

## **Forensic DNA Phenotyping: starting point to prediction model in Pernambuco population, Brazil**

**Fenotipagem por DNA Forense: um ponto de partida para o modelo de predição na população de Pernambuco, Brasil**

**Fenotipado de DNA Forense: un punto de partida para el modelo de predicción de población de Pernambuco, Brasil**

Received: 09/22/2021 | Reviewed: 10/01/2021 | Accept: 10/09/2021 | Published: 10/11/2021

### **Juliana Maria de Souza**

ORCID: <https://orcid.org/0000-0002-4423-9350>  
University of Pernambuco, Brazil  
E-mail: [juliana.mariadesouza@hotmail.com](mailto:juliana.mariadesouza@hotmail.com)

### **Michael Lopes Bastos**

ORCID: <https://orcid.org/0000-0001-6581-2067>  
Institute SENAI of Innovation, Brazil  
E-mail: [michaellb17@gmail.com](mailto:michaellb17@gmail.com)

### **Bárbara de Oliveira Silva**

ORCID: <https://orcid.org/0000-0003-4158-5650>  
University of Pernambuco, Brazil  
E-mail: [barbaraoliveirabiomol@gmail.com](mailto:barbaraoliveirabiomol@gmail.com)

### **Karla Giselle Gomes de Lima**

ORCID: <https://orcid.org/0000-0002-4760-9054>  
University of Pernambuco, Brazil  
E-mail: [karllaag@hotmail.com](mailto:karllaag@hotmail.com)

### **Giwellington Silva Albuquerque**

ORCID: <https://orcid.org/0000-0001-6610-963X>  
University of Pernambuco, Brazil  
E-mail: [giwbiomedico@gmail.com](mailto:giwbiomedico@gmail.com)

### **Renata Santos de Oliveira**

ORCID: <https://orcid.org/0000-0002-1567-0897>  
University of Pernambuco, Brazil  
E-mail: [renata.so.oliveira@hotmail.com](mailto:renata.so.oliveira@hotmail.com)

### **Lúisa Priscilla Oliveira de Lima**

ORCID: <https://orcid.org/0000-0002-4119-0491>  
University of Pernambuco, Brazil  
E-mail: [luh.lima.116@gmail.com](mailto:luh.lima.116@gmail.com)

### **Edileine Dellalibera**

ORCID: <https://orcid.org/0000-0002-8943-183X>  
Oswaldo Cruz University Hospital, Brazil  
E-mail: [edileine.del@gmail.com](mailto:edileine.del@gmail.com)

### **Anthony José da Cunha Carneiro Lins**

ORCID: <https://orcid.org/0000-0002-7153-841X>  
Catholic University of Pernambuco, Brazil  
E-mail: [thonylins@gmail.com](mailto:thonylins@gmail.com)

### **Maria Tereza Cartaxo Muniz**

ORCID: <https://orcid.org/0000-0001-9498-5223>  
University of Pernambuco, Brazil  
E-mail: [tereza.cartaxo@upe.br](mailto:tereza.cartaxo@upe.br)

## **Abstract**

The study of Externally Visible Characteristics (EVC) of pigmentation associated with SNPs (Single Nucleotide Polymorphisms) has become a target in the forensic field due to the possibility of phenotypically characterizing an individual. In Brazil, there are few data that shows the evaluation of some these markers, so further studies are necessary to understand better the pigmentation process related to genetic markers. The aim of this study was to test the association between 8 SNPs present in HIrisplex tool and EVC to provide a starting point for the development of prediction models for heterogeneous populations like the one in Pernambuco. Were evaluated 176 individuals by associations between self-reported eye, hair and skin color data and polymorphisms. Artificial intelligence tools were used for the prediction models. Significant associations were found between rs1800404 (*OCA2*), rs6058017 (*ASIP*),

rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*) with (EVC). The prediction models evaluated showed satisfactory prediction rates, rates above 60% for skin color and above 70% for eyes and hair. The associations found in our data show the importance of SNPs evaluation used in DNA Phenotyping, because of its ability to provide new information in the context of criminal investigations. Our data indicate that is possible to use molecular information to predict phenotypes in miscigenated populations, like the Brazilian population. These polymorphisms could be possible phenotypic predictors for the Pernambuco population.

**Keywords:** Pigmentation genes; Miscigenated Population; Brazil; SNPs; Phenotypic prediction; Artificial intelligence.

### Resumo

O estudo das Características Externamente Visíveis (CEV) da pigmentação associada a SNPs (Single Nucleotide Polymorphisms) tornou-se um alvo na área forense devido à possibilidade de caracterizar fenotipicamente um indivíduo. No Brasil, poucos são os dados que mostram a avaliação de alguns desses marcadores, portanto, mais estudos são necessários para entender melhor o processo de pigmentação relacionado aos marcadores genéticos. O objetivo deste estudo foi testar a associação entre 8 SNPs presentes na ferramenta HIrisplex e as CEV para fornecer um ponto de partida para o desenvolvimento de modelos de predição para populações heterogêneas como a de Pernambuco. 176 indivíduos foram avaliados por meio de associações entre dados autorreferidos de cor dos olhos, cabelo e pele e os polimorfismos. Ferramentas de inteligência artificial foram utilizadas para os modelos de predição. Foram encontradas associações significativas entre rs1800404 (*OCA2*), rs6058017 (*ASIP*), rs16891982 (*SLC45A2*) e rs1426654 (*SLC24A5*) com as CEV. Os modelos de predição avaliados apresentaram índices de predição satisfatórios, acima de 60% para cor da pele e acima de 70% para olhos e cabelos. As associações encontradas em nossos dados mostram a importância da avaliação de SNPs utilizados na Fenotipagem por DNA, por sua capacidade de fornecer novas informações no contexto de investigações criminais. Os dados encontrados indicam que é possível usar informações moleculares para prever fenótipos em populações miscigenadas, como a brasileira. Esses polimorfismos podem ser possíveis preditores fenotípicos para a população de Pernambuco.

**Palavras-chave:** Genes de pigmentação; População miscigenada; Brasil; SNPs; Preditores fenotípicos; Inteligência artificial.

### Resumen

El estudio de las Características Externamente Visibles (CEV) de la pigmentación asociada a los SNP (Single Nucleotide Polymorphisms) se ha convertido en un objetivo en el área forense debido a la posibilidad de caracterizar fenotípicamente a un individuo. En Brasil, hay pocos datos que muestren la evaluación de algunos de estos marcadores, por lo que se necesitan más estudios para comprender mejor el proceso de pigmentación relacionado con los marcadores genéticos. El objetivo de este estudio fue probar la asociación entre 8 SNP presentes en la herramienta HIrisplex y el CEV para proporcionar un punto de partida para el desarrollo de modelos de predicción para poblaciones heterogéneas como la de Pernambuco. Se evaluaron 176 individuos a través de asociaciones entre los datos autoinformados sobre el color de ojos, cabello y piel y polimorfismos. Se utilizaron herramientas de inteligencia artificial para los modelos de predicción. Se encontraron asociaciones significativas entre rs1800404 (*OCA2*), rs6058017 (*ASIP*), rs16891982 (*SLC45A2*) y rs1426654 (*SLC24A5*) con CEV. Los modelos de predicción evaluados presentaron tasas de predicción satisfactorias, superiores al 60% para el color de piel y superiores al 70% para ojos y cabello. Las asociaciones encontradas en nuestros datos muestran la importancia de evaluar los SNP utilizados en el fenotipado de DNA, debido a su capacidad para aportar nueva información en el contexto de las investigaciones penales. Nuestros datos indican que es posible utilizar información molecular para predecir fenotipos en poblaciones mixtas, como la brasileña. Estos polimorfismos pueden ser posibles predictores fenotípicos para la población de Pernambuco.

**Palabras clave:** Genes de pigmentación; Población mixta; Brasil; SNP; Preditores fenotípicos; Inteligencia artificial.

## 1. Introduction

Forensic DNA Phenotyping (FDP) is the prediction of appearance traits from DNA present in biological samples (Queirós, 2019). The first studies with FDP began in the 2000s (Koops & Schellekens, 2006), This technology can be applied in criminal investigation scenarios, such as identifying missing persons, when DNA samples collected at crime scenes do not match the profiles stored in forensic DNA databases and there are no other leads for investigation or available eyewitnesses. This technique can be used also to identifying people in mass disasters (Queirós, 2019).

This tool has been used mainly in the prediction of Externally Visible Characteristics (EVC), such as skin, eyes and hair color (CEV) (Chaitanya et al., 2018) and more recently in predicting face shape (Qiao et al., 2018). It is believed that in the future it may also be used to predict other characteristics of biological sample donors such as age, baldness, fingerprints, as well as predisposition to certain diseases, facilitating the search for treatment (Scudder, McNevin, Kelty, Walsh, & Robertson, 2018). The FDP advantage over other commonly used methodologies, such as Short Tandem Repeat (STR) for example, is that there is no need for a reference sample for comparative analysis, but it is based on the association between genotypes and phenotypes, based on molecular markers already validated for the interest phenotype (Kayser, 2015).

FDP is made from determination of specific genetic polymorphisms, Single Nucleotide Polymorphisms (SNPs), which are alterations (deletions, insertions, substitutions) in a single nucleotide in DNA sequence, most often found in intergenic spaces, but can also be found in coding regions. These alterations are expressed in phenotype of each individual. Polymorphisms analysis in genes that encode eyes, skin and hair color are already a well accepted technique and used in several European countries, aiming at phenotypic prediction from biological samples. In Netherlands for example, it is already a reality, successfully used and regulated by law since 2003, in United Kingdom some phenotypic predictions are already used, based on existing laws. (Kayser & Schneider, 2009). Currently in Netherlands there is a tool called HirisPlex, which has three prediction models, 6 SNPs for predicting eye color, 22 SNPs for predicting hair color and 36 SNPs for predicting skin color, representing the first forensic tool validated for simultaneous prediction of eye, hair and skin color from DNA (Chaitanya et al., 2018).

Externally Visible Characteristics prediction, such as eye, hair and skin colors has been based on SNPs of the genes involved in melanogenesis, process by which melanin is produced. Melanin is the main protein involved in regulation of human pigmentation, more than 120 genes are involved somehow in this pathway (Branicki, Brudnik, Kupiec, Wolańska-Nowak, & Wojas-Pelc, 2007; Serre, Busutil, & Botto, 2018; Sturm & Duffy, 2012; Virmond, Robert, Brito, & Massuda, 2016), among them, *ASIP*, *SLC45A2*, *SLC24A5*, *MC1R*, *OCA2* and *TYR* are described in the literature for their fundamental role in melanogenesis (Bonilla et al., 2005; Graf, Voisey, Hughes, & van Daal, 2007; Lamason, 2005; Makova & Norton, 2005).

*ASIP* gene is located on chromosome 20, encoding the Agouti Signaling Protein, with 132 amino acids, it acts in pigmentation regulation due to its antagonistic role towards *MC1R*, inhibiting eumelanin synthesis (Bonilla et al., 2005; de Araújo Lima, de Toledo Gonçalves, & Fridman, 2015; Kanetsky et al., 2002). The polymorphism rs6058017 is found in 3'UTR region, described as a replacement of guanine by adenine (G> A) at position 8.818 in *ASIP* gene (G8818A). When the A allele is present, the activation of the melanogenesis process is blocked, resulting in pheomelanin synthesis. When homozygous genotype GG is present, it causes less expression of this gene and consequently less antagonism to *MC1R*, leading to eumelanine formation related to dark phenotypes (Bonilla et al., 2005; de Araújo Lima et al., 2015).

The *SLC45A2* gene (Solute carrier family 45, member 2), located on chromosome 5, encodes MATP protein (protein-associated transported protein) with 530 amino acids. This protein is within the melanosomal membrane, its function is linked to the traffic and tyrosinase processing to maintain the melanosome, as well as its pH. (Bin et al., 2015; Graf et al., 2007; J Park, AH Talukder, 1 SA Lim, 2 K Kim, 1 K Pan, 1 B Melendez, 1 SD Bradley, 1 KR Jackson, 1 JS Khalili, 1 J Wang, 1 C Creasy, 1 BF Pan, 3 SE Woodman, 1 C Bernatchez, 1 D Hawke, P Hwu, KM Lee, J Roszik, G Lizée, Aboud, Kang, Cutting, & Bennett, 2016). The polymorphism rs16891982, located in exon 5 in *SLC45A2*, is characterized by a cytosine-guanine exchange, resulting in a non-synonymous exchange of leucine for phenylalanine at amino acid position 374, this polymorphism is associated with normal variations of human pigmentation, as it regulates the MATP (Sawitzki et al., 2017; Soejima & Koda, 2006).

*SLC24A5* gene (Solute carrier family 24, member 5) is located on the long arm of chromosome 15, encoding the NCKX5 protein (sodium / calcium / potassium exchanger 5), described as a solute transportable protein involved in melanosome maturation, melanin produced type and control of melanosomal pH (Lamason, 2005). The SNP rs1426654 is located in coding region of the *SLC24A5* gene, characterized by a substitution of guanine for adenine in exon 3, this exchange results in amino acid change, a substitution of alanine by threonine, resulting in decreased of the ion exchange scale and pheomelanine synthesis (Cook et al., 2009; Jackson, 2006; Lamason, 2005).

*MC1R* (Melanocortin-1 Receptor) is the best characterized among the genes that regulate pigmentation in humans. It is located on chromosome 16, composed of a single exon. It encodes a receptor protein located on cell surface of melanocytes and it plays an important role in melanogenesis. (Makova & Norton, 2005). The polymorphism rs885479 is characterized by a substitution of guanine by adenine, causing an amino acid exchange of arginine for glycine, also known as variant R163Q (Fernandez et al., 2007).

The *OCA2* gene encodes a P protein, with 12 transmembrane domains. It is believed to be involved in anion transport, regulation of melanosomal pH and also involved in the processing and transport of internal proteins such as tyrosinase and tyrosinase-associated proteins. (Bellono, Escobar, Lefkovith, Marks, & Oancea, 2014). The rs1800407 polymorphism is located at exon 13 in *OCA2* gene at amino acid position 419, the allele A mutant encodes the amino acid glutamine instead of arginine, is mainly associated with iris phenotypic variation (Jannot et al., 2005; Rebbeck et al., 2002). Rs1800404 is located at exon 10 in *OCA2* gene and mutant allele A encodes a synonym substitution, this polymorphism is most commonly related to skin pigmentation (Adhikari et al., 2019; Crawford et al., 2017).

The *TYR* gene encodes tyrosinase, a key enzyme in the melanogenesis process, catalyzes the three most important steps in melanin production, the oxidation of dopamine tyrosine, dopaquinone dopamine, and 5,6-dihydroxyindyl to 5,6-indolequinone. Melanin production and yours yield are fully related to *TYR* activity and expression. (Feng et al., 2015). It is located on chromosome 11, contains 5 exons and covers about 65kb of genomic DNA encoding tyrosinase, containing 529 amino acids. Almost 200 mutations have been found in this gene, including mutations that are associated with albinism, as well as some polymorphic variations are also associated with changes in eye, hair and skin coloration. (Chaitanya et al., 2018; K & Purohit, 2013).

The allele A of rs1393350 in the *TYR* gene is associated with decreased tyrosinase activity, resulting in a slightly pigmented phenotype, sun-sensitive skin. It was also associated with susceptibility to blue eye color instead of green and blond hair color instead of brown (Jacobs et al., 2015; Sulem et al., 2007). The polymorphic allele A of rs1042602 (*TYR*) is associated with the absence of freckles and found at a frequency of 35% in European populations, ancestral allele C is found most often in East Asia (Sulem et al., 2007).

SNP analysis with association tests has been shown a tool for finding a relationship between genetic markers and phenotypic characteristics in a given population. (Virmond et al., 2016). Prediction models for CEV have been proposed for several populations, with the purpose of finding a model that has a better percentage of accuracy to predict the skin, eyes and hair color and even the three simultaneously (Chaitanya et al., 2018; Walsh et al., 2017). The use of Artificial Intelligence tools can bring a differential to these prediction models, making the comparison of algorithms, to evaluate which one presents a better performance in predicting CEV from the information of SNPs (Zaorska, Zawierucha, & Nowicki, 2019). Therefore, the aim of this study was to test the association between 8 SNPs present in HIrisplex tool and Externally Visible Characteristics (EVC) to provide a starting point for the development of prediction models for heterogeneous populations like the one in Pernambuco.

## 2. Methodology

### 2.1 Study population

It was performed an analytical cross-sectional study. It was evaluated 176 samples from healthy individuals and volunteers, including teachers, students and collaborators from the University of Pernambuco, Brazil, as well as volunteers from neighboring neighborhoods. Survey participants were females, males and both older than 18 years. All participants received information about the research and signed the Informed Consent Form, allowing the use of their biological material and collected information.

### 2.2 Data collect

The volunteers answered a questionnaire where they self-reported eye, skin and hair color. For hair classification, the participants were asked about the original hair color at 15 years of age. In order to obtain a more accurate pattern and result, the volunteers were presented with a grid of possible hair colors including red, blond, brown and black, so that among the options, the individual pointed to the closest color they had at the age reference (de Araújo Lima et al., 2015). For the eyes color report was presented a table with different colors that are grouped in three shades: blue, green and brown. (Walsh et al., 2011). For skin color information, the participants self-reported within four categories: White, brown, yellow and black. In order to avoid bias in the self-reported information, the same interviewers were responsible for collecting all the data, which were readjusted when a large discrepancy was noted using the observation of two other researchers. This information was used to make associations with genotypes (de Araújo Lima et al., 2015).

### 2.3 DNA isolation, amplification and genotyping

It was collected 4mL of peripheral blood, by venipuncture, in tubes treated with EDTA or it was collected buccal mucosa cells with sterile cytological brushes, performing circular movements on the inside of the cheeks to capture the cells. The brush was packed in 2mL tubes with 1mL of absolute ethanol. DNA extraction was performed according to the Mini Salting-out method (Whikehart, 2003). DNA samples were amplified by PCR (*Polymerase Chain Reaction*) reaction.

Genotyping was performed by PCR-RFLP analysis using the following primers and restriction enzymes, respectively, foward 5'- GTGCTCAGCCTCAACTGCT -3'; reverse 5'- CACACCTTCGAGCAGCCTA -3'; *MbiI* for rs6058017 (*ASIP*) (Kanetsky et al., 2002), foward 5'- GGAAAACACGGAGTTGCTGCA -3', reverse 5'- CTTTCAGATTGTGTACCGCGGG -3'; *PstI* for rs16891982 (*SLC45A2*) (Yun, Gu, Rajeevan, & Kidd, 2014); foward 5'- GATTGTCTCAGGATGTTGCAGG -3'; reverse 5'- GATCCAAGGATGGTGCTAATGC -3'; *HhaI* for rs1426654 (*SLC24A5*) (Valls, 2012), foward 5'- TGATCACCTGCAGCTCCATG -3' e reverse 5'- CACCCAGATGGCCGCAACGGCT -3'; *HindIII* for rs885479 (*MC1R*) (Dbniak et al., 2006), foward 5'-GTGTGTGCTAACTCCCTCGT-3 ' ; reverse 5'-AGCGAAAGCCTGAATCCTGG-3'; *HaeIII* for rs1800404 (*OCA2*)(Andrade, Fracasso, Strazza Júnior, Simões, & Mendes-Junior, 2017), foward 5'- AAGGCTGCCTCTGTTCTACG-3', reverse 5'- CGATGAGACAGAGCATGATGA-3'; *SmaI* for rs1800407 (*OCA2*) (Gu, Yun, Zhang, Yang, & Hou, 2011); foward 5'- CAACACCCATGTTTAACGACA-3'; reverse 5'- GCTTCATGGGCAAATCAAT-3'; *MboI* for rs1042602 (*TYR*) (Giebel & Spritz, 1990; Walsh et al., 2013) and foward 5'- TCATTTGTAAGACCACACAGAATT-3' and reverse 5'-GCTTGGATTTCCAGACATCATC -3'; *EcoRI* for rs1393350 (*TYR*) (Yun et al., 2014).The digestion product was analyzed by 3% agarose gel electrophoresis.

### 2.4 Statistical analysis

Hardy-Weinberg Equilibrium (EHW) was used to estimate the difference between the observed and expected frequencies by OEGE (Online Encyclopedia for Genetic Epidemiology studies) (“OEGE - Online Encyclopedia for Genetic

Epidemiology studies,” n.d.). The association between genotypes and phenotypic characteristics were made by category and grouped. Fisher's exact test and odds ratio (OR) with 95% confidence intervals (CI) were calculated according to Lima et al. (de Araújo Lima et al., 2015). The associations were considered significant if  $p < 0.05$  and  $OR > 1$  according to Andrade et al. (Andrade et al., 2017). Williams' G and Chi-square tests were also calculated for analysis between groups. All tests are performed using the GraphPad 5.0 software (*GraphPad Prism version 5.04*, n.d.) and Bioestat 5.0 (AYRES, Manuel; Ayres Jr, Manuel; Ayres, Daniel Lima; Santos, 2007).

## 2.5 Prediction Model

Prediction modeling was performed on 137 individuals using the principal Component Analysis (PCA), followed to unsupervised learning using K-means method. After it was used the supervised learning comparing six different algorithms: Random forest, KNeighborsClassifier, DecisionTreeClassifier, RadiusNeighborsClassifier, ExtraTreesClassifier, Support Vector Classifier. For which one of these algorithms it was observed a global accuracy score and eye, hair and skin accuracy score.

## 2.6 Ethical considerations

The study was submitted to Plataforma Brasil for consideration by the Ethics Committee of the Oswaldo Cruz University Hospital (HUOC) that approved the research protocol, with CAAE number: 69723017.8.0000.5192.

## 3. Results

We analyzed 176 individuals with mean age of 39.9 years, 55.1% of the sample were female and 44.9% male. About the self-reported information of eye color, 68.7% said they had dark brown eyes, 22.7% light brown, 7.4% green and 1.1% blue. For hair color, 63.1% said they had black hair, 25.6% brown and 11.3% blond and for skin color 50.6% said they were black, 27.8% brown, 17% white and 4,5% yellow. For genotype analyzes 172 individuals were evaluated for rs6058017 (*ASIP*), 154 for rs16891982 (*SLC45A2*), 176 for rs1426654 (*SLC24A5*) and 170 for rs885479 (*MC1R*).

### 3.1 rs6058017 (*ASIP*)

No deviations were observed in Hardy-Weinberg equilibrium for rs6058017 (*ASIP*) polymorphism. The allele A was found at a frequency of 76.7% and allele G, 23.3%. The most frequent genotype was AA, followed by GA and GG, 59.9%, 33.7% and 6.4% respectively. No significant associations were observed between rs6058017 and eye and hair color in this population (Tables 1A, 2A, 4A and 5A), but some significant associations were found with skin color. Significant associations were observed between the genotypes GA, AA and combined analysis of GA + AA with brown skin ( $p = 0.0007$ ; OR: 29.17), ( $p < 0.0001$ ; OR: 45.5), ( $p < 0.0001$ ; OR: 37.33), respectively (Table 3A). The genotype AA and joint analysis of GA + AA were associated with white skin ( $p = 0.009$ ; OR: 18.67), ( $p = 0.013$ ; OR: 13.42), also respectively (Table 3A). In grouped analysis the genotypes GA ( $p = 0.002$  OR: 9.52), AA ( $p = 0.0006$ ; OR: 11.13) and their sum ( $p = 0.0005$ ; OR: 10.50) were associated with the non-black group (Table 6A).

**Table 1A.** Analysis of SNPs in relation to the color of eyes.

Gene (Polymorphism)	Genotype	Dark Brown* n (%)	Light Brown n (%)	OR (CI)	p*	Green n (%)	OR (CI)	p*	Blue n (%)	OR (CI)	p*	
<b>ASIP (rs6058017)</b>	GG	10(5,8)	1(0,6)	1,0		0(0)	1,0		0(0)	1,0		
	GA	35(20,3)	16(9,3)	4,57 (0,53-38,84)	0,26	7(4,1)	4,43 (0,23 - 84,35)	0,32	0(0)			
	AA	73(42,4)	23(13,4)	3,15 (0,38 - 25,96)	0,44	5(2,9)	1,57 (0,08 - 30,54)	1,0	2(1,2)	0,71 (0,03 - 15,94)	1,0	
	n:172	GA+AA	108(62,8)	39(22,7)	3,61 (0,44 - 29,15)	0,29	12(7,0)	2,41 (0,13 -43,86)	0,59	2(1,2)	0,49 (0,02 - 10,7)	1,0
	p**				0,45				0,22			0,75
Trend Test***				0,49				0,19			0,72	
<b>SLC45A2 (rs16891982)</b>	CC	36(23,4)	10(6,5)	1,0		2(1,3)	1,0		0(0)	1,0		
	CG	59(38,3)	20(13,0)	1,22(0,51-2,89)	0,82	9(5,9)	2,74(0,56-13,43)	0,32	1(0,6)	1,84(0,07-46,42)	1,0	
	GG	13(8,4)	3(1,9)	0,83(0,19-3,49)	1,0	1(0,6)	1,38(0,11-16,59)	1,0	0(0)			
	n:154	CG+GG	72(46,7)	23(14,9)	1,15(0,49-2,67)	0,83	10(6,5)	2,5(0,51-12,02)	0,33	1(0,6)	1,51(0,05-38,03)	1,0
	p**				0,93				0,57			0,85
Trend Test***				0,93				0,55			0,85	
<b>SLC24A5 (rs1426654)</b>	GG	6(3,4)	0(0)	1,0		1(0,5)	1,0		0(0)	1,0		
	GA	71(40,3)	10(5,7)	1,90 (0,10 - 36,45)	1,00	5(2,8)	0,42 (0,04 - 4,23)	0,42	0(0)			
	AA	44(25)	30(17,0)	8,91 (0,48 - 164,2)	0,07	7(4,0)	0,95 (0,09 - 9,17)	1,0	2(1,1)	0,73 (0,03 - 16,9)	1,0	
	n:176	GA + AA	115(65,3)	40(22,7)	4,55 (0,25 - 82,79)	0,33	12(6,8)	0,62 (0,06 - 5,64)	0,51	2(1,1)	0,28 (0,01 - 6,49)	1,0
	p**				0,0004				0,57			0,34
Trend Test***				0,0002				0,62			0,5	
<b>MC1R (rs884579)</b>	GG	101(59,4)	35(20,6)	1,0		12(7,0)	1,0		2(1,2)	1,0		
	GA	17(10,0)	3(1,70)	0,50 (0,14 - 1,84)	0,40	0(0)	0,23 (0,01 - 4,10)	0,36	0(0)	1,16 (0,05 - 25,2)	1,0	
	AA	0(0)	0(0)			0(0)			0(0)			
	n:170											
	p**				0,29				0,15			0,56
Trend Test***				0,42				0,27			0,68	

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*; Williams's G Test (Trend Test).  
Source: Authors.

**Table 2A.** Analysis of SNPs in relation to hair color.

Gene (Polymorphism)	Genotype	Black* n (%)	Brown n (%)	OR (CI)	p*	Blond n (%)	OR (CI)	p*
<b>ASIP (rs6058017)</b> n:172 p** Trend Test***	GG	3(1,7)	7(4,1)	1,0		1(0,6)	1,0	
	GA	11(6,4)	37(21,5)	1,44 (0,31 - 6,53)	0,69	10(5,8)	2,72 (0,24 - 30,68)	0,60
	AA	29(16,9)	66(38,4)	0,97 (0,23 - 4,04)	1,0	8(4,6)	0,81 (0,07 - 9,08)	1,0
	GA+AA	39(22,7)	103(59,9)	1,13 (0,27 - 4,60)	1,0	19(11,0)	1,46 (0,14 - 15,01)	1,0
					0,81			0,23
				0,81				0,26
<b>SLC45A2 (rs16891982)</b> n:154 p** Trend Test***	CC	12(7,8)	31(20,1)	1,0		5(3,2)	1,0	
	CG	22(14,3)	59(38,3)	1,03(0,45-2,37)	1,0	8(5,2)	0,87(0,23-3,26)	1,0
	GG	1(0,6)	12(7,8)	4,64(0,54-39,74)	0,25	4(2,6)	9,6(0,84-108,8)	0,11
	CG+GG	23(14,9)	71(46,1)	1,19(0,52-2,70)	0,67	12(7,8)	1,25(0,35 -4,39)	1,0
					0,47			0,13
				0,40			0,16	
<b>SLC24A5 (rs1426654)</b> n:176 p** Trend Test***	GG	2(1,1)	5(2,8)	1,0		0(0)	1,0	
	GA	26(14,8)	52(29,5)	0,80 (0,14 - 4,40)	1,0	8(4,5)	1,60 (0,06 - 36,8)	1,0
	AA	17(9,6)	54(30,7)	1,27 (0,22 - 7,15)	1,0	12(6,8)	3,57 (0,15 0 81,10)	0,50
	GA + AA	43(24,4)	106(60,2)	0,98 (0,18 - 5,28)	1,0	20(11,3)	2,35 (0,10 - 51,38)	1,0
					0,66			0,35
				0,67			0,33	
<b>MC1R (rs884579)</b> n:170 p** Trend Test***	GG	38(22,3)	96(56,5)	1,0		16(9,4)	1,0	
	GA	6(3,5)	12(7,0)	0,79 (0,27 - 2,26)	0,78	2(1,2)	0,79 (0,14 - 4,35)	1,0
	AA	0(0)	0(0)			0(0)		
					0,66			0,78
					0,87			0,88

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*: Williams's G Test (Trend Test).

Source: Authors



**Table 3A.** Analysis of SNPs in relation to skin color.

Gene (Polymorphism)	Genotype	Black* n (%)	Brown n (%)	OR (CI)	p*	White n (%)	OR (CI)	p*	Yellow n (%)	OR (IC)	p*
<b>ASIP (rs6058017)</b> n:172 p** Trend Test***	GG	7(4,1)	1(0,6)	1,0		1(0,6)	1,0		0(0)	1,0	
	GA	6(3,5)	25(14,5)	29,17 (2,99 - 284,4)	<b>0,0007</b>	7(4,1)	8,16 (0,76 - 86,72)	0,08	2(1,2)	5,76 (0,23 - 143,5)	0,46
	AA	6(3,5)	39(22,7)	45,50 (4,72 - 438,3)	<b>&lt;0,0001</b>	16(9,3)	18,67 (1,87 - 185,5)	<b>0,009</b>	6(3,5)	15,0 (0,70 - 320,9)	0,04
	GA+AA	12(7,00)	64(37,2)	37,33 (4,20 - 331,8)	<b>&lt;0,0001</b>	23(13,4)	13,42 (1,47 - 122,2)	<b>0,013</b>	8(4,7)	10,20 (0,51 - 203,5)	0,06
					<b>&lt;0,0001</b>			<b>0,02</b>			<b>0,13</b>
				<b>0,0004</b>			<b>0,02</b>			<b>0,06</b>	
<b>SLC45A2 (rs16891982)</b> n:154 p** Trend Test***	CC	12(7,8)	31(20,1)	1,0		4(2,6)	1,0		1(0,6)	1,0	
	CG	14(9,1)	44(28,6)	1,21(0,49-2,98)	0,81	24(15,6)	5,14(1,38-19,06)	<b>0,01</b>	7(4,5)	6,0 (0,64 -55,98)	0,11
	GG	1(0,6)	6(3,9)	2,32(0,25-21,38)	0,65	10(6,5)	30,0 (2,86-313,7)	<b>0,001</b>	0(0)	2,77(0,07-103,9)	1,0
	CG+GG	15(9,75)	50(32,5)	1,29(0,53-3,11)	0,65	34(22,1)	6,8 (1,88 -24,57)	<b>0,002</b>	7(4,5)	5,6(0,60-52,03)	0,21
					0,86			<b>0,002</b>			0,31
				0,86			<b>0,002</b>			0,32	
<b>SLC24A5 (rs1426654)</b> n:176 p** Trend Test***	GG	4(2,3)	1(0,6)	1,0		1(0,6)	1,0		1(0,6)	1,0	
	GA	23(13,1)	43(24,4)	7,47 (0,78 - 70,92)	0,06	18(10,2)	3,13 (0,32 - 30,51)	0,38	2(1,1)	0,34 (0,02 0 4,80)	0,43
	AA	3(1,7)	45(25,6)	60,0 (5,0 - 719,4)	<b>0,0006</b>	30(17,0)	40,0 (3,30 - 485,7)	<b>0,002</b>	5(2,8)	6,66 (0,48 - 91,39)	0,26
	GA + AA	26(14,8)	88(50,0)	13,54 (1,44 - 126,6)	<b>0,01</b>	48(27,2)	7,38 (0,78 - 69,59)	0,06	7(3,9)	1,07 (0,10 - 11,24)	1,0
					<b>0,0001</b>			<b>&lt;0,0001</b>			0,01
				<b>0,0001</b>			<b>&lt;0,0001</b>			0,03	
<b>MC1R (rs884579)</b> n:170 p** Trend Test***	GG	27(15,9)	80(47,0)	1,0		38(22,3)	1,0		5(2,9)	1,0	
	GA	3(1,8)	7(4,1)	0,78 (0,19 - 3,26)	0,71	7(4,1)	1,65 (0,39 - 6,99)	0,73	3(1,8)	5,40 (0,83 - 34,82)	0,09
	AA	0(0)	0(0)			0(0)			0(0)		
											0,058
						0,74			0,48		0,058
					0,96			0,72		0,20	

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*, Williams's G Test (Trend Test).  
Source: Authors.

**Table 4A.** Analysis of SNPs in relation to grouped eye color.

Gene (Polymorphism)	Genotype	Brown* n (%)	Non- Brown n (%)	OR (IC)	p*	
<b>ASIP (rs6058017)</b> n:172 p** Trend Test***	GG	11(6,4)	0(0)	1,0		
	GA	51(29,5)	7(4,0)	3,35 (0,17 - 62,98)	0,58	
	AA	96(55,8)	7(4,0)	1.788 (0,095 - 33,41)	1,0	
	GA+AA	147(85,4)	14(8,1)	2,26 (0,12 - 40,40)	0,60	
						0,49
					0,39	
<b>SLC45A2 (rs16891982)</b> n:154 p** Trend Test***	CC	46(29,9)	2(1,3)	1,0		
	CG	79(51,3)	10(6,5)	2,91 (0,61 - 13,88)	0,21	
	GG	16(10,4)	1(0,6)	1,43 (0,12 - 16,95)	1,0	
	CG+GG	95(61,7)	11(7,1)	2,66 (0,56 - 12,52)	0,34	
						0,51
					0,47	
<b>SLC24A5 (rs1426654)</b> n:176 p** Trend Test***	GG	6(3,4)	1(0,5)	1,0		
	GA	81(46,0)	5(2,8)	0,37 (0,037 - 3,7)	0,38	
	AA	74(42,0)	9(5,1)	0,72 (0,07 - 6,77)	0,57	
	GA + AA	155(88,0)	14(7,9)	0,54 (0,06 - 4,8)	0,47	
						0,63
					0,67	
<b>MC1R (rs884579)</b> n:170 p** Trend Test***	GG	136(80)	14(8,2)	1,0		
	GA	20(11,7)	0(0)	0,93 (0,19 - 4,38)	1,0	
	AA	0(0)	0(0)			
						0,92
						0,77

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*; Williams's G Test (Trend Test).  
Source: Authors.

**Table 5A.** Analysis of SNPs in relation to grouped hair color.

Gene (Polimorphism)	Genotype	Non -Blond* n (%)	Blond n (%)	OR (IC)	p*	
<b>ASIP (rs6058017)</b> n:172 p** Trend Test***	GG	1(0,6)	10(5,8)	1,0		
	GA	10(5,8)	48(27,9)	0,48 (0,05 - 4,18)	0,67	
	AA	8(4,6)	95(55,2)	1,18 (0,13 - 10,50)	1,0	
	GA+AA	18(10,4)	143(83,1)	0,79 (0,09 - 6,57)	1,0	
						0,33
					0,38	
<b>SLC45A2 (rs16891982)</b> n:154 p** Trend Test***	CC	43(31,1)	5(3,2)	1,0		
	CG	81(52,6)	8(5,2)	0,84 (0,26 - 2,75)	0,76	
	GG	13(8,4)	4(2,6)	2,64 (0,61 - 11,33)	0,22	
	CG+GG	94(61,0)	12(7,8)	1,09 (0,36 - 3,31)	1,0	
						0,37
					0,48	
<b>SLC24A5 (rs1426654)</b> n:176 p** Trend Test***	GG	7(3,4)	0(0)	1,0		
	GA	78(44,3)	8(4,5)	1,62 (0,08 - 31,02)	1,0	
	AA	71(40,3)	12(6,8)	2,62 (0,14 - 48,91)	0,58	
	GA + AA	149(84,6)	20(11,3)	2,05 (0,11 - 37,40)	1,0	
						0,5
					0,45	
<b>MC1R (rs884579)</b> n:170 p** Trend Test***	GG	134(78,8)	16(9,4)	1,0		
	GA	18(10,6)	2(1,2)	0,93 (0,19 - 4,38)	1,0	
	AA	0(0)	0(0)			
						0,92
						0,77

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*; Williams's G Test (Trend Test).  
Source: Authors.

**Table 6A.** Analysis of SNPs in relation to grouped skin color.

Gene (Polimorphism)	Genotypes	Black* n (%)	Non-Black n (%)	OR (IC)	p*
<b>ASIP (rs6058017)</b> n:172 p** Trend Test***	GG	7(4,1)	4(2,3)	1,0	
	GA	9(5,2)	49(28,5)	9,52 (2,30 - 39,40)	<b>0,0020</b>
	AA	14(8,1)	89(51,7)	11,13 ( 2,87 - 42,99)	<b>0,0006</b>
	GA+AA	23(13,3)	138(80,2)	10,50 (2,84 - 38,75)	<b>0,0005</b>
					<b>&lt;0,001</b>
					<b>0,0048</b>
<b>SLC45A2 (rs16891982)</b> n:154 p** Trend Test***	CC	12(7,8)	36(23,4)	1,0	
	CG	14(9,1)	75(48,7)	1,7 (0,74 - 4,25)	0,25
	GG	1(0,65)	16(10,4)	5,3 (0,63 - 44,6)	0,15
	CG+GG	15(9,74)	91(59,1)	2,0 (0,86 - 4,74)	0,11
					0,21
					0,22
<b>SLC24A5 (rs1426654)</b> n:176 p** Trend Test***	GG	4(22,7)	3(1,7)	1,0	
	GA	23(13,1)	63(35,8)	3,65 (0,75 - 17,58)	0,18
	AA	3(1,7)	80(45,4)	35,56 (5,37 - 235,3)	<b>0,0004</b>
	GA + AA	26(14,8)	143(81,2)	7,33 (1,54 - 34,71)	<b>0,016</b>
					<b>&lt;0,0001</b>
					<b>&lt;0,0001</b>
<b>MC1R (rs884579)</b> n:170 p** Trend Test***	GG	27(15,9)	123(72,3)	1,0	
	GA	3(1,8)	17(10,0)	1,24 (0,34 - 4,54)	1,0
	AA	0(0)	0(0)		
					0,74
					0,98

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*; Williams's G Test (Trend Test).  
Source: Authors.

### 3.2 rs16891982 (*SLC45A2*)

Ancestral allele C was found more frequently, 60.1%, followed by mutant G, 39.9%. Heterozygotes frequency was higher in this population, 57.8%, followed by genotypes CC and GG, 31.1% and 11%, respectively. Were found deviations in Hardy-Weinberg equilibrium for this polymorphism ( $X^2$  6,45;  $p = 0,01$ ). No significant associations were observed between rs16891982 and eye and hair color (Tables 1A, 2A, 4A e 5A). However, significant associations were found between genotypes CG, GG and combined analysis of CG+GG with white skin color, ( $p = 0,01$  OR: 5.14), ( $p = 0,001$  OR: 30.0), ( $p = 0,002$  OR: 6.8), respectively (Table 3A). In grouped analyzes no significant associations were found with skin color (Table 6A).

### 3.3 rs1426654 (*SLC24A5*)

Were found deviations in Hardy-Weinberg equilibrium for rs1426654 ( $X^2$  7,12;  $p = 0,007$ ). The polymorphic allele A was found in high frequency 71,6% and the ancestral G, 28,4%. It was observed a higher heterozygotes frequency, 48.9%, followed by AA and GG, 47.1% and 4%, respectively. Significant associations were found between this polymorphism and skin color. The genotype AA ( $p = 0,0006$  OR: 60.0) was strongly associated with brown skin. Joint analysis of GA + AA also showed an association with brown skin ( $p = 0,01$  OR: 13.54). Strong associations were also found between genotype AA and white skin

( $p = 0.002$  OR: 40.0) (Table 3A). In grouped analysis, the homozygous mutant AA was strongly associated with non-black skin color ( $p = 0.0004$  OR: 35.56), Joint analysis of GA + AA also showed an association with non-black skin ( $p = 0.01$  OR: 7.33) (Table 6A). No significant associations were found between this polymorphism and eye and hair color (Tables 1A, 2A, 4A and 5A).

### 3.4 rs885479 (*MC1R*)

It was observed a high frequency of the ancestral allele G, 94.1%, while the frequency of the polymorphic allele A was 5.9%. No individuals were found with homozygous genotype AA. The homozygous genotype GG was the most frequent, 88.2%, followed by heterozygous genotype GA, 11.8%, although no deviations were observed in Hardy-Weinberg equilibrium. No significant association was found between rs885479 and eye, hair and skin color (Tables 1A – 6A).

### 3.5 *OCA2* gene

It was observed a higher frequency of mutant allele A in rs1800404, 56.3% in relation to ancestral allele G, 43.6% of the *OCA2* polymorphism. It was observed higher frequency of heterozygous genotype GA, followed by AA and GG, respectively, 45.7%, 33.5% and 20.8%. No deviations in Hardy-Weinberg equilibrium were observed for this polymorphism. Associations were found between allele A in homozygous and green eyes ( $p = 0.0084$ ; OR: 17.39), when analyzed together with the heterozygous genotype frequency, a significant association was also found. ( $p = 0.03$ ; OR: 10.03) (Table 1B). In the grouped analysis the genotype AA was associated with non-brown eye color group, as well as when it was added with heterozygous genotype GA, ( $p = 0.01$ ; OR: 15.80) and ( $p = 0.04$ ; OR: 9.23), respectively (Table 4B).

Significant associations were found between rs1800404 and the hair color. When compared to the black hair group, the GA ( $p = 0.01$ ; OR: 15.0) and AA genotypes ( $p = 0.01$ ; OR: 15.32) were associated with blond hair. The sum of GA and AA genotypes also reinforces the association ( $p = 0.01$ ; OR: 15.0). (Table 2B) In the grouped analysis the heterozygous genotype GA ( $p = 0.02$ ; OR: 11.03), the homozygous mutant AA ( $p = 0.01$ ; OR: 14.01) and their sum ( $p = 0.01$ ; OR: 12.01) were also associated with blond hair color when compared to the non-blond group (Table 5B).

The analysis of GA, AA and GA + AA genotypes show the association of rs1800404 with brown skin color ( $p = 0.0002$ ; OR: 6.8), ( $p = 0.003$ ; OR: 5.89), ( $p < 0.0001$ ; OR: 6.45), respectively and a strong association with white skin color ( $p < 0.0001$ ; OR: 21.25), ( $p < 0.0001$ ; OR: 40.80), ( $p < 0.0001$ ; OR: 28.77), also respectively, when compared to the black skin color group (Table 3B). In the grouped analysis the genotypes GA ( $p < 0.0001$ ; OR: 7.94), AA ( $p < 0.0001$ ; OR: 9.48) were correlated with the non-black skin color group, when compared to the black color group, the sum of GA and AA genotypes reinforces the association with the non-black group ( $p < 0.0001$ ; OR: 8; 53) (Table 6B).

It was observed a high frequency of the ancestral G allele of rs1800407 in this population, 93.2%, while the mutant allele A frequency was 6.8%. No individuals with the polymorphic allele in homozygous AA were found, while the ancestral allele in homozygous GG was the most frequent 86.4%, followed by the heterozygous genotype GA, 13.6%, although no deviations were observed in Hardy-Weinberg equilibrium. No significant associations were found between rs1800407 and eye, hair and skin color (Tables 1B – 6B).

**Table 1B.** Analysis of SNPs in relation to the color of eyes.

Gene (Polymorphism)	Genotype	Dark Brown* n (%)	Light Brown n (%)	OR (CI)	p*	Green n (%)	OR (CI)	p*	Blue n (%)	OR (CI)	p*
<b>OCA2 (rs1800404)</b> n:173 p** Trend Test***	GG	32(18,5)	4(2,3)	1,0		0(0)	1,0		0(0)	1,0	
	GA	52(30,0)	22 (12,7)	3,38 (1,068 - 10,72)	0,03	4(2,3)	5,57 (0,29 - 107,0)	0,29	1(0,6)	1,85 (0,07-47,0)	1,00
	AA	35 (20,2)	13 (7,5)	2,97 (0,87 - 10,06)	0,1	9(5,2)	17,39 (0,97 - 311,1)	<b>0,0084</b>	1(0,6)	2,74 (0,10 - 69,8)	1,00
	GA+AA	87(58,3)	35(20,2)	3,21 (1,05 - 9,77)	0,04	13(7,5)	10,03 (0,57 - 173,7)	<b>0,03</b>	2(1,1)	1,85 (0,08 - 39,7)	1,00
						0,16			<b>0,028</b>		
					0,12			<b>0,0087</b>			0,76
<b>OCA2 (rs1800407)</b> n:155 p** Trend Test***	GG	90(58,6)	32(20,6)	1,0		10(6,4)	1,0		2(1,3)	1,0	
	GA	17(11)	3(1,9)	0,49 (0,13 - 1,80)	0,40	1(0,6)	0,52 (0,06 - 4,41)	1,0	0(0)	1,03 (0,04 - 22,50)	1,0
	AA	0(0)	0(0)			0(0)			0(0)		
						0,28			0,55		0,53
						0,40			0,87		0,72
<b>TYR (rs1042602)</b> n:176 p** Trend Test***	CC	46(26,1)	19(10,8)	1,0		4(2,3)	1,0		0(0)	1,0	
	CA	63(35,8)	18(10,2)	0,69 (0,32 - 1,46)	0,34	9(5,1)	1,64 (0,47 - 5,66)	0,55	2(1,1)	3,66 (0,17 - 78,14)	0,51
	AA	12(6,8)	3(1,7)	0,60 (0,15 - 2,39)	0,54	0(0)	0,41 (0,02 - 8,20)	0,57	0(0)		
	CA+AA	75(42,6)	21(11,9)	0,67 (0,32 - 1,39)	0,35	13(7,4)	1,99 (0,61 - 6,48)	0,29	2(1,1)	3,07 (0,14 - 65,6)	0,52
						0,68			0,37		0,63
					0,70			0,23		0,55	
<b>TYR (rs1393350)</b> n:176 p** Trend Test***	GG	89(50,5)	28(15,9)	1,0		9(5,1)	1,0		2(1,1)	1,0	
	GA	32(18,2)	12(6,8)	1,19 (0,54 - 2,62)	0,68	4(2,3)	1,23 (0,35 - 4,29)	0,74	0(0)	0,55 (0,02 - 11,79)	1,0
	AA	0(0)	0(0)			0(0)			0(0)		
						0,66			0,73		0,39
						0,81			0,99		0,97

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*, Williams's G Test (Trend Test).  
Source: Authors.

**Table 2B.** Analysis of SNPs in relation to hair color.

Gene (Polymorphism)	Genotype	Black* n (%)	Brown n (%)	OR (CI)	p*	Blond n (%)	OR (CI)	p*
<b>OCA2 (rs1800404)</b> n:173 p** Trend Test***	GG	12(6,9)	24(13,4)	1,0		0(0)	1,0	
	GA	17(9,8)	52(30,)	1,5 (0,63 - 3,7)	0,36	10(5,8)	15,0 (0,80 - 280,7)	<b>0,01</b>
	AA	15(8,6)	34(19,7)	1,13 (0,45 - 2,84)	0,81	9(5,2)	15,32 (0,80 - 290,0)	<b>0,01</b>
	GA+AA	32(18,4)	86(49,7)	1,34 (0,60 - 3,0)	0,52	19(11,0)	15,0 (0,83 - 267,9)	<b>0,01</b>
					0,77			<b>0,08</b>
					0,78			<b>0,01</b>
<b>OCA2 (rs1800407)</b> n:155 p** Trend Test***	GG	34(21,9)	84(54,2)	1,0		16(10,3)	1,0	
	GA	5(3,2)	14(9,0)	1,13 (0,37 - 3,39)	1,0	2(1,3)	0,85 (0,14 - 4,86)	1,0
	AA	0(0)	0(0)			0(0)		
					0,82			0,85
					0,96			0,80
<b>TYR (rs1042602)</b> n:176 p** Trend Test***	CC	22(12,5)	43(24,4)	1,0		4(2,3)	1,0	
	CA	21(11,9)	58(32,9)	1,41 (0,69 - 2,89)	0,36	13(7,4)	3,40 (0,95 - 12,13)	0,08
	AA	2(1,1)	10(5,7)	2,55 (0,51 - 12,71)	0,32	3(1,7)	8,25 (1,02 - 66,2)	0,06
	CA+AA	23(13,0)	68(38,6)	1,51 (0,75 - 3,04)	0,28	16(9,1)	3,82 (1,10 - 13,25)	0,03
					0,51			0,08
					0,52			0,08
<b>TYR (rs1393350)</b> n:176 p** Trend Test***	GG	34(19,3)	78(44,3)	1,0		16(9,1)	1,0	
	GA	11(6,2)	33(18,7)	1,30 (0,59 - 2,88)	0,56	4(2,3)	0,77 (0,21 - 2,80)	0,76
	AA	0(0)	0(0)			0(0)		
					0,5			0,69
					0,63			0,94

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*: Williams's G Test (Trend Test).  
Source: Authors.

**Table 3B.** Analysis of SNPs in relation to skin color.

Gene (Polymorphism)	Genotype	Black* n (%)	Brown n (%)	OR (CI)	p*	White n (%)	OR (CI)	p*	Yellow n (%)	OR (IC)	p*
<b>OCA2 (rs1800404)</b> n:173 p** Trend Test***	GG	17(9,8)	15(8,7)	1,0		2(1,2)	1,0		2(1,2)	1,0	
	GA	8(4,6)	48(27,7)	6,8 (2,44 - 18,88)	<b>0,0002</b>	20(11,5)	21,25(3,96 - 113,9)	<b>&lt; 0,0001</b>	3(1,7)	3,18 (0,44 - 23,02)	0,32
	AA	5(2,9)	26(15,0)	5,89 (1,80 - 19,23)	<b>0,003</b>	24(13,9)	40,80(7,06 - 235,7)	<b>&lt; 0,0001</b>	3(1,7)	5,10 (0,65 - 39,57)	0,13
	GA+AA	13(7,5)	74(42,8)	6,45 (2,59 - 16,04)	<b>&lt; 0,0001</b>	44(25,4)	28,77(5,86 - 141,2)	<b>&lt; 0,0001</b>	6(3,4)	3,92 (0,67 - 22,71)	0,23
					<b>&lt; 0,0001</b>			<b>&lt; 0,0001</b>			0,34
				<b>0,0002</b>			<b>&lt; 0,0001</b>			0,34	
<b>OCA2 (rs1800407)</b> n:155 p** Trend Test***	GG	22(14,2)	70(45,2)	1,0		35(22,6)	1,0		7(4,5)	1,0	
	GA	2(1,3)	11(7,1)	1,72 (0,35 - 8,40)	0,72	7(4,5)	2,20 (0,41 - 11,57)	0,46	1(0,7)	1,57 (0,12 - 20,07)	1,0
	AA	0(0)	0(0)			0(0)			0(0)		
					0,49				0,34		0,72
					0,73				0,55		0,71
<b>TYR (rs1042602)</b> n:176 p** Trend Test***	CC	15(8,5)	37(23,9)	1,0		14(7,9)	1,0		3(1,7)	1,0	
	CA	14(7,9)	45(25,6)	1,30 (0,55 - 3,04)	0,66	28(15,9)	2,14 (0,81 - 5,65)	0,14	5(2,8)	1,78 (0,35 - 8,90)	0,69
	AA	1(0,6)	7(4,0)	2,83 (0,32 - 25,11)	0,66	7(4,0)	7,5 (0,81 - 68,97)	0,10	0(0)	1,47 (0,04 - 44,46)	1,0
	CA+AA	15(8,5)	52(29,6)	1,40 (0,61 - 3,22)	0,52	35(19,9)	2,50 (0,96 - 6,44)	0,09	5(2,8)	1,66 (0,33 - 8,26)	0,69
					0,72				0,11		0,83
				0,71				0,11		0,83	
<b>TYR (rs1393350)</b> n:176 p** Trend Test***	GG	25(14,2)	61(34,6)	1,0		36(20,4)	1,0		6(3,4)	1,0	
	GA	5(2,8)	28(15,9)	2,29 (0,79 - 6,62)	0,15	13(7,4)	1,80 (0,57 - 5,70)	0,41	2(1,1)	1,66 (0,25 - 10,78)	0,62
	AA	0(0)	0(0)			0(0)			0(0)		
					0,11				0,31		0,58
					0,17				0,45		0,97

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*, Williams's G Test (Trend Test).

Source: Authors..



**Table 4B.** Analysis of SNPs in relation to grouped eye color.

Gene (Polymorphism)	Genotype	Brown* n (%)	Non- Brown n (%)	OR (CI)	p*
<b>OCA2 (rs1800404)</b> n:173 p** Trend Test***	GG	36(20,8)	0(0)	1,0	
	GA	74(42,8)	5(2,9)	5,38 (0,28 - 100,2)	0,32
	AA	48(22,8)	10(5,80)	15,80(0,89 - 278,8)	<b>0,011</b>
	GA+AA	122(70,5)	15(8,7)	9,23(0,53 - 158,3)	<b>0,04</b>
					<b>0,03</b>
					<b>0,0095</b>
<b>OCA2 (rs1800407)</b> n:155 p** Trend Test***	GG	122(78,7)	12(7,7)	1,0	
	GA	20(12,9)	4(2,6)	2,03 (0,59 - 6,93)	0,26
	AA	0(0)	0(0)		
					0,24
					0,45
<b>TYR (rs1042602)</b> n:176 p** Trend Test***	CC	65(36,9)	4(2,8)	1,0	
	CA	81(46,0)	11(6,2)	2,20 (0,67 - 7,25)	0,27
	AA	15(8,5)	0(0)	0,46 (0,02 - 9,19)	1,0
	CA+AA	96(54,5)	11(6,2)	1,86 (0,56 - 6,10)	0,41
					0,32
				0,20	
<b>TYR (rs1393350)</b> n:176 p** Trend Test***	GG	117(66,5)	11(6,2)	1,0	
	GA	44(25,0)	4(2,3)	0,96 (0,29 - 3,19)	1,0
	AA	0(0)	0(0)		
					0,95
					0,80

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*; Williams's G Test (Trend Test).  
Source: Authors.

**Table 5B.** Analysis of SNPs in relation to grouped hair color.

Gene (Polimorphism)	Genotype	Non -Blond* n (%)	Blond n (%)	OR (IC)	p*
<b>OCA2 (rs1800404)</b> n:173 p** Trend Test***	GG	36(20,8)	0(0)	1,0	
	GA	69(39,9)	10(5,8)	11,03(0,62 - 193,7)	<b>0,02</b>
	AA	49(28,3)	9(5,2)	14,01(0,78 - 248,7)	<b>0,01</b>
	GA+AA	118(68,2)	19(11,0)	12,01(0,70 - 204,0)	<b>0,01</b>
					0,1
					<b>0,018</b>
<b>OCA2 (rs1800407)</b> n:155 p** Trend Test***	GG	118(76,1)	16(10,3)	1,0	
	GA	19(12,2)	2(1,3)	0,77 (0,16 - 3,65)	1,0
	AA	0(0)	0(0)		
					0,74
					0,96
<b>TYR (rs1042602)</b> n:176 p** Trend Test***	CC	65(36,9)	4(2,3)	1,0	
	CA	79(44,9)	13(7,4)	2,67 (0,83 - 8,59)	0,12
	AA	12(6,8)	3(1,7)	4,06 (0,80 - 20,5)	0,10
	CA+AA	91(51,7)	16(9,1)	2,85 (0,91 - 8,94)	0,08
					0,22
					0,19
<b>TYR (rs1393350)</b> n:176 p** Trend Test***	GG	112(63,6)	16(9,1)	1,0	
	GA	44(25,0)	4(2,3)	0,63 (0,20 - 2,01)	0,59
	AA	0(0)	0(0)		
					0,43
					0,6

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*; Williams's G Test (Trend Test).

Source: Authors.

**Table 6B.** Analysis of SNPs in relation to grouped skin color.

Gene (Polimorphism)	Genotypes	Black* n (%)	Non-Black n (%)	OR (IC)	p*
<b>OCA2 (rs1800404)</b> n:173 p** Trend Test***	GG	17(9,8)	19(11,0)	1,0	
	GA	8(4,6)	71(41,0)	7,94(2,97 - 21,19)	<0,0001
	AA	5(2,9)	53(30,6)	9,48(3,07 - 29,26)	<0,0001
	GA+AA	13(7,5)	124(71,6)	8,53(3,57 - 20,35)	<0,0001
					<0,0001
<b>OCA2 (rs1800407)</b> n:155 p** Trend Test***	GG	22(14,2)	112(72,2)	1,0	
	GA	2(1,3)	19(12,2)	1,86 (0,40 - 8,59)	0,53
	AA	0(0)	0(0)		0,41
					0,61
<b>TYR (rs1042602)</b> n:176 p** Trend Test***	CC	15(8,5)	54(30,7)	1,0	
	CA	14(7,9)	78(44,3)	1,54 (0,69 - 3,46)	0,30
	AA	1(0,6)	14(7,9)	3,88 (0,47 - 32,02)	0,28
	CA+AA	15(8,5)	92(52,2)	1,70 (0,77 - 3,75)	0,21
					0,38
<b>TYR (rs1393350)</b> n:176 p** Trend Test***	GG	25(14,2)	103(58,5)	1,0	
	GA	5(2,8)	43(24,4)	2,08 (0,74 - 5,81)	0,18
	AA	0(0)	0(0)		0,15
					0,21

\* Reference group; OR: Odds Ratio; CI: Confidence Interval; p\*: odds Ratio; p\*\*: Chi-Square Test; p\*\*\*; Williams's G Test (Trend Test).  
Source: Authors.

### 3.6 TYR

Hardy-Weinberg equilibrium deviations were found for rs1042602 ( $X^2$  4.17 and  $p = 0.04$ ) and for rs1393350 ( $X^2$  4.38 and  $p = 0.03$ ). A higher frequency of ancestral allele C of rs1042602 was observed, 65.3% and polymorphic allele A, 34.7%. The most frequent genotype was the heterozygous genotype CA, followed by CC and AA genotypes, 52.3%, 39.2% and 8.5% respectively. The ancestral allele G of rs1393350 was found in high frequency, 86.3%, followed by mutant allele A, 13.7%. No individuals with the homozygous genotypes AA was identified in this population, the frequencies found for GG and GA genotypes of rs1393350 were 72.7% and 27.3%, respectively. No significant associations were found between rs1393350, rs1042602 with the eye, hair and skin color (Tables 1B – 6B).

### 3.7 Prediction Model

For each of the six tested algorithms, a global precision score was generated, which would be the probability of general correctness of the three characteristics and a score for eye, skin and hair color. The KNeighborsClassifier and DecisionTreeClassifier algorithms showed the best performance for general accuracy, with 43.7%, for the eye color, the RadiusNeighborsClassifier and Support Vector Classifier algorithms obtained the best performance with 73.8%, the latter also

presented the best performance for the skin color, 61.9% and for hair color RadiusNeighborsClassifier and Support Vector Classifier with 76.2% (Table 7).

**Table 7.** Performance of prediction models developed for six different algorithms evaluated for CEV.

Prediction Parameters	Prediction model type					
	Random Forest	Kneighbors	DecisionTree	RadiusNeighbors	ExtraTrees	Support Vector
General hit probability	38,9%	43,7%	43,7%	34,1%	41,3%	29,4%
Precision score - eyes	64,3%	64,3%	64,3%	73,8%	64,3%	73,8%
Precision score - skin	54,8%	47,6%	47,6%	47,6%	50%	61,9%
Precision score - hair	64,3%	57,1%	57,1%	76,2%	61,9%	76,2%

Source: Authors.

#### 4. Discussion

Self-reported information on eye, hair and skin color is a methodology already found in some studies in Brazilian population, such as France et al. who used self-reported skin color to make associations with genotypes (Leite, Fonseca, França, Parra, & Pereira, 2011a), Durso et al in studies with biogeographic ancestry (Durso et al., 2014), Lima et al. and Fridman et al. also used aiming at DNA phenotyping (de Araújo Lima et al., 2015; Fridman, Cardena, Lima, & Gonçalves, 2015). Santos et al (2018) identified the association of these characteristics with genotypes and diseases development (Gomes et al., 2018). Regarding the frequency of self-reported EVCs in the present study, dark brown color was the most frequent to eyes, black color to hair and the most frequent skin color was black. This data corroborate with the finds of Lima et al. (de Araújo Lima et al., 2015) to eye color and differ to hair and skin color, where brown hair and white skin were more frequent. This difference can be explained because these samples come from different regions, the sample evaluated by Lima et al. was from southeastern Brazil.

These proportions vary in different regions of Brazil due to the peculiarities of their colonization, there was a greater contribution of native Americans in north region and Africans in northeast region, in south region a low contribution was observed for both. (Salzano & Sans, 2014). In southeast region, the ancestral proportions show a larger European contribution (78.5%), followed by African (14.7%) and Amerindian (6.7%), in northeast region the European contribution is 42.9% while the African corresponds to 50.8%, Amerindian appears with 6.4% (Kehdy et al., 2015). These differences highlight the importance to evaluate genetic markers in populations from different regions of Brazil, since the Brazilian population is characterized as one of the most heterogeneous in the world. (Salzano & Sans, 2014).

The most frequent genotype of rs6058017 was AA, followed by GA and GG, 59.9%, 33.7% and 6.4% respectively. Our data corroborate with Kanetsky et al.(Kanetsky et al., 2002) who also found a higher frequency of genotype AA (78%) in a sample of the USA population and corroborate with Lima et al, they found genotype AA at a frequency of 70.8% in the southeastern Brazilian population,(de Araújo Lima et al., 2015). In a study conducted by Bonilla et al in African American population, the most common genotype was GG (66%)(Bonilla et al., 2005). In our study the mutant allele A was found at a frequency of 76.7% and the ancestral allele G, 23.3%, corroborating the data found by Lima et al. in southeastern Brazil, where the allele A was found more frequently, 84% and allele G, 16% (de Araújo Lima et al., 2015). The allele A is also found at a high frequency in Caucasians (93%) and Indians (76%) from Australia. (MEZIANI et al., 2005; Sulem et al., 2008; Zeigler-Johnson et al., 2004).

According to Maronas et al, the polymorphism rs6058017 (*ASIP*) is one of the most important markers of dark pigmentation, the allele G is found in almost 100% of individuals with this pigmentation and is often described in association with dark phenotypes in eye, hair and skin color analysis, the allele A has the opposite function, being present in individuals with light pigmentation (Maroñas et al., 2014). In the present study we found significant association between rs6058017 and skin color. The genotypes GA and AA were associated with brown skin. When one copy of allele A is present there is about 29 times more chance for an individual to have brown skin, when compared to black skin. When two copies of allele A are present, this chance increases to 45.5 times. The joint analysis of GA+AA shows a chance of 37.33, reinforcing the strength of allele (Table 3A). Genotype AA was also associated with 18.67 times more chance of presenting white skin, the joint analysis of GA+AA was also significant, but the chance drops to 13.42 times when compared to black skin (Table 3A). The above data show a tendency of the allele A to be strongly associated with non-black skin color, a result also observed in the grouped analysis, where carriers of genotype GA presenting 9.52 times more chance to have non-black skin color and carriers of two copies of allele A presenting 11.13 more chance to have non-black skin, the combined analysis of GA+AA shows a 10.50 times more chance of having non-black skin color, this sum shows the strength of this allele (Table 6A).

Zaorska et al found the association between rs6058017 and skin color in studies using neural networks methodology in a Polish population. They report the influence of this SNP on skin sensitivity to the sun and the freckles presence. (Zaorska et al., 2019). Our data corroborate the findings of Lima et al. where the homozygous genotype AA was related to white and brown skin, this genotype was associated with 8.6 times more chance of presenting white skin and 5.1 more chance of presenting brown skin in a southeastern Brazil population. No significant association were found with eye and hair color, as in the present study (de Araújo Lima et al., 2015). Durso et al did not find any association between this SNP and skin color in Rio de Janeiro - Brazil population (Durso et al., 2014).

In Asians the ancestral allele G appears at a frequency between 10 and 28% of the population, in Africans this allele appears at a frequency of approximately 80%, suggesting the association of this allele with features such as dark eye, skin and hair color, which are mostly observed in these individuals (Bonilla et al., 2005; Kanetsky et al., 2002; Voisey et al., 2006). Our data corroborate with the above cited, indicating a tendency of allele A to be associated with non-black skin pigmentation and G allele with black skin pigmentation.

The ancestral allele C of rs16891982 (*SLC45A2*) was found more frequently, 60.1%, followed by allele mutant G, 39.9%, corroborating with data found by Fracasso et al., where the frequency found for allele C was 64.33% in southeastern Brazil population (Fracasso et al., 2017). According to Sawitzki et al, that analyzed data from various populations in the world on ALFRED (The Allele Frequency Database), the C allele was found at a frequency of 88% in populations with a high melanin index and 6% in populations with a low melanin index. The G allele is found at a frequency of 11% in populations with a high melanin index and 93% in those with a low melanin index. (Sawitzki et al., 2017). In the present study one copy of G allele was associated with 5.14 times more chance to having white skin. When the two copies are present, this chance increases to 30 times when compared to black skin. The joint analysis of GA+AA was also significant; the chance was 6.8 time (Table 3A). Thus, the allele G seems to be associated with phenotypes that present a low melanin index, such as white skin.

Fridman et al reported an association between genotypes CG (OR: 3.08) and GG (OR: 16.35) with white skin in a southeastern Brazil population, the chance of individual presenting white skin increases when it carries the genotype GG, as in the present study (Fridman et al., 2015). In our sample, the most frequent genotype was CG at a frequency of 57.8%, followed by genotypes CC and GG, 31.1% and 11%, respectively. Hernando et al. performed a study in southeastern Europe population in Spain and reported that GG genotype was the most recurrent in this population. The ancestral C allele appears to have a protective effect on skin sensitivity to the sun. (Hernando et al., 2018). Soejima et al described that in European

populations the allele G is found at high frequency and the allele C at low frequency, the opposite happens in South Africa populations, the ancestral allele C is more recurrent and allele G appears at very low frequencies (Soejima & Koda, 2006). According to Leite et al, individuals with lighter skin tend to have a higher proportion of European ancestry, while individuals with darker skin a greater proportion of African ancestry (Leite et al., 2011a). Our data follow the same trend as the results mentioned above, due to the migration processes in Brazil, resulting in the crossing of populations from different continents and different exchanges of genetic information. (Hart et al., 2013).

The presence of Allele C in rs16891982 encodes the amino acid leucine which plays an important role in proton transport providing an optimal pH within the melanosomes, which allows the tyrosinase activity and the adequate eumelanine production, abundant pigment in dark pigmented individuals. When the allele G is present, encoding another amino acid, phenylalanine, there may be a change in pH and eumelanine synthesis (Sawitzki et al., 2017), this may be associated with the appearance of low melanin characteristics such as skin, eyes and light hair. Our data corroborate the above and also suggest an association of the allele G with fair skin pigmentation.

Hardy-Weinberg equilibrium deviations were found for rs16891982 (*SLC45A2*) ( $X^2$  6,45;  $p = 0,01$ ) and rs1426654 (*SLC24A5*) ( $X^2$  7,12;  $p = 0,007$ ), in both cases a higher heterozygotes frequency was found, a factor that may have caused this deviation, due to the interethnic mixture and a population substructure, a common event in a population with high miscegenation. (Chen, Cole, & Grond-Ginsbach, 2017; Wilde et al., 2014).

The polymorphic allele A of rs1426654 (*SLC24A5*) was found at a frequency of 71.6%, the ancestral G, 28.4%. Lima et al also found the high frequency of allele A, 79% and allele G, 21% (de Araújo Lima et al., 2015). Lamason et al describe the important role of rs1426654 in pigmentation, African and East Asian populations have a high frequency (93 to 100%) of ancestral allele G, these populations have dark phenotypes, allele A is found at a frequency of 98 to 100 % in European populations, characterized by lighter phenotypes. The allele G promotes a higher eumelanine production, favoring the occurrence of dark pigment. (Lamason, 2005). Canfield et al describe that the occurrence of allele A in individuals living near the equator line is lower, according to the authors. This is a characteristic that justifies the natural selection action that led to the reduction of melanization in Europe (Canfield et al., 2013). The present study population had a higher frequency of heterozygotes, 48.9%, followed by AA and GG, 47.1% and 4%, respectively, differing from data found by Lima et al in southeastern Brazil, where AA was the most frequent genotype, 67.5%, followed by GA and GG, 23.4% and 9.1%, respectively. Data from Lima et. al also report a strong association between genotype AA with fair skin (OR: 47.8), when compared to the black skin group, the heterozygous genotype GA was also associated with individuals from the non-black skin color group (de Araújo Lima et al., 2015).

In the present study significant associations were found between rs1426654 and skin color, genotype AA was strongly associated with brown skin color ( $p = 0.0006$  OR: 60.0), individuals carrying genotype AA presented 60 times more chance of having brown skin when compared to the black skin group. And when performing the joint analysis of GA + AA was also found an association ( $p = 0.01$  OR: 13.54) with brown skin (Table 3A). Homozygous genotype AA was strongly associated with white skin, individuals carrying two copies of allele A presented 40 times more chance to have white skin color ( $p = 0.002$  OR: 40.0) (Table 3A). No significant associations were found to heterozygous genotype GA, these were only found in joint analysis of GA+AA (Table 3A), indicating a relationship between the genotype AA and non-black individuals. The homozygous genotype GG of rs1425564 was associated with an increase in melanin index in Brasilia population (Brazil West Center). This work also reports the strong association between rs1426654 and skin color. (Leite, Fonseca, França, Parra, & Pereira, 2011b). However, the effect of alleles A and G together is not well understood, and further studies are necessary to explain this relationship (Durso et al., 2014).

This tendency of genotype AA to be associated with non-black individuals was also found in the grouped analysis. Homozygous genotype AA was strongly associated with non-black skin color ( $p = 0.0004$  OR: 35.56), and in joint analysis of GA + AA also found an association with non-black skin ( $p = 0.01$  OR: 7.33) (Table 6A). These data corroborate with the studies performed by Maronas et al., who described rs1426654 as the first marker to differentiate black and non-black individuals (Maroñas et al., 2014). In south Brazil, a high frequency of allele A was reported in individuals with low melanin index, in these same individuals the frequency of allele G was very low. (Sawitzki et al., 2017). Our data also indicate this trend observed in literature cited above, where allele A is associated with light skin pigmentation and G allele with dark skin pigmentation.

We found a high frequency of ancestral allele G to rs885479, 94.1%, whereas the polymorphic allele A appears at a frequency of 5.9%, since no individuals with the genotype AA were found, genotype GG was the most frequent, 88.2%, followed by heterozygous genotype GA, 11.8%. Our data corroborate with those contained in ENSEMBL ([www.ensembl.org](http://www.ensembl.org)) ("Ensembl," n.d.) and NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) ("NCBI," n.d.), both report that the global frequency of allele A is low, 19%, allele G reaches 80%. The polymorphic allele is very recurrent in East Asian populations, reaching frequencies of up to 73%. (Motokawa et al., 2006; SHI, LU, LUO, XIANG-YU, & ZHANG, 2001). In African populations allele G appears at very high frequencies, over 90% and allele A at very low frequencies (Deng & Xu, 2018). This polymorphism has been associated in some studies with skin color pigmentation. (Adhikari et al., 2019; Chaitanya et al., 2018; Hart et al., 2013). No significant associations were observed between rs885479 and EVCs in present study (Tables 1A – 6A), this may be due to the low allele A frequency and the number of individuals analyzed, requiring a more representative sample to find possible associations.

Our data indicate that rs1800404 (*OCA2*) appears to play an important role in pigmentation and in defining Externally Visible Characteristics. We found associations between the allele A of rs1800404 and light phenotype color, while the ancestral allele G with dark phenotypes. Our data corroborate with Andrade et al. in a study conducted in Southeastern Brazil (Fracasso et al., 2017). Homozygous genotype AA was strongly associated with non-brown eye color, ( $p = 0.01$ ; OR: 15.80), presenting 15 times more chances of having non-brown eyes (Table 4B). Adding GA and AA genotypes, a correlation was also found, but with a lower chance of 9.23 times, to occurrence non-brown eyes, ( $p = 0.04$ ; OR: 9.23) (Table 4B). From this data it can be inferred that one copy of allele A is not sufficient to determine the non-brown eye color but the presence of the genotype AA.

In the analyzes by category, were observed associations of the homozygous genotype AA with 17.39 times more chance to have green eyes ( $p = 0.0084$ ; OR: 17.39). When was analyzed the genotypes GA and AA together, a significant association was also found ( $p = 0.03$ ; OR: 10.03) (Table 1B), in this case two copies of allele A seem to be necessary for the development of feature. Andrade et al. in a study conducted in Southeast Brazil also obtained similar results, the genotype AA was associated with green eye color and blond hair (Andrade et al., 2017).

Our data corroborate with Andrade's findings in association with blond hair, the genotypes GA ( $p = 0.01$ ; OR: 15.0) and AA ( $p = 0.01$ ; OR : 15.32) were associated with 15 times more chance of having blond hair (Table 2B). Joint analysis of GA + AA reinforces the association with blond hair and the strength of allele A, one copy of the allele A seems to be sufficient to determine this characteristic ( $p = 0.01$ ; OR: 15.0) (Table 2B). The same trend was found in grouped analysis, GA, AA genotypes and joint analysis of GA + AA were also associated with blond hair color when compared to the non-blond group, ( $p = 0.02$ ; OR: 11.03), ( $p = 0.01$ ; OR: 14.01), ( $p = 0.01$ ; OR: 12.01), respectively (Table 5B). Differing from the Adhikari et al. that performed an analysis in Latin Americans, no significant associations were found between rs1800404 and hair color, this SNP was only associated with eye and skin color (Adhikari et al., 2019).

Genotypes GA, AA and GA + AA of rs1800404 were associated with non-black skin color in the grouped analysis ( $p < 0.0001$ ; OR: 7.94); ( $p < 0.0001$ ; OR: 9.48), ( $p < 0.0001$ ; OR: 8.53), respectively (Table 6B). One copy of the allele A gives 7.94 times more chance of having a non-black skin color and when two copies of the allele A are present (genotype AA) this chance increases to 9.48 times, the joint analysis again reinforces the strength of the allele A. In the category analyzes this SNP was related to brown skin color and most strongly associated with white skin color. The associations found between brown skin and GA, AA and GA + AA were ( $p = 0.0002$ ; OR: 6.8), ( $p = 0.003$ ; OR: 5.89), ( $p < 0.0001$ ; OR: 6.45), respectively. To white skin we found ( $p < 0.0001$ ; OR: 21, 25), ( $p < 0.0001$ ; OR: 40.80), ( $p < 0.0001$ ; OR: 28.77), also respectively (Table 3B). When one copy of allele A is present there is 21.25 times more chance to have white skin, the presence of genotype AA gives 40.8 times more chance of having skin white, showing a strong association between rs1800404 and skin color.

According to Crawford et al. the ancestral allele G is associated with dark skin pigmentation and is common in most Africans, while the polymorphic allele A is related to light skin pigmentation, more common in Europeans with a frequency > 70%. In the present study the polymorphic allele was found at a frequency of 56.3% (Crawford et al., 2017) In Southeast Brazil this same association was found, the allele G related to dark skin and the allele A with light skin, corroborating with the data found in our study. (Andrade et al., 2017). The literature relates most commonly rs1800404 with skin pigmentation, this SNP showed a significant effect on skin color in African American and Afro-Caribbean populations. (Deng & Xu, 2018).

The frequency of allele G, the most common in African populations, was 43.6% in our study, and the A allele, 56.3%, which is more common in European populations. This genotypic and phenotypic variety found in Brazil is the result of a mixture of Africans, Europeans and Native Americans, for about five centuries, generating one of the most mixed populations in the world. It is important to highlight that the variation in the colonization and occupation process of the Brazilian regions and states, created different degrees of genetic mixture throughout the country, the northeast was the cradle of portuguese colonization and where most slaves landed and settled, in northeast a Greater African ancestry is observed, while in the south a greater European ancestry, making it important to study these molecular markers in different regions and states in the same country, as well to elucidate more specific markers for each population (Magalhães da Silva et al., 2015; Pena et al., 2011; Souza, Resende, Sousa, & Brito, 2019).

There are few data in Brazil relating rs1800404 with Externally Visible Characteristics, our data show that this SNP may be a possible phenotypic predictor for the Pernambuco population. Importantly, phenotypic prediction is a complex process that involves many genetic and environmental factors and a set of markers is required to predict a characteristic, but rs1800404 can be an important marker and along with other markers that are associated with the same traits, it can be used for phenotypic prediction.

No significant associations were found between rs1800407 (*OCA2*) and the externally visible characteristics evaluated (Table 1B – 6B), studies bring association of derivative allele A with light iris pigmentations, such as blue eyes (Andersen et al., 2016; Pośpiech et al., 2016; Walsh et al., 2011). However, in the present study the derived allele A was found at a very low frequency, 6.8%, while the ancestral allele was found at a frequency of 93.2%. This low frequency of allele A along with the sample size in this paper may be the reason for not finding significant associations. No individuals with genotype AA were found, directly reflecting the low frequency of the polymorphic allele in this population. What is expected because the allele A is more restricted mainly to Europe in a percentage of (0-11%), Italian (9.7%), Portuguese (7.5%), outside this region is found in low frequencies, Afro Americans (1.7%), Chinese (0.9%) (Andersen et al., 2016; Donnelly et al., 2012; Gu et al., 2011). A study conducted by Andrade et. al. in Brazil shows the polymorphic allele at a frequency of approximately 8% and allele G was associated with dark skin color (Andrade et al., 2017).



Deviations in the Hardy-Weinberg equilibrium were found to rs1042602 (*TYR*) and rs1393350 (*TYR*), ( $X^2$  4.17 and  $p = 0.04$ ), ( $X^2$  4.38 and  $p = 0.03$ ), respectively. These deviations may be associated with heterozygosis excess to rs1042602 (Chen et al., 2017), or as a result of population stratification, a common event in mixed populations, such as Brazil (de Araújo Lima et al., 2015). The most frequent genotype in rs1042602 was CA, followed by CC and AA genotypes, 52.3%, 39.2%, 8.5% respectively, as also found by Hernando et al in a sample of the Spanish population. (Hernando et al., 2018). Genotypic frequencies of rs1393350 were GG (72.7%) and GA (27.3%), no individuals with genotype AA were found, which may also have contributed to the deviation in EHW.

The ancestral allele C of rs1042602 was found at a frequency of 65.3%, while polymorphic A at a frequency of 34.7%. Sawitzki et al studying a population in south Brazil found the allele C at a frequency of 86% and allele A, 13% in people with high melanin content (from African populations). In people with low melanin content (from European populations) allele C was found at a frequency of 62% and allele A at a frequency of 37%. Performing an analysis with populations from around the world with data from ALFRED, Sawitzki et al also found the allele C in higher frequencies and our findings corroborate the data. (Sawitzki et al., 2017).

No associations were found between rs1042602 and eye, hair and skin color in the present study (Table 1B – 6B), unlike the studies by Adhikari et al, in which this SNP was associated with eye and hair pigmentation (Adhikari et al., 2019). When the allele C is replaced by the polymorphic allele A, this results in a decrease of about 40% in the catalytic activity of tyrosinase, directly affecting melanin production. Thus the presence of allele A is strongly associated with individuals with low melanin content. (Sawitzki et al., 2017). There are studies showing association of rs1042602 with the absence and presence of freckles (Kukla-Bartoszek et al., 2019), as well as in various populations with eye, skin and hair color (Chaitanya et al., 2018). Durso et al also found an association between rs1042602 and skin color in a Rio de Janeiro population (Durso et al., 2014). The reason for not finding significant associations with the CEV in our work may be due to the sample size, and maybe more individuals may need to be evaluated to find significant associations.

The ancestral allele G of rs1393350 (*TYR*) was observed at a frequency of 86.3%, followed by the mutant allele A, 13.7%, this result is expected as the genotype AA has a lower frequency in many populations worldwide (Walsh et al., 2011). This polymorphism has been associated with eye, skin and hair color, but especially eye color (Virmond et al., 2016; Walsh et al., 2011; Yun et al., 2014) In our studies no significant associations were observed between rs1393350 and skin, hair and eye color (Table 1B -6B). Pospiech et al found the polymorphic allele at a frequency of about 22% and also found no association with eye color. The allele A was associated with increased susceptibility to blond instead of brown hair by 1.29 times, a higher chance of having blue eyes instead of green eyes by 1.29 times and an increase in skin sensitivity to the sun in European population (Sulem et al., 2007). In our study, no association may have been found by the low frequency of the allele A in our sample. Hohl et al in studies conducted in Buenos Aires population also found no significant associations between rs1393350 and eye color, he comments that the genetic context of iris color in Argentina is different from other countries due to population miscegenation (Hohl, Bezus, Ratowiecki, & Catanesi, 2018), as is the case in Brazil, according to Hohl, markers already elucidated in European countries should be carefully applied to populations with different characteristics.

The tested algorithms for a possible prediction model, showed encouraging results, since it is a heterogeneous population, because most of the studies that propose a prediction model were carried out in homogeneous populations. The algorithms that presented better global performance, were above 40%, better performance for eye color above 70%, hair color above 75% and skin color above 61.9% (Table 7), lower values than those found in studies by Maroñas and Walsh (Maroñas et al., 2014; Walsh et al., 2017) who found 70 to 97% prediction values for skin color, while Valenzuela et al (Valenzuela et al., 2010) found a prediction value lower than that found in the present study 45.7%. Zaorska et al (Zaorska et al., 2019) in their

studies in a Polish population, with 150 individuals they found a value of 58.3% for skin color with the Random Forest algorithm, one of which was tested in the present study, we found a value of 54.8% (Table 7), a value lower than that found with the support vector algorithm, 61.9% (Table 7), which showed the best performance for skin color.

The different values found in mentioned studies may be due to differences in the sample size, in the algorithm used or because of the characteristics of the studied populations, which are homogeneous, different from this study, which is heterogeneous. The values found in the tested algorithms for a future prediction model for this population, are satisfactory, since some of the values found exceed 70% for eye and hair color and skin compared to a population of similar sample size, the one studied by Zaorska(Zaorska et al., 2019), our model performed better. These results point to an encouraging starting point for the construction of a prediction model for the heterogeneous Pernambuco population and others heterogeneous populations.

These associations found in our data and the values found in prediction models show the importance of SNPs evaluation used in DNA Phenotyping, which is emerging as a promising forensic tool, because of its possible ability to provide new information in the context of criminal investigations. Through this technique it is possible to infer information about Externally Visible Characteristics, as well as ethnicity and information about biogeographic ancestry (Queirós, 2019). However, data are still scarce in non-European and mixed populations such as Brazil, and further studies with a larger number of individuals are needed. However, getting associations of SNPs with CEVs and knowing how significant they can be, as found in our study, can help in phenotypic prediction analysis in heterogeneous populations.

## 5. Conclusion

Our results suggest that rs6058017, rs16891982 and rs1426654 polymorphisms may be used as phenotypic predictors of skin color and rs1800404 may be used as phenotypic predictor to skin, hair and eye color in Pernambuco population, Brazil, along with other markers that are also associated with these same characteristics. The prediction models evaluated showed satisfactory prediction rates, rates above 60% for skin color and above 70% for eyes and hair. This associations and models can assist in forensic DNA phenotyping in Pernambuco population and others mixed populations. The results showed a starting point for the development of prediction models for heterogeneous populations. However, further studies are necessary to confirm the phenotypic power of this SNPs and the prediction model. The markers studied in this study can be evaluated in other populations from different regions of Brazil, as well as other CEV markers can be added to the panel already used in this study.

## Acknowledgments

This Study was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

## References

- Adhikari, K., Mendoza-Revilla, J., Sohail, A., Fuentes-Guajardo, M., Lampert, J., Chacón-Duque, J. C., & Ruiz-Linares, A. (2019). A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nature Communications*, *10*(1), 358. <https://doi.org/10.1038/s41467-018-08147-0>
- Andersen, J. D., Pietroni, C., Johansen, P., Andersen, M. M., Pereira, V., Børsting, C., & Morling, N. (2016). Importance of nonsynonymous OCA2 variants in human eye color prediction. *Molecular Genetics & Genomic Medicine*, *4*(4), 420–430. <https://doi.org/10.1002/mgg3.213>
- Andrade, E. S., Fracasso, N. C. A., Strazza Júnior, P. S., Simões, A. L., & Mendes-Junior, C. T. (2017). Associations of OCA2 - HERC2 SNPs and haplotypes with human pigmentation characteristics in the Brazilian population. *Legal Medicine*, *24*, 78–83. <https://doi.org/10.1016/j.legalmed.2016.12.003>

- Ayres, M.; Ayres Jr, M.; Ayres, D. L.; Santos, A. de A. dos S. (2007). *Bioestat 5.0*. Belém; Sociedade Civil Mamirauá: cnpq.
- Bellono, N. W., Escobar, I. E., Lefkovich, A. J., Marks, M. S., & Oancea, E. (2014). An intracellular anion channel critical for pigmentation. *ELife*, 3, e04543. <https://doi.org/10.7554/eLife.04543>
- Bin, B.-H., Bhin, J., Yang, S. H., Shin, M., Nam, Y.-J., Choi, D.-H., & Lee, T. R. (2015). Membrane-Associated Transporter Protein (MATP) Regulates Melanosomal pH and Influences Tyrosinase Activity. *PLOS ONE*, 10(6), e0129273. <https://doi.org/10.1371/journal.pone.0129273>
- Bonilla, C., Boxill, L.-A., Donald, S. A. M., Williams, T., Sylvester, N., Parra, E. J., & Kittles, R. A. (2005). The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. *Human Genetics*, 116(5), 402–406. <https://doi.org/10.1007/s00439-004-1251-2>
- Branicki, W., Brudnik, U., Kupiec, T., Wolańska-Nowak, P., & Wojas-Pelc, A. (2007). Determination of Phenotype Associated SNPs in the MC1R Gene. *Journal of Forensic Sciences*, 52(2), 349–354. <https://doi.org/10.1111/j.1556-4029.2006.00361.x>
- Canfield, V. A., Berg, A., Peckins, S., Wentzel, S. M., Ang, K. C., Oppenheimer, S., & Cheng, K. C. (2013). Molecular Phylogeography of a Human Autosomal Skin Color Locus Under Natural Selection. *G3 & Genes|Genomes|Genetics*, 3(11), 2059–2067. <https://doi.org/10.1534/g3.113.007484>
- Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., & Walsh, S. (2018). The HIRISplex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Science International: Genetics*, 35(March), 123–135. <https://doi.org/10.1016/j.fsigen.2018.04.004>
- Chen, B., Cole, J. W., & Grond-Ginsbach, C. (2017). Departure from Hardy Weinberg Equilibrium and Genotyping Error. *Frontiers in Genetics*, 8(OCT), 1–6. <https://doi.org/10.3389/fgene.2017.00167>
- Cook, A. L., Chen, W., Thurber, A. E., Smit, D. J., Smith, A. G., Bladen, T. G., & Sturm, R. A. (2009). Analysis of Cultured Human Melanocytes Based on Polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P Loci. *Journal of Investigative Dermatology*, 129(2), 392–405. <https://doi.org/10.1038/jid.2008.211>
- Crawford, N. G., Kelly, D. E., Hansen, M. E. B., Beltrame, M. H., Fan, S., Bowman, S. L., & Tishkoff, S. A. (2017). Loci associated with skin pigmentation identified in African populations. *Science*, 358(6365), eaan8433. <https://doi.org/10.1126/science.aan8433>
- D??bniak, T., Scott, R., Masojc, B., Serrano-Fernández, P., Huzarski, T., Byrski, T., & Lubinski, J. (2006). MC1R common variants, CDKN2A and their association with melanoma and breast cancer risk. *International Journal of Cancer*, 119(11), 2597–2602. <https://doi.org/10.1002/ijc.22210>
- de Araújo Lima, F., de Toledo Gonçalves, F., & Fridman, C. (2015). SLC24A5 and ASIP as phenotypic predictors in Brazilian population for forensic purposes. *Legal Medicine*, 17(4), 261–266. <https://doi.org/10.1016/j.legalmed.2015.03.001>
- Deng, L., & Xu, S. (2018). Adaptation of human skin color in various populations. *Hereditas*, 155(1), 1. <https://doi.org/10.1186/s41065-017-0036-2>
- Donnelly, M. P., Paschou, P., Grigorenko, E., Gurwitz, D., Barta, C., Lu, R.-B., Kidd, K. K. (2012). A global view of the OCA2-HERC2 region and pigmentation. *Human Genetics*, 131(5), 683–696. <https://doi.org/10.1007/s00439-011-1110-x>
- Durso, D. F., Bydlowski, S. P., Hutz, M. H., Suarez-Kurtz, G., Magalhães, T. R., & Junho Pena, S. D. (2014). Association of Genetic Variants with Self-Assessed Color Categories in Brazilians. *PLoS ONE*, 9(1), e83926. <https://doi.org/10.1371/journal.pone.0083926>
- Ensembl. (n.d.). Retrieved from [www.ensembl.org](http://www.ensembl.org)
- Feng, H., Xia, X., Li, C., Song, Y., Qin, C., Zhang, Y., & Lan, X. (2015). TYR as a multifunctional reporter gene regulated by the Tet-on system for multimodality imaging: an in vitro study. *Scientific Reports*, 5(1), 15502. <https://doi.org/10.1038/srep15502>
- Fernandez, L., Milne, R., Bravo, J., Lopez, J., Avilés, J., Longo, M., ... Ribas, G. (2007). MC1R: three novel variants identified in a malignant melanoma association study in the Spanish population. *Carcinogenesis*, 28(8), 1659–1664. <https://doi.org/10.1093/carcin/bgm084>
- Fracasso, N. C. de A., de Andrade, E. S., Wiesel, C. E. V., Andrade, C. C. F., Zanão, L. R., da Silva, M. S., & Mendes-Junior, C. T. (2017). Haplotypes from the SLC45A2 gene are associated with the presence of freckles and eye, hair and skin pigmentation in Brazil. *Legal Medicine*, 25, 43–51. <https://doi.org/10.1016/j.legalmed.2016.12.013>
- Fridman, C., Cardena, M. M. S. G., Lima, F. D. A., & Gonçalves, F. D. T. (2015). Is it possible to use Forensic DNA phenotyping in Brazilian population? *Forensic Science International: Genetics Supplement Series*, 5, e378–e380. <https://doi.org/10.1016/j.fsigs.2015.09.150>
- Giebel, L. B., & Spritz, R. A. (1990). RFLP for Mbol in the human tyrosinase (TYR) gene detected by PCR. *Nucleic Acids Research*, 18(10), 3103–3103. <https://doi.org/10.1093/nar/18.10.3103-a>

- Gomes, M. B., Gabrielli, A. B., Santos, D. C., Pizarro, M. H., Barros, B. S. V., Negrato, C. A., & Silva, D. A. (2018). Self-reported color-race and genomic ancestry in an admixed population: A contribution of a nationwide survey in patients with type 1 diabetes in Brazil. *Diabetes Research and Clinical Practice*, 140, 245–252. <https://doi.org/10.1016/j.diabres.2018.03.021>
- Graf, J., Voisey, J., Hughes, I., & van Daal, A. (2007). Promoter polymorphisms in the MATP ( SLC45A2 ) gene are associated with normal human skin color variation. *Human Mutation*, 28(7), 710–717. <https://doi.org/10.1002/humu.20504>
- GraphPad Prism version 5.04. (n.d.). Retrieved from ([www.graphpad.com](http://www.graphpad.com)).
- Gu, Y., Yun, L., Zhang, L., Yang, F., & Hou, Y. (2011). The potential forensic utility of two single nucleotide polymorphisms in predicting biogeographical ancestry. *Forensic Science International: Genetics Supplement Series*, 3(1), e105–e106. <https://doi.org/10.1016/j.fsigss.2011.08.052>
- Hart, K. L., Kimura, S. L., Mushailov, V., Budimlija, Z. M., Prinz, M., & Wurmbach, E. (2013). Improved eye- and skin-color prediction based on 8 SNPs. *Croatian Medical Journal*, 54(3), 248–256. <https://doi.org/10.3325/cmj.2013.54.248>
- Hernando, B., Sanz-Page, E., Pitarch, G., Mahiques, L., Valcuende-Cavero, F., & Martinez-Cadenas, C. (2018). Genetic variants associated with skin photosensitivity in a southern European population from Spain. *Photodermatology, Photoimmunology & Photomedicine*, 34(6), 415–422. <https://doi.org/10.1111/phpp.12412>
- Hohl, D. M., Bezus, B., Ratowiecki, J., & Catanesi, C. I. (2018). Genetic and phenotypic variability of iris color in Buenos Aires population. *Genetics and Molecular Biology*, 41(1), 50–58. <https://doi.org/10.1590/1678-4685-gmb-2017-0175>
- J Park, AH Talukder, 1 SA Lim, 2 K Kim, 1 K Pan, 1 B Melendez, 1 SD Bradley, 1 KR Jackson, 1 JS Khalili, 1 J Wang, 1 C Creasy, 1 BF Pan, 3 SE Woodman, 1 C Bernatchez, 1 D Hawke, P Hwu, KM Lee, J Roszik, G Lizée, and C. Y., Aboud, K., Kang, H., Cutting, L. E., & Bennett, A. (2016). Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. In S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, & W. Wells (Eds.), *Cancer Immunol Res*. <https://doi.org/10.1007/978-3-319-46720-7>
- Jackson, I. J. (2006). Pigmentary Diversity: Identifying the genes causing human diversity. *European Journal of Human Genetics*, 14(9), 979–980. <https://doi.org/10.1038/sj.ejhg.5201659>
- Jacobs, L. C., Liu, F., Pardo, L. M., Hofman, A., Uitterlinden, A. G., Kayser, M., & Nijsten, T. (2015). IRF4, MC1R and TYR genes are risk factors for actinic keratosis independent of skin color. *Human Molecular Genetics*, 24(11), 3296–3303. <https://doi.org/10.1093/hmg/ddv076>
- Jannot, A.-S., Meziani, R., Bertrand, G., Gérard, B., Descamps, V., Archimbaud, A., & Melan-Cohort. (2005). Allele variations in the OCA2 gene (pink-eyed-dilution locus) are associated with genetic susceptibility to melanoma. *European Journal of Human Genetics*, 13(8), 913–920. <https://doi.org/10.1038/sj.ejhg.5201415>
- K, B., & Purohit, R. (2013). Mutational analysis of TYR gene and its structural consequences in OCA1A. *Gene*, 513(1), 184–195. <https://doi.org/10.1016/j.gene.2012.09.128>
- Kanetsky, P. A., Swoyer, J., Panossian, S., Holmes, R., Guerry, D., & Rebbeck, T. R. (2002). A Polymorphism in the Agouti Signaling Protein Gene Is Associated with Human Pigmentation. *The American Journal of Human Genetics*, 70(3), 770–775. <https://doi.org/10.1086/339076>
- Kayser, M. (2015). Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics*, 18, 33–48. <https://doi.org/10.1016/j.fsigen.2015.02.003>
- Kayser, M., & Schneider, P. M. (2009). DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic Science International: Genetics*, 3(3), 154–161. <https://doi.org/10.1016/j.fsigen.2009.01.012>
- Kehdy, F. S. G., Gouveia, M. H., Machado, M., Magalhães, W. C. S., Horimoto, A. R., Horta, B. L., & Tarazona-Santos, E. (2015). Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proceedings of the National Academy of Sciences*, 112(28), 8696–8701. <https://doi.org/10.1073/pnas.1504447112>
- Koops, B.-J., & Schellekens, M. H. M. (2006). Forensic DNA Phenotyping: Regulatory Issues. *SSRN Electronic Journal*, 158–202. <https://doi.org/10.2139/ssrn.975032>
- Kukla-Bartoszek, M., Pośpiech, E., Woźniak, A., Boroń, M., Karłowska-Pik, J., Teisseyre, P., & Branicki, W. (2019). DNA-based predictive models for the presence of freckles. *Forensic Science International: Genetics*, 42(June), 252–259. <https://doi.org/10.1016/j.fsigen.2019.07.012>
- Lamason, R. L. (2005). SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. *Science*, 310(5755), 1782–1786. <https://doi.org/10.1126/science.1116238>
- Leite, T. K. M., Fonseca, R. M. C., França, N. M. de, Parra, E. J., & Pereira, R. W. (2011a). Genomic Ancestry, Self-Reported “Color” and Quantitative Measures of Skin Pigmentation in Brazilian Admixed Siblings. *PLoS ONE*, 6(11), e27162. <https://doi.org/10.1371/journal.pone.0027162>

- Leite, T. K. M., Fonseca, R. M. C., França, N. M. de, Parra, E. J., & Pereira, R. W. (2011b). Genomic Ancestry, Self-Reported “Color” and Quantitative Measures of Skin Pigmentation in Brazilian Admixed Siblings. *PLoS ONE*, *6*(11), e27162. <https://doi.org/10.1371/journal.pone.0027162>
- Magalhães da Silva, T., Sandhya Rani, M. R., de Oliveira Costa, G. N., Figueiredo, M. A., Melo, P. S., Nascimento, J. F., & Blanton, R. E. (2015). The correlation between ancestry and color in two cities of Northeast Brazil with contrasting ethnic compositions. *European Journal of Human Genetics*, *23*(7), 984–989. <https://doi.org/10.1038/ejhg.2014.215>
- Makova, K., & Norton, H. (2005). Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. *Peptides*, *26*(10), 1901–1908. <https://doi.org/10.1016/j.peptides.2004.12.032>
- Maroñas, O., Phillips, C., Söchtig, J., Gomez-Tato, A., Cruz, R., Alvarez-Dios, J., & Lareu, M. V. (2014). Development of a forensic skin colour predictive test. *Forensic Science International: Genetics*, *13*, 34–44. <https://doi.org/10.1016/j.fsigen.2014.06.017>
- Meziani, R., Descamps, V., Gerard, B., Matichard, E., Bertrand, G., Archimbaud, A., & Bassetseguin, N. (2005). Association study of the g.8818A&gt;G polymorphism of the human agouti gene with melanoma risk and pigmentary characteristics in a French population. *Journal of Dermatological Science*, *40*(2), 133–136. <https://doi.org/10.1016/j.jdermsci.2005.08.001>
- Motokawa, T., Kato, T., Hashimoto, Y., Hongo, M., Ito, M., Takimoto, H., & Katagiri, T. (2006). Characteristic MC1R polymorphism in the Japanese population. *Journal of Dermatological Science*. <https://doi.org/10.1016/j.jdermsci.2005.10.006>
- NCBI. (n.d.). [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)
- OEGE - Online Encyclopedia for Genetic Epidemiology studies. (n.d.). <http://oege.org/>
- Pena, S. D. J., Di Pietro, G., Fuchshuber-Moraes, M., Genro, J. P., Hutz, M. H., Kehdy, F. de S. G., & Suarez-Kurtz, G. (2011). The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. *PLoS ONE*, *6*(2), e17063. <https://doi.org/10.1371/journal.pone.0017063>
- Pośpiech, E., Karłowska-Pik, J., Ziemkiewicz, B., Kukla, M., Skowron, M., Wojas-Pelc, A., & Branicki, W. (2016). Further evidence for population specific differences in the effect of DNA markers and gender on eye colour prediction in forensics. *International Journal of Legal Medicine*, *130*(4), 923–934. <https://doi.org/10.1007/s00414-016-1388-2>
- Qiao, L., Yang, Y., Fu, P., Hu, S., Zhou, H., Peng, S., & Tang, K. (2018). Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction. *Journal of Genetics and Genomics*, *45*(8), 419–432. <https://doi.org/10.1016/j.jgg.2018.07.009>
- Queirós, F. (2019). The visibilities and invisibilities of race entangled with forensic DNA phenotyping technology. *Journal of Forensic and Legal Medicine*, *68*(April), 101858. <https://doi.org/10.1016/j.jflm.2019.08.002>
- Rebbeck, T. R., Kanetsky, P. A., Walker, A. H., Holmes, R., Halpern, A. C., Schuchter, L. M., & Guerry, D. (2002). P gene as an inherited biomarker of human eye color. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *11*(8), 782–784. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12163334>
- Salzano, F. M., & Sans, M. (2014). Interethnic admixture and the evolution of Latin American populations. *Genetics and Molecular Biology*, *37*(1 suppl 1), 151–170. <https://doi.org/10.1590/S1415-4752014000200003>
- Sawitzki, F. R., Rodenbusch, R., Gubert, D. W., Soares, D., Santos, B., Filipe, E., & Silva, A. (2017). *SM Gr up Analysis of Eight SNPs in South Brazilian Subjects with Different Skin and Eye Melanin Content*. *1*(2), 1–6.
- Scudder, N., McNevin, D., Kelty, S. F., Walsh, S. J., & Robertson, J. (2018). Forensic DNA phenotyping: Developing a model privacy impact assessment. *Forensic Science International: Genetics*, *34*(June 2017), 222–230. <https://doi.org/10.1016/j.fsigen.2018.03.005>
- Serre, C., Busuttill, V., & Botto, J.-M. (2018). Intrinsic and extrinsic regulation of human skin melanogenesis and pigmentation. *International Journal of Cosmetic Science*, *40*(4), 328–347. <https://doi.org/10.1111/ics.12466>
- Shi, P., Lu, X. M., Luo, H. R., Xiang-Yu, J.-G., & Zhang, Y. P. (2001). Melanocortin-1 receptor gene variants in four Chinese ethnic populations. *Cell Research*, *11*(1), 81–84. <https://doi.org/10.1038/sj.cr.7290070>
- Soejima, M., & Koda, Y. (2006). Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. *International Journal of Legal Medicine*, *121*(1), 36–39. <https://doi.org/10.1007/s00414-006-0112-z>
- Souza, A. M. de, Resende, S. S., Sousa, T. N. de, & Brito, C. F. A. de. (2019). A systematic scoping review of the genetic ancestry of the Brazilian population. *Genetics and Molecular Biology*, *42*(3), 495–508. <https://doi.org/10.1590/1678-4685-gmb-2018-0076>

- Sturm, R. A., & Duffy, D. L. (2012). Human pigmentation genes under environmental selection. *Genome Biology*, 13(9), 248. <https://doi.org/10.1186/gb-2012-13-9-248>
- Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Jakobsdottir, M., & Stefansson, K. (2008). Two newly identified genetic determinants of pigmentation in Europeans. *Nature Genetics*, 40(7), 835–837. <https://doi.org/10.1038/ng.160>
- Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., & Stefansson, K. (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genetics*, 39(12), 1443–1452. <https://doi.org/10.1038/ng.2007.13>
- Valenzuela, R. K., Henderson, M. S., Walsh, M. H., Garrison, N. A., Kelch, J. T., Cohen-Barak, O., & Brilliant, M. H. (2010). Predicting phenotype from genotype: Normal pigmentation. *Journal of Forensic Sciences*. <https://doi.org/10.1111/j.1556-4029.2009.01317.x>
- Valls, J. H. (2012). *Identificación y caracterización de regiones cromosómicas asociadas a melanoma esporádico en el uruguay mediante el método de mapeo por mestizaje : MCI-R como posible gen candidato*. Facultad de Medicina UDELAR.
- Virmond, M. B., Robert, A. W., Brito, P. B., & Massuda, T. Y. C. (2016). Fenotipagem forense pelo DNA através de SNPs. *Revista Brasileira de Criminalística*, 5(2), 37. <https://doi.org/10.15260/rbc.v5i2.128>
- Voisey, J., Gomez-Cabrera, M. D. C., Smit, D. J., Leonard, J. H., Sturm, R. A., & van Daal, A. (2006). A polymorphism in the agouti signalling protein (ASIP) is associated with decreased levels of mRNA. *Pigment Cell Research*, 19(3), 226–231. <https://doi.org/10.1111/j.1600-0749.2006.00301.x>
- Walsh, S., Chaitanya, L., Breslin, K., Muralidharan, C., Bronikowska, A., Pospiech, E., & Kayser, M. (2017). Global skin colour prediction from DNA. *Human Genetics*, 136(7), 847–863. <https://doi.org/10.1007/s00439-017-1808-5>
- Walsh, S., Liu, F., Ballantyne, K. N., van Oven, M., Lao, O., & Kayser, M. (2011). IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics*, 5(3), 170–180. <https://doi.org/10.1016/j.fsigen.2010.02.004>
- Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., & Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics*, 7(1), 98–115. <https://doi.org/10.1016/j.fsigen.2012.07.005>
- Whikehart, D. R. (2003). Nucleic Acids. In *Biochemistry of the Eye* (Vol. 16, pp. 191–229). <https://doi.org/10.1016/B978-0-7506-7152-1.50011-4>
- Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterländer, M., & Burger, J. (2014). Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proceedings of the National Academy of Sciences*, 111(13), 4832–4837. <https://doi.org/10.1073/pnas.1316513111>
- Yun, L., Gu, Y., Rajeevan, H., & Kidd, K. K. (2014). Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations. *International Journal of Legal Medicine*, 128(3), 447–453. <https://doi.org/10.1007/s00414-013-0953-1>
- Zaorska, K., Zawierucha, P., & Nowicki, M. (2019). Prediction of skin color, tanning and freckling from DNA in Polish population: linear regression, random forest and neural network approaches. *Human Genetics*, 138(6), 635–647. <https://doi.org/10.1007/s00439-019-02012-w>
- Zeigler-Johnson, C., Panossian, S., Gueye, S. M., Jalloh, M., Ofori-Adjei, D., & Kanetsky, P. A. (2004). Population Differences in the Frequency of the Agouti Signaling Protein g.8818A&gt;G Polymorphism. *Pigment Cell Research*, 17(2), 185–187. <https://doi.org/10.1111/j.1600-0749.2004.00134.x>