**ORIGINAL ARTICLE**

# Prediction of eye and hair pigmentation phenotypes using the HIrisPlex system in a Brazilian admixed population sample

Thássia Mayra Telles Carratto[1] · Letícia Marcorin[2] · Guilherme do Valle-Silva[1] ·
Maria Luiza Guimarães de Oliveira[2] · Eduardo Antônio Donadi[3] · Aguinaldo Luiz Simões[2] · Erick C. Castelli[4] ·
Celso Teixeira Mendes-Junior[1]

## Abstract

Human pigmentation is a complex trait, probably involving more than 100 genes. Predicting phenotypes using SNPs present in those genes is important for forensic purpose. For this, the HIrisPlex tool was developed for eye and hair color prediction, with both models achieving high accuracy among Europeans. Its evaluation in admixed populations is important, since they present a higher frequency of intermediate phenotypes, and HIrisPlex has demonstrated limitations in such predictions; therefore, the performance of this tool may be impaired in such populations. Here, we evaluate the set of 24 markers from the HIrisPlex system in 328 individuals from Ribeirão Preto (SP) region, predicting eye and hair color and comparing the predictions with their real phenotypes. We used the HaloPlex Target Enrichment System and MiSeq Personal Sequencer platform for massively parallel sequencing. The prediction of eye and hair color was accomplished by the HIrisPlex online tool, using the default prediction settings. Ancestry was estimated using the SNP*for*ID 34-plex to observe if and how an individual's ancestry background would affect predictions in this admixed sample. Our sample presented major European ancestry (70.5%), followed by African (21.1%) and Native American/East Asian (8.4%). HIrisPlex presented an overall sensitivity of 0.691 for hair color prediction, with sensitivities ranging from 0.547 to 0.782. The lowest sensitivity was observed for individuals with black hair, who present a reduced European contribution (48.4%). For eye color prediction, the overall sensitivity was 0.741, with sensitivities higher than 0.85 for blue and brown eyes, although it failed in predicting intermediate eye color. Such struggle in predicting this phenotype category is in accordance with what has been seen in previous studies involving HIrisPlex. Individuals with brown eye color are more admixed, with European ancestry decreasing to 62.6%; notwithstanding that, sensitivity for brown eyes was almost 100%. Overall sensitivity increases to 0.791 when a 0.7 threshold is set, though 12.5% of the individuals become undefined. When combining eye and hair prediction, hit rates between 51.3 and 68.9% were achieved. Despite the difficulties with intermediate phenotypes, we have shown that HIrisPlex results can be very helpful when interpreted with caution.

**Keywords** SNPs · Massively parallel sequencing · DNA phenotyping · Forensic DNA phenotyping · Forensic genetics

✉ Celso Teixeira Mendes-Junior
ctmendes@ffclrp.usp.br

[1] Departamento de Química, Laboratório de Pesquisas Forenses e Genômicas, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, Ribeirão Preto, SP 14040-901, Brazil

[2] Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP 14049-900, Brazil

[3] Divisão de Imunologia Clínica, Departamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP 14048-900, Brazil

[4] Departamento de Patologia, Faculdade de Medicina de Botucatu, Unesp - Universidade Estadual Paulista, Botucatu, SP 18618-970, Brazil

# Introduction

Human pigmentation is a polygenic trait [1] that is primarily determined by the presence of melanin in the epidermis, iris, and hair [2]. Among the many different genes involved in pigmentation, *TYR*, *TYRP1*, *OCA2*, *SLC45A2*, *SLC24A5*, *MC1R*, *ASIP*, *KITLG*, *HERC2*, *SLC24A4*, *IRF4*, *TPNC2*, and *BNC2* have been widely studied [1–11]. Many studies investigating associations between SNPs and pigmentation phenotypes have been done in admixed populations [12–25], including the Brazilian one. It has already been reported that some polymorphic sites in *SLC24A5*, *SLC45A2*, *OCA2*, and *HERC2* may influence pigmentation of Brazilians and can be considered for forensic applications [21–24].

Eye color was the first characteristic to have a validated prediction method. The IrisPlex system is composed of six SNPs highly correlated with the human iris color [26, 27]. According to the IrisPlex developers, accuracy to determine blue and brown eyes is above 90%. However, the tool presents some limitations regarding intermediate eye color predictions, possibly due to imprecise phenotype characterization as well as unidentified SNPs that may influence this phenotype [26–28]. Various studies have evaluated the IrisPlex system in populations from Europe (Italy, Portugal, and Slovenia) and other countries/continents (Asia, Brazil, USA, and Venezuela) [29–35], corroborating that IrisPlex is not suitable to predict intermediate eyes. This issue is particularly critical in admixed populations since they present a higher frequency of intermediate eye phenotypes due to the genetic contribution of Europeans (with eye color variation) and non-Europeans (with little eye color variation).

In 2013, the HIrisPlex system was developed aiming for the simultaneous prediction of hair (based on 22 markers) and eye colors (based on the earlier published IrisPlex system) [36, 37]. One limitation for predicting hair color is the age-dependent changes in hair color; i.e., sometimes hair becomes darker from childhood to adulthood. The molecular basis of hair darkening is not clear yet, and the HIrisPlex model is expected to predict the individual's hair color from early childhood instead of the darker hair color of advanced childhood [38]. However, it should be emphasized that many of the individuals who were blond as a child remain blond as an adult [36, 37]. Moreover, additional markers are necessary in order to improve the prediction and raise the performance of the system.

Other predictive tools for pigmentation traits have been proposed. The HIrisPlex-S system was developed to predict skin color based on 36 SNPs, in addition to the previous models for eye and hair color prediction [39, 40].

Hart et al. proposed models consisting of eight SNPs for eye and skin color prediction. These models were developed with North-American volunteers and presented an error rate of approximately 5% for eye color prediction, while no errors were observed for skin color prediction, although 38% of the results were inconclusive (i.e., they were not light or not dark) [41]. Allwood et al. also proposed a model for eye color prediction using classification trees. This tool was developed aiming to be useful in New Zealand, which includes European descendants and other minority groups, such as indigenous people, Polynesians, and multiple groups of Asian ancestry. The overall accuracy for the Allwood model was 79% [42]. There are also the Snipper sets for eye, skin, and hair color prediction [43–45]. Although these tools were developed from European association studies, the sets for eye color prediction were evaluated in samples from Northeastern Brazil and Venezuela [34]. A better performance was observed for brown and blue eyes, but the tools presented difficulties in predicting intermediate eyes, achieving low sensitivity levels [34].

HIrisPlex remains as the most widely studied Forensic DNA Phenotyping (FDP) tool so far. However, HIrisPlex sensitivity in admixed populations needs to be deeper evaluated. Admixed American populations present particular genetic and demographic histories, and therefore, the evaluation of the informativeness of this tool in admixed urban samples is necessary to verify the overall amplitude and reliability of its use in the forensic practice.

The Brazilian urban population is characterized by different levels of African, European, and Amerindian admixture proportions [46] and provides an ideal setting to evaluate HIrisPlex sensitivity in admixed samples. Thus, here we have sampled 328 Brazilian individuals with different phenotypes, called genotypes using NGS (next-generation sequencing) and predicted phenotypes using the HIrisPlex online tool. Then, we compared the predictions and the real phenotypes to assess HIrisPlex performance in admixed samples. We have also calculated ancestry proportions for each individual, in order to evaluate if the admixed nature of this Brazilian sample could somehow interfere with the predictions performed.

# Materials and methods

## Population sample

This study was approved in its ethical aspects by the research ethics committee of this institution (Comitê de Ética em Pesquisa, FFCLRP-USP), according to protocol *CAAE # 25696413.7.0000.5407*.

We collected blood samples from 328 volunteers from the city of Ribeirão Preto and proximities, Southeastern Brazil, consisting of 159 women and 169 men with ages ranging from 18 to 72 years. Of those, 197 individuals were

randomly sampled, while the remaining 131 were invited because they present phenotypes that are not common in Brazil. At least two independent observers (members of our research group) assigned each volunteer in groups concerning three categories of eye color (blue, intermediate [green/hazel], and brown), and four categories of hair color (red, blond, brown, and black). These data were used to compare HIrisPlex predictions with actual phenotypes from volunteers.

In order to evaluate the effect of age-dependent hair color changes in HIrisPlex predictions, the volunteers indicated whether or not their hair color changed from childhood to adulthood, what kind of change have happened, and the estimated age in which this change has occurred.

## Laboratory analysis

DNA was extracted using the salting-out protocol [47]. We assessed genomic DNA quality using agarose gel electrophoresis for integrity, NanoDrop spectrophotometry (Thermo Fisher Scientific Inc.) for purity, and Qubit™ dsDNA BR fluorimetric assay (Thermo Fisher Scientific Inc.) for concentration. We also normalized DNA samples to 5 ng/μL to achieve an ideal concentration for the DNA sequencing library preparation.

We prepared sequencing libraries using a customized HaloPlex Target Enrichment System (Agilent Technologies, Inc.) protocol that included probes to capture the exonic and regulatory regions of various genes involved in pigmentation, including the 24 HIrisPlex markers and the 34 SNP-*for*ID 34-plex Ancestry-Informative Markers (AIMs) [48, 49]. The probe panel was designed using the SureDesign tool (Agilent Technologies, Inc.) and the hg19/GRCh37 human genome as reference.

Following the manufacturer instructions, 5 μL of each sample were digested by eight different pairs of enzymes to create libraries of DNA fragments. These fragments were captured using HaloPlex biotinylated probes, and indices were incorporated for sample identification. Lastly, the captured fragments were amplified by PCR using the Herculase II Fusion polymerase in the SureCycle 8800 thermocycler (Agilent Technologies, Inc.). The amplified fragments were purified using AMPure XP magnetic beads (Beckman Coulter) and resuspended in Tris–HCL buffer (pH 8.0). Each sample library was kept at − 20 °C until sequencing. DNA libraries were quantified before sequencing using *Qubit® 2.0 Fluorometer* (Thermo Fisher Scientific) and *2100 Bioanalyzer* (Agilent Technologies, Inc.). A pool of DNA libraries, consisting of up to 96 samples (4 nmol/L–Supplementary Table 1), is then diluted to 16 pM as recommended by the manufacturer (protocol available on https ://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq /miseq-denature-dilute-libraries-guide-15039740-10.pdf) and inserted as input for paired-end sequencing using the MiSeq Reagent kit V3 (600 cycles), in the MiSeq Personal Sequencer (Illumina Inc.) [50].

## Genotype calling

Sequencing adaptors were trimmed using the cutadapt software [51]. We aligned sequencing reads using the BWA-MEM algorithm (Burrows-Wheeler) [52] and the human reference genome GRCh37/hg19. Genotyping was performed using the GATK HaplotypeCaller in the GVCF mode [53]. VCF files were further processed by *vcfx checkpl* (version 2.0b, available at www.castelli-lab.net/apps/vcfx) to introduce missing alleles when the genotype likelihood was under 99%, assuring that only high-quality genotypes are used for prediction purposes. It was not possible to use HaplotypeCaller to call the InDel rs312262906, which is rare in our population. However, the Integrative Genomics Viewer (IGV) software [54] allows the user to visualize the NGS data in a base-to-base resolution. Thus, it was possible to view the InDel locus and determine whether the individuals had this polymorphism or not. Only one individual presented this allele, in heterozygous form. As two markers (rs16891982 and rs1805009) have C > G mutations, genotype calls were made evaluating surrounding sequence information and the human reference genome. We also compare the frequency of their alleles with global frequencies from the 1000 Genomes dataset [55] (Supplementary Table 2). Due to Brazil's historic and demographic backgrounds, and to ancestry analysis performed in this sample [56, 57], it is expected that allele frequencies will be closer to European's frequencies than to those of other populations. Since the number of samples differed in the five sequencing runs performed (Supplementary Table 1), we calculate a weighted average considering the total number of reads and the number of samples for each run to determine the read depth.

## 1000 genomes population data

For comparisons, allele frequencies were calculated for the five super population groups from the 1000 Genomes Project dataset [55]. Sample sizes and names of the subpopulations that compose each group can be found in Supplementary Table 3.

## Statistical analysis

The GENEPOP 4.51 software [58] was used to estimate allele and genotype frequencies, observed and expected heterozygosity, and Hardy–Weinberg equilibrium (HWE).

The ancestry proportions of each individual were inferred using the SNP*for*ID 34-plex AIMs [48, 49] and the STRUCTURE 2.3.4 program [59], as published elsewhere [56]. For this purpose, it was applied the admixture model with correlated allele frequencies, 100,000 burn-in steps followed by 100,000 Markov Chain Monte Carlo interactions, three clusters (k = 3) representing the three main Brazilian ancestral components, in 100 independent runs. A total of 1,412 individuals were used to represent the parental populations: 404 Europeans (TSI, FIN, GBR, IBS), 504 Africans (YRI, LWK, GWD, MSL, ESN), and 504 East Asians (CHB, JPT, CHS, CDX, KHV). Since there is not a Native American population in the 1000 Genomes dataset, we used the East Asian population in order to replace it.

Eye and hair color phenotype predictions were performed by the HIrisPlex online tool (available on https://hirisplex.erasmusmc.nl/). Based on the six markers from the original IrisPlex system, the tool estimates probabilities for each eye color phenotype: blue, intermediate, and brown. For eye color, we performed predictions in two levels: with no threshold, i.e., the phenotype with higher probability is assumed to be the actual individual's phenotype, and applying a 0.7 threshold, in which a given eye color is assumed only when it is estimated with a probability equal or higher than 0.7 [26, 27]. Based on 22 markers, hair color (blond, brown, red, and black) and shade (light and dark) were also predicted. The prediction results were compared with actual phenotypic data from each volunteer, which was obtained based on observations made while sampling. Parameters such as AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for each phenotypic characteristic. SPSS 20.0 (IBM) software was used to draw receiver operating characteristic (ROC) curves and to calculate the associated AUC. The overall sensitivity (hit rate of combined eye and hair prediction) of HIrisPlex system was also obtained.

## Results

Table 1 presents the allele frequencies of the 24 HIrisPlex markers in a sample of 197 randomly chosen Brazilian individuals. The rs28777 marker is the only one not fitting Hardy–Weinberg expectations ($p = 0.0305$, Table 1). The Supplementary Table 2 presents the allele frequencies for the entire sample and for the five super populations of the 1000 Genomes dataset.

Regarding hair color, we excluded eight individuals because they were bald or presented grey hair, making it impossible to verify their real hair color. Assuming four hair categories (blond, brown, black, and red) and that the highest probability value determines the actual hair color of the individual, the comparison of predicted outcomes with actual phenotypic data revealed an overall sensitivity rate of 0.691. Tables 2 and 3 present the results of hair color classification and the corresponding statistical values for AUC, sensitivity, specificity, PPV, and NPV for each phenotypic category. HIrisPlex predicted brown and red hair with higher sensitivity, showing values below 0.620 for the other two categories. Notwithstanding that, all AUCs values were above 0.700.

As mentioned before, the HIrisPlex system may present some difficulties in predicting individuals of blond/brown hair due to age-dependent changes in hair color. Sixty-six out of the 320 volunteers declared that their hair has darkened over the years. While this process was not strong enough to change hair color category of 35 individuals (23 with blond and 12 with brown hair), 25 individuals with blond hair in childhood turned into brown-haired adults, while 6 adults with black hair used to have blond (1) or brown (5) hair while young. Figure 1 provides a detailed description of the 66 cases of age-dependent changes in hair color, showing the individual hair color in childhood, the current phenotype of these individuals, the predictions by HIrisPlex, and whether or not the prediction is in agreement with the childhood and/or adulthood phenotype.

For eye color, we compared predictions with or without using a probability threshold level of 0.7 to verify which one provides better results, as described in the Materials and Methods section. Tables 4 and 5 present the results of eye color classification obtained using the six IrisPlex markers from the HIrisPlex system and the corresponding statistical values for AUC, sensitivity, specificity, PPV, and NPV for each phenotypic category. Predictions without using the 0.7 threshold revealed an overall sensitivity of 0.741. By establishing a 0.7 threshold, overall sensitivity achieves 0.791, but 12.5% of the individuals become undefined. In both analyses, no individual presented the highest probability for intermediate eyes category, so none of them was classified as having intermediate eyes. As expected, HIrisPlex reveals a better performance when predicting extreme phenotypes (blue and brown).

Results of HIrisPlex prediction for eye (using the two approaches evaluated) and hair color (following the guide or not) for each individual were considered together for determination of overall combined prediction sensitivity. When the 0.7 threshold was not applied, the total hit rate was only 51.3%. When applying the 0.7 threshold, the hit rate increased for 63.9%.

In order to evaluate the impact of ancestry on the sensitivities presented above, the SNP*for*ID 34-plex AIMs was used to infer individual heritage. Our total sample presents European ancestry as major contributor (70.5%), followed by African (21.1%) and Native American/East Asian (8.4%) contributions. Ancestry proportions for individuals stratified according to eye and hair phenotypes are

**Table 1** Description of the 24 markers from the HIrisPlex System surveyed in a random sample of 197 individuals from Ribeirão Preto, Brazil, and probabilities of adherence to the Hardy–Weinberg equilibrium (*p*HWE)

| SNP | $n^a$ | Read depth (weighted average) [b] | Reference allele (frequency) [c] | Observed heterozygosity | Expected heterozygosity | *p*HWE |
|-----|------|------|------|------|------|------|
| rs16891982 | 197 | 482.9 | C (0.383) | 0.4112 | 0.4739 | 0.0710 |
| rs28777 | 196 | 113.4 | C (0.314) | 0.3622 | 0.4317 | **0.0305** |
| rs12203592 | 197 | 284.4 | C (0.939) | 0.1015 | 0.1147 | 0.1491 |
| rs4959270 | 196 | 269.5 | C (0.569) | 0.4439 | 0.4918 | 0.1910 |
| rs683 | 194 | 131.5 | C (0.466) | 0.5000 | 0.4990 | 1.0000 |
| rs1042602 | 197 | 402.7 | C (0.668) | 0.4112 | 0.4450 | 0.3350 |
| rs1393350 | 197 | 549.6 | G (0.817) | 0.2741 | 0.2995 | 0.2368 |
| rs12821256 | 197 | 352.0 | T (0.959) | 0.0711 | 0.0781 | 0.2718 |
| rs12896399 | 197 | 147.5 | G (0.713) | 0.3807 | 0.4101 | 0.3827 |
| rs2402130 | 197 | 251.4 | G (0.269) | 0.3959 | 0.3943 | 1.0000 |
| rs1800407 | 197 | 238.4 | C (0.942) | 0.1168 | 0.1102 | 1.0000 |
| rs12913832 | 197 | 184.1 | A (0.637) | 0.4315 | 0.4636 | 0.3566 |
| rs312262906 | 197 | 419.4 | C (1.000) | 0.0000 | 0.0000 | - |
| rs1805005 | 196 | 545.8 | G (0.903) | 0.1837 | 0.1755 | 1.0000 |
| rs1805006 | 196 | 310.5 | C (0.992) | 0.0153 | 0.0152 | 1.0000 |
| rs2228479 | 197 | 312.3 | G (0.952) | 0.0964 | 0.0920 | 1.0000 |
| rs11547464 | 197 | 621.2 | G (0.997) | 0.0051 | 0.0051 | - |
| rs1805007 | 197 | 623.5 | C (0.992) | 0.0152 | 0.0152 | 1.0000 |
| rs1110400 | 197 | 623.4 | T (0.997) | 0.0051 | 0.0051 | - |
| rs1805008 | 197 | 635.5 | C (0.980) | 0.0406 | 0.0399 | 1.0000 |
| rs885479 | 197 | 628.9 | G (0.934) | 0.1117 | 0.1236 | 0.1955 |
| rs1805009 | 197 | 624.9 | G (0.980) | 0.0406 | 0.0399 | 1.0000 |
| rs201326893 | 197 | 623.7 | C (1.000) | 0.0000 | 0.0000 | - |
| rs2378249 | 197 | 293.0 | G (0.104) | 0.1878 | 0.1869 | 1.0000 |

Values in boldface indicate significant deviations from equilibrium expectations

[a]Number of individuals successfully genotyped

[b]Weighted average of the number of reads obtained for 328 samples split in five sequencing assays (Supplementary Table 1)

[c]Frequency of the reference allele in the reference genome hg19

depicted in Table 6. The admixed nature of this Brazilian sample is evident even in individuals with blue eyes and blond hair. European ancestry decreases among darker phenotypes, being as low as 48.4% among individuals with black hair.

## Discussion

The analysis of the HIrisPlex set of markers was proposed using the *SNaPshot* Multiplex (*Thermo Fisher Scientific*) minisequencing procedure [37]. However, here we have

**Table 2** HIrisPlex classification for each different hair color category, based on the analysis of 320 individuals from a Brazilian admixed population sample

| True hair phenotype | Predicted hair phenotype (%) | | | |
|-----|------|------|------|------|
| | Red | Blond | Brown | Black |
| Red (*n*=11) | 72.7 | 0.00 | 27.3 | 0.0 |
| Blond (*n*=49) | 4.1 | 61.2 | 32.6 | 2.1 |
| Brown (*n*=174) | 0.0 | 7.5 | 78.2 | 14.4 |
| Black (*n*=86) | 0.0 | 0.0 | 45.4 | 54.7 |

**Table 3** HIrisPlex summary statistics (AUC, sensitivity, specificity, PPV, and NPV) for each different hair color category, based on the analysis of 320 individuals from a Brazilian admixed population sample

| Hair color category | Summary statistics | | | | |
|-----|------|------|------|------|------|
| | AUC | Sensitivity | Specificity | PPV | NPV |
| Red | 0.992 | 0.727 | 0.994 | 0.800 | 0.990 |
| Blond | 0.888 | 0.612 | 0.952 | 0.698 | 0.931 |
| Brown | 0.700 | 0.782 | 0.603 | 0.701 | 0.698 |
| Black | 0.841 | 0.547 | 0.889 | 0.644 | 0.842 |

**Fig. 1** HIrisPlex predictions for the individuals who reported age-dependent changes in hair color. (**a**) Prediction compatible with both childhood and adulthood phenotype; (**b**) prediction compatible only with childhood phenotype; (**c**) prediction compatible only with adulthood phenotype; (**d**) prediction incompatible with either childhood or adulthood
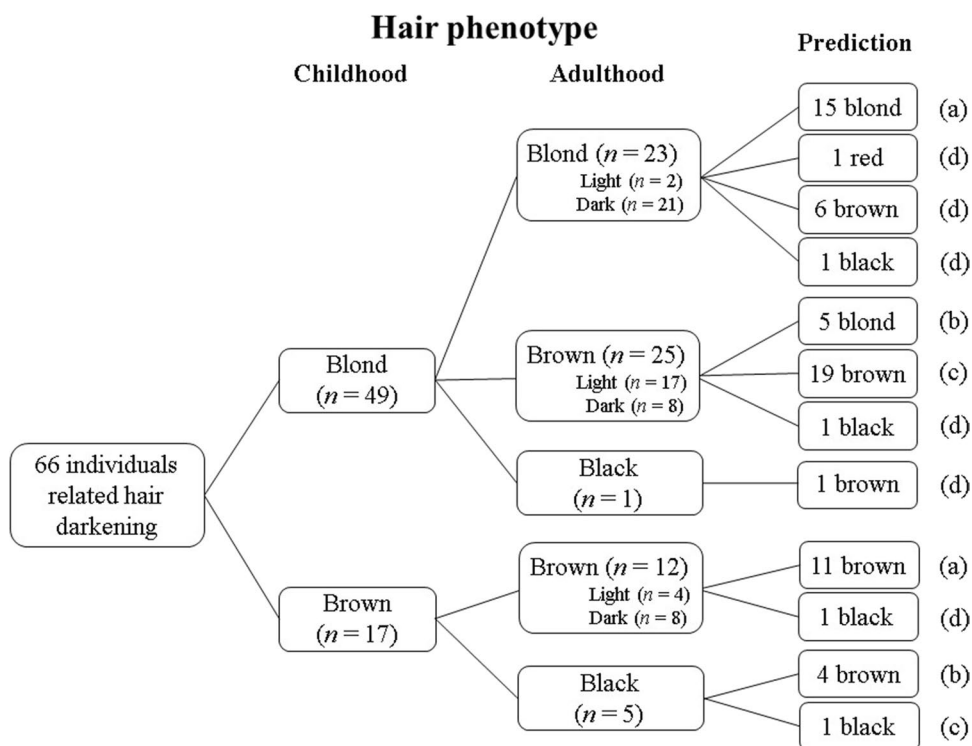


evaluated these 24 markers using an NGS platform (Illumina Inc.), which provides high accuracy for genotype calls [57, 60]. Since we aimed for sequencing with a high depth of coverage, we achieve averages ranging from 113.4 to 635.5 reads (Table 1). Therefore, we successfully called 17 out of 24 loci in all samples, with complete profiles for 320 out of the 328 individuals. Overall, there were only 10 missing genotypes, spread across eight samples (seven samples presented only one missing genotype, while one sample presented three missing genotypes). Moreover, the high depth of coverage allowed the call of reliable genotypes. We observed only one deviation from Hardy–Weinberg equilibrium (Table 1), which can be attributed to chance. However, it should be noted that this marker also presents HWE deviations in TSI and MSL populations from the 1000 Genomes Project. It is important to emphasize that the allelic frequencies obtained in this study are, on average, more similar to those of AMR and EUR groups from the 1000 Genomes Project. Except for two markers (rs1805009 and rs2378249), allele frequencies are always placed between the values observed in EUR and AFR/EAS/SAS, being usually closer to EUR (Supplementary Table 2).

The HIrisPlex website displays the AUC, sensitivity, specificity, PPV, and NPV values, associated with their standard deviations of 1000 cross-validation tests using a

**Table 4** HIrisPlex classification for each different eye color category, with or without the 0.7 probability threshold, based on the analysis of 328 individuals from a Brazilian admixed population sample

| True eye phenotype | Predicted eye phenotype (%) | | | |
|---|---|---|---|---|
| | Blue | Intermediate | Brown | Undefined |
| Blue ($n=37$) | 86.5 | - | 13.5 | - |
| Intermediate ($n=78$) | 57.7 | 0.0 | 42.3 | - |
| Brown ($n=213$) | 0.9 | - | 99.1 | - |

| True eye phenotype | Predicted eye phenotype (%) (0.7 threshold) | | | |
|---|---|---|---|---|
| | Blue | Intermediate | Brown | Undefined |
| Blue ($n=37$) | 81.1 | - | 8.1 | 10.8 |
| Intermediate ($n=78$) | 44.9 | 0.0 | 25.6 | 29.5 |
| Brown ($n=213$) | 0.9 | - | 92.5 | 6.6 |

**Table 5** HIrisPlex summary statistics (AUC, sensitivity, specificity, PPV, and NPV) for each different eye color category, based on the analysis of 328 individuals from a Brazilian admixed population sample

| Eye color category No threshold | Summary Statistics | | | | |
|---|---|---|---|---|---|
| | AUC | Sensitivity | Specificity | PPV | NPV |
| Blue | 0.891 | 0.865 | 0.838 | 0.405 | 0.980 |
| Intermediate | 0.751 | 0.000 | 1.000 | - | 0.762 |
| Brown | 0.933 | 0.991 | 0.670 | 0.847 | 0.975 |

| Eye color category 0.7 threshold | Summary Statistics | | | | |
|---|---|---|---|---|---|
| | AUC | Sensitivity | Specificity | PPV | NPV |
| Blue | 0.891 | 0.909 | 0.854 | 0.448 | 0.986 |
| Intermediate | 0.751 | 0.000 | 1.000 | - | 0.808 |
| Brown | 0.933 | 0.990 | 0.739 | 0.895 | 0.970 |

**Table 6** Overall ancestry proportions inferred for individuals stratified in phenotypic groups

| Phenotypic category | Ancestry proportions (%) | | |
|---|---|---|---|
| | European | African | East Asian/ Native American |
| Blue eyes ($n=37$) | 88.0 | 7.9 | 4.1 |
| Intermediate eyes ($n=78$) | 83.9 | 11.4 | 4.7 |
| Brown eyes ($n=213$) | 62.6 | 26.9 | 10.5 |
| Red hair ($n=11$) | 85.9 | 9.8 | 4.3 |
| Blond hair ($n=49$) | 89.0 | 7.8 | 3.2 |
| Brown hair ($n=174$) | 75.4 | 16.3 | 8.4 |
| Black hair ($n=86$) | 48.4 | 39.4 | 12.2 |

database with more than 1500 global individuals for the four hair color categories (Supplementary Table 4). Although no statistical test was performed, it is noteworthy that blond and red hair present higher AUC, specificity, PPV, and NPV values in the present study. Red hair also presented higher sensitivity. Brown hair presented here higher sensitivity and PPV, while black hair also presents higher AUC and sensitivity, in spite of lower specificity and NPV. Despite these interesting findings, the overall system sensitivity for hair color prediction is much lower among Brazilian admixed samples (0.691) than in Europeans (0.790) [36]. We must emphasize that our sample size is much lower than the one used by HIrisPlex, with a larger proportion of intermediate phenotypes and a small number of samples for some phenotypic categories (such as red hair). Moreover, no admixed populations were used to estimate those parameters available in the HIrisPlex website. We can also observe that when individuals with blond and red hair are incorrectly predicted, the prediction result was always brown. Although it is surprising, this cannot be attributed to incorrect phenotype assessment by the observers, since the individuals have also self-reported as red hair. Notwithstanding that, most of red haired people were predicted correctly. Again, the low number of red hairs in our sample ($n=11$) may represent a limitation in the results analysis. Erroneous predictions involving individuals with brown hair results resulted in black hair and vice versa (Table 2). This may be because sometimes it is difficult to distinguish dark brown from black hair.

Although our total sample has a major European contribution, when we analyze only the subset of individuals with black hair, the European ancestry decreases to a much lower proportion. It is expected that samples with higher non-European proportions will present higher probabilities for dark hair, since light derived alleles from Europeans may be less common in such samples [36]. We can observe that the highest admixture rate is associated with the category with lower sensitivity. Notwithstanding, this is consistent

with the idea that admixture introduces African and East Asian/Native American alleles that are associated with darker phenotypes, increasing considerably black hair sensitivity when compared to the developer's values (0.547 vs. 0.333–Supplementary Table 4).

Regarding HIrisPlex application in forensics practice, we can observe that, except for red hair, all the categories presented low PPV values (Table 3), which means that the probability of a prediction outcome be correct is around 64–70%. However, some considerations may minimize the impact of these PPV values and optimize HIrisPlex's predictions in a police investigation. For instance, blond outcome is assigned to 13 individuals with brown hair (Table 2), and most (10) of them presented light brown phenotype. Therefore, when a blond prediction is obtained, it is important to expand the search to individuals with light brown hair too. Similar can be done for black hair. Twenty-five out of the 26 individuals wrongly assigned as black hair actually present brown hair, and specifically dark brown for most (24) of them. Thus, when a black outcome is obtained, the police search should include individuals with dark brown hair as well. Red hair is the most accurate color predicted, and only two blond individuals were incorrectly assigned with this phenotype. Finally, brown hair is the most complex category, because it was wrongly assigned to all other hair colors (Table 2). Thereby, if a brown prediction is made, it is important to keep in mind that the suspect could have hair color varying from dark blond to black.

Investigating the impact of age-dependent changes in hair color on prediction (Fig. 1), we notice that many individuals that were blond in childhood remained blond in adulthood, not influencing the prediction results. Regarding the individuals that had their hair changed from blond to brown/black color, 76% of them had their predictions compatible with their current hair color, and only 20% had blond hair color prediction. We can also observe that most of the individuals with brown hair in childhood still present it today. From the five individuals remaining (~ 30%) in which brown hair darkened to black, four of them were predicted as brown (80%), compatible with childhood and one of them was predicted as black hair (20%), compatible with adulthood. From these data, we can conclude that there is not an expressive influence of age-dependent changes on the results, because: (a) only a few individuals (31 out of 320, i.e., 9.69%) experienced hair color changes that were strong enough to lead them to another category, and (b) when a change to another category occurs, most of them (20 out of 31, i.e., 64.52%) had predictions compatible with what is observed during adulthood.

The HIrisPlex website also displays the AUC, sensitivity, specificity, PPV, and NPV values for the three eye color categories, based on 1000 cross-validation tests using a database with more than 9000 individuals from around the world

(Supplementary Table 4). Regarding blue eyes, except for NPV, the remaining parameters (AUC, sensitivity, specificity, and PPV) obtained in the Brazilian sample are lower than those presented on the website. This finding could be partly due to the limited number of blue eyes in our sample ($n = 37$). HIrisPlex did not predict any individual with intermediate eyes, leading to a sensitivity zero, specificity one, and the impossibility of estimating PPV. It has also led to a much lower NPV. This is not an unforeseen outcome, given the previous knowledge of the difficulties in predicting this phenotype. Finally, regarding brown eyes, it is interesting to highlight that in our study, the tool presented higher sensitivity, PPV, and NPV values, together with lower AUC and specificity.

An in-depth analysis of these parameters (Table 5) reveals that when the tool classifies an individual as having brown eyes, this prediction is reliable (PPV = 0.847). This parameter, as well as specificity, is not strongly affected by the fact that 33 individuals with intermediate eyes are wrongly classified with brown eyes (Table 4) since the number of sampled individuals with brown eyes is large ($n = 213$). On the other hand, HIrisPlex shows low reliability when predicting blue eyes (PPV = 0.405) (Table 5). This is because 45 out of the 78 individuals with intermediate eye color are predicted as having a blue eye (Table 4), and the number of sampled individuals with blue eyes is very small ($n = 37$) compared to other categories. As there was no intermediate eye prediction, all individuals with blue eyes wrongly classified by HIrisPlex are assigned to the brown eye category and vice versa (Table 4). Because of HIrisPlex's difficulty in predicting intermediate eyes, overall sensitivity for eye color prediction (0.741) is rather low.

Following the same reasoning presented for black hair, it is expected that samples with higher levels of non-European ancestry present higher probabilities for dark eye color, since they are more likely to have the majority of causal genotypes displaying a pair of ancestral alleles [36]. In fact, when we analyze the ancestry of the subset of individuals with brown eyes, they present lower European contribution and higher levels of African and Native American/East Asian ancestry. The sensitivity for brown eyes is improved by this admixed pattern, since the higher African and East Asian/Native American contributions within this group provide alleles associated with eumelanin production, boosting sensitivity for this category when compared to the observation among Europeans (0.991 vs. 0.935–Supplementary Table 4). When applying the 0.7 probability threshold, although there is a slight increase in values of most of the parameters (Table 5), there are no expressive changes in comparison with the developers' values. However, the sensitivities obtained for blue and brown eyes can be considered satisfactory. A 6.75% increase was observed for overall sensitivity (from 0.741

to 0.791). The use of this threshold seems to be a good choice since it increases blue eyes prediction sensitivity (5.1%), and only 12.5% of individuals are set as undefined (Table 5).

In our sample, there were 78 individuals with intermediate eyes: 18 of them presented hazel eyes, which is a little lighter than brown, while the remaining 60 presented green eyes, which can vary from light to dark green. Fifteen out of the 18 hazel eyes and only 18 out of 60 green eyes were predicted as brown. Most of these individuals had darker tones. Thinking about forensic practice, according to these results, we could suggest that, when a brown eye is predicted, authorities should keep in mind that it could be a hazel or a dark green eye as well. The same can be stated for blue eyes, since 42 out of 60 intermediate eyes were predicted as blue, and most of them presented light intermediate tones. Therefore, when blue eyes are predicted, it would be important to remind that this could indicate green eyes as well, and the search should include people in the blue light intermediate eyes spectrum. This approach could be used in order to minimize the impact of errors involving intermediate phenotypes, while markers that can identify them remain unknown. The use of the 0.7 thresholds also helps to decrease the impact of intermediate eyes failure, since almost 30% of them become undefined when applying it, leaving predictions of blue and brown eyes more reliable.

Some studies have evaluated the IrisPlex system in admixed and non-admixed populations [29–34], and the tool was not able to predict intermediate eyes for any individual in all of them, achieving a sensitivity zero and specificity one. These studies, as well as the present one, are in agreement with the statement described in the original IrisPlex studies [26, 27]: the HIrisPlex system can predict blue and brown eyes accurately, but it presents some difficulties in predicting intermediate eyes, pointing to the necessity of additional markers strongly associated with this phenotype. The evaluation of the IrisPlex system is especially important in admixed populations outside Europe because this genetic background leads to a higher presence of intermediate phenotypes.

As discussed above, HIrisPlex predicts blue and brown eyes competently and shows a reasonable prediction performance when it comes to hair color, with higher sensitivities for red and brown hair. However, its use in admixed populations must be addressed with caution. Such populations, due to their variable degrees of European and non-European genetic inheritance, tend to present more intermediate phenotypes than most autochthonous populations. Then, such predictive difficulty is especially problematic and has a stronger impact when used in admixed populations. Even though markers capable of helping in intermediate phenotypes predictions are still being studied, the enhancement of the HIrisPlex system with such markers would increase undoubtedly the reliability of prediction in Brazil and,

perhaps, in other South-American countries that share a similar ancestry background [61].

The European, African, and Amerindian biogeographical groups compose the ancestry background of other Latin American populations but in different proportions. Countries such as Mexico, Guatemala, Peru, and Ecuador display a more significant Amerindian contribution in their populations, while countries such as Cuba, Chile, Colombia, Puerto Rico, Venezuela, Argentina, and Uruguay have a major European influence. African ancestry is more common in the Caribbean area, including the Bahamas, Haiti, and Jamaica. Finally, Brazil presents a considerable variability among its regions: although European ancestry is predominant everywhere, we can observe a higher Amerindian contribution in the north of the country, while the northeast has a strong African contribution and the south and southeast (where the samples of this study were collected) have a higher European influence [34, 61–64]. Further studies in other Latin populations are required to evaluate HIrisPlex's performance.

The need of additional predictive markers is more pronounced in such admixed countries, which, in addition to presenting a higher frequency of intermediate phenotypes and a more diverse spectrum of pigmentation phenotypes, have particular demographic histories, which may result in a new genetic background characterized by the epistatic interaction of different ancestry-specific alleles. For example, two variants in the *MFSD12* gene that are strongly associated with skin color have been recently identified in studies involving African and Latin American admixed populations [13, 25, 65].

## Conclusion

In conclusion, the present study evaluated HIrisPlex performance in a different admixed background and reinforced its performance in predicting blue and brown eyes, despite its already known limitations in predicting intermediate eye color, which decreases the system sensitivity. Red hair was the most accurately predicted phenotype among hair colors. Despite the limitations for the other colors prediction, as age-dependent changes and probably the lack of additional predictive markers, the results obtained support its utility if addressed carefully. Finally, it is important to emphasize that there might be other factors influencing pigmentation, such as new combinations of alleles derived from different biogeographical ancestries, in case of admixed populations. Besides that, additional studies are important to identify new genetic markers that could be used to complement and enhance the HIrisPlex system predictions, providing more accurate outcomes. The improvement on prediction of intermediate eyes and dark blond/light brown hair phenotypes would be very important for admixed population, since they present high frequency of these phenotypes.

**Data availability** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Code availability** Not applicable.

## Declarations

**Ethics approval** This study was approved in its ethical aspects by the research ethics committee of this institution (Comitê de Ética em Pesquisa, FFCLRP-USP), according to protocol CAAE # 25,696,413.7.0000.5407.

**Consent to participate** Informed consent was obtained from all individual participants included in the study.

**Consent for publication** In the informed consent form signed by all the volunteers, there was a clause explaining that the research results would be published in scientific journals.

**Conflicts of interest** The authors declare no competing interests.

## References

1. Sturm RA (2009) Molecular genetics of human pigmentation diversity. Hum Mol Genet 18:R9–R17. https://doi.org/10.1093/hmg/ddp003
2. Liu F, Wen B, Kayser M (2013) Colorful DNA polymorphisms in humans. Semin Cell Dev Biol 24:562–575. https://doi.org/10.1016/j.semcdb.2013.03.013
3. Kayser M (2015) Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes.

Forensic Sci Int Genet 18:33–48. https://doi.org/10.1016/j.fsigen.2015.02.003

4. Parra EJ (2007) Human pigmentation variation: evolution, genetic basis, and implications for public health. Am J Phys Anthropol 134:85–105. https://doi.org/10.1002/ajpa.20727

5. Branicki W, Liu F, van Duijn K et al (2011) Model-based prediction of human hair color using DNA variants. Hum Genet 129:443–454. https://doi.org/10.1007/s00439-010-0939-8

6. Sulem P, Gudbjartsson DF, Stacey SN et al (2008) Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet 40:835–837. https://doi.org/10.1038/ng.160

7. Han J, Kraft P, Nan H et al (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4:e1000074. https://doi.org/10.1371/journal.pgen.1000074

8. Sulem P, Gudbjartsson DF, Stacey SN et al (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 39:1443–1452. https://doi.org/10.1038/ng.2007.13

9. Branicki W, Brudnik U, Draus-Barini J et al (2008) Association of the SLC45A2 gene with physiological human hair colour variation. J Hum Genet 53:966–971. https://doi.org/10.1007/s10038-008-0338-3

10. Donnelly MP, Paschou P, Grigorenko E et al (2012) A global view of the OCA2-HERC2 region and pigmentation. Hum Genet 131:683–696. https://doi.org/10.1007/s00439-011-1110-x

11. Kayser M, Liu F, Janssens ACJW et al (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet 82:411–423. https://doi.org/10.1016/j.ajhg.2007.10.003

12. Beleza S, Johnson NA, Candille SI et al (2013) Genetic architecture of skin and eye color in an African-European admixed population. PLoS Genet 9:e1003372. https://doi.org/10.1371/journal.pgen.1003372

13. Lona-Durazo F, Hernandez-Pacheco N, Fan S et al (2019) Meta-analysis of GWA studies provides new insights on the genetic architecture of skin pigmentation in recently admixed populations. BMC Genet 20:59. https://doi.org/10.1186/s12863-019-0765-5

14. Zhang M, Song F, Liang L et al (2013) Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. Hum Mol Genet 22:2948–2959. https://doi.org/10.1093/hmg/ddt142

15. Lamason RL (2005) SLC24A5, a putative cation exchanger, affects pigmentation in Zebrafish and Humans. Science (80- ) 310:1782–1786. https://doi.org/10.1126/science.1116238

16. Quillen EE, Bauchet M, Bigham AW et al (2012) OPRM1 and EGFR contribute to skin pigmentation differences between Indigenous Americans and Europeans. Hum Genet 131:1073–1080. https://doi.org/10.1007/s00439-011-1135-1

17. Hernandez-Pacheco N, Flores C, Alonso S et al (2017) Identification of a novel locus associated with skin colour in African-admixed populations. Sci Rep 7:44548. https://doi.org/10.1038/srep44548

18. Lloyd-Jones LR, Robinson MR, Moser G et al (2017) Inference on the genetic basis of eye and skin color in an admixed population via Bayesian linear mixed models. Genetics 206:1113–1126. https://doi.org/10.1534/genetics.116.193383

19. Norton HL, Edwards M, Krithika S et al (2016) Quantitative assessment of skin, hair, and iris variation in a diverse sample of individuals and associated genetic variation. Am J Phys Anthropol 160:570–581. https://doi.org/10.1002/ajpa.22861

20. Hohl DM, Bezus B, Ratowiecki J, Catanesi CI (2018) Genetic and phenotypic variability of iris color in Buenos Aires population. Genet Mol Biol 41:50–58. https://doi.org/10.1590/1678-4685-gmb-2017-0175

21. de Cerqueira CCS, Hünemeier T, Gomez-Valdés J et al (2014) Implications of the admixture process in skin color molecular

22. Andrade ES, Fracasso NCA, StrazzaJúnior PS et al (2017) Associations of OCA2 - HERC2 SNPs and haplotypes with human pigmentation characteristics in the Brazilian population. Leg Med 24:78–83. https://doi.org/10.1016/j.legalmed.2016.12.003

23. de AguiarFracasso NC, de Andrade ES, Wiezel CEV et al (2017) Haplotypes from the SLC45A2 gene are associated with the presence of freckles and eye, hair and skin pigmentation in Brazil. Leg Med 25:43–51. https://doi.org/10.1016/j.legalmed.2016.12.013

24. de Araújo LF, de Toledo GF, Fridman C (2015) SLC24A5 and ASIP as phenotypic predictors in Brazilian population for forensic purposes. Leg Med 17:261–266. https://doi.org/10.1016/j.legalmed.2015.03.001

25. Adhikari K, Mendoza-Revilla J, Sohail A et al (2019) A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. Nat Commun 10:358. https://doi.org/10.1038/s41467-018-08147-0

26. Walsh S, Lindenbergh A, Zuniga SB et al (2011) Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet 5:464–471. https://doi.org/10.1016/j.fsigen.2010.09.008

27. Walsh S, Liu F, Ballantyne KN et al (2011) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5:170–180. https://doi.org/10.1016/j.fsigen.2010.02.004

28. Liu F, van Duijn K, Vingerling JR et al (2009) Eye color and the prediction of complex phenotypes from genotypes. Curr Biol 19:R192–R193. https://doi.org/10.1016/j.cub.2009.01.027

29. Yun L, Gu Y, Rajeevan H, Kidd KK (2014) Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations. Int J Legal Med 128:447–453. https://doi.org/10.1007/s00414-013-0953-1

30. Dario P, Mouriño H, Oliveira AR et al (2015) Assessment of Iris-Plex-based multiplex for eye and skin color prediction with application to a Portuguese population. Int J Legal Med 129:1191–1200. https://doi.org/10.1007/s00414-015-1248-5

31. Salvoro C, Faccinetto C, Zucchelli L et al (2019) Performance of four models for eye color prediction in an Italian population sample. Forensic Sci Int Genet 40:192–200. https://doi.org/10.1016/j.fsigen.2019.03.008

32. Kastelic V, Pośpiech E, Draus-Barini J et al (2013) Prediction of eye color in the Slovenian population using the IrisPlex SNPs. Croat Med J 54:381–386. https://doi.org/10.3325/cmj.2013.54.381

33. Dembinski GM, Picard CJ (2014) Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. Forensic Sci Int Genet 9:111–117. https://doi.org/10.1016/j.fsigen.2013.12.003

34. Freire-Aradas A, Ruiz Y, Phillips C et al (2014) Exploring iris colour prediction and ancestry inference in admixed populations of South America. Forensic Sci Int Genet 13:3–9. https://doi.org/10.1016/j.fsigen.2014.06.007

35. Pneuman A, Budimlija ZM, Caragine T et al (2012) Verification of eye and skin color predictors in various populations. Leg Med 14:78–83. https://doi.org/10.1016/j.legalmed.2011.12.005

36. Walsh S, Chaitanya L, Clarisse L et al (2014) Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet 9:150–161. https://doi.org/10.1016/j.fsigen.2013.12.006

37. Walsh S, Liu F, Wollstein A et al (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet 7:98–115. https://doi.org/10.1016/j.fsigen.2012.07.005

38. Kukla-Bartoszek M, Pośpiech E, Spólnicka M et al (2018) Investigating the impact of age-depended hair colour darkening

during childhood on DNA-based hair colour prediction with the HIrisPlex system. Forensic Sci Int Genet 36:26–33. https://doi.org/10.1016/j.fsigen.2018.06.007

39. Walsh S, Chaitanya L, Breslin K et al (2017) Global skin colour prediction from DNA. Hum Genet 136:847–863. https://doi.org/10.1007/s00439-017-1808-5

40. Chaitanya L, Breslin K, Zuñiga S et al (2018) The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. Forensic Sci Int Genet 35:123–135. https://doi.org/10.1016/j.fsigen.2018.04.004

41. Hart KL, Kimura SL, Mushailov V et al (2013) Improved eye- and skin-color prediction based on 8 SNPs. Croat Med J 54:248–256. https://doi.org/10.3325/cmj.2013.54.248

42. Allwood JS, Harbison S (2013) SNP model development for the prediction of eye colour in New Zealand. Forensic Sci Int Genet 7:444–452. https://doi.org/10.1016/j.fsigen.2013.03.005

43. Ruiz Y, Phillips C, Gomez-Tato A et al (2013) Further development of forensic eye color predictive tests. Forensic Sci Int Genet 7:28–40. https://doi.org/10.1016/j.fsigen.2012.05.009

44. Söchtig J, Phillips C, Maroñas O et al (2015) Exploration of SNP variants affecting hair colour prediction in Europeans. Int J Legal Med 129:963–975. https://doi.org/10.1007/s00414-015-1226-y

45. Maroñas O, Phillips C, Söchtig J et al (2014) Development of a forensic skin colour predictive test. Forensic Sci Int Genet 13:34–44. https://doi.org/10.1016/j.fsigen.2014.06.017

46. Pena SDJ, Di Pietro G, Fuchshuber-Moraes M et al (2011) The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. PLoS One 6:e17063. https://doi.org/10.1371/journal.pone.0017063

47. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215–1215. https://doi.org/10.1093/nar/16.3.1215

48. Phillips C, Salas A, Sánchez JJ et al (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet 1:273–280. https://doi.org/10.1016/j.fsigen.2007.06.008

49. Fondevila M, Phillips C, Santos C et al (2013) Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies. Forensic Sci Int Genet 7:63–74. https://doi.org/10.1016/j.fsigen.2012.06.007

50. Ravi RK, Walton K, Khosroheidari M (2018) MiSeq: a next-generation sequencing platform for genomic analysis. Methods Mol Biol 1706:223–232. https://doi.org/10.1007/978-1-4939-7471-9_12

51. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10. https://doi.org/10.14806/ej.17.1.200

52. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698

53. McKenna A, Hanna M, Banks E et al (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110

54. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017

55. Consortium T 1000 GP (2015) A global reference for human genetic variation. Nature 526:68–74. https://doi.org/10.1038/nature15393

56. de Oliveira MLG, Veiga-Castelli LC, Marcorin L et al (2018) Extended HLA-G genetic diversity and ancestry composition in a Brazilian admixed population sample: implications for HLA-G transcriptional control and for case-control association studies. Hum Immunol 79:790–799. https://doi.org/10.1016/j.humimm.2018.08.005

57. Valle-Silva G, do, Souza FDN de, Marcorin L, et al (2019) Applicability of the SNPforID 52-plex panel for human identification and ancestry evaluation in a Brazilian population sample by next-generation sequencing. Forensic Sci Int Genet 40:201–209. https://doi.org/10.1016/j.fsigen.2019.03.003

58. Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8:103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

59. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

60. Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351. https://doi.org/10.1038/nrg.2016.49

61. Salzano FM, Sans M (2014) Interethnic admixture and the evolution of Latin American populations. Genet Mol Biol 37:151–170. https://doi.org/10.1590/S1415-47572014000200003

62. Ruiz-Linares A, Adhikari K, Acuña-Alonzo V et al (2014) Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. PLoS Genet 10:e1004572. https://doi.org/10.1371/journal.pgen.1004572

63. Rodrigues-Soares F, Peñas-Lledó EM, Tarazona-Santos E et al (2020) Genomic ancestry, CYP 2D6, CYP 2C9, and CYP 2C19 *Latin Americans. Clin Pharmacol Ther 107:257–268. https://doi.org/10.1002/cpt.1598

64. Norris ET, Wang L, Conley AB et al (2018) Genetic ancestry, admixture and health determinants in Latin America. BMC Genomics 19:861. https://doi.org/10.1186/s12864-018-5195-7

65. Crawford NG, Kelly DE, Hansen MEB et al (2017) Loci associated with skin pigmentation identified in African populations. Science (80- ) 358:eaan8433. https://doi.org/10.1126/science.aan8433