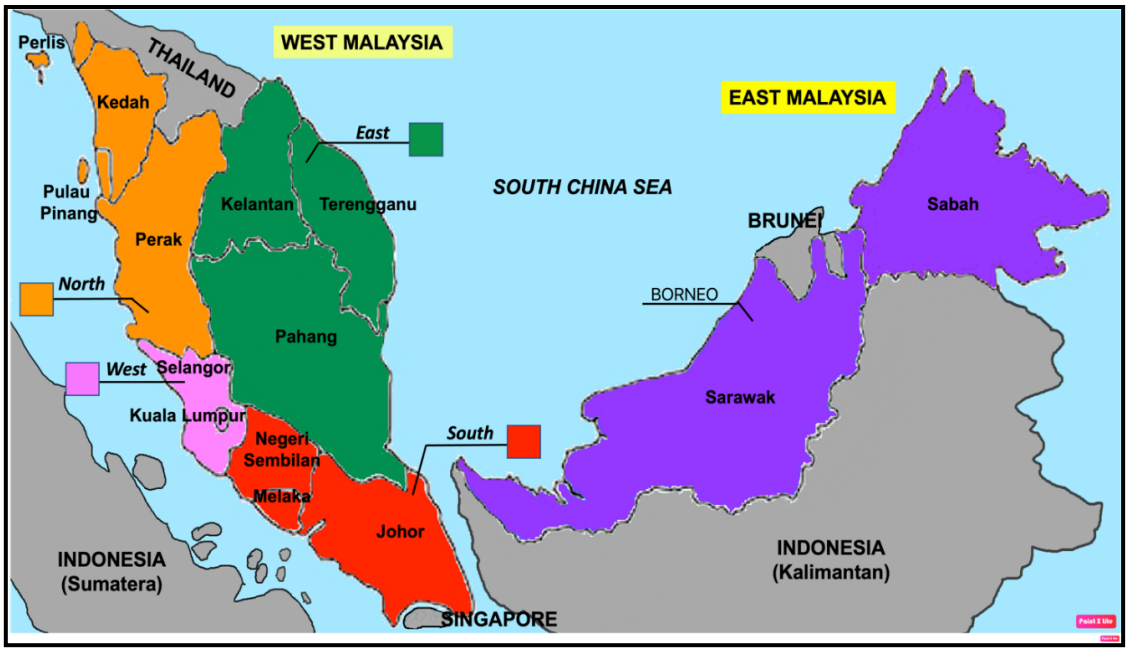

Figure 1.1 Map showing southeast Asian country of Malaysia with two major regions of Malaysia, Peninsular Malaysia (West Malaysia) and East Malaysia. This map was modified from Federal Department of Town and Country Planning, Malaysia, (2010)

The population groups in Malaysia are linguistically, culturally and religiously diverse. However, the whole Malaysian population is generally divided into Bumiputera (i.e., Orang Asal or original people) and non-Bumiputera (Cheu, 2020; Nagaraj *et al.,* 2015) for administrative purposes (Table 1.1). The Bumiputera for example includes three Orang Asli groups (i.e., Semang, Senoi and Proto-Malays), Malays and other native groups of Sabah and Sarawak. Other populations including Chinese and Indians are grouped as non-Bumiputera. Both; Bumiputera and non-Bumiputera are composed of people from different origins and affiliated with different language and cultural groups (refer to sub-section 2.1 for details). This is a rather simple classification that does not consider past population histories of various population groups in Malaysia. The three Orang Asli, for instance, were shown to have unique genetic fractions based on genome-wide single nucleotide polymorphism (SNP) and whole genome sequence data (Deng *et al.,* 2014; Hatin *et al.,* 2014; Norhalifah *et al.,* 2016; Rophina *et al.,* 2023). These findings collaborated well with archaeology and linguistic findings, which show different origins and migration patterns of Orang Asli Semang, Senoi and Proto-Malays (Hill *et al.,* 2006; Ahmad et *al.,* 2021; Baer, 2014; Aghakhanian *et al.,* 2015).

Knowledge about various genetic fractions in Malaysia not only helps to elucidate population history but is of direct practical relevance for medical treatments and diagnosis. For instance, population genetic data from regions that determine human responses to infectious agents have helped in mapping disease loci (Zotz *et al.,* 1998; Ramachandra *et al.,* 2007; Liu *et al.,* 2021; Abegaz, 2021) and in designing donor registry for blood transfusion and organ transplant (Wen *et al.,* 2019; Karimi *et al.,* 2019; Jovanovic *et al.,* 2021; Stimac *et al.,* 2023).

In this study, well-characterised genetic samples from Malay, Chinese and Indian individuals were collected and systematically genotyped for *blood groups*, *human platelet antigen (HPA)*, *human neutrophil antigen* (*HNA*) *hemochromatosis* (*HFE*), *cytokine* and *human leukocyte antigen G* (*HLA-G*) to identify the presence of specific genetic variations of polymorphisms associated with these markers. Furthermore, these samples were examined using either polymerase chain reaction restriction-fragment length polymorphism (PCR-RFLP), PCR-sequence specific amplification (PCR-SSP) and Sanger sequencing. These newly collected population genetic datasets from the three ethnic groups in Peninsular Malaysia are anticipated to fill up the remaining gaps of our understanding of human and health in this part of the world.

Table 1.1 Bumiputera and non-Bumiputera in Malaysia

Bumiputera	Non-Bumiputera
Malays	**Chinese**
Negrito	Hokkien
Senoi	Khek (Hakka)
Proto Malay	Cantonese
Dusun	Teochew
Kadazan	Hainanese
Kwijau	Kwongsai
Bajau	Foochow/ Hokchiu
Iranun	Henghua
Murut	Hokchia
Rang Sungei	Other Chinese
Sulu/Suluk	**Indians**
Bisaya	Indian Tamil
Rungus	Malayali
Sino-native	Sikh/ Punjabi
Kadayan	Telegu
Tidong	Sri Lankan
Tambanuo	Tamil Singalese
Idahan	Bangladeshi
Dumpas	Pakistani
Mangkaak	Other Indian
Minokok	**Others**
Maragang	Indonesian
Paitan	Thai
Rumanau	Filipino
Lotud	Myanmar
Cocos Islander	Japanese
Iban/ Dayak Laut	Korean
Bidayuh/ Dayak Darat	Other Asian
Melanau	Eurasian
Kenyah	European
Lun Bawang/ Murut	Others
Penan	
Kajang	
Kelabit	

Data was obtained from Nagaraj *et al.,* (2015)

1.2 Problem statement

Peninsular Malaysia received multiple settlements by people of different genetic lineages from the last 50,000 years ago (Norhalifah *et al.,* 2016). Thus, all aspects of the highly diverse and complex genetic makeup in the country should be fully examined for ancestry and their potential applications in health. The latter include for blood transfusion and organ transplantation (Ruderman, 2023). Unlike Malay sub-ethnic groups and Orang Asli groups (Semang, Senoi and Proto-Malays;

Manaf *et al.,* 2016; Syafawati *et al.,* 2016; Norhalifah *et al.,* 2016), there are very few genetic studies that have been conducted on the general Malay population (i.e., Deutro-Malays), Chinese and Indians; refer Musa *et al.,* (2012) and Tan *et al.,* (2012) for blood group and HPA data. While these earlier surveys on Malays, Chinese and Indians provide preliminary information, further studies are needed to generate data for the previously unreported blood group and HPA loci. Nonetheless, no HNA, HFE, cytokine and HLA-G data for Malays, Chinese and Indians in Malaysia has ever been reported and will only make available for the first time via this study.

1.3 Significance of the study

Findings from earlier studies using genomic samples of Malay sub-ethnic groups and Orang Asli showed that these population groups have a whole different genetic repertoire, which reflects their unique origins. These include genetic data from genes that code for blood group antigen, HNAs, HPAs, HFEs, cytokines and HLAs (Abd Ghani *et al.,* 2015; Manaf *et al.,* 2015; Manaf *et al.,* 2016; Syafawati *et al.,* 2016; Wan Syafawati *et al.,* 2015; Norhalifah *et al.,* 2015; Norhalifah *et al.,* 2016). It is important to note that these genes are markers of tissue identity and can be used to determine disease susceptibility (Bakanay *et al.,* 2013, Muller *et al.,* 2012; Eksteen *et al.,* 2017; Pelusi *et al.,* 2016; Lan *et al.,* 2021; Uboldi *et al.,* 2004; Alelign *et al.,* 2018; Gilbert *et al.,* 1998; Gould & Auchincloss 1999; Nowak, 2008). Therefore, the blood group, HNA, HPA, HFE, cytokine gene SNP and HLA-G population datasets generated in this study for Malays, Chinese and Indians provide valuable insight for ancestry analyses and serve as reference standards for future mapping of disease loci, disease association studies and for other medical related applications, including for designing blood and organ donation programmes.

1.4 Objectives of the study

The objectives of this study are:

To genotype genes code for blood group antigens, HNAs, HPAs, HFE variants, cytokines and HLA-G molecules in Malays, Chinese and Indians in Peninsular Malaysia.

Specific

i. To develop blood group, HNA, HPA, HFE, cytokine and HLA-G population datasets for Malays, Chinese and Indians.

ii. To determine the ancestral fractions of population groups in Peninsular Malaysia with the population datasets.

iii. To compare genetic relationships between population groups in Peninsular Malaysia using the blood group, HNA, HPA, HFE, cytokine and HLA-G population datasets.

iv. To evaluate blood group, HNA, HPA, HFE, cytokine and HLA-G allelic diversity in Peninsular Malaysia and their implifcations for health.

1.5 Hypotheses

i. Population structure in Peninsular Malaysia is diverse as the region occupied by people of different ancestries.

ii. The population groups in Peninsular Malaysia exhibit distinct risk profiles for diseases linked to with blood group antigens, HNA, HPA, HFE, cytokine and HLAs.

CHAPTER 2

LITERATURE REVIEW

2.1 Population History in Peninsular Malaysia

Peninsular Malaysia is populated by people of different ancestries including Orang Asli, Chinese, Indians and Arabs (Norhalifah *et al.,* 2016). The Orang Asli was the earliest to settle in the region, as early as ~50 kya (Hill *et al.,* 2006; Deng *et al.,* 2021). There are 18 tribes of Orang Asli and they are all classified into three subgroups which are Semang or Negrito, Senoi and Proto-Malays (Aghakhanian *et al.,* 2015). These three Orang Asli groups have distinct features and migration histories as inferred using anthropological, social science and genetic studies; (Figure 2.1). The Semang represent descendants of the first migration 'Out of Africa' by modern human via coastal of India to Peninsular Malaysia 50-70 kya (Bellwood *et al.,* 1995; Macaulay, 2005). The Austro-Asiatic Senoi people migrated south from mainland Asia 5-7 kya while the Proto-Malays who speak Austronesians languages are believed to arrive in multiple series of migrations between 4-5 kya to various part of Asia Pasific including Malaysia (Norhalifah *et al.,*2016; Chambers and Edinur, 2013; Dentan, 1999; Hill *et al.,* 2006; Heiske *et al.,* 2021). Around 300-400 years ago, descendants of Austronesian people from nearby regions like Java, Kalimantan, Sulawesi and Sumatera such as Jawa, Bugis and Rawa moved to Peninsular Malaysia and since then become established as communities in Selangor, Johor and Perak (Fathil *et al*., 2018; Hill *et al*., 2006; Parthipan & Ishar, 2002; SharifahNanyRahayuKarmilla *et al*., 2017).

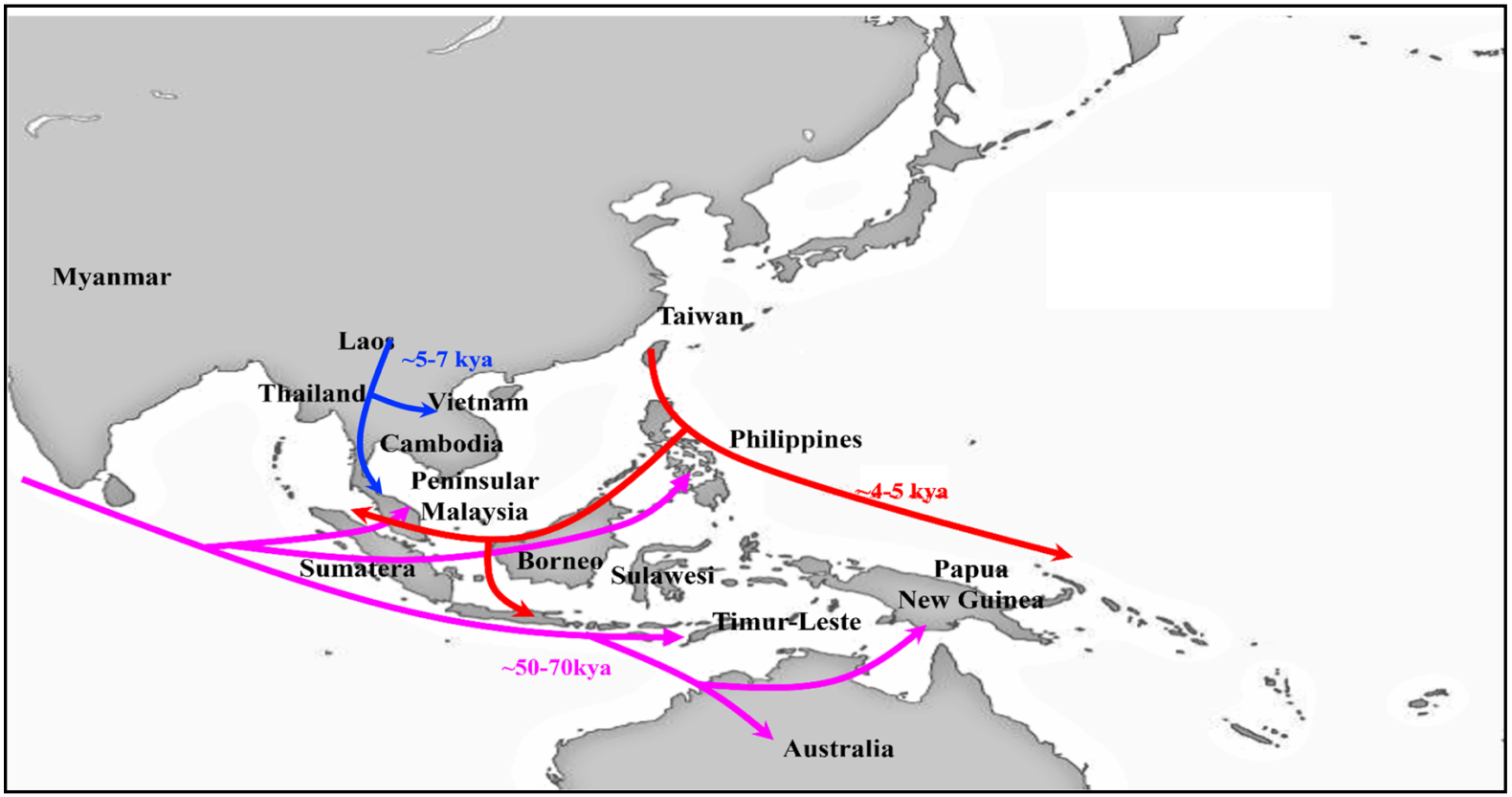

Figure 2.1 Migration history of the Orang Asli ancestors to Asia Pacific region. The migration pattern of Orang Asli Semang is shown by the pink arrow, blue for Orang Asli Senoi and red for Orang Asli Proto-Malays. This figure was adapted and modified from Norhalifah *et al.,* (2016)

Currently, the Semang, Senoi and Proto-Malays which make up less than 1% of the population in Peninsular Malaysia live alongside with three other major ethnic groups; the Malays, Chinese and Indians (Hajar *et al.,* 2020a; Hajar *et al.,* 2020b; Hajar *et al.,* 2020c). The Malays who represent 60% of the Malaysian population are admixed of Proto-Malays with Malay subethnic groups and other ethnicities (Hatin *et al.,* 2014; Yahya *et al.,* 2017). This admixture occurred due to intermarriages between local Malays and traders and immigrants from India, China, and Southeast Asia since the era of Malacca Empire and British colonization (Andaya, 2008; Wain, 2012; Wey & Harun, 2018).

The arrival of Chinese immigrants to Peninsular Malaysia has a significant impact on the Malaysian's demographic and cultural evolution (Kim, 1998). Initials, Chinese migrated to the Malay Peninsula for economic purposes and engaged with local communities (Lockard, 2010). Commercial connections between China and Malay Peninsula (Malacca) were further enhanced during the Ming and Qing dynasties (1368-1912) and witnessed an increases number of Chinese migration settled in the region (Wade & Chin, 2018). Intermarriage between Chinese merchants who migrated to Malacca from the 15th century onward and indigenous women contributed to the formation of a distinct society known as the Peranakan Chinese, or the Baba and Nyonya community, all of which endure in present-day Malacca, Penang and Terengganu (Suryadinata & Iseas, 2022; Moore, 2022). In the middle of 19th century, significant influx of Chinese immigrants to Peninsular Malaysia occurred, particularly to the states of Perak, Selangor and Negeri Sembilan (Kim, 1998). This migration was largely driven by the discovery of tin deposits in these areas (Tarmiji *et al.,* 2019). Over time, Chinese immigrants and their descendants became an integral

part of Malaysian society. They brought with them their unique cultural practices, language and traditions.

Indians influences in Peninsular Malaysia began around AD 110, when certain parts of pre-colonial Malaysia like Kadaram (Kedah) were still part of the Greater India Kingdom (Sandhu & Kemial Singh, 2006; Raja Sulong *et al.,* 2019; Akhir *et al.,* 2022). Traces of Indian civilisation in Peninsular Malaysia can readily be seen at archaeological excavation sites in at Bujang Valley, located near the present day Merbok in Kedah (Murphy, 2017). Artefacts from this area were date back to more than 2,535 years ago (Balakrishnan, 2017; Nadarajan, 2011. The Indian influence continued during the Malacca Empire (15^{th}-16^{th} centuries), a powerful maritime kingdom that controlled the strategic port in island of Southeast Asia (Kim, 1992) Hall, 2004). The Malacca sultanate maintained diplomatic and trade relations with various Indian states, including the Chola and Vijayanagara empires. Notably, Tamil Muslim traders exchanged Indian textiles and rice for tin, pepper and aromatic woods. They were appointed as royal merchants by Malay rulers and were highly integrated into the society of Malacca (Brown, 1994). During British era in Peninsular Malaysia, many Indian immigrants, mainly Tamils came to Peninsular Malaysia, as labours in plantations and tin mines (Selvaratnam, 2021; Vadlamudi, 2016; Kamble, 2007; Watson, 2012).

2.2 Genetic markers

Genetic markers are variable regions in human genome. They serve as powerful tools in human identification, population study and health inferences (Wang *et al.,* 2021; Heiske *et al.,* 2021). In the biomedicine, genetic markers can be used to assess the risk disease development and determine tissue compatibility between

donors and recipients for transplantation and blood transfusion (Yuan *et al.,* 2014; Saleh *et al.,* 2018; Brown and Navarrete, 2011; Thye *et al.,* 2010; Edinur *et al.,* 2016). In forensic science and crime investigations, highly variable markers such as short tandem repeats (STRs) are routinely used for identification due to their high power of discrimination to differentiate people, even between individuals whoe are closesly related (Rapley & Whitehouse, 2007; Wyner *et al.,* 2020; Hakim *et al.,* 2020). Variability in human genome is also used for population study as people of similar ancestry are shown to share a common genetic fraction (Slack, 2023; Bergström *et al.,* 2021; King & Motulsky, 2002). Therefore, many genetic markers have been screened in various population groups including for gene codes in human platelet antigens (HPAs), human neutrophil antigens (HNAs), cytokines, human homeostatic iron regulator **(**HFE) protein and human leukocyte antigens (HLAs) refer Abd Gani *et al.,* 2015; Musa *et al.,* 2015; Thakral *et al.,* 2010; Xia *et al.,* 2011; Azizi *et al.,* 2021; Manaf *et al.,* 2015; Manaf *et al.,* 2016; Syafawati *et al.,* 2016; Tan *et al.,* 2012; Costeas *et al.,* 2003; Norhalifah *et al.,* 2016; Dhaliwal *et al.,* 2007; Edinur *et al.,* 2009; Tao *et al.,* 2020; Bodmer, 2015. The following sub-sections described the genetics of blood group antigens, HPAs, HNAs, cytokines, HFE and HLA-G molecules, population data repositories and diseases linked to these genetic markers.

2.2.1 Blood group system

There are more than 360 antigens were identified on the surface of red blood cells (RBCs). Most of these antigens were grouped into 45 blood group systems by International Society of Blood Transfusion (ISBT) while others such as Cost, li and Er is pending approval from ISBT (ISBT, 2021). In addition to ABO and Rhesus, other clinically relevant blood group systems such as Kell, Kidd, Duffy and MNS were also

shown to be polymorphic and clinically relevant (Hosoi, 2008). The important and relevant blood groups are described in the following sub-sections.

2.2.1(a) ABO

The molecular basis of ABO blood groups has been elucidated through gene cloning and sequencing in 1990 (Yamamoto, 2021). This resulted in the discovery *A* and *B* alleles of *ABO* gene located on chromosome 9q34.1-q34.2 (Yamamoto & Yamamoto, 2001). These *A* and *B* alleles code for *alpha 1,3-N-acetylgalactosaminyltransferase* and *alpha 1,3-D-galactosyltransferase* which catalyze the production of A and B antigens from a common precursor (H substance) expressed by the gene on chromosome 9 (Harmening and Firestone, 2005). A single deletion (261G) on the background of *ABO* gene produces *O* allele and resulted on no enzymatic changes on the H antigen. Therefore, those with O blood group have H antigens on the surface of their RBCs, instead of A or B or both A&B antigens for those with A, B, or AB phenotypes respectively (Yamamoto *et al.,* 1990Rhesus

The Rh blood group system is determined by the presence or absence of the Rhesus antigens (D, C, c, E, and e) on the surface of RBCs (Agre and Cartron, 1991). Individuals are identified as Rhesus positive when the D antigen is present on the surface of their RBCs, irrespective of C, c, E or e antigens. The D antigen is coded by the *RHD* gene while C, c, E and e antigens are expressed by the *RHCE* gene. Both genes are located on the chromosome 1 (Carton, 1994; Westhoff, 2007; Avent & Reid, 2000). Several polymorphisms in the *RHD* genes including deletion, gene hybridization and SNPs cause a complete absence, weak or partial expression of D antigen (Flegel, 2011; Cruz *et al.,* 2011). In contrast, C and c and E and e antigens are associated with single nucleotide polymorphisms in *RCHE* genes (Mouro *et al.,* 1993; Tanaka *et al.,* 2001).

2.2.1(b) Kell

The Kell blood group system is the third most common blood group to cause transfusion reactions after the Rh and ABO blood group systems (Marsh & Redman, 1990; Rath *et al.,* 2014; Lee, 1997). Antigens of Kell blood group system are encoded by the 19 exons of *KEL* gene located on the short arm of chromosome 7. The *KEL* gene determines the specifications of more than 35 Kell antigens, which include the highly immunogenic K and k (cellano) antigens. The K and k alleles differ by single nucleotide changes (T to C) at position 578 (Stephen *et al.,* 2012; Lee *et al.,* 1995). The K-k+ phenotype is relatively more common in Europeans and Blacks then other races, which are 98% and 91%, respectively (Dean, 2005).

2.2.1(c) Kidd

The Kidd blood group system was discovered in 1951 by Allen, Diamond and team (Allen *et al.,* 1951). It consists of two main antigens (Jk^a and Jk^b) which differs by 838G→A on exon 9 (Lawicki *et al.,* 2017; Chu *et al.,* 2017). The changes result in the coding of aspartic for *JKA* allele and asparagine for *JKB* allele (Hamilton 2019; Lawicki *et al.,* 2017). The Jk_{null} phenotype is extremely rare and characterised by the absence of both Jk^a and Jk^b antigens. This phenotype is associated with mutation on the 3'-acceptor splice site of intron 5, SNPs (342G>A, 230G>A) and deletion at nucleotide positions 647 and 648 (Ekman & Hessner, 2000; Zhang *et al.,* 2023).

2.2.1(d) Duffy

The Duffy blood group system is encoded by the *antigen chemokine receptor 1* (*ACKR1*) gene located on chromosome 1q23.2 (Saha *et al.,* 2023). The Fy^a and Fy^b antigens are encoded by co-dominant *FY*A* (*FY*01*) and *FY*B* (*FY*02*) alleles which differ by 125G>A nucleotide substitution (Langhi Júnior *et al.,* 2022). Thus, individuals can be either Fy(a+b), Fy(a-b+) or Fy(a+b+). The other Duffy antigen is

encoded by the *FY*X* allele and is associated with weak Fy^b expression (Arndt *et al.,* 2015). The *FY*X* allele is similar to F^xB, except for a mutation at nucleotide 265 from cytosine to thymine (Langhi Júnior *et al.,* 2022; Oscar Pogo & Chaudhuri, 2000).

2.2.1(e) MNS

The MNS antigens are expressed on glycophorin A (GPA, CD235a), glycophorin B (GPB, CD235b), or hybrid glycophorin molecules (Storry *et al.,* 2022). The genes encoding for MNS antigens, *GYPA* and *GYPB*, are located on chromosome 4q28-q31.1 (Selorm *et al.,* 2020). There is high nucleotide sequence similarity between *GYPA* and *GYPB* genes which code for M and N and S and s antigens, respectively (Palacajornsuk, 2020; Willemetz *et al.,* 2015). The *M* and *N* alleles are differentiated by three SNPs (59C→T, 71G→A, and 72G→T), while the S and s alleles differ by a single 143 C>T nucleotide change (Dean, 2005).

2.2.1(f) Dombrock

The Do^a and Do^b antigens of Dombrock blood group system are coded by the *Do* gene on the long arm of chromosome 12 (Eiberg & Mohr, 1996). These antigens have been implicated in causing delayed hemolytic transfusion reactions (DHTRs) and classified as to 14th blood group system by the ISBT (Reid and Lomas-Francis, 2004; Lomas-Francis and Reid, 2020). Individuals may be classified as Do(a+b-), Do(a-b+), or Do(a+b+) based on the presence or absence of Do^a and Do^b antigens (Halawani *et al.,* 2022). Do^a and Do^b alleles are distinguished by the 378C>T, 624T>C and 793A>G nucleotide changes (Gubin *et al.,* 2000; Rios *et al.,* 2001).

2.2.1(g) Colton

Colton is the 15th blood group system classified by the ISBT and coded by the *aquaporin 1* gene (Sutton *et al.,* 2019). The *aquaporin* gene is located on chromosome

7 and a SNP at nucleotide position 45 from adenosine to guanine produced the two Colton blood group alleles, Co^a and Co^b respectively (Agre *et al.,* 1995; Leo *et al.,* 1997). The Co^a and Co^b alleles determine the three Colton phenotypes; Co(a+b-), Co(a+b+) and Co(a-b+). The null Colton phenotype has been reported to cause mild hemolytic disease of the fetus and newborn (HDFN) and this is due to a mutation at nucleotide position 601 that results in a frameshift and premature termination of the protein (Crottet, 2019; Joshi *et al.,* 2001).

2.2.1(h) Lutheran

The Lutheran blood group system is determined by the *LU* gene, also known as *basal cell adhesion molecule (BCAM*) gene (Daniels, 2020). The *BCAM* gene is located on chromosome 19q13.2 and encodes 18 Lutheran blood group antigens including Lu^a, Lu^b, Lu^8 and Lu^{14}. The Lu^a/Lu^b and Lu^8/Lu^{14} antigens are due to Arg77His and Met204Lys amino acid changes, respectively (Daniels, 2020). Other Lutheran antigens including Lu^5, Lu^{12}, Lu^{13}, Lu^{16} and Lu^{21} that are all caused by SNPs.

2.2.1(i) Cartwright

The Cartwright blood group system consists of two antigens, Yt^a and Yt^b. These antigens econded by the *acetylcholinesterase (ACHE)* gene located on chromosome 7q (George, 2020). The Yt^a and Yt^b alleles differ by the 1196A>C SNP that causes amino acid changes from histidine for Yt^a and asparagine for Yt^b(George, 2020). Unlike ABO and Rhesus, Cartwright blood group system is not routinely tested, either in donor or patient as alloimmunization against the Yt^a and Yt^b is rare (George, 2020).

2.2.1(j) Vel

The Vel blood group antigens are expressed on the surface of RBCs and encoded by *SMIM1* gene on chromosome 1p36 (Cvejic *et al.,* 2013; Storry, 2014). This gene produces small integral membrane protein 1 which carries Vel antigens. The Vel- is caused by a 17-bp deletion in exon 3 of *SMIM1* gene (c.64_80del) (Storry *et al.,* 2017; Marcia Regina Dezan *et al.,* 2019). Individual with Vel- antigen may develop severe acute haemolytic transfusion reactions when transfused with Vel+ blood and may also triggers HDFN (Daniels, 2013). The Vel- antigen is very rare among Europeans, Africans and Asians (Storry & Peyrard, 2019).

2.2.2 Human neutrophil antigen (HNA)

Human neutrophil antigens (HNAs) are a group of glycoproteins expressed on the surface of human neutrophil granulocytes (Gogri *et al.,* 2022). The International Society of Blood Transfusion Working Party on Granulocyte Immunobiology (ISBT GIWP) officially assigned 14 HNA alleles into HNA-1, -2, -3, -4 and -5 systems (Flesch & Reil, 2018). The allelic variations within HNA-1 to HNA-5 systems are determined by the *FCGR3B, CD177, SLC44A2, ITGAM* and *ITGAL* genes, respectively; Browne *et al.,* (2020); Bux (2008); Flesch *et al.,* (2016) and refer Table 2.1 for molecular bases of HNA systems. As shown in Table 2.1, SNPs are responsible for the observed variations in HNA-1, -3, -4 and -5 systems. However, the HNA-2a and -2b of HNA-2 system is associated with a defect during HNA-2 transcription (Storch *et al.,* 2014). HNA antigens play important roles in immune responses and including in inflammation and disease infectious (Browne et al, 2020).

Table 2.1 Molecular bases of the currently defined HNA systems

System	Antigens	Genes	Chromosome location	Variants/SNPs of interest		rs Number	Expression
				Glycoprotein/ Amino acid change	Nucleotide Change		
HNA-1	HNA-1a HNA-1b HNA-1c	*FCGR3B*	1	CD16b/Asp↔Asn_{82} Asn↔Ser_{65} Asp↔Ala_{78}	227A→G 147T→C 266A→C	396991 527909462 rs5030738	Neutrophils
HNA-2	HNA-2a HNA-2b	*CD177*	19	–	–	121917105	Neutrophils, monocytes
HNA-3	HNA-3a HNA-3b	*SLC44A2*	19	CTL2/ Arg↔Gln_{154}	461G→A	6061252	Neutrophils, granulocytes, monocytes, lymphocytes, platelets, kidney and placental, spleen, lymph node and endothelial cells
HNA-4	HNA-4a HNA-4b	*ITGAM*	19	CD11b.CD18/ Arg↔His_{61}	302A→G	1143679	Granulocytes, monocytes, T-lymphocytes
HNA-5	HNA-5a HNA-5b	*ITGAL*	16	CD11a.CD18/ Arg↔Thr_{766}	2466C→G	2230433	Granulocytes, monocytes, T-lymphocytes

Data obtained from Bux (1999); ISBT Granulocyte Antigen Working Party (1999); Fung & Minchinton (2011); Veldhuisen *et al.,* (2014); Flesch *et al.,* (2016). Abbreviation: CD=cluster of differentiation, CTL=choline transporter-like protein

2.2.3 Human platelet antigens (HPA)

Human platelet antigens (HPAs) are alloantigens expressed on the platelet membrane (Curtis & McFarland, 2013). These antigens play various functions including in hemostasis, inflammation and immune responses (Bode *et al.,* 2014). A total of 35 HPA systems have been identified and expressed on *GPIIb, GPIIIa, GPIa*, *GPIIa, GPIbα, GPIbβ, IX* and *CD109* genes; Table 2.2 and Figure 2.2. These HPA systems were approved and classified by the International Platelet Immunology Nomenclature Committee of the International Society of Blood Transfusion (ISBT). Each HPA system are bi allelic and assigned by either "*a*" or "*b*" allele (Newman, 1994). These *a* and *b* alleles are differentiated by SNPs, except for *HPA-14a* and *HPA-14b* of HPA-14 system which are due to three nucleotide deletion (AAG) at position 1909-1911 (Santoso *et al.,* 2002). Nonetheless, only HPA-1 to -6 and HPA-15 are polymorphic in many population groups while the rest are fixed, either for *a* or *b* allele (Rožman, 2002; Ghevaert *et al.,* 2009).

Table 2.2 Molecular information of the 35 recognized HPA systems

System	Antigens	Genes	Chromosome location	Variants/SNPs of interest		rs Number	Expression
				Glycoprotein / Amino acid change	Nucleotide changes		
HPA-1	**HPA-1a** **HPA-1b**	***ITBG3***	**17**	**GPIIIa** **Leu ↔ Pro_{33}**	**176T→C**	5918	**Platelets and neutrophils**
HPA-2	**HPA-2a** **HPA-2b**	***GPIBA***	**17**	**GPIb** **Thr ↔ Met_{145}**	**482C→T**	6065	**Platelets**
HPA-3	**HPA-3a** **HPA-3b**	***ITGA2B***	**17**	**GPIIb** **IIe ↔ Ser_{843}**	**2621T→G**	5911	**Platelets**
HPA-4	**HPA-4a** **HPA-4b**	***ITGB3***	**17**	**GPIIIa** **Arg ↔ Gln_{143}**	**506G→A**	5917	**Platelets and neutrophils**
HPA-5	**HPA-5a** **HPA-5b**	***ITGA2***	**5**	**GPIa** **Glu ↔ Lys_{505}**	**1600G→A**	1801106	**Platelets and T-lymphocyte**
HPA-6	HPA-6a HPA-6b	*ITGB3*	17	GPIIIa Arg ↔ Gln_{489}	1544G→A	13306487	Platelets and neutrophils
HPA-7	HPA-7a HPA-7b	*ITGB3*	17	GPIIIa Pro ↔Ala_{407}	1297C→G	12198448	Platelets and neutrophils
HPA-8	HPA-8a HPA-8b	*ITGB3*	17	GPIIIb Arg ↔Cys_{636}	1984C→T	51219882	Platelets
HPA-9	HPA-9a HPA-9b	*ITGA2B*	17	GPIIIa Val↔Met_{837}	2602G→A	4988902	Platelets and neutrophils
HPA-10	HPA-10a HPA-10b	*ITGB3*	17	GPIIIa Arg↔Gln_{62}	263G→A	200358667	Platelets and neutrophils
HPA-11	HPA-11a HPA-11b	*ITGB3*	17	GPIIIa Arg↔His_{33}	1976G→A	77302275	Platelets and neutrophils
HPA-12	HPA-12a HPA-12b	*ITGB3*	22	GPIbβ Gly↔Glu_{15}	119G→A	3752857	Platelets
HPA-13	HPA-13a HPA-13b	*ITGA2*	5	GPIIIa Thr↔ Met_{799}	2483C→T	79923242	Platelets and neutrophils

Table 2.2 Continued

System	Antigens	Genes	Chromosome location	Variants/SNPs of interest		RS Number	Expression
				Glycoprotein / Amino acid change	Nucleotide changes		
HPA-14	HPA-14a HPA-14b	*ITGB3*	17	GPIIIa DeltaLys$_{611}$	1909_1911del AAG		Platelets and neutrophils
HPA-15	**HPA-15a HPA-15b**	***CD109***	**8**	**CD109 Ser ↔ Tyr^{703}**	**2108A→C**	rs10455097	**Platelets, T-lymphocyte and monocytes**
HPA-16	HPA-16a HPA-16b	*ITGB3*	17	GPIIIa Thr↔IIe$_{140}$	497C→T	rs74708909	Platelets and neutrophils
HPA-17	HPA-17a HPA-17b	*ITGB3*	17	GPIIIa Thr ↔ Met$_{195}$	662C→T	rs770992614	Platelets and neutrophils
HPA-18	HPA-18a HPA-18b	*ITGA2*	5	GPIa Gln ↔ His$_{716}$	2235G→T	rs267606593	Platelets and T-lymphocyte
HPA-19	HPA-19a HPA-19b	*ITGB3*	17	GPIIIa Lys↔Gln$_{137}$	487A→C	rs80115510	Platelets and neutrophils
HPA-20	HPA-20a HPA-20b	*ITGA2B*	17	GPIIb Thr↔Met$_{619}$	1949C→T	rs78299130	Platelets
HPA-21	HPA-21a HPA-21b	*ITGB3*	17	GPIIIa Glu↔Lys$_{628}$	1960G→A	rs70940817	Platelets
HPA-22	HPA-22a HPA-22b	*ITGA2B*	17	GPIIb Lys↔Thr$_{164}$	584A→C	rs14281190	Platelets
HPA-23	HPA-23a HPA-23b	*ITGB3*	17	GPIIIa Arg↔Trp$_{622}$	1942C→T	rs139166528	Platelets
HPA-24	HPA-24a HPA-24b	*ITGA2B*	17	GPIIb Ser↔Asn$_{472}$	1508G→A	rs281864910	Platelets
HPA-25	HPA-25a HPA-25b	*ITGA2*	5	GPIa Thr↔Met$_{1087}$	3347C→T	rs771035051	Platelets and T-lymphocyte

Table 2.2 Continued

System	Antigens	Genes	Chromosome location	Variants/SNPs of interest		RS Number	Expression
				Glycoprotein / Amino acid change	Nucleotide changes		
HPA-26	HPA-26a HPA-26b	*ITGB3*	17	GPIIIa Lys↔Asn$_{580}$	1818G→T	1156382155	Platelets and neutrophils
HPA-27	HPA-27a HPA-27b	*ITGA2B*	17	GPIIb Leu↔Met$_{841}$	2614C→A	149468422	Platelets
HPA-28	HPA-28a HPA-28b	*ITGA2B*	17	GPIIb Val↔Ile$_{740}$	2614C→A	368953599	Platelets
HPA-29	HPA-29a HPA-29b	*ITGB3*	17	GPIIIa Thr↔Met$_{7}$	98C→T	544276308	Platelets and neutrophils
HPA-30	HPA-30a HPA-30b	*ITGA2B*	17	GPIIb Gln↔Gly$_{806}$	2511G→C	377753373	Platelets
HPA-31	HPA-31a HPA-31b	*GP9*	3	GPIX Pro↔Leu$_{107}$	368C→T	20229101	Platelets
HPA-32	HPA-32a HPA-32b	*ITGB3*	17	GPIIIa Asp↔Ser$_{148}$	521A→G	879083862	Platelets and neutrophils
HPA-33	HPA-33a HPA-33b	*ITGB3*	17	GPIIIa Asp↔Gln$_{432}$	1373A→G	155572829	Platelets and neutrophils
HPA-34	HPA-34a HPA-34b	*ITGB3*	17	GPIIIa Arg↔Trp$_{91}$	349C→T	77748046	Platelets and neutrophils
HPA-35	HPA-35a HPA-35b	*ITGB3*	17	GPIIIa Arg↔His$_{479}$	1514A→G	779974422	Platelets and neutrophils

This table was modified from Robinson *et al.,* (2012). The HPA loci that were screened in the present study are bold. Nucleotide changes are shown from a to b alleles. The data was obtained from Newman *et al.,* (1989); Kuijpers *et al.,* (1990); Lyman *et al.,* (1990); Wang *et al.,* (1992); Santoso *et al.,* (1993); Wang *et al.,* (1993); Kuijpers *et al.,* (1993); Santoso *et al.,* (1994); Noris *et al.,* (1995); Peyruchaud *et al.,* (1997); Simsek *et al.,* (1997); Sachs *et al.,* (2002); Santoso *et al.,* (1990); Santoso *et al.,* (2002); Schuh *et al.,* (2002); Jallu *et al.,* (2002); Stafford *et al.,* (2008); Bertrand *et al.,* (2009); Peterson *et al.,* (2010); Peterson *et al.,* (2011); Jallu *et al.,* (2012); Kroll *et al.,* (2011); Sachs *et al.,* (2012); Jallu *et al.,* (2013); Poles *et al.,* (2013); Sullivan *et al.,* (2014); Wihadmadyatami *et al.,* (2015); Jallu *et al.,* (2017); Sullivan *et al.,* (2017); Poles *et al.,* (2018); Bertrand *et al.,* (2019)

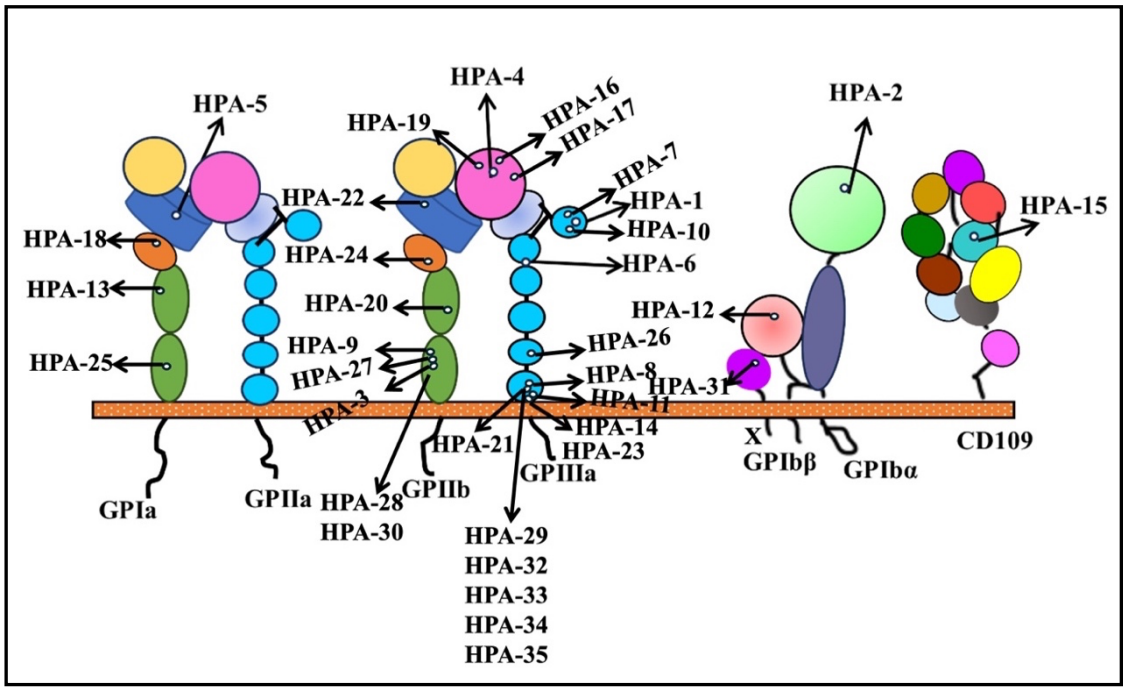

Figure 2.2 Expression of 35 HPA systems on platelet glycoproteins. This figure was adapted from Curtis & McFarland, (2013)

2.2.4 Hemochromatosis (HFE)

The *HFE* gene on the short arm of chromosome 6 codes for a protein that binds to beta-2 microglobulin (Feder *et al.,* 1996; Barton *et al.,* 2015). The latter is coded by the *BMP6* gene on the same chromosome 6 (Alvarenga *et al.,* 2020). The HFE is an integral membrane protein that is similar to MHC class I-type proteins and important in regulating the uptake of circulating iron by controlling the interaction between the transferrin receptor and transferrin (Barton *et al.,* 2015; Mónaco *et al.,* 2023). The *HFE* gene contains 7 exons spanning over 12 kb and composed of 343 amino acids (Feder *et al.,* 1996). Exon 1 codes for signal peptide, exons 2 to 4 code for an extracellular transferrin receptor-binding region (α1, α2 and α3), exon 5 codes for transmembrane region I, exon 6 codes for a short cytoplasmic region that forms a complex with beta 2-microglobulin (β2M) and exon 7 not translated into part of HFE molecule. Thus, the full-length transcript represented by 6 exons (Dorak, 2009). The

receptor-binding sites for extracellular transferrins are located in the α1 and α2 domains, which interact with Transferrin Receptor I (TFRI) (Lebrón *et al.,* 1998). On the other hand, the α3 domain binds to Transferrin Receptor 2 (TFR2) (Figure 2.4), a membrane protein similar to TFR1 but exhibiting lower affinity in transporting iron to the plasma (Graham *et al.,* 2007). The regulation of iron metabolism involves intricate interactions between transferrin, HFE and TFRI (Townsend & Drakesmith 2002). The formation of the HFE-TFRI complex inhibits the production of the iron-regulating hormone hepcidin, preventing iron absorption in liver cells (Nemeth & Ganz 2009). Conversely, the binding of TFRI to transferrin enables iron entry into liver cells while concurrently increasing hepcidin production (Ganz, 2016). In summary, TFR1 and TFR2 are crucial for the proper regulation of iron homeostasis, and both receptors are involved in pathways that ultimately influence hepcidin levels. In Hereditary Hemochromatosis, the malfunction of these pathways (due to mutations in HFE, TFR1, or TFR2) disrupts the normal regulation of iron absorption. The result is excessive iron absorption and iron overload in the body. The three common mutations in coding region of HFE gene are C282Y (rs1800562), H63D (rs1799945) and S65C (rs3002468). They are located in exons 2 and 4 (Feder *et al.,* 1996; Katsarou *et al.,* 2019; Le Gac & Férec 2005; Mura *et al.,* 1999).

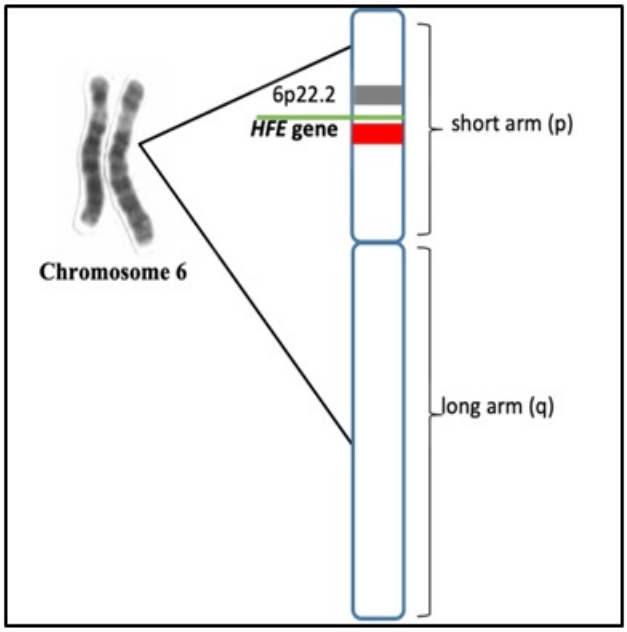

Figure 2.3 Location of the *HFE* gene locus on the short arm of chromosome 6p22.2

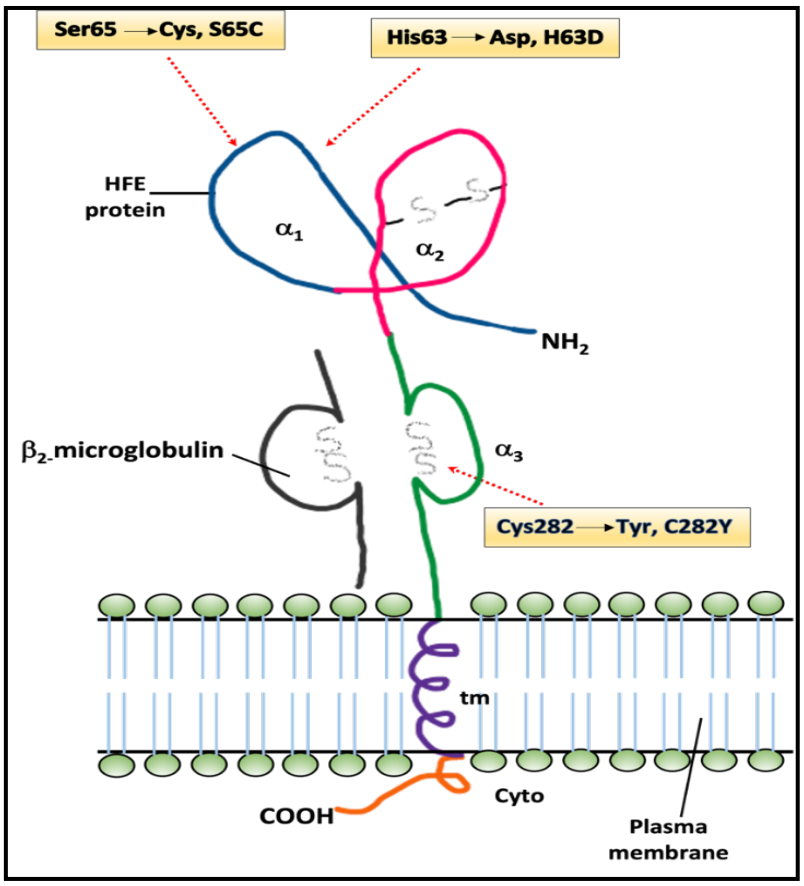

Figure 2.4 Structure of HFE protein. The three SNPs that affect protein structure associated with hereditary haemochromatosis are shown in yellow boxes