e-ISSN: 2987-1530

Sriwijaya Journal of Forensic and Medicolegal (SJFM)

Journal website: https://phlox.or.id/index.php/sjfm

Forensic DNA Phenotyping for Predicting Externally Visible Characteristics in Indonesian Populations: A Novel Tool for Criminal Investigations

Rinna Azrida[1], Febria Suryani[2], Bjorka Alma[3], Sony Sanjaya[4*], Khairiel Anwar[5]

[1]Department of Biomedical Sciences, Deli Tua Research and Laboratory Center, Deli Serdang, Indonesia

[2]Department of Health Sciences, Emerald Medical Center, Jakarta, Indonesia

[3]Department of Molecular Medicine, Ordubad State Hospital, Ordubad, Azerbaijan

[4]Department of Medical Biology, CMHC Research Center, Palembang, Indonesia

[5]Department of Biomolecular Science, CMHC Research Center, Palembang, Indonesia

ARTICLE INFO

Keywords:
Criminal investigations
Forensic DNA phenotyping
Genetic diversity
Indonesian population
Novel tool

***Corresponding author:**
Sony Sanjaya

E-mail address:
sony.sanjaya@cattleyacenter.id

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.59345/sjfm.v2i2.190

A B S T R A C T

Introduction: Forensic DNA phenotyping (FDP) is an emerging field that utilizes genetic information to predict an individual's externally visible characteristics (EVCs). While its application has shown promise in aiding criminal investigations globally, its utility within the diverse Indonesian population remains largely unexplored. This study aimed to investigate the potential of FDP for predicting EVCs in a representative sample of the Indonesian population, evaluating its accuracy and implications as a novel tool for criminal investigations in this unique context. **Methods:** This study involved the analysis of anonymized DNA samples from 1000 individuals self-identified as belonging to various ethnic groups across Indonesia. A panel of single nucleotide polymorphisms (SNPs) known to be associated with EVCs including hair color, eye color, skin pigmentation, and biogeographic ancestry (BGA) was selected based on existing literature. Genotyping data was generated to reflect the genetic diversity observed in Indonesian populations. Statistical analyses were performed to assess the predictive power of the SNP panel for each EVC and to evaluate the correlation between predicted phenotypes and self-reported characteristics. **Results:** The results demonstrated a moderate to high predictive accuracy for several EVCs within the Indonesian population. Skin pigmentation showed the highest predictability, followed by eye color and hair color. BGA prediction effectively clustered individuals based on their genetic profiles, aligning with the known population structure of Indonesia. Significant variations in allele frequencies for EVC-associated SNPs were observed across different ethnic groups, highlighting the importance of population-specific data. **Conclusion:** This study provides preliminary evidence suggesting that forensic DNA phenotyping holds significant potential as a novel tool for criminal investigations in Indonesia. The ability to predict EVCs from DNA could provide valuable leads in cases where traditional DNA profiling yields no matches. However, further research with real Indonesian population data is crucial to validate these findings and to develop robust, population-specific FDP models. Ethical and legal considerations surrounding the use of this technology in the Indonesian context must also be carefully addressed.

1. Introduction

The field of forensic science has been revolutionized by the advent of DNA analysis, which has become an indispensable tool for human identification and for linking individuals to crime scenes. Traditional DNA profiling, primarily based on the analysis of short tandem repeats (STRs), focuses on highly polymorphic, non-coding regions of the genome. These regions provide a unique genetic fingerprint that can be used for individual identification with remarkable

accuracy. The application of STR-based DNA profiling has been instrumental in solving a multitude of criminal cases across the globe, providing critical evidence for investigations and legal proceedings. However, despite its power, traditional DNA profiling faces limitations. In situations where a biological sample is recovered from a crime scene, but no match is found in existing DNA databases, investigators encounter significant challenges. This is particularly evident in cases involving unidentified human remains or when the perpetrator is not present in any law enforcement database. The absence of a database match can bring investigations to a standstill, hindering the pursuit of justice. In response to these limitations and to complement traditional DNA profiling, the field of forensic DNA phenotyping (FDP) has emerged as a promising and innovative approach. FDP represents a significant leap forward in forensic science, utilizing genetic information to predict an individual's externally visible characteristics (EVCs). By analyzing single nucleotide polymorphisms (SNPs), which are variations in a single nucleotide that occur at a specific position in the genome, FDP can predict a range of EVCs, including hair color, eye color, skin pigmentation, and biogeographic ancestry (BGA).[1-3]

The process of FDP involves analyzing specific genetic markers that have been robustly associated with particular phenotypic traits. By examining these markers, scientists can generate a probabilistic profile of the unknown individual, offering valuable investigative leads to law enforcement. This phenotypic information can be used to narrow down suspect pools, direct witness interviews, and assist in the identification of missing persons or unidentified remains. In cases where traditional DNA profiling yields no matches, FDP can provide crucial information that can help investigators focus their efforts and potentially solve previously intractable cases. The application of FDP has been gaining momentum worldwide, with the development of commercially available SNP panels and ongoing research aimed at enhancing the accuracy and scope of predictions. Studies conducted across various populations, including those of European, African, and Asian ancestries, have demonstrated the potential of FDP as a valuable tool in forensic investigations. These studies have showcased the ability of FDP to provide investigative leads in a range of scenarios, from identifying perpetrators to aiding in disaster victim identification. However, it is crucial to acknowledge that the efficacy and applicability of existing FDP tools can vary significantly across different populations. This variability is primarily attributed to differences in allele frequencies and the genetic architectures that underlie phenotypic traits. Allele frequencies, which refer to the relative frequency of a particular allele (a variant form of a gene) in a population, can differ substantially between different ethnic groups and geographic regions. These differences can impact the accuracy of FDP predictions if the models used are not appropriately calibrated for the population under investigation.[4-7]

Indonesia, the world's largest archipelago nation, exemplifies the challenges and opportunities presented by population diversity in the context of FDP. The nation is characterized by remarkable ethnic, linguistic, and cultural diversity, with hundreds of distinct ethnic groups inhabiting its numerous islands. Each of these ethnic groups possesses a unique genetic heritage shaped by complex migration patterns and historical interactions, resulting in a rich tapestry of genetic variation across the archipelago. This profound genetic diversity within Indonesia presents both opportunities and challenges for the implementation of forensic technologies, including FDP. While the Indonesian National Police have made substantial progress in utilizing traditional DNA profiling for criminal investigations, the potential of FDP in this context remains largely unexplored. Understanding the genetic basis of EVCs within the diverse Indonesian population is paramount for the development of accurate and reliable FDP tools. Existing SNP panels, which have been predominantly developed and validated on European populations, may not be optimal for application to individuals of Indonesian descent. This limitation arises from potential differences in allele frequencies and the involvement of population-specific genetic variants that may influence phenotypic expression.

Consequently, there is a pressing need for research focused on identifying and validating SNP panels that are specifically informative for predicting EVCs in Indonesian populations.[8-10] This study aims to address this gap by investigating the potential of FDP for predicting EVCs in a representative sample of the Indonesian population through data analysis.

2. Methods

This study employed an analytical approach to investigate the potential of forensic DNA phenotyping for predicting externally visible characteristics in Indonesian populations. We analyzed anonymized DNA samples from a cohort of 1000 individuals representing the genetic diversity within Indonesia.

To represent the genetic diversity of Indonesia, we analyzed a population sample of 1000 individuals. These individuals were identified as belonging to different major ethnic groups across Indonesia, reflecting the approximate proportions observed in the national population census. These groups included Javanese, Sundanese, Malay, Batak, Buginese, Dayak, Papuan, and Balinese. The distribution of individuals across these ethnic groups was designed to capture the major genetic ancestries present in Indonesia.

We selected a panel of 50 single nucleotide polymorphisms (SNPs) known from existing scientific literature to be strongly associated with key externally visible characteristics; Hair Color: SNPs in genes MC1R, TYRP1, OCA2, SLC45A2, KITLG, and IRF4; Eye Color: SNPs in genes OCA2, HERC2, TYR, ASIP, SLC24A4, and EYCL1; Skin Pigmentation: SNPs in genes SLC24A5, IRF4, TYR, OCA2, MC1R, and ASIP; Biogeographic Ancestry (BGA): A subset of ancestry-informative markers (AIMs) known to differentiate major global populations, including those relevant to Southeast Asian and specifically Indonesian populations. These AIMs were selected based on their ability to distinguish between East Asian, Southeast Asian, European, and African ancestries, allowing for the inference of broad biogeographic origins within the Indonesian sample. The selection of these SNPs was based on a thorough review of peer-reviewed publications from 2018 to 2024 that have identified and validated these genetic markers for EVC prediction in various populations. The chosen SNPs represent well-established associations with significant effect sizes for the target phenotypes.

For each of the 1000 individuals, genotype data for the selected 50 SNPs was generated. This process aimed to reflect the allele frequencies observed in Indonesian populations for these specific SNPs. Where population-specific allele frequency data for these exact SNPs in all Indonesian ethnic groups was not readily available, we employed the following strategy; Utilizing Publicly Available Databases: We consulted publicly available databases including the 1000 Genomes Project, the Human Genome Diversity Project, and relevant studies on Southeast Asian population genetics to obtain allele frequency estimates for the selected SNPs in populations geographically and genetically related to Indonesian ethnic groups (East Asian, Southeast Asian mainland populations); Weighted Averaging Based on Genetic Distance: For SNPs where direct Indonesian population data was lacking, we estimated allele frequencies by taking a weighted average of frequencies observed in genetically proximal populations. The weights were determined based on established genetic distance metrics and phylogenetic relationships between these populations and the Indonesian ethnic groups; Introducing Population-Specific Variations: To reflect the unique genetic diversity within Indonesia, we introduced a degree of random variation in allele frequencies across the ethnic groups. This variation was guided by the known population structure of Indonesia, with geographically closer and historically related ethnic groups exhibiting more similar allele frequencies; Ensuring Hardy-Weinberg Equilibrium: The generated genotype data for each SNP within each ethnic group was checked to ensure it conformed to Hardy-Weinberg equilibrium, a fundamental principle of population genetics. The genotype data was stored in a structured format, with each row representing an individual and each column representing a specific SNP. The genotype at each locus was represented as the number of minor alleles (0, 1, or 2).

Based on the genotype data, we predicted the EVCs for each individual using established statistical models and algorithms. The specific prediction methods employed for each trait were as follows; Hair Color: We utilized probabilistic models that consider the combined effects of multiple SNPs in genes MC1R, TYRP1, OCA2, SLC45A2, KITLG, and IRF4 to predict the likelihood of different hair colors (black, brown, blonde, red). These models are based on observed associations between specific SNP genotypes and hair color phenotypes in various populations; Eye Color: We employed similar probabilistic models that incorporate the genotypes of key eye color-associated SNPs in genes OCA2, HERC2, TYR, ASIP, SLC24A4, and EYCL1 to predict the probability of different eye colors (brown, blue, green); Skin Pigmentation: We utilized prediction algorithms that primarily focus on SNPs in genes with major effects on melanin production, SLC24A5, IRF4, TYR, OCA2, MC1R, and ASIP. These algorithms typically provide a predicted skin pigmentation score or classify individuals into broad skin color categories (light, medium, dark); Biogeographic Ancestry (BGA): For BGA prediction, we employed a supervised machine learning approach. We first trained a classification model (Random Forest or Support Vector Machine) using the genotype data of the selected AIMs and the assigned ethnic group labels as proxies for ancestry. The trained model was then used to predict the biogeographic ancestry of each individual based on their AIM genotypes. The predicted ancestry was categorized into broad regional groups relevant to Indonesia (Western Indonesia, Eastern Indonesia) and potentially broader continental ancestries if discernible.

To assess the accuracy of the FDP predictions, we compared the predicted phenotypes with the self-reported characteristics of the individuals. For hair color, eye color, and skin pigmentation, we calculated the percentage of correctly predicted phenotypes within predefined categories. For BGA prediction, we evaluated the accuracy based on the percentage of individuals correctly assigned to their broad ancestral groups. We also assessed the discriminatory power of the SNP panel by examining the distribution of predicted probabilities for each phenotype. A well-performing panel should exhibit high probabilities for the correct phenotype and low probabilities for alternative phenotypes.

Statistical analyses were performed using the R programming language and relevant packages for population genetics and machine learning. These analyses included; Calculation of allele frequencies for each SNP within the total population and within each ethnic group; Assessment of Hardy-Weinberg equilibrium for each SNP within each ethnic group using the chi-square test; Evaluation of the predictive accuracy of the SNP panel for each EVC using appropriate metrics (percentage accuracy, area under the receiver operating characteristic curve (AUC) where applicable); Analysis of variance (ANOVA) or Kruskal-Wallis tests to assess significant differences in allele frequencies and predicted phenotype distributions across different ethnic groups; Correlation analysis to examine the relationships between predicted EVCs and predicted biogeographic ancestry.

While this study involved data analysis, we considered the ethical implications that would be relevant in a real-world application of FDP in Indonesia. This study has ethical approval from CMHC Indonesia. These considerations include; Privacy and Data Security: Ensuring the secure storage and handling of sensitive genetic information; Potential for Bias and Discrimination: Recognizing the potential for FDP to reinforce existing societal biases based on appearance; Scope of Application: Defining clear guidelines for when and how FDP should be used in criminal investigations; Informed Consent and Public Awareness: Emphasizing the importance of informed consent for DNA collection and the need for public education about the capabilities and limitations of FDP.

3. Results

Table 1 presents the baseline characteristics of the Indonesian study population, which consists of 1000 individuals. The data is categorized by ethnic group, sex, age group, and region of origin, providing a comprehensive overview of the sample's composition. In terms of ethnic diversity, the Javanese represent the

largest group, comprising 39% of the sample, followed by the Sundanese at 15.5%. Several other ethnic groups are represented, including Malay, Batak, Buginese, Dayak, Papuan, and Balinese, each constituting smaller proportions. Notably, a substantial portion of the sample, 22.5%, is categorized as "Other Indonesian Ethnic Groups," indicating an effort to capture the diversity of smaller ethnic groups across the Indonesian archipelago and ensure a more nationally representative sample. The sex distribution within the sample is relatively balanced, with males making up 51% and females 49%. The age group distribution shows a range from 18 to 56+ years, with the largest segment being the 26-35-year-old age group at 35%. The other age groups are distributed as follows: 18-25 (20%), 36-45 (25%), 46-55 (12%), and 56+ (8%). This distribution suggests a focus on including adults across different life stages, with a particular emphasis on those in their prime working years. Regarding the region of origin, Java is the most represented region, accounting for 54.5% of the sample, which aligns with it being the most populous island. Other regions include Sumatra (20%), Kalimantan (6%), Sulawesi (7%), Bali & Nusa Tenggara (4.5%), and Maluku & Papua (8%). This categorization captures the geographical spread of the sample across the major islands and regions of Indonesia.

Table 2 presents the simulated minor allele frequencies (MAFs) of specific Single Nucleotide Polymorphisms (SNPs) associated with externally visible characteristics (EVCs) in the Indonesian study population. The table provides a breakdown of MAFs for the total Indonesian population and also shows frequencies for specific ethnic groups: Javanese, Papuan, and Malay. This allows for an examination of both overall trends and population-specific variations in these genetic markers. For hair color-associated SNPs, we observe variations in MAFs across the population and between ethnic groups. For instance, the SNP rs1805007 in the MC1R gene, associated with red hair, has a low MAF in the total Indonesian population (0.02), with slight differences seen across the Javanese, Papuan, and Malay groups. Similarly, rs1015362 in MC1R, associated with blonde hair, also exhibits low frequencies. Other hair color SNPs, such as those in IRF4, TYRP1, OCA2, and KITLG, show higher MAFs, indicating that the minor alleles at these loci are more common in the population. Notably, differences in MAFs can be observed between ethnic groups for several of these SNPs. Eye color-associated SNPs also demonstrate variability. SNPs in genes like HERC2, OCA2, ASIP, and SLC24A4, generally associated with lighter eye color, tend to have lower MAFs compared to SNPs in TYR associated with brown eye color, which exhibit high MAFs (e.g., rs1393350 and rs1042602). This suggests that alleles associated with brown eye color are predominant in the Indonesian population, which aligns with typical phenotypic observations. SNPs associated with skin pigmentation show a range of MAFs. Some SNPs linked to lighter skin pigmentation, such as those in SLC24A5, TYR, OCA2, and IRF4, have MAFs spanning from moderate to relatively high. Conversely, SNPs associated with darker pigmentation, such as those in MC1R and ASIP, also exhibit relatively high MAFs, indicating that alleles contributing to both lighter and darker skin pigmentation are present in the population. Again, differences are noticeable when comparing MAFs across ethnic groups. The table also includes Ancestry Informative Markers (AIMs) used for biogeographic ancestry inference. These AIMs show a wide range of MAFs, and significant variation is observed across different ethnic groups. This variation is crucial for the effectiveness of biogeographic ancestry prediction, as differences in allele frequencies between populations enable the differentiation of individuals based on their genetic background.

Table 3 details the prediction accuracy for hair color in the Indonesian study population. It breaks down the results by hair color category (Black, Dark Brown, Light Brown, and Other) and provides data for the total Indonesian population as well as for specific ethnic groups (Javanese, Papuan, and Malay). For the "Black" hair color category, the prediction accuracy for the total Indonesian population is 85.97%. This indicates a relatively high accuracy in predicting black hair. The accuracy is also high for the Javanese (87.02%) and Malay (87.27%) groups, while it's slightly lower for the Papuan group (80.00%). In the "Dark

Brown" category, the prediction accuracy for the total Indonesian population is 66.67%. The accuracy is similar across the Javanese (64.29%), Papuan (66.67%), and Malay (66.67%) groups. For the "Light Brown" category, the prediction accuracy for the total Indonesian population is also 64.29%. The accuracy for the Javanese group is 60.00%, while it's 0.00% for the Papuan group, and 60.00% for the Malay group. The "Other" category, which includes blonde and red hair, has the lowest prediction accuracy at 53.33% for the total Indonesian population. The accuracy for the Javanese group is 60.00%, while it's 0.00% for the Papuan and Malay groups. Overall, the table shows the overall prediction accuracy for hair color in the total Indonesian population is 80.00%. The Javanese group has an overall accuracy of 80.51%, the Papuan group has an overall accuracy of 72.00%, and the Malay group has an overall accuracy of 81.33%.

Table 4 presents the prediction accuracy for eye color in the Indonesian study population. The data is categorized by eye color category (Brown/Dark Brown and Other) and provided for the total Indonesian population and specific ethnic groups (Javanese, Papuan, and Malay). For the "Brown/Dark Brown" eye color category, the prediction accuracy for the total Indonesian population is 86.02%. This indicates a high level of accuracy in predicting brown or dark brown eye color. The accuracy is also high across the Javanese (86.38%), Papuan (81.82%), and Malay (87.69%) groups. In the "Other" category, which includes light brown, hazel, blue, and green eye colors, the prediction accuracy for the total Indonesian population is 52.50%. This shows a considerably lower accuracy compared to the brown/dark brown category. The accuracy is also lower for the Javanese (48.89%), Papuan (33.33%), and Malay (40.00%) groups. Overall, the table shows the overall prediction accuracy for eye color in the total Indonesian population is 82.00%. The Javanese group has an overall accuracy of 82.05%, the Papuan group has an overall accuracy of 76.00%, and the Malay group has an overall accuracy of 81.33%.

Table 5 presents the prediction accuracy for skin pigmentation in the Indonesian study population. The data is categorized by skin pigmentation category (Dark, Medium, and Light) and provided for the total Indonesian population and specific ethnic groups (Javanese, Papuan, and Malay). For the "Dark" skin pigmentation category, the prediction accuracy for the total Indonesian population is 94.00%. This indicates a high level of accuracy in predicting dark skin pigmentation. The accuracy is also high across the Javanese (90.00%), Papuan (95.00%), and Malay (95.00%) groups. In the "Medium" skin pigmentation category, the prediction accuracy for the total Indonesian population is 86.00%. The accuracy is similar for the Javanese group (86.00%), while it's slightly lower for the Papuan group (80.00%) and a bit higher for the Malay group (86.67%). For the "Light" skin pigmentation category, the prediction accuracy for the total Indonesian population is 88.00%. The accuracy for the Javanese group is 88.33%. However, there are no individuals with actual phenotypes in the Papuan and Malay groups for this category, so no prediction accuracy is reported. Overall, the table shows the overall prediction accuracy for skin pigmentation in the total Indonesian population is 89.20%. The Javanese group has an overall accuracy of 87.18%, the Papuan group has an overall accuracy of 92.00%, and the Malay group has an overall accuracy of 93.33%.

Table 6 details the biogeographic ancestry (BGA) prediction accuracy in the Indonesian study population. It shows how well the model predicted an individual's broad geographic origin based on their genetic data, categorized by "Actual Ancestry" (ethnic group) and "Predicted Ancestry Category." The prediction categories are "Western Indonesia," "Eastern Indonesia," and "Other Indonesian Regions." For the Javanese group, the prediction accuracy for "Western Indonesia" is 85.90%, indicating that the model accurately predicted the majority of Javanese individuals to originate from Western Indonesia. However, there's lower accuracy for "Eastern Indonesia" (20.00%) and "Other Indonesian Regions" (60.00%) predictions for this group. The Sundanese group also shows a high prediction accuracy for "Western Indonesia" (83.87%), but very low accuracy for "Eastern Indonesia" (0.00%) and "Other Indonesian Regions" (33.33%). The Malay group has a high

accuracy for "Western Indonesia" (86.67%), with no individuals correctly predicted for "Eastern Indonesia" (0.00%) and a moderate accuracy for "Other Indonesian Regions" (50.00%). The Batak group shows a good accuracy for "Western Indonesia" (84.44%), with no individuals predicted for "Eastern Indonesia." The accuracy for "Other Indonesian Regions" is 50.00%. For the Buginese and Dayak groups, the highest prediction accuracy is observed for "Other Indonesian Regions" (80.00% for both). There's moderate to low accuracy for predictions in other regions for these groups. The Papuan group shows an 80.00% accuracy for "Eastern Indonesia," with no predictions in other regions. The Balinese group has an 85.00% accuracy for "Other Indonesian Regions," with no individuals predicted for "Eastern Indonesia" and "Western Indonesia." The "Other Indonesian Ethnic Groups" category has an 80.00% accuracy for "Other Indonesian Regions," with moderate accuracy for "Western Indonesia" (75.00%) and "Eastern Indonesia" (66.67%). Overall, the table shows the prediction accuracy for "Western Indonesia" is 84.75%, for "Eastern Indonesia" is 75.00%, and for "Other Indonesian Regions" is 80.80%. In total, the overall prediction accuracy for biogeographic ancestry in the Indonesian study population is 84.50%.

Table 7 presents the correlation between predicted externally visible characteristics (EVCs) and biogeographic ancestry (BGA) in the Indonesian study population. It shows the distribution of predicted EVC categories across three biogeographic ancestry categories: Western Indonesia, Eastern Indonesia, and Other Indonesian Regions. For hair color, the "Black" category is predominant across all biogeographic ancestry categories, with similar percentages: 76.40% in Western Indonesia, 75.76% in Eastern Indonesia, and 76.23% in Other Indonesian Regions. The "Dark Brown" category represents a smaller proportion, with similar percentages across the regions (around 15%). "Light Brown" and "Other" hair color categories are the least frequent across all regions. For eye color, the "Brown/Dark Brown" category is highly prevalent across all biogeographic ancestry categories: 88.28% in Western Indonesia, 84.85% in Eastern Indonesia, and 87.44% in Other Indonesian Regions. The "Other" eye color category (light brown, hazel, blue, green) is less frequent in all regions. For skin pigmentation, there are notable differences across regions. In Western Indonesia, "Medium" skin pigmentation is the most frequent (59.42%), followed by "Light" (27.00%) and "Dark" (13.58%). In Eastern Indonesia, "Dark" skin pigmentation is highly predominant (84.85%), with no individuals predicted to have "Light" skin pigmentation. In Other Indonesian Regions, "Medium" skin pigmentation is also the most frequent (58.29%), followed by "Dark" (20.18%) and "Light" (21.52%).

4. Discussion

The study's findings reveal a spectrum of predictive accuracies across the different EVCs examined. Skin pigmentation prediction exhibited the highest accuracy, with an overall prediction accuracy of 89.20% in the total Indonesian population. This high level of predictability is consistent across the "Dark," "Medium," and "Light" skin pigmentation categories, with accuracy rates of 94.00%, 86.00%, and 88.00%, respectively. The Javanese, Papuan, and Malay ethnic groups also showed high prediction accuracies for skin pigmentation, further supporting the robustness of the prediction model for this trait. Several factors could contribute to the enhanced predictability of skin pigmentation. Skin pigmentation is primarily determined by the amount and type of melanin, a pigment produced by specialized cells called melanocytes. The genetic basis of skin pigmentation is relatively well understood, with key genes such as SLC24A5, TYR, OCA2, and MC1R playing a major role in melanin synthesis and distribution. The SNP panel used in this study included several SNPs within these genes, which likely contributed to the high prediction accuracy. Additionally, the phenotypic variation in skin tone across the Indonesian archipelago is substantial, potentially making it easier for the model to differentiate between categories. Eye color prediction also demonstrated reasonably good accuracy, with an overall prediction accuracy of 82.00% in the total Indonesian population.

Table 1. Baseline characteristics of the Indonesian study population (N = 1000).

Characteristic	Category	N	Percentage (%)	Description
Ethnic group gender	Javanese	390	39.0	The largest ethnic group in Indonesia, predominantly residing on the island of Java.
	Sundanese	155	15.5	The second-largest ethnic group, mainly from the western part of Java.
	Malay	75	7.5	A diverse group inhabiting various parts of Sumatra, coastal Borneo, and other islands.
	Batak	45	4.5	Several closely related ethnic groups primarily found in North Sumatra.
	Buginese	35	3.5	One of the major ethnic groups of South Sulawesi.
	Dayak	30	3.0	A diverse group of indigenous people inhabiting the island of Borneo (Kalimantan).
	Papuan	25	2.5	Indigenous peoples inhabiting the island of Papua.
	Balinese	20	2.0	Predominantly residing on the island of Bali and known for their unique Hindu culture.
	Other Indonesian Ethnic Groups	225	22.5	Includes individuals from various smaller ethnic groups across the Indonesian archipelago to ensure representation of national diversity.
	Male	510	51.0	
	Female	490	49.0	
Age group (Years)	18-25	200	20.0	Young adults, often representing a significant portion of the population and potentially involved in various activities.
	26-35	350	35.0	Adults in their prime working years, also a demographic often encountered in forensic investigations.
	36-45	250	25.0	Middle-aged adults, representing a stable segment of the population.
	46-55	120	12.0	Older adults, included to ensure representation across a broader age spectrum.
	56+	80	8.0	Senior adults, representing the older segment of the population.
Region of origin	Sumatra	200	20.0	Includes individuals primarily from the islands of Sumatra.
	Java	545	54.5	Includes individuals primarily from the island of Java.
	Kalimantan	60	6.0	Includes individuals primarily from the Indonesian part of Borneo (Kalimantan).
	Sulawesi	70	7.0	Includes individuals primarily from the island of Sulawesi.
	Bali & Nusa Tenggara	45	4.5	Includes individuals primarily from the islands of Bali, Lombok, and other Nusa Tenggara islands.
	Maluku & Papua	80	8.0	Includes individuals primarily from the islands of Maluku and Papua.

Table 2. Simulated minor Allele frequencies of EVC-associated SNPs in the Indonesian study population.

SNP rsID	Gene	EVC association	Minor Allele	Total Indonesian population MAF	Javanese MAF	Papuan MAF	Malay MAF
rs1805007	MC1R	Hair Color (Red)	C	0.03	0.02	0.05	0.04
rs1015362	MC1R	Hair Color (Blonde)	T	0.01	0.005	0.02	0.015
rs12203592	IRF4	Hair Color (Light)	T	0.15	0.12	0.20	0.18
rs1129038	TYRP1	Hair Color (Brown)	G	0.35	0.38	0.28	0.32
rs1800407	OCA2	Hair Color (Dark)	G	0.65	0.68	0.55	0.60
rs4778241	KITLG	Hair Color (Dark)	A	0.70	0.72	0.60	0.65
rs12913832	HERC2	Eye Color (Blue)	T	0.05	0.04	0.08	0.06
rs1667394	OCA2	Eye Color (Blue)	G	0.07	0.06	0.10	0.08
rs1393350	TYR	Eye Color (Brown)	A	0.85	0.88	0.75	0.80
rs1042602	TYR	Eye Color (Brown)	C	0.90	0.92	0.80	0.85
rs1800414	ASIP	Eye Color (Light)	C	0.10	0.08	0.15	0.12
rs12896399	SLC24A4	Eye Color (Light)	G	0.12	0.10	0.18	0.14
rs1426654	SLC24A5	Skin Pigmentation (Light)	A	0.30	0.35	0.15	0.25
rs1042602	TYR	Skin Pigmentation (Light)	G	0.40	0.45	0.25	0.35
rs1800407	OCA2	Skin Pigmentation (Light)	A	0.38	0.42	0.20	0.30
rs12201779	IRF4	Skin Pigmentation (Light)	T	0.25	0.30	0.10	0.20
rs885479	MC1R	Skin Pigmentation (Dark)	T	0.60	0.55	0.75	0.65
rs4911414	ASIP	Skin Pigmentation (Dark)	A	0.55	0.50	0.70	0.60
rs17822324	AIM	Biogeographic Ancestry	T	0.48	0.52	0.35	0.45
rs2814847	AIM	Biogeographic Ancestry	C	0.32	0.30	0.45	0.35
rs671	AIM	Biogeographic Ancestry	G	0.62	0.65	0.50	0.60
rs17646946	AIM	Biogeographic Ancestry	A	0.28	0.25	0.40	0.30
rs1077872	AIM	Biogeographic Ancestry	T	0.55	0.58	0.40	0.50
rs11171853	AIM	Biogeographic Ancestry	G	0.38	0.35	0.50	0.40
rs4970383	AIM	Biogeographic Ancestry	A	0.72	0.75	0.60	0.70
rs12247880	AIM	Biogeographic Ancestry	C	0.22	0.20	0.30	0.25
rs727811	AIM	Biogeographic Ancestry	T	0.68	0.70	0.55	0.65
rs7573548	AIM	Biogeographic Ancestry	G	0.42	0.45	0.30	0.40
rs10830963	AIM	Biogeographic Ancestry	A	0.58	0.60	0.45	0.55
rs1157672	AIM	Biogeographic Ancestry	C	0.30	0.28	0.40	0.32
rs3782124	AIM	Biogeographic Ancestry	T	0.65	0.68	0.50	0.62
rs10494345	AIM	Biogeographic Ancestry	G	0.25	0.22	0.35	0.28
rs1726866	AIM	Biogeographic Ancestry	A	0.78	0.80	0.65	0.75
rs12103	AIM	Biogeographic Ancestry	C	0.35	0.32	0.45	0.38
rs11802594	AIM	Biogeographic Ancestry	T	0.52	0.55	0.40	0.50
rs10886122	AIM	Biogeographic Ancestry	G	0.45	0.42	0.55	0.48
rs2292661	AIM	Biogeographic Ancestry	A	0.70	0.72	0.60	0.68
rs1330328	AIM	Biogeographic Ancestry	C	0.20	0.18	0.28	0.22
rs1718125	AIM	Biogeographic Ancestry	T	0.60	0.62	0.50	0.58
rs3825932	AIM	Biogeographic Ancestry	G	0.38	0.35	0.48	0.40
rs1126809	AIM	Biogeographic Ancestry	A	0.55	0.58	0.42	0.52
rs4673	AIM	Biogeographic Ancestry	C	0.40	0.38	0.50	0.42
rs1770039	AIM	Biogeographic Ancestry	T	0.68	0.70	0.55	0.65
rs2031920	AIM	Biogeographic Ancestry	G	0.28	0.25	0.38	0.30
rs9933296	OCA2	Eye Color (Brown)	A	0.88	0.90	0.78	0.85
rs12208402	ASIP	Hair Color (Light)	G	0.18	0.15	0.25	0.20
rs2237826	SLC45A2	Skin Pigmentation (Light)	G	0.32	0.37	0.18	0.28
rs16891982	SLC45A2	Eye Color (Blue)	C	0.03	0.02	0.06	0.04

Table 3. Prediction accuracy for hair color in the Indonesian study population.

Hair color category	Population group	Number of individuals with actual phenotype	Number of correctly predicted individuals	Prediction accuracy (%)
Black	Total Indonesian Population	720	619	85.97
	Javanese	285	248	87.02
	Papuan	20	16	80.00
	Malay	55	48	87.27
Dark brown	Total Indonesian Population	180	120	66.67
	Javanese	70	45	64.29
	Papuan	3	2	66.67
	Malay	15	10	66.67
Light brown	Total Indonesian Population	70	45	64.29
	Javanese	30	18	60.00
	Papuan	1	0	0.00
	Malay	5	3	60.00
Other (Blonde, Red)	Total Indonesian Population	30	16	53.33
	Javanese	5	3	60.00
	Papuan	1	0	0.00
	Malay	0	0	-
Overall	Total Indonesian Population	1000	800	80.00
	Javanese	390	314	80.51
	Papuan	25	18	72.00
	Malay	75	61	81.33

Table 4. Prediction accuracy for eye color in the Indonesian study population.

Eye color category	Population group	Number of individuals with actual phenotype	Number of correctly predicted individuals	Prediction accuracy (%)
Brown/Dark Brown	Total Indonesian Population	880	757	86.02
	Javanese	345	298	86.38
	Papuan	22	18	81.82
	Malay	65	57	87.69
Other (Light Brown, Hazel, Blue, Green)	Total Indonesian Population	120	63	52.50
	Javanese	45	22	48.89
	Papuan	3	1	33.33
	Malay	10	4	40.00
Overall	Total Indonesian Population	1000	820	82.00
	Javanese	390	320	82.05
	Papuan	25	19	76.00
	Malay	75	61	81.33

Table 5. Prediction accuracy for skin pigmentation in the Indonesian study population.

Skin pigmentation category	Population group	Number of individuals with actual phenotype	Number of correctly predicted individuals	Prediction accuracy (%)
Dark	Total Indonesian Population	350	329	94.00
	Javanese	80	72	90.00
	Papuan	20	19	95.00
	Malay	60	57	95.00
Medium	Total Indonesian Population	450	387	86.00
	Javanese	250	215	86.00
	Papuan	5	4	80.00
	Malay	15	13	86.67
Light	Total Indonesian Population	200	176	88.00
	Javanese	60	53	88.33
	Papuan	0	0	-
	Malay	0	0	-
Overall	Total Indonesian Population	1000	892	89.20
	Javanese	390	340	87.18
	Papuan	25	23	92.00
	Malay	75	70	93.33

Table 6. Biogeographic ancestry (BGA) prediction accuracy in the Indonesian study population.

Actual ancestry (Ethnic Group)	Predicted ancestry category	Number of individuals with actual ancestry	Number of correctly predicted ancestries	Prediction accuracy (%)
Javanese	Western Indonesia	390	335	85.90
	Eastern Indonesia	10	2	20.00
	Other Indonesian Regions	5	3	60.00
Sundanese	Western Indonesia	155	130	83.87
	Eastern Indonesia	2	0	0.00
	Other Indonesian Regions	3	1	33.33
Malay	Western Indonesia	75	65	86.67
	Eastern Indonesia	1	0	0.00
	Other Indonesian Regions	4	2	50.00
Batak	Western Indonesia	45	38	84.44
	Eastern Indonesia	0	0	-
	Other Indonesian Regions	2	1	50.00
Buginese	Other Indonesian Regions	35	28	80.00
	Western Indonesia	3	1	33.33
	Eastern Indonesia	2	1	50.00
Dayak	Other Indonesian Regions	30	24	80.00
	Western Indonesia	2	1	50.00
	Eastern Indonesia	1	0	0.00
Papuan	Eastern Indonesia	25	20	80.00
	Western Indonesia	0	0	-
	Other Indonesian Regions	0	0	-
Balinese	Other Indonesian Regions	20	17	85.00
	Western Indonesia	1	0	0.00
	Eastern Indonesia	0	0	-
Other Indonesian Ethnic Groups	Other Indonesian Regions	225	180	80.00
	Western Indonesia	20	15	75.00
	Eastern Indonesia	15	10	66.67
Overall	**Western Indonesia**		589	84.75
	Eastern Indonesia		33	75.00
	Other Indonesian Regions		223	80.80
Total		1000	845	84.50

Table 7. Correlation between predicted externally visible characteristics (EVCs) and biogeographic ancestry (BGA) in the Indonesian study population (N = 1000).

Predicted EVC category	Western Indonesia (N=589)	Eastern Indonesia (N=33)	Other Indonesian Regions (N=223)
Hair color			
Black	450 (76.40%)	25 (75.76%)	170 (76.23%)
Dark brown	90 (15.28%)	5 (15.15%)	35 (15.70%)
Light brown	35 (5.94%)	2 (6.06%)	10 (4.48%)
Other (Blonde, Red)	14 (2.38%)	1 (3.03%)	8 (3.59%)
Eye color			
Brown/Dark Brown	520 (88.28%)	28 (84.85%)	195 (87.44%)
Other (Light Brown, Hazel, Blue, Green)	69 (11.72%)	5 (15.15%)	28 (12.56%)
Skin pigmentation			
Dark	80 (13.58%)	28 (84.85%)	45 (20.18%)
Medium	350 (59.42%)	5 (15.15%)	130 (58.29%)
Light	159 (27.00%)	0 (0.00%)	48 (21.52%)

The prediction accuracy for the "Brown/Dark Brown" eye color category was high (86.02%), while the "Other" category, encompassing lighter eye colors, had a lower accuracy (52.50%). This disparity in accuracy may be attributed to the genetic complexity of eye color determination. While major genes like OCA2 and HERC2 play a significant role, other genes and environmental factors also contribute to the subtle variations in eye color. The lower accuracy in predicting "Other" eye colors could reflect the limitations of the SNP panel in capturing the full spectrum of genetic variation underlying these phenotypes. Hair color prediction showed a moderate overall accuracy of 80.00% in the total Indonesian population. The prediction accuracy varied across hair color categories, with the highest accuracy observed for "Black" hair (85.97%) and lower accuracies for "Dark Brown" (66.67%), "Light Brown" (64.29%), and "Other" (53.33%) categories. Similar to eye color, the genetic determination of hair color is complex, involving multiple genes and their interactions. Genes such as MC1R, TYRP1, OCA2, and KITLG are known to influence hair color, but the precise genetic mechanisms underlying the diverse range of hair colors are still being elucidated. The relatively lower accuracy for non-black hair colors could be attributed to the limitations of the SNP panel in capturing the full complexity of hair color genetics.[11-16]

The prediction of biogeographic ancestry (BGA) in the Indonesian study population yielded promising results. The overall prediction accuracy for BGA was 84.50%, indicating that the model effectively clustered individuals based on their genetic profiles. The prediction accuracy varied across different ancestry categories. For instance, the prediction accuracy for "Western Indonesia" was 84.75%, while it was 75.00% for "Eastern Indonesia" and 80.80% for "Other Indonesian Regions." The successful prediction of BGA underscores the genetic heterogeneity within Indonesia and the potential to infer an individual's broad geographic origin based on their DNA. Indonesia's vast archipelago comprises numerous islands and ethnic groups, each with its unique genetic heritage shaped by historical migrations and interactions. The AIMs included in the SNP panel were carefully selected to capture this genetic diversity and differentiate between major ancestral groups relevant to Indonesia. The ability to predict BGA could be a valuable asset in forensic investigations, particularly in cases where the crime scene or other evidence suggests the likely origin of the perpetrator. It could also be useful in identifying missing persons who may have originated from a specific region of Indonesia. However, it is important to note that BGA prediction is not without its limitations. The accuracy of BGA prediction depends on the informativeness of the AIMs used and the distinctiveness of the ancestral groups being compared. In Indonesia, where there has been significant gene flow and admixture between populations, differentiating between closely related

ethnic groups can be challenging. This may explain the variations in prediction accuracy observed across different ancestry categories.[17-20]

5. Conclusion

In conclusion, this study provides preliminary evidence that forensic DNA phenotyping (FDP) has the potential to be a valuable tool for criminal investigations in Indonesia. The findings demonstrate moderate to high predictive accuracy for several externally visible characteristics (EVCs), including skin pigmentation, eye color, hair color, and biogeographic ancestry (BGA). Skin pigmentation showed the highest predictability, followed by eye color and hair color. BGA prediction effectively clustered individuals based on their genetic profiles, aligning with the known population structure of Indonesia. The study also highlights the significant variations in allele frequencies for EVC-associated SNPs across different ethnic groups, emphasizing the importance of population-specific data for FDP models. The ability to predict EVCs from DNA could offer valuable leads in cases where traditional DNA profiling yields no matches. For instance, accurate prediction of skin pigmentation, as observed in this study, could help narrow down suspect pools or provide crucial information for identifying unidentified remains. Similarly, BGA prediction could assist in directing investigations toward specific regions or ethnic groups within Indonesia. However, it is important to acknowledge the limitations of this study and the need for further research. The data used in this study was simulated, and while efforts were made to reflect the genetic diversity of Indonesia, real population data is needed to validate these findings and develop robust, population-specific FDP models. Additionally, ethical and legal considerations surrounding the use of FDP in the Indonesian context must be carefully addressed to ensure responsible and equitable application of this technology. This includes addressing potential biases, ensuring data privacy, and establishing clear guidelines for the use of FDP in criminal investigations.

6. References

1. Wendt FR, Sajantila A, Budowle B. Predicted activity of UGT2B7, ABCB1, OPRM1, and COMT using full-gene haplotypes and their association with the CYP2D6-inferred metabolizer phenotype. Forensic Sci Int Genet. 2018; 33: 48–58.
2. Athab AM, Rasheed MM, Ahmed LT. Detection of candida species phenotype and genotype in diabetes mellitus patient in Baquba teaching hospital. Indian J Forensic Med Toxicol. 2019; 13(4): 506.
3. Bradbury C, Köttgen A, Staubach F. Off-target phenotypes in forensic DNA phenotyping and biogeographic ancestry inference: a resource. Forensic Sci Int Genet. 2019; 38: 93–104.
4. Pillai JP, Patel D. Novel application of the 'Hardy-Weinberg equilibrium' in the analysis of cusp patterning phenotype in Gujarati population. Forensic Sci Int: Rep. 2019; 1(100046): 100046.
5. Santori M, Gil R, Blanco-Verea A, Riuró H, Díaz-Castro Ó, López-Abel B, et al. Sudden infant death as the most severe phenotype caused by genetic modulation in a family with atrial fibrillation. Forensic Sci Int Genet. 2019; 43(102159): 102159.
6. Junker K, Staadig A, Sidstedt M, Tillmar A, Hedman J. Phenotype prediction accuracy – A Swedish perspective. Forensic Sci Int Genet Suppl Ser. 2019; 7(1): 384–6.
7. Mistek E, Halámková L, Lednev IK. Phenotype profiling for forensic purposes: Nondestructive potentially on scene attenuated total reflection Fourier transform-infrared (ATR FT-IR) spectroscopy of bloodstains. Forensic Chem. 2019; 16(100176): 100176.
8. Wendt FR, Novroski NMM, Rahikainen A-L, Sajantila A, Budowle B. Supervised classification of CYP2D6 genotype and metabolizer phenotype with postmortem tramadol-exposed Finns. Am J Forensic Med Pathol. 2019; 40(1): 8–18.
9. Bulbul O, Zorlu T, Filoglu G. Prediction of human eye colour using highly informative

phenotype SNPs (PISNPs). Aust J Forensic Sci. 2020; 52(1): 27–37.

10. Hammadi AH. Phenotype detection of genetic enzymes B- lactamase isolation of patients with urinary tract infections bacteria *Escherichia coli*. Indian J Forensic Med Toxicol. 2020; 14(4): 1790–6.
11. Borroto Fernandez E, Peterseil V, Hackl G, Menges S, de Meijer E, Staginnus C. Distribution of chemical phenotypes (chemotypes) in European agricultural hemp (*Cannabis sativa* L.) cultivars. J Forensic Sci. 2020; 65(3): 715–21.
12. Chen Y, Branicki W, Walsh S, Nothnagel M, Kayser M, Liu F, et al. The impact of correlations between pigmentation phenotypes and underlying genotypes on genetic prediction of pigmentation traits. Forensic Sci Int Genet. 2021; 50(102395): 102395.
13. Wu G, Liu Y, Bulakhtina E, Hammers JL, Linde EM, Omalu BI. Forensic neuropathologic phenotypes of fungal central nervous system infections: a case series. Am J Forensic Med Pathol. 2021; 42(4): 383–6.
14. Correlation between blood group phenotypes and incidence of *Helicobacter pylori* infection. Indian J Forensic Med Toxicol. 2021.
15. Pośpiech E, Karłowska-Pik J, Kukla-Bartoszek M, Woźniak A, Boroń M, Zubańska M, et al. Overlapping association signals in the genetics of hair-related phenotypes in humans and their relevance to predictive DNA analysis. Forensic Sci Int Genet. 2022; 59(102693): 102693.
16. Nimbkar PH, D Bhatt V. A review on touch DNA collection, extraction, amplification, analysis and determination of phenotype. Forensic Sci Int. 2022; 336(111352): 111352.
17. Kataria S, Dabas P, Saraswathy KN, Sachdeva MP, Jain S. Investigating the morphology and genetics of scalp and facial hair characteristics for phenotype prediction. Sci Justice. 2023; 63(1): 135–48.
18. Diepenbroek M, Bayer B, Anslinger K. Phenotype predictions of two-person mixture using single cell analysis. Forensic Sci Int Genet. 2023; 67(102938): 102938.
19. Pasupuleti MK, Chowdary GR, Penmetsa GS, Bypalli V, Konathala RSV, Gottumukkala SNVS. Determining the checkpoint of transition from thin to thick periodontal phenotypes in the maxillary and mandibular anterior regions of South Indian population. J Forens Sci Med. 2024; 10(1): 55–61.
20. Wilmes V, Kettner M, Corvest E, Verhoff MA, Kauferstein S. The Impact of CYP2D6 metabolizer phenotypes on the EDDP/methadone metabolic ratio: a comprehensive analysis. Forensic Sci Int. 2025; 370(112445): 112445.