Clin. Lab. 2024;70:XXX-XXX

ORIGINAL ARTICLE

Application of a 24-SNP Multiplex Genotyping Assay System for Phenotypic Identification of Fujian Han Population

Hong Yu [1, 2, 3], Lili Han [2, 3], Dian Chen [2, 3], Yaocheng Wang [1, 2, 3], Shanglong Liu [2, 3], Suimei Wu [4], Tongtong Zheng [5], Li Lai [1, 2, 3]

[1] *Shengli Clinical College of Fujian Medical University, Fuzhou, Fujian, P. R. China*
[2] *Forensic DNA Laboratory of Fujian Provincial Hospital, Fujian Provincial Key Laboratory of Cardiovascular Disease, Fuzhou, Fujian, P. R. China*
[3] *Central Laboratory of Fujian Provincial Hospital, Fuzhou, Fujian, P. R. China*
[4] *Vascular and Tumor Intervention Department of Fujian Provincial Hospital, Fuzhou, Fujian, P. R. China*
[5] *Operating Room of Fujian Provincial Hospital, Fuzhou, Fujian, P. R. China*

SUMMARY

Background: **This study aimed to evaluate the utility of 24 single nucleotide polymorphism (SNP) loci associated with iris color and hair color in phenotypic identification of the Han Chinese population in Fujian Province. The selected SNPs, known for their strong correlation with specific human phenotypic features, provide valuable reference data for developing a molecular phenotypic identification system.**
Methods: **A multiplex genotyping assay system was established with primers for the 24 SNPs linked to iris color and hair color synthesized based on existing literature. In total, 235 unrelated individuals of Han Chinese ethnicity in Fujian Province were included in this study. PowerStats v12 was employed to calculate forensic parameters associated with the 24 SNP loci, including gene frequencies, genotype frequencies, minor allele frequencies, discrimination power (DP), polymorphism information content (PIC), and observed heterozygosity (Ho). Hardy-Weinberg equilibrium tests were conducted for each locus. The SNP genotyping results were uploaded to the HIrisPlex model (https://HIrisPlex.erasmusmc.nl/) to predict iris and hair colors, and the inferred results were compared with manually assessed images. The accuracy of pigment phenotype inference was evaluated by using ROC curves in SPSS 26.0 software.**
Results: **The accuracy rates of inferring brown iris and black hair phenotypes were 99.6% and 99.5%. The area under the curve (AUC) values were 0.923 and 0.980, respectively.**
Conclusions: **The 24 SNP loci demonstrated high accuracy in inferring iris color and hair color; it seems to be a useful tool for forensic phenotypic identification and anthropological or evolutionary applications. Establishment of suitable pigment classification criteria and optimized prediction models is based on revealing more phenotypic genetic markers.**
(Clin. Lab. 2024;70:xx-xx. DOI: 10.7754/Clin.Lab.2024.240425)

Correspondence:
Li Lai
Shengli Clinical College of Fujian Medical University
Fuzhou
Fujian 350001
P. R. China
Email: amoeba2000@126.com

Manuscript accepted May 19, 2024

KEYWORDS

single nucleotide polymorphism, multiplex genotyping assay, phenotypic identification, individual identification

INTRODUCTION

Human phenotypes encompass a range of externally visible characteristics, including ethnic origin, age, height, body weight, facial features, male pattern bald-

ness, and pigment features such as iris color, hair color, and skin color [1]. These externally visible characteristics (EVCs) are known to be highly conserved and strongly influenced by genetics. Most traits are complex hereditary traits controlled by multiple genes, while only a few are determined by a single gene. In recent decades, the prediction of human phenotypes from DNA has gained significant interest, particularly in the field of forensic genetics [2-4]. Advances in genetic studies on human phenotypes and related theories and technologies, such as DNA molecular markers, have provided valuable insights into the characterization of EVCs. By analyzing the correlation between genetic information and phenotypic features, associations between specific phenotypes and corresponding DNA molecular markers can be identified, providing a crucial theoretical foundation for the use of DNA molecular markers in phenotypic identification [5]. The advent of "smart DNA" [6] has made it increasingly feasible to infer biogeographic ancestry (BGA) and human phenotypes by using SNP genotyping, opening new avenues in forensic genetics.

Both domestic and international studies have identified SNPs that are highly associated with EVCs and have developed corresponding detection systems and inference models. Currently, EVCs, including ethnic origin, age, and pigment features such as iris color, hair color, and skin color, have been extensively studied aspects of human phenotypes [7-9]. Therefore, there is a wealth of research reports focusing on the relationship between pigment features and genetic molecular markers, with well-established inference models and high accuracy, making it a prominent and challenging area in individual identification research.

Research on pigment features primarily focuses on molecular identification of iris color and hair color. The iris comprises the anterior stroma and the posterior pigment epithelium, with iris color determined by the amount of pigment in the pigment epithelial cells within the stroma. In 2010, Walsh [10] and colleagues identified six SNP loci (rs1800407, rs12913832, rs16891982, rs12896 399, rs12203592, and rs1393350) that exhibit the strongest correlation with iris color inference. They successfully developed the IrisPlex system, capable of inferring blue and brown iris colors, and subsequently validated the detection system the following year [11]. Wollstein [12] proposed a fully automated human pigment phenotype quantification system, which quantifies the proportions of eumelanin, pheomelanin, and no pigment in different pigmented regions of the iris. This system significantly enhances the accuracy of iris color inference and replaces traditional iris color classification for phenotype description. Human hair contains eumelanin, a reddish-yellow pheomelanin, and a brown-black eumelanin, which exist in different compound forms. Variations in hair color arise from differences in the content, distribution, and types of these compounds [13].

In this project, we constructed a multi-fluorescence detection system consisting of 24 SNP loci related to iris color and hair color in human phenotypic molecules. The aim was to explore the application value of this composite genotyping system in phenotypic molecular identification of iris color and hair color in the Han population of Fujian. Using this genotyping system, we conducted a population survey of the Han population in Fujian to obtain forensic genetic parameters, including minimum allele frequency, polymorphism information content, and individual identification capability. The genotypes of the 24 SNP loci were uploaded to the international "Iris and Hair Color Inference Model" at Erasmus Medical Center in the Netherlands. The inferred results of iris color and hair color in the Han population of Fujian were then compared with manually assessed image results to evaluate the accuracy of inferring pigment phenotype features. This project represents the first study in China to investigate pigment phenotype features in the Han population of Fujian. It involves the screening of genetic markers that conform to the phenotypic features of Asian populations and the evaluation of the application value of phenotype molecular features in individual identification practice.

MATERIALS AND METHODS

Patients

A total of 235 unrelated individuals of Han Chinese ethnicity from the Fujian region, referred to as CHF, were included in the sample collection. This comprised 85 individuals from eastern Fujian and 50 individuals each from northern, southern, and western Fujian. Each sample consisted of 2 mL of venous blood collected with EDTA anticoagulant. During the sampling process, individuals were required to provide their identification cards and household registers to confirm their Han ethnicity and ancestral residence in Fujian Province for at least three generations. Additionally, a 2D facial photograph was taken using a Canon digital camera (POWER SHOT 3000 IS) for each participant. The sampling process followed the principles of informed consent. This study was approved by the Fujian Provincial Hospital Ethics Committee (approval reference number: K2021 04 61, date of approval: April 15, 2021).

DNA extraction from samples

DNA extraction from the samples was performed by using the Zhongyuan Company nucleic acid extraction kit (magnetic bead method) according to the manufacturer's instructions. The extracted DNA was stored at -20°C.

DNA quality check and quantification

For quality checking, 5 μL of each DNA sample was subjected to 1% agarose gel electrophoresis. The electrophoresis parameters were set at 150 V and 100 mA for 10 - 20 minutes. The concentration and purity of the DNA samples were measured by using the Beckman DU640 nucleic acid protein analyzer. A concentration

measurement was performed by using 1 - 2 μL of each DNA sample, and the OD260 and OD280 values were recorded. The purity of the DNA samples was determined by calculating the ratio of OD260 to OD280, with values between 1.7 and 1.9 considered acceptable. Finally, the samples were diluted to a working concentration of 5 - 10 ng/μL.

Primer synthesis

Based on previously reported primer sequences and system construction methods, the sequence of each locus was obtained by searching for the corresponding rs number. The flanking sequences of each locus, spanning 300 - 500 base pairs upstream and downstream, were selected. Genetic information, such as chromosomal location, minor allele, and major allele, for each locus was also obtained. The detection system included six SNP loci associated with iris color: rs12913832 (located in the HERC2 gene), rs1800407 (in the OCA2 gene), rs12896399 (in the SLC24A4 gene), rs16891982 (in the SLC45A2 gene), rs1393350 (in the TYR gene), and rs12203592 (in the IRF4 gene). The remaining 18 loci were related to hair color, including 11 loci in the MC1R gene and 7 loci in genes such as TYR and EXOC2. The specific information of the loci is presented in Table 1.

SNP site detection

PCR amplification

The PCR reaction system was prepared as follows: a total volume of 10 μL consisting of 1 μL of 10 × Buffer I, 0.8 μL of dNTPs (2.5 μM), 2 μL of primers (5 μM, forward and reverse), 0.1 μL of Hs Taq polymerase (5 U/μL), 2 μL of DNA (10 - 20 ng/μL), and topped up to 10 μL with ddH2O.

The PCR reaction program consisted of an initial denaturation at 94°C for 5 minutes, followed by 10 cycles of denaturation at 94°C for 30 seconds, annealing at a temperature gradually decreasing from 65°C to 55°C for 30 seconds, and extension at 72°C for 30 seconds. This was followed by 30 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 72°C for 10 minutes.

Pre-amplification product quality check and digestion

For quality check, 2 μL of the pre-amplification product was loaded onto a 2% agarose gel for electrophoresis. The electrophoresis parameters were set at 150 V and 100 mA for 10 - 20 minutes. The pre-amplification products from each sample were combined in proportion, and 4 μL of the mixture was used for digestion. The digestion system consisted of 5 μL of diluted ExoI enzyme (10-fold dilution), 4 μL of PCR product, 1.33 μL of SAP (1 U/μL), and 0.27 μL of Exol (20 U/μL). The pre-amplification products were purified at 37°C for 1 hour, followed by 75°C for 20 minutes.

Extension reaction

For the extension reaction, 1.2 μL of the digested pre-amplification product, 2 μL of Primers Mix, 0.5 μL of ABI Mix, 0.4 μL of 10 × Buffer I, and H2O to a final volume of 5 μL were mixed. The extension conditions were as follows: 10 seconds at 96°C, 5 seconds at 50°C, and 30 seconds at 60°C for 30 cycles. The extension products were purified by adding 1 μL of CIP to 6 μL of the extension reaction product, incubating at 37°C for 1 hour, followed by 75°C for 15 minutes. For extension product detection, 1.0 μL of PCR product was mixed with 9 μL of a mixture containing internal molecular weight standards and formamide. The mixture was denatured at 95°C for 5 minutes and loaded onto the ABI3130 genetic analyzer. Data Collection Software V3.0 was used to collect the raw data, and the results were analyzed by using GeneMapper ID Software V5.0.

Uploading genotype results to an online inference model

The genotype results for the 24 SNPs obtained from the 235 samples were submitted to an international online inference model. This model provided inferred probabilities (p-values) for iris color (blue, intermediate, brown) and hair color (red, blond, brown, black). The interpretation of iris and hair colors was based on the instructions provided by the inference model and relevant scientific literature. In the case of iris color inference, the phenotype with the highest probability (p-value) was selected. For hair color inference, a systematic approach was employed, taking into account the highest overall probability (p-value) and the likelihood of different color shades. If the highest probability for hair color exceeded 0.7, it was inferred as black hair. However, if the probability was less than 0.7 and the likelihood of having dark hair was greater than 0.9, it was inferred as black hair; otherwise, it was inferred as dark brown or black. Similarly, if the highest probability for hair color exceeded 0.7 and the likelihood of having light hair was greater than 0.9, it was inferred as blond hair; otherwise, it was inferred as blond or dark blond hair. If the highest probability for hair color exceeded 0.7 and the likelihood of having light hair was greater than 0.9, it was inferred as brown hair; otherwise, it was inferred as brown or dark brown hair. Lastly, if the highest probability for hair color was associated with red, the individual's hair color was inferred as red (Figure 1).

Manual classification of pigment phenotypes

A manual method was employed to classify pigment phenotypes and extract relevant characteristics from 2D photographs. During the manual analysis of the photographs, iris and hair colors were categorized into three groups based on predefined criteria: blue, intermediate (including green, hazel, and light brown), and brown (ranging from light brown to dark brown). Hair colors were divided into four categories: red, blond, brown, and black. Categories such as dyed hair, white hair, hair

Table 1. 24 SNP loci information.

No.	dbSNP rs numbers	Chromosome	Chromosomal position		Gene	Major allele	Minor allele
1	N29insA	16	89985753	Exon	MC1R	C	A
2	rs11547464	16	89986091	Exon	MC1R	G	A
3	rs885479	16	89986154	Exon	MC1R	C	T
4	rs1805008	16	89986144	Exon	MC1R	C	T
5	rs1805005	16	89985844	Exon	MC1R	G	T
6	rs1805006	16	89985918	Exon	MC1R	C	A
7	rs1805007	16	89986117	Exon	MC1R	C	T
8	rs1805009	16	89986546	Exon	MC1R	G	C
9	Y152OCH	16	89986122	Exon	MC1R	C	A
10	rs2228479	16	89985940	Exon	MC1R	G	A
11	rs1110400	16	89986130	Exon	MC1R	T	C
12	rs28777	5	33994716	Intergenic	SLC45A2	A	C
13	rs16891982	5	33987450	Exon	SLC45A2	G	C
14	rs12821256	12	87852466	Intron	KITLG	A	G
15	rs4959270	6	402748	Intron	LOC1	C	A
16	rs12203592	6	341321	Intergenic	IRF4	C	T
17	rs1042602	11	88551344	Exon	TYR	G	T
18	rs1800407	15	25903913	Exon	OCA2	G	A
19	rs2402130	14	91870956	Intron	SLC24A4	A	G
20	rs12913832	15	26039213	Intron	HERC2	C	T
21	rs2378249	20	32681751	Intron	PIGU	T	C
22	rs12896399	14	91843416	Intergenic	LOC1	T	G
23	rs1393350	11	88650694	Intron	TYR	C	T
24	rs683	9	12699305	Exon	TYRP1	T	G

loss, and indistinguishable hair color were designated as category 0. Three individuals participated in the process of classifying phenotypes. Prior to the formal analysis, the three individuals were acquainted with the definitions of pigment phenotype classifications and underwent training using a randomly selected set of 50 2D photographs from the samples to ensure consistent understanding of the phenotypes. Subsequently, each reader independently classified the phenotypes by using a shared display screen. After completion, the recorded results from the three individuals were analyzed, and the average value of their classifications was calculated as the final data for pigment phenotype classification.

Statistical analysis

PowerStats v12 was employed to calculate forensic parameters associated with the 24 SNP loci, including gene frequencies, genotype frequencies, minor allele frequencies, discrimination power (DP), polymorphism information content (PIC), and observed heterozygosity (Ho). Hardy-Weinberg equilibrium tests were conducted for each locus. The accuracy of the prediction model was assessed using SPSS 26.0 software, by calculating the area under the receiver characteristic operating curve (AUC). The receiver characteristic operating curve (ROC) is a curve that represents the relationship between the sensitivity (true positive rate, TPR) and specificity (false positive rate, FPR) of an analysis method, providing a comprehensive measure of the accuracy of the experiment. AUC is a criterion used to evaluate the performance of binary classification prediction models. Based on the AUC value and predefined assessment criteria, the accuracy of the model's predictions regarding eye and hair color in the sample population was evaluated.

RESULTS

Construction of multiplex genotyping assay system for SNPs

In this study, a multiplex genotyping assay system was developed for the analysis of 24 SNP loci. The system utilized primer sequences labeled with four different

Table 2. Allele frequencies and genotype frequencies of the 24 SNP loci (n = 235).

No.	dbSNP rs#	Allele frequencies				Genotype frequencies					
1	N29insA	C	1.000			C/C	1.000				
2	rs11547464	G	1.000			G/G	1.000				
3	rs885479	T	1.000			T/T	1.000				
4	rs1805008	C	1.000			C/C	1.000				
5	rs1805005	G	0.991	A	0.009	G/G	0.983	G/A	0.017		
6	rs1805006	C	1.000			C/C	1.000				
7	rs1805007	C	0.998	T	0.002	C/C	0.996	C/T	0.004		
8	rs1805009	G	0.998	A	0.002	G/G	0.996	G/A	0.004		
9	Y152OCH	C	1.000			C/C	1.000				
10	rs2228479	G	0.740	A	0.260	G/G	0.566	G/A	0.349	A/A	0.085
11	rs1110400	T	1.000			T/T	1.000				
12	rs28777	C	0.883	A	0. 117	C/C	0.774	C/A	0.217	A/A	0.009
13	rs16891982	C	1.000			C/C	1.000				
14	rs12821256	A	1.000			A/A	1.000				
15	rs4959270	C	0.738	A	0.262	C/C	0.583	C/A	0.311	A/A	0. 106
16	rs12203592	C	0.987	A	0.013	C/C	0.974	C/A	0.026		
17	rs1042602	G	0.998	T	0.002	G/G	0.996	G/T	0.004		
18	rs1800407	G	1.000			G/G	1.000				
19	rs2402130	G	0.055	A	0.945	G/A	0. 111	A/A	0.889		
20	rs12913832	C	0.004	T	0.996	C/T	0.009	T/T	0.991		
21	rs2378249	C	0. 164	T	0.836	C/T	0.302	T/T	0.685	C/C	0.013
22	rs12896399	G	0.677	T	0.323	G/G	0.455	G/T	0.443	T/T	0. 102
23	rs1393350	C	1.000			C/C	1.000				
24	rs683	G	0.994	T	0.006	G/G	0.991	G/T	0.004	T/T	0.004

fluorescent tags. With the exception of the rs683 locus, which exhibited non-specific detection peaks, the remaining 23 loci produced specific and high-yield products, yielding clear genotyping results (Figure 2).

Allele frequencies and genotype frequencies of the 24 SNP loci

The allele frequencies and genotype frequencies of the 24 SNP loci were determined by analyzing blood samples from 235 unrelated individuals of Han Chinese ethnicity in the Fujian region. Table 2 presents the obtained results.

Population genetic parameters of the 24 SNP loci

The population genetic parameters of the 24 SNP loci in the Han Chinese population of the Fujian region were calculated by using PowerStats v12 software. The obtained parameters include the discrimination power (DP), ranging from 0.000 to 0.552 with an average DP of 0.117, polymorphism information content (PIC), ranging from 0.000 to 0.310 with an average PIC of 0.065, and observed heterozygosity (Ho), ranging from 0.004 to 1.00 with an average Ho of 0.533.

To assess the genetic equilibrium of the genotype distribution among the population groups, a chi-squared test was performed using SPSS 26.0 software. The genotype distribution of all 24 SNP loci was found to be in accordance with Hardy-Weinberg equilibrium ($p > 0.05$). Please refer to Table 3 for the details of the population genetic parameters of the 24 SNP loci in the Han Chinese population of the Fujian region.

Definition of pigment phenotype classification and sample reading results

Following the manual reading of 2D photos and adhering to predefined rules for pigment phenotype classification, iris and hair color phenotype classifications were determined for the 235 samples. Among the individuals, there were 0, 1, and 234 individuals with blue, intermediate, and brown iris colors, respectively. Referring to Table 4, regarding hair color, there were 0, 0, 37, and 198 individuals with red, blond, brown, and black hair colors, respectively.

Table 3. The allele frequencies and genotype frequencies of the 24 SNP loci were determined by analyzing blood samples from 235 unrelated individuals of Han Chinese ethnicity in the Fujian region.

No.	dbSNP rs#	DP	PIC	Ho
1	N29insA	0.000	0.000	1.00
2	rs11547464	0.000	0.000	1.00
3	rs885479	0.000	0.000	1.00
4	rs1805008	0.000	0.000	1.00
5	rs1805005	0.033	0.020	0.017
6	rs1805006	0.000	0.000	1.00
7	rs1805007	0.008	0.000	0.004
8	rs1805009	0.008	0.000	0.004
9	Y152OCH	0.000	0.000	1.00
10	rs2228479	0.551	0.310	0.349
11	rs1110400	0.000	0.000	1.00
12	rs28777	0.353	0.190	0.217
13	rs16891982	0.000	0.000	1.00
14	rs12821256	0.000	0.000	1.00
15	rs4959270	0.552	0.310	0.311
16	rs12203592	0.050	0.020	0.026
17	rs1042602	0.008	0.000	0.004
18	rs1800407	0.000	0.000	1.00
19	rs2402130	0.197	0.100	0.111
20	rs12913832	0.017	0.01	0.009
21	rs2378249	0.439	0.240	0.302
22	rs12896399	0.586	0.340	0.443
23	rs1393350	0.000	0.000	1.00
24	rs683	0.017	0.01	0.004

Table 4. Classification results of iris and hair color phenotypes in the Han Chinese population of Fujian region.

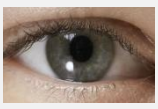
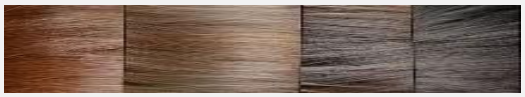
	Iris color			Hair color			
							
Phenotypes	**Blue**	**Intermediate**	**Brown**	**Red**	**Gold**	**Brown**	**Black**
Number	**0**	**1**	**234**	**0**	**0**	**37**	**198**

Table 5. Iris color inference results (n = 235).

Phenotype assessment	Blue (n = 0)	Intermediate (n = 1)	Brown (n = 234)
Number of individuals inferred as intermediate	0	0	0
Number of individuals inferred as brown	0	1	234
Accuracy of the inference	0	0%	99.6%

Table 6. Hair color inference results (n = 235).

Phenotype assessment	Red	Gold	Brown	Black
	n = 0	n = 0	n = 9	n = 226
Number of individuals inferred as red	0	0	0	0
Number of individuals inferred as gold	0	0	0	0
Number of individuals inferred as brown/dark brown	0	0	3	1
Number of individuals inferred as dark brown/black	0	0	6	22
Number of individuals inferred as black	0	0	28	175
Accuracy of the inferences	-	-	24.3%	99.5%

Table 7. Comparison of AUC for pigment phenotype inference.

AUC value	Han Fujian population	European population	Eurasian admixed population
Brown iris color inference	0.923	0.930	0.874
Brown hair color inference	0.500	0.820	0.575
Black hair color inference	0.980	0.870	0.635

Inference results and accuracy of pigment phenotype online model

The SNP genotyping results of the 235 samples were uploaded to the HIrisPlex online inference model (https://HIrisPlex.erasmusmc.nl/) [14] to obtain p-values for phenotype inference (Figure 3). The inferred iris color results showed 0 individuals with blue color, 1 individual with intermediate color, and 234 individuals with brown color, yielding an accuracy of 99.6% (Table 5). Regarding hair color inference, there were 0 individuals with red color, 0 individuals with blonde color, and counts for brown/dark brown and dark brown/black colors were 4 and 28, respectively, while 203 individuals had black hair. The accuracy of inference for brown hair was 24.3%, and for black hair, it was 99.5% (Table 6).

Calculation of accuracy of the international pigmentation phenotype inference model for the Han population in Fujian

The accuracy of the International Pigmentation Phenotype Inference Model in predicting pigmentation phenotypes among the Han population in Fujian was objectively assessed by using ROC curve analysis in SPSS 26.0 software. The AUC values were calculated for each pigmentation phenotype inference, providing a quantitative measure of the model's accuracy. The model demonstrated a high level of accuracy in predicting brown eye color among the 235 individuals from the Han population in Fujian, with an AUC value of 0.923. This performance was comparable to the model's predictions for brown eye color in European populations (AUC = 0.930), indicating a similar level of accuracy. However, the accuracy of the model's predictions for brown eye color in Eurasian admixed populations was lower, with an AUC of 0.874, suggesting relatively lower accuracy compared to the Han population in Fujian. In terms of hair color, the model's predictions for brown and black hair color in the Han population in Fujian had AUC values of 0.500 and 0.980, respectively. These AUC values differed from the model's performance in predicting hair color in European populations (brown hair AUC = 0.820, black hair AUC = 0.870) and Eurasian admixed populations (brown hair AUC = 0.575, black hair AUC = 0.635). Since the majority of the Han population in Fujian had black hair, the AUC value for predicting black hair color was higher compared to European and Eurasian admixed populations (Table 7).

DISCUSSION

The utilization of EVC identification technology, utilizing biological samples, can offer valuable reference clues for cases and assist in narrowing down the pool of suspects, thereby enhancing the success rate of solving intricate cases. At the heart of EVC identification technology lies the international pigment phenotype inference model, which relies on the genotyping results of 24 SNPs associated with iris and hair color in the molecular characterization of the human body. This inference model stands as the most precise system for deducing physical appearance and serves as the foundational

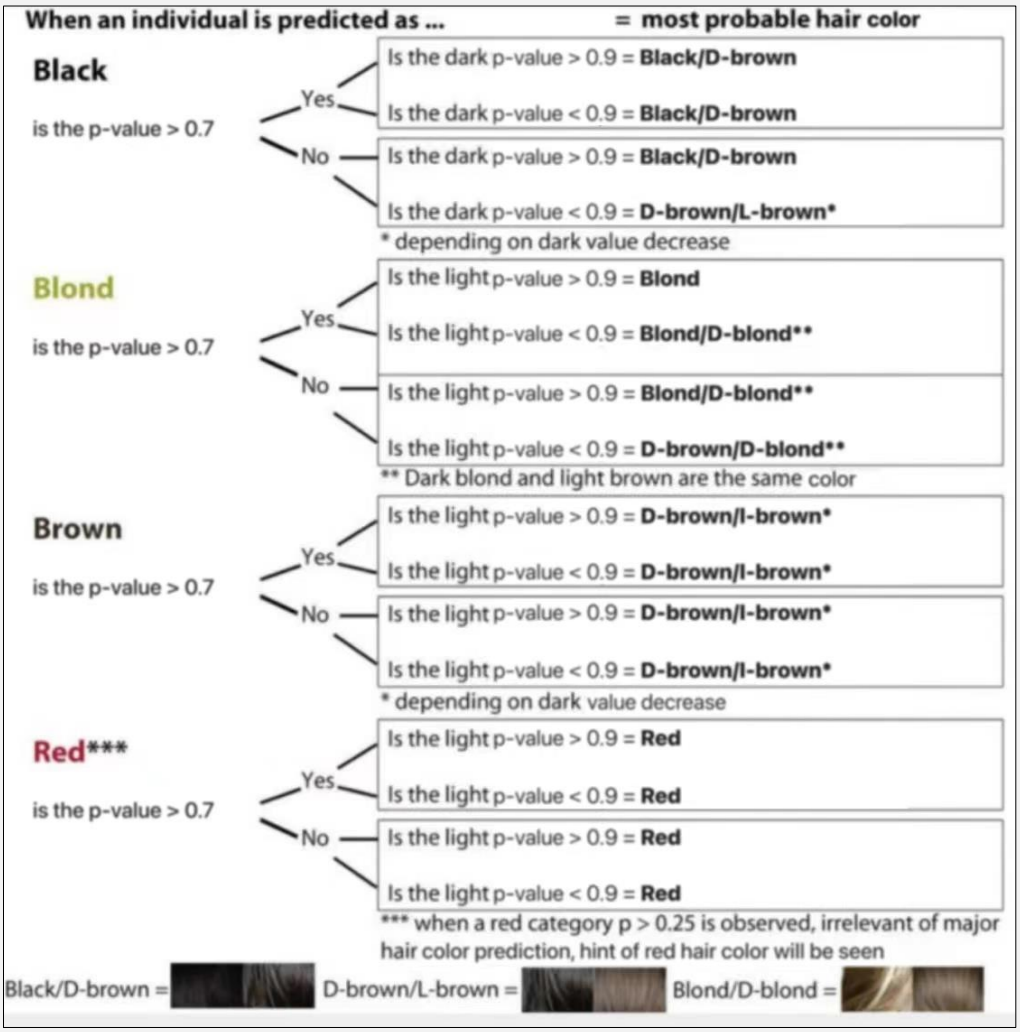


Figure 1. Rules for inferring hair color.

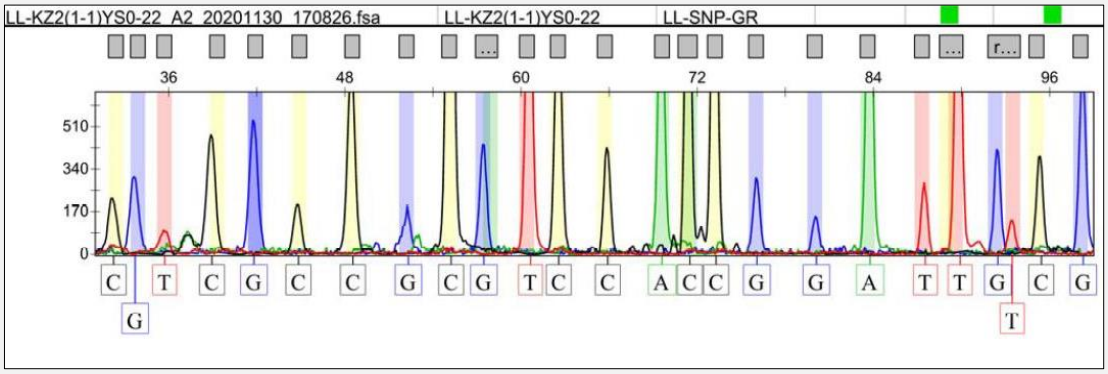


Figure 2. Composite genotyping profiles of 24 SNP loci.

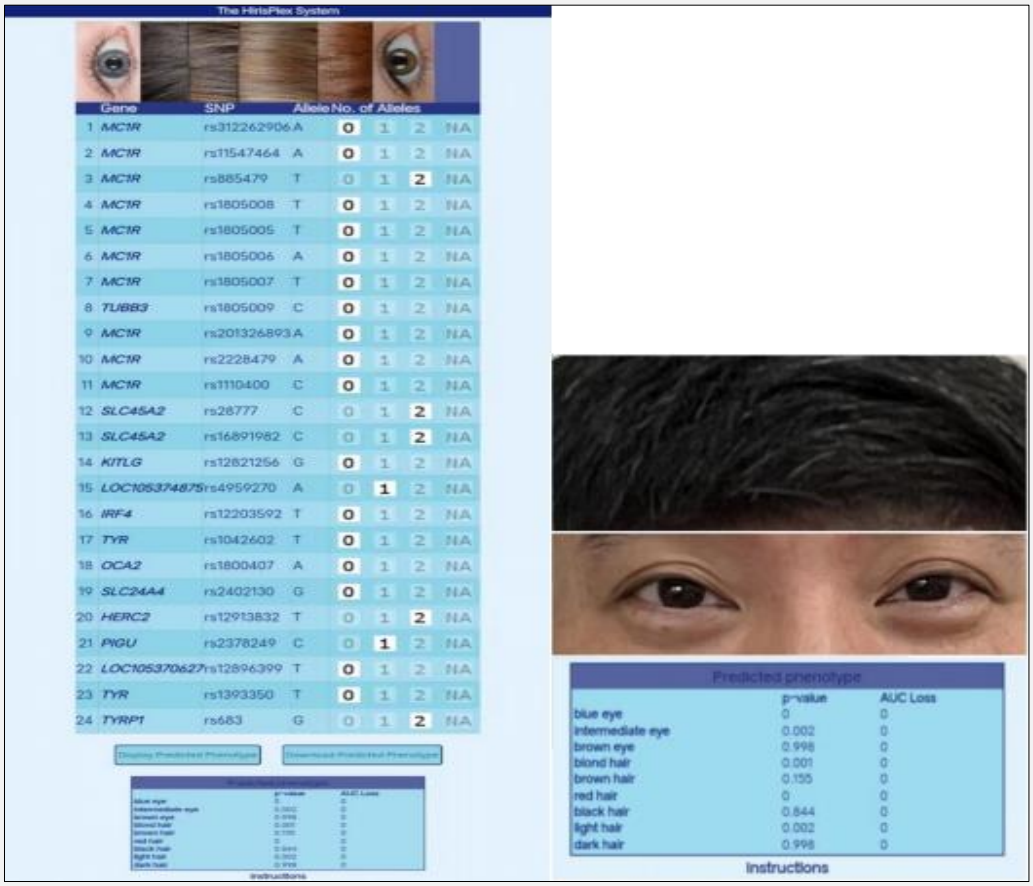


Figure 3. Phenotype inference model and p-values.

pillar of EVC identification technology. However, the majority of research on pigment phenotypes is predominantly based on European populations, with limited studies focusing on East Asian populations. Currently, China lacks a mature EVC identification technology and the population-based validation of inference systems developed from foreign populations.

The constructed composite genotyping system for 24 SNPs of human molecular characterization provided the allele frequencies, genotype frequencies, minimum allele frequencies, and forensic genetic parameters (e.g., DP, PIC, and Ho) for each SNP locus in the Han population of Fujian. The loci rs12896399, rs2228479, rs4959270, and rs2378249 exhibited high DP and PIC values, indicating their significant polymorphism in the Han population of Fujian and their valuable applications.

The statistical results of phenotype inference in this study showed that the detection system and inference model achieved an accuracy of 99.6% for brown irises, 24.3% for brown hair, and 99.5% for black hair. The online inference model demonstrated high accuracy in inferring brown iris color and black hair color in the Han population of Fujian, with rates close to 100%, surpassing the European database. However, the accuracy rates for inferring intermediate iris color and brown hair color were relatively low, possibly due to two reasons: (1) the majority of the Han population in Fujian has brown iris color and black hair, resulting in low phenotypic polymorphism for intermediate iris color and brown hair; (2) the classification of pigment phenotypes in this study relied on manually reading 2D photos, which are susceptible to factors such as camera pixels, photography environment, and lighting, leading to misjudgments of phenotypic pigment features in some samples. Future research can enhance the classification of pigment phenotypes by increasing the sample size and employing more accurate photo acquisition and identification instruments.

The phenotype inference results of this study indicate that the detection system and international pigment phenotype inference model achieved a high accuracy rate of

99.6% for inferring brown iris color and 99.5% for inferring black hair color in the Han population of Fujian, with AUC values of 0.923 and 0.980, respectively.
Validation of 24 SNPs system was carried out based on the HIrisPlex model [15]. After a series of experiments, including statistics of population genetic parameters and case sample tests, the results indicate that the 24 SNPs system demonstrates a high level of accuracy and represents the remarkable tool for simultaneously establishing categorical eye and hair color of a person from DNA [16]. The practical forensic application of the 24 SNPs system is expected to benefit the prediction of human pigmentation traits from DNA in criminal cases, missing person cases, and other avenues of investigation [17]. The system is also the supplementary means when the STR profiling provides no leads on who the unknown crime scene sample donor or the unknown missing person might be. Moreover, forensic DNA phenotyping (FDP) by utilizing SNPs with high DP and PIC can reveal the externally visible characteristics of an unknown individual. We advocate for screening more reliable genetic markers in future prediction modeling for forensic individual identification and bio-geographic ancestry inference.

Acknowledgment:
The authors are grateful to the patients for their participation.

Source of Funds:
This work was supported by the Natural Science Foundation of Fujian Province (2021J01354) and the startup Fund for Scientific Research of Fujian Medical University (grant number: 2020QH1176).

Declaration of Interest:
The authors have no conflicts of interest to declare.

References:

1. Zhang J, Zhang J, Tao R, Yang Z, Zhang S, Li C. Mass spectrometry-based SNP genotyping as a potential tool for ancestry inference and human identification in Chinese Han and Uygur populations. Sci Justice 2019 May;59(3):228-233.(PMID: 31054813)

2. Tully G. Genotype versus phenotype: human pigmentation. Forensic Sci Int Genet 2007;1(2):105-10. (PMID: 19083738)

3. Kayser M, Schneider PM. DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. Forensic Sci Int Genet 2009;3(3):154-61. (PMID: 19414162)

4. Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 2011;12(3):179-92. (PMID: 21331090)

5. Toom V, Wienroth M, M'charek A, et al. Approaching ethical, legal and social issues of emerging forensic DNA phenotyping (FDP) technologies comprehensively: Reply to 'Forensic DNA phenotyping: Predicting human appearance from crime scene material for investigative purposes' by Manfred Kayser. Forensic Sci Int Genet 2016 May;22:e1-e4. (PMID: 26832996)

6. MacLean CE, Lamparello A. Forensic DNA phenotyping in criminal investigations and criminal courts: assessing and mitigating the dilemmas inherent in the science. Recent Adv DNA Gene Seq 2014;8(2):104-12.(PMID: 25687339)

7. Muiños Díaz Y, Saornil MA, Almaraz A, Muñoz-Moreno MF, García C, Sanz R. Iris color: validation of a new classification and distribution in a Spanish population-based sample. Eur J Ophthalmol 2009;19(4):686-9. (PMID: 19551689)

8. Han J, Kraft P, Nan H, et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 2008;4(5):e1000074. (PMID: 18483556)

9. Allwood JS, Harbison S. SNP model development for the prediction of eye colour in New Zealand. Forensic Sci Int Genet 2013; 7(4):444-52. (PMID: 23597786)

10. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 2011;5(3):170-80. (PMID: 20457092)

11. 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):464-71. (PMID: 20947461)

12. Wollstein A, Walsh S, Liu F, et al. Novel quantitative pigmentation phenotyping enhances genetic association, epistasis, and prediction of human eye colour. Sci Rep 2017;7:43359. (PMID: 28240252)

13. Grimes EA, Noake PJ, Dixon L, Urquhart A. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. Forensic Sci Int 2001;122(2-3):124-9. (PMID: 11672965)

14. Carratto TMT, Marcorin L, do Valle-Silva G, et al. Prediction of eye and hair pigmentation phenotypes using the HIrisPlex system in a Brazilian admixed population sample. Int J Legal Med 2021; 135(4):1329-39. (PMID: 33884487)

15. Chaitanya L, Breslin K, Zuñiga 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;35:123-35. (PMID: 29753263)

16. 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. (PMID: 22917817)

17. Breslin K, Wills B, Ralf A, et al. HIrisPlex-S system for eye, hair, and skin color prediction from DNA: Massively parallel sequencing solutions for two common forensically used platforms. Forensic Sci Int Genet 2019;43:102152. (PMID: 31518964)