

Original Research Article

Forensic DNA Phenotyping of Externally Visible Characteristics and Biogeographic Ancestry in Indonesian Populations: A Cross-Sectional Validation Study

Rinna Azrida¹, Febria Suryani², Bjorka Alma³, Sony Sanjaya^{4*}, Khairiel Anwar⁵¹Department of Biomedical Sciences, Deli Tua Research and Laboratory Center, Deli Serdang, Indonesia²Department of Health Sciences, Emerald Medical Center, Jakarta, Indonesia³Department of Molecular Medicine, Ordubad State Hospital, Ordubad, Azerbaijan⁴Department of Medical Biology, CMHC Research Center, Palembang, Indonesia⁵Department of Biomolecular Science, CMHC Research Center, Palembang, Indonesia

ARTICLE INFO

Article history

Received: September 20, 2024

Revised: October 15, 2024

Accepted: November 10, 2024

Available online: November 27, 2024

Published: November 30, 2024

Keywords

Biogeographic ancestry
Externally visible characteristics
Forensic DNA phenotyping
Indonesia
Wallace line

Corresponding author

Sony Sanjaya

E-mail: sony.sanjaya@cattleyacenter.id

ORCID: 0009-0002-3044-4481

DOI

[10.59345/sjfm.v2i2.366](https://doi.org/10.59345/sjfm.v2i2.366)

ABSTRACT

Background. Forensic DNA phenotyping (FDP) infers externally visible characteristics (EVCs) and biogeographic ancestry (BGA) from DNA when no reference profile is available. How well it performs in Indonesian populations has been little characterised.

Objective. To evaluate how accurately a single nucleotide polymorphism (SNP) panel predicts skin, iris and hair colour and broad regional ancestry in Indonesian adults.

Methods. In this cross-sectional study, 1,000 unrelated adults from eight named ethnic groups and a composite group were genotyped at 20 pigmentation SNPs and 26 ancestry-informative markers by multiplex single-base extension. Colour categories were predicted with a multinomial logistic model calibrated to HIrisPlex-S, and ancestry zones (Western, Central/other, Eastern) with a random forest. Accuracy was compared with the no-information rate, the accuracy of always predicting the commonest class.

Results. Skin colour was predicted correctly in 89.2% of participants (no-information rate 45.0%), iris colour in 82.0% (88.0%) and hair colour in 80.0% (72.0%). The area under the curve was 0.96 for dark skin, 0.88 for brown irises and 0.87 for black hair; sensitivity for non-brown irises was 52.5%. Overall, 83.7% of participants were assigned to their expected ancestry zone (no-information rate 66.5%); sensitivity for the Eastern zone was 80.0% (95% CI 60.9–91.1; n = 25).

Conclusion. In this cohort, skin colour prediction clearly exceeded the baseline, whereas iris and hair colour prediction improved little on it. The panel may support investigative leads, but independent test samples are needed before casework use.

Copyright. © 2024 The Author(s). Published by Phlox Institute: Indonesian Medical Research Organization, on behalf of the Sriwijaya Journal of Forensic and Medicolegal. The authors retain copyright of their work.

Access licence. This is an open access article distributed under the terms of the Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International Licence (CC BY-NC-SA 4.0), which permits non-commercial use, sharing, adaptation and reproduction in any medium, provided the original author and source are credited and any derivative work is distributed under the same licence.

1. Introduction

Short tandem repeat (STR) profiling is the standard method for human identification, but it is comparative: it needs a reference sample or a database entry to match.^{1,2} When a crime-scene trace matches no suspect or database record, STR typing gives no information about the donor. Forensic DNA phenotyping (FDP) infers externally visible characteristics (EVCs), mainly eye, hair and skin colour, and biogeographic ancestry (BGA) from single nucleotide polymorphisms (SNPs), and can provide an investigative lead rather than evidence of identity.^{1,2}

The HIrisPlex-S system predicts eye, hair and skin colour from a defined set of SNPs with published model parameters.^{3–5} The VISAGE tools extend this approach to sequencing-based panels for appearance and ancestry.^{6,7} These tools were developed and tested mainly in European

and admixed samples. Pigmentation varies between world populations,⁸ and accuracy can change when a panel is applied to an admixed population.⁹ Iris and hair colour categories are also hard to predict when most people share one dark phenotype.^{10,11}

Indonesia is an archipelago of many ethnic groups whose ancestry changes across Wallacea, between Sundaland in the west and Sahul in the east.^{12–14} Forensic genetics in Indonesia relies mostly on STR typing, and information on how pigmentation and ancestry panels perform across Indonesian ethnic groups is limited.

The objective of this study was to evaluate the accuracy of a panel of 20 pigmentation SNPs and 26 ancestry-informative markers (AIMs) for predicting skin, iris and hair colour and broad regional ancestry in 1,000 Indonesian adults, and to compare each accuracy with the no-information rate.

2. Methods

2.1. Study design and ethical approval

This was a cross-sectional study reported in line with the STROBE statement¹⁵, approved by Ethics Committee of CMHC Research Center, Indonesia (approval number 2024-088), and the work followed the Declaration of Helsinki.¹⁸ All participants gave written informed consent, including consent to de-identification and pseudonymised storage, and to use of the data for forensic phenotypic and ancestry inference.

2.2. Participants

We recruited 1,000 unrelated adults aged 18 years or older through multi-centre sampling. Participants were eligible if they reported three-generation endogamy within their self-declared ethnic group and had no clinically evident pigmentary disorder such as oculocutaneous albinism, vitiligo or piebaldism. We excluded people with chemically bleached or dyed hair within six months, tinted contact lenses or recent ultraviolet phototherapy.

The cohort comprised eight named ethnic groups and a composite stratum: Javanese (n = 390), Sundanese (n = 155), Malay (n = 75), Batak (n = 45), Buginese (n = 35), Dayak (n = 30), Papuan (n = 25), Balinese (n = 20) and other groups (n = 225; Minangkabau, Madurese, Sasak, Ambonese and Torajan). There were 510 men and 490 women. Table 1 and Figure 1 show the sex, age and regional distribution.

2.3. Reference phenotypes

Phenotypes were recorded at a standardised examination under calibrated illumination. Skin colour (Dark, Medium, Light) was based on spectrophotometric reflectance of the unexposed upper inner arm and the von Luschan scale. Iris colour was recorded macroscopically and photographically as Brown/Dark Brown or Other (light brown, hazel, green, blue). Natural scalp hair colour was recorded at the root as Black, Dark Brown, Light Brown or Other (blond, red).

2.4. Genotyping

DNA was extracted from venous blood or buccal swabs with an automated magnetic-bead method (Chemagic DNA Blood Kit). Concentration and purity were measured by spectrophotometry (NanoDrop 2000) and fluorometry (Qubit 4.0); the minimum accepted concentration was 1.0 ng/ μ L, with an A260/A280 ratio of 1.75–1.90. Loci were amplified by multiplex PCR (95 °C for 10 min; 32 cycles of 94 °C for 30 s, 59 °C for 45 s and 72 °C for 45 s; final extension at 72 °C for 10 min) and typed by single-base extension (SNaPshot Multiplex Kit) on an ABI 3500xl Genetic Analyzer with GeneMapper ID-X v1.6. The method is capillary electrophoresis, not microfluidic genotyping.

The panel comprised 20 unique EVC SNPs and 26 AIMs. For hair colour, seven SNPs were typed: MC1R rs1805007, ASIP/ITCH rs1015362, IRF4 rs12203592, HERC2 rs1129038, OCA2 rs1800407, KITLG rs4778241 and ASIP rs12208402. For iris colour, eight SNPs were typed: HERC2 rs12913832, OCA2 rs1667394 and rs9933296, TYR rs1393350 and rs1042602, ASIP rs1800414, SLC24A4 rs12896399 and SLC45A2 rs16891982. For skin colour, seven SNPs were typed: SLC24A5 rs1426654, IRF4

rs12201779, MC1R rs885479, ASIP rs4911414, SLC45A2 rs2237826, TYR rs1042602 and OCA2 rs1800407; the last two are also used for iris colour. The 26 AIMs were ABCC11 rs17822931, ALDH2 rs671, DARC rs2814847 and rs17646946, rs1077872, rs11171853, rs4970383, rs12247880, rs727811, rs7573548, rs10830963, rs1157672, rs3782124, rs10494345, rs1726866, rs12103, rs11802594, rs10886122, rs2292661, rs1330328, rs1718125, rs3825932, rs1126809, rs4673, rs1770039 and rs2031920.

Genotypes were coded as 0, 1 or 2 copies of the tested allele given in Table 2. Calls required a peak height of at least 150 relative fluorescence units and, for heterozygotes, a peak-height ratio above 60%. Each plate included negative PCR controls and positive controls (9947A and 007). Ten per cent of samples were genotyped in duplicate. The authors report a 100% call rate and 100% duplicate concordance.

2.5. Prediction models

Skin, iris and hair colour were predicted with multinomial logistic regression calibrated to HIRISplex-S,⁴ giving the probability $P(T_i | G)$ of each category T_i given the genotype vector G . For skin colour, a participant was assigned to Dark if $P(\text{Dark}) \geq 0.60$, to Light if $P(\text{Light}) \geq 0.60$ and otherwise to Medium. For iris colour, Brown/Dark Brown was assigned if $P \geq 0.50$, and otherwise Other. For hair colour, the category with the highest probability was assigned.

2.6. Ancestry classification

A random forest of 500 trees was trained on the 26 AIMs to assign each person to one of three zones: Western (Javanese, Sundanese, Malay and Batak; Java and Sumatra), Eastern (Papuan; Maluku and Papua) or Central/other (Buginese, Dayak, Balinese and other groups; Sulawesi, Kalimantan and Nusa Tenggara). The expected zone was therefore defined from self-reported ethnic group. Tree depth and minimum split size were tuned by stratified five-fold cross-validation to minimise out-of-bag error.

2.7. Statistical analysis

Analyses used R 4.3.2 and Python 3.11 (scikit-learn, SciPy, pandas). Categorical variables are given as counts and percentages. Hardy–Weinberg equilibrium was tested per locus and ethnic group by exact tests with Benjamini–Hochberg false-discovery-rate correction. Accuracy is the proportion of correct predictions; for each category, the proportion of observed cases predicted correctly is reported as sensitivity (recall). The no-information rate is the proportion of the commonest class. Binomial 95% confidence intervals (Wilson) were calculated for proportions, and percentages are not given for strata with fewer than 10 participants. Discrimination was summarised by the one-vs-rest area under the receiver operating characteristic curve (AUC) with 95% bootstrap confidence intervals (1,000 resamples).

3. Results

3.1. Cohort and allele frequencies

The cohort is described in Table 1 and Figure 1. Java contributed 54.5% of participants, Maluku and Papua 8.0% and Papuans 2.5% (n = 25).

Table 1. Characteristics of the 1,000 participants.

Characteristic	Category	n	%
Ethnic group	Javanese	390	39.0
	Sundanese	155	15.5
	Malay	75	7.5
	Batak	45	4.5
	Buginese	35	3.5
	Dayak	30	3.0
	Papuan	25	2.5
	Balinese	20	2.0
	Other groups	225	22.5
Sex	Male	510	51.0
	Female	490	49.0
Age, years	18–25	200	20.0
	26–35	350	35.0
	36–45	250	25.0
	46–55	120	12.0
	≥ 56	80	8.0
Region of origin	Java	545	54.5
	Sumatra	200	20.0
	Maluku and Papua	80	8.0
	Sulawesi	70	7.0
	Kalimantan	60	6.0
	Bali and Nusa Tenggara	45	4.5

Notes: n, number of participants. Ethnic groups and regions are self-reported. The “other groups” stratum comprises Minangkabau, Madurese, Sasak, Ambonese and Torajan participants.

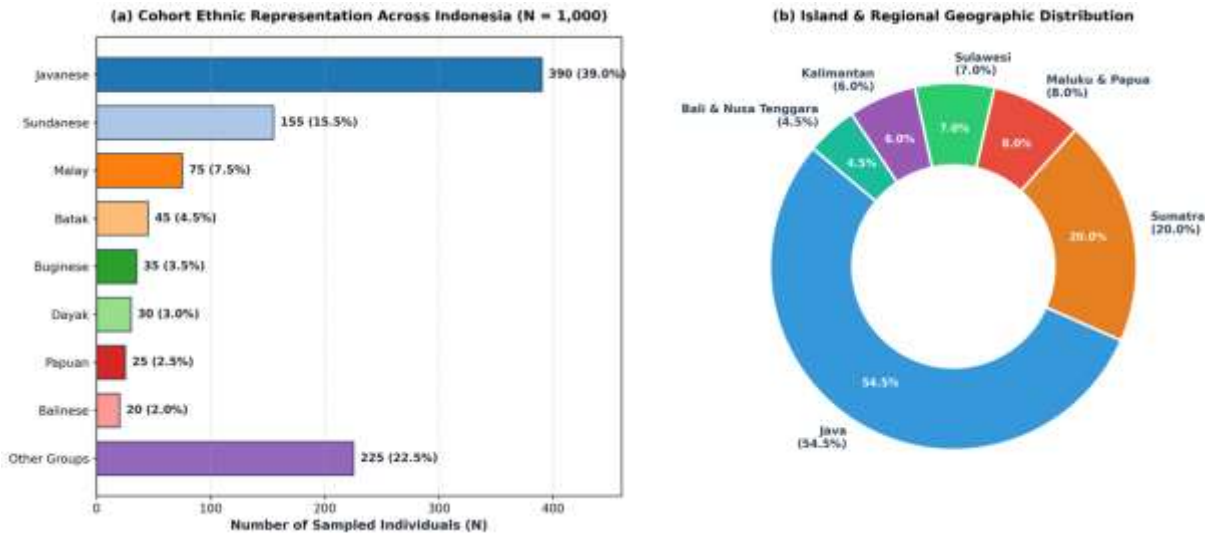


Figure 1. Ethnic and regional composition of the cohort (N = 1,000). (a) Participants per ethnic group, eight named groups and a composite stratum of other groups; (b) participants per region of origin.

Table 2 and Figure 2 give the frequencies reported for 25 of the 46 listed loci (20 EVC SNPs and five AIMs). The values are frequencies of the tested allele, not necessarily the minor allele: ten frequencies exceed 0.50 (for example rs1042602, 0.900; rs9933296, 0.880; rs1393350, 0.850), which a minor-allele frequency cannot do. The A allele at SLC24A5 rs1426654 was reported at 0.350 in Javanese

(95% CI 0.316–0.385) and 0.150 in Papuans (0.074–0.272), and the HERC2 rs12913832 allele at 0.050 overall. authors report no significant departure from Hardy–Weinberg equilibrium at any locus in any stratum after correction; with 20–45 participants in several strata, power to detect departures is low.

Table 2. Reported allele frequencies for 25 of the 46 listed SNPs.

rsID	Gene	Trait	Tested allele	Total cohort	Javanese	Papuan	Malay
rs1805007	MC1R	Hair (red)	C	0.030	0.020	0.050	0.040
rs1015362	ASIP/ITCH	Hair (light)	T	0.010	0.005	0.020	0.015
rs12203592	IRF4	Hair, skin	T	0.150	0.120	0.200	0.180
rs1129038	HERC2	Hair, iris	G	0.350	0.380	0.280	0.320
rs1800407	OCA2	Hair, skin	G	0.650	0.680	0.550	0.600
rs4778241	KITLG	Hair	A	0.700	0.720	0.600	0.650
rs12913832	HERC2	Iris (blue/light)	T	0.050	0.040	0.080	0.060
rs1667394	OCA2	Iris	G	0.070	0.060	0.100	0.080
rs1393350	TYR	Iris	A	0.850	0.880	0.750	0.800
rs1042602	TYR	Iris, skin	C	0.900	0.920	0.800	0.850
rs1800414	ASIP	Iris (light)	C	0.100	0.080	0.150	0.120
rs12896399	SLC24A4	Iris	G	0.120	0.100	0.180	0.140
rs1426654	SLC24A5	Skin (lighter)	A	0.300	0.350	0.150	0.250
rs12201779	IRF4	Skin (lighter)	T	0.250	0.300	0.100	0.200
rs885479	MC1R	Skin (darker)	T	0.600	0.550	0.750	0.650
rs4911414	ASIP	Skin (darker)	A	0.550	0.500	0.700	0.600
rs2237826	SLC45A2	Skin (lighter)	G	0.320	0.370	0.180	0.280
rs16891982	SLC45A2	Iris, skin	C	0.030	0.020	0.060	0.040
rs9933296	OCA2	Iris	A	0.880	0.900	0.780	0.850
rs12208402	ASIP	Hair	G	0.180	0.150	0.250	0.200
rs17822931	ABCC11	Ancestry (AIM)	T	0.480	0.520	0.350	0.450
rs671	ALDH2	Ancestry (AIM)	G	0.620	0.650	0.500	0.600
rs2814847	DARC	Ancestry (AIM)	C	0.320	0.300	0.450	0.350
rs4970383	n/a	Ancestry (AIM)	A	0.720	0.750	0.600	0.700
rs1726866	n/a	Ancestry (AIM)	A	0.780	0.800	0.650	0.750

Notes: The “tested allele” is the allele whose frequency is given, as listed in the submission; it is not necessarily the minor allele and strand is not stated. Genes follow the Editorial Assessment (rs1015362 ASIP/ITCH; rs1129038 HERC2; ABCC11 rs17822931). n/a, gene not given.

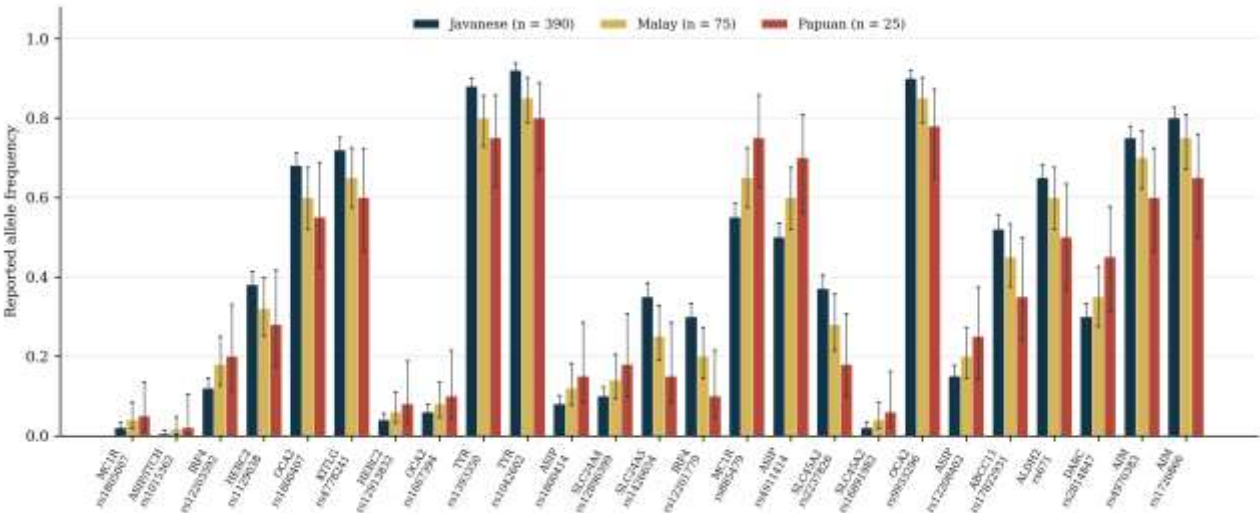


Figure 2. Reported allele frequencies (tested allele, error bars show 95% binomial confidence intervals for 2n alleles) at the 25 loci of Table 2 in Javanese (n = 390), Malay (n = 75) and Papuan (n = 25) participants. Loci are in the order of Table 2.

3.2. Prediction of skin, iris and hair colour

Table 3 summarises accuracy against the no-information rate. Skin colour was predicted correctly in 89.2% of participants, well above the 45.0% obtained by always predicting Medium. Iris colour accuracy (82.0%) was

below the 88.0% obtained by always predicting Brown, and hair colour accuracy (80.0%) was 8.0 percentage points above the 72.0% obtained by always predicting Black. The accuracies are apparent estimates unless the authors confirm that the model was tested on independent data (Methods 2.5).

Table 3. Prediction accuracy compared with the no-information rate.

Trait	Correct / N	Accuracy, % (95% CI)	No-information rate, % (class)	AUC, one-vs-rest (95% CI)
Skin colour	892 / 1,000	89.2 (87.1–91.0)	45.0 (Medium)	Dark 0.96 (0.94–0.98); Light 0.91 (0.88–0.94); Medium 0.89 (0.86–0.92)
Iris colour	820 / 1,000	82.0 (79.5–84.3)	88.0 (Brown)	Brown vs other 0.88 (0.85–0.91)
Hair colour	800 / 1,000	80.0 (77.4–82.4)	72.0 (Black)	Black 0.87 (0.84–0.90); Dark Brown 0.76 (0.71–0.81); Light Brown 0.71 (0.64–0.78)
Ancestry zone	837 / 1,000	83.7 (81.3–85.9)	66.5 (Western)	Not reported

Notes: AUC, area under the receiver operating characteristic curve; CI, confidence interval. The no-information rate is the proportion of the commonest observed class (Medium skin, Brown iris, Black hair, Western zone). Iris colour is a two-class problem, so one AUC is given.

Hair colour is detailed in Table 4. Sensitivity was 85.97% for black hair (619 of 720) but lower for the less common categories (Dark Brown 66.7%, Light Brown 64.3%, Other 53.3%). Percentages are not shown for strata under 10 participants; the Papuan stratum had one Light Brown and

one Other participant, both predicted incorrectly. Iris colour is detailed in Table 5. Sensitivity was 86.0% for brown irises (757 of 880) but only 52.5% for other irises (63 of 120).

Table 4. Sensitivity of hair colour prediction by observed hair colour and group.

Observed hair colour	Group	Observed, n	Correct, n	Sensitivity, % (95% CI)
Black	Total cohort	720	619	86.0 (83.2–88.3)
	Javanese	285	248	87.0 (82.6–90.4)
	Papuan	20	16	80.0 (58.4–91.9)
	Malay	55	48	87.3 (76.0–93.7)
Dark Brown	Total cohort	180	120	66.7 (59.5–73.1)
	Javanese	70	45	64.3 (52.6–74.5)
	Papuan	3	2	2/3
	Malay	15	10	66.7 (41.7–84.8)
Light Brown	Total cohort	70	45	64.3 (52.6–74.5)
	Javanese	30	18	60.0 (42.3–75.4)
	Papuan	1	0	0/1
	Malay	5	3	3/5
Other (blond/red)	Total cohort	30	16	53.3 (36.1–69.8)
	Javanese	5	3	3/5
	Papuan	1	0	0/1
	Malay	0	0	n/a
Overall accuracy	Total cohort	1,000	800	80.0 (77.4–82.4)
	Javanese	390	314	80.5 (76.3–84.1)
	Papuan	25	18	72.0 (52.4–85.7)
	Malay	75	61	81.3 (71.1–88.5)

Notes: Sensitivity (recall) is the proportion of participants with the observed phenotype that were predicted correctly. Counts are shown instead of percentages for n < 10. n/a, not applicable (no participants). The no-information rate is 72.0%.

Table 5. Sensitivity of iris colour prediction by observed iris colour and group.

Observed iris colour	Group	Observed, n	Correct, n	Sensitivity, % (95% CI)
Brown / Dark Brown	Total cohort	880	757	86.0 (83.6–88.2)
	Javanese	345	298	86.4 (82.4–89.6)
	Papuan	22	18	81.8 (61.5–92.7)
	Malay	65	57	87.7 (77.5–93.6)
Other (light brown, hazel, blue, green)	Total cohort	120	63	52.5 (43.6–61.2)
	Javanese	45	22	48.9 (35.0–63.0)
	Papuan	3	1	1/3
	Malay	10	4	40.0 (16.8–68.7)
Overall accuracy	Total cohort	1,000	820	82.0 (79.5–84.3)
	Javanese	390	320	82.1 (77.9–85.5)
	Papuan	25	19	76.0 (56.6–88.5)
	Malay	75	61	81.3 (71.1–88.5)

Notes: Sensitivity (recall) is the proportion of participants with the observed phenotype that were predicted correctly. Counts are shown instead of percentages for n < 10. The no-information rate is 88.0%.

Skin colour is detailed in Table 6. Sensitivity was 94.0% for Dark (329 of 350), 86.0% for Medium (387 of 450) and 88.0% for Light (176 of 200). No Papuan or Malay participant had Light skin.

Table 6. Sensitivity of skin colour prediction by observed skin colour and group.

Observed skin colour	Group	Observed, n	Correct, n	Sensitivity, % (95% CI)
Dark	Total cohort	350	329	94.0 (91.0–96.0)
	Javanese	80	72	90.0 (81.5–94.8)
	Papuan	20	19	95.0 (76.4–99.1)
	Malay	60	57	95.0 (86.3–98.3)
Medium	Total cohort	450	387	86.0 (82.5–88.9)
	Javanese	250	215	86.0 (81.2–89.8)
	Papuan	5	4	4/5
	Malay	15	13	86.7 (62.1–96.3)
Light	Total cohort	200	176	88.0 (82.8–91.8)
	Javanese	60	53	88.3 (77.8–94.2)
	Papuan	0	0	n/a
	Malay	0	0	n/a
Overall accuracy	Total cohort	1,000	892	89.2 (87.1–91.0)
	Javanese	390	340	87.2 (83.5–90.1)
	Papuan	25	23	92.0 (75.0–97.8)
	Malay	75	70	93.3 (85.3–97.1)

Notes: Sensitivity (recall) is the proportion of participants with the observed phenotype that were predicted correctly. Counts are shown instead of percentages for n < 10. The no-information rate is 45.0%.

Discrimination is shown in Figure 3. For skin colour the AUC was 0.96 for Dark (95% CI 0.94–0.98), 0.91 for Light (0.88–0.94) and 0.89 for Medium (0.86–0.92). The AUC for brown irises was 0.88 (0.85–0.91). For hair colour the AUC was 0.87 for Black (0.84–0.90), 0.76 for Dark Brown (0.71–0.81) and 0.71 for Light Brown (0.64–0.78). The submission also gave an AUC of 0.72 for “other” irises;

because iris colour was a two-class problem, the two AUCs must be identical and the second value was deleted. A multiclass Brier score of 0.114 was reported for skin colour only. It cannot be judged without a reference: for the observed class proportions, a forecast that always gives the class frequencies scores 0.635 (summed over classes) or 0.212 (mean over classes).

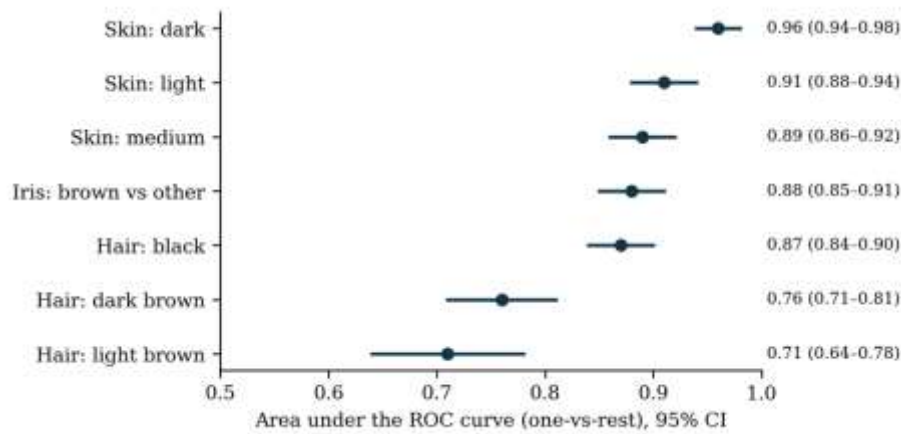


Figure 3. Area under the ROC curve (one-vs-rest) with 95% confidence intervals, as reported by the authors, for skin, iris and hair colour categories.

3.3. Ancestry classification

Table 7 and Figure 4 show how many participants were assigned to the expected zone. Overall, 837 of 1,000 (83.7%) were assigned correctly, against a no-information rate of 66.5% (665 Western participants). Sensitivity was

85.4% for the Western zone (568 of 665), 80.0% for the Eastern zone (20 of 25; 95% CI 60.9–91.1) and 80.3% for the Central/other zone (249 of 310); the balanced accuracy was 81.9%. The Eastern estimate is imprecise because it rests on 25 Papuan participants.

Table 7. Assignment of participants to the expected ancestry zone by ethnic group.

Expected zone	Ethnic group	n	Correctly assigned, n	Sensitivity, % (95% CI)
Western	Javanese	390	335	85.9 (82.1–89.0)
	Sundanese	155	130	83.9 (77.3–88.8)
	Malay	75	65	86.7 (77.2–92.6)
	Batak	45	38	84.4 (71.2–92.3)
	Subtotal	665	568	85.4 (82.5–87.9)
Eastern	Papuan	25	20	80.0 (60.9–91.1)
Central/other	Buginese	35	28	80.0 (64.1–90.0)
	Dayak	30	24	80.0 (62.7–90.5)
	Balinese	20	17	85.0 (64.0–94.8)
	Other groups	225	180	80.0 (74.3–84.7)
	Subtotal	310	249	80.3 (75.5–84.4)
All	All participants	1,000	837	83.7 (81.3–85.9)

Notes: The expected zone is defined from self-reported ethnic group (Methods 2.6). Sensitivity is the proportion of participants in the group assigned to the expected zone. The no-information rate is 66.5% (always Western).

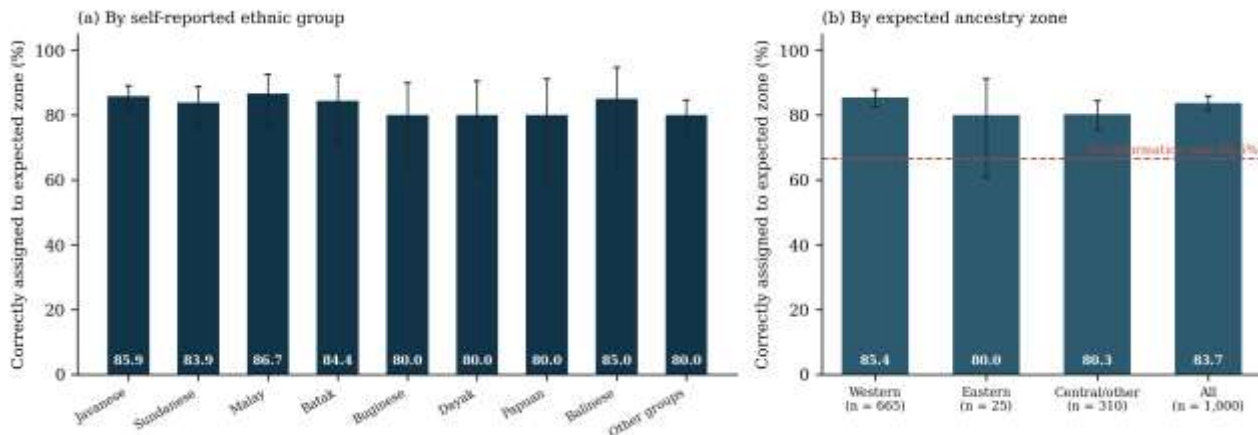


Figure 4. Proportion of participants assigned to the expected ancestry zone, with 95% confidence intervals. (a) By self-reported ethnic group; (b) by expected zone and overall, with the no-information rate (dashed line).

4. Discussion

4.1. Principal findings

In this cohort the panel predicted skin colour well above the baseline (89.2% against 45.0%). Iris colour accuracy (82.0%) was below the proportion of brown irises (88.0%), and hair colour accuracy (80.0%) exceeded the proportion of black hair (72.0%) by only 8 points. Participants were assigned to their expected ancestry zone in 83.7% of cases, against 66.5% for always predicting Western. Unless a held-out test is confirmed, these are apparent estimates.

4.2. Comparison with the literature

The HIRisPlex-S skin model was developed for five skin categories,⁵ and this study used three, so the accuracies cannot be compared directly. The high AUC for dark skin is consistent with the strong pigmentation signal in the loci typed, but the high accuracy partly reflects the shape of the cohort: the commonest skin class was Medium (45.0%).

Iris colour was predicted well only for the dominant brown class. Accuracy and predictive values depend on how common each phenotype is in the population tested,¹⁹ and with 88.0% brown irises a trivial rule already scores higher than the model. Sensitivity for other irises was 52.5% (63 of 120). The iris categories also differ from the blue/brown outcome of IrisPlex,¹⁰ and this study cannot say whether the weak non-brown prediction reflects the genotypes, the small number of such irises or measurement of light-brown irises. For hair, sensitivity fell with each less common colour. Natural blond hair in Melanesians is associated with a TYRP1 variant that is not part of standard panels;²⁰ it could explain misclassified Papuan

participants, but the variant was not typed here and the Papuan stratum had seven misclassified hair colours (18 of 25 correct), so this remains an untested possibility.

Ancestry in Indonesia changes across Wallacea,^{12,14} and Sundaland and Sahul ancestries are well separated in ancient and modern genomes.^{13,14} Here, however, the expected class was defined from self-reported ethnic group. Dayak and Balinese participants, whose homelands lie on the Sunda shelf, were placed in the Central/other zone, and the Eastern class contained only Papuan participants ($n = 25$), so the analysis does not test discrimination across the Wallace line itself. Reported allele frequencies for several loci also look unusual for Southeast Asian populations (Results 3.1); until orientation and strand are checked, biological interpretation of single loci should be avoided.

4.3. Implications for forensic practice

FDP provides an investigative lead, not identification. Under the Indonesian Criminal Procedure Code, a conviction requires at least two valid items of evidence and the judge's conviction (Articles 183–184), and expert statements are one such item (Article 186).²¹ A prediction with the accuracy shown here, particularly for iris and hair colour, should therefore be reported with its uncertainty and baseline. We suggest that laboratories consider likelihood-ratio reporting of genotype evidence, but we did not build or test a likelihood-ratio model, and the worked ratio in the submission ("48.6 times more likely") had no derivation and was removed. Figure 5 shows a proposed workflow whose last two steps were not evaluated.



Figure 5. Conceptual workflow for forensic DNA phenotyping. Steps 1–3 were performed in this study; steps 4 and 5 are proposed and were not evaluated.

Genetic data are classed as specific personal data under Law No. 27 of 2022,²² and participants consented to a research purpose, so use of the genotypes to build a law-enforcement reference set would need a separate legal basis. Ancestry inference is contested among practitioners,²³ and false leads could fall on small groups such as the 25 Papuan participants. Restricting FDP to serious offences or to cases under judicial warrant is a policy proposal, not a result of this study. Disaster victim identification is a possible use, and FDP has been tried on decomposed bodies,²⁴ but all samples here were fresh blood or buccal swabs, so degraded samples were not tested.

4.4. Limitations

First, models and thresholds may have been fitted and tested on the same participants, and no independent test set was used. Second, the iris and skin categories were collapsed and the reference phenotypes came from a single protocol whose cut-offs, examiners and agreement are not reported. Third, strata were small (25 Papuan and 20 Balinese participants), and percentages are therefore unstable. Fourth, ancestry classes were defined from self-reported ethnic group, admixed individuals were excluded

by the endogamy criterion, and no external ancestry reference was used. Fifth, the unmodified HIRisPlex-S model was not run on the same samples, so this study cannot show that a locally calibrated panel is better than an existing one. Sixth, single-base extension covers fewer markers than sequencing-based panels,⁶ and hair texture and facial shape were not studied.²⁵ Finally, hair darkening and greying with age were not analysed. Further validation on independent samples would be needed before any use in casework.

5. Conclusion

In this cohort of 1,000 Indonesian adults, the panel predicted skin colour with 89.2% accuracy and assigned 83.7% of participants to their expected ancestry zone, but iris colour accuracy (82.0%) was below the proportion of brown irises (88.0%) and hair colour accuracy (80.0%) was only 8 points above the baseline of 72.0%. These apparent estimates come from one cohort. Independent testing, comparison with the unmodified HIRisPlex-S model and correction of the allele reporting are required before the panel is used as an investigative lead in casework.

Declarations

Use of generative artificial intelligence. None. The authors declare that no generative artificial intelligence tool was used in the conception, research, analysis, drafting, revision or preparation of this manuscript.

Author contributions. RA: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft. FS: Formal analysis, Investigation, Validation, Visualization. BA: Data curation, Software, Visualization. SS: Conceptualization, Project administration, Supervision, Writing – review & editing. KA: Resources, Validation, Writing – review & editing. All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding. None. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest. None. The authors declare no competing financial or non-financial interests in relation to this work.

Ethical approval. The study was approved by Ethics Committee of CMHC Research Center, Indonesia (approval no. 2024-088, complied with the Declaration of Helsinki).

Consent for publication. Not applicable. No identifiable individual data are reported. Written informed consent to participation was obtained from every participant.

Data availability. The dataset and the analysis scripts supporting the findings are available from the corresponding author on reasonable request.

References

- Kayser M. Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. *Forensic Sci Int Genet.* 2015;18:33-48. doi:10.1016/j.fsigen.2015.02.003
- Kayser M, Branicki W, Parson W, et al. Recent advances in forensic DNA phenotyping of appearance, ancestry and age. *Forensic Sci Int Genet.* 2023;65:102870. doi:10.1016/j.fsigen.2023.102870
- Walsh S, Liu F, Wollstein A, et al. The HIRISplex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet.* 2013;7(1):98-115. doi:10.1016/j.fsigen.2012.07.005
- Chaitanya L, Breslin K, Zúñiga S, et al. The HIRISplex-S system for eye, hair and skin colour prediction from DNA: introduction and forensic developmental validation. *Forensic Sci Int Genet.* 2018;35:123-135. doi:10.1016/j.fsigen.2018.04.004
- Walsh S, Chaitanya L, Breslin K, et al. Global skin colour prediction from DNA. *Hum Genet.* 2017;136(7):847-863. doi:10.1007/s00439-017-1808-5
- Xavier C, de la Puente M, Mosquera-Miguel A, et al. Development and inter-laboratory evaluation of the VISAGE Enhanced Tool for Appearance and Ancestry inference from DNA. *Forensic Sci Int Genet.* 2022;61:102779. doi:10.1016/j.fsigen.2022.102779
- Ruiz-Ramírez J, de la Puente M, Xavier C, et al. Development and evaluations of the ancestry informative markers of the VISAGE Enhanced Tool for Appearance and Ancestry. *Forensic Sci Int Genet.* 2023;64:102853. doi:10.1016/j.fsigen.2023.102853
- Parra EJ. Human pigmentation variation: evolution, genetic basis, and implications for public health. *Am J Phys Anthropol.* 2007;134(S45):85-105. doi:10.1002/ajpa.20727
- Marano LA, Andersen JD, Gonçalves FT, et al. Evaluation of HIRISplex-S system markers for eye, skin and hair color prediction in an admixed Brazilian population. *Forensic Sci Int Genet Suppl Ser.* 2019;7(1):427-428. doi:10.1016/j.fsigs.2019.10.038
- Walsh S, Lindenbergh A, Zuniga SB, et al. Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. *Forensic Sci Int Genet.* 2011;5(5):464-471. doi:10.1016/j.fsigen.2010.09.008
- Pośpiech E, Karłowska-Pik J, Kukla-Bartoszek 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. doi:10.1016/j.fsigen.2022.102693
- Hudjashov G, Karafet TM, Lawson DJ, et al. Complex patterns of admixture across the Indonesian archipelago. *Mol Biol Evol.* 2017;34(10):2439-2452. doi:10.1093/molbev/msx196
- Carlhoff S, Duli A, Nägele K, et al. Genome of a middle Holocene hunter-gatherer from Wallacea. *Nature.* 2021;596(7873):543-547. doi:10.1038/s41586-021-03823-6
- Oliveira S, Nägele K, Carlhoff S, et al. Ancient genomes from the last three millennia support multiple human dispersals into Wallacea. *Nat Ecol Evol.* 2022;6(7):1024-1034. doi:10.1038/s41559-022-01775-2
- von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med.* 2007;4(10):e296. doi:10.1371/journal.pmed.0040296
- Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* 2015;350:g7594. doi:10.1136/bmj.g7594
- Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. doi:10.1136/bmj-2023-078378
- World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053
- Caliebe A, Krawczak M, Kayser M. Predictive values in forensic DNA phenotyping are not necessarily prevalence-dependent. *Forensic Sci Int Genet.* 2018;33:e7-e8. doi:10.1016/j.fsigen.2017.11.006
- Kenny EE, Timpson NJ, Sikora M, et al. Melanesian blond hair is caused by an amino acid change in TYRP1. *Science.* 2012;336(6081):554. doi:10.1126/science.1217849
- Republic of Indonesia. Law No. 8 of 1981 on the Criminal Procedure Code (Kitab Undang-Undang Hukum Acara Pidana, KUHP). Jakarta; 1981.
- Republic of Indonesia. Law No. 27 of 2022 on Personal Data Protection (Undang-Undang Pelindungan Data Pribadi, UU PDP). Jakarta; 2022.
- Granja R, Machado H. Forensic DNA phenotyping and its politics of legitimization and contestation: views of forensic geneticists in Europe. *Soc Stud Sci.* 2020;53(6):850-868. doi:10.1177/0306312720945033
- Fabbri M, Alfieri L, Mazdai L, et al. Application of forensic DNA phenotyping for prediction of eye, hair and skin colour in highly decomposed bodies. *Healthcare (Basel).* 2023;11(5):647. doi:10.3390/healthcare11050647
- Kataria S, Dabas P, Saraswathy KN, et al. Investigating the morphology and genetics of scalp and facial hair characteristics for phenotype prediction. *Sci Justice.* 2023;63(1):135-148. doi:10.1016/j.scijus.2022.12.002