

**ALLELIC VARIATIONS IN THE *EDAR* GENE AND
THEIR INFLUENCE ON HAIR TYPES IN
MALAYSIAN POPULATIONS**

WENDY TAN LEE BOON

UNIVERSITI SAINS MALAYSIA

2025

**ALLELIC VARIATIONS IN THE *EDAR* GENE AND
THEIR INFLUENCE ON HAIR TYPES IN
MALAYSIAN POPULATIONS**

by

WENDY TAN LEE BOON

**Thesis submitted in partial fulfillment of the requirement
for the Bachelor of Science in Forensic Science.**

February 2025

DECLARATION

I hereby declare that the work presented in this dissertation, entitled “Allelic variations in the *EDAR* gene and their influence on hair types in Malaysian populations”, submitted in partial fulfillment of the requirement for the Bachelor of Science in Forensic Science at Universiti Sains Malaysia, is an original piece of research conducted under the guidance of Dr. Nur Waliyuddin Hanis bin Zainal Abidin. The information derived from the literature has been duly acknowledged in the text, and a list of references has been provided. No part of this dissertation has been submitted for any other degree or diploma at this or any other University.

Student Signature,

A handwritten signature in black ink, consisting of a large loop and a cross-like shape, positioned above a horizontal dotted line.

(Wendy Tan Lee Boon)

Date: 27th February 2025

CERTIFICATE

This is to certify that the dissertation entitled, “Allelic variations in the *EDAR* gene and their influence on hair types in Malaysian populations” is the bona fide work of Wendy Tan Lee Boon (159584), carried out under my guidance and supervision from November 2024 to February 2025. This dissertation is submitted to the Universiti Sains Malaysia in partial fulfillment of the requirement for the Bachelor of Science in Forensic Science.

Supervisor,


DR. NUR WAN YUDDIN HANIS ZAINAL ABIDIN
Pensyarah Universiti
Program Sains Forensik
Pusat Pengajian Sains Kesihatan
Kampus Kesihatan, Universiti Sains Malaysia
16150 Kubang Keratan, Kelantan
.....
(Dr. Nur Wanyuddin Hanis bin Zainal Abidin)

Date: 27 Feb 2025

ACKNOWLEDGEMENT

First and foremost, I would like to extend my deepest gratitude to my supervisor, Dr. Nur Waliyuddin Hanis bin Zainal Abidin, for granting me the invaluable opportunity to undertake my final year project under his supervision. His expertise, insightful guidance, and patience throughout this study have been instrumental in helping me overcome challenges and expand my knowledge, making this journey both meaningful and rewarding.

I would also like to express my heartfelt appreciation to Puan Nur Azeelah Abdullah and Encik Muhammad Ismail for their patience, dedication, and willingness to overlook minor mistakes while guiding me in using resources in the Human Identification Unit/DNA (HID) laboratory. Their support made my time in the lab an incredibly positive and enriching experience. I am also grateful to all the forensic laboratory staff for providing their valuable insights during material preparation, which minimized errors throughout the lab work.

My sincere appreciation goes to all participants, especially my friends and coursemates, for their unwavering support, contributions, and active involvement in this study. Their willingness to participate played an important role in the successful completion of this study. Lastly, I am truly grateful to my family for their constant support and encouragement behind the scenes.

TABLE OF CONTENTS

DECLARATION	ii
CERTIFICATE	iii
ACKNOWLEDGEMENT	iv
LIST OF FIGURES	vii
LIST OF TABLES	ix
LIST OF ABBREVIATIONS	xi
LIST OF SYMBOLS.....	xiii
ABSTRAK	xiv
ABSTRACT	xvi
CHAPTER 1: INTRODUCTION.....	1
1.1 Introduction	1
1.1.1 Genetics	1
1.1.2 Single Nucleotide Polymorphism	2
1.1.3 <i>EDAR</i> Gene	4
1.1.4 SNP rs3827760 in the <i>EDAR</i> Gene	5
1.2 Problem Statement.....	5
1.3 Objectives	6
1.3.1 General Objective.....	6
1.3.2 Specific Objectives.....	6
1.4 Study Rationale/ Significance of Study	6
CHAPTER 2: LITERATURE REVIEW	8
2.1 Global Variation of the SNP rs3827760	8
2.2 Phenotypic Trait Associated with the SNP rs3827760	9
2.2.1 Teeth Structure	9
2.2.2 Hair Structure.....	10
2.2.3 Chin Protrusion	12
2.3 Additional SNPs in the <i>EDAR</i> gene.....	13
2.3.1 SNP rs146567337 in the <i>EDAR</i> Gene.....	13
2.3.2 SNP rs12623957 in the <i>EDAR</i> Gene	14
2.4 Silica-Based Solid-Phase Extraction (SPE)	14
2.5 Sanger Sequencing.....	15
CHAPTER 3: METHODOLOGY	17
3.1 Materials	17
3.2 Ethical Compliance	22
3.3 Study Location, Population, and Sample.....	22
3.4 Subject Criteria.....	22
3.5 Sample Collection.....	24
3.6 DNA Extraction	26
3.7 Agarose Gel Electrophoresis (1%)	27
3.8 DNA Quantification (DeNovix)	30
3.9 Polymerase Chain Reaction (PCR)	31

3.10 Agarose Gel Electrophoresis (3%)	34
3.11 DNA Purification	36
3.12 DNA Quantification of Purified PCR Product (Qubit).....	37
3.13 Cycle Sequencing.....	39
3.14 Cycle Sequencing Product Purification (Ethanol Precipitation).....	41
3.15 Sanger Sequencing	42
3.16 Data Analysis using Chromas and Jalview	44
3.17 Statistical Analysis using SPSS (Chi-square test)	45
CHAPTER 4: RESULTS	46
4.1 Sample Collection.....	46
4.2 Agarose Gel Electrophoresis (1%)	47
4.3 DNA Quantification (DeNovix)	48
4.4 Polymerase Chain Reaction (PCR) and Agarose Gel Electrophoresis (3%) ..	50
4.5 DNA Quantification of Purified PCR Product (Qubit).....	52
4.6 Data Analysis using Chromas and Jalview	53
4.7 Statistical Analysis using SPSS (Chi-square test)	59
CHAPTER 5: DISCUSSION	72
5.1 Agarose Gel Electrophoresis (1%)	72
5.2 DNA Quantification (DeNovix)	72
5.3 Polymerase Chain Reaction (PCR) and Agarose Gel Electrophoresis (3%) ..	73
5.4 Allelic Variations in the <i>EDAR</i> Gene	74
5.4.1 SNP rs3827760.....	75
5.4.2 SNP rs146567337	76
5.4.3 SNP rs12623957	77
CHAPTER 6: CONCLUSION AND FUTURE RECOMMENDATIONS	78
6.1 Conclusion.....	78
6.2 Limitations of the Study.....	78
6.3 Recommendation for Future Research.....	80
REFERENCES.....	81
APPENDIX.....	94

LIST OF FIGURES

Figure 3.1: Workflow of the study.	23
Figure 4.1: UV-illuminated 1% agarose TBE gel showing the sizes of extracted gDNA sizes from buccal cells, referenced against the VC Lambda/ HindIII marker.	47
Figure 4.2: Absorbance spectrum graph of the extracted DNA, showing wavelength (nm) plotted against absorbance.	49
Figure 4.3: UV-illuminated 3% agarose TBE gel showing PCR products amplified using the EDAR R2 primer, with sizes determined by comparing to the GeneRuler 50 bp DNA Ladder.	50
Figure 4.4: DNA sequencing chromatogram generated using the SeqStudio™ Genetic Analyzer and visualized with Chromas.	54
Figure 4.5: Targeted observed nucleotide variants in the SNP rs3827760, including a homozygous nucleotide (G) in the first chromatogram, a heterozygous nucleotide (A&G) represented as ‘R’ in the second chromatogram), and a homozygous nucleotide (A) in the third chromatogram.	55
Figure 4.6: Additional observed nucleotide variants in SNP rs146567337, include a homozygous nucleotide (T) in the first chromatogram, a heterozygous nucleotide (G&T) represented as ‘K’ in the second chromatogram and a homozygous nucleotide (G) in the third chromatogram.	56
Figure 4.7: Additional observed nucleotide variants in SNP rs12623957, including a homozygous nucleotide (A) in the first chromatogram and a heterozygous nucleotide (A&G) represented as ‘R’ in both the second and third chromatograms.	57
Figure 4.8: Multiple sequences with low-quality bases trimmed and aligned using Jalview.	58
Figure 4.9: Bar chart showing the frequency of SNP rs3827760 nucleotide variants within each ethnic group.	64
Figure 4.10: Bar chart showing the frequency of SNP rs3827760 nucleotide variants within each hair type.	64

Figure 4.11: Bar chart showing the frequency of SNP rs146567337 nucleotide variants within each ethnic group.	67
Figure 4.12: Bar chart showing the frequency of SNP rs146567337 nucleotide variants within each hair type.	67
Figure 4.13: Bar chart showing the frequency of SNP rs12623957 nucleotide variants within each ethnic group.	70
Figure 4.14: Bar chart showing the frequency of SNP rs12623957 nucleotide variants within each hair type.	70

LIST OF TABLES

Table 3.1: List of materials and equipment used in the study.	17
Table 3.2: Characteristics and representative images for each hair type.....	25
Table 3.3: Primer sequences, product length, and melting temperature of <i>EDAR</i> R2 primers.....	31
Table 3.4: The components volume for positive control, negative control, a single PCR reaction, and ten PCR reactions.....	32
Table 3.5: Thermal cycling conditions for performing PCR.	33
Table 3.6: The volume of components required for a standard assay tube and a sample assay tube.....	37
Table 3.7: The components volume for a control, a single cycle sequencing reaction, and ten cycle sequencing reactions.	39
Table 3.8: Recommended quantities for cycle sequencing based on PCR product size (NimaGen, 2024).....	39
Table 3.9: Thermal cycling conditions for performing cycle sequencing.	40
Table 4.1: Ethnicity and hair types of the participants.	46
Table 4.2: DNA concentration (dsDNA) and purity ratios (A260/280 and A260/230) of the extracted DNA measured using the DeNovix DS-11 spectrophotometer.	48
Table 4.3: DNA concentration in purified PCR products measured using the Qubit® 2.0 fluorometer.	52
Table 4.4: Ethnicity, hair types, and nucleotide variants at SNP rs3827760, SNP rs146567337, and SNP rs12623957 among 30 participants, dataset for SPSS analysis.	60
Table 4.5: Crosstabulation of nucleotide variation in SNP rs3827760 across different ethnic groups.	62
Table 4.6: Crosstabulation of nucleotide variation in SNP rs3827760 across different hair types.....	63

Table 4.7: Crosstabulation of nucleotide variation in SNP rs146567337 across different ethnic groups.....	65
Table 4.8: Crosstabulation of nucleotide variation in SNP rs146567337 across different hair types.	66
Table 4.9: Crosstabulation of nucleotide variation in SNP rs12623957 across different ethnic groups.	68
Table 4.10: Crosstabulation of nucleotide variation in SNP rs12623957 across different hair types.	69
Table 4.11: Association of SNP rs3827760 nucleotide variants with ethnicity and hair type.	71
Table 4.12: Association of SNP rs146567337 nucleotide variants with ethnicity and hair type.	71
Table 4.13: Association of SNP rs12623957 nucleotide variants with ethnicity and hair type.	71

LIST OF ABBREVIATIONS

A	Adenine
Ala	Alanine
Arg	Arginine
C	Cytosine
cDNA	Complementary DNA
CHB	Han Chinese
DNA	Deoxyribonucleic acid
ddNTPs	Dideoxynucleotide triphosphates
dNTPs	Deoxynucleotide triphosphates
dsDNA	Double-stranded DNA
EDA	Ectodysplasin-A
<i>EDAR</i>	Ectodysplasin A receptor
EDARADD	<i>EDAR</i> -associated death domain
EDTA	Ethylenediaminetetraacetic acid
FST	Fixation index
G	Guanine
GC	Guanine-cytosine
HED	Hypohidrotic ectodermal dysplasia
Hi-Di	Highly deionized
IDN	Indonesian
JEPeM	Human Research Ethics Committee
JPN	Japanese
LD	Linkage disequilibrium

mRNA	Messenger ribonucleic acid
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next-generation sequencing
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCR	Polymerase Chain Reaction
Ser	Serine
Ser380Arg	Serine-to-Arginine substitution at position 380 in the <i>EDAR</i> protein
SHG	Scandinavian Hunter-Gatherers
S/N	Signal-to-noise ratios
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SPE	Solid-phase extraction
ssDNA	Single strands
T	Thymine
TBE	Tris-borate-EDTA
THM	Thai-Mai
TNFR	Tumor necrosis factor receptor
USM	Universiti Sains Malaysia
Val	Valine
Val370Ala	Valine-to-Alanine substitution at position 370 in the <i>EDAR</i> protein

LIST OF SYMBOLS

~	Approximately
bp	Base pair
°C	Degrees Celsius
g	Gram
Mg ²⁺	Magnesium ion
μL	Microliter
μM	Micromolar
mL	Milliliter
M	Molar
n	Sample size
ng	Nanogram
nm	Nanometre
%	Percentage
®	Registered trademark
™	Trademark
V	Volt

VARIASI ALEL DALAM GEN *EDAR* DAN PENGARUHNYA TERHADAP JENIS RAMBUT DALAM POPULASI MALAYSIA

ABSTRAK

Ectodysplasin A Receptor (*EDAR*) adalah reseptor faktor nekrosis tumor (TNFR) yang terlibat dalam perkembangan tisu ektoderma, termasuk folikel rambut. Polimorfisme nukleotida tunggal (SNP) rs3827760, yang mengkod *EDAR*:p.(Val370Ala), merupakan mutasi salah erti dalam gen *EDAR* yang mempengaruhi ciri-ciri fenotip secara pleiotropik, terutamanya jenis rambut, dalam populasi Asia Timur. Walau bagaimanapun, taburan genotip serta hubungan SNP ini dengan jenis rambut dalam kalangan kumpulan etnik di Malaysia masih belum diterokai secara meluas. Oleh itu, kajian ini bertujuan untuk menentukan variasi alel SNP rs3827760 dalam populasi Malaysia, khususnya dalam kalangan kumpulan etnik Melayu, Cina, dan India. Sampel DNA (swab bukal) diperoleh daripada 30 peserta, dengan 10 individu daripada setiap kumpulan etnik, dan diekstrak menggunakan kaedah ekstraksi fasa pepejal berasaskan silika (SPE). Tindak balas berantai polimerase (PCR) dijalankan menggunakan primer *EDAR* tersuai, diikuti dengan elektroforesis gel agarosa, pengkuantitian DNA, penjujukan kitaran, dan pemendakan etanol (penulenan). Penjujukan Sanger dijalankan menggunakan SeqStudio™ Genetic Analyzer, manakala kromatogram dianalisis menggunakan perisian Chromas dan Jalview untuk mengenal pasti variasi nukleotida pada lokus SNP rs3827760, rs146567337, dan rs12623957. Ujian khi kuasa dua dijalankan bagi menentukan asosiasi statistik antara SNPs dengan etnik atau jenis rambut. Analisis menunjukkan bahawa SNP rs3827760 mempunyai hubungan yang signifikan dengan etnik (nilai $P <$

0.001) dan jenis rambut (nilai $P = 0.011$). Alel leluhur (alel A) lebih kerap ditemukan dalam kalangan individu India (rambut kerinting), manakala alel terbitan (alel G) lebih dominan dalam kalangan individu Cina (rambut lurus). Individu Melayu dengan rambut lurus atau bergelombang menunjukkan gabungan ketiga-tiga genotip. SNP rs146567337 tidak menunjukkan perkaitan yang signifikan dengan etnik (nilai $P = 0.155$) atau jenis rambut (nilai $P = 0.432$). Walaupun SNP rs12623957 tidak berkait secara signifikan dengan etnik (nilai $P = 0.058$), ia menunjukan korelasi dengan jenis rambut (nilai $P = 0.017$). Oleh itu, SNP rs3827760 dalam gen *EDAR* dikenal pasti sebagai faktor genetik yang paling berkaitan dengan jenis rambut dan etnik dalam kajian ini.

ALLELIC VARIATIONS IN THE *EDAR* GENE AND THEIR INFLUENCE ON HAIR TYPES IN MALAYSIAN POPULATIONS

ABSTRACT

Ectodysplasin A Receptor (*EDAR*) is a tumor necrosis factor receptor (TNFR) that is involved in the development of ectodermal tissues, including hair follicles. The single nucleotide polymorphism (SNP) rs3827760, which encodes *EDAR*:p.(Val370Ala), is a missense mutation in the *EDAR* gene that pleiotropically influences phenotypic traits such as hair type, particularly in East Asian populations. However, the genotype distribution and correlation of this SNP with hair type among Malaysian ethnic groups remain largely unexplored. Thus, this study aimed to determine the allelic variation of SNP rs3827760 in the Malaysian population among the Malay, Chinese, and Indian ethnic groups. DNA samples (buccal swabs) were collected from 30 participants, with 10 individuals from each ethnic group, and extracted using a silica-based solid-phase extraction (SPE) method. Polymerase chain reaction (PCR) was performed using the custom *EDAR* primers, followed by agarose gel electrophoresis, DNA quantification, cycle sequencing, and ethanol precipitation (purification). Sanger sequencing was conducted using the SeqStudio™ Genetic Analyzer, and the chromatograms were analyzed with Chromas and Jalview software to identify nucleotide variations at the SNP loci rs3827760, rs146567337, and rs12623957. A chi-square test was performed to determine the statistical significance of the association between the SNPs and ethnicity or hair type. The analysis showed that SNP rs3827760 was significantly correlated with both ethnicity (P-value <0.001) and hair type (P-value = 0.011). The ancestral allele (allele A) was strongly associated

with Indian (curly hair) individuals, while the derived allele (allele G) was strongly associated with Chinese (straight hair) individuals. Malay individuals with straight or wavy hair exhibited a mix of all three genotypes. SNP rs146567337 showed no significant association with ethnicity (P-value = 0.155) or hair type (P-value = 0.432). SNP rs12623957 was not significantly related to ethnicity (P-value = 0.058) but was correlated with hair type (P-value = 0.017). Therefore, SNP rs3827760 in the *EDAR* gene was the most strongly correlated genetic factor for hair types and ethnicity in this study.

CHAPTER 1: INTRODUCTION

1.1 Introduction

The study of human genetics is important for enhancing the understanding of hereditary traits and their variations among different populations, thus shedding light on the roles that particular gene variations play in both biological functions and physical characteristics.

1.1.1 Genetics

Deoxyribonucleic acid (DNA) is the molecular blueprint for life located in the nucleus of nearly all cells in an organism and carries genetic information necessary for growth, reproduction, and the overall functioning of living organisms. DNA is composed of two strands of nucleotides (smaller repeating units) running in opposing directions that form a double helix (spiral) structure. DNA will replicate before cell division by using each strand as a template for a new complementary strand formation, ensuring genetic uniformity. Each nucleotide has three main components: a phosphate group which provides structural support; deoxyribose, a sugar molecule that links the phosphate group to the nitrogenous base; and nitrogenous base (adenine (A), thymine (T), cytosine (C), and guanine (G)) which pair specifically through hydrogen bonding (A with T, C with G) to form the rungs of the DNA ladder. DNA is tightly wrapped around histone protein, fitting about 2 meters of DNA into a microscope nucleus through multiple levels of coiling and folding, to form a compact organized genetic material called chromosome. Humans typically have 46 chromosomes, arranged in 23 pairs, with one chromosome in each pair inherited from each parent, ensuring redundancy. Therefore, DNA is also known as the carrier of hereditary information as the genetic information is passed down from one generation to the next, maintaining

the continuity of life. Gene, also known as the information encoded in sections of DNA, serves as instructions for protein synthesis through transcription (copying DNA into messenger ribonucleic acid (mRNA)) and translation (assembling amino acids into proteins a ribosome by mRNA). Since different cell types require different proteins to function, DNA also serves as a critical control system in cells to determine the types and amounts of proteins that are generated. There is only about 1% coding DNA in humans, the rest is known as non-coding DNA which plays a role in gene activity regulation (Calladine and Drew, 1997).

Similar to fingerprints, each person's DNA, except homozygous twins, has a unique genetic code that remains constant throughout their lifetime. With only about 0.1 % variation in human DNA sequence, the probability of two unrelated individuals having identical DNA is an astronomical 1 in 594.1 trillion (Nizami et al., 2018). This distinctiveness makes DNA an invaluable tool in forensic science such as crime scene investigation, paternity tests, and human remains identification. By analyzing the biological samples, individuals can be identified with remarkable precision, thus ensuring justice for victims and society by identifying the victim and establishing connections to suspects (Alketbi, 2023).

1.1.2 Single Nucleotide Polymorphism

Polymorphism refers to the presence of two or more variants (alleles) at a specific locus within a population. To be classified as a polymorphism, the genetic variant must occur at a frequency of at least 1% of the population. Polymorphism identification is challenging and complex due to the diverse nature of DNA variants and mutation types, which range from single-base changes to more significant mutations. The simplest form of polymorphism is known as a single nucleotide polymorphism (SNP), which

involves the substitution of one nucleotide for another at a specific genomic site. SNPs are the most abundant type of polymorphism in the genome, occurring approximately every 0.3-1 kilobase, but estimates vary based on the population and genomic region (Schork et al., 2000).

Types of SNPs are classified based on their genomic location and functional implications. SNPs occurring within the coding regions of genes are classified as coding SNPs and are further divided into synonymous SNPs and non-synonymous SNPs. Synonymous SNPs, also called silent mutation, do not change the encoded amino acid sequence and have little or no impact on the phenotype. However, they may still influence mRNA stability or splicing efficiency. In contrast, non-synonymous SNPs (missense or nonsense mutations) are more likely to affect protein function and may contribute to disease because they alter the amino acid sequence, leading to either an amino acid substitution (missense) or premature stop codon (nonsense). Non-synonymous SNPs occur less commonly than synonymous SNPs due to the higher selective pressure to maintain functional proteins (Cargill et al., 1999; Gray et al., 2000). Non-coding SNPs, as the name, occur outside the coding regions. For example, SNPs that occur in promoter, enhancer, and other regulatory regions of the genome are categorized as regulatory SNPs. These SNPs can alter the binding sites of transcription factors or influence enhancer activities, thereby affecting gene expression (Degtyareva et al., 2021). In addition, intronic SNPs occur within intron, and intergenic SNPs are found between genes.

SNPs are significant for studying gene that influencing multifactorial diseases and traits, as well as for gene mapping because they are highly abundant across the genome, offering a dense network of markers for gene mapping and association studies (Kruglyak, 1997); they are widespread throughout the genome, including exons,

introns, promoters, and enhancers, enhancing the prevalence of yielding alleles that are functional or physiologically relevant; neighbouring SNPs often show linkage disequilibrium (LD), creating haplotype diversity that supports genetic linkage and direct association studies (Nickerson et al., 1992); they are simple structured, allowing rapid and efficient genotyping using advanced technologies; they are population dynamics (allele frequencies influenced by population genetics), making them valuable in population-based studies (McKeigue, 1998); they are less mutable during allele transmission across generation, reducing confounding effects and ensuring reliability of LD and association studies (Brookes, 1999); and they can reveal LD patterns with recurrent mutation, gene conversion, or recombination hotspots, aiding studies of traits and diseases (Clark et al., 1998; Schork et al., 2000).

1.1.3 *EDAR* Gene

Ectodysplasin A Receptor (*EDAR*) gene encodes a type I transmembrane protein, specifically a tumor necrosis factor receptor (TNFR), which is essential for the development of ectodermal tissues such as teeth, sweat glands, and hair follicles (Headon and Overbeek, 1999). *EDAR* interacts with its ligand, ectodysplasin-A (EDA), at the TNFR domain (30-148 amino acids), and recruits an adapter protein called *EDAR*-associated death domain (*EDARADD*) through its death domain (358-431 amino acids) to activate the downstream signaling pathways, primarily the nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) pathway which regulates the formation and morphogenesis of structures derived from the ectoderm (Headon et al., 2001; Kumar et al., 2001). Mutations in the *EDAR* gene disrupt the pathway mentioned above, leading to autosomal forms of hypohidrotic ectodermal dysplasia (HED) characterized by hypotrichosis (reduced hair), hypodontia (missing

teeth), and hypohidrosis (reduced sweat glands) (Okita et al., 2019; Andreoni et al., 2021).

1.1.4 SNP rs3827760 in the *EDAR* Gene

SNP rs3827760 (*EDAR*:c.1109T>C or 1540T/C) is one of the single nucleotide variants (SNVs) or polymorphisms in the *EDAR* gene on chromosome 2, located at the 108,897,145th nucleotide in the genomic sequence (chr2:108897145). The variation can manifest as either A>C or A>G (National Center for Biotechnology Information (NCBI), 2022c). A or T is known as the ancestral allele, while C or G is known as the derived allele. SNP rs3827760 is classified as a non-synonymous SNP, specifically a missense mutation, as it results in the substitution of valine (Val) to alanine (Ala) at position 370 (*EDAR*:p.(Val370Ala)) in the death domain of the *EDAR* protein. This amino acid substitution enhances *EDAR* signaling by improving the recruitment efficiency of EDARADD, thereby amplifying the activation of NF-κB, making it a gain-of-function mutation. Furthermore, several phenotypic traits have been linked to rs3827760 variation, including thicker and straighter hair fibers, an increased density of sweat glands on fingertips, alterations in tooth morphology (such as shovel-shaped incisors), changes in ear and chin shapes, and variations in the size of sebaceous glands (Riddell, 2020; Riddell et al., 2020).

1.2 Problem Statement

SNP rs3827760 is a pleiotropic single nucleotide variant (SNV) in the *EDAR* gene that influences various phenotypic traits, including hair morphology, dental structure, and sweat gland density. These variations have been associated with strong selection pressure or demographic expansion. However, studies on the allelic variation in SNP

rs3827760 have primarily focused on the East Asian population, with minimal research on Southeast Asia (SEA), particularly in Malaysia. Malaysia is a multi-ethnic society shaped by a complex demographic history, including migration, admixture, local adaptation, and cultural exchange, resulting in each ethnic group having a unique genetic makeup. The lack of data on Malaysian populations limits the understanding of ethnicity- or trait-specific genotypes. Therefore, this study aims to determine the allelic variation in SNP rs3827760 in the *EDAR* gene among Malaysian populations by examining its distribution across different ethnic groups and hair types.

1.3 Objectives

1.3.1 General Objective

To investigate allelic variations in SNPs of the *EDAR* gene and their influence on hair types in Malaysian populations.

1.3.2 Specific Objectives

- i. To analyze allelic variation in the *EDAR* gene at SNPs rs3827760 (primary focus), rs146567337, and rs12623957 among major Malaysian ethnic groups (Malay, Chinese, and Indian).
- ii. To assess the correlation between SNPs rs3827760, rs146567337, and rs12623957 and hair types (straight, wavy, curly) in Malaysian populations.

1.4 Study Rationale/ Significance of Study

This study aims to explore the relationship between SNP rs3827760 and ethnicity or hair type within the Malaysian population. Establishing this relationship could reveal ethnicity- or trait-specific genotypes, providing valuable insights into the rich tapestry

of genetic diversity, evolution, ancestry, and demographic history of Malaysian populations. Furthermore, these specific genotypes may have forensic applications, aiding in human identification through ethnic classification and phenotypic prediction.

CHAPTER 2: LITERATURE REVIEW

2.1 Global Variation of the SNP rs3827760

The derived allele (G) of SNP rs3827760 is highly prevalent in East Asian and Native American populations. Frequencies close to 90% of the derived allele have been observed in East Asian populations, particularly in Han Chinese (CHB), while it is nearly fixed (close to 100%) in Native American populations. Based on the median-joining haplotype network analysis, the star-like pattern suggested that the high prevalence of the derived allele in East Asian and Native American populations is due to strong selection pressures or demographic expansion in these regions. Besides that, the derived allele shows lower frequencies in SEA populations, including Vietnam, the Philippines, Malaysia, and Indonesia, compared to East Asia. This lower prevalence is likely due to reduced selection pressure or different demographic histories in these regions. Furthermore, the derived allele is virtually absent in the African population and is rare (approximately 4.2% frequency (Xue et al., 2009)) in European populations, suggesting that it evolved after the migration of modern humans out of Africa and that there was limited gene flow to Europe (Riddell et al., 2020).

The derived allele is found in East Asian populations, including CHB and Japanese (JPN), with a combined frequency of 87.6% (Fujimoto et al., 2008a). In other East Asian populations, it has a frequency of 35.18% in Tibetans, 13.78% in Mongolians, and 6.89% in Li population (Tan et al., 2013). In SEA populations, its frequency is 33.9% in the Indonesian (IDN) population, 30.0% in the Thai-Mai (THM) population (Fujimoto et al., 2008a), and 41.7% in Malaysian populations (Ling, 2024). In Melanesian populations, the frequency is 1.0% in the Gidra population (New Guinea) and 10.4% among Solomon Islanders (Fujimoto et al., 2008a). The derived allele is also found in the admixed Latin American population at a frequency of 42% (Adhikari et al.,

2016). Furthermore, Mathieson et al. (2015) significantly discovered that the derived allele of SNP rs3827760 was present in Scandinavian Hunter-Gatherers (SHG) from Motala, Sweden, with a prevalence of 50%, as the derived allele was largely absent in Europeans. Since genomic analyses found no evidence of East Asian ancestry in SHG, the presence of the derived allele in SHG is likely attributed to the complex demographic history of Scandinavia, which may have involved migrations or isolated instances of allele introduction unrelated to East Asian heritage.

Yamaguchi-Kabata et al. (2008) discovered a pronounced differentiation in the allele distribution of SNP rs3827760 between the two main JPN subpopulations, the Hondo and Ryuku clusters. A total of 7,003 JPN individuals were classified into these two main clusters based on genetic analyses. Population structure was then assessed, and SNP rs3827760 was identified using principal component analysis (PCA) and genetic differentiation measures, such as the fixation index (F_{ST}). The T allele (ancestral allele) frequency in the Hondo cluster was 0.222, while in the Ryuku cluster, it was 0.398. The Cochran-Armitage trend test was used to evaluate the difference in genotype frequency, yielding a highly significant P-value of 7.73×10^{-21} .

2.2 Phenotypic Trait Associated with the SNP rs3827760

2.2.1 Teeth Structure

Kataoka et al. (2021) found an association between the SNP rs3827760 and root number/shape in teeth within a JPN population (225 individuals). The effects of *EDAR* 370A varied across tooth types: a decreased number of roots in upper first premolars (UP1s), an increased number of roots in lower first molars (LM1s), and the presence of a C-shaped single root in lower second molars (LM2s). The significant association between the SNP rs3827760 and tooth root number/shape was confirmed using logistic

regression analysis, the P-values of 0.0034, 0.039, and 0.00028, all P-values were less than 0.05. The effects of the SNP rs3827760 on tooth root morphogenesis were stimulated using a reaction-diffusion model. The decrease in UP1 roots or the formation of C-shaped LM2 roots was likely due to the increased activator synthesis, while the increase in LM1 roots was likely due to increased inhibitor synthesis.

Kimura et al. (2009) determined that the SNP rs3827760, specifically the derived 1540C allele, was strongly correlated with the shovel-shaped incisors, a trait prevalent in East Asian populations. This finding was based on a genetic analysis of 202 JPN individuals and confirmed with a P-value of 7.7×10^{-10} using Spearman's correlation test. This derived allele exhibited an additive effect, with the shoveling grade increasing by 0.7 points for every copy of 1540C. Moreover, principal component analysis (PCA) revealed that the allele affected both the proportional dimensions of teeth (mesiodistal to buccolingual diameter ratio; P-value = 0.0026), and overall tooth size (P-value = 0.0049). Regression analyses further validated that the additive effect of the 1540C allele on dental morphology was independent of sex and region.

2.2.2 Hair Structure

Adhikari et al. (2016a) identified that SNP rs3827760 has a strong association with scalp hair shape by assessing the genetic diversity of 6,630 individuals of admixed ancestry (European, Native American, and African) from five Latin American countries (Brazil, Colombia, Chile, Mexico, and Peru). The genome-wide association studies (GWAS) were performed using multivariate linear regression in PLINK v.1.9, resulting in a P-value of 3×10^{-119} , indicating a very strong genetic signal. At the same time, a meta-analysis was performed to ensure the consistency of results across all country samples. Moreover, they discovered that the rs3827760 signal was independent,

confirming its association with hair shape and not influenced by other nearby SNPs associated with different traits, such as beard thickness, determined through conditional analyses. Endo et al. (2018) conducted replication analyses and confirmed the association between the SNP rs3827760 and both eyebrow thickness (thick vs. thin eyebrow) and hair morphology (straight vs. curly hair) through GWAS involving 11,311 JPN females. Genome-wide significance thresholds were used to validate the findings, and meta-analysis was used to ensure consistency and reliability across different stages of the study, emphasizing SNP rs3827760 as a significant marker for genetic studies focused on ectodermal traits in East Asian populations. This association was also corroborated by Ling (2024) in Malaysian populations, with a P-value of 0.0001.

Wu et al. (2016) discovered that the SNP rs3827760 was associated with hair straightness in both CHB and Uyghur (an admixed population of East Asians and Europeans) populations, supported by genome-wide significant P-value of 4.67×10^{-16} (CHB) and 1.92×10^{-12} (Uyghurs), and confirmed through logistic regression analysis. SNP rs3827760 was found in East Asian populations due to strong positive selection. However, since this selection pressure was absent in the Uyghur population after admixture, it was suggested that SNP rs3827760 in the Uyghur population likely originated from a single mutation event in East Asian ancestors and was subsequently passed to the Uyghurs through admixture.

Based on the genetic and statistical analysis in two SEA populations, the IDN population and the THM population, the P-values of 5.5×10^{-3} (IDN) and 9.5×10^{-4} , generated using ANOVA, indicated a significant association between the 1540C allele and increased hair cross-sectional area, contributing to thicker hair (Fujimoto et al., 2008a). However, this derived allele was independent of age, sex, and ethnicity (determined by regression analysis), as well as less frequently present in Melanesian

populations, which have thinner hair. Functional tests (Luciferase Assays) suggested that hair morphology was most likely affected through the activation of the downstream NF- κ B signaling pathway, which was impacted by the 1540C allele. In addition, the evolutionary analysis, including extended haplotype homozygosity (EHH) and long-range haplotype (LRH) tests, suggested that the high frequency of the 1540C allele in East Asian populations was likely due to selective advantages.

A replication study was carried out by Fujimoto et al. (2008b) on 189 JPN individuals to confirm the association between the 1540C allele and hair cross-sectional area. The P-value of 1.4×10^{-5} , generated using ANOVA, indicated a significant association, and regression analysis demonstrated that the association was independent of age and sex. Although there were differences in hair cross-sectional area between the JPN and SEA populations, likely due to genetic or environmental factors, the derived allele still exhibits a codominant effect across both populations. Moreover, the relationship between SNP rs3827760 and hair morphology was also identified using a transgenic mouse model (EdarTg951). The transgenic mice displayed coarser, straighter hair fibers with a circular cross-sectional profile, closely resembling the typical East Asian hair form, validating the gain-of-function mutation associated with SNP rs3827760 (Mou et al., 2008).

2.2.3 Chin Protrusion

Within the admixed Latin American population, Adhikari et al. (2016b) also found that SNP rs3827760 was strongly associated with chin protrusion, with a P-value of 4×10^{-10} , through a GWAS. SNP rs3827760, specifically the derived G allele, which occurred at a frequency of 42% in the admixed Latin American population, was associated with reduced chin protrusion and explained 1.32% of the phenotypic

variation in chin protrusion. This association was further validated through experiments on mouse models with altered *EDAR* expression, which showed consistent results.

Section 2.1 and **Section 2.2** of the literature review demonstrated that SNP rs3827760 has been widely studied in the East Asia population, particularly among the Japanese. However, data on this SNP in SEA, especially Malaysia, remains scarce, with no published studies exploring its distribution or impact within the Malaysian population. In addition, while SNP rs3827760 has been linked to various phenotypic traits, such as tooth and chin morphology, more studies have established a correlation between this SNP and hair morphology, making it a key genetic marker for hair type variation.

2.3 Additional SNPs in the *EDAR* gene

Although SNP rs3827760 is the primary SNP identified in the *EDAR* gene, additional SNPs such as rs146567337 and rs12644248 have also been identified.

2.3.1 SNP rs146567337 in the *EDAR* Gene

SNP rs146567337 (*EDAR*:c.1138A>C), similar to SNP rs3827760, is classified as a missense and gain-of-function mutation. However, it results in the substitution of serine (Ser) to arginine (Arg) at position 380 (*EDAR*:p.(Ser380Arg)) in the death domain of the *EDAR* protein, which may lead to modifications in traits, though the exact phenotypic effects remain unknown (Riddell et al., 2020; National Center for Biotechnology Information, 2022b). Besides that, SNP rs146567337 is less widespread than SNP rs3827760, being more concentrated in southern regions of East and SEA, such as southern China, Vietnam, the Philippines, Malaysia, and IDN, with a derived allele frequency of up to ~5%. SNP rs146567337 also exhibits less haplotype

homozygosity than SNP rs3827760, indicating weaker selective pressure and suggesting that its associated phenotypic traits are more regionally relevant to SEA rather than being broadly advantageous in diverse environments. Furthermore, a double mutant carrying both Val370Ala and Ser380Arg slightly enhances NF- κ B activation compared to the single variants (Riddell et al., 2020).

2.3.2 SNP rs12623957 in the *EDAR* Gene

SNP rs12623957 (*EDAR*:c.1056C>T) is classified as a synonymous or stop-gained mutation, as it does not alter the amino acid sequence of the protein and is phenotypically irrelevant to HED (Wohlfart et al., 2016; National Center for Biotechnology Information, 2022a). Instead of being a causative variant for the phenotypic differences such as tooth shoveling grades, SNP rs12623957 is considered a neighbouring marker of interest due to its association with phenotypic traits, which arises indirectly through its LD with functional variants such as SNP rs3827760 (Kimura et al., 2009).

2.4 Silica-Based Solid-Phase Extraction (SPE)

Silica-based solid-phase extraction (SPE) is a method to isolate and purify DNA from various biological samples by utilizing the adsorption of nucleic acid to silica surfaces in the presence of high concentrations of chaotropic salts (Li et al., 2022). These salts disrupt the structure of the DNA, accelerating the binding of negatively charged DNA molecules to silica particles through a salt bridge mechanism (Doran and Foran, 2014). The extraction process begins with lysing the starting biological samples, followed by binding nucleic acid to a silica-gel membrane in the presence of high concentrations of chaotropic salts. Unwanted contaminants, such as proteins and

cellular debris, are removed through washing. Lastly, pure DNA is eluted from the silica substrate using an elution buffer (Dilley et al., 2019).

This method yields DNA of higher purity compared to organic and Chelex methods (Cavanaugh, and Bathrick, 2018), provides high throughput, and produces consistent results across a variety of sample types, including saliva and diluted blood, with 82% of samples producing conclusive profiles with balanced allelic peaks. However, the method is often ineffective at recovering smaller, degraded DNA fragments, as it preferentially binds high molecular weight DNA, with only approximately 50% of DNA from trace samples being extracted efficiently. It may also lead to DNA loss due to ineffective adsorption of DNA onto the silica substrate or ineffective elution of DNA from the silica substrate (Fan et al., 2017; Dilley et al., 2019). In addition, DNA recovery can be improved by optimizing elution procedures, such as using smaller elution volumes or multiple elutions (Dilley et al., 2019).

2.5 Sanger Sequencing

Sanger sequencing is a method used to determine nucleic acid sequences required for genetic analysis. In this method, the amplified DNA or complementary DNA (cDNA) is annealed to an oligonucleotide primer and then extended by the DNA polymerase enzyme. Deoxynucleotide triphosphates (dNTPs: dATP, dCTP, dGTP, dTTP) and chain-terminating dideoxynucleotide triphosphates (ddNTPs: ddATP, ddCTP, ddGTP, ddTTP) are incorporated at random positions during synthesis by DNA polymerase. During incorporation, the elongation process is halted by the ddNTPs at specific nucleotides as they present in limiting concentrations, producing DNA fragments of various lengths. These terminated fragments can be distinguished as each end is labeled with fluorescent dyes specific to each base (A, T, G, C). After synthesis,

capillary electrophoresis is used to separate the fragments based on size. Then, the automated sequencer detects the fluorescence at the end of each fragment to reconstruct the DNA sequence, which is displayed as chromatograms (electropherograms) showing peaks representing nucleotide signals (Sanger et al., 1977).

Sanger sequencing provides 99.999% base accuracy and is more cost-effective for targeted gene sequencing or small projects compared to next-generation sequencing (NGS). It is suitable for structural variant detection as it can generate long sequences of up to 1000 bp and is useful in forensic cases because it works well with degraded or poor-quality DNA samples (Al-Shuhaib and Hashim, 2023). However, it initiates low-quality regions, resulting in 15-40 bp of unreliable sequence data in primer binding regions, making it challenging to sequence amplicons shorter than 100 bp. Furthermore, distinguishing single nucleotide differences in sequences longer than 900 bp is also challenging (Crossley et al., 2020).

CHAPTER 3: METHODOLOGY

3.1 Materials

An inventory of the materials and equipment used in this study was detailed in **Table 3.1**, specifying the name, brand, and model of each item used in every step of the study, ensuring clarity in the experimental procedures.

Table 3.1: List of materials and equipment used in the study.

No	Materials/ equipment	Brand	Model
A. Sample collection			
1	Sampling swab	-	-
2	Microcentrifuge tube (1.5 mL)	Eppendorf	-
3	Microcentrifuge tube rack	Eppendorf	-
4	Sterile scissors	-	-
5	Consent form	-	-
B. DNA extraction			
1	Buccal swab	-	-
2	GeneJET™ Genomic DNA Purification Kit: - DNA Purification Column - Lysis solution - Wash buffer I (ethanol added) - Wash buffer II (ethanol added) - Elution buffer	Thermo Scientific™	-
3	Phosphate-buffered saline (PBS)	-	-
4	Proteinase K Solution RNA Grade	Invitrogen	-
5	Absolute ethanol	-	-
6	Micropipette (10 µL, 100 µL, 1000 µL)	Chemopharm/ DragonLab	-
7	Micropipette tips (10 µL, 200 µL, 1000 µL)	Axygen®	-
8	Shaking waterbath	Memmert	WNB14L4
9	Vortex mixer	Erla	EVM-6000
10	Spectrafuge 24D Starter Pack	Labnet	24000 (Spectrafuge 24D)
C. Agarose gel electrophoresis (1%)			
1	Extracted DNA	-	-

2	Agarose powder	-	-
3	Tris-borate-EDTA (TBE) buffer (10X)	-	-
4	Deionized water from Arium® Pro DI Ultrapure Water System	Sartorius	H20PRO-DI-T
5	Measuring cylinder (1000 mL)	Simax	-
6	Measuring cylinder (50 mL)	HmbG	-
7	Micropipette (10 µL)	DragonLab	-
8	Micropipette tips (10 µL)	Axygen®	-
9	Spatula	-	-
10	Scott bottle (250 mL)	Duran®	-
11	Parafilm	Bemis	-
12	DyeAll™ staining solution	GeneAll	-
13	VC Lambda/ HindIII marker	Vivantis	-
14	6X TriTrack DNA Loading dye	Thermo Scientific	-
15	Electrophoresis system	Owl	B1/B2
16	Microwave	Elba	EMO-1706
17	Power supply (Power Pack)	Bio-Rad	Power-Pac 3000V
18	UV transilluminator	Uvitec	BTS-26.MS
19	Weighing balance	Sartorius	BSA 224S-CW
D. DNA quantification (DeNovix)			
1	Extracted DNA	-	-
2	Elution buffer from GeneJET™ Genomic DNA Purification Kit	Thermo Scientific	-
3	Distilled water	-	-
4	Micropipette (10 µL)	Chemopharm/ DragonLab	-
5	Micropipette tips (10 µL)	Axygen®	-
6	Kimtech Science™ Kimwipes™ delicate task wipes	Kimberly-Clark Professional	-
7	Microvolume spectrophotometer	DeNovix	DS-11
E. Polymerase Chain Reaction (PCR)			
1	Extracted DNA	-	-
2	Herculase II Fusion DNA Polymerases: • 5X Herculase II Reaction Buffer • Herculase II Fusion DNA Polymerase • 100 mM dNTP Mix	Agilent	-
3	EDAR R2 (forward, reverse) primer (100 µM)	Integrated DNA Technologies (IDT)	-
4	Positive control	-	-
5	Nuclease free water	Thermo Scientific	-

6	Deionized water from Arium® Pro DI Ultrapure Water System	Sartorius	H20PRO-DI-T
7	PCR tubes (0.2 mL)	-	-
8	Micropipette (10 µL, 100 µL, 1000 µL)	Chemopharm/ DragonLab	-
9	Micropipette tips (10 µL, 200 µL, 1000 µL)	Axygen®	-
10	Vortex mixer	Erla	EVM-6000
11	Mini-centrifuge/ vortex	BioSan	Combispin FVL-2400N
12	Thermal cycler (Veriti)	Applied Biosystems	Veriti 9902
13	Laminar Air Flow Cabinet	Erla	CFM-4
F. Agarose Gel Electrophoresis (3%)			
1	PCR product	-	-
2	Agarose powder	-	-
3	Tris-borate-EDTA (TBE) buffer (10X)	-	-
4	Deionized water from Arium® Pro DI Ultrapure Water System	Sartorius	H20PRO-DI-T
5	Measuring cylinder (1000 mL)	Simax	-
6	Measuring cylinder (50 mL)	HmbG	-
7	Micropipette (10 µL)	DragonLab	-
8	Micropipette tips (10 µL)	Axygen®	-
9	Spatula	-	-
10	Scott bottle (250 mL)	Duran®	-
11	Parafilm	Bemis	-
12	DyeAll™ staining solution	GeneAll	-
13	GeneRuler 50 bp DNA Ladder	Thermo Scientific	-
14	6X Loading dye	Vivantis	-
15	Electrophoresis system	Owl	B1/B2
16	Microwave	Elba	EMO-1706
17	Power supply (Power Pack)	Bio-Rad	Power-Pac 3000V
18	UV transilluminator	Uvitec	BTS-26.MS
19	Weighing balance	Sartorius	BSA 224S-CW
G. DNA purification			
1	PCR product	-	-
2	PrimeWay Gel Extraction/ PCR purification Kit: • Gel/ PCR Column • Collection Tube • Wash Buffer (Ethanol added) • BD Buffer • Elution Buffer	1 st BASE	-

3	Microcentrifuge tube (1.5 mL)	Eppendorf	-
4	Micropipette (100 µL, 1000 µL)	Chemopharm/ DragonLab	-
5	Micropipette tips (200 µL, 1000 µL)	Axygen®	-
6	Spectrafuge 24D Starter Pack	Labnet	24000 (Spectrafuge 24D)
H. DNA quantification of purified PCR product (Qubit)			
1	Purified PCR product	-	-
2	Qubit™ dsDNA High Sensitivity (HS) Assay Kits: • Qubit™ dsDNA HS Buffer • Qubit™ dsDNA HS Reagent • Qubit™ dsDNA Standard #1 • Qubit™ dsDNA Standard #2	Invitrogen™	-
3	Assay tube (0.5 mL)	Eppendorf	-
4	Micropipette (10 µL, 100 µL, 1000 µL)	Chemopharm/ DragonLab	-
5	Micropipette tips (10 µL, 200 µL, 1000 µL)	Axygen®	-
6	Kimtech Science™ Kimwipes™ delicate task wipes	Kimberly-Clark Professional	-
7	Qubit® 2.0 Fluorometer	Life Technologies	Qubit 2.0
8	Vortex mixer	Erla	EVM-6000
9	Spectrafuge 24D Starter Pack	Labnet	24000 (Spectrafuge 24D)
I. Cycle sequencing			
1	Purified PCR product	-	-
2	BrilliantDye™ Terminator v3.1 Cycle Sequencing Kit: • 2.5X BrilliantDye™ RR Sequencing Premix • 5X Sequencing Buffer • pGEM Control • -21 M13 Primer	NimaGen	-
3	EDAR R2 forward primer (5 µM)	Integrated DNA Technologies (IDT)	-
4	Micropipette (10 µL, 100 µL, 1000 µL)	Chemopharm/ DragonLab	-
5	Micropipette tips (10 µL, 200 µL, 1000 µL)	Axygen®	-
6	Nuclease free water	Thermo Scientific	-
7	Thermal cycler (Veriti)	Applied Biosystems	Veriti 9902

J. Cycle Sequencing product purification (Ethanol precipitation)			
1	Cycle sequencing product	-	-
2	2M sodium acetate (pH 5.2- 5.5)	-	-
3	Ice-cold 95% ethanol	-	-
4	Ice-cold 70% ethanol	-	-
5	Highly deionized (Hi-Di) formamide	-	-
6	Micropipette (2 µL, 10 µL, 100 µL)	Chemopharm/ DragonLab	-
7	Micropipette tips (10 µL, 200 µL)	Axygen®	-
8	Refrigerated Microcentrifuge	Kubota	03-500
9	Thermal cycler (Veriti)	Applied Biosystems	Veriti 9902
10	Mini-centrifuge/ vortex	BioSan	Combispin FVL- 2400N
K. Sanger sequencing			
1	Purified cycle Sequencing product	-	-
2	Micropipette (10 µL)	DragonLab	-
3	Micropipette tips (10 µL)	Axygen®	-
4	MicroAmp® Optical 8-Tube Strip (0.2 mL)	Applied Biosystems®	-
5	SeqStudio™ Septa 8-Strip	Applied Biosystems®	-
6	Thermal cycler (Veriti)	Applied Biosystems	Veriti 9902
7	Plate spin	Kubota	Plate Spin II
8	SeqStudio™ Genetic Analyzer (DNA Sequencer Systems)	Applied Biosystems	SeqStudio™

3.2 Ethical Compliance

The study was conducted in accordance with approved protocols from the Human Research Ethics Committee (JEPeM) at Universiti Sains Malaysia (USM) Health Campus (USM/JEPeM/KK/23030218; Appendix A). Informed consent was obtained from all participants after they were informed about the purpose and procedure of the study.

3.3 Study Location, Population, and Sample

Buccal swabs and hair samples were collected from students who are Malaysian citizens aged 18 or above at USM. The buccal swab served as the DNA source for SNP analysis, while the hair samples were classified as straight, curly, or wavy to determine the phenotypic effect of SNP variations. A total of 30 samples were collected from three ethnic groups (Chinese, Malay, and Indian), with 10 for each ethnic group, comprising both female and male participants.

3.4 Subject Criteria

Participant consent was the primary eligibility requirement for sample collection in this study. Samples were collected from healthy adults aged 18 years or above who were born in Malaysia. Individuals with any genetic condition or a history of hair treatment were excluded from this study.

Figure 3.1 illustrated the overall experimental workflow of this study, from sample collection to data analysis, aiding in the understanding of the study methodology's progression.

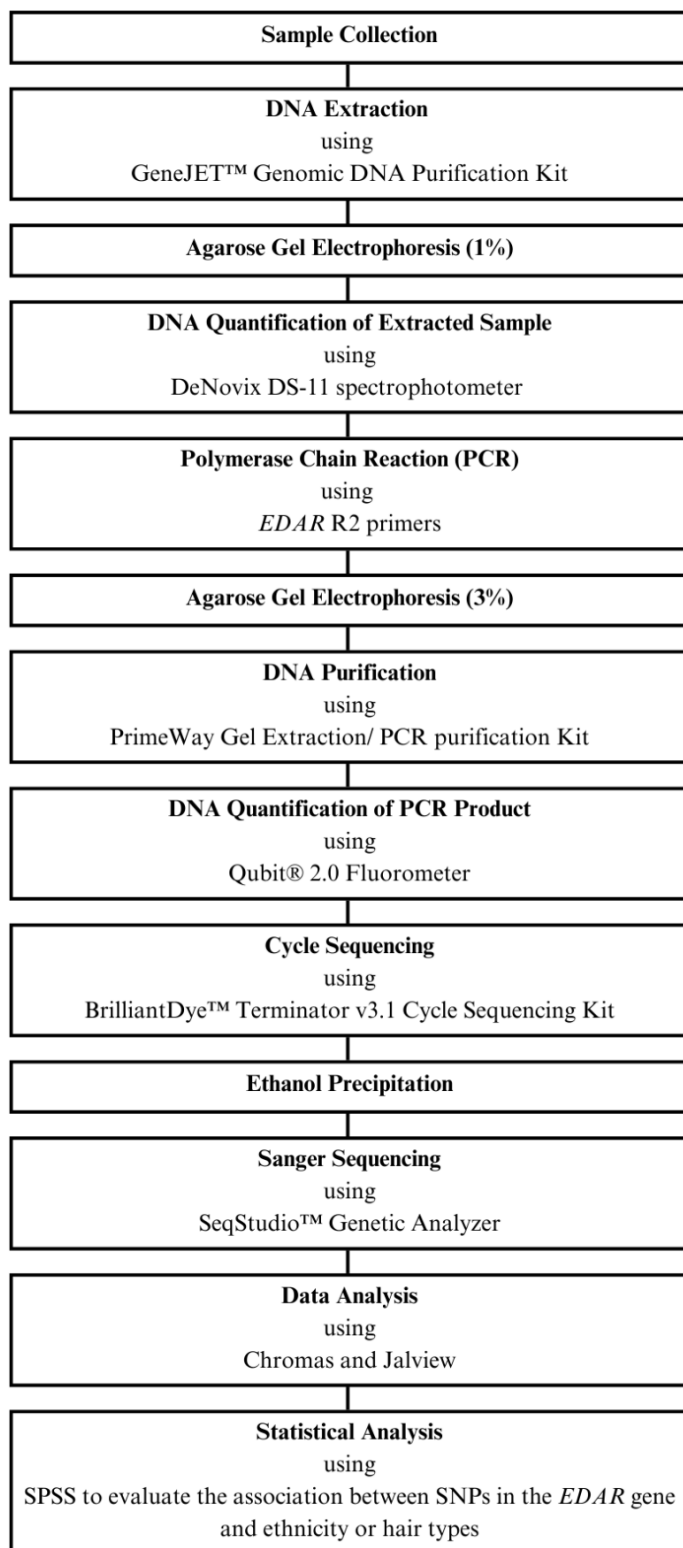


Figure 3.1: Workflow of the study.

3.5 Sample Collection

Before sample collection, the purpose and the procedure of the collection were explained to the participant, and his/her consent was obtained. A 1.5 mL microcentrifuge tube was labeled with the participant's name and the date of collection. A swab, without a transport tube, was removed from its packaging without touching the tip to any surface. The swab was inserted into one side of the mouth, between the inside of the cheek and the upper gum. The swab was pressed gently and twirled against the inside of the inner cheek using an up-and-down motion, moving from front to back and back to front, for at least 30 seconds per swab to ensure sufficient cell collection. The swab was removed from the participant and allowed to air dry for at least 10 minutes. The swab's superior absorption tip was cut using sterile scissors, then placed into the labeled microcentrifuge tube, and the lid was securely closed. The process was repeated on the other side of the cheek using a different swab. Simultaneously, either an image of the participant's scalp hair was taken, or a few strands of hair were collected, depending on the participant's willingness and the convenience of the sample collection location. The collected swab samples were stored at -20 °C until further processing, while the participants' hair types were classified as straight, wavy, or curly based on the Andre Walker Hair Typing System, as shown in **Table 3.2**.