UNIVERSIDADE DE SÃO PAULO FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DEPARTAMENTO DE GENÉTICA

GUILHERME DEBORTOLI

VARIANTES NOS GENES *OCA2* E *HERC2* ASSOCIADAS A FENÓTIPOS CLÁSSICOS DE PIGMENTAÇÃO E ESTRUTURAS SECUNDÁRIAS PRESENTES NA ÍRIS EM AMOSTRA MISCIGENADA DA POPULAÇÃO BRASILEIRA

Ribeirão Preto

2018

GUILHERME DEBORTOLI

VARIANTES NOS GENES *OCA2* E *HERC2* ASSOCIADAS A FENÓTIPOS CLÁSSICOS DE PIGMENTAÇÃO E ESTRUTURAS SECUNDÁRIAS PRESENTES NA ÍRIS EM AMOSTRA MISCIGENADA DA POPULAÇÃO BRASILEIRA

Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, como requisito parcial para a obtenção do título de Doutor em Ciências, área de concentração: Genética

Orientador: Prof. Dr. Celso Teixeira Mendes Junior

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Debortoli, Guilherme

Variantes nos genes *OCA2* e *HERC2* associadas a fenótipos clássicos de pigmentação e estruturas secundárias presentes na íris em amostra miscigenada da população brasileira. Ribeirão Preto, 2018.

202 p. : 28 il. ; 30 cm

Tese de Doutorado, apresentada à Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Área de concentração: Genética.

Orientador: Mendes-Junior, Celso Teixeira.

1. *OCA2.* 2. *HERC2.* 3. Pigmentação humana. 4. Estruturas secundárias da íris. 5. Sequenciamento de nova geração

Dedico este trabalho a meus pais Jorge e Lioni. Meus exemplos para a vida toda. Obrigado pela compreensão, amor e apoio incondicional.

AGRADECIMENTOS

Ao orientador Prof. Dr. Celso Teixeira Mendes Junior, a minha eterna gratidão. Obrigado pela oportunidade de desenvolver este trabalho e por todas as discussões e ensinamentos ao longo destes anos. Guardarei com carinho todos os momentos de aprendizado, procurando sempre espelhar-me na imagem que tenho de você. O de excelente professor, cientista e acima de tudo ser humano.

Ao Prof. Dr. Esteban J. Parra pela supervisão durante o período sanduíche ao qual me recebeu tão gentilmente em seu laboratório. Obrigado pelas valiosas contribuições, e pelas discussões e ensinamentos que certamente me auxiliaram na confecção deste trabalho.

Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por ter concedido a bolsa regular de Doutorado e a bolsa do período sanduíche através do Programa de Doutorado-sanduíche no Exterior (PDSE).

Aos voluntários que gentilmente doaram suas amostras para a realização deste trabalho, estendendo o agradecimento aos funcionários do HEMOCENTRO de Ribeirão Preto pelo auxílio nas coletas.

À Melissa Edwards e David Cha, que gentilmente me auxiliaram nas análises durante o período sanduíche no Canadá.

Ao Prof. Dr. Eduardo Antônio Donadi por fornecer a infraestrutura para a realização deste trabalho e pertinentes contribuições na banca de qualificação. À Juliana, Sandra e Flávia, pelo auxílio nos experimentos.

Aos Professores Dra. Andrea Rita Marrero e Dr. Wilson Araújo da Silva Junior pelas contribuições durante a qualificação do Doutorado.

Ao Prof. Dr. Aguinaldo Luiz Simões, por ceder a infraestrutura e acesso ao Bioanalyzer, e à Cláudia, Elisabete, Rubens e Ana Lúcia pelo auxílio nas coletas e demais procedimentos importantes para a realização deste trabalho.

Ao Prof. Dr. Erick da Cruz Castelli, pela ajuda nas análises de bioinformática e confecção dos scripts utilizados neste trabalho.

Ao Departamento de Genética da FMRP-USP pela oportunidade da realização do Doutorado e a todos os professores, funcionários e colegas da pós-graduação.

À banca examinadora por cordialmente aceitarem em participar na contribuição deste trabalho, através de suas considerações.

Aos meus queridos pais, Jorge e Lioni, irmãos Fabiano e Rodrigo, cunhadas Cheila (sobrinho Diogo ainda por vir) e Priscila, sogros Gercino e Maria, pelo apoio e amor incondicional. Este trabalho tem um pouco de cada um de vocês. Me sinto agraciado por têlos sempre comigo. Mesmo ausente em corpo presente, sei que todos vocês estiveram comigo em pensamento ao longo desta jornada. Amo vocês. À Bruna, por simplesmente ser o meu alicerce ao longo de todos esses anos. Por ter me acompanhado neste caminho e sempre ter estado lá quando eu mais precisei. Obrigado pelo carinho, amor, ensinamentos, compreensão, e fazer da minha vida junto com o Ninho, mais doce.

Aos amigos do Laboratório de Pesquisas Forenses e Genômicas, Nádia, Malu, Alison, Letícia, Thássia, Amanda e Guilherme. Não poderia ter desejado melhor grupo de amigos para se trabalhar. Deixo meus votos de sucesso para todos, e que nossa amizade perdure por muito tempo. Levo cada um de vocês na minha memória com muito carinho.

Aos amigos que fiz no Canadá e tive o prazer de conhecer, Anabela, Carolina, Rishi, Aaron, Bryan e Lida. Obrigado por me receberem tão bem, me fazendo sentir-se acolhido em um lugar totalmente novo.

À todos que de alguma forma contribuíram para que este sonho pudesse ser realizado. Obrigado.

"Se eu vi mais longe, foi por estar sobre ombros de gigantes"

Isaac Newton, 1676

RESUMO

Debortoli G. Variantes nos genes *OCA2* e *HERC2* associadas a fenótipos clássicos de pigmentação e estruturas secundárias presentes na íris em amostra miscigenada da população brasileira [tese]. Ribeirão Preto: Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto; 2018; 214p.

A pigmentação dos olhos, cabelos e pele, bem como presença ou ausência de sardas, está entre os exemplos mais visíveis da variação fenotípica humana. O estudo da diversidade genética em genes de pigmentação tem beneficiado diferentes áreas do conhecimento, como a área da genética e antropologia forense, bem como a área relacionada a saúde e bemestar. Adicionalmente, a presença de estruturas secundárias na íris tem sido reportada como importante fator na percepção de cor de olho observada que um indivíduo pode ter referente a íris e também a fatores de risco para algumas doenças oculares, ainda que as bases genéticas envolvidas nestas características sejam pouco conhecidas. Os genes OCA2 e HERC2 representam dois genes associados à variação normal da pigmentação. Este trabalho avaliou a relação de polimorfismos nas regiões regulatórias e codificantes destes dois genes com os fenótipos de pigmentação e estruturas secundárias presentes na íris encontrados em uma amostra populacional de 340 indivíduos do estado de São Paulo, por meio de sequenciamento de nova geração. Análises de regressão logística e linear para as variáveis qualitativas e quantitativas da cor dos olhos e estruturas secundárias presentes na íris foram realizadas. 170 pontos de variação ao longo das regiões estudadas foram identificados, dos quais 18 estão associadas a pelo menos um fenótipo de pigmentação e estruturas secundárias presentes na íris. Destaca-se a existência de muitos polimorfismos que não se mostrara-se associados quando avaliados independentemente, porém foram associados quando analisados sob a ótica de interações epistáticas, considerada uma possível explicação para a variabilidade encontrada nestes fenótipos, principalmente aqueles intermediários, como a cor dos olhos verdes e mel. O uso de variáveis quantitativas para os olhos revelou pela primeira vez a associação do polimorfismo não sinônimo rs201872292 no gene HERC2 com olhos claros, independente do efeito do polimorfismo rs12913832. Ainda, a associação do polimorfismo rs58358300 localizado em um íntron do gene HERC2 com pigmentação da esclera, o que representa a primeira vez que um polimorfismo é associado a esta característica. Este foi o primeiro estudo no Brasil que se propôs a analisar polimorfismos genéticos em genes candidatos à variação normal da pigmentação humana com estruturas secundárias presentes na íris. Os resultados confirmam a hipótese de que polimorfismos dos genes OCA2 e HERC2 podem contribuir para a formação dos fenótipos clássicos de pigmentação de olhos, pele, cabelos e estruturas secundárias presentes na íris humana dos indivíduos da população brasileira.

Palavras-chave: *OCA2. HERC2.* Pigmentação humana. Estruturas secundárias da íris. Epistasia. Sequenciamento de nova geração.

ABSTRACT

Debortoli G. Variants within *OCA2* and *HERC2* genes associated with classical pigmentation phenotypes and iris features in Brazilian admixed population sample. [thesis]. Ribeirão Preto: University of São Paulo, Ribeirão Preto Medical School; 2018; 214p.

The pigmentation of the eyes, hair and skin, as well as the presence or absence of freckles, are amongst the most visible examples of human phenotypic variation. The study of genetic diversity in pigmentation genes has contributed greatly to the fields of forensics genetics, anthropological genetics and public health. In addition, the presence of iris features has been reported to influence the perception of overall iris color and also consists in risk factors for ocular diseases, although very little is known about the genetic basis of these traits. The OCA2 and HERC2 genes have been associated with normal variation of pigmentation in diverse populations. The present study evaluated the relationship of polymorphisms in the regulatory and coding regions of these two genes with the pigmentation phenotypes and iris features found in a population sample of 340 individuals from the state of São Paulo, Brazil, through next-generation sequencing. Logistic and linear regression analyzes for the qualitative and quantitative variables were performed. A total of 170 points of variation throughout the studied regions were identified, of which 18 were associated with at least one pigmentation phenotype when analyzed as qualitative and/or quantitative variables and iris features. It is worth mentioning that many associations that were not observed when evaluated independently, were indeed associated when analyzed from the perspective of epistatic effects, which is considered a possible explanation for the variability found in these phenotypes, especially those presented as intermediate, such as green and hazel eye colors. The use of quantitative variables to evaluate the eye color, acquired from photographs, revealed for the first time the association of the nonsynonymous mutation rs201872292 in the HERC2 gene with light eyes, independently of the effect of the rs12913832 polymorphism. We highlight the association of the polymorphism rs58358300 located in an intron of the HERC2 gene with sclera pigmentation, which was the first time that a polymorphism is associated with this feature. This was the first study in Brazil to analyze genetic polymorphisms in candidate genes related to normal variation of human pigmentation and iris features by next-generation sequencing. The results confirm the hypothesis that OCA2 and HERC2 genes may contribute to classic pigmentation phenotypes of eyes, skin, hair, freckles and iris features in the Brazilian population.

Keywords: OCA2. HERC2. Human pigmentation. Iris features. Epistasis. Next-generation sequencing

LISTA DE ILUSTRAÇÕES

Figura 1 - Esquema simplificado da síntese de melanina nos melanócitos durante o processo de melanogênese. A produção de eumelanina e feomelanina é determinada pela disponibilidade de cisteína dentro dos melanossomos. (Adaptado de Cichorek et al., 2013).....

Figura 2 - Esquema de superfície de uma íris humana. A íris pode ser dividida em duas zonas que são separadas pela collarette iridiana (linha branca irregular ao redor da pupila). A zona pupilar é a região próxima à pupila (parte interna da collarette iridiana), enquanto que a zona ciliar é a parte mais externa e compreende o restante da íris. Diferentes estruturas na superfície da íris podem ser observadas, como sulcos de contração, nódulos de Wolfflin, criptas de Fuchs e pontos de pigmentação. (Adaptado de Edwards, 2016)...... 26

Figura 12 - Distribuição dos indivíduos de acordo com a presença (Sim) ou ausência (Não) de estruturas secundárias presentes na íris dentro das categorias de cor dos olhos. A: Nódulos de Wolfflin; B: Sulcos de contração; C: Anel de pigmentação; D: Pigmentação na esclera; E: Pontos de pigmentação na íris. O número de indivíduos com a presença de cada estrutura secundária entre as categorias de cor do olho está destacado entre parênteses. 56

Figura 21 - Contribuição ancestral entre os indivíduos que apresentam olhos azuis, verdes, mel, castanho-claros e castanho-escuros. A: Ancestralidade europeia; B: Ancestralidade africana; C: Ancestralidade ameríndia; Valor global de p para o teste de Kruskal-Wallis no

Figura 24 - Correlação entre ancestralidade e índice de melanina para área não exposta ao sol. A: Ancestralidade europeia; B: Ancestralidade africana; C: Ancestralidade ameríndia. 68

Figura 27 - Representação do desequilíbrio de ligação dos genes *OCA2* e *HERC2*. Os blocos em desequilíbrio estão cercados em preto. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel *et al.* (2002). No canto inferior esquerdo estão os nove blocos representados, sendo o três primeiros pertencentes ao gene *OCA2* e os outros seis ao gene *HERC2*. No mapa acima, vermelho indica LOD (*Log of Odds*) \geq 2 e D' = 1; azul indica LOD < 2 e D' = 1; tons de rosa indicam LOD \geq 2 e D' < 1; branco indica LOD < 2 e D' < 1..72

LISTA DE TABELAS

Tabela 2 - Proporção de cada população atribuída a cada um dos três clusters definidospelo programa STRUCTURE.58

 Tabela 3 - Distribuição dos 170 pontos de variação nas regiões dos genes OCA2 e HERC2. ...

 71

Tabela 5 - Associações genéticas encontradas envolvendo alelos de nove polimorfismosanalisados nos genes OCA2 e HERC2 com a pigmentação dos olhos. Valores de psignificativos após correção de Bonferroni para múltiplos testes estão destacados emnegrito.74

Tabela 6 - Associações genéticas encontradas envolvendo alelos de três polimorfismosanalisados nos genes OCA2 e HERC2 com a pigmentação dos cabelos. Valores de psignificativos mesmo após correção de Bonferroni para múltiplos testes estão destacados emnegrito.75

Tabela 8 - Associações genéticas encontradas envolvendo alelos de oito polimorfismosanalisados nos genes OCA2 e HERC2 com características de estruturas secundáriaspresentes na íris. Valores de p significativos mesmo após correção de Bonferroni paramúltiplos testes estão destacados em negrito.77

Tabela 10 - Associações genéticas encontradas envolvendo alelos de sete polimorfismosanalisados nos genes OCA2 e HERC2 com o índice de melanina em região não exposta eexposta ao sol.81

Tabela 12 - Dez modelos de interações de segunda ordem mais significativos compostos porpolimorfismos do gene OCA2 e HERC2 para a dimensão L* do espaço de cor CIELAB.Valores p significativos assinalados em negrito.83

Tabela 20 - Dez modelos de interações de segunda ordem mais significativos compostos por
polimorfismos do gene OCA2 e HERC2 para o índice de melanina em região exposta ao sol.
Valores p significativos assinalados em negrito.90

Tabela 21 - Lista de polimorfismos localizados na região codificante dos genes *OCA2* e *HERC2* associados no presente estudo com base no genoma de referência hg19......100

LISTA DE APÊNDICES

APÊNDICES 123
Apêndice A - Tabelas com as frequências alélicas dos marcadores informativos de ancestralidade entre as populações do 1000genomes utilizados para o controle da estratificação populacional
Apêndice B - Tabela com os 170 polimorfismos analisados dos genes <i>OCA2</i> e <i>HERC2</i> 125
Apêndice C - Tabelas com as comparações par-a-par de desequilíbrio de ligação entre os polimorfismos associados nas análises de regressão logística e linear
Apêndice D - Tabelas com as frequências alélicas, genotípicas, heterozigosidade esperada e observada e aderências ao Equilíbrio de Hardy-Weinberg
Apêndice E - Gráficos bi-dimensionais entre os valores de L*a*b* para cor dos olhos 138
Apêndice F - Diferenciação das medianas do valor de L* e a* para cada um dos grupos heteroclassificados para cor dos olhos
Apêndice G - Imagem do <i>UCSC Genome Browser</i> mostrando a conservação do aminoácido alterado pelo polimorfismo rs6178394C>T em diferentes espécies
Apêndice H - Imagem do <i>UCSC Genome Browser</i> mostrando a conservação do aminoácido alterado pelo polimorfismo rs74005645C>T em diferentes espécies
Apêndice I - Imagem do <i>UCSC Genome Browser</i> mostrando a conservação do aminoácido alterado pelo polimorfismo rs118112076G>C em diferentes espécies143
Apêndice J - Imagem do <i>UCSC Genome Browser</i> mostrando a conservação do aminoácido alterado pelo polimorfismo rs201872292G>A em diferentes espécies
Apêndice K - Imagem do UCSC Genome Browser mostrando a conservação do aminoácido

LISTA DE ANEXOS

ANEXOS	146
Anexo A - Termo de Consentimento Livre e Esclarecido (TCLE)	
Anexo B - Protocolo de análise laboratorial (<i>HaloPlex</i>) para sequenciamento de 1	10va geração 152
Anexo C - Protocolo de processamento computacional de dados de sequenciam geração	ento de nova 160
Anexo D - Resultado do Sequenciamento de Nova Geração: Pré e Pós filtros de c	Jualidade

LISTA DE ABREVIATURAS

3'UTR	3' Untranslated Region
AFR	Populações africanas do 1000Genomes
AIMs	Ancestry Informative Markers
ASIP	Agouti Signaling Protein
BAM	Binary Alignment/Map
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CDS	Coding Sequence
CIELAB	International Commission on Illumination
DHI	5,6-dihydroxyindole
DHICA	DHI-2-caboxylic acid
DL	Desequilíbrio de Ligação
DNA	Deoxyribonucleic acid
DOPA	levodopa
DSCR9	Down Syndrome Critical Region 9
EAS	Populações do leste asiático do 1000Genomes
EF	Escala Fitzpatrick
EUR	Populações europeias do 1000Genomes
ExAC	Exome Aggregation Consortium
GATK	Genome Analysis Toolkit
GRCh37	Genome Reference Consortium human genome build 37
HC	HaplotypeCaller
HERC1	HECT And RLD domain containing E3 ubiquitin protein ligase 1
HERC2	HECT and RLD domain containing E3 ubiquitin protein ligase 2
hg19	Human Genome version 19
HLTF	Helicase-Like Transcription Factor
IBGE	Instituto Brasileiro de Geografia e Estatística
IC 95%	Intervalo de Confiança a 95%
IGV	Integrative Genomics Viewer
InDels	Polimorfismo de Inserção/Deleção
IRF4	Interferon Regulatory Factor 4
Kb	Kilobase
LOD	Log of Odds
LPFG	Laboratório de Pesquisas Forenses e Genômicas
MAF	Minor Allele Frequency
Mb	Megabase
MC1R	Melanocortin 1 Receptor
mRNA	RNA mensageiro

NGS	Next-generation Sequencing
OCA2	Oculocutaneous albinism II
OR	Odds Ratio
pb	pares de base
PCA	Principal Component Analysis
PCR	Polymerase chain reaction
RBP	Ribeirão Preto
RefSeq	Reference Sequence
RGB	Red, Green, Blue
SAM	Sequencing Alignment/Map
SEMA3A	Semaphorin 3A
SLC24A4	Solute Carrier Family 24 Member 4
SLC24A5	Solute Carrier Family 24 Member 5
SLC45A2	Solute Carrier Family 45 Member 2
SNPs	Single Nucleotide Polymorphisms
STRs	Short Tandem Repeats
TCLE	Termo de Consentimento Livre e Esclarecido
TRAF3	TRAF3 Interacting Protein 1
TYR	Tyrosinase
TYRP1	Tyrosinase Related Protein 1
TYRP1	Tyrosinase Related Protein 1
TYRP2	Tyrosinase Related Protein 2
UCSC	University of California, Santa Cruz, Genome Browser
VCF	Variant Calling Format

SUMÁRIO

1. INTRODUÇÃO
1.1. Bases moleculares da pigmentação23
1.1.1. Pigmentação dos olhos25
1.1.2. Estruturas secundárias presentes na íris25
1.1.3. Bases genéticas da pigmentação dos olhos27
1.1.4. Bases genéticas das estruturas secundárias
1.2. OCA2
1.3. <i>HERC2</i>
1.4. Interações epistáticas entre os genes OCA2 e HERC2
1.5. População brasileira como modelo tri-híbrido
2. JUSTIFICATIVA
3. HIPÓTESE
4. OBJETIVOS
5. MATERIAL E MÉTODOS
5.1. Caracterização da amostra e aspectos éticos
5.1.1. Análise quantitativa da cor dos olhos e estruturas secundárias
5.2. Regiões genômicas cobertas pelo sequenciamento de nova geração
5.3. Procedimentos laboratoriais
5.3.1. Extração do DNA genômico (Anexo B - Item 1)
5.3.2. Quantificação do DNA genômico (Anexo B - Item 2)
5.3.3. Preparo das bibliotecas (Anexo B - Itens 3 a 4)
5.3.3.1. Controle de qualidade e quantificação das bibliotecas (Anexo B - Item 5) 44
5.3.3.2. Sequenciamento de nova-geração (Anexo B - Item 6)
5.4. Processamento de análise dos dados45
5.4.1. Primeira etapa de análises45
5.4.1.1. Obtenção dos dados brutos e controle de qualidade (Anexo C - Item 1) 45
5.4.1.2. Trimagem e alinhamento das sequências
5.4.2. Segunda etapa de análises46
5.4.2.1. Determinação de variantes: HaplotypeCaller (Anexo C - Item 1.3)
5.4.2.2. Avaliação e remoção de genótipos de baixa qualidade: VCFx (Anexo C - Item 1.3.2)
5.4.2.3. Anotação de variantes
5.5. Análises estatísticas
5.5.1. Análises de epistasia entre polimorfismos51
5.5.2. Marcadores informativos de ancestralidade52
6. RESULTADOS
6.1. Caracterização populacional e fenótipos de pigmentação

6.1.1. AIMs e controle da estratificação populacional
6.2. Análise quantitativa da cor da íris60
6.2.1. Comparações entre as íris direita e íris esquerda
6.2.2. Distribuição dos valores L*a*b*61
6.3. Análises entre os índices de melanina65
6.3.1. Índice de melanina v s. cor de pele e classificação do IBGE65
6.4. Identificação das variantes genéticas70
6.4.1. Associação genética com fenótipos de pigmentação
6.4.2. Associação genética com estruturas secundárias na íris
6.4.3. Associação genética com os valores quantitativos de CIELAB
6.4.4. Associação genética com índice de melanina
6.5. Análises epistáticas de segunda ordem82
6.5.1. Interações epistáticas: CIELAB (L*, a* e b*)
6.5.2. Interações epistáticas: Índice de melanina
7. DISCUSSÃO
7.1. Amostra populacional91
7.2. Estruturas secundárias presentes na íris93
7.3. Variantes dos genes OCA2 e HERC2 e pigmentação humana
7.3.1. rs12913832A>G: o principal preditor da cor do olhos e outras características fenotípicas
7.3.2. Associações de polimorfismos não-sinônimos localizados na região codificadora dos genes OCA2 e HERC2100
7.3.3. Associações de polimorfismos sinônimos localizados na região codificadora dos genes OCA2 e HERC2
7.3.4. Associações de polimorfismos localizados na região intrônica dos genes <i>OCA2</i> e <i>HERC2</i>
7.3.5. Epistasia e sua importância nos estudos genéticos de pigmentação 109
8. CONCLUSÃO 111
9. REFERÊNCIAS BIBLIOGRÁFICAS 113
APÊNDICES 123
Apêndice A 123
Apêndice B 125
Apêndice C 129
Apêndice D 133
Apêndice E138
Apêndice F140
Apêndice G141
Apêndice H 142
Apêndice I
Apêndice J

ANEXOS	
Апехо А	
Anexo B	
Anexo C	
Anexo D	
MANUSCRITO	

1 INTRODUÇÃO

1.1. Bases moleculares da pigmentação

A pigmentação da pele, olhos e cabelos está entre os exemplos mais visíveis da variação fenotípica humana (Nan *et al.*, 2009). A pigmentação humana é atribuída ao número, tipo e distribuição celular dos melanossomos, que são compartimentos subcelulares produzidos pelos melanócitos, e que sintetizam e armazenam a melanina (biopolímero que atua na absorção de luz). Durante o processo de melanogênese, os melanossomos são secretados para os queratinócitos localizados na junção entre a epiderme e a derme (McEvoy *et al.*, 2006). Já no caso dos melanócitos dos olhos, os melanossomos não são secretados, mas mantidos no citoplasma (Jablonski e Chaplin, 2000; Sturm e Frudakis, 2004; Sturm *et al.*, 2008; Sturm, 2009).

As variações fenotípicas nas cores de pele, cabelo e olhos, observadas em diferentes regiões geográficas, resultam de dois tipos distintos de melanina depositados na epiderme, fios de cabelo e íris: a eumelanina (cuja formação da cor castanho/preta está associada a pigmentação mais escura) e a feomelanina (cuja formação da cor vermelha/amarela está associada a pigmentação mais clara) (Gerstenblith *et al.*, 2007). Em cabelos ruivos há uma predominância das quantidades de feomelanina comparadas às de eumelanina, enquanto nos cabelos escuros as quantidades de eumelanina prevalecem, e nos cabelos loiros há pouca representação de cada tipo (Rees, 2003). Um fato interessante é que, apesar da origem embrionária comum dos melanócitos da pele e do cabelo, os genes que afetam a pigmentação podem expressar-se de maneira independente, gerando combinações conhecidas de cabelos escuros e pele clara ou cabelos claros e pele escura (Sturm *et al.*, 2001).

A síntese de melanina utiliza o aminoácido tirosina como principal substrato. A tirosinase constitui enzima essencial para mediação dos primeiros passos do processo da melanogênese, que envolve a hidroxilação da tirosina em DOPA e a subsequente oxidação desta em DOPAquiona (Parra, 2007). Na presença de cisteína, a DOPAquiona reage com esta formando a DOPAcisteinila, que por conseguinte é oxidada e polimerizada dando origem à feomelanina (Hearing, 2011; Simon *et al.*, 2009). Na ausência de cisteína, a eumelanina é formada através de dois processos. O primeiro processo ocorre quando a DOPAcroma gera a DHI (do inglês, *5,6-dihydroxyindole*), que rapidamente é oxidada e polimerizada para formar a DHI-melanina. O segundo processo ocorre quando há a presença de DOPAcroma tautomerase [codificada pelo gene *TYRP2* (do inglês, *Tyrosinase Related Protein 2*)] formando a DHICA (do inglês, *DHI-2-caboxylic acid*), que por fim, através de subsequentes reações catalíticas mediadas pela tirosinase [codificada pelo gene *TYRP1* (do inglês, *Tyrosinase Related Protein 1*)], forma a DHICA-melanina (Cichorek *et al.*, 2013; del MarmoleBeermann, 1996; KondoeHearing, 2011; Sugumaran, 1991) (Figura 1).



Figura 1 - Esquema simplificado da síntese de melanina nos melanócitos durante o processo de melanogênese. A produção de eumelanina e feomelanina é determinada pela disponibilidade de cisteína dentro dos melanossomos. (Adaptado de Cichorek *et al.*, 2013)

Alguns estudos têm demonstrado que a regulação da via metabólica inerente ao processo de melanogênese deve ser examinada para entender a gama de diferenças fenotípicas em características pigmentares (Sturm, 2009; Yamaguchi *et al.*, 2007). Diferentes proteínas estruturais e enzimáticas descritas como participantes da melanogênese podem ser classificadas como (1) proteínas transmembranas (com funções de transporte de íons e solutos, bem como, outras atividades de tráfego proteico), (2) proteínas com atividades catalíticas (enzimas desidrogenases, hidrolases, proteínas quinases, entre outras), (3) proteínas associadas com atividades motoras do citoesqueleto e (4) proteínas de sinalização celular, proteínas *heat-shock*, e outras associadas com organelas como retículo endoplasmático, mitocôndrias e núcleo celular, constituindo, desta forma, mais de 170 genes implicados neste complexo processo (Chi *et al.*, 2006; Hearing, 1999).

No que se refere aos genes envolvidos, SNPs (do inglês, *Single Nucleotide Polymorphisms*) localizados em regiões exônicas podem alterar as propriedades funcionas de proteínas codificadas através da substituição de aminoácidos específicos, enquanto que as mutações encontradas em regiões regulatórias podem levar a alterações de seus níveis de

expressão, resultando em alterações fenotípicas (cor dos olhos, pele, cabelos e presença de sardas).

1.1.1. Pigmentação dos olhos

Assim como na pele e cabelos, o principal fator que explica a diversidade de pigmentação dos olhos é o tipo e a quantidade de melanina presente na íris. Estudos indicam que olhos claros possuem uma menor concentração de melanina quando comparados a olhos com colorações mais escuras (Hu *et al.*, 2009; Prota *et al.*, 1998).

Adicionalmente, a proporção entre eumelanina e feomelanina também está relacionada a diferentes colorações observadas. A maioria dos olhos contém uma maior quantidade de eumelanina e pouca concentração de feomelanina. De fato, a feomelanina mostra concentrações muito próximas tanto em olhos azuis quanto em olhos castanhos, sendo que os últimos são caracterizados pelos maiores níveis de eumelanina (WielguseSarna, 2005). Entretanto, indivíduos com olhos verdes são únicos no sentido de que estes possuem concentrações de feomelanina maiores em relação a baixas concentrações de eumelanina (Prota *et al.*, 1998).

Apesar da melanina contribuir majoritariamente na explicação da pigmentação da íris, outros fatores que podem contribuir na diversidade de cores de olhos são a distribuição dos melanossomos na íris e a presença ou ausência de estruturas secundárias.

1.1.2. Estruturas secundárias presentes na íris

As estruturas secundárias da íris podem ser definidas como diferentes texturas presentes na superfície da mesma. Estas estruturas, assim como a cor da íris, possuem uma diversidade quanto a sua distribuição e quantidade.

Entre as estruturas presentes na superficie da íris, cinco podem ser citadas: (1) as criptas de Fuchs, (2) os nódulos de Wolfflin, (3) pontos de pigmentação na íris, (4) sulcos de contração e (5) melanose conjuntival (também conhecida como pigmentação na esclera). Criptas de Fuchs são lacunas em forma de diamante que surgem durante o desenvolvimento do olho por reabsorção da membrana pupilar (Purtscher, 1965). Nódulos de Wolfflin consistem em pequenos agrupamentos de colágeno atrofiados que se acumulam na parte mais externa da zona ciliar. Sulcos de contração representam anéis descontínuos que se estendem ao redor da parte mais externa da zona ciliar, e pontos de pigmentação são pequenas regiões de hiper-pigmentação devido ao acúmulo de melanina (Figura 2).

É importante ressaltar que poucos estudos têm voltado atenção para compreender as bases genéticas que possam estar envolvidas na diversidade destas estruturas. Apesar das consequências funcionais destas ainda serem pouco compreendidas, alguns estudos indicam que certas estruturas, como os nódulos de Wolfflin e pontos de pigmentação na íris, podem influenciar na percepção de cor da íris (Liu *et al.*, 2010; Mackey *et al.*, 2011). Além disso, discussões na literatura também indicam que avaliar estas estruturas pode auxiliar a revelar riscos associados a diferentes patologias. Pontos de pigmentação na íris, por exemplo, são considerados marcadores de risco relativo para melanoma ocular (Holly *et al.*, 1990; Horn *et al.*, 1994).



Figura 2 - Esquema de superficie de uma íris humana. A íris pode ser dividida em duas zonas que são separadas pela collarette iridiana (linha branca irregular ao redor da pupila). A zona pupilar é a região próxima à pupila (parte interna da collarette iridiana), enquanto que a zona ciliar é a parte mais externa e compreende o restante da íris. Diferentes estruturas na superficie da íris podem ser observadas, como sulcos de contração, nódulos de Wolfflin, criptas de Fuchs e pontos de pigmentação. (Adaptado de Edwards, 2016).

1.1.3. Bases genéticas da pigmentação dos olhos

De acordo com Larson e Pedersen (2004), a cor dos olhos é uma característica fenotípica com cerca de 98% de herdabilidade, o que significa que a maior parte da diversidade fenotípica existente pode ser explicada por variações genéticas.

Diferentes grupos de pesquisa apontam que a grande maioria da diversidade de pigmentação dos olhos é atribuída a um único locus localizado no íntron 86 do gene *HERC2* (do inglês, *HECT and RLD Domain Containing E3 Ubiquitin Protein Ligase 2*) (Eiberg *et al.*, 2008; Kayser *et al.*, 2008; Sturm *et al.*, 2008). Este polimorfismo conhecido como rs12913832A>G é tido como o principal marcador capaz de explicar majoritariamente a variação de cor existente entre olhos azuis e castanhos. O alelo "A" possui uma frequência elevada em populações com ancestralidade diferentes da europeia, enquanto que o alelo "G" é frequentemente encontrado em populações europeias. Tradicionalmente, acredita-se que este polimorfismo atue sob um modelo de herança dominante/recessiva, de forma que, indivíduos homozigotos para o alelo "A" e heterozigotos "AG" apresentam olhos azuis.

Do ponto de vista molecular, este sítio está inserido em uma região altamente conservada, contendo sítios de ligação de diferentes proteínas regulatórias (Visser *et al.*, 2012). Estudos apontam que este polimorfismo atua na regulação do gene *OCA2* (do inglês, *Oculocutaneous Albinism II*), localizado 10kb à montante (*upstream*) do gene *HERC2*. Acredita-se que esta regulação ocorra por meio de um loop de cromatina de longo alcance que permite a interação entre o sítio do gene *HERC2* com a região promotora do *OCA2*, de forma que o alelo "A" do polimorfismo rs12913832 estaria associado a uma maior expressão do gene *OCA2*, levando ao aparecimento de olhos castanhos, enquanto que o alelo "G" estaria associado a uma baixa expressão, e assim ocasionando o aparecimento de fenótipos mais claros (olhos azuis) (Visser *et al.*, 2012; Visser *et al.*, 2014).

Apesar de ser apontado como o principal sítio de variação responsável por controlar as diferenças entre olhos azuis e castanhos, há discrepâncias entre a associação deste polimorfismo e a pigmentação dos olhos, uma vez que alguns estudos indicam que a relação entre os alelos com olhos azuis ou castanhos nem sempre é verdadeira (Mengel-From *et al.*, 2010; Sturm *et al.*, 2008).

Outros genes também foram associados a diversidade na cor dos olhos, mesmo que de forma mais sutil. É importante ressaltar que a maioria destas associações foi encontrada em populações europeias.

Entre elas referem-se os genes ASIP (do inglês, Agouti Signaling Protein) (Bonilla et al., 2005; Kanetsky et al., 2002); IRF4 (do inglês, Interferon Regulatory Factor 4) (Han et al., 2008, Liu et al., 2010; Walsh et al., 2012); SLC24A4 (do inglês, Solute Carrier Family 24 Member 4) (Sulem et al., 2007; Larsson et al., 2011); SLC24A5 (do inglês, Solute Carrier Family 24 Member 5) (Beleza et al., 2013; Pietroni et al., 2014); SLC45A2 (do inglês, Solute Carrier Family 45 Member 2) (Graf et al., 2007; Liu et al., 2010); TYR (do inglês Tyrosinase)

(Liu et al., 2010; Sulem et al., 2007); TYRP1 (do inglês, Tyrosinase Related Protein 1) (Frudakis et al., 2003).

1.1.4. Bases genéticas das estruturas secundárias

Apesar das bases genéticas de estruturas secundárias presentes na íris serem muito pouco compreendidas, Larson e Pedersen, 2004 descrevem que a herdabilidade destas características são altas (66% para criptas de Fuchs; 78% para sulcos de contração e nódulos de Wolfflin e 58% para pontos de pigmentação na íris) (Larsson e Pedersen, 2004).

Até o momento, três grupos de pesquisa têm voltado seus esforços para identificar variantes genéticas que possam estar ligadas ao aparecimento destas estruturas (Edwards et al., 2016; Larsson et al., 2011; Sturm, 2009), sendo que muitas das variantes associadas aparentemente são diferentes das variantes associadas a pigmentação dos olhos (Larsson et al., 2011). Tais variantes normalmente estão localizadas em genes relacionados à embriogênese e desenvolvimento do olho, quando comparado aos genes para pigmentação dos olhos, que em sua totalidade estão envolvidos com algum processo da melanogênese. Alguns destes genes podem ser citados, como o SEMA3A (do inglês, Semaphorin 3A) e seu polimorfismo rs10235789 fortemente associado a criptas de Fuchs, o qual explica 1,5% da variabilidade observada nesta característica; TRAF3IP1 (do inglês, TRAF3 Interacting Protein 1) e o polimorfismo rs3739070, cuja associação com sulcos de contração representa 1,7% da variabilidade para esta característica (Edwards et al., 2016; Larsson et al., 2011); HERC1 (do inglês, HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase Family Member 1) rs11630290, associado com pontos de pigmentação na íris (Larsson et al., 2011); DSCR9 (do inglês, Down Syndrome Critical Region 9) rs7277820, o qual supostamente pode modular a cor dos olhos controlando a presença e a quantidade de nódulos de Wolfflin na íris (Larsson et al., 2011).

É importante ressaltar que o conhecimento a respeito das bases genéticas que envolvem estas estruturas é escasso e pouco compreendido. Os marcadores supracitados explicam muito pouco da variabilidade existente, o que leva a crer que muitos polimorfismos ainda não identificados possam contribuir para explicar a variabilidade ainda desconhecida.

1.2. OCA2

O gene *OCA2* (assim chamado devido ao fenótipo anormal de pigmentação, albinismo oculocutâneo tipo 2), está localizado na região cromossômica 15q11. De acordo com sua sequência *RefSeq* (NG_009846.1), possui 344.438pb, 24 éxons e codifica a proteína transmembrana melanossomal conhecida como proteína P, composta por 838 aminoácidos.

Proteínas que modulam a produção de melanina nos melanócitos são normalmente encontradas dentro ou próximas aos melanossomos. Dentre estas, destaca-se a atuação da proteína P codificada pelo gene *OCA2*. Acredita-se que tal proteína esteja envolvida no

Introdução 29

transporte de tirosina e regulação do pH melanossomal (Wilde *et al.*, 2014; Yamaguchi e Hearing, 2014).

Estudos revelam que deleções em regiões que englobam este gene no cromossomo 15, como observado nas Síndromes de Prader-Willi e Angelman, estão associadas com a hipopigmentação da pele, cabelo e olhos, e cópias extras destas regiões resultam em hiperpigmentação generalizada da pele (Sturm *et al.*, 2008).

A despeito das mutações associadas com o albinismo oculocutâneo tipo 2 e outras doenças, diversos estudos tem associado polimorfismos dentro ou próximos ao gene *OCA2* com a variação normal da pigmentação, sendo este considerado um dos principais genes na determinação fenotípica da cor dos olhos (Bouakaze *et al.*, 2009; Eiberg *et al.*, 2008; Frudakis *et al.*, 2007; Han *et al.*, 2008; Kayser *et al.*, 2008; Lao *et al.*, 2007; Nan *et al.*, 2009; Shekar *et al.*, 2008; Sulem *et al.*, 2007).

Alelos de diversos sítios polimórficos foram associados com cor de olhos azuis em europeus. Entretanto, dois SNPs são constantemente identificados ou avaliados em diferentes estudos. Os polimorfismos rs1800407 e rs1800414, nos éxons 13 e 19, foram associados a cor dos olhos castanho com tons de verde em europeus e pele clara em indivíduos asiáticos, respectivamente (Branicki *et al.*, 2009; Edwards *et al.*, 2010). Entretanto, dados de estudos recentes demonstram que a influência do gene *OCA2* na variação normal da pigmentação da pele, olhos e cabelos não é determinada exclusivamente por mutações em regiões codificantes de proteína, mas também por mutações que afetam regiões regulatórias que controlam os níveis de expressão deste gene (Visser *et al.*, 2012; Visser *et al.*, 2014).

1.3. HERC2

O gene *HERC2* está localizado na região cromossômica 15q13, 10kb *upstream* em relação ao gene *OCA2*. De acordo com sua sequência RefSeq (NG_016355.1), possui 211.116pb, 93 éxons e codifica a proteína E3 *ubiquitin-protein ligase* HERC2 composta por 4.834 aminoácidos.

Assim como o gene *OCA2*, está majoritariamente ligado a determinação fenotípica da cor dos olhos, mas também influencia a pigmentação dos cabelos e pele. Estudos apontam que o gene *HERC2* está envolvido na regulação do gene *OCA2* (Eiberg *et al.*, 2008; Kayser *et al.*, 2008; Sturm *et al.*, 2008; Visser *et al.*, 2012), através de interações com proteínas que aumentam a ação de fatores de transcrição na região promotora do gene *OCA2*, contribuindo na determinação da cor dos olhos (Visser *et al.*, 2012).

Por não codificar proteínas diretamente envolvidas no processo de melanogênese, o gene *HERC2* não é considerado um gene pigmentar. Entretanto, contém diversos SNPs relacionados a pigmentação. Dentre eles, conforme dito anteriormente, o SNP rs12913832 localizado no íntron 86 (21kb *upstream* em relação ao gene *OCA2*), que é considerado um

dos principais determinantes da variação da cor dos olhos humano, com forte poder preditivo (Branicki *et al.*, 2011; Spichenok *et al.*, 2011).

1.4. Interações epistáticas entre os genes OCA2 e HERC2

Sabendo que a pigmentação é um fenômeno poligênico, e que o efeito de múltiplos genes na determinação de um fenótipo não pode ser limitado a suas atuações independentes, torna-se natural inferir que efeitos epistáticos possam atuar sobre esta característica e deve ser levado em consideração em estudos de traços poligênicos, tais como a determinação fenotípica da cor dos olhos, cujo papel principal é atribuído aos dois genes aqui estudados, *OCA2* e *HERC2* (Pośpiech *et al.*, 2014; Pośpiech *et al.*, 2011).

A existência de interação epistática entre estes dois genes já foi relatada do ponto de vista molecular. A região ao redor do SNP rs12913832 do gene *HERC2* é altamente conservada entre diferentes espécies de animais e está fortemente associada à regulação dos níveis de expressão de OCA2 por conectar-se com a região promotora deste através de um loop de cromatina de longo alcance, atuando como um elemento intensificador (*enhancer*) melanócito-específico. Maiores níveis de expressão de OCA2 em melanócitos foram associados a indivíduos portadores do alelo ancestral rs12913832*A no gene *HERC2* (Visser *et al.*, 2012).

Entretanto, diante da grande quantidade de sítios de variação existentes ao longo dos genes *OCA2* e *HERC2*, observa-se a necessidade de estudos mais abrangentes considerando o impacto das interações epistáticas entre tais genes sobre o amplo espectro de variação que caracteriza a pigmentação humana (Branicki *et al.*, 2009). O estudo da diversidade destes genes em populações altamente heterogêneas fornecerá elementos necessários para a identificação de eventuais interações ainda desconhecidas.

1.5. População brasileira como modelo tri-híbrido

A população brasileira é o produto da miscigenação entre três principais grupos étnicos: ameríndios, europeus e africanos.

A interação com diferentes biomas na América do Sul contribuiu para a diversificação da estruturação social de diferentes tribos indígenas. Esta interação refletiuse em composições genéticas e culturais distintas, dando origem a diferentes grupos populacionais de ameríndios. Populações habitantes da cordilheira dos Andes eram caracterizadas por maiores tamanhos populacionais, maior intensidade de fluxo gênico e, consequentemente, menor diferenciação populacional. Populações localizadas ao oeste da Cordilheira dos Andes, particularmente aquelas localizadas na floresta Amazônica, eram caracterizadas por menores tamanhos populacionais e isolamento em relação a aldeias de tribos distintas, o que acarreta em uma elevada diferenciação genética. Deste modo, quando iniciou-se o contato com imigrantes europeus e africanos (Tarazona-Santos *et al.* 2001), as populações de cada localidade apresentavam características genéticas e culturais bastante variáveis (Ribeiro, 1995).

O processo de miscigenação iniciou entre os anos de 1500 e 1530 com a chegada dos europeus (portugueses) em um continente habitado por aproximadamente dois milhões de nativos americanos (Almeida, 1997; Salzano e Bortolini, 2002). Após a abertura dos portos em 1808, o fluxo de imigrantes europeus foi intensificado com a chegada de italianos, alemães e espanhóis (Salzano e Bortolini, 2002; Venâncio, 2007). Estima-se que cerca de 15 milhões de europeus tenham cruzado o Atlântico até 1880 e que entre 1881 a 1915 o número de migrantes tenha alcançado seu apogeu, chegando a cerca de 31 milhões de indivíduos (Fausto, 2000).

O componente africano na população brasileira foi atribuído à prática escravagista que teve início por volta de 1538 e se manteve por aproximadamente 350 anos (Salzano e Freire-Maia, 1970). Durante este período, milhares de escravos foram embarcados no litoral atlântico da África, em diversos locais da costa que se estende do Senegal ao norte da Namíbia e também no litoral da África Oriental, de Moçambique à Somália. Estes entraram no país principalmente pelo Maranhão, Bahia, Rio de Janeiro e Pernambuco, sendo posteriormente distribuídos para as demais regiões (Tavares, 1988). Estima-se que, até 1880, pelo menos 10 milhões de africanos tenham sido trazidos para o continente americano.

Além de europeus e africanos, o Brasil recebeu imigrantes de outros países, tais como japoneses e sírio-libaneses que saíram de seus países em busca de melhores condições de vida nas Américas (Muniz *et al.*, 2008).

A migração em massa destes diferentes grupos populacionais promoveu padrões genéticos diferentes, havendo o estabelecimento de populações com vários tipos e graus de mistura em toda América. Além disso, a migração em massa promoveu também mudanças demográficas, havendo redução da população nativa em aproximadamente 90% do que existia à época do início da colonização (Bedoya *et al.*, 2006). Este processo migratório que envolveu o Brasil desde a época de seu descobrimento e que perdura há mais de cinco séculos culminou na atual composição da população brasileira e implicou não somente em alterações nas composições genéticas inerentes à miscigenação, como também na inserção de diferentes culturas e costumes.

O aparecimento de diferentes fenótipos de pigmentação na população brasileira pode estar associado à distribuição dos diferentes alelos, polimórficos ou não, que segregam nas populações parentais. Adicionalmente, o processo de miscigenação em terras brasileiras, associado à expansão populacional e consequente acúmulo de mutações de novo e de eventos de recombinação, certamente levou ao surgimento de haplótipos únicos, compostos por alelos que podem interagir de forma sinérgica.

Neste contexto, estudos em genética molecular e de populações tornam-se de grande contribuição na tentativa de caracterizar as bases genéticas da pigmentação de uma população sabidamente tri-híbrida como a brasileira, que possui um confiável registro histórico de miscigenação.

2 JUSTIFICATIVA

A identificação de polimorfismos genéticos associados a pigmentação humana e a busca pela compreensão das bases envolvidas nos mecanismos moleculares responsáveis pela determinação de fenótipos específicos, têm contribuído de forma significativa na área forense e na área relacionada a prática clínica.

A descoberta de polimorfismos funcionais pode ser especialmente informativa em ambas as abordagens (forense e clínica) aqui apresentadas, uma vez que, estes fazem-se necessários para se evitar a possibilidade de uso de polimorfismos que apresentem padrão de associação do tipo população-específicos, o que tornaria sua utilização restrita a populações com padrões de ancestralidade e desequilíbrio de ligação semelhantes aos da população onde os marcadores foram descritos.

Até o momento, poucos estudos na literatura propõem uma análise tão aprofundada destes dois genes (*OCA2* e *HERC2*) em uma mesma amostragem sob a perspectiva aqui apresentada, incluindo análise completa das regiões regulatórias e exônicas com o emprego da técnica de sequenciamento de nova geração. Além disso, destaca-se o fato dos referidos genes ainda não terem sido caracterizados e estudados de forma tão contundente e detalhada em uma população de origem miscigenada tri-híbrida como a brasileira, a qual apresenta condições propícias para a identificação de novos sítios de variação que possam ter um efeito funcional em genes de pigmentação.

3 HIPÓTESE

Polimorfismos localizados em regiões exônicas e regulatórias dos genes *OCA2* e *HERC2*, apresentam correlação com a intensidade de pigmentação dos olhos, pele e cabelos em populações urbanas brasileiras e com a presença de estruturas secundárias presentes na íris. Isto se deve tanto a modificações proteicas, alterações do padrão de expressão gênica e interações epistáticas entre sítios polimórficos distribuídos pelos dois genes.

4 **OBJETIVOS**

O objetivo geral do presente projeto consiste na análise da diversidade genética das regiões codificante e regulatórias do genes *OCA2* e *HERC2* em amostras de 340 indivíduos da população do Estado de São Paulo. Adicionalmente, o estudo permitirá atingir outros objetivos, como por exemplo:

- Determinar as frequências de alelos, genótipos de cada região gênica analisada.
- Identificar novos polimorfismos funcionais.
- Avaliar quantitativamente a cor do olho e da pele por meio de duas diferentes metodologias.
- Identificar estruturas secundárias na íris por meio da análise de imagens digitais.
- Associar os alelos e genótipos encontrados com as características fenotípicas de pigmentação.
- Identificar as possíveis interações epistáticas entre os polimorfismos das regiões avaliadas

5 MATERIAL E MÉTODOS

A metodologia descrita neste trabalho foi executada seguindo protocolos produzidos no Laboratório de Pesquisas Forenses e Genômicas (LPFG) (Anexo B a C), sob coordenação do Prof. Celso Teixeira Mendes Junior, com exceção das análises para a obtenção de medidas quantitativas para a cor dos olhos e também as análises de epistasia. Deste modo, esta seção apresentará uma versão resumida dos protocolos, enfatizando modificações efetuadas nos mesmos.

5.1. Caracterização da amostra e aspectos éticos

Neste trabalho foram avaliados 340 indivíduos voluntários amostrados junto ao Hemocentro de Ribeirão Preto – SP. O presente estudo foi aprovado em seus aspectos éticos pelo Comitê de Ética em Pesquisa da FFCLRP-USP, de acordo com o Processo CEP-FFCLRP CAAE n.º 25696413.7.0000.5407.

Cada indivíduo participante assinou um Termo de Consentimento Livre e Esclarecido (TCLE) permitindo o uso de seu material biológico (10 mL de sangue periférico, mantido em tubo Vacutainer® com anticoagulante EDTA) para estudo. Questionários foram aplicados a cada um dos participantes da pesquisa para obtenção de informações a respeito dos fenótipos de pigmentação e ancestralidade.

Os participantes, composto por indivíduos de ambos os sexos, com idade variando de 18 e 70 anos, foram classificados em grupos de acordo com a cor dos olhos, cabelos e pele, bem como quanto a presença ou ausência de sardas. As classificações foram realizadas por dois membros do LPFG de forma independente para que não houvesse influência sobre as respostas, seguida por diálogo para alcançar uma definição consensual. Desta forma, a classificação fenotípica foi estabelecida pela heteroclassificação.

A classificação dos indivíduos com relação à pigmentação da pele foi realizada conforme o sistema proposto por Fitzpatrick (1988), que avalia a resposta da pessoa em relação a exposição ao sol. A Escala Fitzpatrick (EF) é dividida em seis categorias onde a pontuação mais baixa (I) representa pele muito clara e sensível à radiação ultra-violeta e a maior pontuação (VI) representa pele escura, que nunca se bronzeia em reação à radiação ultra-violeta.

Adicionalmente, a intensidade da pigmentação da pele foi quantificada por meio do espectrofotômetro portátil CMR-2500d e do software SkinAnalysis (*Konica Minolta Holdings, Inc.*), com as especificações de iluminador D65 e observador 2°. O uso deste equipamento permitiu realizar análises quantitativas de modo a tornar mais objetiva a classificação da pele a partir da determinação do índice de melanina de cada indivíduo analisado.

A cor dos olhos e dos cabelos foi caracterizada e dividida de acordo com categorias criadas em nosso laboratório. Estas classificações foram elaboradas a partir de diferentes
classificações, uma vez que não há um consenso quanto a classificação destes fenótipos na literatura (Edwards *et al.*, 2016). Portanto, as categorias para classificação da cor dos olhos foram (1) azul, (2) verde, (3) mel, (4) castanho claro e (5) castanho escuro. Em relação a cor dos cabelos a classificação compreendeu o (1) ruivo, (2) loiro-claro, (3) loiro-escuro, (4) castanho-claro, (5) castanho-escuro e (6) preto. Imagens fotográficas de regiões corporais pequenas e bem delimitadas (olho, raiz dos cabelos, regiões do braço não-expostas ao sol e região da testa exposta ao sol) foram coletadas de modo que estas não permitissem a identificação do doador. Para tal, foi utilizada uma máquina fotográfica *Canon Powershot SX510 HS* com abertura, *shutterspeed* e ISO fixados para f = 3,4, 1/10 e 100, respectivamente, tendo sido utilizada uma distância média de 7 cm entre a câmera e o indivíduo para obtenção das imagens fotográficas. As imagens foram utilizadas para compor um banco de dados, a fim de tornar mais objetiva a classificação de cor dos olhos e cabelos. Sempre que possível, a coleta foi realizada no mesmo local, de modo a tentar minimizar efeitos externos que pudessem interferir na qualidade das fotos e mensuração dos dados do espectrofotômetro.

O controle e cuidado com o local e a iluminação foram primordialmente importantes para as fotografias das íris do indivíduos coletados. Isto porque as imagens fotográficas das íris foram utilizadas para realizar análises quantitativas da cor do olho.

5.1.1. Análise quantitativa da cor dos olhos e estruturas secundárias

Análises quantitativas da cor dos olhos e de estruturas secundárias presentes em ambos os olhos foram realizadas por meio de imagens fotográficas, com o auxílio de um programa desenvolvido no laboratório do Prof. Esteban J. Parra na *University of Toronto, Mississauga, Canada*. Tais análises foram conduzidas em tal laboratório durante o período de estágio no exterior concedido pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

As imagens fotográficas foram capturadas durante a coleta de amostras. Cada participante teve as duas íris (direita e esquerda) fotografadas. Visando manter uma iluminação constante, foi acoplado à câmera um flash externo em forma de anel que projetava uma luz de LED branca, contínua e estável. O flash externo poderia ser ajustado quanto a sua intensidade de luz e quanto ao reflexo desta luz na periferia da pupila de três diferentes formas, sendo uma com o flash completo em forma de anel e duas outras formas dividindo o flash em luz direita e esquerda. Quatro imagens fotográficas foram capturadas para cada indivíduo. Para cada íris foi obtido uma foto com o flash voltado para a parte interna do olho (Figura 3A), e uma foto com o flash voltado para a parte externa do olho (Figura 3B).



Figura 3 - Imagem fotográfica da íris mostrando o posicionamento do flash. **A:** flash interno (voltado para a parte interna do olho, sentido ponto lacrimal). **B:** flash externo (voltado para a parte externa do olho, sentido contrário ao ponto lacrimal).

Para a análise quantitativa através destas fotografias, optamos por utilizar as imagens que possuíam o flash voltado para a parte interna do olho. Isto porque a análise é realizada na metade da íris oposta ao ponto lacrimal e, deste modo, o flash acarretaria em uma menor influência.

O programa para análise quantitativa da cor dos olhos e identificação de estruturas secundárias esta detalhadamente descrito em Edwards *et al.* (2016). O programa consiste em uma aplicação para a *Web*, acessada pelo site *http://iris.davidcha.ca*. Após inserção das fotos no site, cinco etapas são seguidas de modo a avaliar a quantificação pigmentar da íris. Para cada foto, é solicitado ao usuário que identifique (a) o ponto central aproximado da íris, (b) o ponto central aproximado da pupila e, em seguida, (c) desenhe um círculo que melho se ajuste em torno dos limites da esclera, zona pupilar e collerette iridiana (que separa a zonas pupilar e ciliar da íris) (Figura 4).



Figura 4 - Etapas para delimitação da íris e posterior análise de quantificação. O usuário projeta círculos que melhor se ajustem **(A)** à região que delimita a esclera da íris (círculo vermelho), **(B)** em torno da periferia da pupila (círculo rosa), e **(C)** à collarette iridiana (círculo azul). **(D)** Isto permite que o programa identifique a zona pupilar (delimitada pelos círculos azul e rosa) e a zona ciliar (delimitada pelos círculos azul e vermelho).

Após estas definições, (d) o programa recorta uma fatia da íris em um ângulo de 60° da região ciliar e da região pupilar oposta ao ponto lacrimal do olho, e (e) realiza uma mensuração da cor média dos olhos (Figura 5). As seis etapas finais são utilizadas para identificar estruturas secundárias também presentes nos olhos, como (a) sulcos de contração, (b) nódulos de Wolfflin, (c) criptas de Fuch e pigmentos delimitados (d) na íris e (e) na esclera e (f) anel de pigmentação ao redor da pupila.



Figura 5 - Imagem aumentada do corte de diferentes cores de olho (azul, verde, mel, castanho claro e castanho escuro, respectivamente), realizado pelo algoritmo do programa utilizado para análises quantitativas.

Cada conjunto de íris (olho direito e olho esquerdo) foi analisado por um único pesquisador (Guilherme Debortoli). Depois de processar todas as fotografias da íris, a visualização independente das regiões cortadas de cada uma das íris foi realizada pelo analista para verificar visualmente quaisquer evidências gle obstrução ou delimitação incorreta que pudesse dificultar a análise pelo algoritmo.

A metodologia implantada no algoritro para quantificação de cor do alho extrai medidas de cor no espectro de cores RGB (do inglês, *red*, *graen* e *blue*) das fotografias da íris, e transforma estas medidas na escala CIELAB (do inglês, *International Commission on Illumination* LAB). CIELAB é um sistema internacional de cores padronizado e projetado para aproximar a percepção de cor pela visão humana (CIE, 1986), motivo pelo qual, tal espectro de cores foi utilizado. As medidas quantitativas fornecidas pelo espectro de cores CIELAB podem ser observadas em três dimensões diferentes, sendo uma dimensão de brilho (L*) e duas dimensões de cores (a* e b*). A primeira dimensão, L*, é usada para representar o brilho, possuindo valores contínuos de 0 a 100, onde 0 é escuro e 100 é claro. As dimensões de a* e b* representam as escalas cromáticas, com valores negativos de a* indicando verde e positivo indicando vermelho, e valores negativos de b* indicando azul e positivos indicando amarelo.

5.2. Regiões genômicas cobertas pelo sequenciamento de nova geração

O sistema de enriquecimento de alvos *HaloPlex* da *Agilent Technologies*, permitiu a análise das regiões genômicas de interesse em um grande número de amostras, através de um ensaio customizado que envolveu dez genes descritos na literatura como participantes no processo de melanogênese. Dentre eles, os genes *OCA2* e *HERC2* tiveram suas regiões regulatórias [promotora e *3'UTR* (do inglês, *3' Untranslated Region*)], e codificadora (éxons e parte dos íntrons que flanqueiam estes éxons), cobertos pelo referido ensaio (Figuras 6, 7 e

8). A Tabela 1 mostra o detalhamento da cobertura das regiões avaliadas nos genes *OCA2* e *HERC2* aqui estudados.

Adicionalmente, além destas regiões gênicas analisadas no ensaio, um conjunto de 34 AIMs (do inglês, *Ancestry Informative Markers*) que compõe o sistema 34-*plex* do consórcio SNP*for*ID foi incluído no ensaio (Phillips *et al.*, 2007), com a finalidade de controlar possíveis associações espúrias nos testes de associação genética através de regressões logísticas e lineares. A descrição dos 34 polimorfismos que compõe o consórcio SNP*for*ID está inclusa no Apêndice A.

Tabela 1 - Regiões regulatórias e exônicas avaliadas nos genes *OCA2* e *HERC2*. O número de éxons e os tamanhos de cada região foram estimados com base na sequência *RefSeq* de cada gene, utilizando- se o software CLC Sequence Viewer 6.8.2 (CLC bio A/S)

	Promotora	Nº de éxons	Tamanho dos éxons	5'UTR ^a	3' <i>UTR</i> [♭]	Tamanho total	N° de Amplicons
OCA2	~2.500pb	24	3.140pb	110pb	513pb	~5.140pb	30
HERC2	~2.500pb	93	15.340pb	109pb	726pb	~17.340pb	102

^a o tamanho da região 5'UTR de cada gene já se encontra contabilizado na coluna "Tamanho total dos éxons", uma vez que tal região se encontra no primeiro éxon.

^b o tamanho da região 3'*UTR* de cada gene já se encontra contabilizado na coluna "Tamanho total dos éxons", uma vez que tal região se encontra no último éxon.



Figura 6 - Regiões cobertas pelo ensaio *HaloPlex* para o gene *OCA2*. Os éxons estão definidos como barras azuis na vertical ao longo da extensão do gene. Acima, em vermelho estão indicadas as coberturas das regiões (promotora, éxons e 3'UTR).



Figura 7 - Regiões cobertas pelo ensaio *HaloPlex* para o gene *HERC2*. Os éxons estão definidos como barras azuis na vertical ao longo da extensão do gene. Acima, em vermelho estão indicadas as coberturas das regiões (promotora, éxons e *3'UTR*).

Human hg19	¢ chr15	chr15:27,985,2	04-28,594,317 Go 👚 🖣	• • 🖗 🖪 × 🖵 🗆		-
	p13 p12 p11.2	p11.1 q11.2 q12	q13.2 q14 q15.1	q21.1 q21.2 q21.3 q22.1	q22.31 q23 q24.1 q24.3	q25.2 q25.3 q26.1 q26.2 q26.3
	28,000 kb	28,100 kb 	28,200 kb	28,300 kb	28,400 kb 	28,500 kb
RefSeq Genes		· · · ∎ · · ·				
			0002		TIENOZ	

Figura 8 - Regiões cobertas pelo ensaio *HaloPlex* para os genes *OCA2* e *HERC2*. Os éxons estão definidos como barras azuis na vertical ao longo da extensão do gene. Acima, em vermelho estão indicadas as coberturas das regiões (promotora, éxons e *3'UTR*).

5.3. Procedimentos laboratoriais

5.3.1. Extração do DNA genômico (Anexo B - Item 1)

Amostras de sangue foram coletadas e armazenadas em geladeira a 4°C. A extração do DNA foi realizada utilizando o protocolo de *salting-out* (Miller *et al.*, 1988) com modificações.

Resumidamente a extração consistiu em uma lise celular seguido de uma precipitação de proteínas e detritos celulares (os quais foram descartados durante o processo) e, por fim, uma precipitação final de ácidos nucléicos de modo a obter o DNA. Após este processo o DNA foi ressuspendido em 200 µL de água e armazenado a -20°C.

5.3.2. Quantificação do DNA genômico (Anexo B - Item 2)

A qualidade e concentração do DNA genômico foram avaliadas por meio de eletroforese em gel de agarose, NanoDrop[®] ND-1000 (*Thermo Fisher Scientific, Inc.*) e Qubit[®] (*Life Technologies, Inc*).

O NanoDrop[®] foi utilizado para avaliar o grau de contaminação do material genômico por proteínas. Para tal, razões de densidade óptica 260/280 variando de 1,8 a 2,0 foram consideradas adequadas. Posteriormente, foi realizada a eletroforese em gel de agarose, a fim de verificar o grau de degradação do material, considerando como indicativo de degradação qualquer banda abaixo de 2,5kb.

O Qubit[®] foi utilizado para avaliar a concentração de DNA genômico por meio do dsDNA BR Assay (*Invitrogen*, Carlsbad, CA), de modo que ao final do processo fossem obtidas alíquotas de DNA a uma concentração de 5ng/µL, normalizando assim as concentrações para utilização na montagem das bibliotecas.

5.3.3. Preparo das bibliotecas (Anexo B - Itens 3 a 4)

Após a quantificação, as bibliotecas de amplicons para cada uma das 340 amostras foram preparadas seguindo as recomendações do fabricante para o ensaio *HaloPlex* (*Agilent Technologies*). Resumidamente, o protocolo consistiu dos seguintes passos:

1. Digestão do DNA genômico com enzimas de restrição;

2. Hibridização dos fragmentos de DNA a sondas biotiniladas;

3. Captura dos fragmentos-alvo utilizando *beads* magnéticas recobertas por estreptavidina;

- 4. Ligação dos fragmentos circularizados;
- 5. Amplificação por PCR dos fragmentos-alvo capturados para montagem da biblioteca;
- 6. Purificação das bibliotecas com beads magnéticas;
- 7. Quantificação da biblioteca de cada indivíduo (vide item 5.2.3.1);
- 8. Preparo do *pool* contendo até 96 bibliotecas (vide item 5.2.3.1)

9. Sequenciamento na plataforma *MiSeq Personal Sequencer (Illumina, Inc.)* (vide item 5.2.3.2).

5.3.3.1. Controle de qualidade e quantificação das bibliotecas (Anexo B - Item 5)

O controle de qualidade é um importante passo, uma vez que a obtenção de uma qualidade satisfatória das bibliotecas geradas é imperativa para o sucesso do sequenciamento. É ideal que haja uma homogeneidade na distribuição dos tamanhos dos fragmentos, pois, do contrário, isto poderia levar a uma baixa cobertura ou à falha de cobertura em determinadas porções da sequência (Cher, 2011).

Para tal, duas diferentes abordagens foram utilizadas para a verificação da qualidade das bibliotecas:

- 1. Eletroforese capilar (Bioanalyzer 2100, Agilent Technologies);
- 2. Fluorimetria (Qubit[®] 2.0 Fluorometer, *Thermo Fisher Scientific Inc.*);

Após a quantificação, as bibliotecas foram armazenas a -20°C.

5.3.3.2. Sequenciamento de nova-geração (Anexo B - Item 6)

A partir das concentrações das bibliotecas individuais, foram preparadas soluções a 4 nmol/L para cada biblioteca. Dez microlitros de cada amostra foram misturados em um único microtubo, gerando *pool* de até 96 bibliotecas. O *pool* de bibliotecas foi diluído de modo que ao final do processo fosse obtido 600 µl de uma solução a 16pM. Tal material foi então injetado no cartucho de sequenciamento *MiSeq Reagent Kit V3* (600 *cycles*) do *MiSeq*

Personal Sequencer. Estas etapas foram realizadas seguindo especificações do fabricante (*Illumina*, *Inc*).

5.4. Processamento de análise dos dados

O fluxograma de análises no NGS (do inglês, *Next-generation Sequencing*) pode ser sumarizado em duas etapas principais: a primeira etapa de análise, que consiste na geração dos dados no formato *fastq*, aplicação de filtros de qualidade e o alinhamento a um genoma de referência. A segunda etapa compreende o processo de determinação das variantes através de comparação a um genoma de referência. Com os pontos de variação gerados e já mensurados quanto a qualidade, análises subsequentes podem ser realizadas (anotação de variantes e análises de associação).

Os procedimentos realizados no processamento e análise dos dados estão detalhadamente descritos no Anexo C. Aqui será apresentada uma breve descrição destes passos.

5.4.1. Primeira etapa de análises

5.4.1.1. Obtenção dos dados brutos e controle de qualidade (Anexo C - Item 1)

A primeira etapa de análise, os arquivos brutos que são gerados pelo sequenciador são transformados em um formato *fastq* onde é atribuída uma pontuação de qualidade (*phred-score*) (Ewing e Green, 1998). Os dados são então disponibilizados para download no site da *Illumina*, em um ambiente denominado *BaseSpace* (*Illumina*, *Inc*).

A partir dos arquivos *fastq* de cada uma das amostras sequenciadas, é realizada uma série de processos que consistem na avaliação da qualidade antes da trimagem e alinhamento dos fragmentos sequenciados a um genoma de referência. Os softwares *fastq*C (Andrews, 2010) e *SAMstat* (Lassmann *et al.*, 2011) foram utilizados para avaliar a qualidade das *reads*. Com os relatórios gerados por estes dois programas, estratégias podem ser definidas para melhorar a qualidade dos dados caso estes não estejam satisfatórios.

5.4.1.2. Trimagem e alinhamento das sequências

Sequências de adaptadores, bem como bases diagnosticadas com baixo *phred-score* pelo *fastq*C devem ser retiradas das extremidades das *reads*, em um processo referido como trimagem. A partir do que pôde ser observado no relatório gerado pelo *fastq*C, a trimagem foi

realizada buscando pela sequência dos adaptadores *paired-end* utilizados no sequenciamento, com o uso do *cutadapt* (Martin, 2011).

Após processamento das *reads*, estas foram mapeadas contra um genoma de referência utilizando o algoritmo *BWA-MEM* (Li *et al.*, 2009), em função da sua especificidade em lidar com sequências de tamanho entre 70 e 1 Mb, além de rapidez e acurácia. O genoma de referência utilizado foi o *Homo sapiens*, *UCSC*, hg19, disponibilizado pela *Illumina* (*Illumina*, *Inc*). Ao final do processo de alinhamento de cada amostra, um arquivo no formato SAM (Sequencing Alignment/Map) é gerado e posteriormente transformado em arquivo *BAM* (*Binary Alignment/Map*), ordenado e indexado, com o emprego do programa SAMtools (Li *et al.*, 2008, Li *et al.*, 2009).

Uma vez ordenados e indexados, os arquivos *BAM* podem ser visualizados por programas como o IGV (do inglês, *Integrative Genomics Viewer*) (uma ferramenta de visualização e exploração interativa de conjuntos de dados genômicos) (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013).

Os resultados derivados do processamento dos dados brutos, descrevendo a distribuição da qualidade média das bases que compõem as *reads* pré- e pós-trimagem dos adaptadores *paired-end* da *Illumina* e a qualidade de mapeamento ao genoma de referência, podem ser visualizados no Anexo D.

5.4.2. Segunda etapa de análises

5.4.2.1. Determinação de variantes: HaplotypeCaller (Anexo C - Item 1.3)

Na segunda fase de análises acontece a etapa de determinação de variantes que termina com a geração de um arquivo: o VCF (do inglês, *Variant Calling Format*). O VCF é um formato de arquivo de texto padronizado para representar SNPs, InDels, STRs e variações estruturais encontradas em posições específicas em relação a um genoma de referência. O arquivo VCF consiste no resultado da análise realizada pelo programa GATK v.3.5 (do inglês, *Genome Analysis Toolkit*) (DePristo *et al.*, 2011; McKenna *et al.*, 2010; Van der Auwera *et al.*, 2013).

Para realizar a determinação de SNPs e InDels foi utilizada a ferramenta *HaplotypeCaller* (HC) do GATK com o parâmetro --emiteRefConfidence GVCF. O HC considera simultaneamente todas as amostras a serem analisadas e, ao encontrar uma região apresentando sinais de variação (região ativa), o programa descarta o mapeamento preexistente e realiza uma remontagem das *reads* naquela região. O programa realinha então cada haplótipo contra o haplótipo de referência com o objetivo de identificar potenciais sítios de variação. Desta forma, possibilita a nomeação de SNPs e InDels de forma simultânea, sendo mais preciso quando lidando com regiões reconhecidamente

problemáticas, como por exemplo, onde são encontrados diferentes tipos de variações próximas umas das outras.

Os parâmetros --dontUseSoftClippedBases e -drf Duplicate*read* foram adicionados a linha de comando default, pois o primeiro direciona o programa a não analisar bases *softclipped* (bases não alinhadas nas extremidades das *reads*), o que diminui a determinação de falso-positivos, enquanto que o segundo é necessário devido às características do ensaio *HaloPlex*, onde a vasta maioria dos fragmentos em uma determinada região possuem exatamente mesmo tamanho.

5.4.2.2. Avaliação e remoção de genótipos de baixa qualidade: VCFx (Anexo C -Item 1.3.2)

Visando reduzir ainda mais a taxa de falso-positivos, o arquivo VCF gerado pelo HC foi submetido à rotina do script VCFx (ferramenta escrita em linguagem C++) (Castelli *et al.*, 2015), que interroga os genótipos com base na proporção de *reads* que foi utilizado pelo método HC para determinação de variantes. As regras que o programa VCFx utiliza para a interrogação dos genótipos estão descritas no Anexo C.

É importante realizar esta avaliação uma vez que alguns dos genótipos considerados pelo HC podem não ser verdadeiros, principalmente quando a genotipagem se refere a um segmento com baixa cobertura de sequenciamento ou quando, no caso de genótipos heterozigotos, um dos alelos é pouco representado.

A definição de todos os genótipos interrogados, seja como homozigoto ou heterozigoto, foi realizada avaliando visualmente a proporção e qualidade das *reads*, bem como a qualidade da chamada das bases em cada *read*. Quando não foi possível definir o alelo, este foi dado como um alelo *missing*.

5.4.2.3. Anotação de variantes

Após gerar o arquivo .vcf com as informações sobre as variantes encontradas, o processo de anotação e interpretação funcional destas variantes torna-se de extrema importância, uma vez que existem bancos de dados com informações sobre mutações que podem estar associadas a diferentes doenças ou até mesmo a variação fenotípica normal. A forma mais comum de anotação é baseada em vários bancos de dados públicos de variações, como por exemplo o dbSNP. Em termos de predição funcional de variantes, as ferramentas empregam diferentes abordagens que vão desde a análise de uma sequência simples (pequena) até uma região completa de um gene, gerando assim, informações sobre o impacto destas variantes na formação de proteínas (Kassahn *et al.*, 2014; Pabinger *et al.*, 2014).

Para esta finalidade foi empregado o uso do *web-server SNPnexus*, considerado uma ferramenta de anotação funcional robusta para avaliar o potencial significado de variantes candidatas detectadas por sequenciamento e apontar as isoformas de genes/proteínas que podem ser fenotipicamente importantes (Chelala *et al.*, 2009; Dayem Ullah *et al.*, 2012; Dayem Ullah *et al.*, 2013). Ao acessar o *website http://snp-nexus.org*, pode-se escolher entre diferentes anotações. Os parâmetros de anotação utilizados para gerar o relatório do *SNPnexus* deste trabalho foram:

- Regiões genômicas: RefSeq; Ensembl; UCSC
- Efeitos de SNPs não sinônimos na função proteica: SIFT (Kumar et al., 2009); PolyPhen (Adzhubei et al., 2013)
- Elementos regulatórios: Sítios de ligação de fatores de transcrição; *miRBASE* (KozomaraeGriffiths-Jones, 2014); *ENCODE* (*ENCODE*, 2012); *Roadmap Epigenomics Consortium* (Kundaje *et al.*, 2015).
- Conservação: PHAST (Hubisz et al., 2011); GERP++ (Davydov et al., 2010).
- Associação com diferentes fenótipos: *GAD* (Becker *et al.*, 2004); *ClinVar* (Landrum *et al.*, 2014); *COSMIC* (Forbes *et al.*, 2008); *NHGRI* (MacArthur *et al.*, 2017)
- Polimorfismos não codificantes: *CADD* (Kircher *et al.*, 2014); *fitCons* (Gulko *et al.*, 2015); *EIGEN* (Ionita-Laza *et al.*, 2016); *FATHMM* (Shihab *et al.*, 2013)

De modo a complementar as informações geradas pelo relatório do *SNPnexus*, foram também utilizados o *HaploReg* (para confirmar as informações obtidas pelo *SNPnexus*) (Ward e Kellis, 2016) e o *UCSC Genome Browser* (Casper *et al.*, 2018).

5.5. Análises estatísticas

Salvo indicação contrária, todas as análises estatísticas foram conduzidas no programa PLINK 1.9 (Purcell *et al.*, 2007), um programa de análise do genoma projetado para analisar associações entre genótipos e dados fenotípicos. Foram calculadas as frequências alélicas, genotípicas, hetorozigosidade e aderência ao equilíbrio de Hardy-Weinberg. Um mapa de desequilíbrio de ligação foi gerado para toda a extensão que compreende os genes *OCA2* e *HERC2*. Um mapa de desequilíbrio de ligação foi gerado para toda a extensão que compreende os genes *OCA2* e *HERC2*. Um mapa de desequilíbrio de ligação foi gerado para toda a extensão que compreende os genes *OCA2* e *HERC2* com o emprego do programa *Haploview*, utilizando-se os parâmetros default (Barrett *et al.*, 2005).

De modo a testar as associações com as diferentes categorias de características fenotípicas de pigmentação e estruturas secundárias presentes na íris, análises de regressão logística ajustadas para ancestralidade, e assim controlando para possíveis efeitos da estratificação populacional, foram realizadas. O ajuste para a ancestralidade foi realizado adicionando-se como covariáveis no modelo de regressão as estimativas de mistura individual fornecidas após as análises pelo programa *STRUCTURE*. As regressões foram feitas para cada polimorfismo independentemente, levando em consideração um modelo aditivo de herança para o alelo de menor frequência do polimorfismo analisado. A escolha pelo modelo aditivo de herança, também chamado de modelo genético livre, deve-se ao fato de que raramente há uma justificativa biológica para a escolha de um modelo de herança, seja ele dominante, co-dominante ou recessivo (Branicki *et al.*, 2009, Lewis, 2002, Minelli *et al.*, 2005). Portanto, para os efeitos aditivos dos polimorfismos, um *Odds Ratio* (*OR*) > 1 ou < 1 representa o efeito de cada alelo de menor frequência (MAF, do inglês Minor Allele Frequency), ou seja, um *OR* > 1 significa que o MAF aumenta o risco ao fenótipo proporcionalmente ao número de alelos MAF testados.

A linha de comando utilizada no terminal para realizar a análise de regressão logística controlando para a ancestralidade no programa PLINK 1.9 foi:

plink --bfile banco_de_dados --logistic --covar arquivo.cov --out
arquivo de saida

Onde:

--bfile especifica o arquivo de entrada no programa PLINK, isto é, o banco de dados formatado conforme especificado pelo manual do programa.

--logistic especifica para realizar uma regressão logística.

--covar especifica o arquivo formatado com as três covariaveis de estimativas de ancestralidade individual a serem inclusas no modelo, uma para ancestralidade europeia, africana e ameríndia.

--out especifica o nome do arquivo de saída do programa PLINK.

Após esta primeira análise de regressão, o polimorfismo que se mostrou mais associado, ou seja, com o menor valor de p, foi condicionado no modelo de regressão como uma covariável para testar se os outros polimorfismos que também foram associados exercem uma associação independente do polimorfismo condicionado. A linha de comando utilizada no terminal para realizar a análise de regressão logística simultaneamente controlando para a ancestralidade e condicionando para um determinado polimorfismo no programa PLINK 1.9 foi: plink --bfile banco_de_dados --logistic --covar arquivo.cov --condition
nome_polimorfismo --out arquivo_de_saída

Onde:

--condition especifica o nome do polimorfismo a ser condicionado.

A abordagem de análise condicional tem sido utilizada como uma ferramenta para identificar possíveis sinais secundários e independentes de associação (Psychiatric, 2011, Lango Allen *et al.*, 2010, Yang *et al.*, 2012).

Desta forma, se muitos dos polimorfismos que foram associados na primeira etapa da análise de regressão também exibirem um p significativo na segunda etapa, isto pode significar que eles tenham efeitos independentes na determinação do fenótipo em relação ao polimorfismo condicionado. Esta estratégia permite a descoberta de diferentes SNPs que possam estar associados com o fenótipo estudado (Galarneau *et al.*, 2010, Trynka *et al.*, 2011).

Para verificar a independência de associação entre os polimorfismos que continuaram associados após a primeira análise condicional, foram realizados testes de desequilíbrio de ligação par a par para cada um dos polimorfismos em todas as análises de associação. Adicionalmente, para os mesmos polimorfismos associados no presente estudo, foi verificado o desequilíbrio de ligação destes nas populações africanas, europeias e do leste asiático presentes no 1000Genomes, através da ferramenta LDlink acessada pelo site https://analysistools.nci.nih.gov/LDlink/, selecionando a aba "LDmatrix" (MachielaeChanock, 2015, MachielaeChanock, 2018).

Para as análises quantitativas da cor do olho (L*, a* e b*) e cor da pele (índice de melanina), foram realizadas regressões lineares ajustadas para ancestralidade no programa PLINK 1.9 para determinar a associação e o efeito de cada polimorfismo seguindo o mesmo fluxo de trabalho descrito para a análise de regressão logística, com exceção do parâmetro -- logistic, o qual foi alterado para --linear de modo a realizar uma análise de regressão linear. Todos os outros parâmetros da linha de comando permaneceram os mesmos.

O teste não-paramétrico pa*read*o de *Wilcoxon signed rank test with continuity correction* foi empregado para comparar as medianas de IM entre região exposta e não exposta ao sol de indivíduos heteroclassificados com pele clara, intermediária e escura, e também para avaliar se as medianas de L*a*b* diferem entre o olho direito e esquerdo. O teste de Kruskal-Wallis foi utilizado para comparar as medianas das estimativas de ancestralidade europeia, africana e ameríndia entre (1) os grupos de cor de pele clara, intermediária e escura, e também entre os grupos de categorias do IBGE (branco, amarelo, pardo e preto). Para os olhos, o teste foi utilizado para comparar as medianas do valor de b* entre os grupos heteroclassificados de acordo com a cor dos olhos (azul, verde, mel, castanho-claro e castanho-escuro). Os testes foram conduzido no ambiente estatístico R com as funções wilcox.test() e kruskal.test(), para os testes de *Wilcoxon signed rank test with continuity correction* e Kruskal-Wallis, respectivamente.

O teste de correlação de Spearman foi utilizado para verificar a correlação entre o IM e as estimativas de ancestralidade europeia, africana e ameríndia. Assim como os testes de Wilcoxon e Kruskal-Wallis, este foi conduzido no ambiente R com a função cor.test().

5.5.1. Análises de epistasia entre polimorfismos

Análises de epistasia (interações gene-gene) para testar se a associação de dois ou mais polimorfismos contribuem com um efeito aditivo maior que seus efeitos separadamente na determinação dos fenótipos foram executados no ambiente estatístico R (http://www.R-project.org/), com o emprego do pacote mb-mdr (Calle *et al.*, 2010), que utiliza a metodologia em mineração de dados conhecida como Redução da Dimensionalidade Multifatorial (MDR, do inglês *Multifactor Dimensionality Reduction*).

A abordagem MDR foi desenvolvida para identificar interações gene-gene, e seus possíveis efeitos na determinação de fenótipos quantitativos (Hahn *et al.*, 2003, Ritchie *et al.*, 2001).

Esta abordagem busca identificar combinações de genótipos multilocus que são associados a um alto risco para o fenótipo estudado, e combinações associadas a um baixo risco. Assim, o MDR reduz os dados e define uma única variável que incorpora a informação de diferentes loci que podem ser divididos como combinações de alto e baixo risco. Esta nova variável pode ser avaliada quanto à sua capacidade de classificar e prever o risco de apresentar o fenótipo estudado usando testes de permutação para avaliar a significância (Good, 2000). Valores de $p \le 0,05$ foram adotados como limiar de significância após os testes de permutação.

As análises de epistasia com o emprego do pacote mb-mdr foram realizadas nas variáveis quantitativas para cor dos olhos (L*,a* e b*), e variáveis quantitativas para o índice de melanina em região exposta e não exposta ao sol. Apenas os modelos de segunda ordem, ou seja, combinações par a par entre os genótipos de polimorfismos tagSNPs para os genes *OCA2* e *HERC2* foram analisados.

Para a obtenção dos *tagSNPs* o programa *PriorityPruner* (http:// prioritypruner.sourceforge.net/index.html), foi utilizado para selecionar *tagSNPs* a partir de uma lista de polimorfismos que estão em desequilíbrio de ligação, selecionando preferencialmente SNPs definidos como de maior prioridade, como, por exemplo, os SNPs com mais resultados significativos em um estudo de associação. Desta forma, o algoritmo do programa funciona removendo SNPs que estão em desequilíbrio de ligação com os SNPs mais importantes, ou seja, aqueles que apresentaram maior significância nos testes de regressão linear para cada uma das variáveis quantitativas analisadas no presente trabalho. O programa começa selecionando (mantendo) o SNP mais significativo (com o menor valor de p), calcula o desequilíbrio de ligação com SNPs adjacentes e designa os SNPs acima de um determinado R^2 como "marcado", selecionando os para eliminação. Para todas as cinco variáveis quantitativas um limiar acima de R^2 0,75 foi adotado para a obtenção dos *tagSNPs*.

5.5.2. Marcadores informativos de ancestralidade

As estimativas de contribuição ancestral nas amostras do presente estudo foram obtidas com um conjunto de marcadores informativos de ancestralidade presentes no SNP*for*ID 34-*plex* AIMs (Phillips *et al.*, 2007).

As estimativas de ancestralidade foram realizadas no programa *STRUCTURE* (Pritchard *et al.*, 2000), definindo modelo *admixture* com K=3 e considerando como referência um conjunto de 14 populações da terceira fase do projeto *1000Genomes* (total de 1.412 indivíduos). Estes indivíduos foram usados como amostras representativas das populações parentais para a análise de ancestralidade. As populações parentais foram divididas em três grandes grupos: 404 indivíduos europeus representando (1) finlandeses residentes na Finlândia (FIN), (2) britânicos da Inglaterra e Escócia (GBR), (3) habitantes da Toscana na Itália (TSI) e (4) e ibéricos da Espanha (IBS); 504 indivíduos africanos representando (1) Yoruba de Ibadan na Nigéria (YRI), (2) Quos Luhya de Webuye, Quênia (LWK), (3) gambianos da Divisão Oeste, Gâmbia (GWD), (4) Esan da Nigéria (ESN) e (5) os Mende de Serra Leoa (MSL); 504 indivíduos de populações representativas do leste asiático, compreendendo (1) chineses Han de Pequim (CHB), (2) japoneses de Tóquio (JPT), (3) Dai de Xishuangbanna, China (CDX), (4) chineses Han do sul da China (CHS) e (5) indivíduos de Kinh in Ho Chi Minh, Vietnã (KHV).

Uma vez que o projeto *1000Genomes* não inclui um grupo de ameríndios autóctones que pudesse apropriadamente representar a ancestralidade nativa americana, foram utilizadas as populações do leste asiático como substitutas para a população parental nativa americana. Usando um modelo de mistura, proporções de ancestralidades individuais referentes a cada uma das populações parentais (europeus, africanos e asiáticos) foram estimadas.

Uma análise de componente principal com os 34 AIMs do SNP*for*ID para estas mesmas populações foi realizada para verificar a distribuição das populações sobre os dois eixos (componentes) principais. Esta análise foi conduzida no ambiente R com o auxílio do pacote *SNPrelate* (Zheng *et al.*, 2012).

6 RESULTADOS

6.1. Caracterização populacional e fenótipos de pigmentação

A amostra populacional de estudo constituiu em 340 indivíduos, sendo que destes, 173 eram homens e 167 mulheres. Entretanto, oito amostras foram excluídas de todas as análises devido à alta taxa de genótipos não determinados, o que prejudicaria as análises genéticas subsequentes. Portanto, 332 amostras, sendo 168 homens e 164 mulheres, compuseram a amostragem final.

Algumas categorias para a descrição da cor da pele e cor dos cabelos foram agrupadas de modo a sumarizar melhor os resultados obtidos, uma vez que algumas destas mostraram-se com uma baixa representação amostral. Portanto, as cores de pele consideradas na EF como I e II foram classificadas como pele clara; as cores de pele consideradas na EF como III e IV foram classificadas como pele intermediária e por fim, as cores de pele consideradas na EF como V e VI foram classificadas como pele escura. A distribuição das características da cor da pele resultou em (a) 139 indivíduos (41,9%) classificados como pele clara, dos quais 30 (21,6%) com presença de sardas, (b) 137 indivíduos (41,3%) com pele intermediária, dos quais 10 (7,3%) com presença de sardas e (c) 56 indivíduos (16,9%) com pele escura, dos quais 3 (5,4%) apresentaram sardas (Figura 9). No geral, 43 dos indivíduos (13,0%) considerados apresentaram sardas.



Figura 9 - Distribuição dos indivíduos de acordo com a classificação da cor da pele e presença de sardas. O número de indivíduos dentro de cada categoria de cor de pele com presença de sardas está destacado entre parênteses.

Em relação à cor dos cabelos, as categorias loiro-claro e loiro-escuro foram agrupadas como cor de cabelo loiro, enquanto que castanho-claro e castanho-escuro também foram agrupadas como cor de cabelo castanho. A classificação da cor dos cabelos resultou em 10 indivíduos (3,1%) com cabelos ruivos, 48 (14,8%) com cabelos loiros, 181 (55,9%) com cabelos castanhos e 85 (26,2%) com cabelos pretos (Figura 10).



Figura 10 - Distribuição dos indivíduos de acordo com a classificação da cor de cabelo. O número de indivíduos dentro de cada categoria de cor de cabelo está destacado entre parênteses.

A distribuição das características da cor dos olhos resultou em (a) 36 indivíduos classificados com olho azul, (b) 60 indivíduos com olho verde, (c) 19 indivíduos com olho mel, (d) 71 indivíduos com olho castanho-claro, e (e) 146 indivíduos com olho castanho-escuro (Figura 11).



Figura 11 - Distribuição dos indivíduos de acordo com a classificação da cor dos olhos. O número de indivíduos dentro de cada categoria de cor dos olhos está destacado entre parênteses.

As características de estruturas secundárias presentes na íris ou esclera foram classificadas em presença ou ausência de estrutura, sendo elas: (1) nódulos de Wolfflin, (2) sulcos de contração, (3) anel de pigmentação na periferia da íris, (4) pigmentação na esclera e (5) distribuição de pontos de pigmentação na íris. A análise de quantificação da cor do olho e de identificação de estruturas secundárias presentes na íris foi realizada em 225 dos

332 indivíduos que compuseram a amostragem final. Isto porque algumas fotografias da íris não atingiram os limiares mínimos de qualidade e resolução para as análises. Portanto, os resultados abaixo que sumarizam a presença de estruturas secundárias na íris levam em consideração 225 indivíduos.

- A presença de nódulos de Wolfflin ocorreu em 17 indivíduos dentre os classificados com cor de olho azul (47%) e em 20 indivíduos com cor de olho verde (33%). No geral, 37 dos indivíduos (16,4%) considerados apresentaram nódulos de Wolfflin.
- A presença de sulcos de contração ocorreu em 1 indivíduo classificado com cor de olho azul, em 7 indivíduos com cor de olho verde, em 7 indivíduos classificados com cor de olho mel, em 24 indivíduos classificados com cor de olho castanho-claro e em 22 indivíduos classificados com cor de olho castanho-escuro.
- A presença de anel de pigmentação na periferia da íris ocorreu apenas em 11 indivíduos classificados com cor de olho castanho-escuro.
- A presença de pigmentação na esclera ocorreu em 2 indivíduos classificados com cor de olho azul, em 1 indivíduo classificado com cor de olho verde, em 1 indivíduo classificado com cor de olho mel, em 5 indivíduos classificados com cor de olho castanho-claro e em 30 indivíduos classificados com cor de olho castanho-escuro.
- A presença de pontos de pigmentação na íris ocorreu em 13 indivíduos classificados com cor de olho azul, 24 indivíduos classificados com cor de olho verde, em 9 indivíduos classificados com cor de olho mel, em 9 indivíduos classificados com cor de olho castanho-claro e em 2 indivíduos classificados com cor de olho castanhoescuro.



Figura 12 - Distribuição dos indivíduos de acordo com a presença (Sim) ou ausência (Não) de estruturas secundárias presentes na íris dentro das categorias de cor dos olhos. **A:** Nódulos de Wolfflin; **B:** Sulcos de contração; **C:** Anel de pigmentação; **D:** Pigmentação na esclera; **E:** Pontos de pigmentação na íris. O número de indivíduos com a presença de cada estrutura secundária entre as categorias de cor do olho está destacado entre parênteses.

Além das frequências de características fenotípicas pigmentares, as frequências de classificação baseada nos critérios adotados pelo IBGE (Instituto Brasileiro de Geografia e Estatística) estão representadas na Figura 13. De acordo com os critérios adotados pelo IBGE, os indivíduos foram heteroclassificados como brancos, amarelos, pardos, e pretos (214, 11, 63, 44, respectivamente). Quanto à autoclassificação, 212 se consideraram brancos, 1 indígena, 12 amarelos, 65 pardos e 42 indivíduos pretos.



Figura 13 - Distribuição da autoclassificação e heteroclassificação dos indivíduos de acordo com as categorias do IBGE. O número de indivíduos dentro de cada categoria está destacado entre parênteses.

6.1.1. AIMs e controle da estratificação populacional

A Figura 14 mostra a distribuição da ancestralidade genômica de cada indivíduo em três agrupamentos (*clusters*) definidos por meio dos marcadores informativos de ancestralidade contidos no SNP*for*ID 34-*plex*. A Tabela 2 mostra os valores de atribuição das populações a cada um dos agrupamentos gerados pelo *STRUCTURE*. As populações africanas, europeias e do leste asiático do *1000Genomes* foram agrupadas majoritariamente nos *clusters* 3, 1 e 2 respectivamente.

Observa-se que as amostras de Ribeirão Preto (RBP) possuem 71,4% de componente designado como europeu; 9% do componente asiático/ameríndio (representado pelas populações do leste asiático) e 19,6% de contribuição africana.

A Figura 15 mostra a representação dos dois primeiros eixos de uma análise de PCA incluindo indivíduos da amostra RBP, bem como as amostras AFR, EUR e EAS do projeto *1000Genomes*. O objetivo desta análise foi verificar visualmente a distribuição dos indivíduos da amostra RBP nos agrupamentos referentes a AFR, EUR e EAS. A maioria dos indivíduos da amostra RBP se sobrepõe aos indivíduos da amostra EUR, com um gradiente em direção à amostra AFR; poucos indivíduos se sobrepõem à amostra EAS.



Figura 14 - Estrutura populacional da amostra de Ribeirão Preto (RBP) em comparação com populações parentais (EUR, EAS e AFR), obtido a partir do conjunto de AIMs contidos no SNP*for*ID 34-*plex*. Cada barra vertical representa a distribuição da ancestralidade genômica nos três agrupamentos inferidos pelo *STRUCTURE*.

Tabela 2 - Proporção de cada população atribuída a cada um dos três *clusters* definidos pelo programa *STRUCTURE*.

	Agrupamentos inferidos				
População	Europeu	Asiático/Ameríndio	Africano		
1 (AFR)	0,004	0,002	0,994		
2 (EUR)	0,981	0,010	0,008		
3 (EAS)	0,003	0,996	0,001		
4 (RBP)	0,714	0,090	0,196		

EUR = Populações europeias do *1000Genomes*; AFR = Populações africanas do *1000Genomes*; EAS = Populações do leste asiático do *1000Genomes*; RPB = Amostras do presente estudo coletadas em Ribeirão Preto - SP.



Figura 15 - Componentes principais obtidos a partir do conjunto de AIMs (SNP*for*ID 34-*plex*) que relacionam as frequências genotípicas da população analisada (RBP) com populações mundiais (EUR, AFR e EAS). Estes dois primeiros eixos explicam 29,78% e 19,33% (PC1 e PC2, respectivamente), da variância observada.

6.2. Análise quantitativa da cor da íris

As análises quantitativas para obter valores contínuos que permitem a caracterização objetiva da cor do olhos e a identificação de estruturas secundárias presentes na íris foram realizadas em imagens fotográficas de 225 indivíduos dos 332 que compuseram o estudo (Figuras 16 e 17). Devido a qualidade de algumas imagens não foi possível identificar a estrutura secundária conhecida como criptas de Fuchs e, portanto, não foi realizada a identificação e associação destas com os polimorfismos genéticos.



Figura 16 - Imagens fotográficas representativas das íris incluídas no estudo: olho azul, verde, mel, castanho-claro e castanho-escuro, respectivamente.



Figura 17 - Estruturas secundárias presentes na íris e esclera identificadas no presente estudo. **A**: Pontos de pigmentação na íris. **B**: Presença de pigmentação na esclera. **C**: Nódulos de Wolfflin na periferia da íris. **D**: Anel de pigmentação ao redor da íris. **E**: Sulcos de contração. As estruturas estão apontadas pelas flechas vermelhas.

6.2.1. Comparações entre as íris direita e íris esquerda

Tendo em vista que as análises quantitativas para cor dos olhos foram conduzidas nos dois olhos de cada um dos 225 indivíduos incluídos, o teste pareado de *Wilcoxon signed rank test with continuity correction* foi utilizado para verificar se há diferenças nas medianas dos valores de L*, a* e b* entre os olhos direitos e esquerdos.

Os testes revelaram diferenças significativas nas três dimensões (L*, p = 0,01; a*, p = 0,0001; b*, p = 0,0001), evidenciando diferenças nas medianas entre os valores de L*a*b* quando comparado entre os dois olhos.

Desta forma, as análises de associação genética foram realizadas levando em consideração os valores de L*a*b* para apenas um dos olhos, o direito.

6.2.2. Distribuição dos valores L*a*b*

Quando plotados em um gráfico de três dimensões, evidencia-se uma substancial dispersão dos valores de L*a*b* entre as diferentes íris analisadas no presente estudo (Figuras 18). Os gráficos bi-dimensionais (L* vs. a*, L* vs. b* e b* vs. a*), podem ser visualizados no Apêndice C.



Figura 18 - Distribuição das íris nas coordenadas do espaço de cores L*a*b*. Cada círculo representa a íris direita de um indivíduo. A cor dentro de cada círculo equivale à média da cor da íris capturada pela porção extraída pelo algoritmo no espaço de cores L*a*b*.

De modo geral, olhos azuis tendem a mostrar maiores valores de L*, valores de a* próximos a 0 e valores de b* negativos, enquanto que olhos castanhos tendem a menores valores de L*, e maiores valores de a* e b*. Olhos castanho-escuros possuem menores valores de L*a*b* comparados aos olhos castanho-claros. Alguns indivíduos com olhos verdes obtiveram valores sobrepostos aos dos indivíduos com olhos azuis e indivíduos com olhos mel tiveram valores sobrepostos àqueles dos indivíduos com castanho-claros.

Ao plotar a heteroclassificação realizada pela equipe do LPFG para cada um dos 225 indivíduos no mesmo plano tridimensional L*a*b*, observa-se que a heteroclassificação obteve elevada concordância com os valores dispostos no plano L*a*b*, com agrupamento de olhos azuis e verdes nos valores que vão de 0 a valores negativos de b*, e olhos castanhos sendo distribuídos em sua maioria nos valores de 0 a positivos de b* (Figura 20).



Figura 19 - Distribuição das íris nas coordenadas do espaço de cores L*a*b*. Cada círculo representa a íris direita de um indivíduo. A cor dentro de cada círculo equivale à heteroclassificação realizada pela equipe do LPFG, considerando cores azul, verde, mel, castanho-claro e castanho-escuro.

Dentre as três dimensões CIELAB (L*, a* e b*), a variação da cor dos olhos foi mais perceptível e melhor explicada pela dimensão cromática b* (Figura 20). Quando b* foi comparado entre pares de grupos heteroclassificados para a cor dos olhos, observa-se que, exceto para olhos mel *vs.* castanho-claros (ns, p > 0,05), todos os outros grupos tiveram diferenças significativas quanto às suas medianas. Os gráficos para as dimensões L* e a* podem ser visualizados no Apêndice D.



Figura 20 - Diferenciação das medianas do valor de b* entre os grupos heteroclassificados de acordo com a cor dos olhos. Valor de p global do teste de Kruskal-Wallis no topo do gráfico. Linha pontilhada demarca a mediana de b* da amostra total. ns = não significativo; p > 0,05; ** = p < 0,01; **** = p < 0,0001.

A Figura 21 mostra as estimativas de ancestralidade entre os diferentes grupos para cor dos olhos avaliados categoricamente. A ancestralidade europeia foi a que apresentou as maiores proporções em todos os grupos, independentemente das cores de olhos, com estimativas variando de 49% a 92%. Em segundo, a ancestralidade africana apresentou valores que vão de 4% a 29% e, por fim, a ancestralidade ameríndia com valores de 3% a 7%.

O teste de Kruskal-Wallis para a diferenciação das medianas de ancestralidade europeia entre os pares de cor de olhos não resultou em diferenciação significativa quando comparado os grupos (1) olho azul vs. olho verde, (2) olho verde vs. olho mel, (3) olho mel vs. olho castanho claro e (4) olho verde vs. olho castanho claro. Todas as demais comparações evidenciaram diferenças significativas em termos de ancestralidade europeia.

Em relação à ancestralidade africana, não houve diferença nas estimativas quando comparados os grupos (1) olho azul *vs.* olho verde, (2) olho mel *vs.* olho castanho claro, (3) olho verde *vs.* olho castanho claro e (4) olho azul *vs.* olho castanho claro. Todas as demais comparações evidenciaram diferenças significativas em termos de ancestralidade africana.

Por fim, para as comparações envolvendo a ancestralidade ameríndia, não houveram diferenças entre os grupos (1) olho azul vs. olho verde, (2) olho verde vs. olho mel, (3) olho mel vs. olho castanho claro, (4) olho azul vs. olho mel, (5) olho verde vs. olho mel e (6) olho



azul *vs.* olho castanho claro. Apenas as comparações envolvendo olhos castanho-escuros apresentaram diferenças significativas em termos de ancestralidade ameríndia.

Figura 21 - Contribuição ancestral entre os indivíduos que apresentam olhos azuis, verdes, mel, castanho-claros e castanho-escuros. **A:** Ancestralidade europeia; **B:** Ancestralidade africana; **C:** Ancestralidade ameríndia; Valor global de p para o teste de Kruskal-Wallis no top do gráfico; ns = não significativo, p > 0.05; * = p < 0.05; ** = p < 0.01; *** = p < 0.001; **** = p < 0.001. **D:** Valores das medianas de estimativas de ancestralidade plotadas nos gráficos anteriores.

6.3. Análises entre os índices de melanina

6.3.1. Índice de melanina *vs.* cor de pele e classificação do IBGE

A variação índice de melanina em região não exposta ao sol e exposta ao sol foi avaliada contra os diferentes fenótipos de pigmentação de pele (clara, intermediária e escura) (Figuras 22), e categorias de classificação do IBGE (Figuras 23), buscando encontrar diferenças entre os grupos.



Figura 22 - Variação do índice de melanina nos indivíduos de pele clara, intermediária e escura, em área exposta e não exposta ao sol. Valor de probabilidade para o teste de *Wilcoxon signed rank test with continuity correction* acima dos *boxplots* correspondendo a comparações dos índices entre área exposta e não exposta para cada um dos três tipos de pele.



Figura 23 - Variação do do índice de melanina nos indivíduos classificados como branco, amarelo, pardo e preto de acordo com as categorias do IBGE, em área exposta e não exposta ao sol. Valor de probabilidade para o teste de Kruskal-Wallis acima dos *boxplots* correspondendo a comparações dos índices entre área exposta e não exposta para cada uma das quatro categorias.

Um aumento do índice de melanina pode ser observado entre indivíduos de pele clara para os indivíduos de pele escura, bem como de indivíduos classificados como brancos para os classificados como pretos. O teste pa*read*o de *Wilcoxon signed rank test with continuity correction* foi aplicado para verificar se há diferença nas medianas entre área não exposta e exposta ao sol para indivíduos classificados segundo a cor da pele (clara, intermediária e escura). Todas as comparações resultaram em diferenças significativas e, portanto, aceita-se a hipótese alternativa de que há diferença nas medianas entre área exposta e não exposta ao sol, como consequência do bronzeamento.

Dentre os indivíduos classificados pelas categorias do IBGE, há sobreposição da distribuição do índice de melanina entre os indivíduos classificados como brancos e amarelos, o que ocasionou ausência de diferenciação para esta comparação (p > 0,05).

O teste de correlação de Spearman foi realizado para observar a correlação entre o índice de melanina e a ancestralidade estimada por meio do SNP*for*ID 34-*plex* (Figura 24).

Observa-se uma forte correlação negativa entre a ancestralidade europeia e o índice de melanina (r = -0,69, $p = 2,2x10^{-16}$). À medida que a ancestralidade europeia dos indivíduos torna-se maior o índice de melanina diminui, ou seja, a pele apresenta-se mais clara. O contrário pode ser visto para a forte correlação positiva entre a ancestralidade africana e o índice de melanina (r = 0,70, $p = 2,2x10^{-16}$). À medida que a ancestralidade africana dos indivíduos aumenta, aumentam-se os níveis de melanina, ou seja, a pele apresenta-se mais escura. Em compensação, para a ancestralidade ameríndia, observa-se fraca correlação positiva (r = 0,29, $p = 5,6x10^{-8}$) quando comparado às correlações anteriores.



Figura 24 - Correlação entre ancestralidade e índice de melanina para área não exposta ao sol. **A:** Ancestralidade europeia; **B:** Ancestralidade africana; **C:** Ancestralidade ameríndia.

A contribuição ancestral dos indivíduos de pele clara foi de 93% europeia, 4% africana e 3% ameríndia. Para indivíduos de pele intermediária, a composição ancestral foi de 75% europeia, 19% africana e 6% ameríndia. Indivíduos com pele escura apresentaram valores de contribuição europeia em 34%, africana 60% e ameríndia 6% (Figura 25).

O teste de Kruskal-Wallis para a diferenciação das medianas par a par entre os grupos não foi significativo apenas na comparação de pele intermediária vs. pele escura para ancestralidade ameríndia (Figura 25C)



Pele 🛱 Clara 🛱 Intermediária 🛱 Escura

Figura 25 - Contribuição ancestral entre os indivíduos que apresentam pele clara, intermediária e escura. A: Ancestralidade europeia; **B:** Ancestralidade africana; **C:** Ancestralidade ameríndia; Valor global de *p* para o teste de Kruskal-Wallis no top do gráfico; ns = não significativo; p > 0.05; *** = p <= 0.001; **** = p <= 0.0001; D: Valores das estimativas de ancestralidade entre os grupos.

A contribuição ancestral dos indivíduos autoclassificados de acordo com as categorias do IBGE foi de (a) 89% europeia, 8% africana e 3% ameríndia para indivíduos classificados como brancos, (b) 14% europeia, 4% africana e 82% ameríndia para indivíduos classificados como amarelos, (c) 64% europeia, 29% africana e 7% ameríndia para indivíduos classificados como pardos e (d) 28% europeia, 66% africana e 6% ameríndia para indivíduos classificados como pretos (Figura 26). O único indivíduo que se autoclassificou como indígena possui uma estimativa de ancestralidade em 72% europeia, 14% africana e 12% ameríndia. Por apresentar apenas uma observação a categoria indígena não foi apresentada nos gráficos das Figuras 23 e 26. Não houve diferenciação significativa entre as medianas de indivíduos brancos e amarelos para ancestralidade africana (Figura 26B) e, entre pardos e pretos para a ancestralidade ameríndia (Figura 26C).



Figura 26 - Contribuição ancestral entre os indivíduos autoclassificados de acordo com as categorias do IBGE em branco, amarelo, pardo e preto. Valor global de p para o teste de Kruskal-Wallis no top do gráfico; * = $p \le 0.05$; **** = $p \le 0.0001$; ns = não significativo, p > 0.05.

6.4. Identificação das variantes genéticas

A determinação de variantes pela rotina HC encontrou 749 variantes ao longo das regiões cobertas (vide seção 5.2, Figura 8) pelo ensaio. Visando diminuir a taxa de falsopositivos, o arquivo VCF gerado foi submetido à rotina do script VCFx (Castelli *et al.*, 2015), que interroga os genótipos com base na quantidade e proporção de *reads* para determinação de variantes.

Esta avaliação é necessária uma vez que alguns dos genótipos considerados pelo HC podem não ser verdadeiros, principalmente quando a genotipagem se refere a um segmento com baixa cobertura de sequenciamento ou quando, no caso de genótipos heterozigotos, um dos alelos é pouco representado.

Após a aplicação da rotina VCFx, um filtro para a remoção de variantes com MAF (do inglês, *Minor Allele Frequency*) inferiores a 0,01 e máximo de 5% de genótipos não determinados permitido por indivíduo foi utilizado com o emprego do programa PLINK 1.9. A aplicação dos filtros resultou na retirada de oito amostras do total de 340 coletadas e 579 variantes. Portanto, foram avaliadas 170 variantes ao longo dos genes *OCA2* e *HERC2* (Apêndice B) e sua distribuição pode ser vista na Tabela 3.

Com o emprego do programa *Haploview*, um mapa de desequilíbrio de ligação (DL) foi gerado para os 170 polimorfismos que foram analisados neste trabalho. Nove blocos de haplótipos são observados, sendo três no gene *OCA2* e cinco no *HERC2* (Figura 27).

Regiões/Genes	OCA2	HERC2	
Promotora	14	7	
5'UTR	3	0	
CDS	12	24	
Íntrons*	15	90	
3'UTR	1	4	
	45	125	

Tabela 3 - Distribuição dos 170 pontos de variação nas regiões dos genes OCA2 e HERC2.

CDS: do inglês, *Coding Sequence*; *: Regiões intrônicas flanqueadoras aos éxons cobertas pelo ensaio *HaloPlex*.



Figura 27 - Representação do desequilíbrio de ligação dos genes *OCA2* e *HERC2*. Os blocos em desequilíbrio estão cercados em preto. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel *et al.* (2002). No canto inferior esquerdo estão os nove blocos representados, sendo o três primeiros pertencentes ao gene *OCA2* e os outros seis ao gene *HERC2*. No mapa acima, vermelho indica LOD (*Log of Odds*) \ge 2 e D' = 1; azul indica LOD < 2 e D' = 1; tons de rosa indicam LOD \ge 2 e D' < 1; branco indica LOD < 2 e D' < 1.
6.4.1. Associação genética com fenótipos de pigmentação

Primeiramente foram realizadas análises de regressão logística utilizando como covariáveis no modelo somente as estimativas de mistura para controle da estratificação populacional. Sessenta e quatro das 170 variantes foram associadas com algum fenótipo de pigmentação. De modo a identificar possíveis efeitos independentes entre as variantes encontradas e os fenótipos de pigmentação, uma nova análise de regressão logística, desta vez condicionando a análise ao polimorfismo com o menor valor de p resultante da análise anterior. Após esta análise, dos 64 polimorfismos encontrados com alguma associação fenotípica de pigmentação, um menor conjunto de 13 variantes foram associadas com um ou mais fenótipos de pigmentação, sendo (a) nove delas a cor do olho, (b) três a cor dos cabelos, (c) três a cor da pele e (d) uma a presença de sardas.

O DL foi verificado entre as variantes que se mostraram associadas a um mesmo fenótipo de pigmentação, a fim de confirmar a independência das associações resultantes da análise de regressão logística anterior. Foi gerado um mapa de desequilíbrio de ligação para cada uma das combinações par a par entre as variáveis categóricas de pigmentação. Todas as comparações par a par evidenciaram um baixo e/ou nulo DL (Apêndice C). Portanto, ao final de todas as análises de DL e após a consequente constatação de independência entre as variantes associada a um mesmo fenótipo, conclui-se pela real associação de todas as 13 variantes relatadas no parágrafo anterior. A distribuição das associações pelas diferentes regiões gênicas está apresentada na Tabela 4, e o detalhamento das associações está apresentado nas Tabelas 5, 6 e 7. Distribuições das frequências alélicas e genotípicas, heterozigozidade e aderência ao equilíbrio de Hardy-Weinberg, estão apresentadas no Apêndice B.

	OCA2	HERC2	Total	
Promotora	-	-	-	
5'UTR	-	-	-	
CDS	4	5	9	
Íntrons	2	1	3	
3'UTR	-	1	1	
Total	6	7	13	

Tabela 4 - Distribuição dos polimorfismos nas regiões dos genes *OCA2* e *HERC2*, cujos alelos encontram-se associados com algum fenótipo de pigmentação.

18	1 I I I I I I I I I I I I I I I I I I I	I I I I I I I I I I I I I I I I I I I		Pro de la composición de la composicinde de la composición de la composición de la composición de la c			0		
Associação		rsID (posição)	A1	Odds Ratio [IC 95%]	р	RBP	EUR	AFR	EAS
Olho azul									
	<i>v</i> s. não azul	rs12913832 (28365618)	G	8,328 [3,47 - 19,95]	1,97x10 ⁻⁶	0,417	0,866	0,005	0,002
	vs. mel	rs12913832 (28365618)	А	0,088 [0,02 - 0,32]	2,0x10 ⁻⁴	0,582	0,134	0,995	0.998
	<i>vs.</i> castanho claro	rs12913832 (28365618)	А	0,023 [0,00 - 0,08]	2,44x10 ⁻⁸	0,582	0,134	0,995	0,998
	<i>vs.</i> castanho escuro	rs1129038 (28356859)	Т	17,96 [5,75 - 56,05]	6,53x10 ⁻⁷	0,420	0,863	0,005	0,001
Olho verde									
	vs. mel	rs12913832 (28365618)	А	0,206 [0,08 - 0,51]	0,0006	0,582	0,134	0,995	0,998
	<i>vs.</i> castanho claro	rs12913832 (28365618)	А	0,077 [0,03 - 0,18]	1,11x10 ⁻⁸	0,582	0,134	0,995	0,998
		<u>rs12913832</u> (28365618)	А	0,035 [0,02 - 0,06]	3,91x10 ⁻⁹	0,582	0,134	0,995	0,998
	<i>vs.</i> castanho escuro	rs61756152 (28412872)*	А	4,497 [1,33 - 15,21]	0,0155	0,109	0,055	0,091	0,295
		rs1800407 (28230318)*	Т	10,38 [1,50 - 71,68]	0,0176	0,051	0,047	0,000	0,001
Olho mel									
	<i>vs.</i> castanho claro	rs12910433 (28228644)	G	5,949 [1,98 - 17,81]	0,0014	0,497	0,247	0,906	0,846
		<u>rs1129038</u> (28356859)	Т	7,641 [2,52 - 23,09]	0,0003	0,420	0,863	0,005	0,001
	<i>vs.</i> castanho escuro	rs1800407 (28230318)*	Т	8,291 [1,68 - 40,88]	0,0093	0,051	0,047	0,000	0,001
		rs1800419 (28096538)*	А	2,976 [1,17 - 7,54]	0,0215	0,491	0,639	0,376	0,560
Olho castanho claro									
		<u>rs12913832</u> (28365618)	G	2,798 [1,47 - 5,30]	0,0016	0,417	0,866	0,005	0,002
	<i>vs.</i> castanho escuro	rs201872292 (28475625)*	А	2,944 [1,41 - 6,11]	0,0037	0,096	0,055	0,041	0,294
		rs1800401 (28260053)*	А	0,377 [0,15 - 0,93]	0,0356	0,082	0,037	0,124	0,005
Olho castanho escuro									
		<u>rs12913832</u> (28365618)	G	0,182 [0,11 - 0,30]	2,63x10 ⁻¹¹	0,417	0,866	0,005	0,002
		rs201872292 (28475625)*	А	0,356 [0,17 - 0,71]	0,0038	0,096	0,055	0,041	0,294
	<i>vs.</i> não castanho escuro	rs1800401 (28260053)*	А	2,921 [1,33 - 6,41]	0,0075	0,082	0,037	0,124	0,005
		rs74005645 (28447282)*	Т	3,720 [1,40 - 9,86]	0,0083	0,120	0,001	0,536	0,005
		rs1800419 (28096538)*	А	0,597 [0,36 - 0,97]	0,0403	0,491	0,639	0,376	0,560

Tabela 5 - Associações genéticas encontradas envolvendo alelos de nove polimorfismos analisados nos genes *OCA2* e *HERC2* com a pigmentação dos olhos. Valores de p significativos após correção de Bonferroni para múltiplos testes estão destacados em negrito.

Sublinhado = Polimorfismo condicionado para verificação de sinais independentes; * = Polimorfismo associado após análise condicional ao polimorfismo com maior associação (menor valor de *p*); rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; A1 = Alelo de menor frequência (MAF); RBP = Frequência alélica na população de estudo; EUR = Frequência alélica nas populações europeias do *1000Genomes*; AFR = Frequência alélica nas populações do leste asiático do *1000Genomes*;

EUR Associação rsID (posição) A1 Odds Ratio [IC 95%] RBP AFR EAS р Cabelo ruivo vs. loiro rs11636232 (28386626) Т 0,103 [0,01 - 0,95] 0,0450 0,191 0,450 0,002 0,000 **Cabelo loiro** rs12913832 (28365618) G 4,996 [2,67 - 9,33] 4.57x10-7 0.417 0.866 0.005 0.002 vs. castanho rs1800419 (28096538)* А 0,5347 [0,30 - 0,92] 0,491 0,639 0,376 0,560 0,0256 rs12913832 (28365618) G 8,535 [2,43 - 29,88] 0.0007 0.417 0.866 0,005 0.002 vs. preto

Tabela 6 - Associações genéticas encontradas envolvendo alelos de três polimorfismos analisados nos genes *OCA2* e *HERC2* com a pigmentação dos cabelos. Valores de *p* significativos mesmo após correção de Bonferroni para múltiplos testes estão destacados em negrito.

Sublinhado = Polimorfismo condicionado para verificação de sinais independentes; * = Polimorfismo associado após análise condicional ao polimorfismo com maior associação (menor valor de *p*); rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; A1 = Alelo de menor frequência (MAF); RBP = Frequência alélica na população de estudo; EUR = Frequência alélica nas populações europeias do *1000Genomes*; AFR = Frequência alélica nas populações do leste asiático do *1000Genomes*; AFR = Frequência alélica nas populações do leste asiático do *1000Genomes*; AFR = Frequência alélica nas populações do leste asiático do *1000Genomes*;

Tabela 7 - Associações genéticas encontradas envolvendo alelos de quatro polimorfismos analisados nos genes *OCA2* e *HERC2* com a pigmentação da pele e presença de sardas.

Associação		rsID (posição)	A1	Odds Ratio [IC 95%]	p	RBP	EUR	AFR	EAS
Pele clara					-				-
		<u>rs41304383</u> (28116282)	С	4,200 [1,50 - 11,70]	0,0060	0,046	0,050	0,059	0,000
	<i>vs.</i> não clara	rs12913832 (28365618)*	G	1,730 [1,16 - 2,57]	0,0044	0,417	0,866	0,005	0,002
		rs1800404 (28235773)*	Т	1,618 [1,02 - 2,54]	0,0345	0,594	0,842	0,106	0,386
		<u>rs41304383</u> (28116282)	С	3,983 [1,42 - 11,12]	0,0083	0,046	0,050	0,059	0,000
	vs. intermediária	rs12913832 (28365618)*	G	1,795 [1,20 - 2,68]	0,0044	0,417	0,866	0,005	0,002
		rs1800404 (28235773)*	Т	1,617 [1,03 - 2,52]	0,0345	0,594	0,842	0,106	0,386
Presença de sardas									
	<i>vs.</i> ausência	rs118112076 (28474686)	С	4,357 [1,18 - 16,06]	0,0270	0,018	0,016	0,000	0,000

Sublinhado = Polimorfismo condicionado para verificação de sinais independentes; * = Polimorfismo associado após análise condicional ao polimorfismo com maior associação (menor valor de *p*); rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; A1 = Alelo de menor frequência (MAF); RBP = Frequência alélica na população de estudo; EUR = Frequência alélica nas populações europeias do *1000Genomes*; AFR = Frequência alélica nas populações do leste asiático do *1000Genomes*; EAS = Frequência alélica nas populações do leste asiático do *1000Genomes*.

6.4.2. Associação genética com estruturas secundárias na íris

A análise de associação das variantes genéticas com estruturas secundárias da íris foram conduzidas utilizando o mesmo fluxo de trabalho proposto para a análise de associação genética com fenótipos de pigmentação descritos no tópico anterior.

Após a regressão logística ajustada apenas para ancestralidade, 43 das 170 variantes se mostraram associadas com algum fenótipo de estrutura secundária presente na íris.

Utilizando uma segunda etapa de regressão condicionando para o polimorfismo com o menor valor de *p* para evidenciar possíveis sinais independentes que poderiam contribuir para o aparecimento dos diferentes fenótipos, duas variantes foram associadas. O polimorfismo rs12913832A>G foi associado com três fenótipos de estruturas presentes na íris: (a) pontos de pigmentação na íris, (b) nódulos de Wolfflin e (c) sulcos de contração. O polimorfismo rs58358300A>C, também localizado no gene *HERC2* foi associado com pigmentação na esclera. O detalhamento das associações está descrito na Tabela 8.

Tabela 8 - Associações genéticas encontradas envolvendo alelos de oito polimorfismos analisados nos genes *OCA2* e *HERC2* com características de estruturas secundárias presentes na íris. Valores de *p* significativos mesmo após correção de Bonferroni para múltiplos testes estão destacados em negrito.

Associação	rsID (posição)	A1	Odds Ratio [IC 95%]	p	RBP	EUR	AFR	EAS
Pontos de pigmentação na íris	rs12913832 (28365618)	G	2,406 [1,46 - 3,96]	0,0005	0,417	0,866	0,005	0,002
Nódulos de Wolfflin	rs12913832 (28365618)	G	3,661 [1,66 - 7,20]	0,0001	0,417	0,866	0,005	0,002
Pigmentação na esclera	rs58358300 (28419393)	С	9,247 [2,97 - 28,79]	0,0001	0,104	0,001	0,553	0,002
Sulcos de contração	rs12913832 (28365618)	G	0,477 [0,30 - 0,75]	0,0013	0,417	0,866	0,005	0,002

rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; A1 = Alelo de menor frequência (MAF); RBP = Frequência alélica na população de estudo; EUR = Frequência alélica nas populações europeias do *1000Genomes*; AFR = Frequência alélica nas populações africanas do *1000Genomes*; EAS = Frequência alélica nas populações do leste asiático do *1000Genomes*;

6.4.3. Associação genética com os valores quantitativos de CIELAB

As análises de associação genética para os valores de L*a*b* extraídos por imagem fotográficas da íris, foram realizadas por meio de regressão linear utilizando um modelo aditivo de herança para o MAF dos polimorfismos.

Em uma primeira análise, (a) 55 polimorfismos foram associados à dimensão L*, (b) 16 polimorfismos foram associados à dimensão a* e (c) 35 polimorfismos associados à dimensão b*. Após análise condicional para o polimorfismo de menor valor de p, (a) 2 polimorfismos foram associados a dimensão L*, (b) 1 polimorfismo foi associado com a dimensão a* e (c) 2 polimorfismos foram associados a dimensão b* (Tabela 8). Os três polimorfismos estão localizados no gene HERC2, sendo dois em regiões de íntrons e um no éxon 31.

Associação	rsID (posição)	A1	Beta [IC 95%]	р	RBP	EUR	AFR	EAS
τ*	<u>rs12913832</u> (28365618)	G	3,745 [2,18 - 5,30]	4,67x10 ⁻⁶	0,417	0,866	0,005	0,002
L	rs6497271 (28365431)*	А	-5,498 [-8,352,65]	1,96x10⁴	0,066	0,037	0,277	0,005
a*	rs12913832 (28365618)	G	-2,607 [-3,811,40]	3,39x10 ⁻⁵	0,417	0,060	0,005	0,020
L *	<u>rs12913832</u> (28365618)	G	-5,517 [-6,804,22]	6,46x10 ⁻¹⁵	0,417	0,866	0,005	0,002
D	rs201872292 (28475625)*	А	3,950 [1,61 - 6,28]	0,0010	0,086	0,018	0,041	0,294

Tabela 9 - Associações genéticas encontradas envolvendo alelos de três polimorfismos analisados nos genes *OCA2* e *HERC2* com as dimensões L*a*b* do espaço de cores CIELAB.

Sublinhado = Polimorfismo condicionado para verificação de sinais independentes; * = Polimorfismo associado após análise condicional ao polimorfismo com maior associação (menor valor de *p*); rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; A1 = Alelo de menor frequência (MAF); RBP = Frequência alélica na população de estudo; EUR = Frequência alélica nas populações europeias do *1000Genomes*; AFR = Frequência alélica nas populações do leste asiático do *1000Genomes*; AFR = Frequência alélica nas populações do leste asiático do *1000Genomes*.

6.4.4. Associação genética com índice de melanina

Foram realizadas análises de regressão linear para associações genéticas e o índice de melanina para a região não exposta e exposta ao sol. A análise de regressão linear para associações genéticas e o índice de melanina da pele para a região não exposta ao sol evidenciou em uma primeira análise 28 polimorfismos (do total de 170 analisados), associados. Após condicionar para o polimorfismo com o menor valor de p em busca de sinais independentes em uma segunda análise, quatro polimorfismos mantiveram sua associação. Três polimorfismos estão localizados no gene OCA2, e um no gene HERC2. Para o gene OCA2, os polimorfismos estão localizados respectivamente em um íntron, éxon e região promotora. Para o gene HERC2, os dois polimorfismos estão localizados em íntrons.

Para a região exposta ao sol, em uma primeira análise evidenciou-se 13 polimorfismos (do total de 170 analisados), associados. Dois destes polimorfismos foram associados previamente com o índice de melanina para região não exposta ao sol, e uma nova associação no éxon 2 do gene *OCA2* foi observada (Tabela 10).

Tabela 10 - Associações genéticas encontradas envolvendo alelos de quatro polimorfismos analisados nos genes *OCA2* e *HERC2* com o índice de melanina em região não exposta e exposta ao sol.

Associação	rsID (posição)	A1	Beta [IC 95%]	р	RBP	EUR	AFR	EAS
	<u>rs1800414</u> (28197037)	С	-0,323 [-0,530,10]	0,0033	0,016	0,000	0,000	0,596
Índice de melaninaª	rs12913832 (28365618)*	G	-0,055 [-0,090,01]	0,0109	0,497	0,866	0,005	0,002
	rs13313425 (28096842)*	А	-0,083 [-0,160,00]	0,0360	0,078	0,024	0,388	0,038
	<u>rs12913832</u> (28365618)	G	-0,0479 [-0,080,00]	0,0204	0,497	0,866	0,005	0,002
Índice de melanina ^b	rs1800414* (28197037)	С	-0,204 [-0,400,00]	0,0457	0,016	0,000	0,000	0,596
	rs61738394* (28326942)	Т	0,207 [0,01 - 0,39]	0,0309	0,010	0,000	0,012	0,000

^a = Região não exposta ao sol; ^b = Região exposta ao sol; Sublinhado = Polimorfismo condicionado para verificação de sinais independentes; * = Polimorfismo associado após análise condicional ao polimorfismo com maior associação (menor valor de *p*); rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; A1 = Alelo de menor frequência (MAF); RBP = Frequência alélica na população de estudo; EUR = Frequência alélica nas populações europeias do *1000Genomes*; AFR = Frequência alélica nas populações africanas do *1000Genomes*; EAS = Frequência alélica nas populações do leste asiático do *1000Genomes*.

6.5. Análises epistáticas de segunda ordem

Potenciais efeitos de interação epistática envolvendo os marcadores dos genes *OCA2* e *HERC2* foram observados nas variáveis L*, a* e b* para a cor dos olhos, bem como para as variáveis de índice de melanina para região exposta e não exposta ao sol utilizando o método mb-mdr.

Com base nos valores de *p* observados pelos testes de regressão linear envolvendo todos os 170 polimorfismos previamente analisados para cada uma das cinco variáveis quantitativas analisadas, cinco conjuntos independentes de *tagSNPs* foram gerados.

Ademais, todas as análises epistáticas foram realizadas controlando os possíveis efeitos da estratificação populacional devido a ancestralidade, com a inclusão das estimativas de mistura europeia, africana e ameríndia (leste asiático) como covariáveis.

6.5.1. Interações epistáticas: CIELAB (L*, a* e b*)

Interação de segunda ordem abrangendo 93, 92 e 94 *tagSNPs* para as dimensões L*, a* e b* respectivamente, foram examinadas para 225 amostras.

Uma vez que o SNP rs12913832A>G do gene *HERC2*, é tido como o polimorfismo principal na determinação da cor dos olhos, a análise de epistasia foi primeiramente realizada controlando apenas para ancestralidade. Em todas as três variáveis as interações *multilocus* observadas mais significativas envolveram o polimorfismo rs12913832A>G (dados não mostrados). Desta forma, as análises aqui apresentadas foram realizadas controlando para a ancestralidade e também para possíveis efeitos principais proporcionados por este marcador. O controle para efeitos principais para análise de epistasia, quando há conhecimento prévio sobre um polimorfismo principal determinante, faz-se necessário pois se efeitos principais estiverem presentes na análise, pode ser difícil avaliar se uma interação específica foi detectada por causa dos efeitos principais, ou porque de fato é um efeito epistático real.

As Tabelas 11, 13 e 15 apresentam os polimorfismos mais frequentes em diferentes modelos de segunda ordem para as dimensões L*, a* e b*, respectivamente. As Tabelas 12, 14 e 16 apresentam as interações entre os dez modelos mais significativos compostos por polimorfismos do gene *OCA2* e *HERC2* para as três dimensões do espaço de cor CIELAB (L*, a* e b*), respectivamente.

OCA2	-	-	HERC2	_	-
rsID	Posição	n	rsID	Posição	n
rs41304383	28116282	10	rs138923800	28451772	11
rs2594928	28171147	8	rs143333992	28451776	9
rs7175046	28116275	5	snp_157*	28456382	8
rs1800419	28096538	4	rs201182375	28419566	7
rs1800404	28235773	3	snp_266*	28567867	5
rs12592307	28090173	1	rs11636232	28386626	4
rs12910433	28228644	1	rs372420106	28419526	3
rs4778139	28346511	1	rs376640270	28467246	3
			rs11074322	28368995	2
			rs112483884	28419330	2
			rs141441362	28419619	2
			rs4778242	28355991	2
			rs4778243	28356130	2
			rs6497271	28365431	2
			rs77865049	28419644	2
			rs8025035	28377772	2
			rs11631797	28502279	1
			rs201872292	28475625	1
			snp_282*	28568057	1

Tabela 11 - Polimorfismos presentes nos 51 modelos de segunda ordem significativos (p < 0,05) para a dimensão L* dentre as 4.278 comparações *multilocus* avaliadas.

rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; n = número de modelos de interação nos quais o polimorfismo é observado; * = Polimorfismo que não possui rsID

Tabela 12 - Dez	modelos	de intera	ções de	segunda	ordem	mais	significative	os comp	ostos por
polimorfismos	do gene	OCA2 e	HERC2	para a	dimens	ão L*	do espaço	de cor	CIELÃB.
Valores <i>p</i> signif	ficativos a	ssinalad	os em ne	egrito.					

				.						
SN	Ps		Alto	o Risco (H	.)		Baixo	Risco (L)	- Total n
a	b	NH	beta H	a - b	p	NL	beta L	a - b	p	Total p
rs77865049	rs41304383	1	28,03	CT - TC	7,0x10 ⁻⁴	0	NA	-	-	1,1x10 ⁻⁵
rs143333992	rs7175046	1	27,44	AG - CG	0,0010	0	NA	-	-	3,9x10 ⁻⁵
snp_266	rs41304383	1	26,79	GG - TC	0,0013	0	NA	-	-	1,8x10 ⁻⁴
rs143333992	rs201182375	1	26,30	AG - TC	0,0016	0	NA	-	-	1,8x10 ⁻⁴
rs143333992	rs2594928	1	25,30	AG - TT	0,0023	0	NA	-	-	2,4x10 ⁻⁴
rs143333992	rs1800419	1	24,83	AG - AA	0,0025	0	NA	-	-	2,6x10 -4
rs4778242	rs2594928	1	24,47	CT - TT	0,0032	0	NA	-	-	4,4x10 ⁻⁴
rs41304383	rs7175046	1	12,92	TC - CG	0,0020	0	NA	-	-	5,6x10 ⁻⁴
rs201182375	rs2594928	1	12,30	TC - TT	0,0032	0	NA	-	-	0,0012
rs143333992	rs6497271	1	22,36	AG - GG	0,0055	0	NA	-	-	0,0055

a = Polimorfismo 1; b = Polimorfismo 2; NH = Número de células designadas como alto-risco que foram combinadas no modelo; *beta* H = Coeficiente de regressão beta para categoria de alto-risco; NL = Número de células designadas como baixo-risco que foram combinadas no modelo; *beta* L = Coeficiente de regressão beta para categoria de baixo-risco; Total p = Valor de p após 1000 testes de permutação para acessar o nível de significância da interação; NA = Coeficiente beta não calculado por não haver pelo menos uma combinação de baixo-risco assinalada.

OCA2			HERC2		
rsID	Posição	n	rsID	Posição	n
rs41304383	28116282	10	rs113016718	28419713	13
rs2594928	28171147	9	rs138923800	28451772	13
rs1800404	28235773	7	rs201872292	28475625	13
rs7175046	28116275	6	rs143333992	28451776	11
rs1800419	28096538	5	rs200598834	28447419	9
rs56082644	28327170	5	snp_266	28567867	9
rs10852218	28231793	3	rs201182375	28419566	6
rs12910433	28228644	3	rs11636232	28386626	5
rs7495174	28344238	3	rs376640270	28467246	5
rs12592307	28090173	2	rs77865049	28419644	5
rs17652037	28260128	2	rs78970358	28431980	5
rs1800407	28230318	2	rs11074322	28368995	3
rs1800410	28230184	2	rs112483884	28419330	3
rs4778218	28211758	2	rs11631797	28502279	3
rs13313425	28096842	1	rs117504827	28502065	3
rs1800401	28260053	1	rs142399136	28419587	3
rs4640131	28259863	1	rs4778247	28361764	3
rs58843292	28345931	1	rs58164482	28422819	3
rs7170989	28200408	1	rs8024600	28443682	3
rs7175266	28000484	1	snp_157	28456382	3
			rs139335991	28419746	2
			rs141441362	28419619	2
			rs4778243	28356130	2
			rs5811535T	28346814	2
			rs12913832	28365618	1
			rs142477460	28475864	1
			rs7494786	28414665	1
			rs77093318	28419375	1
			snp_264	28567859	1
			snp_268	28567895	1

Tabela 13 - Polimorfismos presentes nos 101 modelos de segunda ordem significativos (p < 0,05) para a dimensão a^{*} dentre as 4.186 comparações *multilocus* avaliadas.

rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; n = número de modelos de interação nos quais o polimorfismo é observado; * = Polimorfismo que não possui rsID

Tabela 14 - Dez modelos de interações de segunda ordem mais significativos compostos por polimorfismos do gene OCA2 e HERC2 para a dimensão a* do espaço de cor CIELAB. Valores p significativos assinalados em negrito.

SN	Ps	Alto Risco (H)					Baix	o Risco (L	,)	Totol m
а	b	NH	beta H	a - b	р	NL	beta L	a - b	р	Total p
rs138923800	rs5811535	1	23,90	TG - AA	1,3x10 ⁻⁴	1	-8,04	TG - AAT	0,0287	1,0x10 ⁻⁴
snp_266	rs143333992	1	25,40	GG - AG	5,8x10 ⁻⁵	1	-5,17	GA - AG	0,0739	2,0x10 ⁻⁴
rs143333992	rs41304383	1	24,95	AG - TC	7,8x10 ⁻⁵	1	-5,08	AG - TT	0,0788	4,0x10 ⁻⁴
rs143333992	rs7175046	1	25,55	AG - CG	5,1x10 ⁻⁵	1	-5,20	AG - CC	0,0719	4,0x10 ⁻⁴
rs143333992	rs1800419	1	25,51	AG - AA	5,3x10 ⁻⁵	0	NA	-	-	4,0x10 -4
rs138923800	rs11636232	1	23,19	TG - CT	2,1x10 ⁻⁴	1	-7,80	TG - CC	0,0335	4,0x10 ⁻⁴
rs143333992	rs201182375	1	25,61	AG - TC	5,1x10 ⁻⁵	1	-5,21	AG - TT	0,0718	6,0x10 ⁻⁴
rs77865049	rs41304383	1	27,01	CT - TC	1,9x10⁻⁵	0	NA	-	-	8,0x10 ⁻⁴
rs11636232	rs2594928	1	11,89	CT - TT	1,7x10 ⁻⁴	1	-3,11	CC - TT	0,0992	8,0x10 ⁻⁴
rs78970358	rs1800419	1	8,67	CT - AG	0,0169	1	-7,35	CT - GG	0,0194	0,0230

a = Polimorfismo 1; b = Polimorfismo 2; NH = Número de células designadas como alto-risco que foram combinadas no modelo; *beta* H = Coeficiente de regressão beta para categoria de alto-risco; NL = Número de células designadas como baixo-risco que foram combinadas no modelo; *beta* L = Coeficiente de regressão beta para categoria de baixo-risco; Total p = Valor de p após 1000 testes de permutação para acessar o nível de significância da interação; NA = Coeficiente beta não calculado por não haver pelo menos uma combinação de baixo-risco assinalada.

Tabela 15 - Polimorfismos mais aparecem em diferentes modelos de segunda ordem para a dimensão b* levando em consideração 4.371 comparações *multilocus*: 28 polimorfismos compondo um total de 29 modelos com p < 0,05 após 1000 permutações para acessar o nível de significância.

OCA2			HERC2			
rsID	Posição	n	rsID	Posição	n	
rs56082644	28327170	8	rs376640270	28467246	6	
rs12910433	28228644	3	rs11636232	28386626	5	
rs2594928	28171147	3	rs117902599	28499434	5	
rs7495174	28344238	3	rs12913832	28365618	3	
rs10852218	28231793	1	rs11631797	28502279	2	
rs12592307	28090173	1	rs139335991	28419746	2	
rs1800404	28235773	1	rs11074322	28368995	1	
rs1800410	28230184	1	rs142399136	28419587	1	
rs1800419	28096538	1	rs201872292	28475625	1	
rs41304383	28116282	1	rs372420106	28419526	1	
rs4778218	28211758	1	rs4778242	28355991	1	
rs58843292	28345931	1	rs4778243	28356130	1	
rs7170989	28200408	1	rs71467328	28518229	1	
			rs78970358	28431980	1	
			snp_282	28568057	1	

rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; *n* = número de modelos de interação nos quais o polimorfismo é observado; * = Polimorfismo que não possui rsID

Tabela 16 - Dez modelos	de interações de	segunda orde	m mais	significativos	compostos por
polimorfismos do gene	OCA2 e HERC2	para a dime	nsão b*	do espaço d	le cor CIELAB.
Valores p significativos a	issinalados em ne	egrito.			

SNPs			Alto Risco (H)			Baixo Risco (L)			M =4=1 =	
a	b	NH	beta H	a - b	р	NL	beta L	a - b	р	Total p
rs11636232	rs1800419	2	6,23 4,21	TT - GG CT - AA	0,0955 0,0173	1	-2,44	CC - AA	0,0851	0,0032
rs11631797	rs117902599	0	NA	-	-	1	-10,31	GA - CA	0,0118	0,0034
rs139335991	rs56082644	1	9,81	GG - TT	0,0680	1	-19,52	GA - TT	0,0097	0,0042
rs56082644	rs7170989	1	10,01	TT - TT	0,0612	1	-19,93	TT - CT	0,0080	0,0078
rs12910433	rs12592307	0	NA	-	-	1	-21,68	AA - TT	0,0035	0,0096
rs142399136	rs56082644	1	10,09	GG - TT	0,0603	1	-20,09	GA - TT	0,0077	0,0180
rs56082644	rs12910433	1	10,25	TT - GG	0,0531	1	-20,41	TT - AG	0,0061	0,0112
rs56082644	rs1800404	1	15,04	TT - TT	0,0452	1	-20,96	TT - TC	0,0050	0,0150
rs11636232	rs10852218	0	NA	-	-	1	-18,83	TT - CT	0,0105	0,0154
rs117902599	rs201872292	0	NA	-	-	1	-16,91	CA - AG	0,0213	0,0290

a = Polimorfismo 1; b = Polimorfismo 2; NH = Número de células designadas como alto-risco que foram combinadas no modelo; *beta* H = Coeficiente de regressão beta para categoria de alto-risco; NL = Número de células designadas como baixo-risco que foram combinadas no modelo; *beta* L = Coeficiente de regressão beta para categoria de baixo-risco; Total p = Valor de p após 1000 testes de permutação para acessar o nível de significância da interação; NA = Coeficiente beta não calculado por não haver pelo menos uma combinação de baixo-risco assinalada.

6.5.2. Interações epistáticas: Índice de melanina

Interação de segunda ordem abrangendo 95 e 97 *tagSNPs* para índice de melanina em região não exposta e exposta ao sol respectivamente, foram examinadas para 327 amostras. Assim como para a análise de epistasia para as dimensões CIELAB, o controle para ancestralidade e efeitos principais também foram empregados.

As Tabelas 17 e 19 apresentam os polimorfismos mais frequentes em diferentes modelos de segunda ordem para o índice de melanina em região não exposta e exposta ao sol, respectivamente. As Tabelas 18 e 20 apresentam as interações entre os dez modelos mais significativos compostos por polimorfismos do gene *OCA2* e *HERC2* para região não exposta e exposta ao sol.

Tabela 17 - Polimorfismos mais frequentes em diferentes modelos de segunda ordem para o índice de melanina em região não exposta ao sol, levando em consideração 4.465 comparações *multilocus*: 73 polimorfismos compondo um total de 175 modelos com p < 0,05 após 1000 permutações para acessar o nível de significância.

OCA2			HERC2			
rsID	Posição	n	rsID	Posição	n	
rs4778139	28346511	15	rs113016718	28419713	14	
rs114561380	28344613	10	rs138923800	28451772	12	
rs7175266	28000484	10	rs58358300	28419393	11	
rs13313425	28096842	9	rs201182375	28419566	10	
rs4640131	28259863	9	rs73362604	28422684	9	
rs10852218	28231793	7	rs113294881	28388194	8	
rs115488780	28346322	7	rs7494786	28414665	8	
rs28595435	28344740	6	snp_151	28456028	8	
rs61738394	28326942	6	snp_268	28567895	8	
rs7170989	28200408	6	rs112483884	28419330	7	
rs2594928	28171147	5	rs142477460	28475864	6	
rs58843292	28345931	5	rs200598834	28447419	6	
rs12910433	28228644	4	rs2525936	28441817	6	
rs1800410	28230184	4	rs6497271	28365431	6	
rs41304383	28116282	4	rs77865049	28419644	6	
rs41534647	28267758	4	rs10162958	28507825	5	
rs1800401	28260053	3	rs139377364	28467383	5	
rs1800419	28096538	3	rs141441362	28419619	5	
rs111707210	28347419	2	rs7403363	28389508	5	
rs112277343	28116987	2	snp_135	28447234	5	
rs12592307	28090173	2	snp_264	28567859	5	
rs1800404	28235773	2	rs112222559	28451737	4	
rs1800416	28171294	2	rs11636232	28386626	4	
rs7175046	28116275	2	rs139335991	28419746	4	
rs17652037	28260128	1	rs4778243	28356130	4	
rs7497270	28344328	1	rs61754657	28413640	4	
			rs74007958	28362029	4	
			snp_262	28567840	4	
			rs12913832	28365618	3	
			rs142399136	28419587	3	
			rs149963581	28457540	3	
			rs4778245	28357230	3	
			rs61757160	28478323	3	
			rs8025035	28377772	3	
			rs113921795	28451756	2	
			rs372420106	28419526	2	

rs59556486

28499436

2

rs78696453	28419650	2
snp_261	28567799	2
rs111681126	28501195	1
rs112371820	28544707	1
rs113970029	28502462	1
rs114694773	28413446	1
rs143333992	28451776	1
rs151156359	28459219	1
rs57727285	28483619	1
snp_176	28467512	1

rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; n = número de modelos de interação nos quais o polimorfismo é observado; * = Polimorfismo que não possui rsID

Tabela 18 - Dez modelos de interações de segunda ordem mais significativos compostos por polimorfismos do gene OCA2 e HERC2 para o índice de melanina em região não exposta ao sol. Valores p significativos assinalados em negrito.

	Baixo Risco (L)			Alto Risco (H)				SNPs			
Total p	р	a - b	beta L	NL	р	a - b	beta H	NH	b	a	
1,4x10 ⁻⁵	0,0036 0,0320	CT - GA CT - AA	-0,61 -0,57	2	0,0052	CT - GG	0,19	1	rs13313425	rs113294881	
2,2 x10 ⁻⁴	8,2x10 ⁻⁴	CT - AG	-0,30	1	0,0861 0,0934	CT - AA CC - AG	0,07 0,08	2	rs4640131	rs112222559	
2,3x10 ⁻⁴	1,7x10 ⁻⁴	TG - GA	-1,00	1	-	-	NA	0	rs74007958	rs138923800	
2,4x10 ⁻⁴	6,8x10 ⁻⁴	GA - AG	-0,91	1	0,0676	GA - GG	0,49	1	rs10162958	snp_268	
2,4x10 ⁻⁴	0,0147	GG - CG	-0,38	1	0,0346	AG - CG	0,28	1	rs115488780	rs10162958	
2,5x10 ⁻⁴	3,0x10 ⁻⁴	CT - AG	-0,43	1	0,0353	CT - AA	0,14	1	rs4640131	rs113294881	
0,0020	0,0067	GA - TC	-0,30	1	0,0443	GG - TC	0,17	1	rs7175266	rs4778243	
0,0031	8,2x10 ⁻⁴	TC - AG	-0,45	1	0,0178	TC - AA	0,23	1	rs4640131	rs114561380	
0,0034	0,0088	GA - CT	-0,50	1	0,0974	GA - CC	0,20	1	rs77865049	snp_268	
0,0100	0,0011	GA - AG	-0,36	1	0,0833	GA - AA	0,10	1	rs4640131	rs74007958	

a = Polimorfismo 1; b = Polimorfismo 2; NH = Número de células designadas como alto-risco que foram combinadas no modelo; *beta* H = Coeficiente de regressão beta para categoria de alto-risco; NL = Número de células designadas como baixo-risco que foram combinadas no modelo; *beta* L = Coeficiente de regressão beta para categoria de baixo-risco; Total p = Valor de p após 1000 testes de permutação para acessar o nível de significância da interação; NA = Coeficiente beta não calculado por não haver pelo menos uma combinação de baixo-risco assinalada.

OCA2			HERC2			
rsID	Posição	n	rsID	Posição	n	
rs1800410	28230184	15	rs138923800	28451772	20	
rs41304383	28116282	14	snp_268	28567895	15	
rs4640131	28259863	10	rs58358300	28419393	14	
rs114561380	28344613	9	rs78696453	28419650	14	
rs41534647	28267758	9	snp_264	28567859	12	
rs10852218	28231793	8	rs77865049	28419644	11	
rs58843292	28345931	8	snp_151	28456028	11	
rs7497270	28344328	8	rs113294881	28388194	9	
rs1800419	28096538	7	rs201182375	28419566	9	
rs111707210	28347419	6	rs8025035	28377772	9	
rs115488780	28346322	6	rs112371820	28544707	7	
rs12910433	28228644	6	rs142477460	28475864	7	
rs4778218	28211758	6	rs73362604	28422684	7	
rs62007492	28344765	6	rs76333042	28506246	7	
rs13313425	28096842	5	rs112222559	28451737	6	
rs1800401	28260053	5	snp_262	28567840	6	
rs28595435	28344740	5	rs139377364	28467383	5	
rs1800404	28235773	4	rs2525936	28441817	5	
rs/1/0989	28200408	4	rs/89/0358	28431980	5	
rs12592307	28090173	3	rs113921795	28451750	4	
rs2594928	28171147	3	rs1962780	28483543	4	
IS/1/5200	28000484	3	IS2428030	28475370	4	
1817052057 ro1800416	20200120	2	1874007956	20302029	4	
rs1800410	20171294	2 1	rs10438451	28309308	3	
rs1800407	28116336	1	rs117504827	28502065	3	
rs56082644	28327170	1	rs4778242	28355991	3	
rs7175046	28116275	1	rs4778243	28356130	3	
101 11 00 10	201102.0	-	snp 282	28568057	3	
			rs10162958	28507825	2	
			rs11636232	28386626	2	
			ro140200126	2000020	2	
			18142399130	20419307	2	
			18149592795	28419095	2	
			rs149963581	28457540	2	
			rs200598834	28447419	2	
			rs7495441	28359744	2	
			rs8024600	28443682	2	
			rs916976	28513513	2	
			snp_266	28567867	2	
			rs112483884	28419330	1	
			rs12913832	28365618	1	
			rs139335991	28419746	1	
			rs143333992	28451776	1	
			rs201872292	28475625	1	
			rs372420106	28419526	1	
			rs7494786	28414665	1	
			ro77002218	2811 1005	1	
			19/1020010	20712010	1	

Tabela 19 - Polimorfismos mais frequentes em diferentes modelos de segunda ordem para o índice de melanina em região exposta ao sol, levando em consideração 4.656 comparações *multilocus*: 75 polimorfismos compondo um total de 193 modelos com p < 0,05 após 1000 permutações para acessar o nível de significância.

rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; n = número de modelos de interação nos quais o polimorfismo é observado; * = Polimorfismo que não possui rsID

Tabela 20 - Dez modelos de interações de segunda ordem mais significativos compostos por polimorfismos do gene OCA2 e HERC2 para o índice de melanina em região exposta ao sol. Valores p significativos assinalados em negrito.

SN	IPs	Alto Risco (H)				Bai	Total n			
а	b	NH	beta H	a - b	p	NL	beta L	a - b	р	Total p
rs138923800	rs74007958	0	NA	-	-	1	-0,92	TG - GA	2,4x10 ⁻⁴	1,3x10 ⁻⁵
rs138923800	rs113294881	0	NA	-	-	1	-0,91	TG - CT	2,5x10 ⁻⁴	1,3x10 ⁻⁵
snp_151	rs138923800	0	NA	-	-	1	-0,90	AG - TG	3,5x10 ⁻⁴	1,3x10 ⁻⁵
snp_268	rs10162958	1	0,72	GA - GG	0,0044	1	-0,80	GA - AG	0,0014	2,7x10 ⁻⁴
snp_268	rs58843292	1	0,73	GA - GA	0,0039	0	NA	-	-	2,7x10 ⁻⁴
rs138923800	rs58358300	1	0,45	TG - CC	0,0717	1	-0,92	TG - AC	2,3x10 ⁻⁴	0,0020
rs138923800	rs2594928	1	0,45	TG - TT	0,0778	1	-0,85	TG - CT	7,5x10 ⁻⁴	0,0040
rs113294881	rs4640131	1	0,12	CT - AA	0,0768	1	-0,34	CT - AG	0,0024	0,0100
rs138923800	rs1800404	0	NA	-	-	1	-0,30	TG - CC	0,0358	0,0358
rs62007492	rs1800404	1	0,49	AA - TT	-	1	-0,87	CG - CT	0,0046	0,0460

a = Polimorfismo 1; b = Polimorfismo 2; NH = Número de células designadas como alto-risco que foram combinadas no modelo; *beta* H = Coeficiente de regressão beta para categoria de alto-risco; NL = Número de células designadas como baixo-risco que foram combinadas no modelo; *beta* L = Coeficiente de regressão beta para categoria de baixo-risco; Total p = Valor de p após 1000 testes de permutação para acessar o nível de significância da interação; NA = Coeficiente beta não calculado por não haver pelo menos uma combinação de baixo-risco assinalada.

7 DISCUSSÃO

7.1. Amostra populacional

A população brasileira é considerada uma população majoritariamente tri-híbrida, com contribuições de populações parentais europeias, africanas e dos ameríndios autóctones que já viviam no continente sul-americano a época do descobrimento do Brasil. Este longo processo de mistura que perdura a mais de 500 anos reflete na heterogeneidade fenotípica relacionada a cor dos olhos, cabelos e pele que observamos ao longo de todo o território nacional.

No que se refere a estes traços fenotípicos, o censo brasileiro realizado pelo IBGE ainda faz uso principalmente da cor da pele para classificar indivíduos etnicamente em cinco categorias: (a) branco, (b) pardo, (c) preto, (d) amarelo e (e) indígena. Alguns estudos na literatura tendem a utilizar estas categorias auto-declaradas para estratificar a amostra quando não há informações da ancestralidade genômica das mesmas.

Esta abordagem é questionada, não apresentando-se como uma estratégia eficiente para controlar a estratificação populacional devido à dissociação entre a mesma e a ancestralidade genômica (Suarez-Kurtz *et al.*, 2007; Pimenta *et al.*, 2006; Santos *et al.*, 2009; Lins *et al.*, 2011; Parra *et al.*, 2003; Pena *et al.*, 2011). Em contrapartida, outros estudos indicam que a cor da pele auto-declarada seria consistente com a contribuição ancestral dominante ou relativa (Blanton *et al.*, 2008; Cardena *et al.*, 2013; Manta *et al.*, 2013; Muniz *et al.*, 2008; Pena *et al.*, 2009; Queiroz *et al.*, 2013; Santos *et al.*, 2010). Isso sugere que a dinâmica demográfica das populações exerce influência sobre a correlação entre pigmentação e ancestralidade genômica.

De acordo com o último censo brasileiro realizado em 2010 pelo IBGE, cerca de 69,8% dos indivíduos na região de Ribeirão Preto se declararam como sendo brancos, seguido de 22,8% pardos, 6,4% pretos, 0,9% amarelos e 0,1% indígenas. No presente trabalho, 63,9% dos indivíduos se declararam brancos, seguido de 19,6% pardos, 12,7% pretos, 3,6% amarelos e 0,3% indígenas, o que consiste em valores bastante semelhantes aos do último censo.

Estudos realizados em amostras de indivíduos da região de Ribeirão Preto (SP), estratificados em brancos, pretos e pardos mostraram uma miscigenação substancial em todos estes três grupos (Ferreira *et al.*, 2006; Muniz *et al.*, 2008). Análise posterior dos dados conduziu a estimativas de contribuição europeia, africana e ameríndia na população de 71,64%, 19,72% e 8,64%, respectivamente (Silva *et al.*, 2018, submetido à *Forensic Science International Genetics*). Tais estimativas de mistura estão em concordância com as estimativas aqui apresentadas, sendo (a) 71,4% de componente europeu, (b) 19,6% de contribuição africana e (c) 9% de componente ameríndio (representado pelas populações do leste asiático) (Tabela 2).

De uma forma geral, a ancestralidade europeia predomina em todas as regiões do país, atingindo suas maiores estimativas nas regiões sudeste e principalmente na região sul, com 79% e 85% respectivamente (Santos *et al.*, 2015). Existem localidades que desviam deste padrão. Na região norte do país a contribuição ameríndia chega a 52% (Guerreiro *et al.*, 1993), e na região nordeste pode-se observar cerca de 50% de contribuição africana em algumas populações (Santos *et al.*, 2015).

No presente trabalho, a relação entre cor de pele auto-declarada e ancestralidade genômica corrobora outros estudos mostrando que, embora tenham sido observadas diferenças significativas entre os quatro grupos de cor auto-declarada (branco, amarelo, pardo e preto), em ancestralidade genômica e pigmentação da pele (mensurada pelo índice de melanina), observou-se também considerável dispersão dentro de cada grupo e substancial sobreposição entre os mesmos (Figuras 23 e 26).

A Figura 23 evidencia um exemplo de sobreposição e possível fator que pode levar um indivíduo a se auto-declarar preto, onde os índices de melanina para região exposta ao sol entre os indivíduos pardos se sobrepõe de maneira quase perfeita ao índice de melanina para região não exposta ao sol em indivíduos que se auto-declararam pretos. O mesmo ocorre com indivíduos declarados brancos, que, quando expõem a pele ao sol, podem ter tons de pele semelhantes aos indivíduos declarados pardos. Em ambos os casos, a capacidade de bronzeamento da pele é um componente importante a ser considerado, pois pode interferir na percepção de cor que leva à classificação desses indivíduos como brancos, amarelos, pardos e pretos. Além disso, na população estudada, não é possível distinguir brancos e amarelos considerando somente a cor da pele. Neste caso, Pena *et al.* (2011) sugere que, não só o tom de pele, mas também outros traços fenotípicos como o formato dos olhos, nariz e lábios assim como o tipo de cabelo, devam ser considerados na classificação em branco, pardo, preto e amarelo.

É interessante observar que a cor da pele tenha aqui se apresentado apenas como um substituto impreciso da ancestralidade. Apesar disso, o índice de melanina se correlaciona fortemente com a ancestralidade europeia (correlação negativa, r = -0,69) e africana (correlação positiva, r = 0,70), conforme observado na Figura 24. Uma possível explicação para essa correlação pode estar relaciona aos fatos de que (1) a ancestralidade genômica difere entre os grupos e (2) polimorfismos que são associados a pigmentação são também bons marcadores de ancestralidade, como, por exemplo, os polimorfismos rs12913832A>G do gene HERC2 e rs16891982C>G do gene SLC45A2, ambos presentes no SNPforID 34-plex. Esta correlação não é perfeita possivelmente em virtude de a pigmentação ser uma característica complexa, onde diferentes genes interagem para determinar, neste caso, a cor da pele. Estes dados indicam que a determinação da ancestralidade baseada na auto-declaração proposta pelo IBGE é vulnerável a classificações espúrias e deve ser evitada. O conceito de ancestralidade e cor auto-declarada não devem ser encarados como sinônimos em estudos genéticos, devendo-se empregar a ancestralidade individual estimada por AIMs sempre que necessário. Ainda assim, não podemos descartar a classificação do IBGE, uma vez que esta pode ser útil em outras áreas onde a ancestralidade genética é

dispensável, como, por exemplo, para se identificar tendências sociais em economia (Magalhães da Silva *et al.*, 2015).

Um ponto importante que deve ser ressaltado é que durante a coleta de amostras, muitas destas foram direcionadas para obtenção de fenótipos menos representados, como cor de olhos azuis, verdes e mel, bem como cabelos ruivos e loiros, e pele preta. A interpretação das estimativas aqui apresentadas deve ser realizada com cautela uma vez que indivíduos com fenótipos de pigmentação mais claros estão associados a um componente ancestral europeu mais elevado, bem como, indivíduos com pele preta associados a um componente africano e, desta forma, aumentando as estimativas dos respectivos componentes observados (Freire-Aradas *et al.*, 2014). Portanto, a seleção não aleatória de amostras para compor este trabalho possivelmente alterou, mesmo que de forma sutil, as estimativas de mistura dos componentes ancestrais.

7.2. Estruturas secundárias presentes na íris

A diversidade presente na cor dos olhos têm sido objeto de estudo de diferentes grupos de pesquisa ao longo da última década, mesmo após a descoberta do polimorfismo rs12913832A>G do gene *HERC2*, descrito como principal marcador a explicar a variabilidade existente entre olhos azuis e castanhos (Sturm *et al.*, 2008; Eiberg *et al.*, 2008).

Entretanto, apenas três grupos de pesquisas distintos procuraram observar a variabilidade presente em outras características presentes na íris, como nódulos de Wolfflin, sulcos de contração, pontos de pigmentação na íris e na esclera. Algumas dessas características podem alterar a percepção de cor da íris (Liu *et al.*, 2010; Mackey *et al.*, 2011) e, ainda que as consequências funcionais de muitas destas características ainda permaneçam desconhecidas, estas podem ser de grande importância na prática forense, no que se refere à predição da pigmentação dos olhos, e também para a área biomédica, onde algumas destas estruturas secundárias tem sido reportadas como marcadores de risco para diferentes patologias, como por exemplo, melanoma uveal, no qual pigmentos na íris consistem em fatores de risco (Holly *et al.*, 1990; Horn *et al.*, 1994). Ademais, faltam estudos em populações diferentes das europeias e asiáticas que possam contribuir para o melhor entendimento da distribuição e diversidade genética vinculadas a estas características.

Seguindo esta linha de raciocínio, a população brasileira pode contribuir grandemente na tentativa de identificar fatores genéticos envolvidos no aparecimento destas características, uma vez que, conforme dito no tópico anterior, é uma população majoritariamente tri-híbrida.

A distribuição dos nódulos de Wofflin foi predominante em indivíduos com olhos claros (Figura 12A). Estas estruturas circulares presentes na parte mais externa da zona ciliar são altamente associadas a cor dos olhos claros em comparação com olhos escuros, conforme observou Kim *et al.* (2012) em amostras do leste asiático, onde nenhum dos 112 indivíduos com olhos escuros apresentou nódulos de Wolfflin. Edwards *et al.* (2015) observou uma maior frequência de nódulos de Wolfflin em indivíduos com ancestralidade europeia (25,8%) quando comparado a indivíduos do leste asiático (0%) e sul asiáticos (0,8%). No presente estudo a ancestralidade média dos indivíduos que apresentaram nódulos de Wolfflin foi de 87,2% europeia, 3,8% ameríndia e 8% africana (dados não mostrados), corroborando os achados a despeito da presença desta característica em indivíduos com ancestralidade europeia. Ainda assim, é importante enfatizar que Falls (1970) sugere que esta característica pode não ser ausente em íris mais escuras, mas sim, dificilmente identificável por estar escurecida devido à elevada quantidade de melanina presente.

Os sulcos de contração foram mais presentes em indivíduos com olhos escuros (Figura 12B). Nos estudos de Larson e Pedersen (2004) e Sidhartha *et al.* (2014) observou-se também que a extensão com que os sulcos de contração se distribuem ao longo da íris está associado a olhos escuros. No presente estudo, a ancestralidade média dos indivíduos que apresentaram sulcos de contração foi de 77% europeia, 8% ameríndia e 14% africana (dados não mostrados), corroborando os resultados obtidos por Quillen *et al.* (2011) que observou uma correlação positiva entre sulcos de contração e ancestralidade europeia em uma amostragem de Brasília (DF). Interessantemente, Quillen *et al.* (2011) associou sulcos de contração com olhos claros, contrastando com o observado em nosso estudo e na literatura [Larson e Pedersen (2004) e Sidhartha *et al.* (2014)].

A distribuição de pontos de pigmentação na íris foi levemente mais frequente em olhos claros (Figura 12E). A ancestralidade média dos indivíduos que apresentaram pontos de pigmentação na íris foi de 81% europeia, 7% ameríndia e 11% africana (dados não mostrados). No estudo de Edwards *et al.* (2015), pontos de pigmentação na íris foram mais frequente em indivíduos com ancestralidade europeia quando comparado a indivíduos do leste e sul asiáticos. Uma vez que pontos de pigmentação são pequenas regiões delimitadas de acúmulo de pigmentos marrons/pretos, é mais provável observar esta característica em indivíduos com olhos claros e, desta forma, mais provável de observar em indivíduos com ancestralidade europeia, onde diferentes cores de olhos são mais frequentes.

A pigmentação da esclera, também chamada de melanose conjuntival, foi mais frequente em indivíduos com olhos escuros, principalmente castanho-escuros (Figura 12D). Apesar de não ser ainda muito bem compreendida, esta característica parece estar associada a populações com olhos escuros, uma vez que a ancestralidade média dos indivíduos que apresentaram esta característica no presente trabalho foi de 45% europeia, 11% ameríndio e 43% africana (dados não mostrados). Edwards *et al.* (2015) evidenciou este mesmo padrão, ao observar uma maior frequência desta característica em populações do sul e do leste asiático (populações com uma frequência elevada de indivíduos com olhos escuros), quando comparado a indivíduos com ancestralidade europeia.

É importante enfatizar que a ancestralidade europeia está associada a estas características, visto que em todas elas, exceto em melanose conjuntival, a ancestralidade europeia dos indivíduos que as possuem é maior que a média do presente estudo, 71,4%. Isto poderia indicar um viés causado pela identificação das estruturas, as quais seriam mais facilmente identificáveis em olhos claros. Porém, sulcos de contração contradizem esta possibilidade de viés, visto que ocorrem mais frequentemente em olhos escuros e, ainda assim, os indivíduos apresentando esta estrutura na íris possuem uma estimativa de ancestralidade europeia superior (77%).

7.3. Variantes dos genes OCA2 e HERC2 e pigmentação humana

No presente trabalho, foi proposto a análise por meio de sequenciamento de nova geração das regiões exônicas e regulatórias dos genes *OCA2* e *HERC2* implicados na via de melanogênese. Esta via compõe um intrincado processo na qual vários genes contribuem para o aparecimento dos diferentes fenótipos de pigmentação (Sturm, 2006).

Grandes consórcios como o ENCODE (ENCODE, 2012), que buscam delinear todos os elementos funcionais codificados no genoma humano, disponibilizam uma crescente de informações a respeito das variantes implicadas em fenótipos complexos como a pigmentação humana, sugerindo que estas variantes não afetam diretamente as proteínas envolvidas mas, possivelmente, os mecanismos que norteiam o processo de expressão destas (Ward e Kellis, 2016). De fato, estudos genéticos sobre a variação normal da pigmentação têm permitido identificar uma grande porção da variância genética observada entre estes fenótipos localizada em regiões não-codificantes do genoma (Albert e Kruglyak, 2015). Desta forma, torna-se natural pensar que novos elementos com funções regulatórias no processo de pigmentação sejam frequentemente identificados, visto que apenas aproximadamente 1,5% do genoma humano é constituído por DNA codificante de proteínas (Ponting e Hardison, 2011; Vernot et al., 2012). Um exemplo muito bem descrito na literatura consiste no polimorfismo rs12913832A>G localizado no íntron 86 do gene HERC2 aqui estudado, o qual está fortemente associado a variação normal da cor dos olhos por influenciar a ligação simultânea do fator de transcrição Helicase-Like (HLTF) no íntron 86 do gene HERC2 e na região promotora do gene OCA2.

Ainda que represente uma pequena proporção do genoma humano, a o DNA codificante de proteínas deve sempre ser levado em consideração, uma vez que variações não sinônimas, isto é, que causam mudança de aminoácido, levam a alterações na estrutura proteica podendo alterar a função dessa proteína, enquanto que as variações sinônimas, que não alteram o aminoácido, podem afetar a conformação, *splicing* e a estabilidade do pré-mRNA e mRNA, eventualmente afetando os níveis de expressão da proteína (Teng *et al.*, 2012).

Neste contexto, procuramos identificar novas variantes além daquelas já descritas na literatura para os genes *OCA2* e *HERC2* que estejam associadas com a pigmentação. Duas

abordagens distintas foram utilizadas buscando melhor caracterizar e reforçar a autenticidade das associações encontradas. A primeira levando em consideração análises de associação através de variáveis categóricas (discretas), comumente avaliadas em estudos de associação com fenótipos de pigmentação (Duffy *et al.*, 2007; Kayser *et al.*, 2008; Sturm *et al.*, 2008; Sulem *et al.*, 2008; Sulem *et al.*, 2007), e a segunda envolvendo variáveis quantitativas obtidas por imagens fotográficas (para avaliação da cor dos olhos) e com o auxílio de um espectrofotômetro portátil capaz de mensurar o índice de melanina na pele (para avaliação da cor da pele) (Andersen *et al.*, 2013; Edwards *et al.*, 2012; Edwards *et al.*, 2010; Leite *et al.*, 2011; Liu *et al.*, 2010; Rawofi *et al.*, 2017).

Além disso, análises epistáticas para as variáveis quantitativas foram realizadas, trazendo à tona associações de polimorfismos que quando analisados de forma individual não apresentam associação com o fenótipo. Isto já era esperado por dois motivos: o primeiro refere-se ao fato de que uma vez que a pigmentação humana é um fenômeno caracterizado pelo envolvimento de mais de 170 genes, é evidente pensar que interações de magnitudes diferentes entre os mais diversos genes envolvidos sejam identificadas neste tipo de abordagem; e segundo, espera-se que o uso de informações categorizadas a partir de características contínuas (como a cor dos olhos, pele e cabelos) simplifique a natureza quantitativa da característica. Portanto, polimorfismos adicionais que contribuem para a variação da pigmentação tornam-se identificáveis se todo o espectro quantitativo de coloração puder ser explorado (Liu *et al.*, 2010).

Considerando todos os 170 polimorfismos aqui analisados, 48 destes tiveram pelo menos uma associação com alguma característica de pigmentação e/ou estrutura secundária presente na íris, seja esta uma característica avaliada por meio de variáveis categóricas, quantitativas, ou por meio da análise exploratória de epistasia (levando em consideração apenas as 10 interações mais significativas encontradas para cada variável quantitativa).

7.3.1. rs12913832A>G: o principal preditor da cor do olhos e outras características fenotípicas

Por muito tempo a cor do olho em humanos foi tratada como um simples caráter mendeliano (Andersen *et al.*, 2016). Tradicionalmente, esta característica sempre foi descrita como uma herança mendeliana dominante, com a cor "castanho" sendo dominante sobre a cor "azul", ainda que diferentes estudos tenham demonstrado inconsistências com esse simples modelo de herança dominante (Frudakis *et al.*, 2007; Sturm e Larsson, 2009).

Os genes *OCA2* e *HERC2* aqui analisados possuem uma extensa lista de estudos na literatura associando-os a variação normal da pigmentação humana, principalmente a variação na cor dos olhos (Andersen *et al.*, 2016; Beleza *et al.*, 2013; Bennett e Lamoreux, 2003; Donnelly *et al.*, 2012; Eiberg *et al.*, 2008; Frudakis *et al.*, 2007; Kayser *et al.*, 2008).

Ainda assim, é importante enfatizar que os trabalhos que analisaram estes dois genes com uma abordagem mais ampla (e não apenas polimorfismos específicos) estão restritos a populações fora do continente sul-americano, principalmente populações europeias (Andersen *et al.*, 2016) e asiáticas (Rawofi *et al.*, 2017). Apesar da vasta maioria destes estudos de pigmentação estar restrita a estas populações, alguns estudos já se propuseram a analisar polimorfismos específicos destes dois genes em populações miscigenadas sulamericanas, incluindo amostragens da população brasileira (Cerqueira *et al.*, 2014; Freire-Aradas *et al.*, 2014), salientando um trabalho recente de nosso grupo (Andrade *et al.*, 2017).

Ademais, até onde sabemos, este é o primeiro trabalho no Brasil que se propôs a analisar a associação com a cor do olhos por meio de uma metodologia objetiva que busca avaliar quantitativamente a mesma. Como mencionado anteriormente, esta metodologia possui vantagens em relação aos tradicionais métodos subjetivos de categorização da cor do olho. A cor da íris é considerado um traço quantitativo inato. Embora as categorias discretas possam capturar diferenças extremas na cor da íris (por exemplo, azul vs. castanho), elas não são capazes de capturar toda a variação da cor dos olhos. Além disso, não há consenso formal sobre quantas categorias devem ser usadas, o que dificulta a confiabilidade de resultados dependentes de avaliações inter e intra-observador de métodos categóricos (Seddon *et al.*, 1990).

Como esperado, os genes *OCA2* e *HERC2* tiveram as suas maiores associações relacionadas a cor dos olhos sendo que, dentre todos os polimorfismos que tiveram seus alelos associados, o principal foi o rs12913832A>G do gene *HERC2* (Tabela 5). Como mencionado anteriormente, o marcador em si está localizado no meio de uma região de ligação para o fator de transcrição *Helicase-Like* (HLTF), que remodela a estrutura da cromatina em torno dos genes. Visser *et al.* (2012) e Sturm *et al.* (2008) observaram que este polimorfismo influencia a regulação do gene *OCA2*. Quando o alelo ancestral "A" está presente, o HLTF é capaz de se ligar a esta região e começar a desenovelar a estrutura da cromatina expondo locais para outras proteínas reguladoras que promovem a transcrição do gene *OCA2*. Em contraste, quando o alelo derivado "G" está presente, o sítio torna-se inacessível para a molécula HLTF. Como resultado, menos proteínas reguladoras são capazes de acessar essa região e a transcrição de *OCA2* é substancialmente reduzida.

A associação com a cor dos olhos para este polimorfismo corrobora os achados na literatura, onde o alelo "G" está associado a maiores chances de possuir olhos claros quando comparado ao alelo "A" (Tabela 5). O mesmo ocorre quando analisado sob a perspectiva de uma variável quantitativa, aqui evidenciada pela distribuição dos indivíduos homozigotos "GG" na coordenada b* do espaço de cores CIELAB (Figura 28), onde claramente há um agrupamento, ainda que disperso, destes indivíduos no eixo negativo que compreende à cor azul. Em relação as outras duas coordenadas (L* e a*), o alelo "G" foi associado a valores positivos do coeficiente beta para a coordenada L*, que representa o brilho/claridade (maiores valores de L* indicam um maior brilho/claridade) [beta = 3,745 (IC 95% = 2,18 - 5,30) p = 4,67x10⁻⁶], de acordo com o esperado para olhos claros (associados ao alelo "G") (Tabela 9). Para a variável a*, o alelo "A" foi associado a valores

positivos do coeficiente beta [beta = 2,607 (IC 95% = 1,40 - 3,81) $p = 3,39 \times 10^{-5}$] (Tabela 9). Valores positivos de a* indicam a cor vermelha, também estando de acordo com o esperado para o alelo "A", associado a olhos castanho-claros e escuros.



Figura 28 - Distribuição dos genótipos do polimorfismo rs12913832 plotados entre os valores a* e b* do espaço de cores CIELAB.

A análise de regressão linear indicou que o total de variância que pôde ser explicada pela variável independente (rs12913832A>G), foi de 36% (R^2 ajustado = 0,36), 14% (R^2 ajustado = 0,14) e 26% (R^2 ajustado = 0,26) para as variáveis dependentes L*, a* e b*, respectivamente (dados não mostrados). Liu *et al.* (2010) e Andersen *et al.* (2013), também encontraram uma variância majoritariamente explicada por este polimorfismo em populações europeias (44%, R^2 ajustado = 0,44 e 68%, R^2 ajustado = 0,68, respectivamente). Ainda que ambos os estudos tenham utilizado metodologias diferentes da aqui apresentada para avaliar quantitativamente a cor dos olhos através de fotografias, isto é, medidas diferentes do espaço de cor CIELAB, as duas metodologias, mais a aqui apresentada, tiveram em comum o fato de apontarem para o rs12913832A>G como o polimorfismo capaz de explicar grande parte da variação existente, corroborando a importância deste polimorfismo na determinação da cor dos olhos.

Apesar da forte relação com a pigmentação dos olhos, associações com outros fenótipos que não a cor do olhos foram também descritas para este polimorfismo. Branicki *et al.* (2009), e Han *et al.* (2008) e Walsh *et al.* (2013) relataram associações com cor dos cabelos e pele, o que também foi encontrado em nosso estudo. Para a cor dos cabelos, o alelo "G" foi associado a cabelos loiros quando comparado a cabelos castanhos [OR = 4,996 (IC 95% = 2,67 - 9,33) $p = 4,57 \times 10^{-7}$] e pretos [OR = 8,535 (IC 95% = 2,43 - 29,88) p =

 $7,0x10^{-4}$] (Tabela 6). Para cor de pele, o alelo "G" se mostrou associado a uma maior chance de ter pele clara *vs.* pele não clara [OR = 1,730 (IC 95% = 1,16 - 2,57), p = 0,0044] e intermediária [OR = 1,795 (IC 95% = 1,20 - 2,68), p = 0,0044] (Tabela 7). Liu *et al.* (2015) observou a associação do alelo "A" a pele escura através de uma metodologia diferente para avaliar a cor da pele quantitativamente. Cerqueira *et al.* (2014) observou em uma amostragem da população brasileira a associação do alelo "C" do polimorfismo rs1129038C>T com pele escura por meio de medidas quantitativas (índice de melanina). Isso representa uma associação equivalente à aqui relatada, visto que o alelo "C" deste polimorfismo, que está localizado na região *3'UTR* do gene *HERC2*, está em um alto desequilíbrio de ligação com o alelo "A" do rs12913832A>G. Han *et al.* (2008) ainda evidencia uma possível associação com a capacidade de bronzeamento da pele. Apesar do alelo "G" mostrar-se associado ao índice de melanina em região não exposta [*OR* -0,055 (IC 95% = -0,09 - -0,01) p = 0,0109] e exposta [*OR* = -0,0479 (IC 95% = -0,08 - 0,00) p = 0,0204] ao sol, tal associação com a capacidade de bronzeamento não foi evidenciada no presente estudo (Tabela 9).

Algumas estruturas presentes na íris também foram associadas ao rs12913832A>G. O alelo "G" foi associado com maiores chances de aparecimento de pontos de pigmentação na íris [OR = 2,406 (IC = 1,46 - 3,96) p = 0,0005] e nódulos de Wolfflin [OR = 3,661 (IC = 1,66 - 7,20) p = 0,0001] (esta última mantendo-se significativa mesmo após a correção de Bonferroni para múltiplos testes). Este mesmo alelo foi associado como um fator protetor para aparecimento de sulcos de contração na íris [OR = 0,477 (IC 95% = 0,30 - 0,75) p = 0,0013] (Tabela 8).

Até o presente momento, apenas três grupos de pesquisa tem voltados seus esforços em caracterizar as bases genéticas que possam estar envolvidas no aparecimento de estruturas presentes na íris (Edwards *et al.*, 2016; Larsson *et al.*, 2011; Sturm e Larsson, 2009). No que diz respeito ao polimorfismo rs12913832A>G, até onde sabemos, esta é primeira vez que sua associação é descrita em estruturas presentes na íris. Ainda que os poucos estudos na literatura tenham associado estas estruturas com polimorfismos localizados em genes relacionados a embriogênese e desenvolvimento dos olhos, nenhum destes observou a possível associação com genes já estabelecidos como participantes na determinação da cor dos olhos.

Apesar das associações aqui descritas, é importante observar que ambas as estruturas (pontos de pigmentação na íris e nódulos de Wolfflin) estão presentes em maior frequência em indivíduos com a cor dos olhos mais claras (azul e verde), onde o alelo "G" deste polimorfismo está também associado (Figura 12A e E) (Tabela 5). Em relação ao sulcos de contração, a qual apresentou-se mais frequente em indivíduos com olhos escuros (Figura 12B; Tabela 8), o alelo "G" se mostrou como um fator protetor para o aparecimento desta estrutura. Desta forma, não se pode descartar que o polimorfismo rs12913832A>G apresente associação em nosso estudo devido à distribuição e frequência com que estas estruturas fazem-se presentes na amostragem. Estudos controlando para possíveis efeitos de confundimento para a cor dos olhos, envolvendo uma maior amostragem independente da aqui apresentada e, onde a distribuição dos indivíduos entre as diferentes categorias para cor dos olhos esteja mais homogênea, podem auxiliar na confirmação destes achados.

7.3.2. Associações de polimorfismos não-sinônimos localizados na região codificadora dos genes *OCA2* e *HERC2*

As regiões codificadoras dos genes aqui estudado apresentaram 11 polimorfismos individualmente associados a pigmentação. Dados obtidos através do relatório gerado pelo *web-server SNPnexus* mostraram que quatro destes polimorfismos são mutações sinônimas e sete são mutações não sinônimas (Tabela 21).

Tabela 21 - Lista de polimorfismos localizados na região codificante dos genes *OCA2* e *HERC2* associados no presente estudo com base no genoma de referência hg19.

Gene (éxon)	rsID (cromossomo:localização)	Aminoácidos
OCA2 (21)	rs1800419 (15:28096538)	Alanina > Alanina
OCA2 (17)	rs1800414 (15:28197037)	Histidina > Arginina
OCA2 (12)	rs1800407 (15:28230318)	Arginina > Glutamina
OCA2 (10)	rs1800404 (15:28235773)	Alanina > Alanina
OCA2 (9)	rs1800401 (15:28260053)	Arginina > Triptofano
OCA2 (2)	rs61738394 (15:28326942)	Glicina > Arginina
HERC2 (78)	rs11636232 (15:28386626)	Glutamina > Glutamina
HERC2 (68)	rs61756152 (15:28412872)	Alanina > Alanina
HERC2 (47)	rs74005645 (15:28447282)	Valina > Metionina
HERC2 (33)	rs118112076 (15:28474686)	Serina > Arginina
HERC2 (31)	rs201872292 (15:28475625)	Treonina > Metionina
	Gene (éxon) OCA2 (21) OCA2 (17) OCA2 (12) OCA2 (10) OCA2 (2) OCA2 (2) HERC2 (78) HERC2 (68) HERC2 (47) HERC2 (33) HERC2 (31)	Gene (éxon)rsID (cromossomo:localização)OCA2 (21)rs1800419 (15:28096538)OCA2 (17)rs1800414 (15:28197037)OCA2 (12)rs1800407 (15:28230318)OCA2 (10)rs1800404 (15:28235773)OCA2 (9)rs1800401 (15:28260053)OCA2 (2)rs61738394 (15:28326942)HERC2 (78)rs11636232 (15:28386626)HERC2 (68)rs61756152 (15:28412872)HERC2 (47)rs74005645 (15:28447282)HERC2 (33)rs118112076 (15:28474686)HERC2 (31)rs201872292 (15:28475625)

REF = Alelo referência; ALT = Alelo alternativo; rsID = Identificação do polimorfismo; Cinza = mutações não sinônimas

As mutações não-sinônimas afetam diretamente a estrutura da proteína formada e possivelmente sua função (Ramensky *et al.*, 2002; Teng *et al.*, 2012). Três dos sete polimorfismos com mutações não-sinônimas encontrados como associados a alguma característica de pigmentação no presente trabalho já foram descritos em outros estudos, sendo todos eles localizados no gene *OCA2*.

O alelo "C" do polimorfismo rs1800414T>C foi associado em nosso estudo com uma diminuição no índice de melanina tanto para região não exposta quanto para a exposta ao sol [beta = -0,323 (IC 95% = -0,53 - -0,10) p = 0,0033] e [beta = -0,204 (IC 95% = -0,40 - 0,00] p = 0,0457], respectivamente). Este polimorfismo é caracterizado pelo alelo alternativo "C", presente em uma alta frequência em indivíduos do leste asiático, e ausente em outros grupos populacionais, conforme estudos conduzidos por Edwards *et al.* (2010) e Eaton *et al.* (2015). Este alelo está presente em apenas nove dos 327 indivíduos analisados quantitativamente para o índice de melanina na presente amostragem. Destes nove indivíduos, oito apresentam uma composição ancestral que varia de 80% a 90% para o componente leste asiático, além de possuírem feições fenotípicas características de

asiáticos, como, por exemplo, a presença de dobra epicântica, e sobrenomes que remetem a populações asiáticas. Isto evidencia outro aspecto importante a respeito da população brasileira, cuja formação, apesar de majoritariamente europeia, africana e ameríndia, também teve a contribuição de outras populações como os japoneses, sírios e libaneses (Salzano, 1986).

O polimorfismo rs1800407C>T é caracterizado pela mudança do aminoácido arginina pela glutamina. Ainda que este apresente uma baixa frequência em todas as três principais populações que compuseram a população brasileira, ele possui um efeito relativamente alto na determinação da cor do olho. De acordo com Walsh *et al.* (2012), este polimorfismo é tido como o segundo com o maior fator preditivo para as diferentes colorações do olhos em populações europeias, estando associado com a penetrância de olhos verdes e mel quando na presença do genótipo heterozigoto AG ou homozigoto AA no sítio rs12913832A>G do gene *HERC2* (Mengel-From *et al.*, 2010; Pośpiech *et al.*, 2011).

Em nosso estudo, o polimorfismo rs1800407C>T mostrou-se associado tanto a olhos verdes quanto a olhos mel (Tabela 5). Ao analisar o efeito conjunto deste polimorfismo com o polimorfismo rs12913832A>G, observou-se que a combinação *multilocus* CC-GG, mostrou-se significante para olhos verdes *vs.* olhos não-verdes para a análise de epistasia com o pacote mb-mdr (dados não mostrados). Este achado corroborou o estudo de Póspiech *et al.* (2011), que observou a associação *multilocus* entre o genótipo "CC" do rs1800407C>T e "GG" do rs12913832A>G com olhos verdes quando comparados a olhos não-verdes. Não foram observadas associações epistáticas do rs1800407C>T e o rs12913832A>G com olhos mel (associação esta também relatada por Póspiech *et al.* (2011)).

O polimorfismo rs1800401G>A é caracterizado pela mudança do aminoácido arginina para triptofano. Alguns estudos não encontraram associação significativa (Andersen *et al.*, 2016; Mengel-From *et al.*, 2010), enquanto que apenas um estudo encontrou associação com olhos castanho quando comparado a olhos não-castanhos (Rebbeck *et al.*, 2002). Em nosso estudo, o alelo "A" foi associado como um fator protetor para o aparecimento de olhos castanho-claro *vs.* castanho-escuro [OR = 0,377 (IC 95% = 0,15 - 0,93) p = 0,0356], e a um fator de risco para olhos castanho-escuro *vs.* olhos não-castanho-escuro [OR = 2,921 (IC 95% = 1,33 - 6,41) p = 0,0075] (Tabela 5). É interessante observar polimorfismos que possam distinguir entre cores mais homogêneas, como, por exemplo, diferentes tons de castanho, uma vez que a grande maioria dos polimorfismos descritos na literatura tem sido associada a distinção entre fenótipos extremos como cor de olho azul *vs.* olho castanho.

Além dos três marcadores não-sinônimos supracitados previamente associados a variação normal da pigmentação humana, outros ainda não relatados mostraram-se associados em nosso estudo: rs61738394C>T, rs74005645C>T, rs118112076G>C e rs201872292G>A.

A mutação não-sinônima rs61738394C>T é caracterizada pela mudança de um aminoácido glicina para arginina e o alelo "T" foi associado com um aumento do índice de melanina em região da pele exposta ao sol [beta = 0,207 (IC 95% = 0,01 - 0,39) p = 0,0309],

ou seja, ao fenótipo de pele escura (Tabela 10). Os genótipos homozigoto "TT" e heterozigoto "CT" foram previamente reportados em indivíduos com albinismo oculocutâneo (Oetting et al., 2005; Rosenmann et al., 2009; Spritz et al., 1997). O relatório gerado pelo web-server SNPnexus indicou que a mutação é benigna, isto é, não acarreta mudanças drásticas na proteína de modo a causar possíveis danos, ainda que uma análise mais detalhada através do UCSC Genome Browser evidencie uma conservação do aminoácido em espécies de primatas (Apêndice E). Aparentemente, este alelo está presente apenas em populações africanas do 1000Genomes, porém com uma frequência extremamente baixa (Tabela 10), similarmente a frequência observada em nosso estudo. Todos os sete indivíduos portadores de tal alelo foram classificados como apresentando ou pele intermediária (pele tipo IV na escala Fitzpatrick), ou pele escura (pele tipo V na escala Fitzpatrick). Uma vez que este polimorfismo não foi associado com o índice de melanina para região não-exposta ao sol, isto é, a cor da pele constitutiva, este poderia estar associado a capacidade de bronzeamento, ainda que este apresente-se em uma baixa frequência alélica estudos mais abrangentes e de maior porte no que se refere a tamanho amostral devem ser conduzidos de modo a confirmar o efeito deste polimorfismo na variação normal da pigmentação da pele.

O polimorfismo rs74005645C>T apresenta uma substituição de um aminoácido valina por metionina e o alelo "T" foi associado com o aparecimento de olhos castanho-escuros quando comparado a olhos não castanho-escuros [OR = 3,720 (IC 95% = 1,40 - 9,86) p = 0,0083] (Tabela 5). Quando observamos a frequência deste alelo nas populações do 1000Genomes, ele é majoritariamente encontrado em populações africanas, com uma frequência de aproximadamente 53%. Não observa-se este alelo em populações europeias e apenas 0,5% em populações do leste asiático. No presente estudo, o alelo apresentou uma frequência de 12% (Tabela 5). De acordo com o relatório gerado pelo *SNPnexus*, a mutação é considerada benigna, e sua conservação não é alta entre diferentes espécies (Apêndice F), o que está de acordo com a classificação benigna da mutação.

Outro polimorfismo não-sinônimo que mostrou-se associado com alguma característica de pigmentação pelas análises de regressão independentes, foi o rs118112076G>C. A mutação é caracterizada pela mudança de um aminoácido serina para arginina, e é considerada benigna pelo relatório gerado pelo *SNPnexus*. Dentre o conjunto de aminoácidos próximos, ele é o que apresenta menor conservação entre espécies de vertebrados, dando suporte à classificação deste como benigno (Apêndice G). O alelo "C" deste polimorfismo foi associado ao aparecimento de sardas em nossa amostragem [*OR* = 4,357 (IC 95% = 1,18 - 16,06) p = 0,0270] (Tabela 7) e sua baixa frequência alélica em nosso estudo (1,8%) é condizente com a encontrada em populações europeias do *1000Genomes* (1,6%). Além disso, cabe ressaltar que a presença de sardas é mais frequente em populações europeias ou a elas relacionadas. Portanto a associação aqui apresentada pode fazer sentido quando observada sob uma ótica da frequência deste alelo em populações europeias, ainda que sua frequência seja extremamente baixa para um possível efeito de causa no aparecimento de sardas.

Por fim, o último polimorfismo não-sinônimo que mostrou-se associado com alguma característica de pigmentação foi o rs201872292G>A. Interessantemente, o alelo "A" foi associado como um fator protetor para o aparecimento de olhos castanho-claro vs. castanho-escuro [OR = 0,377 (IC 95% = 0,15 - 0,93) p = 0,0356], a um fator de risco para olhos castanho-escuro vs. olhos não-castanho-escuro [OR = 2,921 (IC 95% = 1,33 - 6,41) p = 1,330,0075] (Tabela 5), e a um aumento na variável quantitativa b* do espaço de cor CIELAB [beta = 3,950 (IC 95% = 1,61 - 6,28) p = 0,0010] (Tabela 9). O alelo "A" tem uma substancial frequência em populações do leste asiático quando comparado as populações africanas e europeias (29%, 4% e 1,8%, respectivamente). Este alelo foi também mais frequente em populações latinas (32%) que compuseram o ExAC (do inglês, Exome Aggregation Consortium) (Lek et al., 2016), que agregou além das populações que compõe o 1000Genomes, outras populações totalizando 60.706 indivíduos analisados. A frequência deste alelo no presente estudo foi de 8%, o que pode ter sido inflada pelos indivíduos com descendência asiática presentes na amostragem, e também pelo componente ameríndio presente na população brasileira. A mudança de aminoácido causada pela mutação do rs201872292A>G é a de um aminoácido treonina para metionina. O aminoácido ancestral (treonina) é conservado em primatas e em grande parte dos mamíferos placentários (Apêndice H). Segundo a literatura, assume-se que sequências de proteínas observadas em diferentes organismos tenham sobrevivido a seleção natural, e por conseguinte, posições de aminoácidos que mostram-se conservados em diferentes espécies possivelmente possuem um papel funcional importante, com as substituições de aminoácidos observadas em posições conservadas possivelmente acarretando em efeitos prejudiciais nas funções dos genes (Choi, 2012; Cichorek et al., 2013). Além da associação independente reportada na análise de regressão, o rs201872292A>G mostrou-se presentes nos modelos de associação multilocus nas análises de epistasia, sendo observados (em ordem de frequência de ocorrência) nos modelos de interação para as variáveis quantitativas a*, L*, b* e para pele exposta ao sol (Tabelas 11-16, 19 e 20). O relatório gerado pelo SNPnexus forneceu informações conflitantes a despeito do seu efeito previsto na função proteica, sendo indicado como tolerante e possivelmente danoso. O presente trabalho evidencia que a associação encontrada está de acordo com a distribuição alélica mundial. De fato, indivíduos asiáticos e ameríndios possuem uma diversidade muito baixa em relação a cor dos olhos, estando praticamente restrita a coloração castanho e diferentes tonalidades deste (Rawofi et al., 2017).

7.3.3. Associações de polimorfismos sinônimos localizados na região codificadora dos genes OCA2 e HERC2

As mutações sinônimas consistem naquelas em que há a mudança de um único nucleotídeo, sendo que esta não altera o aminoácido gerado, sendo assim conhecidas como mutações silenciosas. Por esse motivo, muitas vezes são negligenciadas em detrimento de variantes não-sinônimas (quando há a mudança do aminoácido formado), visto que estas possuem uma probabilidade maior de ocasionar erros na produção da proteína e possivelmente efeitos danosos ao indivíduo (Agrawal *et al.*, 2012; Sankin *et al.*, 2014). Entretanto, evidências de consequências moleculares funcionais de mutações sinônimas influenciam o processo de expressão gênica, podendo afetar o correto *splicing* de prémRNAs; alterar a estabilidade do mRNA e também a velocidade da síntese proteica (Bali e Bebok, 2015; Hunt *et al.*, 2014).

Quatro polimorfismos envolvendo mutações sinônimas foram associados com características de pigmentação no presente estudo: rs1800419A>G e rs1800404C>T do gene OCA2, e rs11636232C>T e rs61756152G>A do gene *HERC2*. Destes quatro, três foram previamente citados como associados a alguma característica de pigmentação (rs1800419A>G, rs1800404C>T e rs11636232C>T), e apenas o rs61756152G>A não foi reportado (Andrade *et al.*, 2017; Duffy *et al.*, 2007; Freire-Aradas *et al.*, 2014; Jannot *et al.*, 2005; Lins *et al.*, 2011; Mengel-From *et al.*, 2010).

O polimorfismo rs1800419A>G codifica um aminoácido alanina e o alelo "A" foi associado a maiores chances de possui olhos cor mel vs. castanho-escuros [OR = 2,976 (IC 95% = 1,17 - 7,54) p = 0,0215], e a menores chances de possuir olhos castanho-escuros vs. olhos não castanho-escuros [OR = 0.597 (IC 95% = 0.36 - 0.97) p = 0.0403] (Tabela 5). Estas duas associações se mostram consistentes, uma vez que o alelo "A" estaria associado a fenótipos mais claros. Ademais, Duffy et al. (2007) também reportou a associação do alelo "A" com olhos azuis e o alelo "G" com olhos não azuis em amostras de gêmeos da Austrália. Entretanto, o alelo "A" foi também associado no presente estudo a menores chances de aparecimento de cabelos loiros, ou seja, um típico fenótipo claro [OR = 0.5347] (IC 95% = (0,30 - 0,92) p = (0,0256) (Tabela 6). Embora a associação com olhos e cabelos seja aparentemente contraditória, é preciso levar em consideração que a proteína OCA2 pode apresentar papel funcional diferente na pigmentação dos olhos e cabelos. Assim como para outros polimorfismos discutidos anteriormente, o rs1800419A>G também mostrou-se presente nos modelos de associação multilocus nas análises de epistasia, sendo observados (em ordem de frequência de ocorrência) nos modelos de interação para as variáveis quantitativas a*, L* e b* do espaço de cor CIELAB (Tabelas 11-16). As análises de epistasia foram consistentes com os achados das associações independentes, uma vez que vemos o genótipo "AA" sistematicamente interagindo com genótipos de outros SNPs para determinar fenótipos mais claros. A frequência alélica do alelo "A" é praticamente a mesma em europeus e leste asiático, seguidos de africanos (56,3%, 56,5% e 37,1%, respectivamente), e no presente estudo apresentou uma frequência de 49% (Tabela 5).

O polimorfismo rs1800404C>T codifica um aminoácido alanina e teve o alelo "T" associado com a maior chance de apresentar pele clara quando comparado a pele não clara [OR = 1,618 (IC 95% = 1,02 - 2,54) p = 0,0345] e pele intermediária [OR = 1,617 (IC 95% = 1,03 - 2,52) p = 0,0345] (Tabela 7). Em um trabalho de nosso grupo (Andrade *et al.*, 2017), envolvendo outra amostragem da mesma localidade, o alelo "T" também foi associado com fenótipos claros (cor da pele, cabelos e olhos), sendo confirmado também por análise de

haplótipos quando em conjunto com os polimorfismos rs1800416T>G e rs1800407C>T. Na análise de epistasia do presente estudo o rs1800404C>T foi um dos polimorfismos que, assim como o rs201872292G>A citado anteriormente, mais mostrou-se presentes nos modelos de associação multilocus, sendo observado (em ordem de frequência de ocorrência) nos modelos de interação para todas as variáveis quantitativas analisadas (a*, índice de melanina em região exposta ao sol, L*, índice de melanina em região não exposta ao sol e b*) (Tabelas 11-20). Quando observadas as interações epistáticas envolvendo o índice de melanina em regiões exposta e não exposta ao sol, a associação do alelo "T" foi compatível com pele clara. Por outro lado, quando observadas as interações epistáticas do genótipo "TT" para as variáveis L*, a* e b*, houveram modificações no valor do coeficiente beta compatíveis com olhos escuros. O fato de observar um mesmo alelo ("T") associado ao aparecimento de um fenótipo claro e um fenótipo escuro evidencia a complexidade da epistasia per se, bem como da regulação da pigmentação em diferentes tecidos. Ademais, cabe ressaltar que as análises aqui realizadas partem do menor tipo de interação em termos de magnitude, isto é, interações entre dois *loci*, quando na realidade, supõe-se que muito outros estejam atuando de forma sinérgica. A frequência do alelo "T" é maior em populações europeias, chegando a cerca de 84%, seguido de 36% para o leste asiático e 10% em populações africanas (Tabela 7). As frequências tanto no presente estudo, quanto no estudo de Andrade et al. (2017) para a população de Ribeirão Preto foi de 59,4% e 62,3%, respectivamente.

A variante rs11636232C>T teve seu alelo derivado "T" associado a menor probabilidade de surgimento de cabelos ruivos quando comparado a loiros |OR = 0,103| (IC 95% = 0,01 - 0,95) p = 0,0450] (Tabela 6). Mengel-From et al. (2010), associou o alelo "T" com o aparecimento de olhos azuis quando comparado a olhos castanhos em dinamarqueses. No estudo de Andrade et al. (2017), resultados aparentemente contraditórios foram observados, uma vez que o alelo "T" foi associado a fenótipos claros (olhos azuis/verdes) e escuros (cabelos castanhos), o que novamente evidencia a importância de avaliar interações epistática com outros polimorfismos. Freire-Aradas et al. (2014) incluiu o rs11636232C>T em três modelos de predição para cor dos olhos em um estudo com 99 indivíduos de duas populações sul-americanas (79 de Natal, Rio Grande do Norte, e 20 indivíduos de Barquisimeto, Venezuela), enfatizando a importância de possíveis SNPs com menores efeitos contribuindo na classificação de diferentes fenótipos para cor dos olhos. A análise de epistasia revelou este polimorfismo presente nos modelos de associação multilocus, sendo observados (em ordem de frequência de ocorrência), nos modelos de interação para todas as variáveis quantitativas aqui analisadas (a*, b*, L*, índice de melanina em região não exposta ao sol e índice de melanina em região exposta ao sol) (Tabelas 11-20), estando em concordância com as associações independentes para o alelo "T" com fenótipos claros, uma vez que o genótipo "TT" mostrou-se associado a fenótipos claros quando interagindo com outros SNPs. Além disso, o relatório gerado pelo SNPnexus apontou efeitos benignos para esta mutação sinônima, não acarretando em alterações danosas para a proteína. Por fim, não se pode descartar efeitos de amostragem sobre a

discrepância observada envolvendo o alelo "T" entre nosso estudo e os outros relatados na literatura.

O último polimorfismo reportado no presente estudo nas análises de associação independentes para as regiões codificadoras dos genes OCA2 e HERC2 foi o rs61756152G>A. Este polimorfismo teve o alelo "A" associado com maiores chances de possuir olhos verdes em comparação a olhos castanho-escuros [OR = 4,497 (IC 95% = 1,33 - 1,33)]15,21 p = 0,0155] (Tabela 5). A sua frequência nas populações do 1000Genomes foi de 5%, 9% e 29% aproximadamente nas populações europeias, africanas e do leste asiático, e de 10% no presente estudo (Tabela 5). É interessante observar a alta frequência do alelo associado a olhos verdes na população brasileira em populações asiáticas. Uma vez que olhos verdes não ocorrem naturalmente no leste da Ásia, é possível que interações epistáticas estejam relacionadas ao presente achado de associação independente. Embora tal marcador não esteja presente entre as 10 principais interações epistáticas relatadas nas Tabelas 12, 14, 16, 18 e 20, foram observadas interações menos significativas (dados não mostrados). Adicionalmente, não se pode descartar interações com polimorfismos de outros genes, o que não foi considerado no presente estudo. Por fim, o relatório gerado pelo SNPnexus apontou uma predição de possivelmente benigna para esta mutação sinônima, não acarretando maiores danos na função proteica.

7.3.4. Associações de polimorfismos localizados na região intrônica dos genes OCA2 e HERC2

Além de regiões regulatórias como *enhancers* e *insulators*, que são alguns elementos presentes nas proximidades da região promotora dos genes e que estão envolvidos na regulação da expressão gênica (Irimia e Roy, 2014), elementos regulatórios também podem estar presentes em regiões intrônicas, regulando a expressão propriamente dita ou outros processos, tais como o *splicing* que levará a expressão de diferentes isoformas proteicas (Wang *et al.*, 2012).

Nas últimas décadas, algumas funções atribuídas aos íntrons começaram a surgir na literatura. Em particular, regulando a remoção de íntrons e a junção subsequente de éxons (o chamado processamento intrônico), os genes eucarióticos podem gerar múltiplos transcritos, expandindo vastamente a diversidade molecular. Inicialmente sugerido por Gilbert (1978), o processo, conhecido como *splicing* alternativo, parece ser difundido em eucariotos, aparentemente atingindo seu ápice em mamíferos (Barbosa-Morais *et al.*, 2012), no qual 95% dos genes compostos por vários éxons sofrem *splicing* alternativo (Wang *et al.*, 2008, Pan *et al.*, 2008).

No presente trabalho, o ensaio customizado foi planejado de modo a abranger as regiões regulatórias e codificantes, bem como alguns fragmentos de íntrons que apresentavam polimorfismos previamente relatados na literatura (como, por exemplo, o rs12913832A>G do gene *HERC2*). Entretanto, alguns fragmentos intrônicos também foram capturados devido a sua proximidade com as extremidades 5' e 3' dos éxons. Cabe ressaltar que a cobertura de captura (número de sondas distintas) e sequenciamento destes últimos fragmentos é muito menor do que a dos alvos planejados. Portanto, é possível que sejam dados mais susceptíveis a erros de genotipagem.

Cinco polimorfismos localizados em íntrons dos genes *OCA2* e *HERC2* foram associados com alguma característica de pigmentação no presente trabalho. Os polimorfismos rs12910433A>G, rs13313425G>A e rs41304383T>C estão localizados no gene *OCA2*, enquanto que os polimorfismos rs6497271G>A e rs58358300A>C estão localizados no gene *HERC2*.

O polimorfismo rs12910433A>G foi associado com maiores chances de aparecimento de olhos mel contra olhos castanho-claros [OR = 5,949 (IC 95% = 1,98 - 17,81) p = 0,0014] (Tabela 5). A frequência do alelo "G" no presente trabalho foi de 49,7%, e nas populações europeias, africanas e do leste asiático do 1000Genomes foi de 24,7%, 90,6% e 84,6%, respectivamente (Tabela 5). Este polimorfismo foi reportado por Murray et al. (2015), como parte de um bloco haplotípico nas populações do leste asiático, em que os polimorfismos não sinônimos rs1800414T>C e o rs74653330C>T (não reportado no presente trabalho) são associados ao clareamento da pele em asiáticos. Apesar do rs12910433A>G ser encontrado em um alto desequilíbrio de ligação com o rs1800414T>C no leste asiático, eles não se encontram em LD no presente trabalho ($R^2 = 0.02$ e D' = 1.00), o que pode ser decorrente da miscigenação presente na população brasileira ou de outros aspectos demográficos inerentes a ela. A análise de epistasia revelou que o rs12910433A>G mostrou-se presente nos modelos de associação multilocus, sendo observados (em ordem de frequência de ocorrência) nos modelos de interação para todas as variáveis quantitativas analisadas no presente estudo (índice de melanina em região exposta ao sol, índice de melanina em região não exposta ao sol, b*, a* e L*) (Tabelas 11-20). As análises foram consistentes com aquilo observado para as associações independentes. Por fim, o relatório gerado pelo SNPnexus evidenciou que o rs12910433A>G está a 15 nucleotídeos de um sítio de splicing em uma região com sequências de "AG" (Adenina e Guanina), as quais estão associadas a regulação por splicing alternativo (Burset et al., 2000).

A variante rs13313425G>A teve o seu alelo "A" associado com a diminuição no índice de melanina para a região não exposta ao sol [beta = -0,083 (IC 95% = -0,16 - 0,00) p = 0,0360] (Tabela 10). As análises *multilocus* de epistasia também evidenciaram a presença deste polimorfismo nos modelos de interação para as variáveis quantitativas de índice de melanina para região não exposta e exposta ao sol (Tabelas 17-20). A análise de epistasia corrobora a associação independente encontrada para uma diminuição do índice de melanina em região não exposta ao sol com os genótipos heterozigoto "GA" e homozigoto "AA" associados a uma pequena redução do índice de melanina, representada pela diminuição do valor do coeficiente beta para esta variável. Interessantemente, o alelo "A" é substancialmente mais frequente nas populações africanas quando comparada as europeias e do leste asiático do *1000Genomes* (38,8%, 2,4% e 3,8%, respectivamente) e do presente trabalho (7,8%) (Tabela 10).

O polimorfismo rs41304383T>C teve o alelo "C" associado com a maior chance de apresentar pele clara em nosso estudo quando comparado a pele não clara [OR = 4,200 (IC 95% = 1,50 - 11,70) p = 0,0060] e pele intermediária [*OR* = 3,983 (IC 95\% = 1,42 - 11,12) p = 1,500,0083] (Tabela 7). O rs41304383T>C foi também um dos polimorfismos que mais mostrouse presentes nos modelos de associação multilocus nas análises de epistasia, sendo observados nos modelos de interação para todas as variáveis quantitativas analisadas (Tabelas 11-20). A frequência do alelo "C" é baixa nas populações europeias e africanas do 1000Genomes (5% e 5,9%, respectivamente), e nula em populações do leste asiático (Tabela 7). No presente estudo, a frequência do alelo "C" é de 4,6%, e desta forma próxima das frequências europeias e africanas. Assim como para o rs12910433A>G descrito anteriormente, o rs41304383T>C também encontra-se próximo a um sítio de splicing (18 nucleotídeos) e de acordo com o relatório do SNPnexus, também está associado a elementos regulatórios enhancers. Estudos apontam que o splicing alternativo pode ocorrer simultaneamente à transcrição (Matveeva et al., 2016), indicando que os dois processos podem levar a aumento da produção de transcritos OCA2 e/ou moléculas funcionalmente mais ativas.

No gene HERC2, o polimorfismo rs6497271G>A teve o seu alelo "A" associado a uma diminuição do coeficiente beta nas análises de regressão independente para a variável quantitativa L* do espaço de cor CIELAB [beta = -5,498 (IC 95% = -8,35 - -2,65) p = 1,96x10⁻⁴] (Tabela 9), ou seja, uma associação com pouco brilho/claridade. A alta frequência deste em populações africanas do 1000Genomes (27,7%) e baixa frequência em europeus (3,7%) e leste asiáticos (0,5%) (Tabela 9), condiz com a associação encontrada, uma vez que olhos com coloração escura, como por exemplo castanho-escuros, são mais frequentes em populações africanas e olhos castanho-escuros possuem menos brilho/claridade quando comparado a outras cores, conforme evidenciado por Edwards et al. (2015). A análise de epistasia revelou que o polimorfismo rs6497271G>A está presente em modelos de interação multilocus para as variáveis quantitativas L* e região da pele não exposta ao sol (Tabela 11, 12 e 17, 18). Para a associação independente da variável L*, o genótipo "GG" foi associado com o aumento do valor de L* em uma das dez interações mais significantes para esta variável. Uma análise mais detalhada através do UCSC Genome Browser evidenciou que a posição do rs6497271G>A mostra-se extremamente conservada em espécies de primatas e também mamíferos placentários (Apêndice I), sugerindo um possível papel biológico ou funcional, ainda que distante do sítio de *splicing* mais próximo (1061 nucleotídeos), segundo relatório gerado pelo SNPnexus. Por fim, a consulta no HaploReg evidenciou a associação com modificações do estado da cromatina (enhancers), e modificações de histonas (H3K4me1 e H3K27ac_Enh).

O polimorfismo rs58358300A>C teve o alelo derivado "C" fortemente associado ao aparecimento de pigmentação na esclera da íris [OR = 9,247 (IC 95% = 2,97 - 28,79) p = 0,0001] (Tabela 8). Como dito anteriormente no tópico 7.2, esta característica parece estar
associada com indivíduos cuja ancestralidade remetem a populações com predominância de olhos escuros, como pode ser observado no estudo de Edwards et al. (2015), que observou a presença desta característica em maior frequência em indivíduos do sul e leste asiático, quando comparados a indivíduos com ancestralidade europeia. No presente estudo o alelo "C" teve uma frequência total de 10,4% e cerca de 40% quando observado nos indivíduos apresentando esta característica (dados não mostrados). A frequência do alelo "C" é maior nas populações africanas do 1000Genomes (55,3%), seguido de populações do leste asiático (0,2%) e populações europeias (0,1%), o que é consistente com a possível associação desta característica com indivíduos apresentando olhos escuros (Tabela 8). A análise de epistasia revelou ainda que o rs58358300A>C está presente nos modelos de interação para as variáveis de índice de melanina em região não exposta e exposta ao sol (Tabela 17-20), onde o genótipo "CC" foi associado a um aumento no índice de melanina e o genótipo heterozigoto "CA" associado a uma diminuição do índice de melanina. É interessante ressaltar que, de todas as características de estruturas secundárias presentes na íris, nenhum marcador genético foi previamente associado a pigmentação na esclera (Edwards et al., 2016; Larsson e Pedersen, 2004; Larsson et al., 2011). De acordo com o SNPnexus e HaploReg o polimorfismo rs58358300A>C está localizado em uma região de íntron, a 148 nucleotídeos do sítio de splicing mais próximo, e marcação de histonas (H3K4me1) foram detectadas neste locus, sugerindo que este possa atuar como um enhancer, podendo indicar um possível efeito funcional, ainda que mais estudos sejam necessários para a verificação de tal suposição.

Por fim, é importante enfatizar que todas as análises aqui apresentadas foram realizadas controlando para os possíveis efeitos da ancestralidade. Ainda assim, não se pode descartar a possibilidade de que as associações aqui evidenciadas representem falso-positivos. Adicionalmente, embora o desequilíbrio de ligação tenha sido levado em consideração nas análises realizadas, buscando evitar associações redundantes com as variantes que foram identificadas no presente estudo como possivelmente causais, não se pode descartar a possibilidade de que futuramente haja outro entendimento quanto a esta questão, uma vez que a verdadeira variante causal pode não ter sido capturada neste estudo.

7.3.5. Epistasia e sua importância nos estudos genéticos de pigmentação

O termo "epistasia" foi usado pela primeira vez por Bateson (1909), para descrever um efeito de mascaramento pelo qual uma variante ou alelo em um *locus* impede a variante em outro *locus* de manifestar seu efeito.

No presente trabalho, análises de epistasia para avaliar a interação *multilocus* de polimorfismos ao longo das regiões regulatórias e codificadora dos genes *OCA2* e *HERC2* foram realizadas. Este tipo de abordagem tem sido realizada ao longo da última década na tentativa de elucidar os mecanismos genéticos envolvidos na variância normal da

pigmentação em humanos de diversas populações mundiais (Branicki *et al.*, 2009; Pośpiech *et al.*, 2011; Ruiz *et al.*, 2013; Ulivi *et al.*, 2013; Maroñas *et al.*, 2014; Pośpiech *et al.*, 2014). Do ponto de vista de complexidade, é natural pensar que fenótipos de pigmentação, cuja quantidade de genes envolvidos com as diversas vias que envolvem o processo de melanogenese chega a mais de 170, esteja suscetível a tal processo. Considerando a proximidade dos genes *OCA2* e *HERC2*, interações epistáticas entre eles não podem ser descartadas.

Isto pode ser observado pela quantidade de modelos de interação de segunda ordem observados no presente trabalho para as diferentes medidas quantitativas de pigmentação analisadas (Tabelas 11-20). Alguns dos polimorfismos associados independentemente através das análises de regressão logística (para variáveis categóricas) e linear (variáveis quantitativas) foram também observados nos modelos de interação aqui apresentados. Porém, ainda mais interessante é observar que polimorfismos ao longo dos dois genes aqui estudados que não foram associados pelas análises de regressão independentes, foram associados quando avaliados sobre um contexto epistático (*multilocus*). Observação esta, também evidenciada por Frudakis *et al.* (2003).

Em termos de acurácia na predição de algumas categorias de fenótipos de pigmentação, todos os estudos conduzidos na literatura até o momento não são satisfatórios, especialmente em casos onde são avaliados fenótipos intermediários, como, por exemplo, a cor de olhos verdes e mel (Liu *et al.*, 2009; Spichenok *et al.*, 2011; Walsh *et al.*, 2013). Uma solução potencial para aumentar a acurácia da predição de características de pigmentação intermediárias é via interações gene-gene e sua consideração na modelagem de predição. Vários estudos anteriores já relataram que os efeitos epistáticos podem ser relevantes para fenótipos de pigmentação (Akey *et al.*, 2001; Branicki *et al.*, 2009; Eiberg *et al.*, 2008; Liu *et al.*, 2010). Pośpiech *et al.* (2014), relatou a significância estatística dos efeitos epistáticos entre *HERC2* e *OCA2* e entre *HERC2* e *SLC24A4* e revelou uma nova interação entre *HERC2* e *TYRP1* na determinação da cor dos olhos em humanos. Ademais, é importante enfatizar que além de Maroñas *et al.* (2014), que avaliou o desenvolvimento de um teste para uso forense preditivo para cor da pele, onde em sua amostragem continha apenas três indivíduos brasileiros, nenhum outro estudo buscou avaliar os possíveis efeitos epistáticos em genes de pigmentação em brasileiros.

Por fim, a análise de epistasia aqui realizada constituí um bom primeiro passo em direção a busca por interações de outras magnitudes (3x3, 4x4 e assim por diante), refinando ainda mais os achados e possibilitando a adição de outros polimorfismos localizados em genes diferentes dos aqui estudados como por exemplo os genes *ASIP*, *SLC24A5*, *SLC45A2*, *MC1R* (do inglês, *Melanocortin 1 Receptor*) e *IRF4*, descritos também na literatura como importantes no papel da pigmentação.

8 CONCLUSÃO

O presente trabalho teve como principais conclusões:

A cor da pele autodeclarada pode não representar adequadamente a ancestralidade individual, o que pôde ser observado, uma vez que muitos indivíduos que se autoclassificaram como brancos possuíam de fato uma ancestralidade africana substancialmente alta, da mesma forma que indivíduos que se auto-classificaram como pretos tinham valores significativos de ancestralidade europeia, ainda que uma correlação entre cor de pele e ancestralidade tenha sido observada.

A análise quantitativa para cor dos olhos através do uso de fotografias mostrou-se relativamente eficaz. Apesar das dificuldades encontradas devido à pouca qualidade de algumas das fotos utilizadas, a concordância entre as variáveis quantitativas e classificação categórica realizada por nosso laboratório foi considerada satisfatória uma vez que as classificações para olhos azuis e castanho sobrepuseram-se ao esperado no eixo de cor L*a*b* do espaço de cor CIELAB.

Este foi o primeiro trabalho no Brasil utilizando uma metodologia de Sequenciamento de Nova Geração para analisar os genes *OCA2* e *HERC2* aos diversos fenótipos de pigmentação normal na população brasileira, confirmando o polimorfismo do gene *HERC2* rs12913832 como o principal polimorfismo associado a cor dos olhos. Além disso, até onde sabemos, este foi o primeiro trabalho em que associou-se este polimorfismo com estruturas secundárias presentes na íris.

O uso de variáveis quantitativas para os olhos revelou pela primeira vez, até onde sabemos, a associação do polimorfismo não sinônimo rs201872292 no gene *HERC2* com olhos claros, independente do efeito do polimorfismo rs12913832.

Este foi o primeiro estudo no Brasil que se propôs a analisar polimorfismos genéticos em genes candidatos a variação normal da pigmentação humana com estruturas secundárias presentes na íris. Destaca-se a associação do polimorfismo rs58358300 localizado em um íntron do gene *HERC2* com pigmentação da esclera. Esta é a primeira vez que um polimorfismo é associado a esta característica, que apesar de ser a característica menos estudada do ponto de vista de associações com variantes genéticas, dentre todas as estruturas presentes na íris, estudos indicam sua importância clínica como um possível fator de risco para melanomas, ainda que muito raros (Damato e Coupland, 2008; Pinto-Proença *et al.*, 2018).

Os polimorfismos associados através das análises independentes de regressão neste trabalho estão localizados em regiões de íntrons e éxons, com apenas uma observação de associação para região *3'UTR*, com o polimorfismo rs1129038 localizado no gene *HERC2*, extensivamente relatado na literatura como em completo desequilíbrio de ligação com o polimorfismo rs12913832 do mesmo gene.

Alguns dos polimorfismos associados nas análises independentes de regressão logística e linear foram também associados quando analisados sob a perspectiva de interação epistática com outros polimorfismos. Entretanto, é interessante destacar o aparecimento de diversos outros polimorfismos que não tiveram sua associação independente observada, mas que foram associados quando em conjunto com outros polimorfismos, evidenciando a importância desta abordagem em trabalhos futuros, principalmente na tentativa de buscar polimorfismos que desempenhem um papel funcional na predição de fenótipos de pigmentação intermediários, como a cor de olhos verdes e mel, impactando de forma positiva na área da genética forense e identificação humana.

De modo geral, os resultados desse trabalho reafirmam a relevância dos genes *OCA2* e *HERC2* na determinação dos diversos fenótipos de cor do olho, pele, cabelos e estruturas secundárias da íris encontrados na população brasileira.

9 REFERÊNCIAS BIBLIOGRÁFICAS

Adzhubei I, Jordan DM, Sunyaev SR (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics*. Chapter 7:Unit7.20.

Agrawal N, Jiao Y, Bettegowda C *et al.* (2012). Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discovery*. 2:899-905.

Akey JM, Wang H, Xiong M *et al.* (2001). Interaction between the melanocortin-1 receptor and P genes contributes to inter-individual variation in skin pigmentation phenotypes in a Tibetan population. *Human Genetics.* 108:516-520.

Albert FW, Kruglyak L (2015). The role of regulatory variation in complex traits and disease. *Nature Reviews. Genetics.* 16:197-212.

Almeida RH (1997). O diretório dos índios: um projeto de "civilização" no Brasil do século XVIII. Brasília: *Editora UNB*; p. 370.

Andersen JD, Johansen P, Harder S *et al.* (2013). Genetic analyses of the human eye colours using a novel objective method for eye colour classification. *Forensic Science International Genetics*. 7:508-515.

Andersen JD, Pietroni C, Johansen P *et al.* (2016). Importance of nonsynonymous OCA2 variants in human eye color prediction. *Molecular Genetics & Genomic Medicine*. 4:420-430.

Andrade ES, Fracasso NCA, Strazza Júnior PS *et al.* (2017). Associations of OCA2-HERC2 SNPs and haplotypes with human pigmentation characteristics in the Brazilian population. *Legal Medicine* (*Tokyo*). 24:78-83.

Andrews S (2010). *fastq*c: A quality control tool for high throughput sequence data.

Bali V, Bebok Z (2015). Decoding mechanisms by which silent codon changes influence protein biogenesis and function. *The International Journal of Biochemistry & Cell Biology*. 64:58-74.

Barbosa-Morais NL, Irimia M, Pan Q *et al.* (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science*. 338:1587-1593.

Barrett JC, Fry B, Maller J *et al.* (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England)*. 21:263-265.

Bateson W. Mendel's principles of heredity. Cambridge: University Press; 1909

Becker KG, Barnes KC, Bright TJ *et al.* (2004). The genetic association database. *Nature Genetics* ; 36(5):431-432.

Bedoya G, Montoya P, García J *et al.* (2006). Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proceedings of the National Academy of Sciences.* 103:7234-7239.

Beleza S, Johnson NA, Candille SI *et al.* (2013). Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genetics*. 9:e1003372.

Bennett DC, Lamoreux ML (2003). The color loci of mice - a genetic century. *Pigment Cell Research*. 16:333-344.

Blanton RE, Silva LK, Morato VG *et al.* (2008). Genetic ancestry and income are associated with dengue hemorrhagic fever in a highly admixed population. *European Journal of Human Genetics*. 16:762-765.

Bonilla C, Boxill LA, Donald SA *et al.* (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*. 116:402-406.

Bouakaze C, Keyser C, Crubézy E *et al.* (2009). Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. *International Journal of Legal Medicine*. 123:315-325.

Branicki W, Brudnik U, Wojas-Pelc A (2009). Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. *Annals of Human Genetics*. 73:160-170.

Branicki W, Liu F, van Duijn K *et al.* (2011). Model-based prediction of human hair color using DNA variants. *Human Genetics*. 129:443-454.

Burset M, Seledtsov IA, Solovyev VV (2000). Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Research*. 28:4364-4375.

Calle ML, Urrea V, Malats N *et al.* (2010). mbmdr: an R package for exploring gene–gene interactions associated with binary or quantitative traits. *Bioinformatics*. 26:2198-2199.

Cardena MM, Ribeiro-Dos-Santos A, Santos S *et al.* (2013). Assessment of the relationship between self-declared ethnicity, mitochondrial haplogroups and genomic ancestry in Brazilian individuals. *PLoS One.* 8:e62005.

Casper J, Zweig AS, Villarreal C *et al.* (2018). The UCSC Genome Browser database: 2018 update. *Nucleic Acids Research*. 46:D762-D769.

Castelli EC, Mendes-Junior CT, Sabbagh A *et al.* (2015). HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. *Human Immunology*. 76:945-953.

Cerqueira CC, Hünemeier T, Gomez-Valdés J *et al.* (2014). Implications of the admixture process in skin color molecular assessment. *PLoS One*. 9:e96886.

Chelala C, Khan A, Lemoine NR (2009). *SNPnexus*: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics*. 25:655-661.

Cher C (2011). Bioanalyzer: Applications for Next Generation Sequencing : Updates and Tips. 1-49.

Chi A, Valencia JC, Hu Z-z *et al.* (2006). Proteomic and Bioinformatic Characterization of the Biogenesis and Function of Melanosomes research articles. *Journal of Proteome Research*. 5:3135-3144.

Cichorek M, Wachulska M, Stasiewicz A *et al.* (2013). Skin melanocytes: biology and development. *Advances in Dermatology and Allergology*. 30:30-41.

Damato B, Coupland SE (2008). Conjunctival melanoma and melanosis: a reappraisal of terminology, classification and staging. *Clinical & Experimental Ophthalmology*. 36:786-795.

Davydov EV, Goode DL, Sirota M *et al.* (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology*. 6:e1001025.

Dayem Ullah AZ, Lemoine NR, Chelala C (2012). *SNPnexus*: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Research*. 40:W65-70.

Dayem Ullah AZ, Lemoine NR, Chelala C (2013). A practical guide for the functional annotation of genetic variations using *SNPnexus*. *Briefings Bioinformatics*. 14:437-447.

del Marmol V, Beermann F (1996). Tyrosinase and related proteins in mammalian pigmentation. *Federation of European Biochemical Societies Letters*. 381:165-168.

DePristo MA, Banks E, Poplin R *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*. 43:491-498.

Donnelly MP, Paschou P, Grigorenko E *et al.* (2012). A global view of the OCA2-HERC2 region and pigmentation. *Human Genetics*. 131:683-696.

Duffy DL, Montgomery GW, Chen W *et al.* (2007). A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *American Journal of Human Genetics*. 80:241-252.

Edwards M, Bigham A, Tan J *et al.* (2010). Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. *PLoS Genetics.* 6:e1000867.

Edwards M, Cha D, Krithika S *et al.* (2016). Analysis of iris surface features in populations of diverse ancestry. *Royal Society Open Science* 3:150424.

Edwards M, Gozdzik A, Ross K *et al.* (2012). Technical note: quantitative measures of iris color using high resolution photographs. *American Journal of Physical Anthropology.* 147(1):141-9.

Eiberg H, Troelsen J, Nielsen M *et al.* (2008). Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Human Genetics*. 123:177-187.

ENCODE (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*. 489:57-74.

Erlich Y, Mitra PP, delaBastide M *et al.* (2008). Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nature Methods*. 5:679-682.

Ewing B, Green P (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*. 8:186-194.

Fausto BO. Fazer a América: A Imigração em Massa para a América. **Edusp - Editora da Universidade de São Paulo**; 2000:577.

Ferreira LB, Mendes-Junior CT, Wiezel CE *et al.* (2006). Genomic ancestry of a sample population from the state of São Paulo, Brazil. *American Journal of Human Biology*. 18:702-705.

Forbes SA, Bhamra G, Bamford S *et al.* (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). *Current Protocols in Human Genetics*. Chapter 10:Unit 10.11.

Freire-Aradas A, Ruiz Y, Phillips C *et al.* (2014). Exploring iris colour prediction and ancestry inference in admixed populations of South America. *Forensic Science International Genetics*. 13:3-9.

Frudakis T, Terravainen T, Thomas M (2007). Multilocus OCA2 genotypes specify human iris colors. *Human Genetics*. 122:311-326.

Frudakis T, Thomas M, Gaskin Z *et al.* (2003). Sequences associated with human iris pigmentation. *Genetics*. 165:2071-2083.

Galarneau G, Palmer CD, Sankaran VG *et al.* (2010). Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nature Genetics*. 42:1049-1051.

Gerstenblith MR, Goldstein AM, Fargnoli MC *et al.* (2007). Comprehensive evaluation of allele frequency differences of MC1R variants across populations. *Human Mutation*. 28:495-505.

Gilbert W (1978). Why genes in pieces. Nature. 271:501.

Good P (2000). Theory of Permutation Tests. New York, NY: Springer New York; p. 201-214.

Graf J, Voisey J, Hughes I *et al.* (2007). Promoter Polymorphisms in the MATP (SLC45A2) Gene Are Associated With Normal Human Skin Color Variation. *Human Mutation*. 28:710-717.

Guerreiro JF, Santos AKCR, Santos EJM *et al.* (1993). Genetic structure and demography of the human population of Obidos, in the brazilian Amazon. *Brazilian Journal of Genetics*. 16:1075-1084.

Gulko B, Hubisz MJ, Gronau I *et al.* (2015). A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nature Genetics*. 47:276-283.

Hahn LW, Ritchie MD, Moore JH (2003). Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics*. 19:376-382.

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

Hearing VJ (1999). Biochemical Control of Melanogenesis and Melanosomal Organization. *The Society for Investigative Dermatology*. 4

Hearing VJ (2011). Determination of melanin synthetic pathways. *Journal of Investigative Dermatology*. 131:E8-E11.

Holly EA, Aston DA, Char DH *et al.* (1990). Uveal melanoma in relation to ultraviolet light exposure and host factors. *Cancer Research*.

Horn EP, Hartge P, Shields JA et al. (1994). Sunlight and Risk of Uveal Melanoma. Journal of the National Cancer Institute. 86:1476-1478.

Hu DN, Wakamatsu K, Ito S *et al.* (2009). Comparison of eumelanin and pheomelanin content between cultured uveal melanoma cells and normal uveal melanocytes. *Melanoma Research*. 19:75-79.

Hubisz MJ, Pollard KS, Siepel A (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings Bioinformatics*. 12:41-51.

Hunt RC, Simhadri VL, Iandoli M *et al.* (2014). Exposing synonymous mutations. *Trends in Genetics*. 30:308-321.

Ionita-Laza I, McCallum K, Xu B *et al.* (2016). A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nature Genetics*. 48:214-220.

Irimia M, Roy SW (2014). Origin of spliceosomal introns and alternative splicing. **Cold Spring Harbor** *Perspective in Biology*. 6

Jablonski NG, Chaplin G (2000). The evolution of human skin coloration. *Journal of Human Evolution*. 39:57-106.

Jannot AS, Meziani R, Bertrand G *et al.* (2005). Allele variations in the OCA2 gene (pink-eyed-dilution locus) are associated with genetic susceptibility to melanoma. *European Journal of Human Genetics*. 13:913-920.

Kanetsky PA, Swoyer J, Panossian S *et al.* (2002). A polymorphism in the agouti signaling protein gene is associated with human pigmentation. *American Journal of Human Genetics*. 70:770-775.

Kassahn KS, Scott HS, Caramins MC (2014). Integrating Massively Parallel Sequencing into Diagnostic Workflows and Managing the Annotation and Clinical Interpretation Challenge. *Human Mutation*. 35:413-423.

Kayser M, Liu F, Janssens ACJW *et al.* (2008). Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene. *American Journal of Human Genetics*. 82:411-423.

Kircher M, Witten DM, Jain P *et al.* (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*. 46:310-315.

Kondo T, Hearing VJ (2011). Update on the regulation of mammalian melanocyte function and skin pigmentation. *Expert Reviews of Dermatology*. 6:97-108.

Kozomara A, Griffiths-Jones S (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*. 42:D68-73.

Kumar P, Henikoff S, Ng PC (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. 4:1073-1081.

Kundaje A, Meuleman W, Ernst J *et al.* (2015). Integrative analysis of 111 reference human epigenomes. *Nature*. 518:317-330.

Landrum MJ, Lee JM, Riley GR *et al.* (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*. 42:D980-5.

Lango Allen H, Estrada K, Lettre G *et al.* (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 467:832-838.

Lao O, de Gruijter JM, van Duijn K *et al.* (2007). Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Annals of Human Genetics*. 71:354-369.

Larsson M, Duffy DL, Zhu G *et al.* (2011). GWAS findings for human iris patterns: associations with variants in genes that influence normal neuronal pattern development. *American Journal of Human Genetics*. 89:334-343.

Larsson M, Pedersen NL (2004). Genetic correlations among texture characteristics in the human iris. *Molecular Vision*. 10:821-831.

Lassmann T, Hayashizaki Y, Daub CO (2011). SAMStat: monitoring biases in next generation sequencing data. *Bioinformatics*. 27:130-131.

Leite TK, Fonseca RM, de França NM *et al.* (2011). Genomic ancestry, self-reported "color" and quantitative measures of skin pigmentation in Brazilian admixed siblings. *PLoS One*. 6:e27162.

Lek M, Karczewski KJ, Minikel EV *et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 536:285-291.

Lewis CM (2002). Genetic association studies: Design, analysis and interpretation. **Briefings in Bioinformatics**. 3:146-153.

Li H, Handsaker B, Wysoker A *et al.* (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*. 15;25(16):2078-9.

Li H, Ruan J, Durbin R (2008). Mapping short DNA sequencing *reads* and calling variants using mapping quality scores. *Genome Research*. 18:1851-1858.

Lins TC, Vieira RG, Abreu BS *et al.* (2011). Genetic heterogeneity of self-reported ancestry groups in an admixed Brazilian population. *Journal of Epidemiology*. 21:240-245.

Liu F, van Duijn K, Vingerling JR *et al.* (2009). Eye color and the prediction of complex phenotypes from genotypes. *Current Biology : CB*. 19:R192-3.

Liu F, Wollstein A, Hysi PG *et al.* (2010). Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genetics*. 6:e1000934.

MacArthur J, Bowler E, Cerezo M *et al.* (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*. 45:D896-D901.

Machiela MJ, Chanock SJ (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 31:3555-3557.

Machiela MJ, Chanock SJ (2018). LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. *Bioinformatics*. 34:887-889.

Mackey DA, Wilkinson CH, Kearns LS *et al.* (2011). Classification of iris colour: review and refinement of a classification schema. *Clinical & Experimental Ophthalmology* 39:462-471.

Magalhães da Silva T, Sandhya Rani MR, de Oliveira Costa GN *et al.* (2015). The correlation between ancestry and color in two cities of Northeast Brazil with contrasting ethnic compositions. *European Journal of Human Genetics*. 23:984-989.

Manta FS, Pereira R, Caiafa A *et al.* (2013). Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. *Annals of Human Biology*. 40:94-98.

Maroñas O, Phillips C, Söchtig J *et al.* (2014). Development of a forensic skin colour predictive test. *Forensic Science International: Genetics*. 13:34-44.

Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing *reads*. *EMBnetjournal*. 17:pp. 10-12.

Matveeva E, Maiorano J, Zhang Q *et al.* (2016). Involvement of PARP1 in the regulation of alternative splicing. *Cell Discovery*. 2:15046.

McEvoy B, Beleza S, Shriver MD (2006). The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Human Molecular Genetics*. 15 Spec No:R176-81.

McKenna A, Hanna M, Banks E *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 20:1297-1303.

Mengel-From J, Børsting C, Sanchez JJ *et al.* (2010). Human eye colour and HERC2, OCA2 and MATP. *Forensic Science International: Genetics*. 4:323-328.

Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*. 16:1215.

Minelli C, Thompson JR, Abrams KR *et al.* (2005). The choice of a genetic model in the meta-analysis of molecular association studies. *International Journal of Epidemiology*. 34:1319-1328.

Muniz YC, Ferreira LB, Mendes-Junior CT *et al.* (2008). Genomic ancestry in urban Afro-Brazilians. *Annals of Human Biology*. 35:104-111.

Nan H, Kraft P, Hunter DJ *et al.* (2009). Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians. *International Journal of Cancer*. 125:909-917.

Oetting WS, Garrett SS, Brott M *et al.* (2005). P gene mutations associated with oculocutaneous albinism type II (OCA2). *Human Mutation*. 25:323.

Pabinger S, Dander A, Fischer M *et al.* (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in Bioinformatics*. 15:256-278.

Pan Q, Shai O, Lee LJ *et al.* (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*. 40:1413-1415.

Parra EJ (2007). Human pigmentation variation: evolution, genetic basis, and implications for public health. *American Journal of Physical Anthropology*. 45:85-105.

Parra FC, Amado RC, Lambertucci JR *et al.* (2003). Color and genomic ancestry in Brazilians. **PNAS.** 100:177-182.

Pena SD, Bastos-Rodrigues L, Pimenta JR *et al.* (2009). DNA tests probe the genomic ancestry of Brazilians. *Brazilian Journal of Medical and Biological Research*. 42:870-876.

Pena SD, Di Pietro G, Fuchshuber-Moraes M *et al.* (2011). The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One*. 6:e17063.

Phillips C, Salas A, Sánchez JJ *et al.* (2007). Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International Genetics*. 1:273-280.

Pietroni C, Andersen JD, Johansen P *et al.* (2014). The effect of gender on eye colour variation in European populations and an evaluation of the IrisPlex prediction model. *Forensic Science International Genetics.* 11:1-6.

Pimenta JR, Zuccherato LW, Debes AA *et al.* (2006). Color and genomic ancestry in Brazilians: a study with forensic microsatellites. *Human Heredity*. 62:190-195.

Pinto-Proença R, Santos M, Fonseca C *et al.* (2018). Conjunctival melanoma: association of cyclooxygenase-2 tumor expression to prognosis. *Graefe's Archive for Clinical and Experimental* **Ophthalmology**. 256:989-995.

Ponting CP, Hardison RC (2011). What fraction of the human genome is functional. *Genome Res.* 21:1769-1776.

Pośpiech E, Draus-Barini J, Kupiec T *et al.* (2011). Gene-gene interactions contribute to eye colour variation in humans. *Journal of Human Genetics*. 56:447-455.

Pośpiech E, Wojas-Pelc A, Walsh S *et al.* (2014). The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction. *Forensic Science International Genetics.* 11:64-72.

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

Prota G, Hu DN, Vincensi MR *et al.* (1998). Characterization of melanins in human irides and cultured uveal melanocytes from eyes of different colors. *Experimental Eye Research*. 67:293-299.

Psychiatric GWASCBDWG (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*. 43:977-983.

Purcell S, Neale B, Todd-Brown K *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. 81:559-575.

Purtscher E (1965). On the development and morphology of iris crypts. **Acta Ophthalmologica**. 43:109-119.

Queiroz EM, Santos AM, Castro IM *et al.* (2013). Genetic composition of a Brazilian population: the footprint of the Gold Cycle. *Genetics and Molecular Research*. 12:5124-5133.

Ramensky V, Bork P, Sunyaev S (2002). Human non-synonymous SNPs: server and survey. *Nucleic Acids Research*. 30:3894-3900.

Ranka S, Kahveci T, Singh M (2012)). A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine.* 12:414-417.

Rawofi L, Edwards M, Krithika S *et al.* (2017). Genome-wide association study of pigmentary traits (skin and iris color) in individuals of East Asian ancestry. *PeerJ*. 5:e3951.

Rebbeck TR, Kanetsky PA, Walker AH *et al.* (2002). P gene as an inherited biomarker of human eye color. *Cancer Epidemiology Biomarkers & Prevention*. 11:782-784.

Rees JL (2003). Genetics of hair and skin color. Annual Review of Genetics. 37:67-90.

Ribeiro D. O povo brasileiro: A formação e o sentido do Brasil. Companhia das Letras; 1995:467.

Ritchie MD, Hahn LW, Roodi N *et al.* (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *American Journal of Human Genetics*. 69:138-147.

Robinson JT, Thorvaldsdóttir H, Winckler W et al. (2011). Integrative genomics viewer. **Nature Biotechnology**. 29:24-26.

Rosenmann A, Bejarano-Achache I, Eli D *et al.* (2009). Prenatal molecular diagnosis of oculocutaneous albinism (OCA) in a large cohort of Israeli families. *Prenatal Diagnosis*. 29:939-946.

Ruiz Y, Phillips C, Gomez-Tato A *et al.* (2013). Further development of forensic eye color predictive tests. *Forensic Science International Genetics*. 7:28-40.

Salzano FM (1986). Em busca das raízes. Ciência Hoje. 5(25):48-53.

Salzano FM, Bortolini MC (2002). The Evolution and Genetics of Latin American Populations.

Salzano FM, Freire-Maia N (1970). Problems in Human Biology. A Study of Brazilian Populations.

Sankin A, Hakimi AA, Mikkilineni N *et al.* (2014). The impact of genetic heterogeneity on biomarker development in kidney cancer assessed by multiregional sampling. *Cancer Medicine*. 3:1485-1492.

Santos HC, Horimoto AVR, Tarazona-Santos E *et al.* (2015). A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: the Brazilian set. *European Journal of Human Genetics.* 1-7.

Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK *et al.* (2010). Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Human Mutation*. 31:184-190.

Santos RV, Fry PH, Monteiro S *et al.* (2009). Color, race, and genomic ancestry in Brazil: dialogues between anthropology and genetics. *Current Anthropology*. 50:787-819.

Seddon JM, Sahagian CR, Glynn RJ *et al.* (1990). Evaluation of an iris color classification system. The Eye Disorders Case-Control Study Group. *Investigative Ophthalmology & Visual Science*. 31:1592-1598.

Shekar SN, Duffy DL, Frudakis T *et al.* (2008). Linkage and association analysis of spectrophotometrically quantified hair color in Australian adolescents: the effect of OCA2 and HERC2. *Journal of Investigative Dermatology.* 128:2807-2814.

Shihab HA, Gough J, Cooper DN *et al.* (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human Mutation*. 34:57-65.

Simon JD, Peles D, Wakamatsu K *et al.* (2009). Current challenges in understanding melanogenesis: Bridging chemistry, biological control, morphology, and function. *Pigment Cell and Melanoma Research*. 22:563-579.

Spichenok O, Budimlija ZM, Mitchell AA *et al.* (2011). Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Science International: Genetics*. 5:472-478.

Spritz RA, Lee ST, Fukai K *et al.* (1997). Novel mutations of the P gene in type II oculocutaneous albinism (OCA2). *Human Mutation*. 10:175-177.

Sturm RA (2006). A golden age of human pigmentation genetics. Trends in Genetics. 22:464-468.

Sturm RA (2009). Molecular genetics of human pigmentation diversity. *Human Molecular Genetics*. 18:9-17.

Sturm RA, Duffy DL, Zhao ZZ *et al.* (2008). A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *American Journal of Human Genetics*. 82:424-431.

Sturm RA, Frudakis TN (2004). Eye colour: portals into pigmentation genes and ancestry. *Trends in Genetics*. 20:327-332.

Sturm RA, Larsson M (2009). Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res.* 22:544-562.

Sturm RA, Teasdale RD, Box NF (2001). Human pigmentation genes: identification, structure and consequences of polymorphic variation. *Gene*. 277:49-62.

Suarez-Kurtz G, Vargens DD, Struchiner CJ *et al.* (2007). Self-reported skin color, genomic ancestry and the distribution of GST polymorphisms. *Pharmacogenetic Genomics*. 17:765-771.

Sugumaran M (1991). Molecular mechanisms for mammalian melanogenesis. Comparison with insect cuticular sclerotization. *FEBS Letters*. 295:233-239.

Sulem P, Gudbjartsson DF, Stacey SN *et al.* (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genetics*. 39:1443-1452.

Sulem P, Gudbjartsson DF, Stacey SN *et al.* (2008). Two newly identified genetic determinants of pigmentation in Europeans. *Nature Genetics*. 40:835-837.

Tavares LHD (1988). Comércio proibido de escravos.

Teng M, Ichikawa S, Padgett LR *et al.* (2012). regSNPs: a strategy for prioritizing regulatory single nucleotide substitutions. *Bioinformatics*. 28:1879-1886.

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013). Integrative Genomics Viewer (IGV): Highperformance genomics data visualization and exploration. *Briefings in Bioinformatics*. 14:178-192.

Trynka G, Hunt KA, Bockett NA *et al.* (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genetics*. 43:1193-1201.

Ulivi S, Mezzavilla M, Gasparini P (2013). Genetics of eye colours in different rural populations on the Silk Road. *European Journal of Human Genetics*. 21:1320-1323.

Van der Auwera GA, Carneiro MO, Hartl C *et al.* (2013). From *fastq* data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*. 43:11.10.1-33.

Venâncio RP (2007). Presença portuguesa: de colonizadores a imigrantes. Brasil: 500 anos de povoamento. Rio de Janeiro: **IBGE**. p. 232.

Vernot B, Stergachis AB, Maurano MT *et al.* (2012). Personal and population genomics of human regulatory variation. *Genome Research*. 22:1689-1697.

Visser M, Kayser M, Grosveld F *et al.* (2014). Genetic variation in regulatory DNA elements: The case of OCA2 transcriptional regulation. *Pigment Cell and Melanoma Research*. 27:169-177.

Visser M, Kayser M, Palstra R-j (2012). HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Research*. 22:446-455.

Walsh S, Liu F, Wollstein A *et al.* (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International Genetics*. 7:98-115.

Walsh S, Wollstein A, Liu F *et al.* (2012). DNA-based eye colour prediction across Europe with the IrisPlex system. *Forensic Science International Genetics*. 6:330-340.

Wang ET, Sandberg R, Luo S *et al.* (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*. 456:470-476.

Wang Y, Ma M, Xiao X *et al.* (2012). Intronic splicing enhancers, cognate splicing factors and contextdependent regulation rules. *Nature Structural & Molecular Biology*. 19:1044-1052.

Ward LD, Kellis M (2016). *HaploReg* v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Research*. 44:D877-81.

Wielgus AR, Sarna T (2005). Melanin in human irides of different color and age of donors. *Pigment Cell Research*. 18:454-464.

Wilde S, Timpson A, Kirsanow K *et al.* (2014). Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. **PNAS**. 111:4832-4837.

Yamaguchi Y, Brenner M, Hearing VJ (2007). The regulation of skin pigmentation. *Journal of Biological Chemistry*. 282:27557-27561.

Yamaguchi Y, Hearing VJ (2014). Melanocytes and their diseases. **Cold Spring Harbor Perspectives** in **Medicine**. 4

Yang J, Ferreira T, Morris AP *et al.* (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*. 44:369-75, S1.

Yang X, Liu D, Liu F *et al.* (2013). HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics*. 14:33.

Zheng X, Levine D, Shen J *et al.* (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 28:3326-3328.

APÊNDICES

Apêndice A - Tabelas com as frequências alélicas dos marcadores informativos de ancestralidade entre as populações do *1000genomes* utilizados para o controle da estratificação populacional.

			ł				-	δ	
SNP	Cromossomo (pb)	Gene	Alelos	AFR	EUR	EAS	AFR-EUR	AFR-EAS	EUR-EAS
rs2814778	1 (159174683)	DARC	С	0,900	0,000	0,000	0,900	0,900	0,000
	()		Т	0,100	1,000	1,000	-	-	-
rs1573020	1 (36768200)	THRAP3	A	0,559	0,986	0,951	0,427	0,392	0,035
			G	0,441	0,014	0,049	- 0.078	- 0.154	- 0.420
rs2065160	1 (204790977)	-	G	0,318	0,790	0,504	0,270	-	0,432
			C	0.957	0.538	0,000	0 419	0.033	0.452
rs182549	2 (136616754)	MCM6	Ť	0.043	0.462	0.010	-	-	-
006700	0 (7140155)		С	0,615	0,835	0,374	0,220	0,241	0,461
rs896788	2 (7149155)	RINF144A	Т	0,385	0,165	0,626	-	-	-
ro1/08/1/	3 (168645035)		A	0,859	0,550	0,728	0,309	0,131	0,178
181490444	3 (1080+3033)	-	С	0,141	0,450	0,272	-	-	-
rs2026721	4 (159181963)	-	А	0,525	0,073	0,046	0,452	0,479	0,027
101010111	. (10)101900)		G	0,475	0,927	0,954	-	-	-
4540055	4 (20002055)		A	0,134	0,718	0,396	0,584	0,262	0,322
rs4540055	4 (38803255)	ILRI	C	0,513	0,054	0,102	-	-	-
			ſ	0,353	0,228	0,501	- 0.821	-	- 0.857
rs16891982	5 (33951693)	SLC45A2	C	0,030	0,877	0,020	0,821	0,030	0,837
			G	0,944	0,123	0,960	- 0.417	- 0.412	-
rs917118	7 (4457003)	-	A C	0,001	0,204	0,208	0,417	0,413	0,004
			C	0,910	0,750	0,732	- 0.149	- 0.322	- 0.173
rs7897550	10 (17064992)	CUBN	Ť	0.090	0.239	0.412	-	-	-
1070006	10 (0 45550 40)	DADDO	Ċ	0.414	0.007	0.000	0.407	0.414	0.007
rs1978806	10 (34755348)	PARD3	Т	0,586	0,993	1,000	-	-	-
			А	0,316	0,107	0,053	0,209	0,263	0,054
rs5030240	11 (32424389)	WT1	С	0,369	0,696	0,342	-	-	-
			G	0,316	0,197	0,604	-	-	-
rs10843344	12 (29369871)	FAR^2	C	0,966	0,707	0,905	0,259	0,061	0,198
	. (,	2 7 14 1	Т	0,034	0,293	0,095	-	-	-
rs773658	12 (56603834)	RNF41	C	0,435	0,993	0,969	0,558	0,534	0,024
			G	0,505	0,007	0,031	-	-	-

rs2065982	13 (34864240)	-	A G	0,904 0,096	0,951 0,049	0,342 0,658	0,047 -	0,562 -	0,609 -
rs1335873	13 (20901724)	-	T A	0,153 0,847	0,718 0,282	0,649 0,351	0,565 -	0,496 -	0,069 -
rs1886510	13 (22374700)	-	A G	0,236 0,764	0,520 0,480	0,144 0.856	0,284 -	0,092 -	0,376 -
rs10141763	14 (36170607)	RALGAPA1	A T	0,617 0.383	0,914 0.086	0,554 0,446	0,297 -	0,063 -	0,360 -
rs730570	14 (101142890)	-	C T	0,754 0,246	0,143 0,857	0,792 0,208	0,611 -	0,038 -	0,649 -
rs12913832	15 (28365618)	HERC2	A G	0,991 0,009	0,507 0,493	0,980 0,020	0,484 -	0,011	0,473 -
rs1426654	15 (48426484)	SLC24A5	C T	0,912 0,088	0,022 0,978	0,925 0,075	0,890 -	0,013 -	0,903 -
rs3785181	16 (90105333)	GAS8	C T	0,919 0,081	0,939 0,061	0,439 0,561	0,020 -	0,480 -	0,500 -
rs881929	16 (31079371)	ZNF668	G T	0,902 0,098	0,502 0,498	0,108 0,892	0,400 -	0,794 -	0,394 -
rs2304925	17 (75551667)	-	G T	0,525 0.475	0,224 0,776	0,491 0,509	0,301 -	0,034 -	0,267 -
rs1024116	18 (75432386)	-	A G	0,292	0,564	0,121	0,272	0,171	0,443
rs2303798	19 (42410331)	ARHGEF1	C T	0,433	0,991	0,784	0,558 -	0,351	0,207
rs1321333	20 (38849642)	-	C T	0,956	0,540 0,460	0,947 0.053	0,416 -	0,009 -	0,407 -
rs722098	21 (16685598)	-	A G	0,224 0,776	0,783 0.217	0,442 0.558	0,559 -	0,218	0,341 -
rs239031	21 (17710424)	-	C T	0,612 0,388	0,007 0,993	0,057 0,943	0,605 -	0,555 -	0,050 -
rs2572307	21 (25672460)	-	A	0,431	0,007	0,000	0,424	0,431	0,007
rs5997008	22 (26350103)	MYO18B	A	0,569	0,993	0,041	- 0,397	- 0,417	- 0,020
rs2040411	22 (47836412)	-	A G	0,542 0,782 0,218	0,939 0,655 0,345	0,959 0,260 0,740	- 0,127 -	- 0,522 -	- 0,395 -

pb = Pares de base; AFR = Frequência alélica nas populações africanas do *1000Genomes*; EUR = Frequência alélica nas populações europeias do *1000Genomes*; EAS = Frequência alélica nas populações do leste asiático do *1000Genomes*; δ = Diferencial de frequência alélica.

Gene	Cromossomo:Posição	rsID	REF	ALT
OCA2	15:28000484	rs7175266	Т	С
OCA2	15:28090173	rs12592307	С	Т
OCA2	15:28096538	rs1800419	G	А
OCA2	15:28096842	rs13313425	G	А
OCA2	15:28116275	rs7175046	С	G
OCA2	15:28116282	rs41304383	Т	C
OCA2	15:28116336	rs1800418	C	Т
OCA2	15:28116379	rs1800417	A	G
OCA2	15:28116987	rs112277343	т	A
OCA2	15:28171147	rs2594928	Ċ	т
OCA2	15:28171294	rs1800416	т	G
0012	15.28171/16	rs113151152	т Т	C
OCA2	15.28107027	ro1900/1/	т Т	C
OCA2	15.28197037	181000414	I C	
OCA2	15:28200408	187170909	C	1
OCA2	15:28211758	rs4778218	G	A
OCA2	15:28211921	rs1800411	A	G
OCA2	15:28228644	rs12910433	A	G
OCA2	15:28230184	rs1800410	Т	C
OCA2	15:28230318	rs1800407	С	Т
OCA2	15:28230378	rs117886461	G	А
OCA2	15:28231793	rs10852218	С	Т
OCA2	15:28235725	rs1800405	G	А
OCA2	15:28235773	rs1800404	Т	С
OCA2	15:28259863	rs4640131	А	G
OCA2	15:28260053	rs1800401	G	А
OCA2	15:28260128	rs17652037	С	А
OCA2	15:28267758	rs41534647	А	G
OCA2	15:28326942	rs61738394	С	Т
OCA2	15:28327170	rs56082644	С	Т
OCA2	15:28344238	rs7495174	А	G
OCA2	15:28344328	rs7497270	С	Т
OCA2	15:28344613	rs114561380	Т	С
OCA2	15:28344695	rs7496326	С	Т
OCA2	15:28344740	rs28595435	G	С
OCA2	15:28344765	rs62007492	G	А
OCA2	15:28345896	rs73377799	А	G
OCA2	15:28345931	rs58843292	G	A
OCA2	15:28346033	rs60727732	A	AC
OCA2	15:28346322	rs115488780	C	G
OCA2	15:28346511	rs4778139	A	C
-	15:28346655	rs73377800	Δ	т
_	15:28346705	ro73377802	C	G
_	15.283/681/	ro5811525	Δ	ΔT
-	15.00247004	180011000		TI T
-	10:2004/024	18113383970		1
-	15:28347419	rs111/0/210	G	A
HERC2	15:28355991	rs4778242	C	Т
HERC2	15:28356130	rs4778243	G	A
HERC2	15:28356349	rs4778244	А	G
HERC2	15:28356859	rs1129038	С	Т

Apêndice B - Tabela com os 170 polimorfismos analisados dos genes OCA2 e HERC2

HERC2	15:28357230	rs4778245	С	Т
HERC2	15:28357357	rs62007495	G	С
HERC2	15:28358492	rs75620448	С	Т
HERC2	15:28359744	rs7495441	А	G
HERC2	15:28360081	rs7497014	С	Т
HERC2	15:28361764	rs4778247	Т	С
HERC2	15:28362029	rs74007958	G	А
HERC2	15:28365431	rs6497271	G	A
HERC2	15:28365618	rs12913832	A	G
HERC2	15:28366438	rs78809959	C	Т
HERC2	15.28368995	rs11074322	Т	Ċ
HERC2	15:28370102	rs75159657	Ċ	Т
HERC2	15.28370389	rs61585051	C	т
HERC2	15.28377772	rs8025035	т	Ċ
HERC2	15.28378122	rs4778141	G	с т
HERC2	15:28380518	rs4778249	A	т
HERC2	15:28386626	ro11636030	C	т Т
HERC2	15.28380020	ro112004881	C	т Т
HERC2	15.20300194	rs56012620	C	1
HERC2	15.20300274	1830013020	с т	л С
HERC2	15.20309300	187403303	I C	
HERC2	15:26369744	rs57039470	G	A
HERC2	15:26369660	rs1133496	G	A
HERC2	15:28397812	rs/105158	A	G
HERC2	15:28408573	rs10438451	A	C
HERC2	15:28412872	rs61750152	G	A
HERC2	15:28413446	rs114694773	G	A
HERC2	15:28413640	rs61754657	C	T
HERC2	15:28414492	rs55907941	Т	A
HERC2	15:28414665	rs7494786	С	Т
HERC2	15:28419290	rs72714141	С	Т
HERC2	15:28419330	rs112483884	Т	C
HERC2	15:28419375	rs77093318	AT	A
HERC2	15:28419393	rs58358300	А	С
HERC2	15:28419526	rs372420106	А	G
HERC2	15:28419566	rs201182375	Т	C
HERC2	15:28419587	rs142399136	G	А
HERC2	15:28419619	rs141441362	С	Т
HERC2	15:28419644	rs77865049	С	Т
HERC2	15:28419650	rs78696453	А	G
HERC2	15:28419695	rs149592795	Т	С
HERC2	15:28419713	rs113016718	С	Т
HERC2	15:28419746	rs139335991	G	А
HERC2	15:28419974	rs57079108	С	Т
HERC2	15:28420614	rs73379676	С	Т
HERC2	15:28422026	rs6497284	Т	С
HERC2	15:28422471	rs73362603	С	G
HERC2	15:28422609	rs61756153	С	Т
HERC2	15:28422627	rs61756154	G	А
HERC2	15:28422684	rs73362604	С	А
HERC2	15:28422819	rs58164482	С	Т
HERC2	15:28424100	rs9806328	С	Т
HERC2	15:28427801	rs73362608	А	Т
HERC2	15:28427916	rs12591531	G	А

HERC2	15:28427970	rs73362609	G	А
HERC2	15:28431980	rs78970358	С	Т
HERC2	15:28436596	rs7166814	Т	С
HERC2	15:28437061	rs72714143	А	G
HERC2	15:28437441	rs11856675	Т	С
HERC2	15:28441111	rs74005643	Т	С
HERC2	15:28441817	rs2525936	Т	G
HERC2	15:28443658	rs8041209	G	Т
HERC2	15:28443682	rs8024600	А	G
HERC2	15:28446638	rs76296532	А	G
HERC2	15:28447234	snp_135	G	А
HERC2	15:28447282	rs74005645	С	Т
HERC2	15:28447419	rs200598834	G	С
HERC2	15:28451737	rs112222559	С	Т
HERC2	15:28451756	rs113921795	G	Α
HERC2	15:28451762	rs150246661	А	G
HERC2	15:28451772	rs138923800	Т	G
HERC2	15:28451776	rs143333992	А	G
HERC2	15:28456028	snp_151	А	G
HERC2	15:28456306	rs201431906	А	Т
HERC2	15:28456382	snp_157	С	Т
HERC2	15:28457540	rs149963581	G	С
HERC2	15:28457555	rs143863603	С	Т
HERC2	15:28459219	rs151156359	С	Т
HERC2	15:28461161	rs8030794	Т	G
HERC2	15:28464038	rs112306125	Т	С
HERC2	15:28467246	rs376640270	Т	С
HERC2	15:28467383	rs139377364	Т	С
HERC2	15:28467512	snp_176	С	А
HERC2	15:28467557	rs192627690	А	С
HERC2	15:28474686	rs118112076	G	С
HERC2	15:28475370	rs2428650	С	Т
HERC2	15:28475625	rs201872292	G	А
HERC2	15:28475864	rs142477460	С	G
HERC2	15:28478323	rs61757160	Т	С
HERC2	15:28478780	rs62008730	С	G
HERC2	15:28478799	rs72714147	G	A
HERC2	15:28483543	rs1962780	С	Т
HERC2	15:28483617	rs60422914	Т	C
HERC2	15:28483619	rs57727285	G	A
HERC2	15:28492190	rs76160479	C	Т
HERC2	15.28493688	rs57189560	A	Т
HERC2	15.28493737	rs78460234	A	G
HERC2	15.28494032	rs2240204	G	A
HERC2	15.28499434	rs117902599	C	A
HERC2	15:28499436	rs59556486	A	C
HERC2	15:28499459	rs8036480	T	C
HERC2	15:28501195	rs111681126	C I	A
HERC?	15.28502065	rs117504297	G G	Δ
HERC?	15.28502000	rs11631707	G	Δ
HERCO	15.28502479	rs113070000	C C	л т
HEPCO	15.20502+02	rs160500/1	G	1
HEPCO	15.20302744	1810930941 rs76222040	4	A C
heruz	15:20500240	18/0333042	A	U

HERC2	15:28507825	rs10162958	А	G	
HERC2	15:28507925	rs78374687	С	Т	
HERC2	15:28508467	rs74877971	А	G	
HERC2	15:28510895	rs2240202	G	А	
HERC2	15:28513513	rs916976	А	G	
HERC2	15:28516084	rs8039195	Т	С	
HERC2	15:28518229	rs71467328	С	Т	
HERC2	15:28544707	rs112371820	С	Т	
HERC2	15:28567799	snp_261	С	Т	
HERC2	15:28567840	snp_262	А	G	
HERC2	15:28567859	snp_264	G	А	
HERC2	15:28567867	snp_266	G	А	
HERC2	15:28567895	snp_268	G	А	
HERC2	15:28568000	snp_276	Т	А	
HERC2	15:28568057	snp_282	С	Т	

rsID = Identificação do polimorfismo; REF = Alelo referência; ALT = Alelo alternativo; - = polimorfismos em região intergênica.

Apêndice C -Tabelas com as comparações par-a-par de desequilíbrio de ligação entre os polimorfismos associados nas análises de regressão logística e linear

Associações	rsID	DL	Haplótipo	Frequência
Olho verde <i>vs.</i> castanho-escuro	rs12913832 rs61756152	<i>R</i> ² = 0,08 D' = 1,00	GA AA GG AG	0,0000 0,1099 0,4172 0,4729
Olho verde <i>vs.</i> castanho-escuro	rs12913832 rs1800407	$R^2 = 0.01$ D' = 0.59	GT AT GC AC	0,0087 0,0425 0,4085 0,5403
Olho verde <i>vs.</i> castanho-escuro	rs1800407 rs61756152	<i>R</i> ² = 0,20 D' = 0,68	TA CA TG CG	0,0365 0,0733 0,0148 0,8754
Olho mel <i>vs.</i> castanho-escuro	rs1129038 rs1800407	$R^2 = 0,00$ D' = 0,07	TT CT TC CC	0,0092 0,0420 0,4110 0,5378
Olho mel <i>vs.</i> castanho-escuro	rs1129038 rs1800419	$R^2 = 0,00$ D' = 0,07	TA CA TG CG	0,2220 0,2689 0,1982 0,3109
Olho mel <i>vs.</i> castanho-escuro	rs1800407 rs1800419	$R^2 = 0.01$ D' = 0.54	TA CA TG CG	0,0394 0,4516 0,0118 0,4972
Olho castanho-claro <i>vs.</i> castanho-escuro; Olho castanho-escuro <i>vs.</i> não castanho- escuro; Valores de b* (CIELAB)	rs12913832 rs201872292	<i>R</i> ² = 0,07 D' = 1,00	GA AA GG AG	0,0000 0,0964 0,4172 0,4864
Olho castanho-claro <i>vs.</i> castanho-escuro Olho castanho-escuro <i>vs.</i> não castanho- escuro	rs12913832 rs1800401	$R^2 = 0.01$ D' = 0.47	GA AA GG AG	0,0180 0,0648 0,3991 0,5180
Olho castanho-claro <i>vs.</i> castanho-escuro Olho castanho-escuro <i>vs.</i> não castanho- escuro	rs201872292 rs1800401	$R^2 = 0,00$ D' = 0,06	AA GA AG GG	0,0128 0,0700 0,0835 0,8336
Olho castanho-escuro <i>vs.</i> não castanho- escuro	rs12913832 rs74005645	<i>R</i> ² = 0,09 D' = 1,00	GT AT GC AC	0,0000 0,1205 0,4172 0,4623
Olho castanho-escuro <i>vs.</i> não castanho- escuro; Cabelo loiro <i>vs.</i> castanho	rs12913832 rs1800419	<i>R</i> ² = 0,00 D' = 0,06	GA AA GG AG	0,2179 0,2731 0,1993 0,3097
Olho castanho-escuro <i>vs.</i> não castanho- escuro	rs201872292 rs74005645	<i>R</i> ² = 0,01 D' = 1,00	AT GT AC GC	0,0000 0,1205 0,0964 0,7831
Olho castanho-escuro <i>vs.</i> não castanho- escuro	rs201872292 rs1800419	<i>R</i> ² = 0,00 D' = 0,20	AA GA AG GG	0,0572 0,4337 0,0392 0,4699
Olho castanho-escuro <i>vs.</i> não castanho- escuro	rs1800401 rs74005645	<i>R</i> ² = 0,04 D' = 0,25	AT GT AC GC	0,0287 0,0918 0,0541 0.8254
Olho castanho-escuro <i>vs.</i> não castanho- escuro	rs1800401 rs1800419	<i>R</i> ² = 0,01 D' = 0,37	AA GA AG GG	0,0253 0,4657 0,0576 0,4514

			TA	0,0454
Olho castanho-escuro vs. não castanho-	rs74005645	$R^2 = 0,00$	CA	0,4455
escuro	rs1800419	D' = 0,20	TG	0,0751
			CG	0,4340
			CG	0,0196
Pele clara <i>v</i> s. não clara;	rs41304383	$R^2 = 0,00$	TG	0,3975
Pele clara <i>v</i> s. intermediária	rs12913832	D' = 0,00	CA	0,0271
			TA	0,5558
			CC	0,0163
Pele clara <i>vs.</i> não clara;	rs41304383	$R^2 = 0,00$	TC	0,3888
Pele clara <i>v</i> s. intermediária	rs1800404	D' = 0,13	СТ	0,0304
			TT	0,5645
			GC	0,0819
Pele clara <i>v</i> s. não clara;	rs12913832	$R^2 = 0,12$	AC	0,3232
Pele clara vs. intermediária	rs1800404	D' = 0,51	GT	0,3353
			AT	0,2596
			CG	0.0000
Índice de melanina em região não exposta e	rs1800414	$R^2 = 0.01$	TG	0.4172
exposta ao sol	rs12913832	D' = 1.00	CA	0.0196
-		,	ТА	0,5632
			CA	0,0000
Índice de melanina em região não exposta	rs1800414	$R^2 = 0.00$	ТА	0,0846
ao sol	rs13313425	D' = 1,00	CG	0,0196
			TG	0,8958
			GA	0,0027
Índice de melanina em região não exposta	rs12913832	$R^2 = 0.05$	AA	0,0819
ao sol	rs13313425	D' = 0.92	GG	0,4157
			AG	0,4997
			GT	0,0000
Índias de melonine en norião erresto es col	rs12913832	$R^2 = 0.00$	AT	0,0105
indice de melanina em região exposta ao sol	rs61738394	D' = 1,00	GC	0,4172
			AC	0,5723
			СТ	0,0000
Í. 1	rs1800414	$R^2 = 0.12$	TT	0,0105
Indice de melanina em região exposta ao sol	rs61738394	D' = 0,51	CC	0,0196
			TC	0,9699
			AA	0,0572
Olho castanho-escuro <i>v</i> s. não castanho-	rs201872292	$R^2 = 0.00$	GA	0,4337
escuro	rs1800419	D' = 0,20	AG	0,0392
			GG	0,4699
			AT	0,0287
Olho castanho-escuro vs. não castanho-	rs1800401	$R^2 = 0.04$	GT	0,0918
escuro	rs74005645	D' = 0,25	AC	0,0541
			GC	0,8254
			AA	0,0253
Olho castanho-escuro vs. não castanho-	rs1800401	$R^2 = 0.01$	GA	0,4657
escuro	rs1800419	D' = 0,37	AG	0,0576
			GG	0,4514
			ТА	0,0454
Olho castanho-escuro vs. não castanho-	rs74005645	$R^2 = 0,00$	CA	0,4455
escuro	rs1800419	D' = 0,20	TG	0,0751
			CG	0,4340
			CG	0,0196
Pele clara <i>v</i> s. não clara;	rs41304383	$R^2 = 0,00$	TG	0,3975
Pele clara <i>v</i> s. intermediária	rs12913832	D' = 0,00	CA	0,0271
			TA	0,5558
			CC	0,0163
Pele clara vs. não clara;	rs41304383	$R^2 = 0,00$	TC	0,3888
Pele clara vs. intermediária	rs1800404	D' = 0,13	CT	0,0304
			TT	0,5645
Pele clara <i>u</i> s, não clara:	rs12913832	$R^2 = 0.10$	GC	0,0819
Pele clara <i>us</i> , intermediária	rs1800404	D' = 0.51	AC	0,3232
i die data vo. intermediaria	101000101	D 0,01	GT	0,3353

			AT	0,2596
			CG	0,0000
Îndice de melanina em região não exposta e	rs1800414	$R^2 = 0,01$	TG	0,4172
exposta ao sol	rs12913832	D' = 1,00	CA	0,0196
			TA	0,5632
		\mathbf{D}^{2}	CA	0,0000
Indice de melanina em região não exposta	rs1800414	$R^2 = 0,00$		0,0846
a0 501	1813313423	D = 1,00	CG TG	0,0190
			GA	0,0938
Índice de melanina em região não exposta	rs12913832	$P^2 = 0.05$	AA	0.0819
ao sol	rs13313425	D' = 0.92	GG	0 4157
			AG	0.4997
			GT	0,0000
Índica da malanina em ragião errocata os sol	rs12913832	$R^2 = 0,00$	AT	0,0105
mulce de melannia em região exposta ao sor	rs61738394	D' = 1,00	GC	0,4172
			AC	0,5723
			CT	0,0000
Índice de melanina em região exposta ao sol	rs1800414	$R^2 = 0,12$	TT	0,0105
	rs61738394	D' = 0,51	CC	0,0196
				0,9699
	001070000	Σ^2 0.00	AA	0,0572
Olno castanno-escuro vs. nao castanno-	rs201872292	$R^2 = 0.00$	GA	0,4337
esculo	181000419	D = 0,20	AG	0,0392
				0,4099
Olho castanho-escuro <i>u</i> s, não castanho-	rs1800401	$P^2 = 0.04$	GT	0,0287
escuro	rs74005645	D' = 0.25	AC	0.0541
		2 0,20	GC	0.8254
			AA	0,0253
Olho castanho-escuro <i>v</i> s. não castanho-	rs1800401	$R^2 = 0.01$	GA	0,4657
escuro	rs1800419	D' = 0,37	AG	0,0576
			GG	0,4514
			TA	0,0454
Olho castanho-escuro <i>vs.</i> não castanho-	rs74005645	$R^2 = 0,00$	CA	0,4455
escuro	rs1800419	D' = 0,20	TG	0,0751
			CG	0,4340
Dele clara un pão clara;	ro41204282	$D^2 = 0.00$	CG TG	0,0196
Pele clara vs intermediária	rs12913832	R' = 0.00 D' = 0.00	CA	0,0271
	1812910002	D 0,00	ТА	0,5558
			CC	0.0163
Pele clara vs. não clara;	rs41304383	$R^2 = 0.00$	TC	0,3888
Pele clara vs. intermediária	rs1800404	D' = 0,13	СТ	0,0304
			TT	0,5645
			GC	0,0819
Pele clara <i>vs.</i> não clara;	rs12913832	$R^2 = 0,12$	AC	0,3232
Pele clara vs. intermediária	rs1800404	D' = 0,51	GT	0,3353
			AT	0,2596
			CG	0,0000
Indice de melanina em região não exposta e	rs1800414	$R^2 = 0,01$	TG	0,4172
exposta ao sol	rs12913832	D' = 1,00	CA	0,0196
				0,5632
Índice de melanina em região não ornesto	rs 1800/11	$P^2 = 0.00$		0,0000
ao sol	rs13313425	R' = 0,00 D' = 1.00	CG	0,0840
40.001		2 1,00	ТG	0.8958
			GA	0,0027
Índice de melanina em região não exposta	rs12913832	$R^2 = 0.05$	AA	0,0819
ao sol	rs13313425	D' = 0,92	GG	0,4157
			AG	0,4997
Indice de melanina em região exposta ao sol	rs12913832		GT	0,0000

rs61738394	$D^2 = 0.00$	AT	0,0105
	$R^{-} = 0,00$	GC	0,4172
	D - 1,00	AC	0,5723
		CT	0,0000
rs1800414	$R^2 = 0.12$	TT	0,0105
rs61738394	D' = 0,51	CC	0,0196
		TC	0,9699
	rs61738394 rs1800414 rs61738394	rs61738394 $R^2 = 0,00$ D' = 1,00rs1800414 $R^2 = 0,12$ rs61738394rs61738394D' = 0,51	$\begin{array}{c} rs61738394 \\ rs61738394 \\ rs61738394 \\ rs1800414 \\ rs61738394 \\ D' = 0,51 \\ CC \\ TC \\ \end{array}$

DL = Valores de R^2 e D' (medidas desequilíbrio de ligação)

Apêndice D-Tabelas com as frequências alélicas, genotípicas, heterozigosidade esperada e observada e aderências ao Equilíbrio de Hardy-Weinberg.

Frequências alélicas e genotípicas, heterozigose observada (Ho) e esperada (He), e aderência ao Equilíbrio de Hardy-Weinberg (pEHW) dos nove pontos de variação associados a pelo menos um fenótipo de pigmentação para indivíduos agrupados para cor dos olhos. Valores de p significativos estão destacados em negrito.

Ponto de variaç	ão	Olho azul (2n=72)	Olho verde (2n=120)	Olho mel (2n=38)	Olho castanho- claro (2n=142)	Olho castanho- escuro (2n=292)	Total (2n=664)
	A/A A/G	0,2222	0,3167	0,4211	0,2113	0,1781	0,2289 0.5241
	G/G	0.1944	0.3000	0.1053	0.1831	0.2877	0.2470
rs1800419	A	0,5104	0,5115	0,5444	0,4677	0,4588	0,4910
(28096538)	G	0,4896	0,4885	0,4556	0,5323	0,5412	0,5090
(,	Но	0,5833	0,3833	0,4737	0,6056	0,5342	0,5241
	He	0,4996	0,4999	0,4501	0,4996	0,4940	0,4998
	pEHW	0,5037	0,0740	1,0000	0,0994	0,4024	0,4419
	A/A A/G	0,0000 0,1111	0,0000 0,0167	0,0000 0,0000	0,0000 0,1268	0,0345 0,2207	0,0151 0,1390
	G/G	0,8889	0,9833	1,0000	0,8732	0,7448	0,8459
rs13313425 (28096842)	Α	0,0260	0,0382	0,0500	0,1181	0,1271	0,0846
	G	0,9740	0,9618	0,9500	0,8819	0,8729	0,9154
	Но	0,1111	0,0167	0,0000	0,1268	0,2207	0,1390
	He	0,1049	0,0165	0,0000	0,1187	0,2477	0,1549
	pEHW	1,0000	1,0000	1,0000	1,0000	0,1821	0,0696
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	T/C	0,0278	0,1167	0,3158	0,1268	0,0753	0,1024
	C/C	0,9722	0,8833	0,6842	0,8732	0,9247	0,8976
rs1800407	Т	0,0417	0,0611	0,0833	0,0461	0,0330	0,0512
(28230318)	С	0,9583	0,9389	0,9167	0,9539	0,9670	0,9488
	Но	0,0278	0,1167	0,3158	0,1268	0,0753	0,1024
	He	0,0274	0,1099	0,2659	0,1187	0,0725	0,0972
	pEHW	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	A/A A/G	0,0000 0,0556	0,0167 0,0500	0,0000 0,1053	0,0000 0,1127	0,0205 0,2192	0,0120 0,1416
rs1800401	G/G	0,9444	0,9333	0,8947	0,8873	0,7603	0,8464
(28260053)	Α	0,0365	0,0496	0,0556	0,1060	0,1099	0,0828
	G	0,9635	0,9504	0,9444	0,8940	0,8901	0,9172

	Но	0,0556	0,0500	0,1053	0,1127	0,2192	0,1416
	He	0,0540	0,0799	0,0997	0,1063	0,2264	0,1519
	pEHW	1,0000	0,0830	1,0000	1,0000	0,7114	0,2614
	T/T	0,8056	0,7167	0,1579	0,0423	0,0137	0,2410
rs1129038 (28356859)	T/C	0,1389	0,2167	0,7368	0,6620	0,2740	0,3584
	C/C	0,0556	0,0667	0,1053	0,2958	0,7123	0,4006
	Т	0,8438	0,5802	0,4056	0,2235	0,2940	0,4202
	С	0,1562	0,4198	0,5944	0,7765	0,7060	0,5798
	Но	0,1389	0,2167	0,7368	0,6620	0,2740	0,3584
	He	0,2188	0,2888	0,4986	0,4679	0,2560	0,4873
	<u> </u>	0,0696	0,0644	0,0749	0,0007	0,5311	0,0000
	G/G G/A	0,8333	0,7167	0,1579	0,0423	0,0137	0,2440
		0,1309	0,2107	0,0042	0,0479	0,2005	0,4096
10010000	n, n C	0,0270	0,5762	0,1075	0,0099	0,7200	0,1050
rs12913832	G A	0,0342	0,3703	0,3944	0,2100	0,2940	0,5828
(20303010)	Ho Ho	0,1430	0,4237	0,0030	0,7834	0,7000	0,3828
	HU	0,1389	0,2107	0,0042	0,0479	0,2003	0,3404
	Не	0,1755	0,2888	0,5000	0,4642	0,2463	0,4863
	<u>pEHW</u>	0,2747	0,0644	0,1876	0,0009	0,7386	0,0000
	A/A	0,0000	0,0167	0,0000	0,0423	0,0411	0,0301
	A/G	0,0278	0,1000	0,2632	0,3099	0,1301	0,1596
	G/G	0,9722	0,8833	0,7368	0,6479	0,8288	0,8102
rs61756152	Α	0,0469	0,1374	0,1833	0,1359	0,0879	0,1099
(28412872)	G	0,9531	0,8626	0,8167	0,8641	0,9121	0,8901
	Но	0,0278	0,1000	0,2632	0,3099	0,1301	0,1596
	He	0,0274	0,1244	0,2285	0,3166	0,1898	0,1957
	pEHW	1,0000	0,2205	1,0000	1,0000	0,0015	0,0525
	T/T	0,0000	0,0000	0,0000	0,0000	0,1027	0,0452
	T/C	0,0000	0,0167	0,0000	0,0845	0,2945	0,1506
	C/C	1,0000	0,9833	1,0000	0,9155	0,6027	0,8042
rs74005645	Т	0,0000	0,0267	0,0000	0,1820	0,2005	0,1205
(28447282)	С	1,0000	0,9733	1,0000	0,8180	0,7995	0,8795
	Но	-	0,0167	-	0,0845	0,2945	0,1506
	He	-	0,0165	-	0,0809	0,3750	0,2119
	pEHW	-	1,0000	-	1,0000	0,0137	0,0000
	A/A	0,0000	0,0167	0,0000	0,0282	0,0274	0,0211
rs201872292	A/G	0,0278	0,0667	0,2632	0,3099	0,1233	0,1506
(28475625)	G/G	0,9722	0,9167	0,7368	0,6620	0,8493	0,8283
	Α	0,0365	0,1221	0,1722	0,1198	0,0742	0,0964

G Ho	0,9635 0,0278	0,8779 0,0667	0,8278 0,2632	0,8802 0,3099	0,9258 0,1233	0,9036 0,1506
He	0,0274	0,0950	0,2285	0,2991	0,1622	0,1742
pEHW	1,0000	0,1228	1,0000	1,0000	0,0552	0,0515

Frequências alélicas e genotípicas, heterozigose observada (Ho) e esperada (He), e aderência ao Equilíbrio de Hardy-Weinberg (pEHW) dos quatro pontos de variação associados a pelo menos um fenótipo de pigmentação para indivíduos agrupados para cor dos cabelos. Valores de p significativos estão destacados em negrito.

Ponto de variação		Ruivo (2 <i>n</i> =20)	Loiro (2 <i>n</i> =96)	Castanho (2 <i>n</i> =362)	Preto (2 <i>n</i> =170)	Total (2 <i>n</i> =664)
	A/A	0,2000	0,2292	0,2210	0,2353	0,2289
	A/G	0,4000	0,4167	0,5912	0,4588	0,5241
	G/G	0,4000	0,3542	0,1878	0,3059	0,2470
rs1800419	Α	0,4310	0,5000	0,5000	0,5000	0,4910
(28090538)	G	0,5690	0,5000	0,5000	0,5000	0,5090
	Но	0,4000	0,4167	0,5912	0,4588	0,5241
	He	0,4800	0,4922	0,4995	0,4975	0,4998
	pEHW	0,5732	0,3769	0,0174	0,5142	0,4419
	G/G	0,3000	0,7292	0,1989	0,0353	0,2440
	G/A	0,5000	0,2292	0,4088	0,2824	0,3464
	A/A	0,2000	0,0417	0,3923	0,6824	0,4096
rs12913832	G	0,7931	0,4956	0,3308	0,3308	0,4172
(28305018)	Α	0,2069	0,5044	0,6692	0,6692	0,5828
	Но	0,5000	0,2292	0,4088	0,2824	0,3464
	He	0,4950	0,2637	0,4813	0,2907	0,4863
	pEHW	1,0000	0,3033	0,0550	0,7154	0,0000
	T/T	0,0000	0,1250	0,0552	0,0118	0,0542
	T/C	0,2000	0,4792	0,2707	0,1765	0,2741
	C/C	0,8000	0,3958	0,6740	0,8118	0,6717
rs11636232	Т	0,3190	0,2271	0,1617	0,1617	0,1913
28386626	С	0,6810	0,7729	0,8383	0,8383	0,8087
	Но	0,2000	0,4792	0,2707	0,1765	0,2741
	He	0,1800	0,4633	0,3086	0,1800	0,3094
	pEHW	1,0000	1,0000	0,0961	0,5884	0,0590
rs77093318	A/A	0,2000	0,0000	0,0222	0,0241	0,0243
(28419375)	A/AT	0,2000	0,0417	0,1556	0,1566	0,1398
	AT/AT	0,6000	0,9583	0,8222	0,8193	0,8359

Α	0,0690	0,0833	0,1008	0,1008	0,0942	
AT	0,9310	0,9167	0,8992	0,8992	0,9058	
Но	0,2000	0,0417	0,1556	0,1566	0,1398	
Не	0,4200	0,0408	0,1800	0,1838	0,1707	
pEHW	0,1331	1,0000	0,0805	0,1902	0,0638	

Frequências alélicas e genotípicas, heterozigose observada (Ho) e esperada (He), e aderência ao Equilíbrio de Hardy-Weinberg (pEHW) dos quatro pontos de variação associados a pelo menos um fenótipo de pigmentação para indivíduos agrupados para cor da pele e presença de sardas. Valores de p significativos estão destacados em negrito.

Ponto de variação		Clara (2 <i>n</i> =278)	Intermediária (2 <i>n</i> =274)	Escura (2 <i>n</i> =112)	Presença de Sardas (2 <i>n</i> =86)	Total (2 <i>n</i> =664)
	C/C	0,0000	0,0000	0,0000	0,0000	0,0000
	C/T	0,1367	0,0584	0,1179	0,0698	0,0934
	T/T	0,8633	0,9416	0,8821	0,9302	0,9066
rs41304383	С	0,0684	0,0292	0,0357	0,0349	0,0467
(28110282)	Т	0,9316	0,9708	0,9643	0,9651	0,9533
	Но	0,1367	0,0584	0,1179	0,0698	0,0934
	He	0,1273	0,0567	0,1110	0,0673	0,0890
	pEHW	0,7414	0,5499	1,0000	1,0000	1,0000
	C/C	0,0647	0,2117	0,5179	0,1163	0,2018
	C/T	0,3453	0,4672	0,4107	0,3953	0,4066
1000404	T/T	0,5899	0,3212	0,0714	0,4884	0,3916
rs1800404	С	0,2374	0,4453	0,7232	0,3140	0,4051
(28235773)	Т	0,7626	0,5547	0,2768	0,6860	0,5949
	Но	0,3453	0,4672	0,4107	0,3953	0,4066
	He	0,3621	0,4940	0,4004	0,4308	0,4820
	pEHW	0,5627	0,5487	1,0000	0,7220	0,0044
	G/G	0,4173	0,1460	0,0536	0,2326	0,2440
	G/A	0,3813	0,3650	0,2143	0,4186	0,3464
rs12913832	A/A	0,2014	0,4891	0,7321	0,3488	0,4096
(28365618)	G	0,6079	0,3285	0,1607	0,4419	0,4172
(20000010)	Α	0,3921	0,6715	0,8393	0,5581	0,5828
	Но	0,3813	0,3650	0,2143	0,4186	0,3464
	He	0,4767	0,4412	0,2698	0,4932	0,4863
	pEHW	0,0165	0,0424	0,1301	0,3574	0,0000
rs118112076	C/C	0,4173	0,1460	0,0536	0,0000	0,0000

Apêndices 137

	C/G G/G	0,3813 0,2014	0,3650 0,4891	0,2143 0,7321	0,0930 0,9070	0,0361 0,9639
(09474696)	C	0,4692	0,2798	0,4795	0,0181	0,0181
(28474080)	G	0,5308	0,7202	0,5205	0,9819	0,9819
	Но	0,3813	0,3650	0,2143	0,0930	0,0361
	He	0,4767	0,4412	0,2698	0,0887	0,0355
	pEHW	0,0206	0,0518	0,1301	1,0000	1,0000

Frequências alélicas e genotípicas, heterozigose observada (Ho) e esperada (He), e aderência ao Equilíbrio de Hardy-Weinberg (pEHW) dos dois pontos de variação associados a estruturas secundárias presentes na íris. Valores de p significativos estão destacados em negrito.

Ponto de variaçã	io	Pontos de pigmentação na íris (2 <i>n</i> =114)	Nódulos de Wolfflin (2 <i>n</i> =74)	Pigmentação na esclera (2 <i>n</i> =78)	Sulcos de contração (2 <i>n</i> =122)	Total (2 <i>n</i> =450)
	G/G	0,5263	0,7297	0,0513	0,2131	0,3289
	G/A	0,4035	0,2162	0,2564	0,4098	0,3689
	A/A	0,0702	0,0541	0,6923	0,3770	0,3022
rs12913832	G	0,7281	0,8378	0,1795	0,5820	0,5024
(28365618)	Α	0,2719	0,1622	0,8205	0,4180	0,4976
	Но	0,4035	0,2162	0,2564	0,4098	0,3689
	He	0,3960	0,2717	0,2945	0,4866	0,4996
	pEHW	1,0000	0,2154	0,5775	0,2903	0,0001
	C/C	0,0000	0,0000	0,2051	0,0328	0,0356
	C/A	0,0351	0,0000	0,3846	0,0656	0,0978
	A/A	0,9649	1,0000	0,4103	0,9016	0,8667
rs58358300	С	0,01754	0,0000	0,3974	0,0656	0,0844
(28419393)	Α	0,98246	1,0000	0,6026	0,9344	0,9156
	Но	0,0351	-	0,4790	0,0656	0,0978
	He	0,0345	-	0,3846	0,1225	0,1546
	pEHW	1,0000	1,0000	0,3124	0,0141	0,0000



Apêndice E -Gráficos bi-dimensionais entre os valores de L*a*b* para cor dos olhos.

Dispersão das íris no plano b* e L* do espectro de cor L*a*b*



Dispersão das íris no plano b* e a* do espectro de cor L*a*b*



Dispersão das íris no plano a
* e L* do espectro de cor L*a*b*





Diferenciação das medianas do valor de L* e a* para cada um dos grupos heteroclassificados como cor de olho azul, verde, mel, castanho claro e castanho escuro. Valor de p global do teste de Kruskal-Wallis no topo do gráfico. Linha pontilhada demarca a mediana de L* e a* em seus respectivos gráficos. ns = não significativo, p > 0.05; ** = p < 0.01; *** = p < 0.001; *** = p < 0.0001.

Apêndice G-Imagem do UCSC Genome Browser mostrando a conservação do aminoácido alterado pelo polimorfismo rs6178394C>T em diferentes espécies.



20 bases Scale - hai9 chr15: Simple Nucleotide Polymorphisms (dpSNP 150) Found in >= 17 of Samples ns74005645 RefSea gene predictions from NCBI RefSeq Curated Placenta' Mammal Basewise by Mammal Cons -4 Multiz Alignments of 46 Vertebrates Gaps D D V Human G b e b ъĭ D E ×. ν. Е 1.1 Chimp . 1. M Gorilla Orangutan Rhesus Baboon Marmoset Tansien Mouse_lemur Bushbaby Tree_shrew - - - - -_ _ _ _ Mouse C . . Rat C . M Kangaroo rat Guinea_pig Squirre 1 . Rabbit . . Pika Ċ. Alpaca - C G . . . A C à : Do lph in · V. 8 . Cow Horse - C G . M Cat . . a second s Dog Ŵ. -A . - ċ ò Microbat - **A** . . L Megabat ÷Ū. Ď Hedgehog : = = : D c c . . 8 8 . Shreŵ Elephant f Rock hynax N Tenrec fl D. ÷ – Armadillo A T Sloth A Mallabu = Opossum 10.1 E . . . T. . D Platupus = Chicken A . Zebra_finch A T ie. E I Lizand . т с + + + т 📫 X_tropicalis T Tetraodon = = = = = = = = = G M Fugu = = = Stickleback - -_ _ _ _ _ _ _ _ _ _ _ _ _ _ G L E - . A G P L . . . Medaka Ď Zebrafish = E F L E Lamprey , G G - ý L 4.49 GERP Scores for Mammalian Alignments GERP -8.97

Apêndice H -Imagem do UCSC Genome Browser mostrando a conservação do aminoácido alterado pelo polimorfismo rs74005645C>T em diferentes espécies.

Scale 20 bases | hg19 28,474,675 28,474,680 28,474,685 28,474,690 28,474,695 A G T A A G A A A T T C T T G C T C G C C A G T 28,474,665 G C A C A (74,665| 28,474,670| C A G A T C C chr15: 28,474,700 28,474,705 28,474,710 T T A A A T T G T A T C T TGGAA Simple Nucleotide Polymorphisms (dbSNP 150) Found in >= 12 of Samples RefSeq gene predictions from NCBI RefSed Curated Placenta l Mammal Basewise Conservation by Phylo Mammal Cons -4 Multiz Alignments of 46 Vertebrates Gap: N Human L D. Chimp Gorilla = = - - - - -- -- - - - --- - - - -Orangutan Rhesus Baboon Marmoset Tarsier Mouse_lemur Bushbaby Tree_shrew . . . H Ĥ. Mouse Rat Kangaroo_rat Guinea_pig Rabbit . Ĥ. Pika Alpaca . . . A. Do 1ph in Cow 8 Horse orse Cat N N Ń N N Ń N N Ń N N Ń N Ń N Ń N N Ń N Ń N Ń N N Ń N N Ń N Ń N Ń N Ń N Ń N Ń N Ń N Ń N Ń N Ń N Ń N Ń Dog . Microbat ŝ Ĥ. Megabat . . . 111 111 Hedgehog = - - -Shrew = = = E lephant Rock_hynax Tenrec . . . Armadillo Sloth Wallaby Opossum Platypus _ _ + _ _ + _ + _ + _ + _ + _ + _ + + - - -- - -Chicken 8 Zebra_f inch . . . Lizard . . . X_tropicalis - + - - + - - - - - - - -Tetraodon + = - - --= = = = = - -- - - - - -- -- - -- - -Fugu S , . E Stickleback . . Μ Medaka . . . Zebraf ish Lamprey = 4.41 - - - - - -- - - - -= = = + = = -÷., GERP Scores for Mammalian Alignments GERP -8.47

Apêndice I -Imagem do UCSC Genome Browser mostrando a conservação do aminoácido alterado pelo polimorfismo rs118112076G>C em diferentes espécies.

Apêndice J-Imagem do UCSC Genome Browser mostrando a conservação do aminoácido alterado pelo polimorfismo rs201872292G>A em diferentes espécies.


Apêndice K-Imagem do UCSC Genome Browser mostrando a conservação do aminoácido alterado pelo polimorfismo rs6497271G>A em diferentes espécies.



ANEXOS

Anexo A -Termo de Consentimento Livre e Esclarecido (TCLE)

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Nome da pesquisa: Sequenciamento de nova geração de genes envolvidos na biossíntese de melanina em amostra da população brasileira

Pesquisador responsável: Prof. Dr. Celso Teixeira Mendes Junior

Endereço para contato: Av. Bandeirantes, 3900; CEP: 14040-901, Ribeirão Preto - SP Departamento de Química, FFCLRP-USP Telefone: (16) 3602-0417

Prezado (a) doador (a),

A pigmentação humana é determinada pela presença de um pigmento denominado melanina. Proteínas produzidas por mais de 120 genes estão envolvidas na produção de melanina em células específicas e na distribuição desta melanina para outras células de nosso organismo. Variações normais nos genes envolvidos nestes processos levam aos diferentes tons de cor de cabelos, olhos e pele que vemos nas pessoas.

Nesse estudo, faremos a análise da sequência de algumas regiões do DNA que podem estar relacionadas à determinação das características físicas dos brasileiros, como a cor de olhos, pele e cabelos, tipo de cabelos e presença de sardas. Com isso buscamos identificar as variações (polimorfismos do DNA) e os mecanismos responsáveis pela determinação da cor das pessoas. Embora o conhecimento a ser produzido neste estudo não lhe trará benefícios diretos, os resultados desta pesquisa podem trazer grande contribuição para o desenvolvimento de terapias para tratar problemas de pele.

Para a realização desta pesquisa não haverá a necessidade de procedimentos que possam comprometer sua saúde. Caso concorde em participar da pesquisa, serão coletados cerca de 10 mL de seu sangue (aproximadamente uma colher de sopa). A coleta será feita por um profissional habilitado e altamente capacitado nessa atividade. Serão ainda obtidas imagens fotográficas de regiões pequenas e bem delimitadas de seu corpo (olho, raiz dos cabelos e regiões do braço não-expostas ao sol), as quais não permitirão sua identificação, mantendo, assim, o caráter confidencial

da informação. Adicionalmente, um colorímetro portátil, equipamento frequentemente usado em pesquisas dermatológicas, será usado para definir a cor de sua pele e dos seus cabelos com maior precisão, em um procedimento rápido indolor e livre de qualquer desconforto. Em todo este procedimento não existem riscos previsíveis além de um pequeno desconforto durante a coleta de sangue.

Os seguintes aspectos lhe são esclarecidos para que você decida sobre sua participação voluntária:

a.) Esta pesquisa não lhe trará qualquer custos. Visando lhe proporcionar maior conforto e comodidade, a coleta será realizada uma única vez em data e local a combinar.

b.) Não haverá nenhuma forma de reembolso em dinheiro, uma vez que sua participação na pesquisa não proporcionará nenhum gasto.

c.) Você terá a garantia de receber a resposta a qualquer pergunta ou esclarecimento a qualquer dúvida acerca dos procedimentos, riscos, benefícios e outros relacionados com a pesquisa. Isso poderá ser feito a qualquer momento. Nós temos o compromisso de proporcionar informações atualizadas durante o estudo, ainda que esta possa influenciar a sua vontade de continuar autorizando a sua participação.

d.) Você terá a liberdade de retirar o seu consentimento a qualquer momento sem que isso lhe traga prejuízo algum ou qualquer tipo de constrangimento.

e.) Você terá garantida a segurança de que não será identificado e que será mantida a sua privacidade e o caráter confidencial das informações concedidas.

f.) A amostra biológica será coletada com total responsabilidade e custo dos pesquisadores. A coleta de seu sangue será única e feita com material descartável, sendo necessário apenas um pequeno furo com agulha em veia do seu braço, o que pode, de fato, ser um procedimento desconfortável, mas apenas pela "picada" com a agulha.

g.) O sangue doado por você não será empregado na realização de pesquisas de outra natureza sem sua autorização.

h.) Será aplicado um questionário visando uma melhor compreensão de sua origem e características físicas. Toda informação obtida será mantida em sigilo.

i.) Os resultados da pesquisa serão apresentados em reuniões científicas (simpósios e congressos) e serão publicados em revistas científicas especializadas.

Anexos 148

Assinatura do pesquisador responsável			
Eu,		, RG nº,	
residente na		n°	
, cid	ade de	, telefone	
	abaixo assinado,	tendo recebido as informações acima e ciente dos meus	
direitos, concordo	em participar da pesqui	sa por livre e espontânea vontade. Declaro ainda que recebi	
uma cópia deste te	ermo de consentimento	assinada pelo pesquisador responsável.	

Tendo ciência do exposto acima, assino abaixo.

Ribeirão Preto, _____de 20____.

Assinatura do doador

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

PARA ARMAZENAMENTO DE MATERIAL BIOLÓGICO

Nome da pesquisa: Sequenciamento de nova geração de genes envolvidos na biossíntese de melanina em amostra da população brasileira

Pesquisador responsável: Prof. Dr. Celso Teixeira Mendes Junior

Endereço para contato: Av. Bandeirantes, 3900; CEP: 14040-901, Ribeirão Preto - SP Departamento de Química, FFCLRP-USP Telefone: (16) 3602-0417

Prezado (a) doador (a),

Como você concordou em participar da pesquisa "Sequenciamento de nova geração das regiões regulatórias de dez genes envolvidos na biossíntese de melanina em amostra da população brasileira", solicito agora uma nova autorização para que possamos armazenar parte do material biológico processado, para uso em pesquisas futuras desta mesma natureza, isto é, envolvendo a análise de outros genes e marcadores genéticos (polimorfismos do DNA).

Caso concorde, seu material biológico será armazenado sob minha responsabilidade em um freezer localizado em meu laboratório, compondo Biorrepositório devidamente registrado junto ao Comitê de Ética em Pesquisa desta Instituição. Este Biorrepositório será futuramente transformado em Biobanco, conforme determinação do Conselho Nacional em Saúde. Os seguintes aspectos lhe serão esclarecidos para que você decida sobre esta autorização voluntária:

a.) Você tem a liberdade de decidir se seu material será armazenado.

b.) Você terá a liberdade de retirar o seu consentimento a qualquer momento sem que isso lhe traga prejuízo algum ou qualquer tipo de constrangimento.

c.) O armazenamento não lhe trará qualquer custo. A amostra biológica será armazenada com total responsabilidade e custo dos pesquisadores.

d.) Toda nova pesquisa a ser realizada com o material armazenado será submetida para aprovação do Comitê de Ética em Pesquisa (CEP) institucional e, quando for o caso, da Comissão Nacional de Ética em Pesquisa (CONEP).

e.) Caso deseje, você tem o direito de ser consultado sempre que houver intenção de uso de seu material biológico em pesquisas futuras, para que eu autorize ou não autorize sua nova utilização.

f.) Você terá a garantia de receber a resposta a qualquer pergunta ou esclarecimento a qualquer dúvida acerca do seu material biológico, condições de armazenamento e sua utilização em outros projetos de pesquisa. Isso poderá ser feito a qualquer momento. Nós temos o compromisso de proporcionar informações atualizadas durante o estudo, ainda que esta possa influenciar a sua vontade de continuar autorizando a sua participação.

g.) Você terá garantida a segurança de que não será identificado e que será mantida a sua privacidade e o caráter confidencial das informações concedidas. Toda informação obtida será mantida em sigilo.

h.) Você tem o direito de ser informado sobre eventual perda ou destruição de suas amostras biológicas, bem como sobre eventual encerramento do Biorrepositório.

i.) Os resultados de novas pesquisas serão apresentados em reuniões científicas (simpósios e congressos) e serão publicados em revistas científicas especializadas.

Assinatura do pesquisador responsável

Eu,								_, RG nº				,
residente	na											_ nº
	,	cidade	de								, tele	efone
		<u></u>	abaixo	assinado,	tendo	recebido	as	informações	acima	e ciente	dos i	meus

direitos, concordo com o armazenamento de meu material biológico por livre e espontânea vontade.

Declaro ainda que:

 () desejo exercer o direito de ser consultado sempre que houver intenção de uso de meu material biológico em pesquisas futuras. () dispenso o direito de ser consultado quando houver intenção de uso de meu material biológico em pesquisas futuras, autorizando automaticamente seu uso em estudos desta mesma natureza, isto é, envolvendo a análise de outros genes e marcadores genéticos (polimorfismos do DNA).

Atesto que recebi uma cópia deste termo de consentimento assinada pelo pesquisador responsável. Tendo ciência do exposto acima, assino abaixo.

Ribeirão Preto, _____de 20____.

Assinatura do doador

Anexo B - Protocolo de análise laboratorial (HaloPlex) para sequenciamento de nova geração



Laboratório de Pesquisas Forenses e Genômicas Departamento de Química – FFCLRP/USP

Protocolo de análise laboratorial (HaloPlex) para sequenciamento de nova geração

Oliveira MLG, Marcorin L, Pereira ALE, Debortoli G,

Fracasso NCA, Silva GV, Mendes-Junior CT

Versão 1 - Outubro de 2016

Adaptado de: Oliveira MLG. Sequenciamento de nova geração do gene *IRF4*: identificação de variações associadas a fenótipos de pigmentação na população brasileira. Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo. 2016. 151p.

1. EXTRAÇÃO DE DNA GENÔMICO	2
2. QUANTIFICAÇÃO E VERIFICAÇÃO DA QUALIDADE DO DNA	2
3. PREPARO DOS REAGENTES PARA ENRIQUECIMENTO DE ALVO	3
4. ENRIQUECIMENTO DO ALVO E PREPARO DA BIBLIOTECA	3
4.1. Digestão e hibridação do DNA	4
4.2. Captura, ligação, eluição e PCR	5
4.3. Purificação da biblioteca	6
5. CONTROLE DE QUALIDADE E QUANTIFICAÇÃO DA BIBLIOTECA	6
6. SEQUENCIAMENTO	7
7. REFERÊNCIAS	8

1. EXTRAÇÃO DE DNA GENÔMICO

Após a coleta, a extração do DNA deve ser realizada a partir das amostras de sangue total utilizando o protocolo de *salting-out* (Miller, Dykes, & Polesky, 1988) com modificações.

O sangue coletado e mantido em tubo *Vacutainer*® com EDTA é transferido para um tubo *falcon* graduado de 50mL, ao qual é adicionado tampão de Lise I gelado (Tris-HCl 0,01M pH 7,5; Sacarose 0,3M; MgCl₂ 0,005M; Triton 1%; H₂O qsp) até completar o volume de 45mL. De forma a lisar as hemácias, o conteúdo é delicadamente misturado por inversão e centrifugado durante 15 minutos a 5000rpm em centrífuga 5430 R (Eppendorf AG). A extração prossegue com o botão de células obtido, o qual deve ser ressuspendido em 4,5mL de tampão de Lise II (NaCl 0,075M, EDTA 0,024M, H2O qsp, pH 8), 125µL de SDS 10% e 1,1mL de perclorato de sódio (5M) e vortexado a temperatura ambiente por 10 segundos para rompimento dos glóbulos brancos. Para lisar as proteínas, são adicionados 2mL de NaCl 6M (saturado) ao tubo *falcon*, que é vortexado a temperatura ambiente por 15 segundos. A seguir, o conteúdo é centrifugado por 15 minutos a 5000rpm em centrífuga 5430 R (Eppendorf AG). O sobrenadante é transferido para um novo tubo *falcon* de 50mL, ao qual é adicionado 5mL de isopropanol absoluto a temperatura ambiente (Merck Millipore, Darmstadt, Alemanha) e homogeneizado por inversão até que o DNA se separe e fique suspenso na solução.

Finalmente, o DNA é retirado com o auxílio de uma pipeta estéril e transferido para um microtubo de 1,5mL contendo 1mL de etanol (Merck Millipore, Darmstadt, Alemanha) 70% gelado. O microtubo é então centrifugado por 5 minutos a 5000rpm em centrífuga MiniSpin Plus (Eppendorf AG), e o sobrenadante descartado a seguir. O microtubo é deixado em centrífuga a vácuo (Concentrator plus, Eppendorf AG) durante 5min para remoção completa do etanol. Por fim, o DNA aderido ao microtubo é dissolvido em 200µL de água MilliQ, mantido na geladeira por 1 semana e posteriormente armazenado a -20°C no banco de amostras do laboratório.

2. QUANTIFICAÇÃO E VERIFICAÇÃO DA QUALIDADE DO DNA

O NanoDrop® ND-1000 (Thermo Fisher Scientific Inc.) é utilizado para avaliar o grau de contaminação do material genômico por proteínas e para estimativa inicial da concentração de DNA. É importante que as amostras possuam uma razão OD 260/280 variando de 1,8 a 2,0. Além disso, é realizada a eletroforese em gel de agarose, a fim de verificar o grau de degradação do material, considerando como indicativo de degradação qualquer banda abaixo de 2,5kb (*Agilent Technologies*).

Na sequência, as amostras são quantificadas através de fluorescência por meio do $Qubit^{TM} dsDNA BR Assay$ (Life Technologies). Para isso, o DNA genômico é diluído em água MilliQ em uma proporção de 1:10. Inicialmente, é preciso fornecer ao instrumento uma curva padrão. Para tanto, o corante (fluoróforo) para dsDNA é diluído no buffer do Qubit (1:200) e 10µL dos padrões (1 e 2) são diluídos em 200µL da solução anterior. Realizada a leitura dos padrões, uma curva padrão é estabelecida, a partir da qual é calculada a concentração da amostra original, com base no volume adicionado para leitura.

Conhecendo agora a concentração das amostras, essas são diluídas à concentração de 5ng/µL, no intuito de normalizar as amostras a uma concentração compatível para a montagem das bibliotecas de fragmentos.

3. PREPARO DOS REAGENTES PARA ENRIQUECIMENTO DE ALVO

Regiões alvo, incluindo a sequência promotora, CDS (*Coding DNA Sequence*), 5'UTR e 3'UTR de genes humanos, além de outras regiões de interesse, foram submetidas ao ensaio *in silico* de desenho de sondas para captura de DNA alvo utilizando o aplicativo *SureDesign* da *Agilent Technologies* (https://earray.chem.agilent.com/suredesign/home.htm). Para desenho das sondas, o *SureDesign* tomou como referência as coordenadas dos genes em questão junto ao *release* hg19/GRCh37 do genoma humano. Visando a cobertura de 284.367pb, um total de 12.131 sondas foi desenhado, abrangendo 99,63% de tal extensão. Devido à distribuição dos sítios de restrição utilizados posteriormente no preparo das bibliotecas, as sondas desenhadas capturam uma região de 488.658pb, que garante o sequenciamento parcial dos introns flanqueadores dos exons alvo.

4. ENRIQUECIMENTO DO ALVO E PREPARO DA BIBLIOTECA

Após a quantificação, cada amostra de DNA é processada individualmente, gerando uma biblioteca contendo todos os amplicons. A captura das regiões-alvo utiliza reagentes do kit de enriquecimento de alvo customizado *HaloPlex* (*Agilent Technologies*)

e segue o protocolo sugerido pelo fabricante. Resumidamente, o protocolo consiste dos seguintes passos: 1) digestão do DNA genômico com enzimas de restrição; 2) hibridação dos fragmentos de DNA a sondas biotiniladas cujos finais são complementares a esses fragmentos-alvo (durante esse passo os fragmentos são circularizados e índices, que irão identificar cada indivíduo durante o sequenciamento, são adicionados); 3) captura dos fragmentos-alvo utilizando *beads* magnéticas recobertas por estreptavidina; 4) ligação dos fragmentos circularizados; 5) amplificação por PCR dos fragmentos-alvo capturados para montagem da biblioteca.

4.1. Digestão e hibridação do DNA

A digestão das amostras é feita conforme indicações do fabricante, com algumas modificações (*HaloPlex Target Enrichment System For Illumina Sequencing, Agilent Technologies*, 2013). Um total de 225ng (45 μ L de uma solução 5ng/ μ L) de DNA genômico de cada amostra e do *Enrichment Control DNA* (ECD), no volume total de 45 μ L, é separado e reservado em gelo.

Digestão: Para o mix de restrição, preparado para lotes de 12 amostras (11 amostras de DNA genômico e outra de ECD), são misturados 476µL de *RE buffer* e 11,9µL de BSA e esse conteúdo é distribuído em 8 tubos. Em cada um dos 8 tubos é adicionado um par de enzimas de restrição (7µL de cada). Cada amostra é digerida pelos 8 pares de enzima separadamente. Para isso, cada amostra é distribuída em 8 tubos (5µL por tubo) onde são adicionados 5µL dos mixes contendo as enzimas de restrição resultando, para um total de 12 amostras, em 96 tubos de digestão que são incubados no termociclador (*SureCycler 8800 Thermal Cycler, Agilent Technologies*) a 37°C por 30 minutos.

<u>Hibridação</u>: Para o mix de hibridação são misturados 650µL da *Hybridization Solution* e 260µL das sondas *Haloplex* em um tubo e 70µL dessa solução é distribuída por amostra. Adicionalmente, são diluídos nessa solução 10µL de um dos índices disponíveis em cada um dos doze tubos (índices diferentes são utilizados para cada amostra, permitindo sua identificação após a mistura das amostras para o sequenciamento).

O DNA digerido de cada amostra (80µL somando todos os 8 tubos de digestão) é então transferido para os microtubos contendo o mix de hibridação (70µL) e seus índices

correspondentes (10μL). Essas soluções são incubadas a 54°C durante 16 horas. O tempo de duração da hibridação é determinado de acordo com a densidade de sondas do ensaio. Segundo instruções do fabricante (*Haloplex, Agilent Technologies*), ensaios que atingem cobertura de entre 500kb - 5Mb exigem 16 horas de hibridação, enquanto que ensaios com cobertura de entre 1-500kb exigem apenas 3 horas de hibridação. Apesar de o presente ensaio estar voltado para a cobertura de apenas 488,7kb, os ensaios iniciais demonstraram a necessidade de um tempo de hibridação de 16 horas.

4.2. Captura, ligação, eluição e PCR

<u>Captura:</u> A captura dos fragmentos hibridados é feita com o uso de 520µL de *beads* magnéticas recobertas por estreptavidina. Com o auxílio de uma placa magnética (*DynaMag*^{TM-2} *Magnet, Life Technologies*), o sobrenadante da solução de *beads* é removido e substituído por 520µL de *Capture Solution*. Deste conteúdo, 40µL é adicionado a cada solução de DNA hibridizado (160µL) e após 15min de incubação em temperatura ambiente, os tubos contendo as soluções são colocados junto a uma placa magnética de 96 poços (*Agencourt SPRIPlate 96R, Beckman Coulter*) para a remoção do sobrenadante. A seguir, adiciona-se 100µL de *Wash Solution* por amostra, e essa solução é incubada no termociclador a 46°C por 10min. Após esse período, o sobrenadante é novamente removido.

<u>Ligação</u>: Procedendo com a ligação e circularização dos fragmentos alvo, adiciona-se 617,5µL de *Ligation Solution* e 32,5µL de DNA ligase e distribuí-se 50µL dessa solução a cada amostra hibridada e capturada. A solução é então incubada a 55°C por 10min.

<u>Eluição</u>: A seguir, lavamos o conteúdo de DNA ligado com 100µL de SSC buffer, retirando o sobrenadante ao final da etapa com o auxílio da placa magnética. Para eluir o DNA das *beads* adicionamos 25µL de NaOH (50mM) e encubamos em temperatura ambiente por 1min..

<u>PCR</u>: Com o auxílio da placa magnética, coletamos o DNA eluído num total de 20µL por amostra e o transferimos para tubos contendo mix de PCR. Tal mix para a reação de PCR é preparado com 209,3µL de água MilliQ, 130µL de 5X *Herculase II Reaction Buffer*, 5,2µL de dNTPs (100mM, 25mM para cada dNTP), 13µL de cada um dos *primers forward* e *reverse* (25µM), 6,5µL de ácido acético (2M) e 13µL da polimerase *Herculase*

II Fusion, somando um total de 30μL por amostra. O volume final da reação é de 50μL por amostra. As condições de ciclagem da PCR seguem uma etapa inicial de desnaturação por 10 minutos a 98°C, seguida por 20 ciclos de 98°C por 30 segundos, 60°C por 30 segundos e 72°C por 1 minuto, e um passo final a 72°C por 10 minutos. Após o término da reação de PCR o produto é mantido a 8°C.

4.3. Purificação da biblioteca

Os produtos de PCR são então purificados utilizando *AMPure XP beads* (*Beckman Coulter*) de acordo com as instruções da *Agilent Technologies*. Um total de 40µL do produto de PCR é misturado a 100µL de *AMPure XP beads* mais 40µL de água MilliQ por amostra. A seguir, essa solução é colocada junto à placa magnética para remoção do sobrenadante e é lavada duas vezes com álcool 70%, que é completamente removido ao final. Então, adicionamos 40µL de Tris-HCl (pH = 8,0) para eluir o DNA capturado pelas *AMPure XP beads*. O produto final de 40µL é armazenado a -20 °C.

5. CONTROLE DE QUALIDADE E QUANTIFICAÇÃO DA BIBLIOTECA

A qualidade das bibliotecas é um dos principais determinantes para o sucesso da corrida de sequenciamento. Bibliotecas de fragmentos que não correspondem à distribuição de tamanho recomendada podem levar a uma baixa cobertura ou à falha de cobertura em determinadas porções das sequências alvo (Cher, 2011).

O controle de qualidade das bibliotecas é feito através do sistema de eletroforese capilar 2100 Bioanalyzer (Agilent Technologies) utilizando um chip de alta-sensibilidade. A eletroforese é capaz de separar os fragmentos com base no tamanho, permitindo dessa forma avaliar a qualidade das bibliotecas geradas. É esperado para este ensaio que as bibliotecas apresentem um perfil de picos variando de 175 a 625pb, conforme previsto por um eletroferograma padrão (Figura 1). A partir da seleção manual da faixa de fragmentos gerados (175 a 625pb) é possível obter a concentração (em pg/ μ L e pmol/L) referente a este intervalo, sendo a molaridade utilizada para a quantificação de cada uma das bibliotecas para composição adequada do *pool* de bibliotecas para o sequenciamento. Adicionalmente, as bibliotecas individuais podem ser quantificadas por fluorescência (*Qubit*TM *dsDNA BR Assay, Life Technologies*) para composição adequada do *pool* de bibliotecas para o sequenciamento.



Figura 8: Perfil de picos esperado no eletroferograma para as bibliotecas de fragmento deste ensaio. Tal eletroferograma é produzido pela *Agilent Technologies* durante a avaliação do kit *Haloplex* customizado por meio de ensaio *in silico* de desenho de sondas para captura de DNA alvo utilizando o aplicativo *SureDesign*. Para leitura correta no *2100 Bioanalyzer* é crítico que os 13 picos previstos no *ladder* sejam identificados, que os picos do menor (35pb) e do maior (10380pb) marcador apresentem um tempo de corrida correspondente entre as amostras analisadas e o *ladder* e que as linhas de base sejam planas (Cher, 2011).

6. SEQUENCIAMENTO

Como descrito no item anterior, a partir das concentrações obtidas no 2100 Bioanalyzer e Qubit, prepara-se soluções a 4nM de biblioteca, de maneira que cada biblioteca tenha a mesma representatividade. O 2100 Bioanalyzer é utilizado também para quantificação deste pool final de bibliotecas. O pool final de bibliotecas pode ser adicionalmente quantificado por fluorescência (Qubit[™] dsDNA BR Assay, Life Technologies) ou PCR em tempo real (ViiA[™] 7 Real-Time PCR System, Applied Biosystems). Na PCR quantitativa em tempo real (qPCR), através da construção de uma curva padrão, é possível determinar a quantidade absoluta de DNA nas amostras a serem sequenciadas (Applied Biosystems ViiA[™] 7 Real-Time PCR System. User Guide. Life Technologies, 2010, 262p).

A seguir, 5µL do *pool* de bibliotecas a 4nM são desnaturados pela adição de 5µL de NaOH (0,2M) e 990µL *de Hybridization Buffer* (HT1) para obtenção de uma biblioteca

a 20pM, que é diluída com HT1 para que se chegasse à 600µL de uma solução de concentração 16pM. Essa solução é inserida no cartucho de sequenciamento do *MiSeq Personal Sequencer* (*Preparing Libraries for Sequencing on the MiSeq*, California, USA, *Illumina*, Inc, 2013, 14p).

Antes de iniciar o sequenciamento é preciso carregar uma *SampleSheet* ou planilha de amostras (*.csv) contendo as informações necessárias para configurar a corrida, as quais correspondem à lista de amostras a serem sequenciadas e às sequências dos índices utilizados em cada amostra.

Durante o sequenciamento, as moléculas de DNA fita simples se ligam à superficie da *flow cell* por meio dos adaptadores presentes em suas extremidades. O sequenciamento ocorre a partir da síntese da fita complementar ao fragmento ligado na *flow cell* utilizando didesoxinucleotídeos marcados com fluoróforos específicos para cada base (A, T, C e G). O *software* Bustard relaciona então a fluorescência emitida após cada incorporação ao didesoxinucleotídeo correspondente com o *cluster* de fragmentos fixados à região que emitiu o sinal. No caso do sequenciamento usando reagentes *MiSeq Reagent Kit V3 (600 cycle)*, são realizados 301 ciclos em leituras do tipo *paired-end* (cada fragmento é sequenciado tanto na fita *forward* quanto na *reverse*) perfazendo um total de 2x301 ciclos de leitura, além de 8 ciclos referentes à leitura do índice incorporado em cada fragmento capturado, o que relaciona de maneira unívoca o fragmento a uma determinada amostra.

7. REFERÊNCIAS

- Cher, C. (Field A. S. (2011). Bioanalyzer Applications for Next Gen Sequencing: Updates and Tips, 1–49.
- Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16(3), 1215. https://doi.org/10.1093/nar/16.3.1215

Anexo~C -Protocolo de processamento computacional de dados de sequenciamento de nova geração



Protocolo de processamento computacional de dados de sequenciamento de nova geração

Oliveira MLG, Marcorin L, Pereira ALE, Debortoli G,

Fracasso NCA, Silva GV, Castelli EC, Mendes-Junior CT

Versão 2 – março de 2018

1

Adaptado de: Oliveira MLG. Sequenciamento de nova geração do gene *IRF4*: identificação de variações associadas a fenótipos de pigmentação na população brasileira. Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo. 2016. 151p.

1. ANÁLISE DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO	
1.1. Análise primária	2
1.2. Análise secundária	3
1.2.1. Avaliação da qualidade	3
1.2.2. Trimagem	3
1.2.3. Alinhamento ao genoma de referência	5
1.2.3.1. SAM	7
1.2.3.2. BAM	8
1.3. Análise terciária	9
1.3.1. Determinação de variantes	9
1.3.1.1. VCF	12
1.3.2. Remoção de genótipos de baixa qualidade	14
2. REFERÊNCIAS	15

1. ANÁLISE DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO

O fluxograma de análises no NGS pode ser sumarizado em três fases principais: a fase primária, que consiste na geração dos dados no formato FastQ, a fase secundária, que compreende a aplicação de filtros de qualidade e o alinhamento a um genoma de referência e a fase terciária, que compreende o processo de determinação das variantes através de comparação das sequências alvo a um genoma de referência.

1.1. Análise primária

O processamento dos dados segue passos gerais pré-estabelecidos, de acordo com o objetivo pretendido, e tem início no *base calling* (ou a interpretação da imagem obtida a partir do sequenciador), quando são gerados os dados brutos do sequenciamento no formato *base call* (*.bcl). Estes dados são então disponibilizados através do ambiente de computação em nuvem da *Illumina*, *BaseSpace* (<u>https://basespace.illumina.com</u>), no formato FastQ (*.fastq.gz), o qual possui uma pontuação de qualidade associada (*Qscore*) a cada base sequenciada.

Essa pontuação de qualidade, também denominada *Phred-score* por ter sido introduzida como convenção pelo programa Phred (Ewing, Hillier, Wendl, & Green, 1998), se relaciona à probabilidade de erro na determinação de uma base. A medida se baseia na fórmula a seguir, onde dado a determinação de uma base X, a probabilidade de que X não seja verdadeira (P(X)) é expressa pela relação Q(X) = -10 log₁₀ (P(X)) ou P(X) = $10^{-Q(X)/10}$. Assim, um Q30 indica probabilidade de ocorrência de um erro a cada 1.000 bases determinadas.

Cada arquivo FastQ é nomeado como no exemplo a seguir, onde PIG0001 representa o nome da amostra, S1 representa o número da amostra e indica neste exemplo que esta é a primeira amostra listada na *SampleSheet*, L001 indica o número da linha que a amostra ocupa na *flow cell*, sendo sempre o mesmo, já que a *flow cell* utilizada possui apenas uma linha, R1 e R2 indicam que o arquivo é composto por *reads forward* e *reverse*, respectivamente, enquanto o último segmento 001 sempre recebe essa denominação:

PIG0001_S1_L001_R1_001.fastq.gz PIG0001 S1 L001 R2 001.fastq.gz

1.2. Análise secundária

Essa segunda etapa consiste na aplicação de filtros de qualidade e posterior alinhamento dos fragmentos sequenciados a um genoma de referência.

1.2.1. Avaliação da qualidade

Em um primeiro momento, é importante verificar a qualidade das *reads* sequenciadas, a fim de garantir que os dados não apresentem problemas que poderiam afetar o passo seguinte de mapeamento. Para tanto, pode ser empregado o *software* FastQC (Andrews S., 2010), o qual utiliza um arquivo FastQ para gerar um relatório completo da qualidade das sequências (Figura 1a). A partir dos resultados apresentados é possível decidir por estratégias que melhorem a qualidade dos dados gerados.



Figura 1: Gráfico de distribuição da qualidade por base no programa FastQC. Cada coluna apresenta um *boxplot* com os valores de qualidade média das bases incorporadas em todas as *reads* em um determinado ciclo. Como pode ser observado em (a) ocorre uma diminuição da qualidade (Q<30) das *reads* não-processadas a partir dos ciclos 150-154. O eixo Y apresenta os *scores* de qualidade, sendo dividido conforme cores que indicam: verde (boa qualidade), laranja (qualidade razoável) e vermelho (baixa qualidade). (b) Após processadas pelo *cutadapt*, foi obtida uma maior proporção das *reads* com qualidade desejável, ainda que a partir dos ciclos 225-229 tenha ocorrido uma diminuição da qualidade, o que, no entanto, é comum para ensaios da *Illumina*.

1.2.2. Trimagem

Sequências de adaptadores, bem como bases diagnosticadas com baixo *Q-score* devem ser retiradas das extremidades das *reads*, em um processo referido como trimagem. As sequências obtidas podem ser trimadas de diversas maneiras: por tamanho, qualidade ou a partir de sequências conhecidas.

A trimagem por qualidade parte do princípio de que é preferível trabalhar com sequências menores de maior qualidade, removendo sequências de bases na extremidade da *read* que tenham uma qualidade abaixo do limite estabelecido como adequado. A partir do que pôde ser observado no relatório gerado pelo FastQC optou-se por realizar a trimagem apenas pelos adaptadores *paired-end TruSeq da Illumina* utilizados no sequenciamento.

No processo de sequenciamento *Illumina*, todos os *clusters* são sequenciados seguindo o número de ciclos previstos pelo kit de reagentes utilizado (no caso, 2x301 ciclos). Mesmo que um *cluster* inclua um pequeno fragmento a ser sequenciado, os ciclos de sequenciamento não são interrompidos, continuando até alcançar o número total de ciclos. Como consequência, as *reads* geradas como *output* podem conter, além das sequências alvo (caso essas sejam menores que o número de ciclos), as sequências dos adaptadores do outro extremo do fragmento. Caso ainda restem ciclos após o sequenciamento dos adaptadores, leituras adicionais continuam sendo feitas, embora com qualidade extremamente baixa, uma vez que nucleotídeos inexistentes são aleatoriamente identificados (possivelmente devido à fluorescência de diferentes nucleotídeos não incorporados e não retirados durante a lavagem). Logo, é preciso encontrar e remover esses adaptadores, o que resultará, consequentemente, na remoção das bases incorporadas aleatoriamente.

Em função da construção realizada pelo *Haloplex* (*PCR primer --- Illumina adaptor --- TARGET --- Illumina adaptor --- Barcode --- PCR primer*), ao remover as sequências dos adaptadores de uma *read* podemos remover tudo o que existe depois dessas sequências, neste caso os *primers* da PCR e o *barcode*, quando presente.

A trimagem pelos adaptadores pode ser realizada com uso do *cutadapt* (Martin, 2011) utilizando a seguinte linha de comando (Tabela 1):

cutadapt -e 0.2 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A AGATCGGAAGAGCGTCGTGTGGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT -o sample.trimmed.r1.fastq -p sample.trimmed.r2.fastq input.raw.r1.fastq input.raw.r2.fastq

Comandos utilizados no cutadapt	Função do comando executado
-e 0.2	indica o nível máximo de tolerância de erros (<i>mismatches</i> , inserções e deleções) entre as sequências dos adaptadores e sequências presentes nas <i>reads</i> , para que uma sequência adaptadora seja efetivamente trimada.
-a	indica que o parâmetro a seguir se refere à sequência do adaptador <i>forward</i> .
-A	indica que o parâmetro a seguir se refere à sequência do adaptador <i>reverse</i> .
-0	indica o arquivo <i>output</i> da primeira <i>read</i> do par.
-p	indica o arquivo <i>output</i> da segunda <i>read</i> do par.

Tabela 1: Descrição dos componentes da linha de comando executada no cutadapt

Tal procedimento resultou em melhorias significativas na qualidade média das *reads* processadas (Figura 1b).

1.2.3. Alinhamento ao genoma de referência

Estando as *reads* processadas, o próximo passo é mapeá-las contra um genoma de referência. A maior parte dos programas de alinhamento disponíveis usam um dentre os seguintes algoritmos de mapeamento: baseado em *hash* ou na transformação de Burrows-Wheeler (BWT). Estes algoritmos irão construir um arquivo de alinhamento com base nos dados de *single-end* e em uma análise seguinte utilizarão os dados de *paired-end* para refinar os alinhamentos gerados.

Para fazer o alinhamento pode ser utilizado o BWA (Li & Durbin, 2009), o qual utiliza um algoritmo do tipo BWT, que tem como principal vantagem o menor consumo de memória. O BWA é um pacote de softwares, que consiste em três algoritmos, que são chamados por diferentes comandos: aln/samse/sampe para o BWA-backtrack, bwasw para o BWA-SW e men para o BWA-MEM.

Antes de realizar o alinhamento, porém, é preciso indexar o genoma de referência por meio do índice FM (Ferragina, Manzini, Veli, & Navarro, 2004). Neste caso, utilizamos um genoma já indexado (*Homo sapiens*, UCSC, hg19), disponibilizado pela *Illumina* através do *iGenomes* em: <u>https://support.illumina.com/sequencing/</u> <u>sequencing_software/igenome.html</u>. Pode-se também indexar localmente o genoma de interesse por meio de ferramentas específicas. Para tanto, é necessário gerar três diferentes arquivos, arquivo índice, arquivo fasta índice e um dicionário de sequência, os quais podem ser obtidos por meio dos programas BWA, Samtools e Picard, respectivamente.

O algoritmo utilizado para realizar o alinhamento é o BWA-MEM em função da sua especificidade para sequências com tamanho entre 70 e 1Mb e por ser mais rápido e acurado que o BWA-SW. A seguinte linha de comando deve ser utilizada (Tabela 2):

```
bwa mem -t 6 -R
'\@\R\G\\tID:sampleID\tLB:sampleLB\tSM:tag\tPL:illumina\tPU:samp
leP' /path/to/reference/hg19.fa
sample.trimmed.r1.fastq sample.trimmed.r2.fastq
> mapped.sample.sam
```

Tabela 2: Descrição dos componentes da linha de comando executada no BWA.

Comandos utilizados no BWA	Função do comando executado
-t	number of threads (número de processos com um endereço/espaço compartilhado no Linux).
-R	indica que o parâmetro a seguir se refere à linha do cabeçalho de determinado grupo de <i>reads</i> .
\t	comando convertido em TAB no output SAM, a fim de separar adequadamente os campos.
\@\R\G\\	identifica o cabeçalho de cada grupo de <i>reads</i> . Permite ainda que o arquivo BAM possa ser lido pelo GATK.
ID	identifica a qual grupo de <i>reads</i> cada <i>read</i> pertence. O ID permite que ao invés de ter que lidar com múltiplos grupos de dados, se passe a lidar com apenas um. Assim, a informação sobre o grupo de <i>reads</i> permite a identificação dos dados de diferentes experimentos, ainda que eles estejam combinados em um arquivo único. Cada linha @RG deve conter um ID único. Logo, todas as <i>reads</i> de um mesmo grupo são consideradas como parte da mesma corrida de sequenciamento e indicam a linha que ocupam nessa corrida. Por exemplo: ID: FLOWCELL1. LANE2.
LB	identificador da biblioteca de DNA.
SM	nome da amostra sequenciada em determinado grupo de <i>reads</i> . Todo dado que tiver um mesmo valor SM será tratado como pertencendo a uma mesma amostra.

Comandos utilizados no BWA	Função do comando executado
PL	identifica a plataforma de sequenciamento utilizada.
PU	denominação alternativa do grupo de <i>reads</i> . Mantém a informação sobre a linha onde a amostra se encontra na <i>flow cell</i> . Essa designação não é exigida pelo GATK, mas prevalece sobre o ID caso seja executada uma recalibração de base.
/path/to/reference/hg19.fa	identifica o caminho a ser seguido para consulta ao genoma de referência
sample.trimmed.r1.fastq	<i>input</i> para o mapeamento, contendo a primeira <i>read</i> do par
sample.trimmed.r2.fastq	<i>input</i> para o mapeamento, contendo a segunda <i>read</i> do par
> mapped.sample.sam	indica o redirecionamento do resultado do mapeamento para o arquivo SAM especificado.

Continuação Tabela 2: Descrição dos componentes da linha de comando executada no BWA.

Como *output* do comando acima é gerado um arquivo SAM (*.sam). Porém, uma vez que os arquivos SAM são arquivos de texto que ocupam muito espaço no disco rígido, e para que as informações sobre mapeamento pudessem ser visualizadas, esse arquivo pode ser convertido para o formato BAM (*.bam).

1.2.3.1. SAM

O formato SAM (*Sequencing Alignment/Map*) (Li et al., 2009) é um arquivo baseado em texto, utilizado como *output* por diversos programas de alinhamento. Cada linha no arquivo SAM armazena informações sobre uma *read* mapeada contra um genoma de referência.

Arquivos SAM possuem uma seção com o cabeçalho (que é opcional) e uma seção de alinhamento. As linhas no cabeçalho iniciam com "@" seguido por um código de duas letras. Por sua vez, cada *read* alinhada à referência é representada em uma linha da seção de alinhamento.

Podem estar presentes ainda linhas opcionais com marcações predefinidas (*The SAM/BAM Format Specification Working Group, Sequence Alignment/Map Format*

Specification, 2015, http://samtools.github.io/hts-specs/SAMv1.pdf). Algumas destas marcações são geradas pelo BWA (aquelas começando com "X").

1.2.3.2. BAM

O formato BAM (*Binary Alignment/Map*) é a versão binária compactada de um arquivo SAM, indexável e amplamente utilizado para representar as sequências após o alinhamento. A compactação em blocos (chamada BGZF) de um arquivo BAM permite além da diminuição do espaço de armazenamento o carregamento parcial do arquivo, o que é bastante utilizado em programas de visualização como o IGV ou o *UCSC Genome Browser*.

A conversão do arquivo SAM em BAM gera um arquivo com as mesmas informações e pode ser realizada através do programa *SAMtools* (<u>http://samtools.sourceforge.net/</u>), o qual fornece diversas ferramentas para manipulação desses arquivos.

Para visualizar o arquivo BAM e permitir sua utilização pelos *softwares* de determinação de variantes, é preciso ordenar os alinhamentos de acordo com a posição genômica. O comando a seguir é utilizado para realizar a conversão de SAM em BAM e a ordenação do arquivo BAM:

samtools sort mapped.sample.sam sorted.mapped.sample.bam

O nome do arquivo gerado pelo comando seguiu o formato a seguir, onde PIG0001 representa o nome da amostra e *sorted* indica que o arquivo BAM gerado se encontra ordenado:

PIG0001.sorted.bam

Arquivos BAM armazenam as mesmas informações que um arquivo SAM e possuem uma seção de cabeçalho, que contém informações gerais sobre o arquivo como nome, tamanho da amostra, método de alinhamento utilizado e uma seção com o alinhamento propriamente dito contendo nome, sequência e qualidade da *read*, além de marcações personalizadas.

Os arquivos BAM devem ainda ser indexados, permitindo uma pesquisa rápida no arquivo ordenado. Um índice de correspondência do arquivo BAM em questão em relação às coordenadas de referência será identificado pela terminação (*.bam.bai). O comando a seguir permitiu gerar um arquivo indexado:

samtools index sorted.mapped.sample.bam

O output do comando acima pode ser nomeado como a seguir:

PIG0001.sorted.bam.bai

Uma vez ordenados e indexados, os arquivos BAM puderam ser visualizados através de um visualizador externo. Neste trabalho utilizamos o *Integrative Genomics Viewer* (IGV) 2.3.67, um programa em java disponibilizado pelo *Broad Institute* (Robinson et al., 2011; Thorvaldsdottir, Robinson, & Mesirov, 2013).

1.3. Análise terciária

1.3.1. Determinação de variantes

Em um projeto de sequenciamento, o passo seguinte ao mapeamento é a detecção das diferenças entre as regiões genômicas sequenciadas e o genoma de referência (determinação de variantes). Um programa amplamente utilizado para análise dos dados de NGS é o *Genome Analysis Toolkit* (GATK) v.3.5, disponibilizado pelo *Broad Institute* (DePristo et al., 2011; McKenna et al., 2010; Van der Auwera et al., 2013).

O programa apresenta uma estrutura modular e é dividido em diversas análises, as quais são reunidas junto a um *framework* em um arquivo java (*.jar). Para ter acesso à lista de análises disponíveis é utilizado o comando:

java -jar GenomeAnalysisTK.jar

Já para realizar uma análise específica é utilizado o comando:

java -jar GenomeAnalysisTK.jar -T <analysis> <arguments*>

Para realizar a determinação de SNPs e indels as ferramentas *HaplotypeCaller* com o parâmetro --emitRefConfidence GVCF e *UnifiedGenotyper* do GATK podem ser utilizadas. O *HaplotypeCaller* considera simultaneamente todas as amostras a serem analisadas e, ao encontrar uma região apresentando sinais de variação (região ativa), o programa descarta o mapeamento preexistente e realiza uma remontagem das *reads* naquela região. O programa realinha então cada sequência contra a sequência de referência com o objetivo de identificar potenciais sítios de variação.

Comparado ao *UnifiedGenotyper*; o *HaplotypeCaller* possibilita a nomeação de SNPs e indels de forma simultânea. Isso permite que o *HaplotypeCaller* seja mais preciso quando lidando com regiões reconhecidamente problemáticas, como por exemplo, onde são encontrados diferentes tipos de variações próximas umas das outras (Van der Auwera et al., 2013).

Por outro lado, o *UnifiedGenotyper* possui filtros menos restringentes que possibilitam encontrar variantes com frequências populacionais mais baixas, principalmente em áreas de baixa cobertura, como *singletons* e alelos raros em polimorfismos multialélicos. Porém, essa característica pode resultar em uma taxa alta de falsos positivos. Além disso, essa ferramenta faz a leitura das sequências ponto a ponto, não sendo aplicável à identificação de indels. Por essas e outras razões, o próprio site do GATK descontinuou as atualizações e aposentou a ferramenta *UnifiedGenotyper* recomendando o uso do *HaplotypeCaller* em seu lugar.

Devido a utilização do parâmetro --emitRefConfidence GVCF, é gerado para cada amostra separadamente um arquivo gVCF (do inglês, *genomic variant calling format*), onde estão dispostas um registro para cada posição (ou intervalo de interesse) independentemente de uma variante ser detectada naquele sítio ou não. O uso do gVCF é importante pois permite a adição de outros conjuntos de amostra, caso necessite realizar

uma análise conjunta. Desta forma, pode-se adicionar os arquivos gVCFs de diferentes coortes e realizar a análise de chamada de variantes simultaneamente nestes conjuntos de amostra, aumentando a acurácia nesta determinação.

As seguintes linhas de comando são utilizadas para realizar a nominação das variantes (Tabela 3):

HaplotypeCaller:

```
java -jar GenomeAnalysisTK.jar \
    -R reference.fasta \
    -T HaplotypeCaller \
    -I listofsamples.list \
    --emiteRefConfidence GVCF \
    -dbsnp dbSNP.vcf \
    -L targets.interval_list \
    -o output.raw.snps.indels.g.vcf \
    -dontUseSoftClippedBases \
    -drf DuplicateRead
```

UnifiedGenotyper:

```
java -jar GenomeAnalysisTK.jar \
    -T UnifiedGenotyper \
    -R reference.fasta \
    -I sample1.bam [-I sample2.bam ...] \
    --dbsnp dbSNP.vcf \
    -o snps.raw.vcf \
    -stand_call_conf [50.0] \
    -stand_emit_conf 10.0 \
    [-L targets.interval_list]
```

Tabela 3: Descrição dos componentes da linha de comando executada no GATK

Comandos utilizados no GATK	Função do comando executado
-R	indica o arquivo com o genoma de referência
-T	indica qual ferramenta do GATK será utilizada,
	isto é, HaplotypeCaller ou UnifiedGenotyper
I-	indica arquivo de texto com lista de amostras
	(formato BAM) a ser utilizado pelo GATK
dbsnp	indica arquivo (dbSNP.vcf) contendo a identificação dos rs de cada ponto de variação oficialmente descrito, a ser utilizada para preenchimento da coluna ID do <i>output</i>
-stand_call_conf	é o limiar mínimo de confiança (na escala Phred) utilizado para separar determinações de genótipo realizadas com maior e menor confiança. Apenas determinações de genótipo emitidas com confiança ≥ limiar mínimo são

Continuação Tabela 3: Descrição dos componentes da linha de comando executada no GATK

Comandos utilizados no GATK	Função do comando executado
-stand_emit_conf	permite emitir determinações de genótipo com baixa qualidade como resultados filtrados
emitRefConfidence GVCF	indica a produção do arquivo g.vcf para o HaplotypeCaller
-L	direciona o GATK a restringir o processamento dos dados aos intervalos genômicos especificados em um arquivo <i>"targets.interval_list"</i> , que indica o cromossomo e as coordenadas das posições de início e fim em relação ao genoma de referência
-0	indica o <i>output</i> , que é um arquivo VCF ou gVCF (Variant Call Format) contendo os sítios que o algoritmo (UnifiedGenotyper ou HaplotypeCaller, respectivamente) identificou como sendo variáveis
dontUseSoftClippedBases	direciona o programa a não analisar bases <i>soft</i> <i>clipped</i> (bases não alinhadas nas extremidades das <i>reads</i>), o que diminui a determinação de falso-positivos
-drf	desabilita a aplicação de filtros à análise (DuplicateRead neste caso)

Após gerar os arquivos g.vcf para cada amostra um passo é realizado para combinar os arquivos em um único arquivo vcf com a linha de comando:

```
java -jar GenomeAnalysisTK.jar \
    -T GenotypeGVCFs \
    -R reference.fasta \
    -I listofsamples.list \
    -o cohort.g.vcf
```

Tabela 4: Descrição dos componentes da linha de comando executada no GATK

Função do comando executado
indica o arquivo com o genoma de referência
indica qual ferramenta do GATK será utilizada,
isto é, GenotypeGVCFs
indica arquivo de texto com lista de amostras
(formato BAM) a ser utilizado pelo GATK
indica o output, que é um arquivo gVCF
(Variant Call Format) contendo os sítios que o

algoritmo (HaplotypeCaller) identificou como
sendo variáveis

1.3.1.1. VCF

O VCF é um formato de arquivo que contém informações sobre variantes encontradas em posições específicas em relação a um genoma de referência. O arquivo VCF possui um cabeçalho, bastante útil, uma vez que contém uma descrição para os campos contidos no arquivo (Tabela 5):

Informações presentes no cabeçalho	Descrição das abreviações apresentadas		
Fileformat	versão do formato de arquivo e a versão de		
	variant caller utilizado		
FILTER	identifica quais filtros são aplicados aos dados		
FORMAT e INFO	essas linhas definem as anotações contidas na		
	colunas correspondentes a essas informações		
	no arquivo VCF		
GATKCommandLine	contém a linha de comando utilizada pelo GATK para realizar a genotipagem das amostras. Essa linha de comando especifica todos os parâmetros utilizados pelo <i>variant</i>		
	genoma de referência e do arquivo BAM		
Contig lines e Reference	contém informações como os nomes dos		
	fragmentos, tamanho e genoma de referência		
	que é utilizado como <i>input</i> para gerar o arquivo		
	BAM		

 Tabela 5: Descrição dos componentes da seção de cabeçalho de um arquivo VCF

Além do cabeçalho, um arquivo VCF contém as linhas com os dados das amostras, onde estão contidas em cada linha informações sobre uma única variante. As linhas com os dados são listadas seguindo o designado no cabeçalho do arquivo VCF. As oito primeiras colunas do arquivo (até INFO) representam as propriedades observadas dos sítios variantes (ou invariantes). Informações específicas à amostra são apresentadas na coluna FORMAT e nas seguintes.

As nove primeiras colunas são necessárias para o formato VCF, ainda que possam estar vazias. As informações contidas em cada coluna estão apresentadas na Tabela 6.

Tabela 6: Descrição dos componentes do corpo de um arquivo VCF

Informações presentes no corpo do	Descrição das abreviações apresentadas		
arquivo			
CHROM e POS	contém as coordenadas genômicas nas quais as		
ID	campo opcional, cuio conteúdo se baseia nas		
	informações contidas em um banco de dados de referência como o dbSNP		
	identificam o alelo referência e alternativo(s)		
REF e ALT	observado(s) na(s) amostra(s). No caso de		
	inserções, o alelo ALT fornece a sequência		
	inserida, bem como a base anterior à inserção.		
	Para deleções, o alelo ALI e representado pela		
OTAL	base anterior a deleção		
QUAL	(REF/ALT) exista de fato neste local indicado. No entanto, por ser este valor dependente da quantidade de dados analisada ele não representa uma medida útil para avaliação da		
	qualidade de determinação de uma variante		
FILTER	dados analisados. Se a variante nassou pelos		
	filtros aplicados o valor indicado é PASS. Caso		
	nenhum filtro tenha sido aplicado será		
	indicado "."		
INFO	informações descrevendo cada variante		
DB	associado ao dbSNP		
H3	associado ao HapMap3		
VALIDATED	validado por experimentos contínuos		
NA	número total de alelos nos genótipos		
	identificados		
AC	contagem alélica dentre os genótipos do alelo		
	ALT, na mesma ordem em que listados		
SVTYPE	tipo de variante estrutural (DEL para deleção,		
	DUP para duplicação, INV para inversão)		
END	posição final ocupada pela variante		
IMPRECISE	indica que a posição ocupada pela variante não		
	e precisa		
CIPOS/CIEND	intervalo de confiança em torno da posição e		
	posição final para variantes localizadas com		
FORMAT	formação a ordem das informaçãos contidas nas		
FORMAI	colunas seguintes (amostras).		
GT	indica o genótino da amostra em cada sítio Em		
	diploides, são indicados os dois alelos de cada		
	indivíduo, sendo "0" indicativo do alelo		
	referência e "1" do alelo alternativo. O		
	separador (/) indica que os genótipos não estão		
	faseados e () indica genótipos em fase		
AD	representa o número de reads que dá suporte a		
	cada um dos alelos (alternativo e/ou		
	referência) reportados		
DP	indica o número total de <i>reads</i> que representam		
	o genótipo reportado		

GQ	indica qual a confiança (escala Phred) da determinação de genótipo realizada Valores
	muito baixos são indicativos de pouca
	confiança no genótipo identificado
PL	é a probabilidade normalizada dos genótipos
	possíveis ou a probabilidade de o genótipo não
	estar correto. Assim, quanto menor o valor
	apresentado aqui, melhor
HP	identificadores de haplótipos eventualmente
	determinados pelo ReadBackedPhasing
PQ	qualidade de faseamento de haplótipos
	eventualmente determinados pelo
	ReadBackedPhasing

1.3.2. Remoção de genótipos de baixa qualidade

Alguns dos genótipos considerados pelo *HaplotypeCaller* e *UnifiedGenotyper* podem não ser verdadeiros, principalmente quando a genotipagem se refere a um segmento com baixa cobertura de sequenciamento ou quando, no caso de genótipos heterozigotos, um dos alelos é pouco representado.

Visando diminuir a taxa de falso-positivos, genótipos de baixa cobertura são interrogados, tomando como *input* o arquivo VCF obtido pelo *HaplotypeCaller* e *UnifiedGenotyper*.

Neste processo, o arquivo VCF é tratado pela ferramenta VCFx (versão 1.0) (Castelli et al., 2015), disponível em http://www.castelli-lab.net/apps/apps_vcfx.php, que interroga alelos segundo as regras:

- Para genótipos homozigotos inferidos em segmentos cobertos por oito ou menos *reads*, é introduzido um alelo interrogado (vcfx *alpha* = 8). A definição final desse tipo de genótipo (homozigoto ou heterozigoto) será inferida posteriormente.
- Genótipos heterozigotos em que um dos alelos está altamente subrepresentado (proporção de *reads* igual ou menor que 1%), são considerados homozigotos para o alelo mais representado (vcfx *beta* = 0.01). Esse procedimento minimiza a influência de *reads* mapeadas incorretamente e da alta taxa de erros de sequenciamento geralmente introduzidos em dados de NGS. Essa correção só pode ser aplicada em situações de alta cobertura (100 ou mais *reads*).
- Para genótipos heterozigotos em que um dos alelos está relativamente subrepresentado (proporção de *reads* entre 1% e 20%), um alelo interrogado é

introduzido representando o alelo subrepresentado (vcfx delta = 0.20). A definição final desse tipo de genótipo (homozigoto ou heterozigoto) será inferida posteriormente.

 Para genótipos heterozigotos inferidos com no máximo cinco *reads* em que um dos alelos está subrepresentado (Proporção de *reads* ente 20% e 40%), um alelo interrogado é inserido representando o alelo subrepresentado. No entanto, quando os dois alelos têm a mesma representatividade, ambos são considerados interrogados (vcfx *gamma* = 0.50). A definição final desse tipo de genótipo (homozigoto ou heterozigoto) será inferida posteriormente.

A linha de comando utilizada no VCFx é indicada a seguir: vcfx checkpl input=input.vcf output=output.vcf

Após a análise pelo VCFx, se forem realizadas análises por ambas as ferramentas (*HaplotypeCaller* e *UnifiedGenotyper*) as variantes que apresentaram genótipos distintos nos dois arquivos VCF obtidos através do *HaplotypeCaller* e *UnifiedGenotyper*, são analisadas manualmente para (a) verificar qual dos dois algoritmos fez a identificação correta do genótipo e (b) nos casos de genótipos interrogados por somente um dos algoritmos, para resolver variantes que são interrogadas por conta do número de *reads* consideradas pelos algoritmos, principalmente no caso do *HaplotypeCaller* que, devido aos filtros extremamente rigorosos, acaba descartando um número considerável de *reads* de suas análises. Para isso, é preciso analisar os arquivos BAM das amostras nos pontos interrogados. Como descrito anteriormente, o programa IGV pode ser utilizado para a visualização, avaliação de qualidade e contagem das *reads* alternativas que cobrem determinada posição. Considerando somente as bases chamadas com qualidade mínima Q30 em *reads* mapeadas com qualidade mínima Q30, as regras empregadas pelo VCFx são novamente aplicadas, agora manualmente para determinar se essas variantes continuariam interrogadas ou se seriam resolvidas.

Após essa etapa de maturação do arquivo, as informações contidas nos dois arquivos são unidas em um único arquivo VCF (VCF consenso) que é utilizado nas análises posteriores.

Após a análise pelo VCFx, se apenas a ferramenta *HaplotypeCaller* com o parâmetro --emitRefConfidence GVCF foi utilizada, os pontos de variação interrogados das amostras devem ser visualizados no programa IGV. É preciso analisar os arquivos BAMs destas amostras para avaliar a qualidade e contagem de *reads* alternativas que cobrem determinada posição (interrogada pelo programa VCFx). Isto porque o *HaplotypeCaller* aplica filtros extremamente rigorosos, descartando um número considerável de *reads* para suas análises. Portanto, analisar ponto a ponto os arquivos BAM muitas vezes pode auxiliar na compreensão do porquê as variantes foram interrogadas pelo programa VCFx.

Após a maturação manual o arquivo VCF é atualizado manualmente nas posições onde houveram alelos interrogados reavaliados e resolvidos.

2. REFERÊNCIAS

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Castelli, E. C., Mendes-Junior, C. T., Sabbagh, A., Porto, I. O. P., Garcia, A., Ramalho, J., ... Donadi, E. A. (2015). HLA-E coding and 3' untranslated region variability

determined by next-generation sequencing in two West-African population samples.HumanImmunology,76(12),945–953.https://doi.org/10.1016/j.humimm.2015.06.016

- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using nextgeneration DNA sequencing data. *Nature Genetics*, 43(5), 491–8. https://doi.org/10.1038/ng.806
- Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research*, 8(3), 175–185. https://doi.org/10.1101/gr.8.3.175
- Ferragina, P., Manzini, G., Veli, M., & Navarro, G. (2004). An Alphabet-Friendly FM-Index, 150–160.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R.

(2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G.,
 & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26. https://doi.org/10.1038/nbt.1754
- Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. https://doi.org/10.1093/bib/bbs017
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In *Current Protocols in Bioinformatics* (Vol. 11, p. 11.10.1-11.10.33). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471250953.bi1110s43

Anexo D - Resultado do Sequenciamento de Nova Geração: Pré e Pós filtros de qualidade

Os resultados apresentados aqui derivam da Dissertação de Mestrado intitulada "Avaliação da variabilidade genética do gene MITF e suas associações com fenótipos de pigmentação em amostra da população brasileira" (Marcorin, 2017).

O sequenciamento de nova geração totalizou cinco ensaios para as 340 amostras, sendo que algumas amostras que apresentaram baixa cobertura de reads quando sequenciadas pela primeira vez, foram re-sequenciadas.

É comum no ensaio de sequenciamento *MiSeq* da *Illumina* que haja uma queda na qualidade das bases incorporadas nos últimos ciclos, principalmente em amplicons maiores que o total de ciclos de sequenciamento (301) (Yang *et al.*, 2013; Erlich *et al.*, 2008). Porém, as reads foram processadas a fim de remover os adaptadores presentes em suas extremidades. Além disso, o software cutadapt remove integralmente as reads que não foram trimadas (que não possuíam identificação, como adaptadores e índices). Como mostrado na Tabela 1, a porção restante das reads possui uma qualidade média mais alta que das bases das reads nos dados brutos, e a quantidade de reads descartadas corresponde a menos de 5% do total.

Tabela 1 - Distribuição de *reads* de acordo com a qualidade média das bases que as compõem, pré e pós trimagem dos adaptadores *paired-end* da Illumina. Os dados representando cada sequenciamento são, respectivamente, das amostras PIG0001, PIG0100, PIG0261, PIG0307 e PIG0341, as quais foram tomadas aleatoriamente. Dados obtidos com FastQC (Andrews, 2010).

			PHRED score	
		< 20	20 < x < 30	≥ 30
1° sequenciamento pré trimagem		26.003 (3,4%)	407.581 (54,7%)	311.031 (41,7%)
n=48 pós trimagem		5.600 (0,7%)	49.156 (6,8%)	658.521 (92,3%)
2° sequenciamento pré trimager		12.491 (3,9%)	187.314 (58,6%)	119.417 (37,4%)
n=96	pós trimagem	6.656 (2,1%)	16.611 (5,3%)	288.331 (92,5%)
3° sequenciamento	pré trimagem	1.849 (0,9%)	64.931 (31,9%)	136.377 (67,1%)
n=87	pós trimagem	1.483 (0,7%)	8.047 (4,0%)	190.555 (95,2%)
4° sequenciamento	pré trimagem	2.439 (0,7%)	163.240 (52,1%)	147.425 (47,0%)
n=80	pós trimagem	1.000 (0,3%)	16.877 (5,5%)	286.369 (94,1%)
5° sequenciamento	pré trimagem	4.628 (1,1%)	154.957 (38,1%)	246.567 (60,7%)
n=61	pós trimagem	1.723 (0,4%)	27.111 (6,7%)	372.679 (92,8%)

Após a trimagem, as reads foram mapeadas a um genoma humano de referência (hg19). Como mostrado na Tabela 2, os cinco ensaios tiveram em torno de 89% das reads mapeadas com qualidade mínima Q30 (probabilidade de erro de 1 em 1000).

Tabela 2 - Qualidade de mapeamento ao genoma de referência. Os dados representando cada sequenciamento (lote) são, respectivamente, das amostras PIG0001, PIG0100, PIG0261, PIG0307 e PIG0341, as quais foram tomadas aleatoriamente. Dados obtidos com SAMStat (Lassmann *et al.*, 2011).

PHRED score	1° lote (n=48)	2° lote (n=96)	3° lote (n=87)	4° lote (n=80)	5° lote (n=61)
>= 30	1.270.414 (88,8%)	546.220 (87,5%)	356.713 (88,0%)	569.837 (90,7%)	745.211 (91,9%)
20 < x < 30	5.721 (0,4%)	2.318 (0,4%)	1.498 (0,4%)	2.234 (0,4%)	3.615 (0,4%)
10 < x < 20	3.512 (0,2%)	1.441 (0,2%)	1.457 (0,4%)	955 (0,2%)	1.365 (0,2%)
3 < x < 10	3.167 (0,2%)	1.960 (0,3%)	1.775 (0,4%)	1.062 (0,2%)	1.437 (0,2%)
< 3	434 (0%)	197 (0%)	358 (0,1%)	296 (0%)	497 (0,1%)
Não mapeadas	147.241 (10,3%)	72.271 (11,6%)	43.615 (10,8%)	53.561 (8,5%)	58.992 (7,3%)
Total	1.430.489 (100%)	624.407 (100%)	405.416 (100%)	627.945 (100%)	811.117 00%)

MANUSCRITO

Requisito regimental da Pós-Graduação da Faculdade de Medicina de Ribeirão Preto – USP determina que seja encaminhado juntamente com a Tese ao menos um manuscrito derivado dos dados do trabalho. Assim, é apresentada a seguir a versão de um manuscrito que atende a exigência citada.

VARIANTES DOS GENES *OCA2* E *HERC2* ASSOCIADAS COM A C*OR* DOS OLHOS E ESTRUTURAS SECUNDÁRIAS PRESENTES NA ÍRIS EM AMOSTRA DA POPULAÇÃO BRASILEIRA

Guilherme Debortoli¹, Esteban J. Parra², Letícia Marcorin¹, Alison Luis Eburneo Pereira¹, Nádia Carolina de Aguiar Fracasso¹, Maria Luiza Guimarães de Oliveira¹, Juliana Doblas Massaro³, Eduardo Antônio Donadi³, Aguinaldo Luíz Simões¹, Erick da Cruz Castelli⁴, Celso Teixeira Mendes-Junior⁵.

¹ Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto - SP, Brasil.

² Department of Anthropology, University of Toronto Mississauga, Mississauga, Ontario L5L1C6, Canada

³ Departamento de Clínica Médica, Divisão de Imunologia Clínica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto - SP, Brasil.

⁴ Departamento de Patologia, Faculdade de Medicina, Universidade Estadual Paulista, 18618-687, Botucatu - SP, Brasil.

⁵ Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto - SP, Brasil.

* Corresponding author. Tel.: +55-16-3315-0417.

E-mail address: ctmendes@ffclrp.usp.br (C. T. Mendes-Junior)

Conflitos de interesse:

Os autores declaram não haver conflitos de interesse.
RESUMO

A pigmentação humana é um traço fenotípico complexo, com mais de 170 genes associados a via de biossíntese da melanogênese. Dentre estes, os genes OCA2 e HERC2 são considerados os principais genes associados à variação normal da pigmentação dos olhos em humanos. Foi realizado no presente trabalho o sequenciamento de nova geração das regiões regulatórias e codificadoras destes dois genes, associando polimorfismos genéticos com a cor dos olhos através de medidas qualitativas e quantitativas por meio de fotografias, e também quanto a presença de estruturas secundárias na íris em 340 indivíduos de uma população miscigenada do estado de São Paulo-SP, Brasil. Onze polimorfismos foram associados a um ou mais fenótipos de cor dos olhos quando avaliados em variáveis categóricas, quantitativas e estruturas secundárias presentes na íris. O polimorfismo rs12913832 do gene HERC2 foi o que mais mostrou-se associado em diferentes fenótipos, tendo o seu alelo "G" associado a olhos azuis/verdes. Adicionalmente, este polimorfismo foi pela primeira vez, até onde sabemos, associado com estruturas secundárias presente na íris, como nódulos de Wolfflin, sulcos de contração e pontos de pigmentação na íris (p = $1,0x10^{-4}, p = 5,0x10^{-4}, e p = 0,0013$, respectivamente). Outros polimorfismos também foram associados a cor dos olhos. O polimorfismo rs58358300 do gene HERC2 foi associado a pigmentação da esclera ($p = 1,0x10^{-4}$), até então ainda não reportado. Neste trabalho, evidenciamos a associação de polimorfismos ainda não reportados na literatura com algumas características de pigmentação, o que reforça a necessidade de estudos em populações altamente miscigenadas como a brasileira, a qual apresenta condições propícias para a identificação de novos sítios de variação que possam ter um efeito funcional em genes de pigmentação.

Palavras-chave: *HERC2*; *OCA2*; pigmentação; cor dos olhos; população brasileira; polimorfismos.

INTRODUÇÃO

A pigmentação dos olhos está entre os exemplos mais visíveis da variação fenotípica humana ¹. A pigmentação humana é atribuída ao número, tipo e distribuição celular dos melanossomos, que são compartimentos subcelulares produzidos pelos melanócitos, e que sintetizam e armazenam a melanina (biopolímero que atua na absorção de luz).

O principal fator que explica a diversidade de pigmentação dos olhos é o tipo e a quantidade de melanina presente na íris. Estudos indicam que olhos claros possuem uma menor concentração de melanina quando comparados a olhos com colorações mais escuras ^{2,3}. Apesar da melanina contribuir majoritariamente na explicação da pigmentação da íris, outros fatores que podem contribuir na diversidade de cores de olhos são a distribuição dos melanossomos na íris e a presença ou ausência de estruturas secundárias na íris, que podem ser definidas como diferentes texturas presentes na superfície da mesma. Estas

Manuscrito 182

estruturas, assim como a cor da íris, possuem uma diversidade quanto a sua distribuição e quantidade.

De acordo com Larson e Pedersen (2004), a cor dos olhos é uma característica fenotípica com cerca de 98% de herdabilidade, o que significa que a maior parte da diversidade fenotípica existente pode ser explicada por variações genéticas.

Diferentes grupos de pesquisa apontam que a grande maioria da diversidade de pigmentação dos olhos é atribuída a um único *locus* localizado no íntron 86 do gene HERC2 ⁴⁻⁶. Este polimorfismo conhecido como rs12913832A>G é tido como o principal marcador capaz de explicar majoritariamente a variação de cor existente entre olhos azuis e castanhos. O alelo "A" possui uma frequência elevada em populações com ancestralidade diferentes da europeia, enquanto que o alelo "G" é frequentemente encontrado em populações europeias. Tradicionalmente, acredita-se que este polimorfismo atue sob um modelo de herança dominante/recessiva, de forma que, indivíduos homozigotos para o alelo "A" e heterozigotos "AG" apresentam olhos castanhos, enquanto que indivíduos homozigotos para o alelo "G" apresentam olhos azuis.

Entre as estruturas presentes na superfície da íris, cinco podem ser citadas: (1) as criptas de Fuchs, (2) os nódulos de Wolfflin, (3) pontos de pigmentação na íris, (4) sulcos de contração e (5) melanose conjuntival (também conhecida como pigmentação na esclera). Criptas de Fuchs são lacunas em forma de diamante que surgem durante o desenvolvimento do olho por reabsorção da membrana pupilar 7. Nódulos de Wolfflin consistem em pequenos agrupamentos de colágeno atrofiados que se acumulam na parte mais externa da zona ciliar. Sulcos de contração representam anéis descontínuos que se estendem ao redor da parte mais externa da zona ciliar, e pontos de pigmentação na íris e pigmentação na esclera são pequenas regiões de hiper-pigmentação devido ao acúmulo de melanina. Até o momento, três grupos de pesquisa têm voltado seus esforços para identificar variantes genéticas que possam estar ligadas ao aparecimento destas estruturas ⁸⁻¹⁰, sendo que muitas das variantes associadas aparentemente são diferentes daquelas associadas a pigmentação dos olhos ⁹. Alguns polimorfismos associados a estruturas secundárias podem ser citados, como o SEMA3A (do inglês, Semaphorin 3A) e seu polimorfismo rs10235789 fortemente associado a criptas de Fuchs, o qual explica 1,5% da variabilidade observada nesta característica; TRAF3IP1 (do inglês, TRAF3 Interacting Protein 1) e o polimorfismo rs3739070, cuja associação com sulcos de contração representa 1,7% da variabilidade para esta característica 8,9; HERC1 (do inglês, HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase Family Member 1) rs11630290, associado com pontos de pigmentação na íris ⁹; DSCR9 (do inglês, Down Syndrome Critical Region 9) rs7277820, o qual supostamente pode modular a cor dos olhos controlando a presença e a quantidade de nódulos de Wolfflin na íris ⁹.

Ademais, é importante ressaltar que estudos envolvendo tanto a variação normal da cor dos olhos como a caracterização genética de polimorfismos associados ao aparecimento de estruturas secundárias presentes na íris não foram extensivamente estudas em populações altamente miscigenadas, principalmente aquelas pertencentes ao continente sul-americano.

MATERIAL E MÉTODOS

Coleta de amostra e aspectos éticos

Foram coletadas amostras 340 indivíduos voluntários amostrados junto ao Hemocentro de Ribeirão Preto – SP. O presente estudo foi aprovado em seus aspectos éticos pelo Comitê de Ética em Pesquisa da FFCLRP-USP, de acordo com o Processo CEP-FFCLRP CAAE n.º 25696413.7.0000.5407. Cada indivíduo participante assinou um Termo de Consentimento Livre e Esclarecido (TCLE) permitindo o uso de seu material biológico (10 mL de sangue periférico, mantido em tubo Vacutainer® com anticoagulante EDTA) para estudo. Questionários foram aplicados a cada um dos participantes da pesquisa para obtenção de informações a respeito dos fenótipos de pigmentação e ancestralidade. Os participantes, composto por indivíduos de ambos os sexos, com idade variando de 18 e 70 anos, foram classificados em grupos de acordo com a cor dos olhos. As classificações foram realizadas por dois membros do Laboratórios de Pesquisas Forenses e Genômicas (LPFG) de forma independente para que não houvesse influência sobre as respostas, seguida por diálogo para alcançar uma definição consensual. Desta forma, a classificação fenotípica foi estabelecida pela heteroclassificação.

Caracterização dos fenótipos qualitativos e quantitativos

A cor dos olhos caracterizada e dividida de acordo com categorias criadas em nosso laboratório. Estas classificações foram elaboradas a partir de diferentes classificações, uma vez que não há um consenso quanto a classificação destes fenótipos na literatura. Portanto, as categorias para classificação da cor dos olhos foram (1) azul, (2) verde, (3) mel, (4) castanho claro e (5) castanho escuro. Imagens fotográficas de ambas as íris foram coletadas de modo que estas não permitissem a identificação do doador. Para tal, foi utilizada uma máquina fotográfica Canon Powershot SX510 HS com abertura, shutterspeed e ISO fixados para f = 3,4, 1/10 e 100, respectivamente, tendo sido utilizada uma distância média de 7 cm entre a câmera e o indivíduo para obtenção das imagens fotográficas. As imagens foram utilizadas para compor um banco de dados, a fim de tornar mais objetiva a classificação de cor dos olhos e identificação de estruturas secundárias presentes na íris. Sempre que possível, a coleta foi realizada no mesmo local, de modo a tentar minimizar efeitos externos que pudessem interferir na qualidade das fotos. O controle e cuidado com o local e a iluminação foram primordialmente importantes para as fotografias das íris do indivíduos coletados. Isto porque as imagens fotográficas das íris foram utilizadas para realizar análises quantitativas da cor do olho. A pigmentação da íris foi avaliada quantitativamente usando um programa personalizado, projetado para extrair digitalmente tanto a pupila quanto a esclera e reter apenas a íris. O programa recorta uma fatia da íris em um ângulo de 60° da região ciliar e da região pupilar oposta ao ponto lacrimal do olho, e realiza uma mensuração da cor média dos olhos nas coordenadas do CIELAB. Informações detalhadas sobre este programa foram descritas em⁸. Este programa ainda permite anotar a presença ou ausência de estruturas secundárias presentes na íris. Estas análises foram realizadas em 225 dos 340 indivíduos que compuseram a amostra, devido a algumas fotografias não terem atingido um limiar mínimo de qualidade para a análise. As íris foram analisadas por um único pesquisador (G.D.).

Sequenciamento das regiões dos genes OCA2 e HERC2

As regiões regulatórias e codificantes dos genes OCA2 e HERC2, além de outras regiões genômicas de interesse, como regiões que incluem um conjunto de 34 AIMs (do inglês, *Ancestry Informative Markers*) que compõe o sistema 34-*plex* do consórcio SNP*for*ID foi incluído no ensaio ¹¹, foram selecionados através de um ensaio *in silico* usando a ferramenta SureDesign (*Agilent Technologies, Inc.*). Para este ensaio o genoma de referência utilizado foi o hg19/GRCh37, que resultou em um conjunto de sondas capazes de capturar as regiões de interesse (~22.480 pb). É importante enfatizar que devido ao desenho das sondas, partes limites dos íntrons que flanqueiam os éxons alvos também foram capturadas. Adicionalmente, além destas regiões gênicas analisadas no ensaio, um conjunto de 34 AIMs que compõe o sistema 34-*plex* do consórcio SNP*for*ID foi incluído no ensaio ¹¹, com a finalidade de controlar possíveis associações espúrias nos testes de associação genética.

O DNA foi extraído usando o protocolo *Salting-Out*¹². A qualidade das amostras foi avaliada em termos de integridade, pureza e concentração por eletroforese em gel de agarose, espectrofotometria NanoDrop (*Thermo Fisher Scientific, Inc.*) e ensaio fluorimétrico Qubit® dsDNA BR (*Thermo Fisher Scientific, Inc.*), respectivamente. Todas as amostras foram normalizadas para uma concentração de 5 ng/µL para alcançar uma concentração ideal de DNA para a preparação das bibliotecas de sequenciamento.

As bibliotecas de DNA foram preparadas utilizando o protocolo *HaloPlex Target Enrichment System* (Agilent Technologies, Inc., Santa Clara, CA). Resumidamente, 200 ng de cada amostra de DNA foram digeridos por oito pares diferentes de enzimas de restrição. Os fragmentos produzidos foram capturados utilizando sondas biotiniladas HaloPlex e os índices de identificação de cada amostra foram introduzidos nesta etapa. Os fragmentos capturados foram então amplificados por PCR no termociclador SureCycler 8800 (*Agilent Technologies, Inc.*), de acordo com o protocolo do fabricante. Após a amplificação, foram utilizadas esferas magnéticas AMPure XP (Beckman Coulter, Beverly, MA) num passo final de purificação. As bibliotecas de DNA foram ressuspensas em tampão Tris-HCl (pH 8,0) e armazenadas a -20 para o posterior sequenciamento. A avaliação da qualidade das bibliotecas foi verificada pelo 2100 Bioanalyzer (*Agilent Technologies, Inc.*), que, juntamente com o fluorímetro Qubit® 2.0 (*Thermo Fisher Scientific, Inc.*), também foi empregado para a quantificação de cada biblioteca. Um *pool* de bibliotecas, consistindo em até 96 amostras (4 nmol/L) foi então diluído para 16 pM e inserido como entrada para sequenciamento usando o kit MiSeq Reagent V3 (600 ciclos), no sequenciador *MiSeq (Illumina, Inc.*).

Processamento dos dados pós-sequenciamento

A trimagem foi realizada buscando pela sequência dos adaptadores *paired-end* utilizados no sequenciamento, com o uso do cutadapt¹³. Após processamento das *reads*, estas foram mapeadas contra um genoma de referência utilizando o algoritmo BWA-MEM¹⁴. O genoma de referência utilizado foi o *Homo sapiens*, UCSC, hg 19, disponibilizado pela Illumina

(Illumina, Inc). Os arquivos BAM (Binary Alignment/Map), foram ordenados e indexados, com o emprego do programa SAMtools (http://samtools.sourceforge.net). A determinação dos genótipos foi feita utilizando o programa GATK 3.7 (do inglês, *Genome Analysis Toolkit*) no modo GVCF, aplicado para cada amostra separadamente. Após a geração dos arquivos separados para cada amostra, estes foram concatenados com o algoritmo GenotypeGVCFs do GATK, e assim produzindo um VCF final para as análises posteriores ¹⁵⁻¹⁷. Estes mesmos procedimento foram utilizados para gerar o arquivo VCF final incluídos no sistema 34-*plex* do consórcio SNP*for*ID ¹¹.

Análises estatísticas

Marcadores informativos de ancestralidade

Os genótipos do SNP*for*ID 34-*plex* foram utilizados para estimar as contribuições de ancestralidade nas amostras do presente estudo. A ancestralidade foi estimada pelo programa STRUCTURE (Pritchard *et al.*, 2000), definindo K = 3 e considerando a população do projeto *1000Genomes* como referência. No total, 1.412 indivíduos do projeto *1000Genomes* (fase 3) foram usados para representar as populações parentais em nossa análise de ancestralidade: 404 europeus (TSI, FIN, GBR, IBS), 504 africanos (YRI, LWK, GWD, MSL, ESN) e 504 do leste asiático (CHB, JPT, CHS, CDX, KHV). Como o projeto *1000Genomes* não inclui um grupo autóctone ameríndio que poderia representar adequadamente a ancestralidade dos nativos americanos em nossa análise, usamos o leste asiático como *proxy* para a população parental nativa americana. Usando o modelo de mistura, tanto a população quanto os as estimativas de ancestralidade individuais de cada população contribuinte (AFR, EAS e EUR) foram identificadas.

Análises de associação genética

Salvo indicação contrária, todas as análises estatísticas foram conduzidas no programa PLINK 1.9¹⁸. Foram calculadas as frequências alélicas, genotípicas, hetorozigosidade e aderência ao equilíbrio de Hardy-Weinberg. De modo a testar as associações com as diferentes variáveis qualitativas para a cor dos olhos, e variáveis quantitativas para o espaço de cor CIELAB, bem como estruturas secundárias presentes na íris, análises de regressão logística e linear ajustadas para ancestralidade, e assim controlando para possíveis efeitos da estratificação populacional, foram realizadas. O ajuste para a ancestralidade foi realizado adicionando-se como covariáveis no modelo de regressão as estimativas de mistura individual fornecidas após as análises pelo programa STRUCTURE. Um modelo aditivo de herança para o alelo de menor frequência do polimorfismo analisado. Após esta primeira análise de regressão, o polimorfismo que se mostrou mais associado, ou seja, com o menor valor de p, foi condicionado no modelo de regressão, sendo adicionado como uma covariável para testar se os outros polimorfismos que também foram associados exercem uma associação independente do polimorfismo condicionado. O teste de Kruskal-Wallis foi utilizado para comparar as medianas do valor de L*, a* e b* entre os grupos heteroclassificados de acordo com a cor dos olhos. O teste de Wilcoxon foi empregado para comparar as medianas das estimativas de ancestralidade europeia, africana e ameríndia dentro dos grupos que apresentaram ou não as diferentes estruturas secundárias na íris, Os testes foram conduzido no ambiente estatístico R.

RESULTADOS

A amostra populacional de estudo constituiu em 340 indivíduos, sendo que destes, 173 eram homens e 167 mulheres. Entretanto, oito amostras foram excluídas de todas as análises devido à alta taxa de genótipos não determinados, o que prejudicaria as análises genéticas subsequentes. Portanto, 332 amostras, sendo 168 homens e 164 mulheres, compuseram a amostragem final. A distribuição das características da cor dos olhos nos 332 indivíduos amostrados pode ser observado na Figura 1 e para os 225 indivíduos analisados para estruturas secundárias presentes na íris, a distribuição está representada na Figura 2.

A Tabela 1 mostra os valores de atribuição das populações a cada um dos agrupamentos gerados pelo STRUCTURE. As populações africanas, europeias e do leste asiático do *1000Genomes* foram agrupadas majoritariamente nos *clusters* 3, 1 e 2 respectivamente. Observa-se que as amostras de Ribeirão Preto (RBP) possuem 71,4% de componente designado como europeu; 9% do componente asiático/ameríndio (representado pelas populações do leste asiático) e 19,6% de contribuição africana.

As análises quantitativas para obter valores contínuos que permitem a caracterização objetiva da cor do olhos e a identificação de estruturas secundárias presentes na íris foram realizadas em imagens fotográficas que estão representadas nas Figuras 3 e 4. As variáveis quantitativas para cor dos olhos do espaço de cores CIELAB quando plotadas em um gráfico de três dimensões, evidencia uma substancial dispersão dos valores de L*a*b* entre as diferentes íris analisadas no presente estudo Figura 5. Olhos azuis tendem a mostrar maiores valores de L*, valores de a* próximos a 0 e valores de b* negativos, enquanto que olhos castanhos tendem a menores valores de L*, e maiores valores de a* e b*. Olhos castanho-escuros possuem menores valores de L*a*b* comparados aos olhos castanhoclaros. Alguns indivíduos com olhos verdes obtiveram valores sobrepostos aos dos indivíduos com olhos azuis e indivíduos com olhos mel tiveram valores sobrepostos àqueles dos indivíduos com castanho-claros. Ao plotar a heteroclassificação realizada pela equipe do LPFG para cada um dos 225 indivíduos no mesmo plano tridimensional L*a*b*, observa-se que a heteroclassificação obteve elevada concordância com os valores dispostos no plano L*a*b*, com agrupamento de olhos azuis e verdes nos valores que vão de 0 a valores negativos de b*, e olhos castanhos sendo distribuídos em sua maioria nos valores de 0 a positivos de b* (Figura 6). Dentre as três dimensões CIELAB (L*, a* e b*), a variação da cor dos olhos foi mais perceptível e melhor explicada pela dimensão cromática b* (Figura 7). Quando b* foi comparado entre pares de grupos heteroclassificados para a cor dos olhos, observa-se que, exceto para olhos mel vs. castanho-claros (ns, p > 0.05), todos os outros grupos tiveram diferenças significativas quanto às suas medianas. A Figura 8 compara as estimativas de ancestralidade dos indivíduos que apresentaram ou não as diferentes estruturas secundárias presentes na íris. Exceto para pigmentação na esclera, a ancestralidade europeia foi maior entre os grupos que apresentaram ou não a característica analisada.

A determinação de variantes pela rotina HaplotypeCaller do GATK encontrou 749 variantes ao longo das regiões cobertas pelo ensaio, sendo que destas, 579 foram removidas após aplicação de um filtro para a remoção de variantes com MAF (do inglês, Minor Allele Frequency) inferiores a 0,01. Portanto, foram avaliadas 170 variantes ao longo dos genes OCA2 e HERC2. Destas 170, onze variantes foram associadas a um ou mais fenótipos de cor dos olhos quando avaliados em variáveis categóricas, quantitativas e estruturas secundárias presentes na íris. A distribuição das associações com cor de olhos por variáveis qualitativas pelas diferentes regiões gênicas está apresentada na Tabela 2, e o detalhamento das associações está apresentado na Tabelas 3. Duas variantes foram associadas para estruturas secundárias presentes na íris. O polimorfismo rs12913832A>G foi associado com três fenótipos de estruturas presentes na íris: (a) pontos de pigmentação na íris, (b) nódulos de Wolfflin e (c) sulcos de contração. O polimorfismo rs58358300A>C foi associado com pigmentação na esclera (Tabela 4). As análises de associação genética para os valores de L*a*b* extraídos por imagem fotográficas da íris, evidenciaram dois polimorfismos foram associados a dimensão L*, (b) um polimorfismo foi associado com a dimensão a* e (c) dois polimorfismos foram associados a dimensão b* (Tabela 5). O rs12913832A>G foi o mais associado entre as três dimensões, sendo que o genótipo "GG", relatado na literatura como associado a cor dos olhos azuis mostrou uma distribuição na coordenada b* do espaço de cores CIELAB (Figura 9), onde claramente há um agrupamento, ainda que disperso, destes indivíduos no eixo negativo que compreende à cor azul.

DISCUSSÃO

A diversidade presente na cor dos olhos tem sido objeto de estudo de diferentes grupos de pesquisa ao longo da última década, mesmo após a descoberta do polimorfismo rs12913832A>G do gene *HERC2*, descrito como principal marcador a explicar a variabilidade existente entre olhos azuis e castanhos ^{4,6}.

Entretanto, apenas três grupos de pesquisas distintos procuraram observar a variabilidade presente em outras características presentes na íris, como nódulos de Wolfflin, sulcos de contração, pontos de pigmentação na íris e na esclera. Algumas dessas características podem alterar a percepção de cor da íris (Liu *et al.* (2010) e Mackey *et al.* (2011)) e, ainda que as consequências funcionais de muitas destas características ainda permaneçam desconhecidas, estas podem ser de grande importância na prática forense, no que se refere à predição da pigmentação dos olhos, e também para a área biomédica, onde algumas destas estruturas secundárias tem sido reportadas como marcadores de risco para diferentes patologias, como por exemplo, melanoma uveal, no qual pigmentos na íris consistem em fatores de risco ^{19,20}. Ademais, faltam estudos em populações diferentes das europeias e asiáticas que possam contribuir para o melhor entendimento da distribuição e diversidade genética vinculadas a estas características.

Seguindo esta linha de raciocínio, a população brasileira pode contribuir grandemente na tentativa de identificar fatores genéticos envolvidos no aparecimento destas características, uma vez que, conforme dito no tópico anterior, é uma população majoritariamente trihíbrida.

A distribuição dos nódulos de Wofflin foi predominante em indivíduos com olhos claros (Figura 2). Estas estruturas circulares presentes na parte mais externa da zona ciliar são altamente associadas a cor dos olhos claros em comparação com olhos escuros, conforme observou Kim *et al.* (2012) em amostras do leste asiático, onde nenhum dos 112 indivíduos com olhos escuros apresentou nódulos de Wolfflin. Edwards *et al.* (2015) observou uma maior frequência de nódulos de Wolfflin em indivíduos com ancestralidade europeia (25,8%) quando comparado a indivíduos do leste asiático (0%) e sul asiáticos (0,8%). No presente estudo a ancestralidade média dos indivíduos que apresentaram nódulos de Wolfflin foi de 87,2% europeia, 3,8% ameríndia e 8% africana, corroborando os achados a despeito da presença desta característica em indivíduos com ancestralidade europeia. Ainda assim, é importante enfatizar que Falls (1970) sugere que esta característica pode não ser ausente em íris mais escuras, mas sim, dificilmente identificável por estar escurecida devido à elevada quantidade de melanina presente.

Os sulcos de contração foram mais presentes em indivíduos com olhos escuros (Figura 2). Nos estudos de Larson e Pedersen (2004) e Sidhartha *et al.* (2014) observou-se também que a extensão com que os sulcos de contração se distribuem ao longo da íris está associado a olhos escuros. No presente estudo, a ancestralidade média dos indivíduos que apresentaram sulcos de contração foi de 77% europeia, 8% ameríndia e 14% africana, corroborando os resultados obtidos por Quillen *et al.* (2011) que observou uma correlação positiva entre sulcos de contração e ancestralidade europeia em uma amostragem de Brasília (DF). Interessantemente, Quillen *et al.* (2011) associou sulcos de contração com olhos claros, contrastando com o observado em nosso estudo e na literatura [Larson e Pedersen (2004) e Sidhartha *et al.* (2014)].

A distribuição de pontos de pigmentação na íris foi levemente mais frequente em olhos claros (Figura 2). A ancestralidade média dos indivíduos que apresentaram pontos de pigmentação na íris foi de 81% europeia, 7% ameríndia e 11% africana. No estudo de Edwards *et al.* (2015), pontos de pigmentação na íris foram mais frequente em indivíduos com ancestralidade europeia quando comparado a indivíduos do leste e sul asiáticos. Uma vez que pontos de pigmentação são pequenas regiões delimitadas de acúmulo de pigmentos marrons/pretos, é mais provável observar esta característica em indivíduos com olhos claros e, desta forma, mais provável de observar em indivíduos com ancestralidade europeia, onde diferentes cores de olhos são mais frequentes.

A pigmentação da esclera, também chamada de melanose conjuntival, foi mais frequente em indivíduos com olhos escuros, principalmente castanho-escuros (Figura 2). Apesar de não ser ainda muito bem compreendida, esta característica parece estar associada a populações com olhos escuros, uma vez que a ancestralidade média dos indivíduos que apresentaram esta característica no presente trabalho foi de 45% europeia, 11% ameríndio e 43% africana. Edwards *et al.* (2015) evidenciou este mesmo padrão, ao observar uma maior frequência

desta característica em populações do sul e do leste asiático (populações com uma frequência elevada de indivíduos com olhos escuros), quando comparado a indivíduos com ancestralidade europeia.

È importante enfatizar que a ancestralidade europeia está associada a estas características, visto que em todas elas, exceto em melanose conjuntival, a ancestralidade europeia dos indivíduos que as possuem é maior que a média do presente estudo, 71,4%. Isto poderia indicar um viés causado pela identificação das estruturas, as quais seriam mais facilmente identificáveis em olhos claros. Porém, sulcos de contração contradizem esta possibilidade de viés, visto que ocorrem mais frequentemente em olhos escuros e, ainda assim, os indivíduos apresentando esta estrutura na íris possuem uma estimativa de ancestralidade europeia superior (77%).

Por muito tempo a cor do olho em humanos foi tratada como um simples caráter mendeliano ²¹. Tradicionalmente, esta característica sempre foi descrita como uma herança mendeliana dominante, com a cor "castanho" sendo dominante sobre a cor "azul", ainda que diferentes estudos tenham demonstrado inconsistências com esse simples modelo de herança dominante ^{22,23}.

Os genes OCA2 e HERC2 aqui analisados possuem uma extensa lista de estudos na literatura associando-os a variação normal da pigmentação humana, principalmente a variação na cor dos olhos ^{4,5,21,22,24-26}. Ainda assim, é importante enfatizar que os trabalhos que analisaram estes dois genes com uma abordagem mais ampla (e não apenas polimorfismos específicos) estão restritos a populações fora do continente sul-americano, principalmente populações europeias ²¹ e asiáticas ²⁷. Apesar da vasta maioria destes estudos de pigmentação estar restrita a estas populações, alguns estudos já se propuseram a analisar polimorfismos específicos destes dois genes em populações miscigenadas sul-americanas, incluindo amostragens da população brasileira ^{28,29}, salientando um trabalho recente de nosso grupo ³⁰.

Ademais, até onde sabemos, este é o primeiro trabalho no Brasil que se propôs a analisar a associação com a cor do olhos por meio de uma metodologia objetiva que busca avaliar quantitativamente a mesma. Como mencionado anteriormente, esta metodologia possui vantagens em relação aos tradicionais métodos subjetivos de categorização da cor do olho. A cor da íris é considerado um traço quantitativo inato. Embora as categorias discretas possam capturar diferenças extremas na cor da íris (por exemplo, azul *vs.* castanho), elas não são capazes de capturar toda a variação da cor dos olhos. Além disso, não há consenso formal sobre quantas categorias devem ser usadas, o que dificulta a confiabilidade de resultados dependentes de avaliações inter e intra-observador de métodos categóricos ³¹.

Dentre todos os polimorfismos que tiveram seus alelos associados, o principal foi o rs12913832A>G do gene *HERC2* (Tabela 3), que é considerado como o principal capaz de explicar a variação existente entre olhos azuis e castanhos. Visser *et al.* (2012) e Sturm *et al.* (2008) observaram que este polimorfismo influencia a regulação do gene OCA2. A associação com a cor dos olhos para este polimorfismo corrobora os achados na literatura, onde o alelo "G" está associado a maiores chances de possuir olhos claros quando comparado ao alelo "A" (Tabela 3). O mesmo ocorre quando analisado sob a perspectiva de

uma variável quantitativa, aqui evidenciada pela distribuição dos indivíduos homozigotos "GG" na coordenada b* do espaço de cores CIELAB (Figura 9), onde claramente há um agrupamento, ainda que disperso, destes indivíduos no eixo negativo que compreende à cor azul. Em relação as outras duas coordenadas (L* e a*), o alelo "G" foi associado a valores positivos do coeficiente beta para a coordenada L*, que representa o brilho/claridade (maiores valores de L* indicam um maior brilho/claridade) [*beta* = 3,745 (IC 95% = 2,18 - 5,30) $p = 4,67 \times 10^{-6}$], de acordo com o esperado para olhos claros (associados ao alelo "G") (Tabela 5). Para a variável a*, o alelo "A" foi associado a valores positivos do coeficiente beta at a cordo com o esperado para olhos claros do coeficiente beta a cordenada a cor vermelha, também estando de acordo com o esperado para o alelo "A", associado a olhos castanho-claros e escuros.

A análise de regressão linear indicou que o total de variância que pôde ser explicada pela variável independente (rs12913832A>G), foi de 36% (R² ajustado = 0,36), 14% (R² ajustado = 0,14) e 26% (R² ajustado = 0,26) para as variáveis dependentes L*, a* e b*, respectivamente (dados não mostrados). Liu *et al.* (2010) e Andersen *et al.* (2013), também encontraram uma variância majoritariamente explicada por este polimorfismo em populações europeias (44%, R² ajustado = 0,44 e 68%, R² ajustado = 0,68, respectivamente). Ainda que ambos os estudos tenham utilizado metodologias diferentes da aqui apresentada para avaliar quantitativamente a cor dos olhos através de fotografias, isto é, medidas diferentes do espaço de cor CIELAB, as duas metodologias, mais a aqui apresentada, tiveram em comum o fato de apontarem para o rs12913832A>G como o polimorfismo capaz de explicar grande parte da variação existente, corroborando a importância deste polimorfismo na determinação da cor dos olhos. Além do polimorfismo rs12913832A>G outros também foram associados.

O polimorfismo rs1800407C>T é caracterizado pela mudança do aminoácido arginina pela glutamina. Ainda que este apresente uma baixa frequência em todas as três principais populações que compuseram a população brasileira, ele possui um efeito relativamente alto na determinação da cor do olho. De acordo com Walsh *et al.* (2012), este polimorfismo é tido como o segundo com o maior fator preditivo para as diferentes colorações do olhos em populações europeias, estando associado com a penetrância de olhos verdes e mel quando na presença do genótipo heterozigoto AG ou homozigoto AA no sítio rs12913832A>G do gene HERC2 ^{32,33}. Em nosso estudo, o polimorfismo rs1800407C>T mostrou-se associado tanto a olhos verdes quanto a olhos mel (Tabela 3).

O polimorfismo rs1800401G>A é caracterizado pela mudança do aminoácido arginina para triptofano. Alguns estudos não encontraram associação significativa ^{21,32}, enquanto que apenas um estudo encontrou associação com olhos castanho quando comparado a olhos não-castanhos ³⁴. Em nosso estudo, o alelo "A" foi associado como um fator protetor para o aparecimento de olhos castanho-claro *vs.* castanho-escuro [*OR* = 0,377 (IC 95% = 0,15 - 0,93) *p* = 0,0356], e a um fator de risco para olhos castanho-escuro *vs.* olhos não-castanho-escuro [*OR* = 2,921 (IC 95% = 1,33 - 6,41) *p* = 0,0075] (Tabela 3). É interessante observar polimorfismos que possam distinguir entre cores mais homogêneas, como, por exemplo, diferentes tons de castanho, uma vez que a grande maioria dos polimorfismos descritos na

literatura tem sido associada a distinção entre fenótipos extremos como cor de olho azul $\upsilon s.$ olho castanho.

O polimorfismo rs74005645C>T apresenta uma substituição de um aminoácido valina por metionina e o alelo "T" foi associado com o aparecimento de olhos castanho-escuros quando comparado a olhos não castanho-escuros [OR = 3,720 (IC 95% = 1,40 - 9,86) p = 0,0083) (Tabela 3). Quando observamos a frequência deste alelo nas populações do 1000Genomes, ele é majoritariamente encontrado em populações africanas, com uma frequência de aproximadamente 53%. Não observa-se este alelo em populações europeias e apenas 0,5% em populações do leste asiático. No presente estudo, o alelo apresentou uma frequência de 12% (Tabela 3). De acordo com o relatório gerado pelo *SNPnexus*, a mutação é considerada benigna, e sua conservação não é alta entre diferentes espécies, o que está de acordo com a classificação benigna da mutação.

O polimorfismo não-sinônimo rs201872292G>A teve o alelo "A" foi associado como um fator protetor para o aparecimento de olhos castanho-claro vs. castanho-escuro |OR| = 0,377 (IC 95% = 0,15 - 0,93) p = 0,0356), a um fator de risco para olhos castanho-escuro vs. olhos não-castanho-escuro [OR = 2,921 (IC 95% = 1,33 - 6,41) p = 0,0075] (Tabela 3), e a um aumento na variável quantitativa b* do espaço de cor CIELAB [beta = 3,950 (IC 95% = 1,61 -6,28) p = 0,0010] (Tabela 5). O alelo "A" tem uma substancial frequência em populações do leste asiático quando comparado as populações africanas e europeias (29%, 4% e 1,8%, respectivamente). Este alelo foi também mais frequente em populações latinas (32%) que compuseram o Exome Aggregation Consortium³⁵, que agregou além das populações que compõe o 1000Genomes, outras populações totalizando 60.706 indivíduos analisados. A frequência deste alelo no presente estudo foi de 8%, o que pode ter sido inflada pelos indivíduos com descendência asiática presentes na amostragem, e também pelo componente ameríndio presente na população brasileira. A mudança de aminoácido causada pela mutação do rs201872292A>G é a de um aminoácido treonina para metionina. O aminoácido ancestral (treonina) é conservado em primatas e em grande parte dos mamíferos placentários. Segundo a literatura, assume-se que sequências de proteínas observadas em diferentes organismos tenham sobrevivido a seleção natural, e por conseguinte, posições de aminoácidos que mostram-se conservados em diferentes espécies possivelmente possuem um papel funcional importante, com as substituições de aminoácidos observadas em posições conservadas possivelmente acarretando em efeitos prejudiciais nas funções dos genes 36,37.

O polimorfismo rs1800419A>G codifica um aminoácido alanina e o alelo "A" foi associado a maiores chances de possui olhos cor mel *vs.* castanho-escuros [OR = 2,976 (IC 95% = 1,17 - 7,54) p = 0,0215], e a menores chances de possuir olhos castanho-escuros *vs.* olhos não castanho-escuros [OR = 0,597 (IC 95% = 0,36 - 0,97) p = 0,0403] (Tabela 3). Estas duas associações se mostram consistentes, uma vez que o alelo "A" estaria associado a fenótipos mais claros. Ademais, Duffy *et al.* (2007) também reportou a associação do alelo "A" com olhos azuis e o alelo "G" com olhos não azuis em amostras de gêmeos da Austrália. Entretanto, o alelo "A" foi também associado no presente estudo a menores chances de aparecimento de cabelos loiros, ou seja, um típico fenótipo claro [OR = 0,5347 (IC 95% =

0,30 - 0,92) p = 0,0256] (dados não mostrados). Embora a associação com olhos e cabelos seja aparentemente contraditória, é preciso levar em consideração que a proteína OCA2 pode apresentar papel funcional diferente na pigmentação dos olhos e cabelos. A frequência alélica do alelo "A" é praticamente a mesma em europeus e leste asiático, seguidos de africanos (56,3%, 56,5% e 37,1%, respectivamente), e no presente estudo apresentou uma frequência de 49%.

O polimorfismo rs61756152G>A teve o alelo "A" associado com maiores chances de possuir olhos verdes em comparação a olhos castanho-escuros [OR = 4,497 (IC 95% = 1,33 - 15,21) p = 0,0155] (Tabela 3). A sua frequência nas populações do 1000Genomes foi de 5%, 9% e 29% aproximadamente nas populações europeias, africanas e do leste asiático, e de 10% no presente estudo (Tabela 3). É interessante observar a alta frequência do alelo associado a olhos verdes na população brasileira em populações asiáticas. Uma vez que olhos verdes não ocorrem naturalmente no leste da Ásia, é possível que interações epistáticas estejam relacionadas ao presente achado de associação independente. É importante enfatizar que não se pode descartar interações com polimorfismos de outros genes, o que não foi considerado no presente estudo. Por fim, o relatório gerado pelo SNPnexus apontou uma predição de possivelmente benigna para esta mutação sinônima, não acarretando maiores danos na função proteica.

O polimorfismo rs12910433A>G foi associado com maiores chances de aparecimento de olhos mel contra olhos castanho-claros [OR = 5,949 (IC 95% = 1,98 - 17,81) p = 0,0014] (Tabela 3). A frequência do alelo "G" no presente trabalho foi de 49,7%, e nas populações europeias, africanas e do leste asiático do *1000Genomes* foi de 24,7%, 90,6% e 84,6%, respectivamente (Tabela 4). Este polimorfismo foi reportado por Murray *et al.* (2015), como parte de um bloco haplotípico nas populações do leste asiático, em que os polimorfismos não sinônimos rs1800414T>C e o rs74653330C>T (não reportado no presente trabalho) são associados ao clareamento da pele em asiáticos. Apesar do rs12910433A>G ser encontrado em um alto desequilíbrio de ligação com o rs1800414T>C no leste asiático, eles não se encontram em DL no presente trabalho ($R^2 = 0,02$ e D' = 1,00), o que pode ser decorrente da miscigenação presente na população brasileira ou de outros aspectos demográficos inerentes a ela.

O polimorfismo rs6497271G>A teve o seu alelo "A" associado a uma diminuição do coeficiente beta nas análises de regressão independente para a variável quantitativa L* do espaço de cor CIELAB [*beta* = -5,498 (IC 95% = -8,35 - -2,65) $p = 1,96 \times 10^{-4}$] (Tabela 5), ou seja, uma associação com pouco brilho/claridade. A alta frequência deste em populações africanas do *1000Genomes* (27,7%) e baixa frequência em europeus (3,7%) e leste asiáticos (0,5%) (Tabela 4), condiz com a associação encontrada, uma vez que olhos com coloração escura, como por exemplo castanho-escuros, são mais frequentes em populações africanas e olhos castanho-escuros possuem menos brilho/claridade quando comparado a outras cores, conforme evidenciado por Edwards *et al.* (2015).

O polimorfismo rs58358300A>C teve o alelo derivado "C" fortemente associado ao aparecimento de pigmentação na esclera da íris [OR = 9,247 (IC 95% = 2,97 - 28,79) p = 0,0001] (Tabela 4). Como dito anteriormente no tópico 7.2, esta característica parece estar

associada com indivíduos cuja ancestralidade remetem a populações com predominância de olhos escuros, como pode ser observado no estudo de Edwards et al. (2015), que observou a presença desta característica em maior frequência em indivíduos do sul e leste asiático, quando comparados a indivíduos com ancestralidade europeia. No presente estudo o alelo "C" teve uma frequência total de 10,4% e cerca de 40% quando observado nos indivíduos apresentando esta característica (dados não mostrados). A frequência do alelo "C" é maior nas populações africanas do 1000Genomes (55,3%), seguido de populações do leste asiático (0,2%) e populações europeias (0,1%), o que é consistente com a possível associação desta característica com indivíduos apresentando olhos escuros (Tabela 4). É interessante ressaltar que, de todas as características de estruturas secundárias presentes na íris, nenhum marcador genético foi previamente associado a pigmentação na esclera ^{8,9,38}. De acordo com o SNPnexus e Haploreg o polimorfismo rs58358300A>C está localizado em uma região de íntron, a 148 nucleotídeos do sítio de splicing mais próximo, e marcação de histonas (H3K4me1) foram detectadas neste locus, sugerindo que este possa atuar como um enhancer, podendo indicar um possível efeito funcional, ainda que mais estudos sejam necessários para a verificação de tal suposição. É importante enfatizar que todas as análises aqui apresentadas foram realizadas controlando para os possíveis efeitos da ancestralidade. Ainda assim, não se pode descartar a possibilidade de que as associações aqui evidenciadas representem falso-positivos. Adicionalmente, embora o desequilíbrio de ligação tenha sido levado em consideração nas análises realizadas, buscando evitar associações redundantes com as variantes que foram identificadas no presente estudo como possivelmente causais, não se pode descartar a possibilidade de que futuramente haja outro entendimento quanto a esta questão, uma vez que a verdadeira variante causal pode não ter sido capturada neste estudo. A análise quantitativa para cor dos olhos através do uso de fotografias mostrou-se relativamente eficaz. Apesar das dificuldades encontradas devido à pouca qualidade de algumas das fotos utilizadas, a concordância entre as variáveis quantitativas e classificação categórica realizada por nosso laboratório foi considerada satisfatória uma vez que as classificações para olhos azuis e castanho sobrepuseram-se ao esperado no eixo de cor L*a*b* do espaço de cor CIELAB. Este foi o primeiro trabalho no Brasil utilizando uma metodologia de Sequenciamento de Nova Geração para analisar os genes OCA2 e HERC2 aos diversos fenótipos de pigmentação normal na população brasileira, confirmando o polimorfismo do gene HERC2 rs12913832 como o principal polimorfismo associado a cor dos olhos. Além disso, até onde sabemos, este foi o primeiro trabalho em que associou-se este polimorfismo com estruturas secundárias presentes na íris. O uso de variáveis quantitativas para os olhos revelou pela primeira vez, até onde sabemos, a associação do polimorfismo não sinônimo rs201872292 no gene HERC2 com olhos claros, independente do efeito do polimorfismo rs12913832. Este foi o primeiro estudo no Brasil que se propôs a analisar polimorfismos genéticos em genes candidatos a variação normal da pigmentação humana com estruturas secundárias presentes na íris. Destaca-se a associação do polimorfismo rs58358300 localizado em um íntron do gene HERC2 com pigmentação da esclera. Esta é a primeira vez que um polimorfismo é associado a esta característica, que apesar de ser a característica menos estudada do ponto de vista de associações com variantes genéticas, dentre todas as estruturas presentes na íris, estudos indicam sua importância clínica como um possível fator de risco para melanomas, ainda que muito raros ^{39,40}.

Agradecimentos

O presente trabalho teve apoio financeiro do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, #448242/2014-1), FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, #2013/15447-0). C.T.M.J. (309572/2014-2) e E.A.D. (304931/2014-1) contam com apoio financeiro colaborativo de CNPq. G.D. usufruiu de bolsa de doutorado PROEX concedida pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), e bolsa de doutorado sanduíche concedida pela CAPES, processo 88881.134344/2016-01.

REFERÊNCIAS BILIOGRÁFICAS

1. Nan, H., Kraft, P., Hunter, D. J. & Han, J. Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians. *International journal of cancer. Journal international du cancer* **125**, 909-917 (2009).

2. Hu, D. N., Wakamatsu, K., Ito, S. & McCormick, S. A. Comparison of eumelanin and pheomelanin content between cultured uveal melanoma cells and normal uveal melanocytes. *Melanoma Res* **19**, 75-79 (2009).

3. Prota, G., Hu, D. N., Vincensi, M. R., McCormick, S. A. & Napolitano, A. Characterization of melanins in human irides and cultured uveal melanocytes from eyes of different colors. *Exp Eye Res* **67**, 293-299 (1998).

4. Eiberg, H. et al. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Human genetics* **123**, 177-187 (2008).

5. Kayser, M. et al. Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene. *The American Journal of Human Genetics* **82**, 411-423 (2008).

6. Sturm, R. A. 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* **82**, 424-431 (2008).

7. Purtscher, E. On the development and morphology of iris crypts. Acta Ophthalmologica **43**, 109-119 (1965).

8. Edwards, M., Cha, D., Krithika, S., Johnson, M. & Parra, E. J. Analysis of iris surface features in populations of diverse ancestry. *R Soc Open Sci* **3**, 150424 (2016).

9. Larsson, M. et al. GWAS findings for human iris patterns: associations with variants in genes that influence normal neuronal pattern development. *Am J Hum Genet* **89**, 334-343 (2011).

10. Sturm, R. A. Molecular genetics of human pigmentation diversity. *Human Molecular Genetics* **18**, 9-17 (2009).

11. Phillips, C. et al. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet* **1**, 273-280 (2007).

12. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 16, 1215 (1988).

13. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, pp. 10-12 (2011).

14. Li, H., Handsaker, B., Wysoker, A., Fennell, T. & Ruan..., J. The sequence alignment/map format and SAMtools. ... (2009).

15. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491-498 (2011).

16. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).

17. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11.10.1-33 (2013).

18. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).

19. Holly, E. A., Aston, D. A., Char, D. H., Kristiansen, J. J. & Ahn, D. K. Uveal melanoma in relation to ultraviolet light exposure and host factors. *Cancer Research* (1990).

20. Horn, E. P., Hartge, P., Shields, J. A. & Tucker, M. A. Sunlight and Risk of Uveal Melanoma. *JNCI Journal* of the National Cancer Institute **86**, 1476-1478 (1994).

21. Andersen, J. D. et al. Importance of nonsynonymous OCA2 variants in human eye color prediction. *Mol Genet Genomic Med* **4**, 420-430 (2016).

22. Frudakis, T., Terravainen, T. & Thomas, M. Multilocus OCA2 genotypes specify human iris colors. *Human genetics* **122**, 311-326 (2007).

23. Sturm, R. A. & Larsson, M. Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res* **22**, 544-562 (2009).

24. Beleza, S. et al. Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genet* **9**, e1003372 (2013).

25. Bennett, D. C. & Lamoreux, M. L. The color loci of mice - a genetic century. *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society* **16**, 333-344 (2003).

26. Donnelly, M. P. et al. A global view of the OCA2-HERC2 region and pigmentation. *Human genetics* **131**, 683-696 (2012).

27. Rawofi, L. et al. Genome-wide association study of pigmentary traits (skin and iris color) in individuals of East Asian ancestry. *PeerJ* **5**, e3951 (2017).

28. Cerqueira, C. C. et al. Implications of the admixture process in skin color molecular assessment. *PLoS One* **9**, e96886 (2014).

29. Freire-Aradas, A. et al. Exploring iris colour prediction and ancestry inference in admixed populations of South America. *Forensic Sci Int Genet* **13**, 3-9 (2014).

30. Andrade, E. S., Fracasso, N. C. A., Strazza Júnior, P. S., Simões, A. L. & Mendes-Junior, C. T. Associations of OCA2-HERC2 SNPs and haplotypes with human pigmentation characteristics in the Brazilian population. *Leg Med (Tokyo)* **24**, 78-83 (2017).

31. Seddon, J. M., Sahagian, C. R., Glynn, R. J., Sperduto, R. D. & Gragoudas, E. S. Evaluation of an iris color classification system. The Eye Disorders Case-Control Study Group. *Invest Ophthalmol Vis Sci* **31**, 1592-1598 (1990).

32. Mengel-From, J., Børsting, C., Sanchez, J. J., Eiberg, H. & Morling, N. Human eye colour and HERC2, OCA2 and MATP. *Forensic Science International: Genetics* **4**, 323-328 (2010).

33. Pośpiech, E., Draus-Barini, J., Kupiec, T., Wojas-Pelc, A. & Branicki, W. Gene-gene interactions contribute to eye colour variation in humans. *Journal of human genetics* **56**, 447-455 (2011).

34. Rebbeck, T. R. et al. P gene as an inherited biomarker of human eye color. *Cancer Epidemiol Biomarkers Prev* **11**, 782-784 (2002).

35. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291 (2016).

36. Choi, Y. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. **the ACM Conference**, 414-417 (2012).

37. Cichorek, M., Wachulska, M., Stasiewicz, A. & Tymińska, A. Skin melanocytes: biology and development. *Postępy dermatologii i alergologii* **30**, 30-41 (2013).

38. Larsson, M. & Pedersen, N. L. Genetic correlations among texture characteristics in the human iris. *Mol Vis* **10**, 821-831 (2004).

39. Damato, B. & Coupland, S. E. Conjunctival melanoma and melanosis: a reappraisal of terminology, classification and staging. *Clin Exp Ophthalmol* **36**, 786-795 (2008).

40. Pinto-Proença, R. et al. Conjunctival melanoma: association of cyclooxygenase-2 tumor expression to prognosis. *Graefes Arch Clin Exp Ophthalmol* **256**, 989-995 (2018).

FIGURAS



Figura 1 - Distribuição dos indivíduos de acordo com a classificação da cor dos olhos. O número de indivíduos dentro de cada categoria de cor dos olhos está destacado entre parênteses.



Figura 2 - Distribuição dos indivíduos de acordo com a presença (Sim) ou ausência (Não) de estruturas secundárias presentes na íris dentro das categorias de cor dos olhos. **A:** Pontos de pigmentação na íris. **B:** Pigmentação na esclera. **C:** Sulcos de contração. **D:** Nódulos de Wolfflin. O número de indivíduos com a presença de cada estrutura secundária entre as categorias de cor do olho está destacado entre parênteses.



Figura 3 - Imagens fotográficas representativas das íris incluídas no estudo: olho azul, verde, mel, castanho-claro e castanho-escuro, respectivamente.



Figura 4 - Estruturas secundárias presentes nas íris e escleras identificadas no presente estudo. A: Pontos de pigmenação na íris. B: Presença de pigmentação na esclera. C: Nódulos de Wolfflin na periferia da íris. D: Sulcos de contração. As estruturas estão apontadas pelas flechas vermelhas.



Figura 5 - Distribuição das íris nas coordenadas do espaço de cores $L^*a^*b^*$. Cada círculo representa a íris direita de um indivíduo. A cor dentro de cada círculo equivale à média da cor da íris capturada pela porção extraída pelo algoritmo no espaço de cores $L^*a^*b^*$.



Figura 6 - Distribuição das íris nas coordenadas do espaço de cores L*a*b*. Cada círculo representa a íris direita de um indivíduo. A cor dentro de cada círculo equivale à heteroclassificação realizada pela equipe do LPFG, considerando cores azul, verde, mel, castanho-claro e castanho-escuro.



Figura 7 - Diferenciação das medianas do valor de b* entre os grupos heteroclassificados de acordo com a cor dos olhos. Valor de *p* global do teste de Kruskal-Wallis no topo do gráfico. Linha pontilhada demarca a mediana de b* da amostra total. ns = não significativo, p > 0.05; ** = p < 0.01; **** = p < 0.001.





Sulcos de Contração



С







Figura 8 - Composição ancestral dentro dos grupos que apresentaram ou não as características de estruturas secundárias presentes na íris. **A:** Pigmentação na esclera. **B:** Sulcos de contração. **C:** Pontos de pigmentação na íris. **D:** Nódulos de Wolfflin.



Figura 9 - Distribuição dos genótipos do polimorfismo rs12913832 plotados entre os valores a* e b* do espaço de cores CIELAB.

TABELAS

Tabela 1 - Proporção de cada população atribuída a cada um dos três *clusters* definidos pelo programa STRUCTURE.

	Agrupamentos inferidos					
População	Europeu	Asiático/Ameríndio	Africano			
1 (AFR)	0,004	0,002	0,994			
2 (EUR)	0,981	0,010	0,008			
3 (EAS)	0,003	0,996	0,001			
4 (RBP)	0,714	0,090	0,196			

EUR = Populações europeias do *1000Genomes*; AFR = Populações africanas do *1000Genomes*; EAS = Populações do leste asiático do *1000Genomes*; RPB = Amostras do presente estudo coletadas em Ribeirão Preto - SP.

Tabela 2 - Distribuição dos polimorfismos nas regiões dos genes *OCA2* e *HERC2*, cujos alelos encontram-se associados cor dos olhos e estruturas secundárias presentes na íris

	OCA2	HERC2	Total
Promotora	-	-	-
5'UTR	-	-	-
CDS	3	3	6
Íntrons	1	3	4
3'UTR	-	1	1
Total	5	7	11

Associação		rsID (posição)	A1	Odds Ratio [IC 95%]	р	RBP	EUR	AFR	EAS
Olho azul									
	vs. Olho não azul	rs12913832 (28365618)	G	8,328 [3,47 - 19,95]	$1,97 \times 10^{-6}$	0,417	0,866	0,005	0,002
	vs. Olho mel	rs12913832 (28365618)	А	0,088 [0,02 - 0,32]	2,0x10 ⁻⁴	0,582	0,134	0,995	0.998
	<i>vs.</i> Olho castanho claro	rs12913832 (28365618)	А	0,023 [0,00 - 0,08]	2,44x10 ⁻⁸	0,582	0,134	0,995	0,998
	<i>vs.</i> Olho castanho escuro	rs1129038 (28356859)	Т	17,96 [5,75 - 56,05]	6,53x10 ⁻⁷	0,420	0,863	0,005	0,001
Olho verde									
	vs. Olho mel	rs12913832 (28365618)	А	0,206 [0,08 - 0,51]	0,0006	0,582	0,134	0,995	0,998
	<i>vs.</i> Olho castanho claro	rs12913832 (28365618)	А	0,077 [0,03 - 0,18]	1,11x10 ⁻⁸	0,582	0,134	0,995	0,998
		<u>rs12913832</u> (28365618)	А	0,035 [0,02 - 0,06]	3,91x10 ⁻⁹	0,582	0,134	0,995	0,998
	<i>vs.</i> Olho castanho escuro	rs61756152 (28412872)*	А	4,497 [1,33 - 15,21]	0,0155	0,109	0,055	0,091	0,295
		rs1800407 (28230318)*	Т	10,38 [1,50 - 71,68]	0,0176	0,051	0,047	0,000	0,001
Olho mel									
	<i>vs.</i> Olho castanho claro	rs12910433 (28228644)	G	5,949 [1,98 - 17,81]	0,0014	0,497	0,247	0,906	0,846
		<u>rs1129038</u> (28356859)	Т	7,641 [2,52 - 23,09]	0,0003	0,420	0,863	0,005	0,001
	<i>vs.</i> Olho castanho escuro	rs1800407 (28230318)*	Т	8,291 [1,68 - 40,88]	0,0093	0,051	0,047	0,000	0,001
		rs1800419 (28096538)*	А	2,976 [1,17 - 7,54]	0,0215	0,491	0,639	0,376	0,560
Olho castanho claro									
		<u>rs12913832</u> (28365618)	G	2,798 [1,47 - 5,30]	0,0016	0,417	0,866	0,005	0,002
	<i>vs.</i> Olho castanho escuro	rs201872292 (28475625)*	А	2,944 [1,41 - 6,11]	0,0037	0,096	0,055	0,041	0,294
		rs1800401 (28260053)*	А	0,377 [0,15 - 0,93]	0,0356	0,082	0,037	0,124	0,005
Olho castanho escuro									
		<u>rs12913832</u> (28365618)	G	0,182 [0,11 - 0,30]	$2,63 \times 10^{-11}$	0,417	0,866	0,005	0,002
		rs201872292 (28475625)*	А	0,356 [0,17 - 0,71]	0,0038	0,096	0,055	0,041	0,294
	<i>v</i> s. Olho não castanho escuro	rs1800401 (28260053)*	А	2,921 [1,33 - 6,41]	0,0075	0,082	0,037	0,124	0,005
		rs74005645 (28447282)*	Т	3,720 [1,40 - 9,86]	0,0083	0,120	0,001	0,536	0,005
		rs1800419 (28096538)*	А	0,597 [0,36 - 0,97]	0,0403	0,491	0,639	0,376	0,560

Tabela 3 - Associações genéticas encontradas envolvendo alelos de nove polimorfismos analisados nos genes OCA2 e HERC2 com a pigmentação dos olhos.

Sublinhado = Polimorfismo condicionado para verificação de sinais independentes; * = Polimorfismo associado após análise condicional ao polimorfismo com maior associação (menor valor de p); rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; A1 = Alelo de menor frequência (MAF); RBP = Frequência alélica na população de estudo; EUR = Frequência alélica nas populações europeias do 1000Genomes; AFR = Frequência alélica nas populações africanas do 1000Genomes; EAS = Frequência alélica nas populações do leste asiático do 1000Genomes.

Tabela 4 - Associações genéticas encontradas envolvendo alelos de oito polimorfismos analisados nos genes OCA2 e HERC2 com características de estruturas secundárias presentes na íris.

Associação	rsID (posição)	A1	Odds Ratio [IC 95%]	р	RBP	EUR	AFR	EAS
Pontos de pigmentação na íris	rs12913832 (28365618)	G	2,406 [1,46 - 3,96]	0,0005	0,417	0,866	0,005	0,002
Nódulos de Wolfflin	rs12913832 (28365618)	G	3,661 [1,66 - 7,20]	0,0001	0,417	0,866	0,005	0,002
Pigmentação na esclera	rs58358300 (28419393)	С	9,247 [2,97 - 28,79]	0,0001	0,104	0,001	0,553	0,002
Sulcos de contração	rs12913832 (28365618)	G	0,477 [0,30 - 0,75]	0,0013	0,417	0,866	0,005	0,002

Sublinhado = Polimorfismo condicionado para verificação de sinais independentes; * = Polimorfismo associado após análise condicional ao polimorfismo com maior associação (menor valor de *p*); rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; A1 = Alelo de menor frequência (MAF); RBP = Frequência alélica na população de estudo; EUR = Frequência alélica nas populações europeias do *1000Genomes*; AFR = Frequência alélica nas populações africanas do *1000Genomes*; EAS = Frequência alélica nas populações do leste asiático do *1000Genomes*.

Tabela 5 - Associações genéticas encontradas envolvendo alelos de três polimorfismos analisados nos genes OCA2 e HERC2 com as dimensões L*a*b* do espaço de cores CIELAB.

Associação	rsID (posição)	A1	Beta [IC 95%]	p	RBP	EUR	AFR	EAS
T+	<u>rs12913832</u> (28365618)	G	3,745 [2,18 - 5,30]	4,67x10 ⁻⁶	0,417	0,866	0,005	0,002
L*	rs6497271 (28365431)*	А	-5,498 [-8,352,65]	1,96x10 ⁻⁴	0,066	0,037	0,277	0,005
a*	rs12913832 (28365618)	G	-2,607 [-3,811,40]	3,39x10 ⁻⁵	0,417	0,060	0,005	0,020
1.4	<u>rs12913832</u> (28365618)	G	-5,517 [-6,804,22]	6,46x10 ⁻¹⁵	0,417	0,866	0,005	0,002
)^	rs201872292 (28475625)*	А	3,950 [1,61 - 6,28]	0,0010	0,086	0,018	0,041	0,294

Sublinhado = Polimorfismo condicionado para verificação de sinais independentes; * = Polimorfismo associado após análise condicional ao polimorfismo com maior associação (menor valor de *p*); rsID = Nome de identificação do polimorfismo no banco de dados dbSNP; A1 = Alelo de menor frequência (MAF); RBP = Frequência alélica na população de estudo; EUR = Frequência alélica nas populações europeias do *1000Genomes*; AFR = Frequência alélica nas populações africanas do *1000Genomes*; EAS = Frequência alélica nas populações do leste asiático do *1000Genomes*.