SMGroup

SM Journal of Forensic Research and Criminology

Research Article

Analysis of Eight SNPs in South Brazilian Subjects with Different Skin and Eye Melanin Content

Fernanda Rosa Sawitzki[1], Rodrigo Rodenbusch[1,2], Diego Wordell Gubert[1], Deborah Soares Bispo Santos Silva[1,3], Eduardo Filipe Avila Silva[1,4] and Clarice Sampaio Alho[1]*

[1]*Faculty of Biosciences, Laboratory of Human and Molecular Genetics, PUCRS, Brazil*
[2]*Scientific and Technological Development Center (CDCT), State Foundation of Production and Research in Health (FEPPS), Porto Alegre-RS, Brazil*
[3]*Forensic Sciences Institute, North Carolina State University, USA*
[4]*Department of Federal Police, Porto Alegre-RS, Brazil*

Article Information

Received date: Jun 07, 2017
Accepted date: Aug 14, 2017
Published date: Aug 18, 2017

***Corresponding author**

Clarice Sampaio Alho, Laboratório de Genética Forense, PUCRS. Av. Ipiranga 6681, 90619-900, Porto Alegre, RS, Brazil, Tel: (+55) (51) 3320-3568; Email: csalho@pucrs.br

Keywords Forensic; Human pigmentation; Phenotype; SNaPshot; SNP

Abbreviations AUC: Area under the Receiver Operating Characteristic Curve; DNA: Deoxyribonucleic Acid; EVC: Externally Visible Characteristics; FDA: Factorial Discriminant Analyses; GP: Genetic Probability; HMC: High Melanin Content; LMC: Low Melanin Content; PCR: Polymerase Chain Reaction; PGL: Calculation of Pathway Genetic Load; RGB: Red, Green, Blue; SNP: Single Nucleotide Polymorphism; STR: Short Tandem Repeat

Abstract

LMC (Low Melanin Content) and HMC (High Melanin Content) subjects have respectively low and high melanin content in both skin and eyes; LMC has white skin and blue eyes and HMC has dark skin and eyes. Comparative investigation between frequencies of genetic variants in LMC subjects *versus* HMCsubjects may indicate which polymorphic variant is associated with melanin synthesis in skin and eyes. Coordinately, studies with Snow White-Like (SW) subjects may be informative to reveal any tissue-specific expression, since these individuals have white skin and dark eyes.

The LMC - HMC - SW model was used to analyze the allelic distribution of eight biallelic SNPs in pigment-related genes in admixed South Brazilian individuals. Based on allele frequencies of different human populations, allele "L" was used for the alleles associated with low melanin content populations (LMC subjects), and allele "H" was used for the alleles associated with high melanin content populations (HMC subjects). Allelic distribution of eight SNPs showed that 100% of LMC subjects (N=73) had less than eight H alleles, and 82% of HMC subjects (N=61) had eight or more H alleles. The AUC (Area under the Receiver Operating Characteristic Curve) value was 0.99, and the calculation of PGL (Pathway Genetic Load) and GP (Genetic Probability) showed that the SNP set presented 93% and 91% concordance between DNA genotype and phenotypes, respectively. Factorial Discriminant Analyses (FDA) performed in the SW group (light skin and dark eyes; N=116) showed a positive association between SNPs rs16891982 (SLC45A2), rs8045560 (MC1R), rs1426654 (SLC24A5), rs2733832 (TYRP1) and rs1042602 (TYR), and the LMC cluster for skin phenotype, and a positive association between SNPs rs4778138 (OCA2), rs12913832 (HERC2) and rs916977 (HERC2), and the HMC cluster for eye phenotype. The understanding of gene functionin externally visible characteristics is important for the prediction of skin and eye colors in humans; the analyses presented here are an important contribution to the forensic DNA phenotyping scenario.

Communication

Forensic DNA Phenotyping (FDP) aims to predict appearance traits of a sample donor from DNA evidence left at a crime scene; a good FDP strategy may be especiallyuseful in cases where the police have no other investigative leads, include reference DNA profiling. Skin and eye colors are important Externally Visible Characteristics (EVCs), since they are highly heritable genetic traits and are the most obvious and distinguishable externally visible characteristics to be used in human identification [1]. Hair melanin content is more prone to change by intrinsic or extrinsic factors; it is age-dependent, changing during late childhood, adolescence, adulthood, and old age, and becoming darker, gray or white at any given stage [1]. These changes, added to the occurrence of baldness, make hair a trait less useful to forensic DNA phenotyping applications.

There are hundreds of genes involved in the pigmentation process; however, for forensic purposes, a small number of SNPs have already been proved to be sufficient in the prediction of pigmentation patterns with high confidence [1]. In order to find a set of candidate SNPs able to predict EVCs, we first selected seven human genes linked to melanin synthesis (Figure 1): 1- *OCA2*: encodes the P protein involved in anion transport and in the melanosomal pH regulation, operating in the Tyrosinase (TYR) and the Tyrosinase Related Protein 1 (TYRP1) functions [2-5]. 2- *HERC2*: HERC domain and RCC1-Like domain 2 gene is located 10 Kb upstream of *OCA2* and itacts as a regulatory enhancer of *OCA2* region. The *HERC2* function is still unknown [1,2]. 3- *SLC45A2*: Solute carrier family 45 members 2 gene encodes MATP protein and regulates MATP function, with a crucial role in the trafficking processing and intracellular Tyrosinase (TYR) and protons transporting, controlling the intramelanosomal pH and the activity of Tyrosinase (TYR) [6]. 4- *MC1R*: transcribes the melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor),

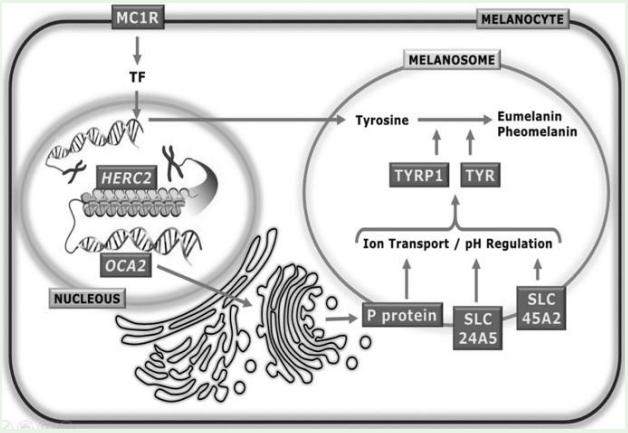

Figure 1: Pigmentation-related genes selected in this study and gene products/function in melanosome and melanocyte structures.

a G-protein located in the melanocyte membrane which is directly responsible for eumelanin/pheomelanin synthesis regulation [7]. 5- *SLC24A5*: solute carrier family 24, member 5, is responsible for accumulating Ca^{+2} in the melanosome [8]. 6- *TYRP1*: encodes the tyrosinase-related protein 1, which is necessary to melanin production [6-9]. 7- *TYR*: transcribes the tyrosinase, a crucial enzyme for the initial melanogenesis process [10]. After gene selection, SNPs were analyzed in each one of the genes using NCBI (National Center for Biotechnology; https://www.ncbi.nlm.nih.gov/snp). A large number of SNPs were identified per gene (*OCA2*: 44456; *HERC2*: 24579; *SLC45A2*: 5349; *MC1R*: 1238; *SLC24A5*: 2751; *TYRP1*: 3020; *TYR*: 13736) and SNP selection was done based on the following criteria: biallelic polymorphisms (i.e. no single insertion/deletion variants), validation by 1000G (One Thousand Genomes Project; http://www.internationalgenome.org) and by HapMap (International Haplotype Map Project; http://www.hapmap.ncbi.nlm.nih.gov), and with global **Minor Allele Frequency (MAF)** distinguishable as a common polymorphism (i.e. with no rare variant).

Next, we examined the allelic frequencies data of each selected SNP in different human populations using ALFRED (ALlele FREquency Database; https://alfred.med.yale.edu). In this step, we selected from ALFRED two sets of samples from extremely distinguishable subjects according to the melanin content in skin and eyes (subjects with low melanin content *versus* subjects with high melanin content), and discerned the allelic frequencies of each SNP in these two sets of samples. Thus, we selected less than 20 pre-candidate SNPs that independently could be able to distinguish those two sets of subjects (LMC *versus* HMC) by allelic frequencies. Simultaneously, we analyzed published studies with significant data on genotype-phenotype association. Finally, eight SNPs in seven genes were chosen. At each locus, allele L was used for each allele strongly associated with LMC subjects, and allele H for each allele strongly associated with HMC subjects. Table 1 presents the selected SNPs and their allele frequencies based on the populations frequencies available on ALFRED, and also data on sensibility, specificity, predictive values, relative ratio, and odds ratio (the database used to construct Table 1 is available in the supplemental file). These loci were tested in previously published genotype-phenotype association studies (Box 1). Since the eight SNPs in pigmentation-related genes presented remarkable results in all of the parameters, they were considered robust markers for color prediction.

1- HMC (High Melanin Content) subjects from populations of Africans, as available on ALFRED websites. 2- LMC (Low Melanin Content) subjects from populations of Europeans, as available on

BOX 1: Eigth locus in seven pigmentation-related genes that were previously tested in genotype-phenotype association studies.

- ***OCA2*:** The rs4778138 is a G>A SNP in *OCA2* intron region strongly associated with eye color [6].
- ***HERC2*:** The rs12913832 is an A>G SNP in *HERC2*intronic area, within *OCA2* enhancer region. The derived allele G of this SNP is associated with blue iris phenotype, being common in Europeans, particularly those of northwestern and eastern European descent [4]. The allele A (or allele T, if in the complementary DNA strand) allows chromatin opening and the *OCA2* transcription factors recruitment, which leads to darker iris, while the allele G (or allele C, if in complementary DNA strand) maintains the chromatin closed, being less effective in recruiting the aforementioned factors, which results in lighter iris [4]. The derived allele G of rs916977, other A>G SNP, in about 10 Kb far from rs12913832, has been associated with blue eyes [1].
- ***SLC45A2***: The rs16891982 is the 1122C>G SNP in exon 5 of the *SLC45A2* gene. This SNP results in the non-synonymous substitution of Leucine (allele C; codon TTG on the coding/reverse DNA strand) by Phenylalanine (allele G; codon TTC on the coding/reverse DNA strand) at amino acid 374 (Leu374Phe). Rs16891982 is associated with human skin pigmentation normal variation, since it regulates the MATP function; the Leu374 variant (allele C – complementary to G on codon TTG on the coding/reverse DNA strand) plays an important role in the transport of protons, resulting in an optimum intramelanosomal pH, which allows the activity of tyrosinase (TYR) and, consequently, the adequate production of eumelanin (brow-black spectrum), while the derived Phe374 variant (allele G – complementary to C on codon TTC on the coding/reverse DNA strand) may change the transport, the pH and the synthesis of the pigment [7]. A strong association between rs16891982 and eye color has been reported [1].
- ***MC1R***: The ancestral allele C of rs8045560 (C>T SNP) is associated with darker skin populations [7].
- ***SLC24A5*:** The rs1426654 (G>A SNP in the coding region of *SLC24A5* gene) results in non-synonymous substitution of Alanine (GCA) by Threonine (ACA) at amino acid 111 Ala111Thr. This SNP has shown evidence of natural selection; the derived allele A (Thr111) is associated with light skin pigmentation and is common in Europe, southwest Asia, and central Asia [1]. The ancestral allele (G allele; Ala111) is associated with darker skin [8].
- ***TYRP1***: The ancestral allele C of rs2733832 (C>T SNP) is associated with darker skin populations [6].
- ***TYR***: The rs1042602 (C>A SNP) results in the non-synonymous substitution of Serine (TCT) by Tyrosine (TAT) at amino acid 192 (Ser192Tyr) and in a reduction of about 40% in the catalytic activity of tyrosinase [10].

Table 1: Allele frequencies and other parameters of the selected SNPs based on population data from ALFRED.

RS (Gene; SNP)		HMC Populations[1]		LMC Populations[2]		P Value	Sensit	Specif	PV-1	PV-2	RR	OR
rs4778138 (*OCA2;* SNP G>A)	**Allele H (G)**	0.745	(1135/1524)	0.164	(957/5826)	<0.0001	0.744	0.836	0.542	0.926	7.31	14.81
	Allele L (A)	0.256	(390/1524)	0.836	(4870/5826)							
rs12913832 (*HERC2*; SNP A>G)	**Allele H (A)**	0.957	(4484/4688)	0.285	(8155/28286)	<0.0001	0.957	0.712	0.355	0.990	35.36	54.26
	Allele L (G)	0.043	(204/4688)	0.712	(20132/28286)							
rs16891982 (*SLC452A;* SNP C>G)	**Allele H (C)**	0.881	(3592/4076)	0.067	(1850/27594)	<0.0001	0.881	0.933	0.660	0.982	35.77	103.28
	Allele L (G)	0.119	(484/4076)	0.933	(25744/27594)							
rs8045560 (*MC1R;* SNP C>T)	**Allele H (C)**	0.915	(1440/1574)	0.440	(1730/3932)	<0.0001	0.915	0.560	0.454	0.943	7.92	13.68
	Allele L (T)	0.085	(134/1574)	0.560	(2202/3932)							
rs1426654 (*SLC24A5;* SNP G>A)	**Allele H (G)**	0.725	(2890/3984)	0.011	(52/4502)	<0.0001	0.725	0.988	0.982	0.803	4.98	226.12
	Allele L (A)	0.275	(1094/3984)	0.989	(4451/4502)							
rs2733832 (*TYR*P1; SNP C>T)	**Allele H (C)**	0.917	(396/432)	0.415	(537/1292)	<0.0001	0.917	0.584	0.424	0.954	9.33	15.47
	Allele L (T)	0.083	(36/432)	0.585	(755/1292)							
rs1042602 (*TYR;* SNP C>A)	**Allele H (C)**	0.946	(1982/2096)	0.702	(3614/5146)	<0.0001	0.946	0.298	0.354	0.930	5.11	7.37
	Allele L (A)	0.054	(114/2096)	0.298	(1532/5146)							
rs916977 (*HERC2;* SNP A>G)	**Allele H (A)**	0.890	(1116/1254)	0.214	(778/3638)	<0.0001	0.890	0.786	0.589	0.954	12.80	29.73
	Allele L (G)	0.110	(138/1254)	0.786	(2860/3638)							

ALFRED websites. 'Allele H': allele strongly associated with HMC people. 'Allele L': allele strongly associated with LMC people. Sensit: sensitivity. Specif: specificity. PV-1: predictive value for the presence of allele H in HMC populations. PV-2: Predictive Value for the presence of allele L in LMC populations. RR: Relative Ratio. OR: Odds Ratio.

To amplify the flanking SNP regions by multiplex PCR in a SNaPshot® System ABI Prism (Applied Biosystems®), primer sequences were designed for each locus for conventional PCR and for the SNapShot multiplex system (Table 2). Labeled products were separated in an ABIPRISM®3130xl Genetic Analyzer (Applied Biosystems®) and analyzed in GeneMapper ID software version 3.2.1 (Applied Biosystems®).

In this study, first we analyzed 134 southern Brazilian subjects belonging to categories: LMC subjects (white skin and blue eyes; N = 73) and HMC subjects (dark skin and eyes; N = 61).Each skin color participant of was identified using Fitzpatrick score and the amount of red (R), green (G), and blue (B) pigments in an inner and hairless portion (below elbow) of the right arm using the *color analyzer*ACR-1023 (Instrutherm, São Paulo, Brazil). Each eye

Table 2: Primer sequences per locus for conventional PCR and for the SNapShot multiplex system.

RS (Gene; SNP)	Primer Sequence
rs916977 (*HERC2;* SNP A>G)	F-5'- ttgttttggcaaactccaca-3' / R-5'- ttttccaattgccctgacat-3' (PCR Fragment: 200pb)
SNapShot (35pb):	ggcaaactccacagtggggatgcagtttgagtaga**T/C**
rs4778138 (*OCA2;* SNP G>A)	F-5'- ggattcaaaaagaaagtctcaagg-3' / R-5'- gctctccttttgataccagca-3' (PCR Fragment: 146bp)
SNapShot (42pb):	tattgaactgaatgaaagtgaaaatataacatatcaaaattg**G/A**
rs12913832 (*HERC2*; SNP A>G)	F-5'- cagctccatcaatgtggtca-3' / R-5'- ctgatgatgatagcgtgcaga-3' (PCR Fragment: 253pb)
SNapShot (56pb):	ttcatggctctctgtgtctgatccaagaggcgaggccagtttcatttgagcattaa**A/G**
rs16891982 (*SLC452A;* SNP C>G)	F-5'- ccaagttgtgctagaccagaaa-3' / R-5'- cctcaacagcctccaatctc-3' (PCR Fragment: 211pb)
SNapShot (64pb):	aagacatccttaggagagagaaagacttacaagaataaagtgaggaaaacacggagttgatgca**C/G**
rs8045560 (*MC1R;* SNP C>T)	F-5'- aacgatgtttgtggtcagca-3' / R-5'- actcaaggcatctggaatgg-3' (PCR Fragment: 289pb)
SNapShot (73pb):	cacccacccctttttccatgggggatctgcactcatctccagggaagatggttgggagataaccccagtctgc**C/T**
rs1426654 (*SLC24A5;* SNP G>A)	F-5'- gccttccctcacccttcta-3' / R-5'- aggatggtgctaatgccaat-3' (PCR Fragment: 425pb)
SNapShot (90pb):	tagttgtaaagacatactctttcactttattaggcataacaatcatttcatttatgttcagcccttggattgtctcaggatgttgcaggc**G/A**
rs2733832 (*TYR*P1; SNP C>T)	F-5'- atgaccctgctgttcgaagt-3' / R-5'- cttcttgcctgcattttcaa-3' (PCR Fragment: 334pb)
SNapShot (98pb):	ccaaatgatcctatttttgtcctcctgcacaccttcacagatgcagtctttgatgaatggctgaggagatacaatgctggtaagacattttcatatgc**C/T**
rs1042602 (*TYR;* SNP C>A)	F-5'- actgcaagtttggctttgg-3' / R-5'- gcttcatgggcaaaatcaat-3' (PCR Fragment: 307pb)
SNapShot (106pb):	ggccaaatgaaaaatggatcaacacccatgtttaacgacatcaatatttatgacctctttgtctggatgcattattatgtgtcaatggatgcactgcttgggggat**C/A**

Table 3: Phenotype data from 134 LMC and HMC Southern Brazilian subjects.

		LMC	HMC
Total	N (%)	73 (54.82)	61 (45.18)
Sex	Male [N (%)]	40 (54.80)	32 (52.46)
	Female [N (%)]	33 (45.20)	29 (47.54)
Age	Years [Mean (SD)]	36.6 (13.6)	30.4 (8.7)
Skin Fitzpatrick Score	Type 1 [N (%)]	30 (41.09)	0
	Type 2 [N (%)]	43 (58.91)	0
	Type 5 [N (%)]	0	26 (42.62)
	Type 6 [N (%)]	0	35 (57.38)
Skin R-G-B Color	Red [Mean (SD)]	88.4 (13.1)	49.1 (12.9)
	Green [Mean (SD)]	70.3 (13.2)	35.3 (10.2)
	Blue [Mean (SD)]	61.8 (13.8)	27.9 (8.2)
Eye Type	Light blue [N (%)]	50 (68.49)	0
	Dark blue [N (%)]	23 (31.51)	0
	Dark brown [N (%)]	0	39 (63.94)
	Black [N (%)]	0	22 (36.06)
Eye R-G-B Color	Red [Mean (SD)]	99.4 (12.5)	59.2 (19.1)
	Green [Mean (SD)]	95.9 (12.5)	32.6 (11.7)
	Blue [Mean (SD)]	137.9 (25.2)	42.8 (14.8)
Auto Declared Origin	European	66 (90.41	13 (21.31)
	African	0	26 (37.70)
	Amerindian	0	2 (0.03)
	Mixed	4 (5.48)	10 (16.39)
	Unknown	3 (4.11)	10 (16.39)

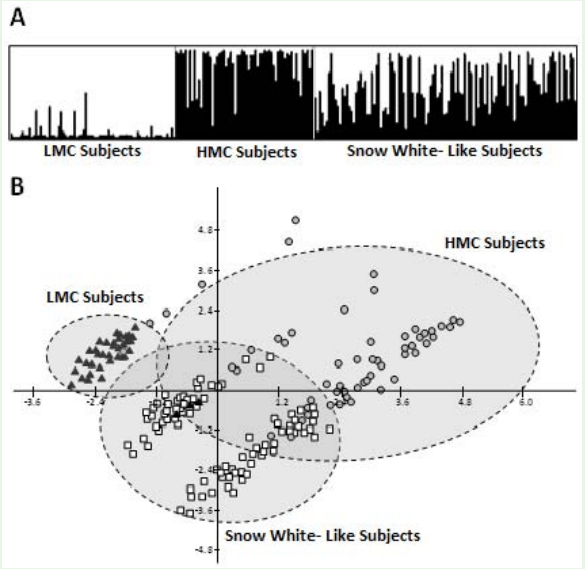

Figure 2: **(A)** Genotype structure distribution of allele H and allele L in LMC (triangle), HMC (circle), and Snow White-Like (square) subjects from South of Brazil; *K*=2 clusters using *Structure Software*. **(B)** Factorial Discriminant Analysis comparison among LMC, HMC, and Snow White-Like genotypes from South of Brazil to SNPs *OCA2* rs4778138, *HERC2* rs12913832, *HERC2* rs12913832 *SLC45A2* rs16891982, *MC1R* rs8045560, *SLC24A5* rs1426654, *TYRP1* rs2733832, *TYR* rs1042602 using PAST 3.05 (http://folk.uio.no/ohammer/past).

color participant was classified and photographed and RGB values weremeasured in both eyes by COLOURS software [https://www.colours-software-pvt-ltd]. Phenotype data of these subjects are presented in table 3.

Table 4: Allele frequencies and other parameters of the selected SNPs in 134 South Brazilian subjects with HMC (N=61), or LMC (n=73) patterns.

RS (Gene; SNP)		HMC SB People[1]		LMC SB Peolple[2]		P Value	Sensit	Specif	PV-1	PV-2	RR	OR
rs4778138 (*OCA2;* SNP G>A)	**Allele H (G)**	0.607	(74/122)	0.075	(11/146)	<0.0001	0.607	0.892	0.871	0.655	2.52	12.75
	Allele L (A)	0.393	(48/122)	0.925	(135/146)							
rs12913832 (*HERC2*; SNP A>G)	**Allele H (A)**	0.828	(101/122)	0.021	(3/146)	<0.0001	0.828	0.979	0.971	0.872	7.58	229.25
	Allele L (G)	0.172	(21/122)	0.979	(143/146)							
rs16891982 (*SLC452A;* SNP C>G)	**Allele H (C)**	0.279	(34/122)	0.021	(3/146)	<0.0001	0.279	0.980	0.919	0.622	2.43	18.67
	Allele L (G)	0.721	(88/122)	0.993	(145/146)							
rs8045560 (*MC1R;* SNP C>T)	**Allele H (C)**	0.746	(91/122)	0.390	(57/146)	<0.0001	0.746	0.612	0.615	0.744	2.40	4.63
	Allele L (T)	0.254	(31/122)	0.616	(90/146)							
rs1426654 (*SLC24A5;* SNP G>A)	**Allele H (G)**	0.672	(82/122)	0.000	(0/146)	<0.0001	0.672	1.000	1.000	0.785	4.65	<1000.0
	Allele L (A)	0.328	(40/122)	1.000	(146/146)							
rs2733832 (*TYR*P1; SNP C>T)	**Allele H (C)**	0.770	(94/122)	0.479	(70/146)	<0.0001	0.770	0.521	0.573	0.731	2.13	3.64
	Allele L (T)	0.230	(28/122)	0.521	(76/146)							
rs1042602 (*TYR;* SNP C>A)	**Allele H (C)**	0.869	(106/122)	0.623	(91/146)	<0.0001	0.869	0.377	0.538	0.775	2.39	4.00
	Allele L (A)	0.131	(16/122)	0.377	(55/146)							
rs916977 (*HERC2;* SNP A>G)	**Allele H (A)**	0.697	(85/122)	0.027	(4/146)	<0.0001	0.697	0.973	0.955	0.794	4.65	82.13
	Allele L (G)	0.303	(37/122)	0.979	(143/146)							

Table 5: Genotype frequencies and Chi-Square Test result of High Melanin Content (HMC), Low Melanin Content (LMC), HMC Subjects, LMC Subjects, and Snow White-Like (SW-L) subjects according HH (homozygote to allele H), LL (homozygote to allele L), and HL (heterozygote) per locus.

RS (Gene; SNP)		HMC	LMC	HMC Subj	LMC Subj	SW Subj
rs4778138 (*OCA2;* SNP G>A)	**HH**	0.55	0.03	0.36	0.00	0.15
	HL	0.38	0.27	0.49	0.15	0.54
	LL	0.06	0.70	0.15	0.85	0.31
	N	762	2913	61	73	116
	P			>0.0001	>0.0001	0.53
rs12913832 (*HERC2*; SNP A>G)	**HH**	0.92	0.08	0.75	0.00	0.47
	HL	0.08	0.40	0.15	0.04	0.00
	LL	0.00	0.52	0.10	0.96	116
	N	2344	14143	61	73	
	P			**0.262**	<0.0001	
rs16891982(*SLC452A;* SNP C>G)	**HH**	0.84	0.19	0.56	O.15	0.19
	HL	0.15	0.49	0.38	0.48	0.59
	LL	0.01	0.32	0.06	0.37	0.22
	N	787	1966	61	73	116
	P			>0.0001	**0.120**	
rs8045560 (*MC1R;* SNP C>T)	**HH**	0.84	0.00	0.05	0.15	0.19
	HL	0.15	0.13	0.46	0.48	0.59
	LL	0.01	0.87	0.49	0.37	0.22
	N	2038	13797	61	73	116
	P			>0.0001	**0.102**	
rs1426654 (*SLC24A5;* SNP G>A)	**HH**	0.52	0.00	0.46	0.00	0.00
	HL	0.40	0.02	0.43	0.00	0.20
	LL	0.08	0.98	0.11	1.00	0.80
	N			61	73	116
	P			>0.0001	**0.178***	
rs2733832 (*TYRP1*; SNP C>T)	**HH**	0.84	0.17	0.58	0.22	0.25
	HL	0.15	0.49	0.39	0.52	0.55
	LL	0.01	0.34	0.03	0.26	0.20
	N	216	646	61	73	116
	P			>0.0001	**0.650**	
rs1042602 (*TYR;* SNP C>A)	**HH**	0.89	0.49	0.77	0.37	0.35
	HL	0.10	0.42	0.20	0.51	0.49
	LL	0.01	0.09	0.03	0.12	0.16
	N	1048	2573	61	73	116
	P			>0.0001	**0.845**	
rs916977 (*HERC2;* SNP A>G)	**HH**	0.79	0.04	0.49	0.00	0.24
	HL	0.20	0.34	0.41	0.05	0.47
	LL	0.01	0.62	0.10	0.95	0.29
	N	627	1819	61	73	116
	P			>0.0001	>0.0001	

*The association chi-square test was calculated with the allele data.

There was a highly significant association between alleles H and HMC Brazilian subjects, and alleles L and LMCsubjects. All other parameters (sensibility, specificity, predictive values, relative ratio, odds ratio) were equally remarkable (Table 4). To observe the cumulative distribution of alleles H and alleles L in the two phenotypically different color groups, we plotted the frequency data and noticed LMC subjects presented between zero and seven alleles H, while HMC subjects presented between four and 14 alleles H. No subject had 15 or 16 alleles H; the average of alleles H among the LMCsubjects was 3.3 (SD=1.49), 96% (70/73) of them had less than five alleles H, and 100% had seven or less alleles H. The average of alleles H among the HMCsubjectswas 12.0 (SD=3.78), 82% (56/61) of them had eight or more alleles H. We also measured the ability of the SNP panel correctly classify those with and without the phenotypic trait (pigmentation) and obtained very high AUC (area under the receiver operating characteristic curve) value of 0.9908 in the prediction of both HMC and LMC phenotypes (SE= 0.0088; IC 95% = 0.99736-1.0). The weighted Pathway Genetic Load (PGL) scores, which tested multiple loci cumulative effect on the phenotype [11], showed HMC subjects ranged from 65.29 to 870.92 (median= 567.37; mean= 584.63; SD=262.46) and LMC from zero to 148.40 (median= 38.31; mean= 44.80; SD=30.93). There was a significant difference between the two groups (P<0.0001; Mann-Whitney Test). Genotype Probability (GP) in all 134 South Brazilian subjects showed that 80% (49/61) of HMC genotypes had the highest chance to belong to HMC group (rates from 1.95 to $8.27x10^7$; mean= 7.92×10^6), and 100% (73/73) of LMC genotypes had the highest chance to belong to the LMC group (rates from $1.46x10^4$ to $7.77x10^{12}$; mean= 2.02×10^{11}). This GP analysis using the current eight SNPs genotyping system showed 91% (122/134) concordance between predicted and observed phenotype (methodology details are available on the supplemental file).

1- HMCSB People: South Brazilian subjects with dark skin and eyes. 2- LMCSB People: South Brazilian subjects with Llight skin and blue eyes. We called as allele H for the allele strongly associated with people with high melanin content (HMC; from African populations, see table 1), and as allele L for the allele strongly associated with people with low melanin content (LMC; from European populations, see Table 1). Sensit: sensitifity. Specif: specificity. PV-1: predictive value to hazard of HMC people to have allele H. PV-2: predictive value to hazard of LMC people to have allele L. RR: Relative Ratio. OR: Odds Ratio.

It is still unclear when a pigment-related gene variant is expressed in both skin and iris, or when it has tissue-specific expression. In order to identify differences on genotype-phenotype associations, we studied 116 Snow White-Like subjects (SW – light skin and dark eyes) from South Brazil. Genotype data from HMC, LMC and SW subjects were verified by structuring [Suppl 4] and Factorial Discriminant Analysis (FDA). Snow White-Like subjects were grouped in an intermediate cluster, as expected since they have light skin as LMC subjects and dark eyes as HMCsubjects (Figure 2).

Association Chi-square test was performed comparing SW genotypes with genotype distribution in ALFRED-HMC, ALFRED-LMC, HMC, and LMC Southern Brazilian samples. There were no significant differences between SW genotype frequencies and HMC data in *HERC2* rs12913832 (both have dark eyes), and there were no significant differences between SW genotype frequencies and LMC

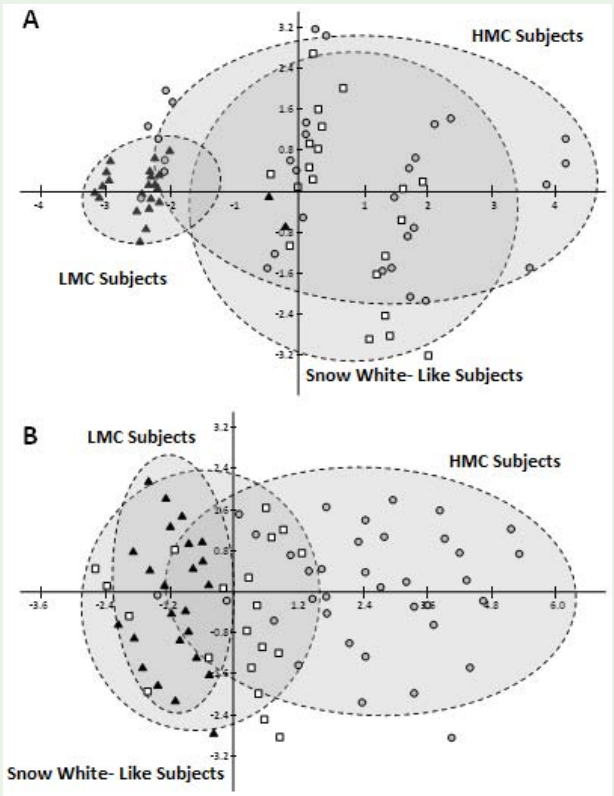

Figure 3: Factorial Discriminant Analysis comparison among LMC (triangle), HMC (circle), and Snow White-Like (square) genotypes using PAST 3.05 (http://folk.uio.no/ohammer/past) to SNPs **(A)** *OCA2* rs4778138, *HERC2* rs12913832, *HERC2* rs12913832 and **(B)** *SLC45A2* rs16891982, *MC1R* rs8045560, *SLC24A5* rs1426654, *TYRP1* rs2733832, *TYR* rs1042602.

data in *SLC45A2* rs16891982, *MC1R* rs8045560, *SLC24A5* rs1426654, *TYRP1* rs2733832, *TYR* rs1042602 (both have light skin) (Table 5).

Based on these results, two definitive FDA were performed using two separate clusters of loci to group the ones able to express melanin in iris (*HERC2* rs12913832, *OCA2* rs4778138, *HERC2* rs916977, we added *OCA2* and *HERC2* SNPs since they are strongly linked to rs12913832), and the ones able to express melanin in skin (*SLC45A2* rs16891982, *MC1R* rs8045560, *SLC24A5* rs1426654, *TYRP1* rs2733832, and *TYR* rs1042602) (Figure 3). SNPs *HERC2* rs12913832, *OCA2* rs4778138, *HERC2* rs916977 clustered 90% of SW subjects with HMC group and *SLC45A2* rs16891982, *MC1R* rs8045560, *SLC24A5* rs1426654, *TYRP1* rs2733832, *TYR* rs1042602 clustered 82% of SW subjects with LMC group.

Conclusion

TheLMC - HMC - SW model was used to analyze eight biallelic SNPs in seven pigment-related genes in an admixture South Brazilian samples. Data revealed that SNPs rs16891982 (*SLC45A2*), rs8045560 (*MC1R*), rs1426654 (*SLC24A5*), rs2733832 (*TYRP1*), rs1042602 (*TYR*) were associated simultaneously with LMC and Snow White clusters, indicating a link of these SNPs with skin phenotype. SNPs rs4778138 (*OCA2*), rs12913832 (*HERC2*), rs916977 (*HERC2*) were associated with HMC and Snow White clusters, demonstrating a link with eye phenotype. The understanding of gene tissue-specific expression in the externally visible characteristics tissue-specific is important for the prediction of skin and eye color in humans; we believe our analysis is an important contribution to the forensic DNA phenotyping scenario.

Supplemental File

Methodology details are available on the Supplemental File.

Acknowledgments

This work was supported by CAPES Brazil | **Edital 25/2014 | Pró-Forenses, and CNPqBrasil | Edital INCT**. Fernanda Rosa Sawitzki, Rodrigo Rodenbusch, Diego WordellGubert, Eduardo Felipe Ávila da Silva and Deborah Soares Bispo Santos Silva weresupportedby CAPES. The authors would like to thank to Mayara Jorgens Prado, Bruna Schroeder, PietraGraebin and Eduardo Wildemann Capelesso for collection and organization of phenotype data.

References

1. Kayser M. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet. 2015; 18: 33-48.
2. Sturm RA, Duffy DL, Zhao ZZ, Leite FPN, Stark MS, Hayward NK, et al. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet. 2008; 82: 424-431.
3. Kondo T, Hearing VJ. Update on the regulation of mammalian melanocyte function and skin pigmentation. Expert Rev Dermatol. 2011; 6: 97-108.
4. Visser M, Kayser M, Palstra RJ. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res. 2012; 22: 446-455.
5. Valenzuela RK, Ito S, Wakamatsu K, Brilliant MH. Prediction model validation: normal human pigmentation variation. J Forensic Res. 2011; 2: 139.
6. Liu F, Wenb B, Kayser M. Colorful DNA polymorphisms in humans. Semin Cell DevBiol. 2013; 24: 562-575.
7. Gerstenblith MR, Goldstein AM, Fargnoli MC, Peris K, Landi MT. Comprehensive evaluation of allele frequency differences of MC1R variants across populations. Human Mutat. 2007; 28: 495-505.
8. Vierkötter A, Krämer U, Sugiri D, Morita A, Yamamoto A, Kaneko N, et al. Development of lentigines in German and Japanese women correlates with variants in the SLC45A2 gene. J Invest Dermatol. 2012; 132: 733-736.
9. Alaluf S, Atkins D, Barrett K, Carter N. Ethnic variation in tyrosinase and TYRP1 expression in photoexposed and photoprotected human skin. Pigment Cell Res. 2003; 16: 35-42.
10. Chaki M, Sengupta M, Mondal M, Bhattacharya A, Mallick S, Bhadra R, et al. Molecular and functional studies of tyrosinase variants among Indian oculocutaneous albinism type 1 patients. J Invest Dermatol. 2011; 131: 260-262.
11. Huebinger RM, Garner HR, Barber RC. Pathway genetic load allows simultaneous evaluation of multiple genetic associations. Burns. 2010; 36: 787-792.